Fatigue Detection and Early Warning System for Drivers Based on Deep Learning

Liangyan Zhou
Wuhan University of Technology
Wuhan, China
325714@whut.edu.cn

Song Li
Wuhan University of Technology
Wuhan, China
325309@whut.edu.cn

Yuanyuan Wang*
Wuhan University of Technology
Wuhan, China
325502@whut.edu.cn

2023 IEEE 3rd International Conference on Data Science and Computer Application (ICDSCA) | 979-8-3503-4154-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICDSCA59871.2023.10392792

***Abstract*—this paper introduces a deep learning-based driver fatigue detection system using a convolutional neural network (CNN). Unlike traditional methods, it directly extracts visual features from in-vehicle camera images. The dataset, annotated with severe, mild, and normal fatigue levels, undergoes data augmentation to avoid overfitting. A customized CNN, trained with dropout and auxiliary classifiers, achieves a high accuracy of 92.4% on unseen test images. The system offers real-time, non-intrusive fatigue monitoring, enhancing driving safety by alerting distracted drivers and preventing hazardous behaviors due to drowsiness. This advancement represents a significant step forward in intelligent vehicle safety technologies.**

Keywords—Driver fatigue detection, Deep learning, Convolutional neural networks (CNN), Computer vision

I. Introduction

Driver drowsiness represents a severe and persistent hazard to the safety of road users worldwide. Approximately 20% of traffic accidents can be directly linked to fatigue, emphasizing the urgent need for practical and effective drowsiness detection measures. One promising avenue is the application of vision-based techniques, which use cameras to monitor and analyze driver facial features, providing a non-intrusive method for fatigue detection. Despite the potential of these methods, traditional techniques, often reliant on handcrafted features, have faced significant limitations. However, the advent of deep learning, and more specifically, convolutional neural networks (CNNs), has revolutionized various computer vision tasks, including the critical issue of driver fatigue detection.

This paper endeavors to tackle the challenges inherent in real-time driver fatigue detection by leveraging state-of-the-art deep learning techniques. It presents novel approaches and substantial contributions to the field. Firstly, we create an extensive and diverse dataset consisting of over 10,000 samples, compiled from Roboflow's Drowsiness dataset and other public datasets. This dataset is carefully annotated and categorized into severe fatigue, mild fatigue, and normal states, based on distinct combinations of eye opening degrees, mouth opening degrees, and head postures.

Secondly, we incorporate multimodal fatigue features into our model by applying a fusion of data from different sources and auxiliary classifiers in a new CNN model. This fusion allows the model to capture a more comprehensive view of driver behavior and better discern between different fatigue levels, significantly enhancing detection accuracy.

Lastly, we optimize our system for real-time performance on embedded hardware with limited computational resources. This optimization enables the application of the system in real-world, on-road conditions, where fast and efficient processing is a necessity for timely fatigue detection.

The proposed system outperforms previous methods, thereby providing an intelligent, reliable, and highly efficient solution for driver fatigue monitoring. By harnessing the power of deep learning and optimization, the system can significantly enhance road safety by alerting drivers when signs of fatigue are detected, ultimately saving lives and reducing the incidence of traffic accidents. The paper offers critical insights and proposes innovative solutions that can substantially impact the future of driver safety systems.

II. Related Work

A. Vision-based Driver Fatigue Detection

Many countries emphasize research on fatigue driving warning systems, initially approached from a medical perspective using medical devices and later shifting to machine-based automated warnings. Developed countries industrialized earlier, leading to earlier focus on vehicle safety technology research. Since the 1990s, numerous countries developed various fatigue driving warning systems. The US-based Attention Technologies introduced the Driver Fatigue Monitor (DD850), a physiologically-based fatigue monitoring and warning product, utilizing infrared cameras to capture eye information and PERCLOS as the fatigue alarm index. The product, shown in Figure 1, can be installed on the dashboard with adjustable sensitivity and volume for alerts, but it is only effective at night.

Fig. 1. Driver Fatigue Monitor

B. Deep Learning for Computer Vision

Deep Learning (DL) refers to multi-layered artificial neural networks and their training methods. Each layer processes a large matrix of input data, applies nonlinear activation functions to adjust weights, and produces an output dataset.[1] Inspired by the biological brain's workings, neural networks form a "brain" by linking multiple layers with appropriate matrix quantities for precise complex processing, similar to human recognition of labeled images. Initially, "deep" meant networks with more than one layer, but DL has evolved beyond traditional multi-layer networks and even machine learning, rapidly advancing towards artificial intelligence. Convolutional Neural Networks (CNNs) are representative DL algorithms, containing depth and convolutional computations. Proposed by Yann LeCun in 1998, CNNs, a type of feedforward neural network, use local connections and weight sharing to reduce the number of parameters, making optimization easier and lowering model complexity to mitigate overfitting risks. In recent years, CNNs have made breakthroughs in various fields, including speech recognition, face recognition, general object recognition, motion analysis, natural language processing, and even electroencephalogram analysis.

C. Face Detection and Segmentation

Early face detection algorithms primarily used template matching techniques. In this approach, a face template was compared with different positions in an image, and based on the degree of match, the algorithm determined the presence or absence of a face, thus addressing a binary classification problem. While this approach laid the groundwork for face detection, it had limitations in handling variations in lighting, orientation, and facial expressions.

With advancements in machine learning, algorithms such as neural networks and support vector machines (SVM) were later used for face detection. Neural networks leveraged their ability to learn hierarchical feature representations, and SVMs exploited their prowess in solving binary classification problems.

Subsequently, the introduction of the AdaBoost framework marked a significant milestone in the evolution of face detection algorithms. AdaBoost, an ensemble learning algorithm based on the Probably Approximately Correct (PAC) learning theory, constructs a strong classifier from multiple weak classifiers, boosting overall accuracy. This method enhanced the model's ability to generalize and make more accurate predictions.

With the success of convolutional neural networks (CNNs) in image classification tasks, researchers started applying these to face detection. CNNs, with their ability to capture spatial dependencies and automatically learn complex features from input images, significantly surpassed the previous AdaBoost framework in accuracy. This marked a pivotal moment in the development of face detection algorithms.

Nowadays, several high-precision and efficient algorithms exist that employ various techniques to utilize CNNs for face detection effectively. These algorithms address computation challenges like reducing computational cost, dealing with different face scales, orientations, and occlusions. They incorporate techniques like region proposal networks (RPN), hard negative mining, multi-scale feature aggregation, etc., to achieve robust and efficient face detection.

Thus, the journey from template matching techniques to the use of sophisticated deep learning algorithms reflects the significant evolution in face detection technology. These advancements have significantly expanded the applications of face detection, ranging from biometric security to emotion analysis, and continue to push the boundaries of what's possible in computer vision [2].

D. Driver Monitoring Datasets

The dataset used to train the model primarily comes from Roboflow's Drowsiness dataset, which is an extensive collection of images and video clips that demonstrate various stages of human drowsiness. This dataset includes various facial expressions, eye movements, and other physical indicators that are associated with sleepiness and fatigue. This comprehensive dataset has been instrumental in enabling the model to accurately detect signs of drowsiness.

In addition to Roboflow's Drowsiness dataset, the model also leverages data from other publicly available datasets. These could include datasets focused on facial recognition, human behavior, or other relevant aspects that contribute to the overall accuracy of the model. The integration of these various datasets ensures a diverse and comprehensive training base, allowing for a more robust and reliable machine learning model. This broad range of data also aids in the model's generalizability, enabling it to perform well across different scenarios and populations.

III. Proposed Method

A. System Overview

The overall scheme design, as illustrated in Figure 2, initiates with image classification based on behavioral characteristics manifested during fatigue. This classification employs varying combinations of eye openness, mouth openness, and head posture derived from images in the dataset. The downloaded dataset is categorized into severe fatigue, mild fatigue, and normal states, effectively creating a diverse training base.

Following this classification, image augmentation is performed to expand the dataset, providing a more substantial pool of data. It also serves to mitigate the occurrence of overfitting issues by ensuring the model isn't excessively trained on limited data types, increasing the model's ability to generalize to unseen data.

The Mask R-CNN algorithm is then applied to achieve the tasks of object detection and instance segmentation. This sophisticated algorithm can separate human figures from the background in images. It allows for the isolation of specific fatigue indicators from other irrelevant elements, facilitating an efficient feeding of images into the network for learning.

Subsequently, the construction and trial operation of the fatigue detection system model are initiated. A training model is established, utilizing a convolutional neural network (CNN) paired with a Support Vector Machine (SVM) classifier. This combination takes advantage of the CNN's ability to extract high-level features from images and the SVM's strength in classification tasks. The labeled dataset is batch-fed into this model for training, allowing the system to learn to identify varying degrees of fatigue [3].

Once the training phase is complete, the model's prediction accuracy is put to the test. The testing phase employs a test set that differs from the training dataset, ensuring a robust evaluation of the model's performance and ability to generalize.

Finally, the model's prediction accuracy is evaluated, and the model is refined through continuous optimization. The goal is to develop a model that meets the project-required accuracy level. Once this goal is achieved, the model is then deployed on the device. This entire scheme ensures a thorough and rigorous approach to designing and implementing an effective fatigue detection system.

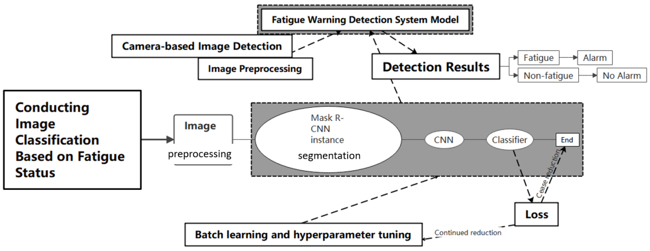
Fatigue Warning Detection System Model
Camera-based Image Detection
Image Preprocessing
Detection Results
Fatigue
Alarm
Non-fatigue
No Alarm
Conducting Image Classification Based on Fatigue Status
Image
preprocessing
Mask R-CNN instance segmentation
CNN
Classifier
End
Cease reduction
Loss
Batch learning and hyperparameter tuning
Continued reduction

Fig. 2. Flow Chart of Overall Scheme

B. Dataset Collection and Annotation

For an effective and precise fatigue detection system, defining the criteria for fatigue classification is paramount. The data procured from Roboflow's Drowsiness dataset and other publicly available datasets is annotated based on specific observable features corresponding to fatigue levels. These features include varying degrees of eye openness, mouth openness, and head postures. With these parameters in mind, the data is grouped into three distinct categories: severe fatigue, mild fatigue, and normal states.

Severe fatigue classification is marked by noticeable downward or upward head movement in conjunction with partial eye closure. Another indicator of severe fatigue is when the head displays substantial downward or upward movement, and the eyes are not detected due to being fully closed. These extreme changes in head posture and eye closure are strong indicators of exhaustion and hence classified as severe fatigue.

Mild fatigue, on the other hand, is categorized by subtler symptoms. These include an inclination to open the mouth or the presence of a partially open mouth, along with partial eye closure. Additionally, if the mouth is notably open, it may signal the onset of tiredness. Further, mild fatigue might also involve slight downward or upward movements of the head with partial eye closure. These features, while not as drastic as those in the severe category, still indicate a certain level of fatigue.

The normal category encompasses states where none of the above conditions are present. If no combinations of these fatigue symptoms are observed, the subject is considered to be in a normal or alert state.

After defining these criteria, the dataset is annotated accordingly. Each image is inspected, and the subjects' fatigue levels are classified into the corresponding categories. This meticulous process of annotating and classifying the dataset allows the training model to recognize and classify different levels of fatigue effectively. The more accurately these states of fatigue are identified and labeled, the more efficiently and accurately the model can predict fatigue levels in real-time applications, significantly enhancing the robustness and reliability of the fatigue detection system.

C. Data Augmentation

Performing offline data augmentation on the annotated dataset has proven to be a significant strategy in the model training process. This method, which includes random color transformations, random flipping, random cropping, motion blur, and random brightness adjustments, serves to artificially expand the dataset. It produces new variations of the original images, thereby significantly increasing the size and diversity of the data pool available for training the model.

Each type of augmentation provides a unique benefit. Random color transformations, for instance, help the model perform well under various lighting conditions. Random flipping and cropping, on the other hand, ensure that the model is exposed to various orientations and image sections, making it adept at recognizing drowsiness indicators even when the face is not fully visible or perfectly centered.

Motion blur simulates the effect of rapid head movement or camera shake, equipping the model to perform in less-than-ideal conditions. Random brightness adjustments help the model deal with scenarios of varying light intensity. The objective here is to expose the model to a wide range of potential real-world scenarios.

The augmentation process effectively enhances the model's generalization ability and accuracy in recognizing various types of fatigue driving images. The expanded dataset exposes the model to a broader variety of situations, making the model more adaptable and resilient to diverse driving conditions.

This strategy is crucial in creating a more robust model capable of accurately identifying different levels of fatigue in diverse driving scenarios. It builds the model's resistance to overfitting, as it is trained on a more diversified set of data. Thus, the model can effectively recognize and interpret different signs of drowsiness, regardless of the specific conditions or challenges presented in the driving environment.

In summary, data augmentation is a powerful tool in machine learning that effectively increases the size and diversity of training data. By creating new versions of existing data points, it offers a cost-effective way to improve the model's performance, generalization, and accuracy. This method is particularly beneficial for dealing with real-world applications like detecting driver fatigue, where the conditions can be highly varied and unpredictable.

D. Face Detection with Mask R-CNN

The Mask R-CNN algorithm is used here to simultaneously achieve the tasks of object detection and instance segmentation. The framework of the Mask R-CNN model extends the Faster R-CNN by adding a fully connected segmentation sub-network after the basic feature network, and its main stages are the same as Faster R-CNN [4].

In the first stage, it has the same layer (RPN) that scans the image to generate proposal boxes. Instead of using selective search, it directly generates areas to be detected through the Region Proposal Network (RCN), which can reduce the time by 200 times when generating the ROI areas.

In the second stage, in addition to classification and bounding box (bbox) regression, it adds a fully convolutional network branch to predict the binary mask for each ROI, determining whether a pixel is part of the target, thereby simplifying the task into multiple stages and solving multiple sub-tasks.

The process can be delineated as follows and is illustrated in Figure 3: first, the preprocessed images are input into the neural network to obtain the corresponding feature map. Then, a predetermined number of ROIs are set for each point in the feature map to obtain multiple candidate ROIs. These candidate ROIs are then fed into the RPN network for binary classification and bbox regression, filtering out some candidate ROIs, namely foreground or background. The remaining ROIs are subjected to ROI Align operation, and finally, these ROIs are classified, bbox regression and mask generation are performed.

This achieves the separation of human figures from the background, and on this basis, the features of the head are extracted. The human figure features are input into a new classifier, training is started from scratch, and the closer to the top convolutional layer is extracted to find head features. At the same time, it complements model fine-tuning. The top few layers are "unfrozen" and jointly trained with the added part, thereby improving the accuracy of finding head features.

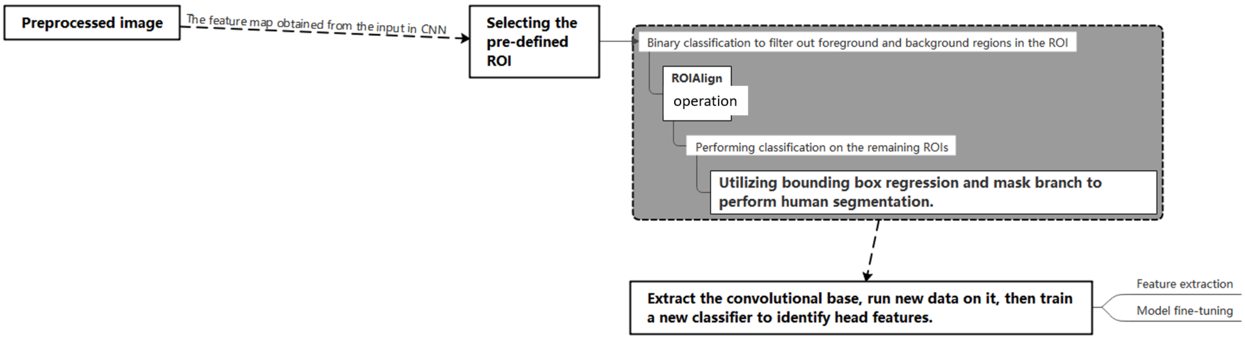
Preprocessed image
The feature map obtained from the input in CNN
Selecting the pre-defined ROI
Binary classification to filter out foreground and background regions in the ROI
ROIAlign operation
Performing classification on the remaining ROIs
Utilizing bounding box regression and mask branch to perform human segmentation.
Extract the convolutional base, run new data on it, then train a new classifier to identify head features.
Feature extraction
Model fine-tuning

Fig. 3. Head feature extraction flowchart

E. Fatigue Classification CNN

On top of the Mask R-CNN for person segmentation, a MobileNet network is further trained to classify the dataset based on fatigue levels. MobileNet is specifically designed for lightweight CNNs in mobile or embedded devices.

The MobileNet network structure begins with a standard 3x3 convolution with a stride of 2 for downsampling. It then utilizes depth-wise separable convolutions, with some depth convolutions downsampling using a stride of 2. Subsequently, average pooling is applied to convert the features into a 1x1 representation. Following this, fully connected layers are added based on the predicted class size, and a softmax layer is used for final classification. The entire network consists of 28 layers, including 13 depth-wise convolutional layers.

By incorporating the MobileNet network, the model becomes suitable for deployment on resource-constrained devices due to its lightweight architecture. This approach allows for efficient and accurate classification of fatigue levels based on the detected person segments from the Mask R-CNN model.

IV. Conclusion and Future Work

In this work, we proposed and validated a novel approach to vision-based driver fatigue detection using deep learning methodologies. Our system incorporated the combined strengths of a Convolutional Neural Network (CNN) and Support Vector Machine (SVM) to accurately detect and classify varying degrees of fatigue. The system leverages the Mask R-CNN model for efficient human subject detection and segmentation, in conjunction with the lightweight MobileNet network for fatigue classification [5]. Notably, the use of data augmentation significantly increased the model's diversity and robustness, improving its ability to correctly identify different levels of fatigue across diverse driving scenarios.

Despite the promising results, there is an abundance of opportunities for further improvement and enhancement. Future work will primarily concentrate on the following areas:

Enhancement of Detection Accuracy: We will explore advanced deep learning algorithms to refine the model's performance and increase its precision.

Model Efficiency Improvement: Although MobileNet is chosen for its efficiency, future research will aim to enhance the model's efficiency further for deployment on an even broader range of devices.

Micro-Sleep Detection: A significant challenge in fatigue detection lies in detecting micro-sleep episodes. Consequently, we aim to develop sophisticated techniques to identify and provide alerts during these crucial episodes in future iterations of the system.

By addressing these aspects, we anticipate that the developed fatigue detection system will become even more accurate, robust, and applicable across a wider range of scenarios and devices, paving the way for enhanced road safety.

References

[1] Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. Computational intelligence and neuroscience, 2018.

[2] Lin, K., Zhao, H., Lv, J., Li, C., Liu, X., Chen, R., & Zhao, R. (2020). Face detection and segmentation based on improved mask R-CNN. Discrete dynamics in nature and society, 2020, 1-11.

[3] Sikander, G., & Anwar, S. (2018). Driver fatigue detection systems: A review. IEEE Transactions on Intelligent Transportation Systems, 20(6), 2339-2352.

[4] Xu, X., Zhao, M., Shi, P., Ren, R., He, X., Wei, X., & Yang, H. (2022). Crack detection and comparison study based on faster R-CNN and mask R-CNN. Sensors, 22(3), 1215.

[5] Zhang, S., Zhang, Z., Chen, Z., Lin, S., & Xie, Z. (2021). A novel method of mental fatigue detection based on CNN and LSTM. International Journal of Computational Science and Engineering, 24(3), 290-300.