A Comparative Analysis of Driver Fatigue Detection Techniques Using Facial Feature Analysis and Advanced Neural Networks

Bingyu Guo, 3772483, Data and Computer Science
Yufan Feng, 4732306, Data and Computer Science
Tian Tan, 4732950, Data and Computer Science
Yaojie Wang, 3771056, Data and Computer Science

Project Source Code and Materials: https://github.com/shirahanesuoh-mayuri/CV-3D-Project

Abstract

Driving fatigue is a significant factor contributing to traffic accidents globally. Effective detection and timely warnings can substantially mitigate this risk, enhancing road safety. This study explores the application of computer vision technologies for the detection of driver fatigue, comparing two primary methods: facial feature analysis and neural network-based detection using Convolutional Neural Networks (CNN) and the You Only Look Once (YOLO) models. Facial landmarks and aspect ratios such as Eye Aspect Ratio (EAR) and Mouth Aspect Ratio (MAR) were analyzed to assess fatigue symptoms. Our comparative analysis includes evaluations based on accuracy, precision, recall, and F1 scores across the models. Results indicate that the YOLOv5 model, despite its simplicity compared to newer models like YOLOv8, performed with perfect accuracy, showcasing its potential for real-time application in diverse driving environments. The study highlights the effectiveness of using advanced machine learning models and facial feature analysis in identifying drowsiness, potentially influencing future improvements in driver assistance systems.

1. Introduction

Inspired by the prevalent issue of driving fatigue, which is a significant contributor to traffic accidents worldwide, this project addresses the urgent need for effective preventive measures. Fatigue impairs a driver's reaction time, alertness, and decision-making abilities, thus increasing the likelihood of accidents. Therefore, timely identification and intervention are crucial to enhance road safety.

The advent of computer vision has introduced a range of practical applications, particularly in monitoring and enhancing driver safety. Among these, driver fatigue detection through facial recognition technology represents a significant leap. By analyzing a driver's facial features and expressions in real time, it is possible to detect signs of fatigue effectively. This technological approach eliminates the need for invasive methods and provides a non-disruptive solution to monitor alertness levels continuously.

While existing technologies have made strides in fatigue detection, challenges such as accuracy in diverse lighting conditions, non-intrusiveness, and real-time processing capabilities persist. Our research aims to address these issues by employing and comparing two principal methods: facial feature analysis and advanced neural network-based detection using CNN and YOLO models. Specifically, the research questions we address include: How effectively can facial feature analysis versus neural network-based models detect driver fatigue under varying conditions? What improvements in detection accuracy and speed can be achieved with the latest YOLO models in real-world scenarios?

By improving the accuracy and reliability of fatigue detection systems, this study aims to contribute significantly to vehicular safety. The practical applications of this research extend beyond personal vehicles to include commercial and public transportation, potentially saving lives and reducing accidents on a global scale.

Therefore, this report is organized as follows: Section 2 reviews related work, highlighting the advancements and limitations of current fatigue detection technologies. Section 3 describes the methods employed, including the specifics of the facial landmark detection and neural network configurations. Section 4 details the experimental setup and datasets used, followed by Section 5, which presents the results and comparative analysis of the different methods. Finally, Section 6 concludes the paper with a discussion of the findings and potential areas for future research.

2. Related work

In the domain of driver fatigue detection, various methodologies have been explored to enhance accuracy and reliability. Traditional methods often rely on physiological signals such as heart rate or eye closure (PERCLOS) but require intrusive sensors. With the advent of deep learning, convolution neural networks (CNNs) have been widely adopted due to their ability to process spatial hierarchies in images, making them ideal for facial feature analysis involved in drowsiness detection.

Recently, attention has shifted towards more sophisticated architectures like YOLOv5 and YOLOv8, which are part of the You Only Look Once (YOLO) family of models. These models are highly regarded for their speed and accuracy in real-time object detection. YOLOv5, being an earlier version, offers a balance between speed and accuracy, whereas YOLOv8 provides advanced features with improved detection capabilities, especially in challenging lighting conditions which are common in driving environments.

In evaluating the performance of fatigue detection models, four key metrics are used: Accuracy, Precision, Recall, and F1 Score. Accuracy measures the overall correctness of the model across all classes, providing a general indication of performance. Precision assesses the model's ability to label as positive a sample that is truly positive, thus reflecting the exactness of the model. Recall, or Sensitivity, indicates the model's capability to find all actual positives, measuring the model's thoroughness. Lastly, the F1 Score is the harmonic mean of Precision and Recall, offering a balance between Precision and Recall in cases where an unequal class distribution might exist. These metrics together provide a comprehensive evaluation framework that helps in understanding each model's strengths and weaknesses in the context of driver drowsiness detection.

"Dlib" is a modern C++ toolkit developed by Davis King. It provides various machine learning algorithms and tools for real-world applications, with a primary focus on computer vision tasks such as face detection. Dlib is known for its high performance, efficiency.

Dlib contains a pre-trained model for predicting 68 facial landmarks. The landmarks include points around the eyes, eyebrows, nose, mouth and jawline.

Our project compares these models by evaluating not just their accuracy, but also precision, recall, and F1 scores. These metrics provide a comprehensive view of model performance, particularly in how well the models identify true drowsy incidents without misclassified awake states. This comparative analysis is crucial as it highlights the strengths and limitations of each method, informing future improvements in fatigue detection systems.

3. Methods

3.1. Direct Method based on Facial Landmarks Detection

3.1.1 Face detection and Landmark Localization using dlib

In facial recognition tasks, we often use the 68 facial landmarks for facial recognition method to detect specific key points on a person's face after face is detected. This is typically done using a shape predictor model, such as the one provided by the dlib library, which has been trained on a large dataset of facial images annotated with these landmarks. The detected landmarks are usually represented as (x, y) coordinates in the image. These points mark significant areas of the face, such as the corners of the eyes, the tip and sides of the nose, the contour of the lips, and the outline of the jaw.

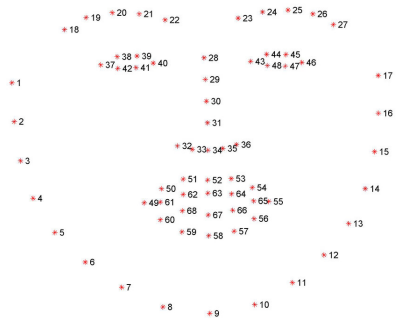

Figure 1: The 68 facial landmarks[1]

3.1.2 The calculation of EAR and MAR

EAR(Eye Aspect Ratio) is a commonly used metric for the assessment of ocular closure. It is calculated by the ratio of vertical and horizontal distances between several key points of the eye.

$$EAR = \frac{\|p_{39} - p_{38}\| + \|p_{41} - p_{42}\|}{2 \times \|p_{40} - p_{37}\|} \tag{1}$$

where, p_37, p_39,...,p_42 are left eye landmarks shown in Figure 1(the calculation for right eye is similar). Similarly, MAR(Mouth Aspect Ratio) is used to assess the degree of mouth opening and is calculated by the ratio of vertical and horizontal distances between key points of the mouth.

$$EAR = \frac{\|p_{57} - p_{53}\| + \|p_{59} - p_{51}\|}{2 \times \|p_{55} - p_{49}\|} \tag{2}$$

where, p_{49}, p_{51},...,p_{57} are mouth landmarks shown in Figure 1.

3.1.3 Fatigue Detection and Assessment

Fatigue is detected by calculating the aspect ratios of the eyes and mouth using their facial landmarks. For each detected face, the EAR and MAR were calculated and used to determine whether fatigue was detected. If either met the conditions for fatigue (EAR below 0.15 or MAR above 0.1), it was determined as ”yes”, otherwise ”no”[2].

3.2. CNN with EAR and MAR

This method is based on the result of 68 facial landmarks which can be decided by the pre-trained model ’shape predictor 68 face landmarks.dat.bz2’ [4] in ’dlib’ [5] package to get these point. Then based on the definition of EAR and MAR, we can calculate EAR and MAR with 68 landmarks.

Because EAR and MAR exhibit significant characteristics in distinguishing drunk driving, and because in our project’s dataset, fatigue driving and non-fatigue driving are a simple binary classification problem, we aim to attempt classification using a trained and saved CNN model.

We plotted the distribution of EAR for the left eye and the right eye and MAR in drowsy and non-drowsy situations. Observing them from various angles, the results indicate the presence of a plane that can effectively separate the two distributions.

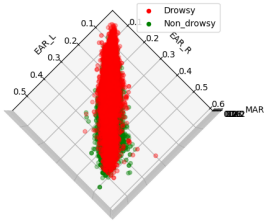

Figure 2: Distribution of EAR,MAR

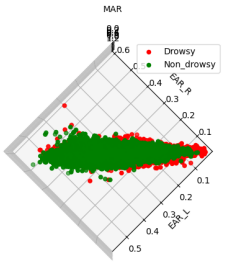

Figure 3: Distribution of EAR,MAR

In our approach, the neural network architecture employed consists of two layers of one-dimensional convolution followed by four fully connected layers. Additionally, a dropout layer is inserted within the fully connected layers.

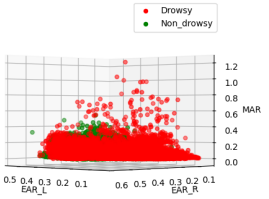

Figure 4: Distribution of EAR,MAR

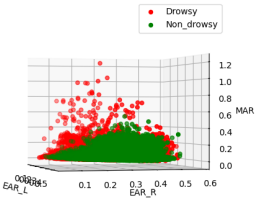

Figure 5: Distribution of EAR,MAR

3.3. YOLOv5

YOLOv5, an iteration in the You Only Look Once series, stands out for its efficiency and accuracy in object detection tasks. In our project, YOLOv5 is utilized to detect signs of fatigue in drivers by analyzing video frames. The model processes images to detect facial features that are indicative of fatigue, such as closed eyes and yawning. [3]

For our drowsiness detection application, YOLOv5 has been trained on a dataset comprising various driver behaviors captured under different lighting conditions. This training helps the model to effectively distinguish between normal and drowsy states. Evaluation metrics such as accuracy, precision, recall, and F1 score have been employed to validate the effectiveness of YOLOv5 in recognizing fatigue with high reliability.

3.4. YOLOv8

In addition to making judgments directly using facial feature vectors, we use YOLOv8 [6] to train face images to determine whether the person in the input image is fatigued. Yolov8 is the latest generation of the You Only Look Once target detection model, which was developed by the Ultralytics research team in 2022. Compared to previous Yolo versions, Yolov8 offers significant improvements in speed and accuracy.

We use YOLOv8 for image classification tasks. It uses a variant of the EfficientNet architecture to perform classification. Yolov8 provides pre-trained models trained on the ImageNet dataset. We fine-tuned the model based on this by adjusting hyperparameters such as learning rate, batch size, and number of iterations to optimize the model for new tasks.

4. Experiments

4.1. Datasets

We use two datasets Driver Drowsiness Dataset (DDD) and Drowsiness Prediction Dataset as our training datasets for the models.

For the test dataset, We choose the Annotated Dataset Sub1 from Dataset-D3S[2]. Because we only focus on dividing the "Drowsy" and "Non-Drowsy", our test data just include these two labels, we combine the subsets 'Yawn' and 'Eye-closed' as "Drowsy", "Natural" as "Non-Drowsy".

4.2. The architecture of CNN model

Our CNN architecture is composed of two convolutional layers that incrementally scale the input dimensions from a 3x1 vector to 64x1 and subsequently to 128x1. Following the expansion, the data is flattened to facilitate the transition through four fully connected layers. Across these layers, we employ the ReLU activation function to introduce non-linearity, enhancing the model's ability to learn complex patterns. At the output stage, the softmax activation function is utilized to normalize the output, providing a probabilistic interpretation suitable for classification tasks. To mitigate the risk of overfitting, we incorporate dropout regularization with a dropout rate of 0.5 between the first and second fully connected layers.

For input data preparation, we use the 'dlib' pre-trained model to extract EAR (Eye Aspect Ratio) and MAR (Mouth Aspect Ratio) values. These are formatted into a 3x1 vector comprising EAR values from both eyes and the MAR value, serving as features to train our CNN model. The training dataset is divided with a test size configuration of 30% and a fixed random state of 42 to ensure reproducibility. We optimize our model over 20 epochs with a batch size of 32, and upon completion, the model's parameters are saved in a '.pth' file for subsequent use.

In the testing phase, EAR and MAR values are once again derived using 'dlib' and stored as ndarray files. We then load these features, along with the saved model parameters, into our CNN framework to perform the classification task. The accuracy of the model is calculated to assess performance, thereby providing insights into the effectiveness of our facial feature-based approach to detect driver fatigue.

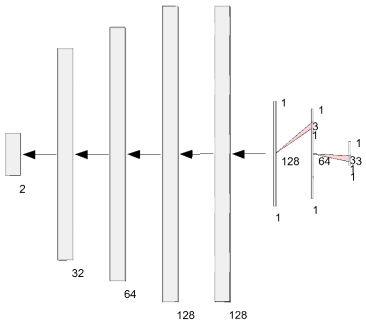

Figure 6: The Architecture of CNN Model

4.3. Model Training Details

YOLOv5 In this project, the YOLOv5 model, specifically the YOLOv5m-cls variant optimized for classification tasks, was employed to detect signs of driver drowsiness using a combined dataset. The model was pre-trained and further fine-tuned using the Adam optimizer with a batch size of 32. Initial learning rates were set to 0.001, adjusted dynamically based on training progress. Training involved 50 epochs with extensive data augmentation techniques such as random scaling, cropping, and flipping to improve generalization across varied driving conditions. The Cross-Entropy loss function was used to enhance classification accuracy by penalizing misclassifications effectively.

YOLOv8 Since the dataset is large enough, we do not augment the data. We use yolov8n-cls.pt as our pre-trained model. We use Adam optimizer and batches of size 32. We train the model for 50 epochs on CPU because of the limited computational resources. YOLOv8 has the function of image pre-processing, so we just specify the image size as 256 in the parameter.

5. Results

5.1. Performance of Yolov5 and Yolov8

5.1.1 Yolov5

In this section, we present the training results of the YOLOv5 model, as implemented in our driver fatigue detection project. The YOLOv5 model underwent a total of 50 epochs of training, which is depicted in Figures 7 and 8, showing the training accuracy and loss respectively. Although the full training spanned 50 epochs, the most significant changes occurred within the first 10 epochs, hence the graphs are zoomed in on this range for clarity. As illustrated in Figure 7, the training accuracy of the YOLOv5 model exhibits a sharp increase in the initial epochs, reaching stability shortly thereafter. This rapid improvement demonstrates the model's quick adaptation to the drowsiness de-

tection task. Post the initial surge, the accuracy levels off, indicating that the model is consistently recognizing drowsy and alert states with high reliability. Figure 8 presents the training loss of our YOLOv5 model. Starting at higher values, there is a noticeable decrease in loss within the first few epochs, which signifies a rapid learning phase. Following this phase, the loss continues to decrease at a slower, more gradual pace. The steady decrease suggests that the model is refining its parameters and improving its predictive capabilities throughout the training period. These figures collectively indicate that the YOLOv5 model achieved a robust understanding of the task within the initial epochs, with further gains in precision occurring at a more incremental rate as training progressed. The results affirm the model's capacity to generalize from the training data effectively, which is essential for reliable real-time drowsiness detection in drivers.

Figure 7: Training Accuracy of YOLOv5

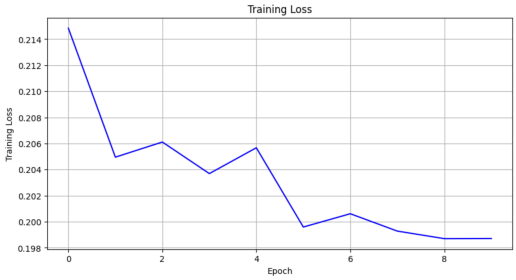

Figure 8: Training Loss of YOLOv5

5.1.2 Yolov8

Based on the description of the yolov8 model parameters in the previous section Experiments, we trained our dataset. The training results are presented in the form of a line graph. In the training accuracy graph, as shown in Figure 9, the accuracy increases rapidly in the first 10 epochs. After 10 epochs, the accuracy increases slowly and stabilizes. Finally, the accuracy reaches about 0.97. It shows that the model is able to classify fatigue images well now. As shown in Figure 10, overall, the training loss values decrease at a linear rate, with the loss decreasing from around 0.3 to below 0.10, indicating that the model's predictions are becoming more accurate. As shown in Figure 11, the validation loss plot initially drops dramatically from 0.63 to 0.35, which suggests that the model quickly begins to generalize to unseen data. After about 10 epochs, the validation loss levels off and slowly increases after the 37th epoch, indicating that overfitting is starting to occur, i.e., the model is starting to learn patterns specific to the training data that do not generalize well to new data.

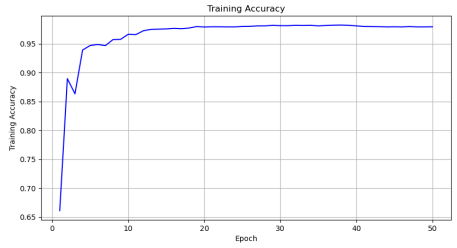

Figure 9: Training Accuracy of YOLOv8

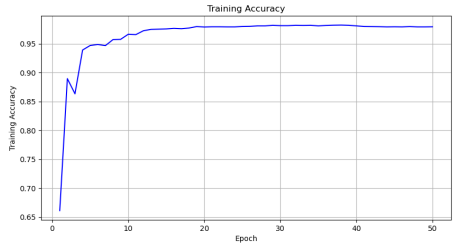

Figure 10: Training Loss of YOLOv8

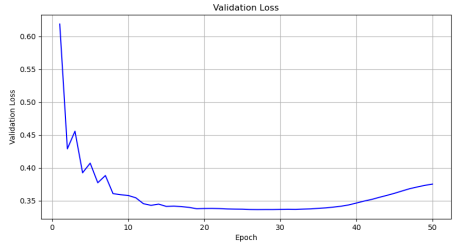

Figure 11: Validation Loss of YOLOv8

5.2. Results on different methods

The performance of these methods is evaluated based on several key metrics: Accuracy, Precision, Recall, and F1 Score. YOLOv8 emerges as the most reliable model, combining high accuracy with real-time processing capabilities. The findings emphasize the importance of model selection based on specific operational contexts and the need

for ongoing evaluation against practical deployment scenarios. This analysis not only informs future improvements in drowsiness detection systems but also guides the selection of appropriate detection technologies based on specific requirements and constraints.

Method	Accuracy	Precision	Recall	F1 score
Direct	0.97	0.95	1.0	0.97
CNN	0.85	0.82	0.95	0.88
YOLOv5	1.00	1.00	1.00	1.00
YOLOv8	0.88	0.95	0.83	0.88

Table 1: Results on different methods.

6. Conclusion

In conclusion, our study demonstrates the Direct Method as the optimal approach for driver fatigue detection when considering both efficiency and reliability. While the neural network-based YOLOv8 model showed promising results, with commendable accuracy and robustness in various testing conditions, it ultimately did not surpass the Direct Method. Its sophisticated architecture, while powerful, requires careful calibration and validation to prevent overfitting and to ensure it can be reliably applied in diverse real-world scenarios.

Nonetheless, YOLOv5's perfect metrics on the test set raise questions about the representativeness of our dataset and the need for further examination to confirm the model's true generalizability. Such exceptional performance could imply a risk of overestimating the model's capabilities and overlooking potential limitations.

Here's some findings when we processing data with 'dlib', we found that 'dlib' can not figure all the facial landmarks of images, because the number of original data in drowsy is 26908 while the processed data with landmarks is 23034, and the number of original data in Non drowsy is 24005 while the processed data with landmarks is 21964.

Moving forward, our research will delve into enhancing the versatility of YOLOv8, addressing the overfitting challenges, and exploring more robust data processing methods. We will also continue to improve the precision of 'dlib' in landmark detection and expand our dataset for a more comprehensive analysis. These steps will contribute to the development of a more dependable and accurate fatigue detection system, suitable for the complexities of everyday driving.

References

[1] Giuseppe Amato, Fabrizio Falchi, Claudio Gennaro, and Claudio Vairo. A comparison of face verification with facial landmarks and deep features. 2018. 2

[2] Isha Gupta, Novesh Garg, Apoorva Aggarwal, Nitin Nepalia, and Bindu Verma. Real-time driver's drowsiness monitoring based on dynamically varying threshold. In *2018 Eleventh International Conference on Contemporary Computing (IC3)*, pages 1–6. IEEE, 2018. 3, 4

[3] Glenn Jocher and Ultralytics. Yolov5. https://github.com/ultralytics/yolov5, 2020. GitHub repository. 3

[4] Vahid Kazemi and Josephine Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1867–1874, 2014. 3

[5] Davis E King. Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, 10:1755–1758, 2009. 3

[6] Ultralytics. Yolov8: Models, inference & training. https://github.com/ultralytics/ultralytics, 2023. 3