WEP Warwick Evans Publishing

Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Research on Human-computer Interaction and Emotion Recognition based on Convolutional Neural Networks

Yuhai Yang

College of Jinjiang, Sichuan University, Meishan, China

yangyuhai@ldy.edu.rs

Abstract. Amidst the swift advancement of intelligent vehicles, computer vision technology has become an increasingly important core technology in the field of autonomous driving, especially in providing safety guarantees for drivers. Given the common Incidence of traffic accidents resulting from driver fatigue. The use of deep learning technology to prevent driver fatigue behavior has become an indispensable protective application in intelligent cockpit human-computer interaction. This article introduces an algorithm designed to detect drowsy driving, which relies on Convolutional neuronal networks are employed to address the challenge of fatigue during driving. The core of the research lies in constructing a model for detecting fatigue depending on YOLOv5. This model uses UTA-RLDD and YawDD datasets and extracts refined image features from video frames as available input data for the model. It can have a better detection effect on the blinking and yawning characteristics of the driver's eyelids and lips, thereby effectively determining whether they are in a fatigue state. Compared with deep learning fatigue detection algorithms, experiments have shown that its average accuracy for eye closure and yawning reaches 97.3% and 91.7%, respectively, which can effectively detect driver fatigue, prevent drivers from entering a fatigue driving state, and improve driving safety. Future research will continue to optimize and expand the model to adapt to more complex driving scenarios and make greater contributions to the development of intelligent vehicles.

Keywords: Human-computer Interaction, Fatigue Check, Convolutional Neural Network.

1. Introduction

In contemporary society, traffic safety issues are increasingly receiving widespread attention. Based on statistics provided by the National Statistics Agency, nearly 200000 traffic accidents occur every year, of which about 20% are caused by fatigue driving [1, 2]. From June to mid July 2022, 17.5% of the major road traffic accidents with more than three fatalities nationwide were suspected of fatigue driving accidents. If a warning is issued to the driver in a timely manner before a traffic accident occurs, it is possible to avoid nearly 90% of traffic accidents [3]. Consequently, to guarantee the safety of drivers and passengers and minimize the frequency of traffic accidents, fatigue driving detection system is crucial in the intelligent cockpit. At present, there are four types of research on fatigue driving detection both domestically and internationally: fatigue detection based on physiological data; Detecting vehicle behavior characteristics; Fatigue detection through computer vision; Fatigue detection through multiple features [4].

In the method of vehicle behavior characteristics, due to the expensive price of the instruments used, the small market size, and the problem of equipment aging leading to a decrease in fatigue detection accuracy over time, the promotion of this method is difficult. However, methods based on driver physiological characteristics require sensors to be attached to the driver's skin for invasive feature signal collection, which can cause interference to the driver's operation. Although this method can detect in real-time and has high accuracy, it may bring a certain degree of discomfort to the driver [5]. With the advancement of technology and the rise of computer vision technology, the method of analyzing driver characteristics through deep learning algorithms has the advantages of high accuracy and low cost.

At present, deep learning algorithms are typically grouped into two primary categories: two-stage algorithms, which involve multiple stages of processing, and single-stage detection algorithms, which perform detection in a single, streamlined process. Among them, despite having high accuracy, the two-stage object detection algorithm, its detection speed is rather sluggish and the computational workload is large.Meeting the real-time requirement presents a challenge, thus hindering effective fatigue driving detection. Relatively speaking, single-stage detection algorithms have more advantages, with fast detection speed and reduced computational resource consumption by directly predicting the bounding boxes and categories of target objects. Among the widely utilized single-stage object detection algorithms include YOLO series, SSD [6], EfficientDet [7], RetinaNet [8], etc. However, compared to two-phase detection algorithms, single- phase detection algorithms may have slight shortcomings in accuracy. Based on the above, this article chooses the YOLOv5 model single-stage detection serving as the framework for detecting fatigue driving of driver's eyes and mouth features [9]. This model, as one of the representatives of single-stage detection algorithms, can ensure a certain accuracy while having a fast detection speed, thus meeting the real-time requirements in actual driving scenarios.

2. Research Methods

2.1. The Dataset Used in the Research

The fatigue driving dataset used in the study was extracted from the UTA-RLDD and YawDD datasets, containing 16246 images that were filtered and preprocessed to train machine learning models that can accurately identify signs of fatigue driving. The dataset comprises of images divided into three distinct parts: firstly, the training set, which encompasses 13719 images; secondly, the validation set, containing 1380 images; and finally, the test set, made up of 1147 images. This distribution guarantees that the model can be comprehensively trained on diverse datasets, ensuring its effective validation and rigorous testing on independent datasets.

This dataset includes images of drivers of different genders, ages, and races under various lighting conditions, as well as various driving behaviors such as yawning, eye closure, expressionless expression, and eye opening. These category labels not only reflect key signs of fatigue, but also include images under normal driving conditions. The diversity of data can improve the generalization ability of fatigue driving detection models.

Use data augmentation techniques in preprocessed images to enhance the robustness of the model. Preprocessing increases the horizontal flipping probability of some images by 50%, enabling the model to better recognize driver behavior in the mirror image. Preprocessing enhances the 90 degree rotation of some images to enable the model to adapt to different head postures in different directions to further bolster the resilience of the model; Simulate visual interference and image quality changes in the real world through random cropping and Gaussian blurring of images. Finally, all images will be uniformly adjusted to 640x640 pixels to ensure consistency of input data (see Figure 1).

This study divided the dataset into four categories: Yawn, close (closed eyes), noYawn, and open (open eyes). Due to the small sample size of key fatigue signs such as "Yawn" and "close", the sample size of the open eye category is the smallest, while the sample size of the "normal" category without fatigue signs is the largest. During the model training process, increasing the weight of a few categories enhances the sensitivity of the model to signs of fatigue.The robustness of the model can be strengthened through various methods like using data augmentation techniques in preprocessed images. This includes a 50% probability of horizontal flipping, enabling the model to recognize driver behavior in the mirror image; 90 degree rotation enhancement allows the model to adapt to different head postures in different directions; And random cropping and Gaussian blur to simulate visual interference and image quality changes in the real world. All images were ultimately uniformly adjusted to 640x640 pixels, ensuring consistency of input data.

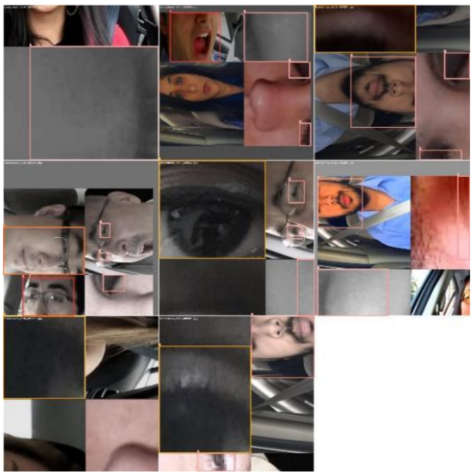

Figure 1. Preprocessed images (Photo credited: Original)

2.2. YOLOv5s Model

There are YOLOv5s YOLOv5 detection network, YOLOv5m, YOLOv5l, YOLOv5x four kinds of models. Among them, YOLOv5s, as shown in Figure 2, which is the network with the least depth and narrowest feature map width in the YOLOv5 series. On this basis, the three subsequent networks undergo continuous deepening and widening processes.The example figure illustrates the composition of the Yolov5 network structure, which is comprised of Input, Backbone, Neck, and Prediction. The input end of the network is constituted by the input part of YOLOv5. Mosaic data enhancement technique is adopted to randomly cut the incoming data and then splicing it, which can alleviate the lack of elements in the training set or enhance the recognition ability. The backbone comprises the Yolov5 feature extraction of the network. The capability of feature extraction has a direct impact on the overall performance of the network. The Focus structure and the CSP structure constitute the main components of the backbone layer in Yolov5. Focus structure can realize subsampling and feature compression of input feature graph. The CSP structure is divided into two parts by the input feature map, a segment is processed by a miniature convolutional network, whereas the second segment undergoes direct processing by the following layer. Subsequently, the two maps of features are merged together to serve as input for the next layer.The Neck part mainly adopts two different structures: SPP and PAN; The Prediction part mainly consists of prediction boxes, and each prediction box is composed of confidence score, class probabilities and bounding box coordinates.

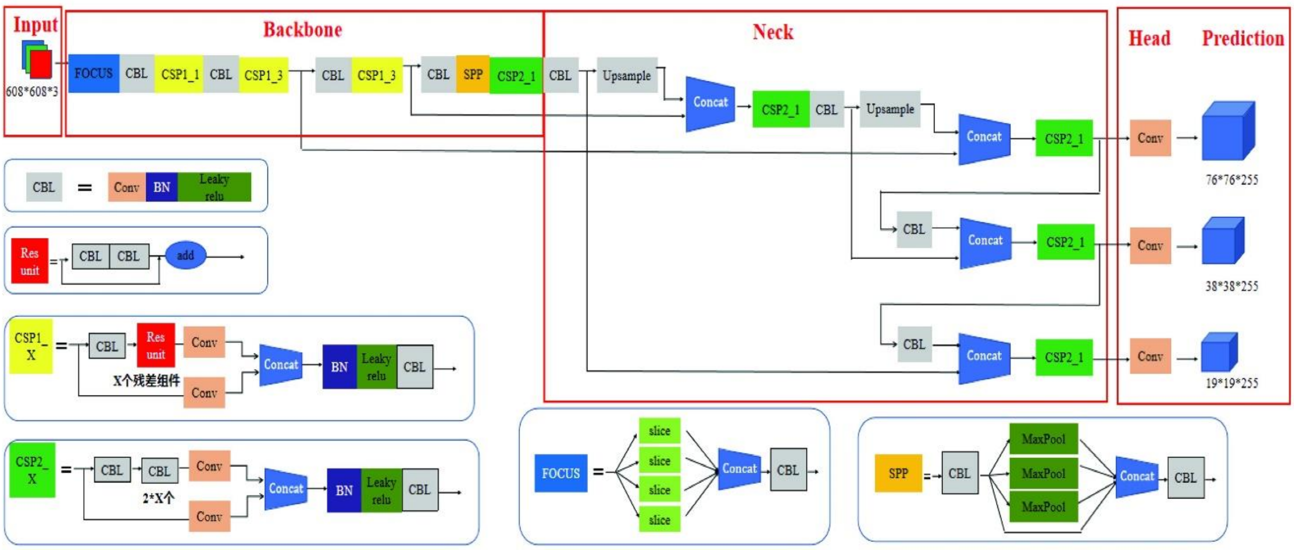

Figure 2. Network structure of YOLOv5s model (Photo credited: Original)

2.3. Model Training

During the network training process, due to the large amount of data, the study divided the data into 120 batches for experimentation. Each batch contains a certain number of images. Train a network model with appropriate errors by updating model parameters through the forward and backward propagation processes of the network. In the experimental stage, the experimental model recognizes images and selects the best model based on the obtained results. The CPU utilized for this experiment is the AMD Ryzen 7 6800H, equipped with Radeon Graphics technology, whereas the GPU is the NVIDIA GeForce RTX 3060 Laptop edition, 6GB of graphics memory, 16GB of running memory, operating system is Ubuntu 18.04, acceleration environment is CUDA12.0, programming language is Pvthon3.10, and deep learning frameworks are Python 1.13.1 and tensorflow 2.15. And, Figure 3 showed the program process diagram.

```
100%|██████████| 755k/755k [01:01<00:00, 12.5kB/s]
                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100%|██████████| 87/87 [00:06<00:00, 13.00it/s]
                   all       1380       1618      0.461      0.824      0.663      0.434
                  Yawn       1380        116      0.688          1      0.973      0.683
                 close       1380        895      0.482      0.978      0.917      0.585
               noYawn       1380         93       0.23      0.978      0.307       0.24
                  open       1380        514      0.445      0.341      0.453      0.229
```

Figure 3. Program process diagram. (Photo credited: Original)

2.4. Evaluating indicator

The indexes used to evaluate the training model are usually Precision, Recall and Average Precision across Categories. Indexes provide a comprehensive assessment of the performance of target detection, with accuracy being defined as the proportion of data correctly classified.

According to the loss function of Yolov5 training, during the process, there are three distinct types of losses that arise: rectangular frame loss, also known as box_loss, confidence loss, commonly referred to as obj_loss, and classification loss, designated as cls_loss. Below in Figure 4, the training curve of the experimental model is exhibited.

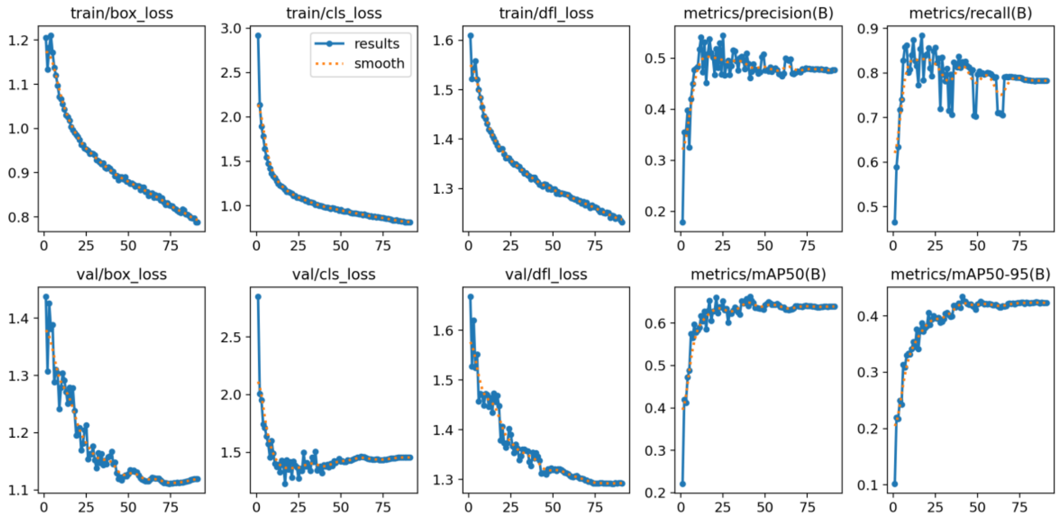

Figure 4. Evaluating indicator(Photo credited: Original)

2.5. Experimental Results

The PR Curve curves on the validation sets of YawDD and UTA-RLDD are shown in the Figure 5. In the PR Curve, recall is represented on the horizontal axis, while accuracy is depicted on the vertical axis. As the PR Curve approaches the topmost right quadrant, it indicates the model can guarantee both high precision and high retrieval capability during prediction, indicating that the prediction results are more accurate. With mAP@0.5 /%As an evaluation metric, the proposed algorithm ultimately achieved 97.3% and 91.7% on the validation sets of y and u, respectively, while noYawn and close only had 30.7% and 45.3%.

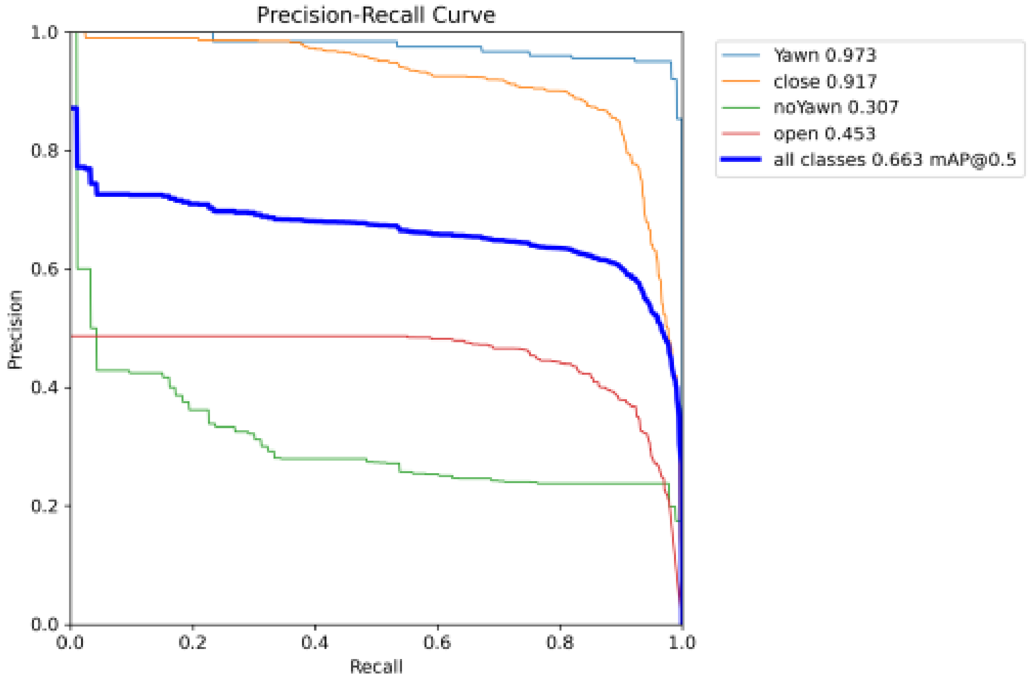

Figure 5. Precision-recall curve(Photo credited: Original)

3. Discussion

Fatigue driving is an increasingly serious traffic safety issue, and the use of deep learning algorithms for automatic detection has emerged as a preeminent research focus. This article uses the lightweight YOLOv5 algorithm to enhance data randomness through image re stitching,enhance the precision and resilience of the model, and achieve remarkable results. However, when delving deeper into this field, several aspects still need to be addressed.

Firstly, although the YOLOv5 algorithm has shown good performance in fatigue driving detection, there is still room for further optimization. For example, introducing more modules that can enhance the algorithm's feature detection capability will help Enhance the precision and retrieval capabilities of the model, the model may exhibit varying degrees of adaptability to the fatigue characteristics of different environments and drivers. Therefore, improving the model's generalization ability is also serves as a crucial avenue for future exploration and research endeavors.

Secondly, the intensity and duration of fatigue behavior are crucial for assessing the driver's fatigue status. Although the current model can recognize obvious fatigue features such as yawning and eye closure, it may not be sensitive enough to mild fatigue or the early stages of fatigue. Therefore, how to further strengthen the capture and analysis of these subtle features holds the key to enhancing the model's detection ability.

In addition, the real-time performance and stability of the model are also important factors that need to be considered in practical applications. Although the YOLOv5 algorithm has a high detection speed, ensuring the stable operation of the model and providing real-time feedback on detection results in complex traffic environments remains a challenge.

Finally, it is imperative that we prioritize the comprehensibility and explainability of the model. Deep learning models often have high black box characteristics, and their decision-making process is difficult to fully understand. In fatigue driving detection, this may lead to a lack of sufficient trust in the output results of the model. Therefore, how to enhance the interpretability of the model through visualization or other methods, and improve the trust of users in model decision-making, is represents a pivotal avenue for future exploration and scientific investigation.

In summary, although the current fatigue driving detection model based on YOLOv5 has achieved certain results, there are still many aspects that need further in-depth research and exploration. We look forward to more innovative solutions in the future to provide stronger support for the practical application of fatigue driving warning systems.

4. Conclusion

This article uses the lightweight YOLOv5 algorithm to construct a fatigue driving detection model, which enriches data diversity through image re stitching and effectively improves the accuracy and robustness of the model. Experimental data shows that the mAP of the model in identifying fatigue features such as yawning and eye closure is as high as 91.3% and 97.3%, respectively, indicating its high recognition accuracy. The study adopts a small volume deep learning model to achieve rapid detection of eye opening and yawning, providing an effective means to determine the fatigue status of drivers, and has certain practical application value.

However, there are still shortcomings in current research, especially the lack of further optimization and improvement of the model. In future work, we will introduce modules that can enhance the algorithm's ability to detect features and deeply optimize the model to further improve its detection accuracy and recall rate. At the same time, we will also strengthen the assessment of the intensity and duration of fatigue behavior, and improve the model's detection ability by adding further analysis of the model's detection results, in order to better apply it to fatigue driving warning systems.

References

[1] D.H Li, Q. Liu, W Yuan et al. Relationship between fatigue driving and traffic accident[J]. Journal of Traffic and Transportation Engineering, 10(2), (2010) 104-109.

[2] X.W Zhang, F.S Wang, H.Y Yang. Fatigue driving detection algorithm based on multi feature information fusion [J/OL]. Journal of Chongqing University of Technology (Natural Science Edition): (2024) 1-11.

[3] R Zhang, T.J Zhu, Z L Zou, et al. A review of research on driver fatigue driving detection methods [J]. Computer Engineering and Applications, 58(21), (2022): 53-66.

[4] Chen L, Xin G, Liu Y, et al. Driver fatigue detection based on facial key points and LSTM[J]. Security and Communication Networks, (2021): 1-9.

[5] W Liu, D Anguelov, D. Erhan, RHAN, et al. Ssd: Single shot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, (2016) 21-37.

[6] M Tan, R Pang, Q.V Le. Efficientdet: Scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2020): 10781-10790.

[7] T.Y Lin, P, Goyal, R. Girshick et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. (2017): 2980-2988.

[8] J. Redmon, S. Divvala, R. Girshick, et al. You only look once: unified, real-time object detection [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). LLas Vegas:IEEE, (2016): 779-788.

[9] Z.H Zheng, J Zhao, Y Li. Research on detecting bearing-cover defects based on improved YOLOv3[J]. IEEE Access, 9 (2021): 10304-10315.