Proceeding of the 18th INFATS & 27th SAE-C/STC, 2024, pp1-10
2024.402

Risk Identification of Distracted Driving Behavior Based on Transfer Learning

Wenbin HE[1], Shuai YANG [1], Zhaohui LIU[2], Wuyi MING[1,*]

[1]*Henan Provincial Key Laboratory of Intelligent Manufacturing of Mechanical Equipment, Zhengzhou University of Light Industry, Zhengzhou, China, 450002*

[2] *Zhengzhou Yutong Bus Co.,Ltd., Zhengzhou, China , 450048*

Email: *hwb@zzuli.edu.cn*, *1603282065@qq.com*, *liuzhc@yutong.com*, *mingwuyi@gmail.com*

Abstract: Driving safety is key to ensuring the safety and health of urban transportation systems. With the continuous advancement of intelligent driving assistance systems, using computer vision and deep learning to detect distracted driving behaviors has become a popular area of research. This study proposes a YOLOv8-CW(YOLOv8-CBAM-WIoU) model framework that integrates YOLOv8, a CBAM (convolutional block attention module), and a loss function module (WIoU), aimed at identifying driver distraction behaviors while driving. This study trains the improved YOLOv8-CW model using the collected distraction dataset, which includes ten types of driver distraction behaviors captured by a camera at a resolution of 3840×2160 pixels. The training results show that the YOLOv8-CW model achieved an average detection precision of 86.1% for distracted driving behavior, which is an improvement of 1.1% compared to the original YOLOv8 model. In addition, we validated the trained model on the public dataset StateFarm(state farm distracted driver detection), achieving an average accuracy of 98.5%. This indicates that the proposed YOLOv8-CW framework has outstanding recognition performance and can effectively identify driver distraction behaviors.

Keywords: Driving safety; Deep learning; Distracted driving behavior

1 Introduction

According to the "Global Status Report on Road Safety 2023", the number of deaths caused by traffic accidents in recent years is 1.19 million [1]. Based on this situation, we need to find a series of effective measures and implement them to reduce traffic accidents and improve driving safety. In traffic accident investigations in the United States, distracted driving by the driver is a major cause of traffic accidents, accounting for 68.3% [2] of accidents and 9.7% [3, 4, 5] of fatal crashes. This investigation provides us with a clear approach to reducing traffic accidents and enhancing driving safety: effectively identifying distracted driving behaviors and preventing them [6]. Therefore, effectively identifying distracted driving behaviors and preventing them is crucial for enhancing driving safety [7].

Distracted driving behavior refers to the driving actions of a driver that differ from normal driving behavior, indicating a lack of focus and resulting in diminished driving ability. Traditional machine learning methods often require manual design and extraction of features, such as HOG (histogram of oriented gradients) and SIFT (scale-invariant feature transform) [8, 9, 10], which may necessitate the knowledge and experience of domain experts. Moreover, the extracted features may not be sufficient to fully represent the complexity of the data[11]. Deep learning is a machine learning approach rooted in artificial neural networks, centered around the concept of modeling and analyzing complex data through hierarchical feature learning and abstraction. Peng et al. [12] introduced a deep learning network that combines a TSD-DLN (time-space dual line network) with a C-AOG (causal and-or graph) for recognizing distracted behaviors. The TSD-DLN merges attention features derived from dynamic optical flow information with spatial features extracted from individual video frames to accurately identify distracted driving postures. Furthermore, the integration of causal knowledge through the C-AOG enhances the robustness of recognition. Shuyan et al. [13]

proposed an effective recognition method utilizing CNNs (convolutional neural networks) and transfer learning. They initially built a CNN model for driving behavior recognition, which achieved an accuracy of only 0.64. However, by fine-tuning the parameters of a pre-trained CNN model with transfer learning techniques, they improved the accuracy to 0.80, reflecting a 16% increase over the original model. Furkan et al. [14] developed a distracted driving detection system aimed at identifying such behaviors. They constructed a vision-based CNN model using transfer learning and fine-tuning methods, while also integrating sensor and image data to create a LSTM-RNN (long short-term memory recurrent neural network) model. This combination of different technologies significantly enhanced the overall performance of their detection system.

Through the study of the aforementioned distracted driving behaviors, it can be seen that distracted driving is diverse in the research of deep learning. As the research on deep learning technology advances, object detection algorithms have become an important research focus. Currently, target detection algorithms utilizing convolutional neural networks can be broadly categorized into two groups. The first category comprises two-stage detection methods, such as RCNN (region-based convolutional neural networks) and Faster R-CNN. The second category includes single-stage detection methods like SSD (single shot multiBox detector) and the YOLO series. The YOLO algorithm approaches the object detection task as a regression problem, which allows for significantly enhanced real-time detection performance. As a result, it has emerged as a prominent focus in the realm of object detection research. This study proposes a deep convolutional neural network based on YOLOv8, called YOLOv8-CW, which utilizes transfer learning algorithms to fine-tune the parameters of the YOLOv8 model for identifying distracted driving behaviors of drivers. First, the CBAM attention mechanism module is added to the backbone network of YOLOv8[15]. CBAM is a lightweight attention mechanism that consists of two components: the channel attention module and the spatial attention module. These components calculate attention weights for both channel and spatial dimensions, thereby improving the feature extraction capability related to distracted driving behavior and increasing the model's efficiency and accuracy. Additionally, the Weighted Intersection over Union (WIoU) provides a more precise evaluation of the similarity between predicted bounding boxes and actual targets, particularly in scenarios involving multiple object parts, which enhances detection accuracy. Consequently, the CIoU loss function in the YOLOv8 head has been substituted with the WIoU loss function[16].

2 Analysis of the Principles of YOLOv8-CW

YOLO is one of the most widely used object detection models, and YOLOv8 is a lightweight model released by ultralytics in 2023. It includes five structures: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x, using the parameters depth_multiple and width_multiple to control these structures. The model consists of an input layer, a backbone, a neck, and a head. It first performs convolution through the backbone to extract important features from the samples, then conducts feature fusion in the neck, and finally utilizes the extracted features for recognition in the head. Although YOLOv8 performs well in terms of detection accuracy and speed, its effectiveness during the convolution process is average, particularly in detecting various types of distracted driving behaviors, where it falls slightly short. Therefore, the original YOLOv8 model has been improved to enhance detection accuracy without sacrificing detection speed: (1) Incorporate the CBAM attention mechanism into Backbone to enhance feature extraction capabilities. (2) The improved YOLOv8-CW model using the attention-based loss function WIoU based on bounding box regression (BBR) is shown in Figure 1.

Recent studies have explored the integration of the CBAM into the YOLOv8 backbone network to enhance real-time detection capabilities. CBAM combines spatial and channel attention mechanisms, allowing for adaptive feature optimization by generating attention maps in both dimensions. These maps are then multiplied with the input feature maps, improving detection performance without significantly burdening the network. The lightweight nature of CBAM facilitates seamless integration into various CNN architectures, enabling end-to-end training. This enhancement effectively improves YOLOv8's detection accuracy and speed, making it better suited for real-time applications [17].

The loss function is crucial in deep learning, measuring the difference between a model's predictions and actual labels. In the context of object detection, traditional Intersection over Union (IoU) assesses the overlap between predicted and real bounding boxes. YOLOv8 employs CIoU as its default loss function, with GIoU and DIoU also available. CIoU enhances traditional IoU by considering the centroid distance, aspect ratio differences, and area variations between the predicted and actual boxes. However, these loss functions can be computationally intensive and may struggle with low-quality data samples, such as those affected by background noise and disharmonized longitudinal width ratio. WIoU [18] introduces a weight parameter to adjust the intersection-parallel ratio based on task requirements, offering greater flexibility for various scenarios. WIoU calculates IoU, then weights the contribution of each box component using the minimum weight and normalizes these values with the maximum weight from both boxes. This approach allows WIoU to more accurately assess similarity between predicted frames and real targets, particularly when multiple object parts are involved, enhancing detection accuracy. Consequently, this study replaces CIoU with WIoU to improve the model's performance.

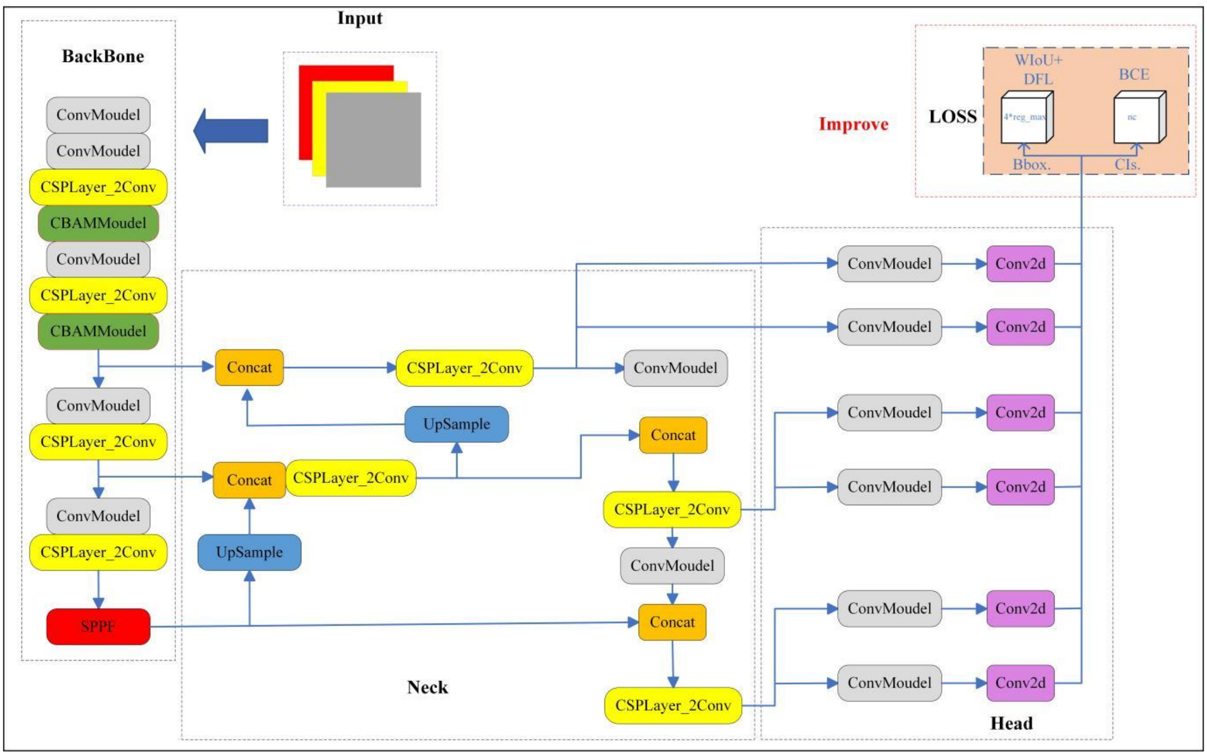


Figure 1. The YOLOv8-CW network model integrates Convolutional Attention (CBAM) and the WIoU loss function

3 Experimental design

3.1 Experimental platform

The research is conducted using the Python programming language along with the PyTorch deep learning framework, and it is performed on a Windows 10 operating system. The experimental platform designed for the study consists of two phases. In the first phase, hardware and software were developed to create a dataset of distracted driving data for drivers. In the second phase, the collected distracted driving dataset is incorporated into the YOLOv8-CW model for transfer learning to train and detect recognition. As shown in Table 1, a summary of the configuration of the experimental platform designed in the laboratory environment is provided.

Table 1. Configuration of the designed experimental platform

Items	Description
CPU	I7-7700
GPU	NVIDIA RTX A4000
CUDA	11.8
Running memory	16GB
Software	Pycharm
Programming language	Python
Deep learning frame work	PyTorch

3.2 Dataset creation

The collection of facial expression data is based on video streams and photos captured from a dual-camera setup in an experimental vehicle, creating the (Distraction) dataset used for training and evaluating distraction detection algorithms. The dataset is randomly split into a training set and a testing set in an 80:20 ratio. The Distraction dataset contains 2,970 images of expressions, with the training set and testing set consisting of 2,373 and 597 images, respectively. This dataset contains a collection of various distracted driving behaviors, with these images serving as indicators of different types of human distracted driving. The images in the dataset are labeled with ten basic distracted driving behaviors: right hand calling, right-hand type, left hand calling, left-hand type, tune the radio, have a drink, talk to passengers, get behind, fix hair and smoking. It includes still images extracted from video sequences and provides a comprehensive representation of distracted driving behavior in various contexts. Use the software Labelimg to annotate the image data for each type of distracted driving behavior, in order to generate XML and TXT files, as shown in Figure 2.

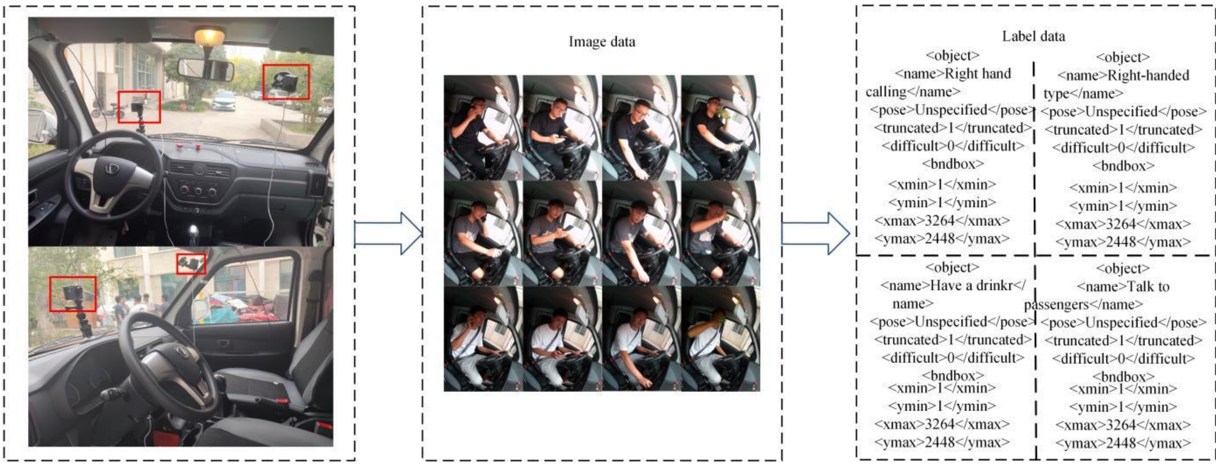


Figure 2. Creation of a dataset for distracted driving behavior

4 Results and Analysis Based on YOLO Experiments

4.1 Evaluation of experimental indicators

This study uses mean average precision (mAP) [19] and floating point operations (FLOPs) [20] as two evaluation metrics to assess the performance of the algorithm. mAP is a commonly used evaluation metric in object detection tasks, representing mean average precision. It averages the AP (average precision) for each category, where AP is a metric used to evaluate the prediction results for each category.

$$AP = \int_0^1 p(r)dr \tag{1}$$

$$mAP = \frac{1}{x}\sum_{i=1}^{x} AP_i \tag{2}$$

$$FLOPs(Conv) = 2 \times W \times H \times (C_i \times K^2 + 1) \times C_o \tag{3}$$

$$FLOPs(Pool) = \frac{W}{S} \times \frac{H}{S} \times C_o \tag{4}$$

In the formula, AP represents the area under the Precision-Recall curve, reflecting the model's performance for a specific category. By averaging the AP values across all categories, we can derive the mAP. The symbols W and H denote the dimensions of the output feature map, C_i represents the number of input channels, K is the kernel size, S indicates the stride, and C_o refers to the number of output channels.

4.2 A comparative experiment based on the YOLO series algorithms

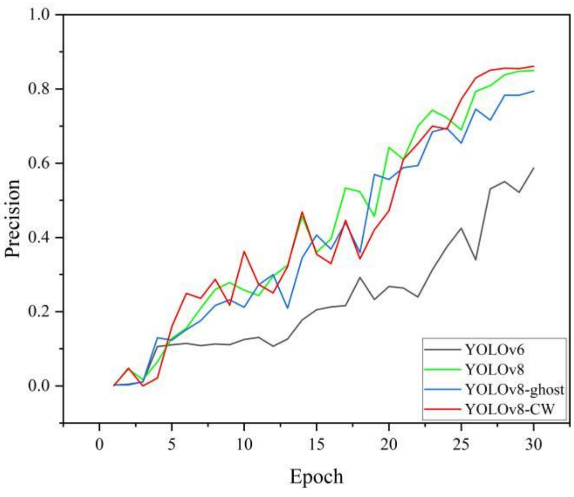
This study conducted comparative experimental research using the YOLO series detection models on the distracted driving behavior dataset, Distraction. In the comparative experiments, all models used the default parameters. The experimental results of the YOLO series algorithms on the distracted driving behavior dataset are shown in Table 2.

In the comparative experiments conducted on the distracted driving dataset, we evaluated the YOLO series models, including YOLOv3, YOLOv5s, YOLOv6, YOLOv8, and its improved version YOLOv8-CW. The experimental results show that the parameter count and floating-point operations (FLOPs) of the YOLOv3, YOLOv5, YOLOv5s, and

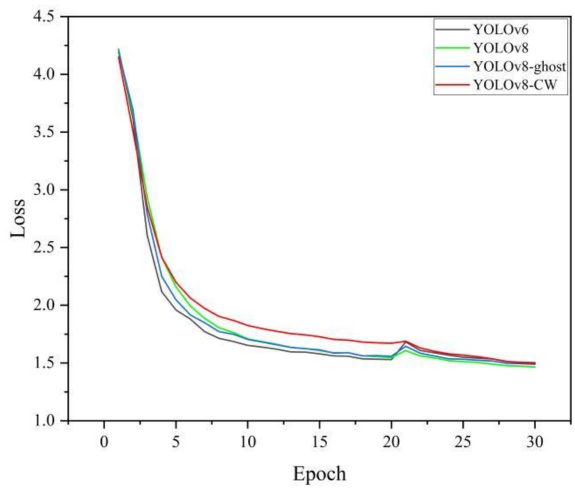
YOLO6 models are relatively large, leading to a significant increase in computational costs. Especially YOLOv3 and YOLOv5s, their parameter count and FLOPs computational complexity are three times that of YOLOv8-CW. However, although the parameter count and FLOPs of other models are lower, their mean Average Precision (mAP) is significantly below that of YOLOv8-CW, with and YOLOv6 achieving mAPs of 0.587, respectively, which represents a decline of up to 27.4% compared to the performance of YOLOv8-CW. The improved YOLOv8-CW model, while not the highest in average precision, achieves a balance between accuracy and computational cost. This improvement is primarily due to the incorporation of the CBAM attention mechanism. This lightweight and versatile module can be easily integrated into any CNN architecture and trained end-to-end alongside the base CNN, significantly boosting the model's detection performance. In addition, the WIoU calculation method used, which does not take aspect ratio into account, further enhances the computation speed, with a time consumption of only 87.2% of that of CIoU, effectively reducing the computational load. These improvements have enabled YOLOv8-CW to excel in both performance and efficiency.

Table 2. Comparison of parameters related to the YOLO series

Model	Params(M)	FLOPs(G)	mAP0.5	mAP0.95	Pre(ms)	Infer(ms)	Post(ms)
YOLOv5	2.5	7.1	0.851	0.653	1.2	1.6	0.8
YOLOv5s	9.1	23.8	0.872	0.68	1.2	2.2	0.8
YOLOv6	4.2	11.8	0.587	0.449	1.2	1.8	0.8
YOLOv8	3.0	8.1	0.850	0.652	1.2	1.7	0.9
YOLOv8-ghost	1.7	5.8	0.794	0.608	1.2	1.9	0.8
YOLOv8-CW	3.0	8.1	0.861	0.651	1.2	1.7	0.8



(a) Precision chart



(b) Loss function chart

Figure 3. Comparison of mAP based on the YOLO family of algorithms

As shown in Figure 3(a), the accuracy curve of the YOLOv8-CW model shows a continuous upward trend, reaching an average precision value of 0.861 at the 30th epoch, surpassing the average precision values of other YOLO models, demonstrating the model's outstanding performance in terms of precision. As shown in Figure 3(b), the loss function curve of the YOLOv8-CW model exhibits a rapid decline followed by a leveling off, indicating a smaller average error between the predicted values and the actual values. The loss between the predicted and actual values has been minimized, indicating the model's strong robustness.

4.3 Validate the model in the public dataset StateFarm

To demonstrate the superiority of the YOLOv8-CW model, this study conducted comparative experiments using the YOLO series detection models on the StateFarm dataset. In the comparative experiments, all models used default parameters.

Table 3. Comparison experiments of the StateFarm dataset

Model	Params(M)	FLOPs(G)	mAP0.5	mAP0.95	Pre(ms)	Infer(ms)	Post(ms)
YOLOv5	2.5	7.1	0.984	0.798	0.8	1.1	1.1
YOLOv5s	9.1	23.8	0.982	0.802	0.8	1.6	1.1
YOLOv6	4.2	11.8	0.985	0.799	0.9	1.0	1.1
YOLOv8	3.0	8.1	0.983	0.802	0.8	1.1	1.0
YOLOv8-ghost	1.7	5.8	0.982	0.798	0.9	1.6	0.9
YOLOv8-CW	3.0	8.1	0.985	0.802	0.9	1.2	0.9

Table 4. A listed comparison of YOLOv8-CW with other approaches.

Author	Year	Network Baseline	Attention Module	Involved Dataset	Accuracy	Train Params
Zhang et al. [21]	2016	VGG-GAP	No	StateFarm	91.3%	140M
Okon et al. [22]	2017	AlexNet+Softmax	No	StateFarm	96.8%	63.2M
Hssayeni et al. [23]	2017	ResNet	No	StateFarm	85%	60M
Hu et al. [24]	2018	VGG16	No	StateFarm	86.6%	33.56M
Janet et al. [25]	2019	Vanilla CNN	No	StateFarm	97.05%	26.05M
Wang et al. [26]	2019	ResNet50	Yes	StateFarm	92.45%	46.16M
Huang et al. [27]	2020	ResNet50+Inception V3+Xception	No	StateFarm	96.74%	72.3M
Dhakate et al. [28]	2020	Inception V3	No	StateFarm	92.9%	25.6M
Lu et al. [29]	2020	Faster R-CNN	Yes	StateFarm	86%	6.53M
Hu et al. [30]	2020	Multi-scale CNN	Yes	StateFarm	96.7%	44.06M
YOLOv8-CW	2024	YOLOv8-CW	Yes	StateFarm	98.5%	3.0M

As shown in Table 3, the average accuracy of YOLOv8-CW on the StateFarm dataset is 98.5%, outperforming other YOLO series. At the same time, its computational parameters and floating-point operations are also lower than most models in the YOLO series mentioned in the paper, ensuring detection accuracy without increasing the computational cost of the model. As shown in Table 4, comparing the training of 10 commonly used models on the dataset, the YOLOv8-CW model outperforms commonly used models like ResNet, Faster R-CNN, and VGG16 by over

10% in average accuracy and reduces the parameter count by a factor of 10. Compared to models without the attention mechanism, the maximum difference in average accuracy is 13.5%, and the maximum difference in parameter count is 137M, significantly reducing the computational cost of the model. Compared to the model with the added attention mechanism, the average precision difference is 12.5%, and the maximum parameter difference is 43.16M, which has improved the model's detection speed. This meets the industry's requirements for lightweight and accuracy, providing the possibility for safe driving assistance deployment for drivers. The test results of the dataset training are shown in Figure 4.

Figure 4. Identification of distracted driving behavior by drivers

5 Conclusion

This study proposes a YOLOv8-CW model framework for identifying distracted driving behaviors in drivers. The key conclusions are as follows: (1) Leveraging the advantages of YOLOv8 and CBAM, YOLOv8-CW can accurately capture fine features and effectively classify distracted behaviors, making it an important tool for identifying driver distraction. (2) Compared to YOLOv3, YOLOv5, YOLOv6, and other models in the YOLO series, YOLOv8-CW demonstrates superior performance in both accuracy and efficiency, especially in complex scenes under varying lighting conditions. Specifically, compared to YOLOv8, YOLOv8-CW can more effectively recognize driver behavior in low gray and high light environments. (3) The training results of the YOLOv8-CW model show that the average detection accuracy of distracted driving behavior reaches 86.1%, demonstrating the model's capability in effectively identifying driver distraction behaviors.

Acknowledgement

This work was supported by the Key Research and Development Projects of Henan Province in 2022(No.221111240200).

References

[1] WHO, *Global status report on road safety 2023*. World Health Organization, 2023.

[2] Thomas D., Feng G., Suzie L., et al., *Driver crash risk factors and prevalence evaluation using naturalistic driving data*. Proceedings of the National Academy of Sciences, 2016,113(10): 2636-2641.

[3] Stewart, Timothy, Overview of Motor Vehicle Crashes in 2020.2022:813-266.

[4] Hasan A. S., Patel D., Jalayer M., Did COVID-19 mandates influence driver distraction Behaviors? A case study in New Jersey. Transportation Research Part F: Traffic Psychology and Behaviour, 2023, 99: 429-449.

[5] NHTSA, National Highway Traffic Safety Administration (NHTSA) 2020 Posted: 2019. In, edited by NHTSA, 2020.

[6] Yang X., Yu Q., Han S., et al., Appearance-posture fusion network for distracted driving behavior recognition. Expert Systems with Applications, 2024, 257,124883.

[7] Zhu Y., Yue L., Zhang Q., et al., Modeling distracted driving behavior considering cognitive processes. Accident Analysis & Prevention, 2024, 202,107602.

[8] Arefin, M. R., Makhmudkhujaev, F., Chae, O., et al., Aggregating CNN and HOG features for real-time distracted driver detection. In 2019 IEEE international conference on consumer electronics ,2019:1–3.

[9] Zhu D., SIFT algorithm analysis and optimization. In 2010 international conference on image analysis and signal processing, 2010:415–419.

[10] Zhao C., Zhang B., He J., et al., Recognition of driving postures by contourlet transform and random forests. IET Intelligent Transport Systems, 2012,6(2):161–168.

[11] Jordan M. I., Mitchell T. M., Machine learning: Trends, perspectives, and prospects. Science, 2015,349(6245): 255-260.

[12] Ping P., Huang C., Ding W., et ai., Distracted driving detection based on the fusion of deep learning and causal reasoning. Information Fusion, 2023, 89:121-142.

[13] Chen S., Yao H., Qiao F., et al., Vehicles driving behavior recognition based on transfer learning. Expert Systems with Applications,2023, 89,119254.

[14] Omerustaoglu F., Sakar C. O., Kar G., Distracted driver detection by combining in-vehicle and image data using deep learning. Applied Soft Computing, 2020, 96,106657.

[15] Chen L., Yao H., Fu J., et al., The classification and localization of crack using lightweight convolutional neural network with CBAM. Engineering Structures, 2023, 275,115291.

[16] Xiong C., Zayed T., Abdelkader E. M., A novel YOLOv8-GAM-Wise-IoU model for automated detection of bridge surface cracks. Construction and Building Materials, 2024, 414, 135025.

[17] Qu J., Sun Q., Qian Z., et al., Hamidreza Zareipour, Fault diagnosis for PV arrays considering dust impact based on transformed graphical features of characteristic curves and convolutional neural network with CBAM modules, Applied Energy, 2024, 355, 122252.

[18] Sun S., Mo B., Xu J., et al., Multi-YOLOv8: An infrared moving small object detection model based on YOLOv8 for air vehicle, Neurocomputing, 2024, 588, 127685.

[19] Ma B., Fu Z., Rakheja S., et al., Distracted Driving Behavior and Driver's Emotion Detection Based on Improved YOLOv8 With Attention Mechanism. in IEEE Access, 2024, 12:37983-37994.

[20] Zhao L., Liu J., Ren Y., et al., YOLOv8-QR: An improved YOLOv8 model via attention mechanism for object detection of QR code defects. Computers and Electrical Engineering, 2024, 118, 109376.

[21] Zhang B., Apply and Compare Different Classical Image Classification Method: Detect Distracted Driver. Stanford CS 229 Project Reports, 2016.

[22] Okon O.D., Meng L., Detecting Distracted Driving with Deep Learning. In Proceedings of the International Conference on Interactive Collaborative Robotics, 2017, 10459,170–179.

[23] Hssayeni M.D., Saxena S., Ptucha R., et al., Distracted Driver Detection: Deep Learning vs Handcrafted Features. IS&T Int. Symp. Electron. Imaging, 2017, 29, 20–26.

[24] Hu Y., Lu M., Lu X. Driving behaviour recognition from still images by using multi-stream fusion CNN. Mach. Vis. Appl., 2018, 30, 851–865.

[25] Jamsheed A., Janet B., Reddy U.S., Real Time Detection of driver distraction using CNN. In Proceedings of the 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, 185–191.

[26] Wang W., Lu X., Zhang P., et al., Driver Action Recognition Based on Attention Mechanism. In Proceedings of the 2019 6th International Conference on Systems and Informatics (ICSAI), 2019, 1255–1259.

[27] Huang C., Wang X., Cao J., et al., HCF: A Hybrid CNN Framework for Behavior Detection of Distracted Drivers. 2020, 8, 109335–109349.

[28] Dhakate K.R., Dash R., Distracted Driver Detection using Stacking Ensemble. In Proceedings of the 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS),2020, 1–5.

[29] Lu M., Hu Y., Lu X., Driver action recognition using deformable and dilated faster R-CNN with optimized region proposals. Appl. Intell., 2019, 50, 1100–1111.

[30] Hu Y., Lu M., Lu X., Feature refinement for image-based driver action recognition via multi-scale attention convolutional neural network. Signal Processing Image Commun. 2020, 81, 115697.