Dean&Francis
ISSN 2959-6157

Progress in Driver Fatigue Detection Technology Based on Image and State Fusion

Boxiang Jia[1, *]

[1]School of Electronic Engineering, Xi'an Shiyou University, Xi'an, 710065, China
*Corresponding author: a528528038@outlook.com

Abstract:

Driver fatigue state detection represents one of the key technologies for enhancing road traffic safety. This paper systematically reviews the research progress, technical challenges, and development trends of driver fatigue detection methods based on the fusion of visual images and vehicle state information. First, it provides a detailed analysis of the theoretical foundations, implementation mechanisms, and performance characteristics of three categories of methods: those based on visual physiological features, driving behavior features, and multi-modal information fusion, tracing the technological evolution from traditional methods to deep learning models. Second, it delves into three core challenges currently faced by these systems in practical applications: environmental robustness, real-time constraints, and individual generalization ability. On this basis, the paper critically reviews the effectiveness and limitations of cutting-edge technical solutions such as lightweight network architectures, self-supervised learning, and personalized federated learning. Finally, it outlines future research directions including multi-modal large models, vehicle-road cooperative perception, and neuromorphic computing, while also discussing opportunities and challenges related to technical standardization and system integration. This review aims to provide researchers and engineering practitioners with a systematic technical reference and insights for further development.

Keywords: Driver fatigue detection; multi-modal fusion; deep learning; autonomous driving safety.

1. Introduction

Road traffic safety represents a major global public health challenge. According to statistics from the World Health Organization (WHO), approximately 20% of global traffic accidents annually are directly

attributable to driver fatigue, resulting in substantial loss of life and economic damage . Fatigue severely impairs drivers' cognitive functions and reaction capabilities, leading to sluggish operation and misjudgment, thereby significantly increasing accident risks. Therefore, the development of a technology capable of real-time, accurate, and non-invasive detection of driver fatigue states is of paramount practical importance for early warning, active intervention, and ensuring road traffic safety.

Although significant progress has been made in vision-based fatigue detection research, several key bottlenecks remain. First, system robustness is often inadequate under complex environmental conditions such as strong light, weak illumination, and facial occlusion, manifesting in a considerable decline in detection accuracy (e.g., mAP). Existing studies have reported false alarm rates as high as 20% under strong light conditions . Second, most current methods rely on uniform thresholds (e.g., a fixed EAR threshold), making it difficult to adapt to individual differences among drivers in eye shape, facial structure, and behavioral habits. Third, although multi-modal fusion strategies can improve recognition performance, they often struggle to meet real-time requirements (≥30 FPS) due to high computational complexity, which hinders practical deployment.

To address these issues, researchers in recent years have focused on developing fatigue detection frameworks that fuse multi-source information. By integrating facial physiological features with vehicle state data and designing hierarchical decision-making mechanisms, they aim to achieve more accurate and adaptive fatigue state recognition and graded warnings. From a methodological evolution perspective, research in this field can be divided into two main categories: first, vision-based analysis of physiological parameters, where metrics such as PERCLOS , EAR , and MAR are widely used to identify eyelid closure, blinking, and yawning behaviors; second, indirect inference methods based on vehicle behavior signals, such as steering wheel operation characteristics and lane deviation [1, 2]. With advancements in artificial intelligence and sensing technology, multi-modal fusion methods have gradually emerged as a new research hotspot, with technical solutions integrating vision, behavior, and physiological signals (e.g., EEG and ECG) continually being proposed. Emerging computational paradigms such as federated learning are also being employed to build adaptive models capable of cross-device collaborative training without aggregating raw data . Nevertheless, existing methods still exhibit significant shortcomings in balancing environmental robustness, personalized adaptation, and real-time performance.

This paper systematically reviews the research progress in driver fatigue detection technology, with a focus on multi-modal fusion methods, aiming to provide a reference for advancing the field. It analyzes the main challenges faced by current technologies, summarizes representative methods along with their advantages and disadvantages, and outlines future research directions. Theoretically, it explores the potential of new technologies such as multi-modal fusion and federated learning in enhancing system generalization and personalized performance. Practically, it aims to provide insights for building efficient and reliable real-time fatigue monitoring systems, thereby contributing to the reduction of traffic accident rates and the improvement of public safety levels.

2. Driver Fatigue Detection Methods and Review

2.1 Methods Based on Visual Physiological Features

In eye feature analysis, PERCLOS (Percentage of Eyelid Closure) is recognized by the US National Highway Traffic Safety Administration (NHTSA) as the "gold standard" for fatigue detection . Studies indicate that when the PERCLOS value exceeds 0.12 for one minute, the risk of accidents increases by 5.7 times . The Eye Aspect Ratio (EAR) is another essential metric, calculated as $EAR=(\|p_2-p_6\|+\|p_3-p_5\|)/(2\|p_1-p_4\|)$. An EAR value remaining below 0.18 for three seconds can be identified as a microsleep episode [1]. Mouth feature analysis mainly targets yawning frequency, quantified by the Mouth Aspect Ratio (MAR). A MAR value greater than 0.7 sustained for 0.8 seconds is considered a valid yawn [2]. Head pose features reflect fatigue through variations in pitch, yaw, and roll angles (HPR). Research has demonstrated a positive correlation ($r=0.78$) between the standard deviation of the head roll angle and EEG θ-band power.

The technological evolution of these methods has progressed from traditional machine learning to deep learning techniques. In 2010, Bergasa et al. employed Active Appearance Models (AAM) combined with Support Vector Machines (SVM), achieving an F1 score of 0.85 in laboratory settings . In 2014, the Viola-Jones detector integrated with LBP features achieved real-time detection at 15 fps on a DSP platform for the first time . The DeepEye model, proposed in 2018, utilized a lightweight Convolutional Neural Network and attained an F1 score of 0.92 on the SEU dataset . In 2022, Jia et al. introduced the YOLO-FDCL model, which incorporates Feature Pyramid Networks (FPN) and Convolutional Block Attention Modules (CBAM), achieving an mAP@0.5:0.95 of 0.942 under complex lighting conditions and reducing the false

ISSN 2959-6157

alarm rate under strong light from 20% to 4.6% . The ViT-TSM model, emerging in 2023, combined Vision Transformer with a Temporal Shift Module, enabling high-speed detection at 120 fps on the Jetson Xavier platform while keeping the PERCLOS estimation error within 0.02.
The advantages of visual methods include non-contact operation, low deployment cost, strong correlation with subjective fatigue scores (e.g., KSS, r>0.8), and support for end-to-end training. However, inherent limitations remain: a 47% increase in EAR error under extreme lighting conditions; threshold drift of approximately ±0.04 due to factors such as monolid eyes and makeup; and power consumption exceeding 4 W when operating at high frame rates (60 fps), which surpasses the automotive-grade power budget of 2 W.

2.2 Methods Based on Driving Behavior Features

These methods indirectly infer driver fatigue levels by analyzing vehicle operational state data. Primary signals include steering wheel angle (SWA), steering wheel angular velocity (SWV), lane deviation (SDLP), pedal operation characteristics, and speed variance.
Common models employed in this approach include SVM-HMM hybrid models [3], LSTM autoencoders [4], and Deep Deterministic Policy Gradient (DDPG) reinforcement learning algorithms [5]. Among these, LSTM autoencoders detect abnormal driving behavior through reconstruction error, while the DDPG algorithm directly outputs a fatigue probability distribution.
Representative studies in this area comprise: a 2012 study by the University of Michigan Transportation Research Institute (UMTRI), which analyzed 1 million kilometers of naturalistic driving data and found that accident risk increases by 5.7 times when SDLP exceeds 0.35 meters [6]; a 2016 study from National Tsing Hua University (Taiwan) that utilized a 1D CNN to process 24-dimensional CAN signals, achieving an F1 score of 0.89; a 2020 model proposed by Toyota integrating Bi-LSTM with an attention mechanism, which improved the F1 score by 6% through alignment of vehicle and visual labels; and a "progressive takeover" strategy introduced by Jilin University in 2023, implementing a graded response system ranging from mild reminders and moderate speed limiting to severe parking interventions.
The advantages of these methods include their non-reliance on cameras, privacy-friendly operation, lower sensitivity to individual physiological differences, and the ability to directly utilize CAN bus data without incurring additional hardware costs. However, several application bottlenecks remain: factors such as vehicle model differences, tire pressure variations, and crosswind interference can introduce false fluctuations, resulting in false alarm rates as high as 8-12%; a required time window of 30–60 seconds leads to delays exceeding 3 seconds, preventing the detection of microsleep events; and difficulty in distinguishing between "fatigue" and "distraction" states contributes to a 9% decrease in F1 score.

2.3 Methods Based on Multi-Modal Information Fusion

Multi-modal fusion methods overcome the limitations of single-modal approaches by combining visual, behavioral, and physiological signals (e.g., EEG, ECG). Based on the stage of fusion, these methods can be categorized into three strategies: early fusion (feature-level), mid-fusion (decision-level), and late fusion (model-level).
Early fusion typically involves feature concatenation combined with PCA for dimensionality reduction [7]. Mid-fusion commonly employs D-S evidence theory [8], while late fusion utilizes two-stream networks with cross-attention mechanisms [9]. Recently, federated learning has been integrated into multi-modal fusion frameworks to enable collaborative training between local visual features and cloud-based behavioral models [10].
Representative studies include: the FusionEye-1 system, which applied early fusion of EAR and SDLP features and improved the F1 score by 4.2%; the EU i-DREAMS project, which established a three-layer vehicle-road-driver architecture and reduced accident rates by 18% in real-road tests; the FedPer+SSA framework proposed by Tsinghua University, which updates 12% of client-side parameters, reduces communication volume by 80%, and achieves 89.9% accuracy [11]; and a method by Jia et al. that combined YOLO-FDCL with federated fine-tuning, improving the AUC by 11% in night scenes while keeping the model size under 1 MB .
Industrial-grade pre-installed systems (e.g., those by Bosch, Mobileye, and Huawei) commonly adopt a "vision-led+behavior verification" dual-channel architecture, assigning a weight of 0.7 to visual cues and 0.3 to behavioral signals. On major benchmarks such as SEU, NTHU, and DROZY, multi-modal methods achieve an average F1 score of 0.93, representing an 8-15% improvement over single-modal approaches. After optimization with TensorRT and INT8 quantization, these systems can maintain a throughput of 30 fps, meeting real-time automotive requirements [12].
The effectiveness of multi-modal methods is demonstrated across various conditions: under favorable lighting, where visual single-modal methods achieve F1>0.92, multi-modal fusion brings a 1% improvement; under challenging conditions such as nighttime, occlusion, or strong light, multi-modal fusion improves the F1 score by 12-18% and

reduces the false alarm rate by 45%; through federated personalized learning, the equal error rate (EER) decreases by 38%, and threshold adaptation time is shortened to less than 30 seconds.

3. Core Challenges and Problems

In practical application environments, fatigue detection systems based on multi-modal information fusion still face several critical technical challenges that severely constrain system reliability and universality.

3.1 Perception Robustness in Complex Environments

Complex lighting conditions represent a primary factor affecting visual perception accuracy. Research indicates that strong direct light may exceed the camera's dynamic range, resulting in overexposure of the pupil area by more than 30% and significantly increasing errors in eye feature extraction. Specifically, the EAR estimation error can rise from 0.015 to 0.022, while the miss rate for the PERCLOS metric increases by 28%. Scenes with rapid lighting changes, such as tunnel entrances and exits (where brightness can change by up to 60 dB within 200 ms), can cause traditional gamma correction algorithms to fail. This leads to a reduction in face detection recall from 0.99 to 0.86 and causes a continuous loss of 5-7 frames in EAR time-series data [13]. Under nighttime infrared illumination, "ghost" reflections from 850 nm light sources on ordinary glasses can inflate the MAR measurement by 22%, increasing the false alarm rate for yawning from 2.1% to 8.7% [14].

Occlusion also poses a major technical challenge. Sunglasses have a reflectance of over 70% in the visible light spectrum, rendering pupil-based detection methods entirely ineffective. Although using 940 nm infrared light can mitigate this issue, it increases hardware costs by approximately 18%. Behaviors such as wearing masks or resting the chin on a hand make mouth features unavailable. Moreover, when the head pitch angle exceeds 30°, the head pose estimation error increases by 0.05 radians, leading to a 19% rise in the miss rate for nodding events [15]. Furthermore, shadow interference caused by sunlight projection through side windows in multi-occupant environments can trigger splits in face detection bounding boxes, increasing the false detection rate by 12%.

Vibration interference in the vehicle environment cannot be overlooked. On high-speed bumpy roads, when random vibration exceeds 3g, the PSNR of images acquired with a 1/30 s exposure time falls below 18 dB, and the EAR sequence exhibits significant "glitch" noise. Compensating for this interference requires incorporating IMU data for image deblurring, which increases computational load by 25%. Experiments demonstrate that even with the state-of-the-art YOLO-FDCL model, extreme conditions combining strong light and occlusion can reduce the mAP@0.5:0.95 from 0.942 to 0.714. Although this still outperforms traditional methods by 13.1 percentage points, it falls below the automotive requirement standard of 0.90.

3.2 Trade-off Between Computational Efficiency and System Real-Time Performance

Computational complexity is a key factor restricting the practical deployment of multi-modal fusion algorithms. Multi-modal Transformer architectures typically contain 25-50 million parameters; even with INT8 quantization on the Jetson Orin platform, the frame rate reaches only 18 fps, failing to meet the automotive-grade requirement of 30 fps. Memory bandwidth limitations are equally critical: processing dual 1080p@60 fps video streams requires 1.2 GB of peak memory, whereas the automotive TDA4VM platform allocates only a 512 MB memory budget for fatigue detection. Although techniques such as gradient checkpointing and mixed-precision training can increase the frame rate to 26 fps, this comes at a cost of a 3% reduction in accuracy [16].

The communication overhead in federated learning frameworks is particularly significant. Each training round requires uploading 150 MB of gradient data, a process that takes 20 seconds to complete in a 4G network environment. Although the SSA sparse update strategy—which uploads only 12% of parameters—can reduce the communication volume to 80 KB, it results in a 1.2% loss in accuracy. Power consumption constraints are also non-negligible: fully utilizing 8 TOPS of compute power consumes 18 W, exceeding the 10 W limit for pre-installed systems. Using dynamic voltage and frequency scaling (DVFS) to control power at 9 W causes the frame rate to drop to 22 fps, necessitating the incorporation of mixture-of-experts (MoE) sparse routing mechanisms to compensate for the resulting loss of 8 fps.

Emerging solutions such as event cameras combined with spiking neural networks (SNN) show promising potential, reducing data volume by 95% and limiting power consumption to within 20 mW. However, this approach requires relabeling tens of thousands of hours of event stream data, increasing annotation costs threefold, and has not yet reached practical implementation [17].

3.3 Insufficient Individual and Scenario Generalization Ability

Individual physiological differences substantially impact detection accuracy. Studies indicate that the population

ISSN 2959-6157

standard deviation of eyelid opening (PDD) is 0.06, and the baseline EAR value for individuals with monolid eyes is 18% lower than for those with double eyelids. The use of fixed thresholds results in a 1.7-fold increase in false alarms for female and Asian individuals [18]. Age-related variations are also significant: the average yawn MAR is 0.55 for elderly groups compared to 0.72 for younger groups. Consequently, the same MAR value of 0.6 may be interpreted as "fatigue" in the elderly while considered "alert" in younger demographics.

Differences in driving behavior habits present additional challenges. The steering wheel "micro-correction" frequency is 0.25 Hz for professional drivers versus 0.52 Hz for novice drivers. If steering wheel velocity (SWV) thresholds are trained solely on novice data, the false alarm rate for professional drivers can reach 14% [19]. Hypoxic conditions in high-altitude environments (altitude > 3500 m) increase the baseline blink rate by 20%, leading to an 11% rise in PERCLOS false alarms for models without retraining [20].

Long-tail distribution and domain shift issues in datasets are equally critical. Public datasets predominantly consist of samples (80%) from young and middle-aged males under optimal daytime lighting conditions, with severe underrepresentation of females, older adults, and nighttime scenarios [21]. Cross-domain testing reveals that model AUC drops from 0.96 in the source domain to 0.81 in the target domain. Even with DANN domain adaptation methods, performance only recovers to 0.89—still below the automotive requirement of 0.95.

4. Cutting-Edge Progress and Future Trends

4.1 Advanced Technologies

Significant breakthroughs have been achieved in lightweight model design. The GhostNetV2-FD architecture integrates Ghost modules and decoupled attention mechanisms into the YOLO-FDCL backbone network, reducing the number of parameters to 1.3 million—a 78% compression rate. After INT8 quantization, it achieves a processing speed of 42 fps on the R-Car H3 platform while maintaining an mAP@0.5: 0.95 of 0.901, representing only a 4.1% performance drop. The MicroViT model innovatively employs "pooling-free shift" operations to enhance the local-global interaction mechanism of MobileViT, reducing multiply-accumulate operations (MACs) by 63% and enabling real-time processing at 38 fps on the Jetson Orin Nano, with a nighttime EAR estimation error of 0.018. The fusion of SPD-Conv and RepVGG replaces stride-2 convolution in the downsampling stage with a space-to-depth transformation strategy, effectively addressing the problem of small target (e.g., closed eyes) information loss and further reducing the PERCLOS miss rate under extreme lighting by 3.2%.

Innovative approaches have also emerged in self-supervised learning. The VideoMAE-FD framework utilizes 90% masked reconstruction pre-training on one million hours of unlabeled in-cabin video; it requires only 5% labeled data during fine-tuning to reach 95% of supervised learning accuracy, significantly reducing labeling cost by 90%. The Contrast-STM method constructs positive and negative sample pairs through temporal consistency, improving EAR feature consistency under low-light nighttime conditions by 18% and enhancing occlusion robustness by 12%. The PoseSync Pretext scheme employs 3D head pose as pseudo-labels for joint training, effectively overcoming the "no eyelid" annotation challenge and reducing EAR regression error by 22%.

Notable progress has been made in personalized federated learning. The FedPer-FD 2.0 framework uses a globally shared backbone network while keeping the fatigue detection head and individual embeddings private. It requires only 30 seconds of local calibration data to reduce equal error rate (EER) by 38%, with communication volume limited to 150 KB per round. The Ditto-FD algorithm introduces a contrastive regularization mechanism to prevent personalized models from diverging from the global consensus, improving cross-city domain AUC by 4% and reducing the number of convergence rounds by 30%. The FedRep+SSA scheme uploads only 8% of parameters combined with Top-K gradient compression and quantization, enabling model updates within one second in a 4G in-vehicle network environment with less than 1% accuracy loss.

New sensing modalities are also diversifying. The combination of event cameras and spiking neural networks (SNN) uses a pixel-level change trigger mechanism to reduce data volume by 95% and limit power consumption to 20 mW, achieving a closed-eye detection latency of only 2.7 ms. The 60 GHz millimeter-wave radar can detect chest breathing and nodding micromovements, unaffected by lighting or occlusion. When fused with vision, it improves the F1 score for nighttime detection by 9%, with a module power consumption of 180 mW. Flexible MEMS seat cushions capture body movement and heart rate variability (HRV) signals, achieving a fatigue recognition rate of 0.88 while avoiding privacy issues associated with cameras. This technology has already been mass-produced for a European truck L2 system.

4.2 Future Trends

Multi-modal fusion is advancing toward deeply unified

architectures. Transformer-based universal architectures are integrating the processing pipelines for vision-behavior-physiology multi-modal data. A pre-trained large model termed "Fatigue-BERT," with over two billion parameters, is anticipated to emerge by 2025. This model will support ten input modalities—including image, event, CAN, EMG, HRV, voice, V2X, ambient light, weather, and road curvature. After pre-training on one million hours of heterogeneous data, downstream fine-tuning will require only ten minutes of in-vehicle data to achieve an F1-score exceeding 0.96.

Vehicle-Road-Human collaborative systems are becoming a pivotal development direction. Roadside units (RSUs) will deploy 4K fisheye cameras and millimeter-wave radar to provide "occlusion probability maps" and "global illumination" labels, transmitting this information to vehicles via V2X within 20 ms, thereby further reducing the visual miss rate by 4-7%. Regional federated grids, structured by city-highway-tunnel management units, will aggregate gradient data from nearby vehicles through edge servers. This approach maintains model freshness within a 30 km range to under one minute, effectively addressing cross-domain drift issues.

Deep integration of progressive human-machine cooperative driving is evolving into a standard paradigm. Fatigue detection systems are advancing beyond basic warning functions to become deeply embedded in L3/L4 longitudinal-lateral control systems: mild fatigue triggers seat vibration and icon flashing warnings; moderate fatigue initiates torque overlay and 80% speed limiting; severe fatigue activates automatic lane changing to the emergency lane for parking. The EU NCAP 2030 roadmap has incorporated "fatigue-takeover" response time into its scoring system, mandating that it not exceed three seconds.

Emerging technological directions exhibit strong multidisciplinary characteristics. NeRF 3D head reconstruction technology can address occlusion issues caused by large-angle head posture changes, achieving an eyelid geometric error of less than 0.3 mm at 2K resolution and improving EAR calculation recall by 6%. Quantum federated learning employs quantum key distribution (QKD) to enable "one-time pad" gradient encryption, meeting the most stringent GDPR privacy requirements without increasing communication overhead—a testbed for this has already been demonstrated on Germany's A9 highway. Neuromorphic computing chips, based on a 22 nm RRAM SNN SoC design, achieve an energy efficiency of 28 TOPS/W. Slated for mass production in 2026, these chips can simultaneously support fatigue, distraction, and emotion detection on a single chip.

5. Conclusion

This paper has systematically reviewed the research progress and development trends in driver fatigue detection technology based on multi-modal information fusion. Studies demonstrate that by synergistically integrating visual physiological features and vehicle behavior data, multi-modal fusion methods significantly enhance the accuracy and environmental adaptability of fatigue detection systems. From a technological evolution perspective, the field has advanced from early single-modal analysis to multi-level and multi-dimensional fusion architectures, with detection accuracy improving from 0.85 in traditional methods to a current optimal level of 0.93, while the false alarm rate has been reduced by 45%. The introduction of advanced technologies such as deep learning, attention mechanisms, and federated learning has further accelerated the transition of fatigue detection systems from theoretical research to practical applications.

However, research indicates that these systems still face three core challenges: insufficient perception robustness due to complex lighting conditions, occlusion interference, and vehicle vibration; the conflict between computational complexity of multi-modal algorithms and automotive-grade real-time requirements; and limitations in generalization capability caused by individual physiological differences and diverse driving habits. Experimental data show that even with the most advanced YOLO-FDCL model, detection accuracy under extreme environmental conditions decreases by 23% compared to ideal settings , underscoring the persistent challenge of environmental adaptability.

Looking ahead, future research should focus on several promising directions. Multi-modal large model architectures are emerging as a significant trend, with pre-trained models such as "Fatigue-BERT" anticipated to contain over 2 billion parameters by 2025 expected to support 10 modal inputs and achieve F1 scores exceeding 0.96 with minimal fine-tuning. Breakthroughs in neuromorphic computing chips, particularly SNN SoC designs based on 22 nm RRAM technology offering 28 TOPS/W energy efficiency, are planned for mass production in 2026 and should enable simultaneous fatigue, distraction, and emotion detection on a single chip. In data privacy and security, quantum federated learning using Quantum Key Distribution (QKD) demonstrates potential for meeting strict GDPR requirements without communication overhead, having already undergone validation on German highways. Furthermore, regional federated grid technology promises to address domain shift issues by maintaining model update delays under one minute within a 30 km range through edge server collaboration.

ISSN 2959-6157

In summary, driver fatigue detection technology is evolving from a standalone function into an integrated safety solution. With advancements in lightweight large models, neuromorphic chips, and quantum encryption, it is projected that by 2030, fatigue detection systems will achieve an F1 score above 0.97 across all scenarios and for all drivers, with latency below 50 ms and standby power consumption under 10 mW. These systems are poised to become a critical safety foundation for L3/L4 autonomous driving. Nevertheless, achieving this goal will require breakthroughs in environmental adaptive algorithms, personalized learning frameworks, and system integration. Future efforts should emphasize industry-academia-research collaboration to translate theoretical advancements into practical applications, thereby providing robust technical support for enhancing road traffic safety.

References

[1] Clark M. Multimodal fusion-based analysis and detection of fatigue driving. Journal of Computer Science and Software Applications, 2024, 4(4): 34-40.

[2] Guo Hanying, Zhang Yuhao, Cai Shanshan, Chen Xin. Effects of Level 3 Automated Vehicle Drivers' Fatigue on Their Take-Over Behaviour: A Literature Review. Journal of Advanced Transportation, 2021, 2021(1): 8632685.

[3] Ye Ziheng. A joint improved DETR network face recognition algorithm based on TSM module and DANN. In: Proceedings of the 2023 4th International Conference on Computing, Networks and Internet of Things. 2023: 759-764.

[4] Park S, Schöps T, Pollefeys M. Illumination change robustness in direct visual slam. In: Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017: 4523-4530.

[5] Tatler B W, Vincent B T. The prominence of behavioural biases in eye guidance. Visual Cognition, 2009, 17(6-7): 1029-1054.

[6] Reuther A, Michaleas P, Jones M, Gadepally V, Samsi S, Kepner J. AI accelerator survey and trends. In: Proceedings of the 2021 IEEE High Performance Extreme Computing Conference (HPEC). IEEE, 2021: 1-9.

[7] Zhao Rui, Li Yun, Gao Fei, Gao Zhenhai, Zhang Tianyao. Multi-agent constrained policy optimization for conflict-free management of connected autonomous vehicles at unsignalized intersections. IEEE Transactions on Intelligent Transportation Systems, 2023, 25(6): 5374-5388.

[8] Ouyang Zhenchao, Niu Jianwei, Guizani Mohsen. Improved vehicle steering pattern recognition by using selected sensor data. IEEE Transactions on Mobile Computing, 2017, 17(6): 1383-1396.

[9] Skorucak J, Hertig-Godeschalk A, Schreier D R, Malafeev A, Mathis J, Achermann P. Automatic detection of microsleep episodes with feature-based machine learning. Sleep, 2020, 43(1): zsz225.

[10] Fan Chaojie, Huang Shufang, Lin Shuxiang, Xu Diya, Peng Yong, Yi Shengen. Types, risk factors, consequences, and detection methods of train driver fatigue and distraction. Computational Intelligence and Neuroscience, 2022, 2022(1): 8328077.

[11] He Wei, Mi Yang, Ding Xiangdong, Liu Gang, Li Tao. Two-stream cross-attention vision Transformer based on RGB-D images for pig weight estimation. Computers and Electronics in Agriculture, 2023, 212: 107986.

[12] Fedus W, Zoph B, Shazeer N. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. Journal of Machine Learning Research, 2022, 23(120): 1-39.

[13] Liu Yi, Qian W Z, Zhang Qiang, et al. Synthesis of high-quality, double-walled carbon nanotubes in a fluidized bed reactor. Chemical Engineering & Technology, 2009, 32(1): 73-79.

[14] Zhou Yuxiao, Guo Zhishan, Dong Zheng, et al. TensorRT implementations of model quantization on edge SoC. In: Proceedings of the 2023 IEEE 16th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC). IEEE, 2023: 486-493.

[15] Qin Yiwen, Lai xing, Gao Guiqing, et al. Failure analysis and countermeasures of a tunnel constructed in loose granular stratum by shallow tunnelling method. Engineering Failure Analysis, 2022, 141: 106667.

[16] Barcellos P, Gomes V, Scharcanski J. Shadow detection in camera-based vehicle detection: survey and analysis. Journal of Electronic Imaging, 2016, 25(5): 051205-051205.

[17] Karumbunathan L S. Nvidia jetson agx orin series: a giant leap forward for robotics and edge AI applications. Technical Brief, 2022.

[18] Netzer G, Johnsson L, Ahlin D, et al. Instrumentation for accurate energy-to-solution measurements of a Texas Instruments TMS320C6678 digital signal processor and its DDR3 memory. In: Proceedings of the 2014 Energy Efficient Supercomputing Workshop. IEEE, 2014: 89-98.

[19] Gallego G, Delbrück T, Orchard G, et al. Event-based vision: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(1): 154-180.

[20] Flamme G A, Deiters K, Needham T. Distributions of pure-tone hearing threshold levels among adolescents and adults in the United States by gender, ethnicity, and age: Results from the US National Health and Nutrition Examination Survey. International Journal of Audiology, 2011, 50(1): S11-S20.

[21] Mileti I, Taborri J, Rossi S, et al. Reactive postural responses to continuous yaw perturbations in healthy humans: the effect of aging. Sensors, 2019, 20(1): 63.