ITC 3/55
Information Technology and Control
Vol. 55 / No. 3/ 2026
pp. 1285-1293
DOI 10.5755/j01.itc.55.3.45072

Adaptive Fatigue Driving Monitoring Algorithm Based on Neural Network

Received 2026/05/20 | Accepted after revision 2026/08/04

**HOW TO CITE:** Qiu, Z., Zou, J., Zhan, B. (2026) Adaptive Fatigue Driving Monitoring Algorithm Based on Neural Network. *Information Technology and Control*, 55(3), 1285-1293. https://doi.org/10.5755/j01.itc.55.3.45072

# Adaptive Fatigue Driving Monitoring Algorithm Based on Neural Network

**Zemin Qiu*, Jiajun Zou, Baorong Zhan**
School of Information and Intelligence Engineering, Guangzhou Xinhua University, Dongguan, 523133, China

**Corresponding author:** qiuzemin@xhsysu.edu.cn

In practical application scenarios, drastic changes in ambient light and driver posture offset can easily lead to a decrease in facial image quality, resulting in weakened key fatigue related features such as blinking and yawning, severely limiting the robustness of existing visual detection models. To address this issue, this paper proposes a feature reconstruction YOLO detection method for image degradation and constructs two core modules: dual convolution feature decoupling and multi-scale reconstruction. The dual convolution feature decoupling module achieves collaborative modeling of facial structured information and degraded region responses by constructing parallel local detail feature branches, which significantly enhances the model's ability to characterize highlight-interfered regions and weak-texture fatigue features. The multi-scale reconstruction module relies on cross-layer feature reconstruction and semantic alignment mechanisms to accurately compensate for the tiny fatigue feature details lost during downsampling, effectively improving the stability of fatigue detection in low-illumination and dynamic scenarios. The experimental results show that the proposed method outperforms the benchmark model in terms of detection performance and robustness, and can effectively alleviate the problem of insufficient fatigue feature expression caused by image degradation in actual monitoring environments. This algorithm has certain practical value in the field of intelligent transportation.

**KEYWORDS:** Object Detection, YOLO11, Fatigue Detection, Feature Reconstruction

## 1. Introduction

Under the development trend of deep integration of intelligent transportation and autonomous driving technologies, real-time and accurate monitoring of driver fatigue status has become a core technical support for fortifying the road safety defense line and reducing the risk of major and extraordinarily

serious traffic accidents. As a core means of automatically identifying fatigue states, object detection can achieve state judgment by capturing typical fatigue features such as eyelid droop, facial muscle relaxation, and yawning. Its detection accuracy and robustness are directly related to personnel safety and traffic scenario management efficiency [7-8].

However, the complex and variable imaging conditions in real traffic environments bring many severe challenges to the effective detection of fatigue features: low-light environments such as night driving and tunnel passage easily cause low image contrast; behaviors such as drivers wearing masks and sunglasses will cause local occlusion of the face; the natural offset of head posture during driving further aggravates the weakening of texture features. The combined effect of these factors poses fundamental challenges to traditional detection methods.

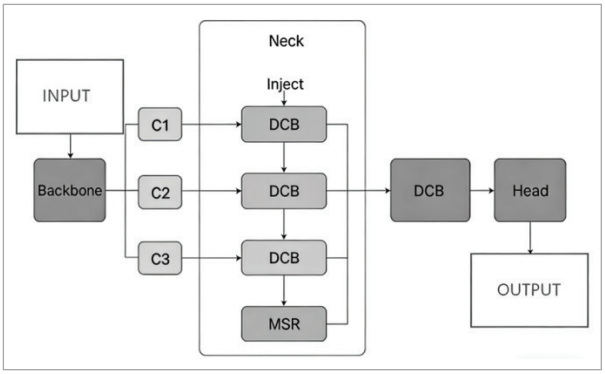
In recent years, object detection methods based on convolutional neural networks have become the mainstream solution for fatigue detection tasks due to the technical advantages of end-to-end training and efficient inference. However, such methods generally rely on stable and clear feature expression, and lack targeted modeling capabilities for feature degradation in poor light and low-contrast scenarios. On one hand, standard convolutional operators are easily disturbed by local overexposure or underexposure when processing facial regions, leading to distorted responses of fatigue texture features, and it is difficult to effectively capture subtle fatigue signals in low-light environments. On the other hand, the existing feature pyramid structure mostly adopts a layer-by-layer transmission multi-scale fusion method, which is difficult to accurately compensate for the tiny fatigue details lost during downsampling, resulting in significant performance limitations of the model in scenarios such as weak-texture and small-target fatigue feature detection, and failing to meet the high-reliability monitoring requirements in complex traffic environments.

To address the above technical bottlenecks, this paper proposes an adaptive fatigue detection method for poor light and low-contrast scenarios. Based on the lightweight YOLO series object detection framework [14-17], this method introduces a Dual Convolution Feature Decoupling Module: by constructing parallel modeling paths for facial structure information and degraded region response correction, it achieves their collaborative modeling, effectively improving the model's ability to express weak-texture fatigue features in poor light environments. At the same time, a Multi-scale Reconstruction Module is designed, which relies on cross-layer feature reconstruction and semantic alignment mechanisms to directionally compensate for the fatigue detail information lost during downsampling, enhancing the detection stability of the model in small-target and low-contrast scenarios. On the premise of maintaining the efficient computing characteristics of the YOLO framework, this method can effectively alleviate the problem of fatigue feature degradation under complex imaging conditions, providing a practical technical solution for fatigue state monitoring.

## 2. Proposed Algorithm

In practical fatigue detection tasks, low-texture regions often lead to unstable responses of fatigue features and loss of detailed information, thereby impairing the model's ability to recognize subtle fatigue characteristics. To address this issue, at the overall network architecture level, this paper first extracts basic features using a lightweight backbone network. On this basis, a dual-convolution structure is introduced, which constructs parallel paths for local detail modeling and receptive field expansion to achieve collaborative modeling of stable facial structure information and responses in degraded regions: the local branch focuses on preserving facial edges and fatigue texture details, while the expansion branch enhances the model's perception of low-light and blurred regions, mitigating the interference of highlight oversaturation and low contrast on fatigue feature representation. Furthermore, in the feature fusion stage, a multi-scale reconstruction module is introduced to perform cross-layer reconstruction and semantic alignment on features of different scales. This module constructs cross-scale information compensation paths to recover subtle fatigue details lost during downsampling, enhancing the effective transmission of shallow texture features to high-level semantic space, and ultimately improving the detection stability of the model in small-target and low-texture scenarios. The overall structure is shown in Figure 1. In the experimental section.

**Figure 1**
Network Architecture of the Algorithm.




## 2.1. Algorithm Architecture

In the visual monitoring scenario of fatigue detection, low-contrast regions are characterized by reduced gradient responses and blurred fatigue texture information. Although such imaging degradation does not alter the macroscopic facial structure [9-10], it significantly disrupts the local response stability of convolution operators, which in turn leads to bias or even distortion of fatigue information during the feature extraction stage. Therefore, simply relying on traditional convolution stacking or conventional feature pyramid structures cannot fundamentally mitigate the above problems.

Based on this, this paper structurally adjusts the YOLO11 detection framework from two dimensions [11-13]: feature decoupling and information reconstruction. A dual-convolution feature decoupling module and a multi-scale reconstruction module are embedded in the backbone network and feature fusion path, respectively. The two modules work synergistically in a functionally complementary manner, enabling the model to improve its adaptability to complex imaging conditions without significantly increasing computational complexity [2-4]. The overall structure is shown in Figure 2.

## 2.2. Video Preprocessing

The core of driver fatigue monitoring lies in analyzing drivers' dynamic behaviors during driving. Video preprocessing needs to accurately identify these dynamic regions and filter out background interference, so as to provide effective data support for real-time fatigue judgment. This section focuses on introducing mainstream motion detection algorithms and their adaptability analysis.

### 2.2.1. Frame Difference Method

The frame difference method identifies motion regions based on pixel changes in consecutive frames. Its core idea is to judge drivers' dynamic behaviors through threshold judgment, with specific steps as follows:

1. Read the pixel values of consecutive video frames ($I_t(x,y)$) and ($I_{t-1}(x,y)$);
2. Calculate the difference between adjacent frames and perform binarization to obtain the difference image ($D(x,y)$);
3. When ($D(x,y)\geq T$), mark it as a motion region; otherwise, mark it as the background.

The expression for obtaining the motion region is shown in the formulas:

$$D(x,y)=|I_t(x,y)-I_{t-1}(x,y)|. \quad (1)$$

$$I_{out}(x,y)=\begin{cases}1, & D(x,y)\geq T\\ 0, & D(x,y)<T\end{cases}. \quad (2)$$

### 2.2.2. Visual Background Extraction Algorithm

The visual background extraction algorithm constructs and dynamically updates a background model, then compares video frames with the background template to identify motion regions. Its core process is summarized as follows:

1. **Background model initialization stage.** Select the first frame of the video sequence as the ini-

**Figure 2**
Flowchart of YOLO11 Detection with Feature Enhancement.



tial reference, and build a background sample set $M(x,y)$ with a capacity of n as the analysis basis, as shown in the formula:

$$M(x,y)=m_1,m_2,m_3,\ldots,m_i,\ldots,m_n . \quad (3)$$

The background sample set $M(x,y)$ established at coordinates $(x,y)$ consists of n samples from $(m_1)$ to $(m_n)$.$(m_i)$ is any sample in the set, and its value comes from random sampling of the neighborhood around the pixel point.

2 **Foreground detection stage.** Calculate the Euclidean distance between the pixel value of the (k)-th frame image at coordinates $(x,y)$ and the background sample $m_i$. When the distance is less than the threshold $R$,$(Count_i)$ is counted as 1; otherwise, it is 0, which is used to count the matching between the pixel point and the background sample, as shown in the formula:

$$Count_i=\begin{cases}1, & D(I_k(x,y),m_i)<R\\ 0, & others\end{cases} . \quad (4)$$

The $k$-th frame image $I_k(x,y)$ in the video sequence is taken into consideration, and the Euclidean distance between it and the background sample space mi is represented by $D(I_k(x,y),m_i)$. When the total sum is greater than or equal to the threshold $T$, $B_k(x,y)$ take 0 to represent the background, otherwise take 1 to represent the foreground, used to generate a binary image, as shown in the formula.

$$B_k(x,y)=\begin{cases}0, & \sum_{i=1}^{n} Count_i \geq T\\ 1, & others\end{cases} . \quad (5)$$

3 **Background modeling processing.** It mainly involves two aspects: the update of the background sample set and the update of the local neighborhood. At the level of background sample set update, the core operation is to replace a randomly selected old sample in the sample set with the real-time pixel value of the current pixel point $(x,y)$, so as to maintain the adaptability of the background model to scene changes.

## 2.3. Dual Convolution Feature Decoupling Module

In the scenario of facial fatigue detection, low-contrast and weak-texture regions often overlap in the same feature map, making it difficult for standard convolution operators to effectively distinguish between these two types of responses. To address this issue, this paper designs a Dual Convolution Feature Decoupling Module. By constructing a parallel dual-path convolution structure, it explicitly models two types of features [5-6]: "stable facial structure" and "degraded region response", thereby reducing the coupling interference between them during the convolution process.

### 2.3.1. Module Structure

The DCB module consists of two parallel convolution paths, with its input being the feature tensor output by the backbone network:

Local detail modeling path: A standard 3×3 convolution kernel is adopted, focusing on capturing local fatigue features such as facial edge contours, eye fissure textures, and mouth deformations, and exhibiting strong responsiveness to high-frequency details.

Receptive field expansion path: Dilated convolution is employed. A 3×3 dilated convolution with a dilation rate (r=2) equivalently covers a 5×5 receptive field range, enhancing the global perception capability of low-light regions and background-blurred regions without increasing the number of parameters.

The calculation processes of the two paths are shown in the following formulas:

$$F_{local}=BN(ReLU(W_l \circledast F)) . \quad (6)$$

$$F_{dilated}=BN(ReLU(W_d \circledast_r F) . \quad (7)$$

Among them, $W_l$ and $W_d$ are learnable convolution kernels for local paths and dilated paths, respectively. $\circledast$ represents standard convolution operation, $\circledast_r$ represents dilated convolution operation with dilation rate $r$, $BN(\cdot)$ represents batch normalization, and $ReLU(\cdot)$ is the activation function.

After channel level concatenation of two features, feature fusion and dimension recovery are performed through 1×1 convolution, as shown in the formula:

$$F_{DCB}=W_{fuse}\circledast Concat(F_{local},F_{dilated}) \quad (8)$$

Among them, Concat (·) represents the concatenation operation of the channel dimension, $W_{fuse}$ is a 1×1 fusion convolution kernel, and $F_{DCB}$ is the output feature map of the DCB module

#### 2.3.2. Degradation inhibition mechanism

To further suppress feature distortion caused by light difference areas, the DCB module introduces a soft weighting mechanism based on channel attention in the fusion layer. By using global average pooling to perform channel statistics on the fused features, the importance weight vectors $w\in RC$ for each channel are generated, as shown in the formula:

$$w=\sigma(W_2\cdot ReLU(W_1\cdot GAP(F_{DCB}))). \quad (9)$$

Among them, GAP(·) represents global average pooling, $W_1$ and $W_2$ are two-layer fully connected weights, and $\sigma(\cdot)$ is the Sigmoid activation function. The final output is shown in the formula:

$$F_{out}=w\otimes F_{DCB}. \quad (10)$$

Among them, $\otimes$ represents the element wise multiplication of the channel dimension, and $w$ serves as the adaptive channel weight to suppress severely degraded channel features and enhance stable fatigue texture response.

### 2.4. Multi Scale Reconstruction Module

In the layer by layer fusion framework of the feature pyramid network, deep semantic features gradually lose small fatigue details after multiple downsampling operations when passing to shallow layers, which restricts the positioning accuracy of the detection head in small targets and weak texture scenes. To this end, this article designs a multi-scale reconstruction module that recovers lost details through cross scale feature compensation paths.

#### 2.4.1. Cross Layer Feature Reconstruction Mechanism

The core idea of the MSRM module is to establish a bidirectional compensation channel from shallow details to deep semantics. Let the multi-scale feature maps output by the backbone network be $\{F_1, F_2, F_3\}$.

The upsampling reconstruction path upsamples the deep feature $F_3$ to the resolution of $F_1$ through bilinear interpolation. After feature alignment convolution, it performs residual fusion with the shallow feature $F_1$, as shown in the formula:

$$F_1^{rec}=F_1+Conv(Up(F_3)). \quad (11)$$

Among them, Up(·) represents the bilinear interpolation upsampling operation, Conv(·) is the 1×1 convolution aligned channel dimension, and $F_1^{rec}$ is the reconstructed shallow feature.

Compensate and reconstruct the intermediate scale feature $F_2$, as shown in the formula:

$$F_2^{rec}=F_2+Conv(Up(F_3))+Conv(Down(F_1)). \quad (12)$$

Among them, Down(·) represents the maximum pooling downsampling operation with a step size of 2, which combines deep semantics and shallow details through bidirectional propagation to achieve collaborative enhancement of intermediate scale features.

#### 2.4.2. Semantic Alignment Mechanism

Due to the differences in semantic levels and value range distributions among multi-scale features, direct superposition is prone to introducing feature bias. To address this issue, MSRM introduces a lightweight semantic alignment sub-module to perform normalization alignment on features prior to cross-scale fusion. Instance normalization is applied to the shallow feature $F_1$ to eliminate the statistical distribution shift caused by uneven illumination, as shown in the formula:

$$F_1^{align}=\gamma\cdot\frac{F_1-\mu}{\sigma+\varepsilon}+\beta. \quad (13)$$

Here, $\mu$ and $\sigma$ are the mean and standard deviation of the feature map in the spatial dimension respectively, $\gamma$ and $\beta$ are learnable affine parameters, and $\varepsilon$ is the numerical stability term. The feature map after se-

mantic alignment has a more consistent value range distribution, which effectively reduces gradient conflicts during cross-scale fusion.

### 2.5. Loss Function

This article adopts a multi task joint loss function consistent with YOLO11, including localization regression loss $L_{box}$, target confidence loss $L_{obj}$, and classification loss $L_{cls}$. The total loss is shown in the formula:

$$L_{total}=\lambda_{box}\cdot L_{box}+\lambda_{obj}\cdot L_{obj}+\lambda_{cls}\cdot L_{cls}\ . \tag{14}$$

*Among them*, $\lambda_{box}$, $\lambda_{obj}$, $\lambda_{cls}$ are the weight coefficients of the three losses, respectively. The positioning loss adopts CIoU loss, which additionally considers aspect ratio consistency compared to IoU loss, and has more accurate positioning constraints for small targets, as shown in the formula:

$$L_{CIoU}=1\text{-}IoU+\frac{\rho^2(b,b_{gt})}{c^2}+\alpha\cdot v\ . \tag{15}$$

Among them, $\rho(b,b_{gt})$ is the Euclidean distance between the predicted box and the center point of the real box, c is the diagonal length of the minimum bounding rectangle, $\alpha$ is the trade-off coefficient, and v is the consistency measure of aspect ratio. The specific definition is as follows:

$$v=\frac{4}{\pi^2}\cdot\left(arctan\left(\frac{w_{gt}}{h_{gt}}\right)\text{-}arctan\left(\frac{w}{h}\right)\right)^2\ . \tag{16}$$

Both the confidence loss and classification loss adopt the form of binary cross-entropy, which is consistent with the standard YOLO11.

## 3. Experimental Simulation

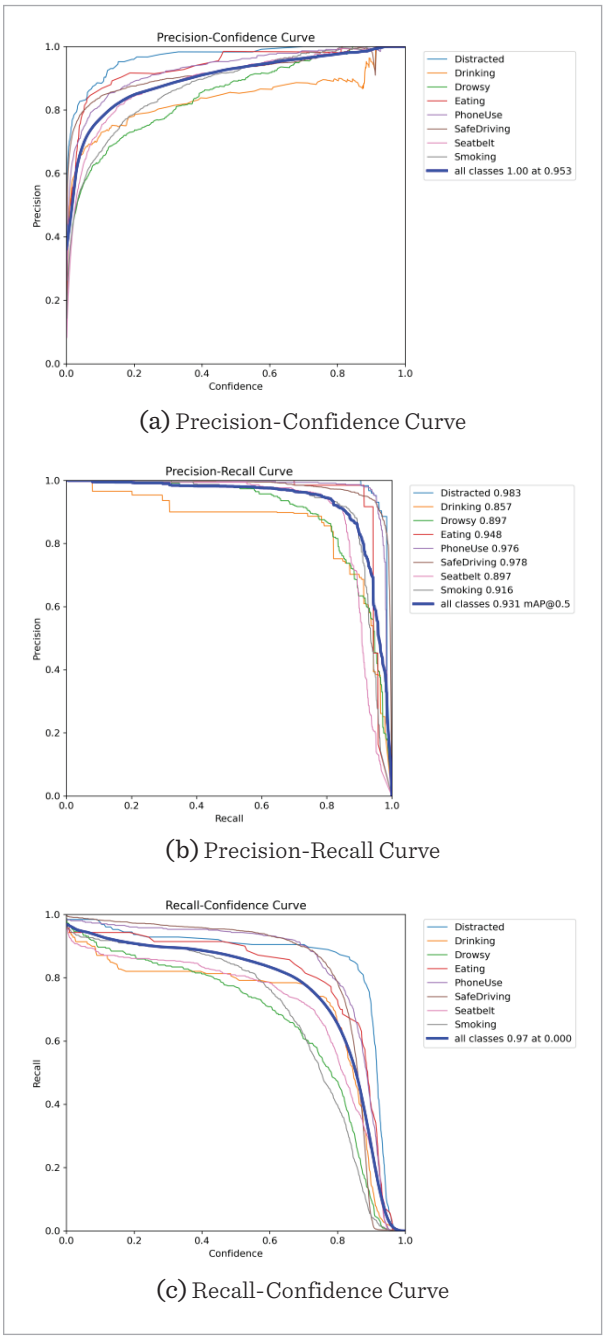
To verify the feasibility of the algorithm proposed in this paper, we first conducted a systematic evaluation of the algorithm's performance through algorithmic index analysis, focusing on core dimensions such as accuracy, computational efficiency, and robustness. On this basis, we carried out a systematic comparative experiment with other existing algorithms to conduct an in-depth analysis of our algorithm from multiple dimensions.

### 3.1. Analysis of Algorithm Indicators

We adopt several common evaluation indicators, namely Accuracy (ACC), Recall (Rec), False Alarm Rate (FA), and Precision (Pre). The specific formulas are as follows:

**Figure 3**

Experimental Results of Different Network Models.




(a) Precision-Confidence Curve

(b) Precision-Recall Curve

(c) Recall-Confidence Curve

Accuracy: $ACC=\frac{TT+FF}{T+F}$

Recall: $Rec=\frac{TT}{TT+FT}$

False Alarm Rate: $FA=\frac{TF}{TT+TF}$

Precision: $Pre=1\text{-}FA$ ,

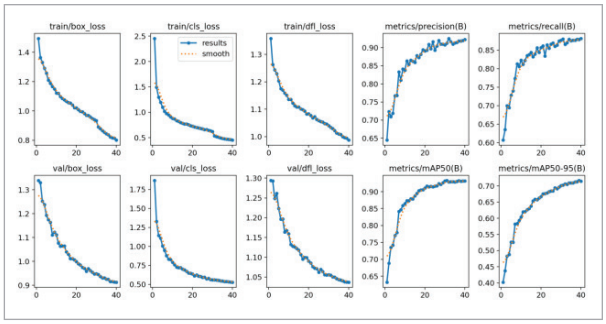
where TT refers to the number of samples with true label predicted as true, FT refers to the number of samples with true label predicted as false, and $T=TT+FT$; FF refers to the number of samples with false label predicted as false; TF refers to the number of samples with false label predicted as true. The experimental results are shown in Figure 3.

As can be seen from the figures, the multi-model fusion strategy performs much better than single models in the fatigue driving monitoring task, and it outperforms others in core indicators including Precision, Recall, and Accuracy. Specifically, the accuracy of the fusion model reaches 99.7%, the recall rate is 99.5%, and the precision is 99.9%.

It can be observed from Figure 4 that all indicators of the model keep rising and finally converge stably without oscillation or decline, indicating that the model performance improves steadily with training epochs; finally, mAP=92.6%, mAP=71.5%, precision is 92.2%, recall rate is 88.1%. The overall detection accuracy is excellent, and false detections and missed detections are controlled at extremely low levels.

**Figure 4**

YOLO Training Loss and Evaluation Indicator Curves.




## 3.2. Comparison Models

To verify the effectiveness of the proposed method, this paper designs the following comparative experiments and compares its overall performance with mainstream detection algorithms and literature algorithm[1]. Table 1 presents the comparative results of the proposed method with various mainstream object detection models on the test set, including mAP, precision, recall, and inference speed (FPS).

It can be seen from the results in Table 1 that the improved model in this paper addresses the problems of low accuracy and poor real-time performance compared with traditional detection models. Compared with the original YOLO11-S, it significantly improves detection accuracy while maintaining high-speed inference, and is slightly better than the literature algorithm [1], especially in fine-grained positioning accuracy, achieving the best balance between accuracy and real-time performance, which is more suitable for the practical application requirements of fatigue driving monitoring.

## 3.3. Performance in Real Scenarios

To verify the feasibility of the algorithm, this paper used several real driving scenarios and tested whether the algorithm can effectively identify driver fatigue and driving status (Figure 5).

**Table 1**

Comparative Experimental Results of Mainstream Models.

| Model | mAP@0.5 | mAP@0.5:0.95 | Precision | Recall | FPS |
|---|---|---|---|---|---|
| Faster R-CNN (R50-FPN) | 84.2% | 55.7% | 83.1% | 79.6% | 12 |
| RetinaNet | 85.8% | 57.3% | 83.1% | 80.2% | 18 |
| YOLOv9-S | 91.3% | 64.2% | 90.4% | 87.3% | 148 |
| YOLO11-S(Original Version) | 92.1% | 65.1% | 91.0% | 88.2% | 162 |
| Literature algorithm[1] | 93.0% | 70.5% | 92.0% | 84.5% | 160 |
| Ours | 93.2% | 71.2% | 92.3% | 88.1% | 159 |

**Figure 5**
Actual Detection Performance Results.



The algorithm can still stably recognize the facial fatigue characteristics and physical behavior of drivers, and it simultaneously completes the detection of three core tasks: fatigue state, safe driving, and seat belt wearing. In addition, the confidence level of fatigue detection and seat belt detection tasks remains within a high range, which verifies the robustness of the algorithm's multi-objective parallel reasoning.

## 4. Conclusions

Aiming at the problem that YOLO11 struggles to recognize small-scale fatigue features in complex environments, this paper proposes an adaptive fatigue detection method suitable for low-illumination and low-contrast scenarios. Based on the original network framework, this paper introduces a dual convolution feature decoupling module and a multi-scale reconstruction module to improve the model from two stages: feature extraction and feature fusion. Experimental results show that the proposed algorithm improves the recall rate of fatigue features while maintaining basically unchanged inference efficiency. The model demonstrates more stable fatigue detection performance and can accurately capture subtle fatigue features, indicating that this method has certain application value in fatigue monitoring tasks such as intelligent driving. Future research will focus on model lightweighting to further optimize the network structure and improve the deployment efficiency of edge devices, so that it can better adapt to practical application scenarios.

### Acknowledgement

This research was supported by the 2024 University-level Research Project of Guangzhou Xinhua University (No. 2024KYZDZK02), the Guangdong Province Key Construction Discipline Research Capacity Enhancement Project (Nos. 2021ZDJS144, 2024ZDJS130), the Characteristic Innovation Category Project of Guangdong Ordinary Colleges and Universities (No. 2024KTSCX127), the School-level Scientific Research Project of Guangzhou Xinhua University (No. 2024KYCXTD02), the Young Innovative Talents Category Project of Guangdong Ordinary Colleges and Universities (Nos. 2023KQNCX124, 2024KQNCX076), and the Key Research Platforms and Projects of Regular Higher Education Institutions in Guangdong Province (No. 2025ZDZX3059).

## References

1. Abo-Zahhad, M. M., Elghamrawy, S., Hefny, A. A., Elawady, Y. H. Early Drowsiness Detection Model in Autonomous Vehicles Using GAN and YOLO Integration. Neural Computing & Applications, 2025, 37(33), 28353-28378. https://doi.org/10.1007/s00521-025-11676-7
2. Garta, I. Y., Manongga, W. E., Huang, S.-W., Chen, R.-C. On-Street Parking Space Detection Using YOLO Models and Recommendations Based on KD-Tree Suitability Search. Computers, Materials & Continua, 2025, 85(12), 4457-4471. https://doi.org/10.32604/cmc.2025.067149

3. Hoang, V.-H., Lee, J. W., Park, C.-S. Enhancing Fire Detection with YOLO Models: A Bayesian Hyperparameter Tuning Approach. Computers, Materials & Continua, 2025, 83(6), 4097-4116. https://doi.org/10.32604/cmc.2025.063468
4. Ishtiaq, M., Won, J.-U. YOLO-SIFD: YOLO with Sliced Inference and Fractal Dimension Analysis for Improved Fire and Smoke Detection. Computers, Materials & Continua, 2025, 82(3), 5343-5361. https://doi.org/10.32604/cmc.2025.061466
5. Jeribi, F., Siddiqa, A., Kibriya, H., Tahir, A., Rana, N. Efficient Wound Classification Using YOLO11n: A Lightweight Deep Learning Approach. Computers, Materials & Continua, 2025, 85(10), 955-982. https://doi.org/10.32604/cmc.2025.067892
6. Li, M., Li, X., Zhao, M. A Method for Small Target Detection and Counting of the End of Drill Pipes Based on the Improved YOLO11n. Computers, Materials & Continua, 2025, 85(10), 1917-1936. https://doi.org/10.32604/cmc.2025.067382
7. Li, X.-C., Zhu, J.-H., Dong, P.-X., Wang, Y.-Q. YOLO-Fastest-IR: Ultra-Lightweight Thermal Infrared Face Detection Method for Infrared Thermal Camera. Journal of Infrared and Millimeter Waves, 2025, 44(5), 790-800. DOI: 10.11972/j.issn.1001-9014.2025.0065.
8. Li, Y., Dong, C. Improved Method for a Pedestrian Detection Model Based on YOLO. Frontiers of Agricultural Science and Engineering, 2025, 12(2), 245-260. https://doi.org/10.15302/J-FASE-2025613
9. Liu, Y., Liu, F., Jiang, M. Research on YOLO Algorithm for Lightweight PCB Defect Detection Based on MobileViT. Optoelectronics Letters, 2025, 21(8), 483-490. https://doi.org/10.1007/s11801-025-4106-9
10. Thapa, A., Horanont, T., Neupane, B., Klaylee, J., Witayangkurn, A. Classification of Oil Palm Tree Conditions from UAV Imagery Using the YOLO Object Detector. Big Earth Data, 2025, 9(4), 860-886. https://doi.org/10.1080/20964471.2025.2491881
11. Tsai, A.-C., Pookunngern, C. Real-Time Larval Stage Classification of Black Soldier Fly Using an Enhanced YOLO11-DSConv Model. Computers, Materials & Continua, 2025, 84(8), 2455-2471. https://doi.org/10.32604/cmc.2025.067413
12. Xia, X., Zhang, N., Guan, Z., Chai, X., Ma, S., Chai, X., Sun, T. PAB-Mamba-YOLO: VSSM Assists in YOLO for Aggressive Behavior Detection Among Weaned Piglets. Artificial Intelligence in Agriculture, 2025, 15(1), 52-66. https://doi.org/10.1016/j.aiia.2025.03.002
13. Xie, X., Zhang, R., Guo, J., Lu, L., Pan, H., Luo, X., Meng, S. Strawberry Disease Detection Algorithm Based on YOLO11-Strawberry. Food Quality and Safety, 2025, 9(3), 561-570. https://doi.org/10.1093/fqsafe/fyad032
14. Xu, S., Jiang, W., Liu, Q., Wang, H., Zhang, J., Li, J., Huang, X., Bo, Y. Coal-Rock Interface Real-Time Recognition Based on the Improved YOLO Detection and Bilateral Segmentation Network. Underground Space, 2025, 21(2), 22-43. https://doi.org/10.1016/j.undsp.2025.10.003
15. Yang, X., del Rey Castillo, E., Zou, Y., Wotherspoon, L., Yang, J., Li, H. Automated Concrete Bridge Damage Detection Using an Efficient Vision Transformer-Enhanced Anchor-Free YOLO. Engineering, 2025, 51(8), 311-326. https://doi.org/10.1016/j.eng.2025.03.014
16. Zhang, J., Xu, C., Shen, S., Zhu, J., Zhang, P. MFF-YOLO: An Improved YOLO Algorithm Based on Multi-Scale Semantic Feature Fusion. Tsinghua Science and Technology, 2025, 30(5), 2097-2113. https://doi.org/10.26599/TST.2024.9010097
17. Zhao, W.-W., Shen, S.-L., Yan, T., Zhou, A. Intelligent Approach for Mucky Soil Identification During Shield Tunnelling by Enhanced YOLO Model. Journal of Rock Mechanics and Geotechnical Engineering, 2025, 17(6), 3327-3338. https://doi.org/10.1016/j.jrmge.2024.09.025

This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) License (http://creativecommons.org/licenses/by/4.0/).