

Research Article

Cite this article: Deng X, Du D, Fang H and Ren Y (2024). Fault diagnosis of proton exchange membrane fuel cells based on improved YOLOv5. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, **38**, e21, 1–10
<https://doi.org/10.1017/S0890060424000295>


Received: 09 March 2024
Revised: 21 June 2024
Accepted: 29 October 2024

Keywords:
EfficientViT; fault diagnosis; PEMFC; YOLOv5

Corresponding author:
Dongshen Du;
Email: dshdu@163.com

This work is supported by the National Natural Science Foundation of China (Grant No. 61873107), and the Qinglan Program to Cultivate Middle-aged and Young Science Leaders of Colleges and Universities of Jiangsu Province, China.

Fault diagnosis of proton exchange membrane fuel cells based on improved YOLOv5

Xiangshuai Deng , Dongshen Du, Huaisong Fang and Yiming Ren

Faculty of Automation, Huaiyin Institute of Technology, Huaian, Jiangsu, China

Abstract

With the widespread application of proton exchange membrane fuel cells (PEMFCs), ensuring the safe and reliable operation of the PEMFCs is becoming more and more important. Timely diagnosis of fault types and the implementation of targeted interventions are crucial for addressing these challenges. In this study, a simulated PEMFC model is firstly introduced by using Fluent, and the effectiveness is validated through experiments involving membrane dry faults, water flooding faults, normal states, and unknown states. Then, a data-driven deep learning convolutional neural network, YOLOv5-CG-AS, is developed, which employs the EfficientViT network as the backbone, incorporating lightweight improvements through the proposed CG-AS attention layer. The results demonstrate that YOLOv5-CG-AS can automatically extract fault features from images for offline fault diagnosis and can perform real-time online diagnosis of multiple parameter curves of PEMFCs. Moreover, the experimental results have validated the feasibility and effectiveness of proposed method and shown the average precision mean Average Precision (mAP) of the trained model reaches 99.50%, superior than other conventional strategies. This has significant implications for advancing fault diagnosis methods, enhancing the reliability and durability of PEMFC systems, and promoting further development in the field.

1. Introduction

Proton exchange membrane fuel cells (PEMFCs), as an emerging clean energy technology, offer advantages such as high efficiency, environmental friendliness, and low noise, making them one of the most promising clean energy sources in recent years (Zheng et al., 2013; Wang et al., 2021). However, the occurrence of faults in fuel cells can lead to reduced reliability and durability, which remains a major challenge hindering the development of fuel cell technology. Therefore, accurate and timely fault diagnosis is crucial for maintaining the performance of PEMFCs and extending their lifespan (Jeong et al., 2019; Sinha and Mondal, 2021).

Although there are various types of faults in proton exchange membrane fuel cells, the most common and typical faults are membrane dry faults and flooding faults (Huang et al., n.d.; Zhang et al., 2020). During the operation of a PEMFC, proton conductivity is closely related to membrane water content; thus, optimal output performance corresponds to a sufficiently hydrated proton exchange membrane. However, excessive internal water content in the cell can lead to flooding faults, while insufficient water content can result in membrane dry faults. Flooding in the gas diffusion layer and channels hinders the transport of reactants to reaction sites, reducing the active surface area of the catalyst due to water coverage. This results in a significant increase in activation and concentration losses in the PEMFC. Membrane dry faults lead to an increase in resistance, causing the PEMFC to generate more heat during operation, further reducing energy conversion efficiency and exacerbating membrane dry faults, potentially leading to membrane tears. This significantly impacts output performance and remaining lifespan.

Currently, there are three main methods for PEMFC fault diagnosis: experimental testing-based, model-based, and data-driven approaches (Allam et al., n.d.; Zhou et al., 2015). Experimental testing-based methods generally have high equipment requirements and often require system shutdown, involving extensive testing and analysis. As a result, they are unsuitable for online diagnosis, continuous system monitoring, and fault prevention. Model-based approaches involve utilizing physical models to simulate and model the system, offering better interpretability and reliability. They allow safe modifications to experimental conditions, mitigating irreversible damage caused by embedded faults in real PEMFC systems (Guo et al., 2014; Huang and Xiang, 2015; Wang et al., 2021; Cui and Xiang, 2023). However, in application scenarios with high real-time requirements, model-based methods may not provide a sufficiently rapid response for fault diagnosis. Data-driven approaches, especially those using deep learning neural networks, leverage large amounts of experimental data for training, automatically learning features, reducing dependence on prior knowledge, and are suitable for diagnosing complex systems, including unknown faults (Zhong et al., 2006; Zhang and Guo, 2021; Darzentas et al., 2022; Lee et al., 2022). They are applicable when the model is incomplete or inaccurate. Data-driven models

require a large number of fault samples as training data, which often face issues such as class imbalance and insufficient labeled fault samples in practical situations (Zhang *et al.*, 2024). Typically, constructing fault models to acquire fault characteristics and generate training samples is a key to addressing this issue (Xiang *et al.*, 2016). Fault models, combined with generative adversarial networks (GAN) and GAN-based domain adaptation (DA) networks, can adjust the original simulated fault samples. Through adversarial training between the refiner and domain discriminator, the samples are made to resemble actual fault samples, which can also solve the problem of insufficient fault samples (Gao *et al.*, 2021; Luo *et al.*, 2022). This paper draws on this concept, using a PEMFC simulation model and embedding faults into the model to obtain a sufficient number of fault samples, thereby addressing the issues of insufficient training data and class imbalance faced by data-driven models. YOLOv5 (You Only Look Once version 5), being a data-driven method, leverages deep learning neural networks to automatically learn features from large amounts of experimental data. This capability is beneficial for capturing intricate patterns and complex relationships within the PEMFC system. Its efficiency and speed make it suitable for real-time fault diagnosis applications, which is particularly advantageous in situations where timely responses are crucial for maintaining the performance and reliability of the PEMFC system.

Based on the above discussion, this paper combines the strengths of model-based and data-driven approaches to diagnose faults in PEMFCs. Initially, a multiphysics coupled PEMFC simulation model is established in the FLUENT environment, designed to acquire data under various PEMFC operating conditions, avoiding irreversible damage to PEMFC during fault data acquisition and issues such as class imbalance and insufficient labeled fault samples in data-driven models. Then, we first enhance the EfficientViT model by incorporating Axial Self-Attention (ASA) to reduce the computational complexity of Cascaded Group Attention (CGA). This improved EfficientViT is then used as the backbone of the YOLOv5 model (Liu *et al.*, n.d.; Ho *et al.*, 2019). It provides a fault diagnosis algorithm suitable for both offline and online PEMFC systems, contributing significantly to the development of PEMFC fault diagnosis. The model's low computational complexity and small size offer a new portable algorithm for embedded devices, presenting a promising possibility for the practical application of portable devices in offline and online PEMFC fault diagnosis.

2. PEMFC simulation model

2.1. Model assumptions

It is well known that the inner chemical reactions of PEMFCs are highly complex owing to involving gases and electrolytes. In order to mitigate computational complexity and enhance computational efficiency, the following assumptions are firstly given:

- Gases are treated as ideal gases and follow the ideal gas law.
- The operating environment of the fuel cell is in a steady state.
- Both the diffusion layer and the catalyst layer are considered porous media.

2.2. Physical model

The physical model of a single-cell PEMFC is illustrated in Figure 1, which is composed of the cathode/anode gas channels, the cathode/anode diffusion layers, the cathode/anode catalyst layers, the cathode/anode current collectors, and the proton exchange membrane (Guo *et al.*, 2014). The model parameters are given in Table 1.

2.3. Control equations

In the process of software modeling, the following equations are primarily utilized: 1) mass conservation equations effective for fluid flow, diffusion, and electrochemical reactions; 2) continuity equations governing fluid transport; 3) the Butler-Volmer equation describing the relationship between current and potential; 4) component conservation equations for gas-phase mixtures; and 5) the Stefan-Maxwell equation describing the gradient of molar fractions in components (Huang *et al.*, n.d.; Guo *et al.*, 2014; Zhou *et al.*, 2015; Zhang *et al.*, 2020). The specific formulas are given by equations (1) to (6).

Mass conservation equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \quad (1)$$

Continuity equation for fluid transport:

$$\frac{\partial(\epsilon \rho)}{\partial t} + \nabla \cdot (\epsilon \rho \mathbf{U}) = 0 \quad (2)$$

The density, denoted by ρ , characterizes the compactness of the fluid. While the velocity vector \mathbf{v} illuminates the dynamic motion of

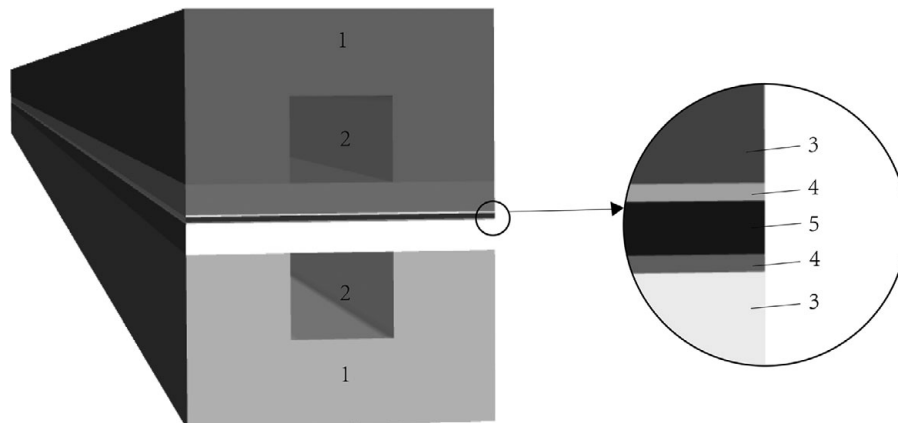


Figure 1. PEMFC model. 1. Collector plate, 2. gas channel, 3. diffusion layer, 4. catalyst layer, and 5. PEM.

Table 1. PEMFC simulation model parameters

Parameter	Value
PEM thickness/m	5×10^{-5}
Catalyst layer thickness/m	1.68×10^{-5}
Gas diffusion layer thickness/m	2×10^{-4}
Length of gas flow field/m	2.5×10^{-2}
Width of gas flow field/m	1×10^{-3}
Height of gas flow field/m	7×10^{-4}
Height of flow plates/m	1.4×10^{-3}
Width of flow plates/m	2.8×10^{-3}
Porosity	0.6
Inlet gas pressure/kPa	101.527
Inlet gas temperature/K	353
Inlet humidity of cathode gas	1.0

the fluid, the operator ∇ finds representation in $\frac{d}{dx} + \frac{d}{dy} + \frac{d}{dz}$. Described by the symbol ϵ , the porosity of the porous medium reflects the openness of the material, with U signifying the vector of velocity for the inner fluid circulating. Within the porous medium, fluid motion is captured by the velocity vector U , while ϵ quantifies the porosity of the medium.

Butler-volmer equation:
(Anode)

$$i_a = i_{0,a} \left\{ \exp \left[\frac{-\alpha_{Rd,a} F (E_a - E_{r,a})}{RT} \right] - \exp \left[\frac{\alpha_{0x,a} F (E_a - E_{r,a})}{RT} \right] \right\} \quad (3)$$

(Cathode)

$$i_c = i_{0,c} \left\{ \exp \left[\frac{-\alpha_{Rd,c} F (E_c - E_{r,c})}{RT} \right] - \exp \left[\frac{\alpha_{0x,c} F (E_c - E_{r,c})}{RT} \right] \right\} \quad (4)$$

Anode current density, denoted as i_a , characterizes the flow of current at the anode, while $i_{0,a}$ represents the anode exchange current density. The transfer coefficient of the anodic reducible substance is given by $\alpha_{Rd,a}$, and $\alpha_{0x,a}$ describes the transfer coefficient of the anodic oxidizable substance. E_a denotes the anode potential, and $E_{r,a}$ signifies the anode equilibrium potential. Additionally, R , T , and F correspond to the gas constant, reaction temperature, and Faraday constant, respectively. Cathode current density, i_c , characterizes the flow of current at the cathode, with $i_{0,c}$ representing the cathode exchange current density. The transfer coefficient of the cathodic reducible substance is denoted as $\alpha_{Rd,c}$, and $\alpha_{0x,c}$ describes the transfer coefficient of the cathodic oxidizable substance. E_c denotes the cathode potential, and $E_{r,c}$ signifies the cathode equilibrium potential.

Species conservation equation:

$$\frac{\partial(\epsilon \rho x_i)}{\partial t} + \nabla(\epsilon \rho x_i U) = \nabla(\rho D_i^{eff} \nabla x_i) + S_i \quad (5)$$

The mass fraction of gas component x_i characterizes the portion of component i in the gas mixture. Component i experiences effective diffusion with a diffusion coefficient D_i^{eff} within the gas mixture. The rate of mass generation (S_i) reflects the outcome of electrochemical reactions involving component i .

Stefan-Maxwell equation:

$$\nabla y_i = RT \sum_j \frac{y_i N_j - y_j N_i}{p D_{ij}^{eff}} \quad (6)$$

The molar fraction of component i in the gas phase, denoted as y_i expresses the relative abundance of component i . Within a differential volume element, N_i represents the average apparent gas phase flux of component i . The partial pressure of components i and j is represented by p . The binary diffusion coefficient for components i and j is denoted as D_{ij}^{eff} , capturing the effective diffusion between the two species.

2.4. Model validation

To verify the accuracy of the established PEMFC simulation model, we compared its generated polarization curves with actual experimental data from an EC-type PEMFC. The results show that the voltage values of the simulation model and the experimental data are highly consistent at multiple current density points, with matching trends and errors within a reasonable range (less than 5%). This consistency and low error indicate that the model has high precision and credibility, effectively simulating the performance of actual fuel cells, as described in Figure 2.

In cases of low porosity, water generated by electrochemical reactions cannot be promptly expelled. The gas diffusion layer may become partially or entirely filled with liquid water, impeding the subsequent transport of oxygen and causing a decline in the efficiency of fuel cell operation, leading to flooding faults. Furthermore, a decrease in membrane humidity can result in reduced conductivity, causing membrane dry faults. Therefore, in the simulation, adjusting the porosity from 0.6 to 0.2 represents the flooding faults, while reducing cathode humidity from 1 to 0.4 indicates the membrane dry faults (Zhou et al., 2015; Sun et al., 2022; Calasan et al., 2024; Ma et al., 2024)

3. YOLOv5 network model

With the benefit of efficiency, accuracy, and user-friendly features, the YOLOv5 model has been widely applied in object detection tasks within the field of computer vision, including object recognition, pedestrian detection, and traffic scene analysis. The model is composed of four components: input, backbone, neck, and prediction (Kim et al., 2022; Liu et al., 2023).

3.1. Input

The input component includes three parts: Mosaic data augmentation method, adaptive anchor calculation, and adaptive image scaling. The Mosaic data augmentation method provides more background information and object context. The adaptive anchor calculation allows the model to adapt to different sizes of targets and image variations. The adaptive image scaling ensures that input images have a consistent size. These techniques collectively contribute to the input component of YOLOv5, enhancing the model's detection performance and robustness.

3.2. Backbone

The backbone network, CSPDarknet53, employs a special CSP connection to fuse low-level and high-level features, enhancing the accuracy of object detection. CSPDarknet53 divides the input

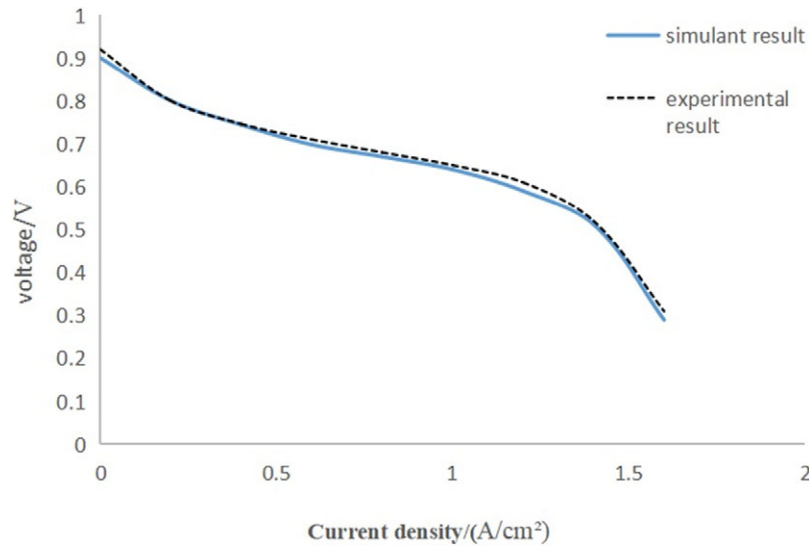


Figure 2. Comparison of experimental and simulated V-I curve.

feature map into two branches: the main branch and the cross-stage connection branch. The main branch is responsible for extracting high-level semantic features, while the cross-stage connection branch focuses on extracting low-level features. The fusion of low-level features with high-level features enhances the overall feature representation capability.

3.3. Neck

The neck section consists of an FPN + PAN structure, primarily used to extract multiscale feature information by aggregating high and low-level features of the image, thereby enhancing the effectiveness of object detection.

3.4. Prediction

The prediction layer mainly consists of a loss function and Non-maximum Suppression (NMS). GIOU_Loss is employed as the loss function to address cases where bounding boxes in previous YOLO versions do not overlap. NMS is utilized to select the box with the highest confidence as the final detection result, thereby enhancing the accuracy and reliability of object detection.

4. Improved YOLOv5 algorithm

4.1. EfficientViT

EfficientViT is a lightweight neural network model proposed (Liu *et al.*, *n.d.*), comprising an efficient memory sandwich layout and cascaded group attention modules. It addresses issues such as excessive memory access time caused by Multihead Self-Attention (MHSA), computational redundancy between attention heads, and inefficient model parameter allocation.

4.2. Improvements to EfficientViT

We construct a new attention layer, CG-AS Attention, by combining Cascaded Group Attention (CGA) and Axial Self-Attention (ASA). By using the EfficientViT block as the base module, the input features of each EfficientViT Block undergo N FFN layers,

followed by the new attention layer CG-AS Attention, and finally, N additional FFN layers to transform the output features. Simultaneously, GSConv is employed before each FFN layer to replace the original DWConv in the model, facilitating interaction between local tokens and introducing inductive bias. Between FFN layers Φ_i^F , CG-AS Attention Φ_i^A is utilized for spatial fusion, with its calculation formula as follows:

$$X_{i+1} = \prod_i^N \Phi_i^F \left(\Phi_i^A \left(\prod_i^N \Phi_i^F(X_i) \right) \right) \quad (7)$$

X_i represents the complete input features of the i -th block. The entire structure transforms into X_{i+1} with N FFN layers, with N FFN layers positioned before and after the cascaded group attention layer CGA. This foundational module reduces the use of attention, alleviating the memory access time consumption issue associated with attention calculations. CG-AS Attention



Figure 3. EC-type proton exchange membrane fuel cell.

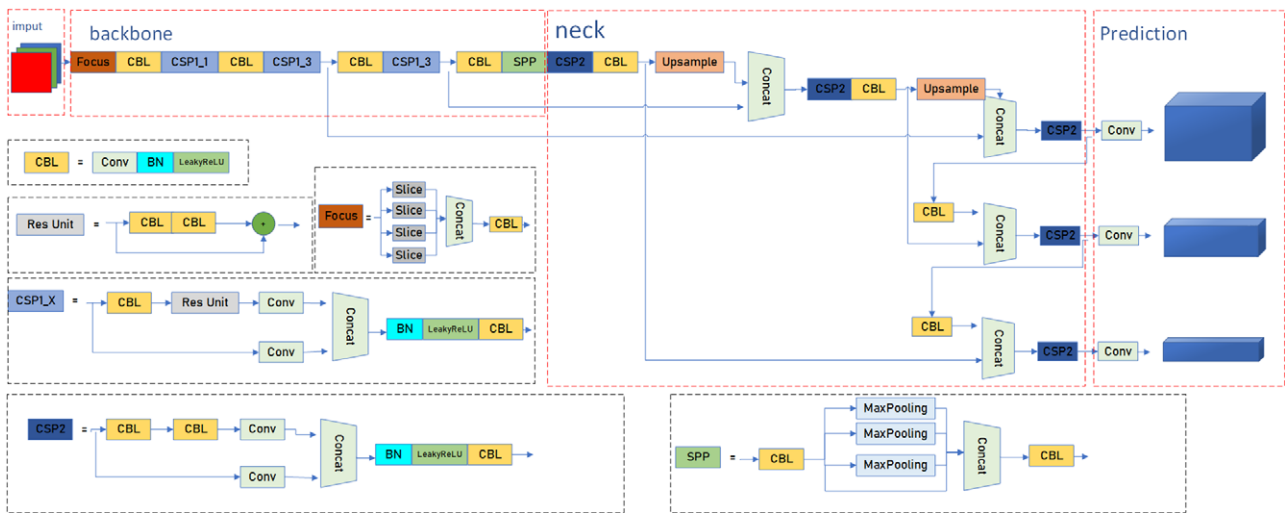


Figure 4. YOLOv5 network model.

employs a method of splitting input features, providing only a portion of the input features to each attention head to reduce computational redundancy. Additionally, attention calculations follow a cascaded approach, with the output of each head added to the subsequent heads, gradually refining feature representations and enhancing the depth and expressive capacity of lightweight networks without introducing additional parameters. Within each subregion, ASA is applied to model the internal confidence and spatial relationships. ASA decomposes the spatial dimensions of the input features into two directions, performs self-attention calculations on each direction separately, and finally adds the outputs of the two directions. This reduces computational complexity while retaining spatial information. In a formal representation, this self-attention can be expressed as:

$$\widehat{Xrow}_{ij} = Attn(X_{ij} W_{ij}^Q, X_{ij} W_{ij}^K, X_{ij} W_{ij}^V) \tag{8}$$

$$\widehat{Xcol}_{ij} = Attn(X_{ij} W_{ij}^Q, X_{ij} W_{ij}^K, X_{ij} W_{ij}^V) \tag{9}$$

$$\widehat{X}_{ij} = \widehat{Xrow}_{ij} + \widehat{Xcol}_{ij} \tag{10}$$

$$\widehat{X}_{i+1} = Concat[\widehat{X}_{ij}]_{j=1:h} W_i^P \tag{11}$$

$$X'_{ij} = X_{ij} + \widehat{X}_{i(j-1)} \tag{12}$$

The output features of self-attention calculations in the row direction are denoted as \widehat{Xrow}_{ij} , while \widehat{Xcol}_{ij} represents the output features in the column direction. The final output feature of axial self-attention is expressed as \widehat{X}_{ij} , where X_{ij} signifies the j -th segment of input features X_i . The total number of heads is represented by h , and W_{ij}^Q , W_{ij}^K , and W_{ij}^V are projection layers that map input features to distinct subspaces. The linear layer W_i^P projects the concatenated output features back to a dimension consistent with the input, ensuring dimensional alignment. The new input feature for the j -th head, denoted as X'_{ij} , replaces X_{ij} , facilitating self-attention to collectively capture local and global relationships, thereby enhancing feature representation. The network is illustrated in Figure 5.

The design choice to split the input features in the CG-AS Attention layer and make only a subset available to each attention

head is motivated by the desire to reduce computational redundancy. Despite the division of features, the cascading nature of the group attention layers ensures that the outputs of previous layers, which contain processed contextual information, are available to subsequent layers. This cumulative processing guarantees that even with the division of input features, the model has access to a comprehensive representation of the context as it progresses through the layers. By partitioning features, each attention head concentrates on different subsets of the input data, enabling the model to process information more efficiently. This method reduces the overlap of computations performed by each head, thus decreasing redundancy. Axial Self-Attention reduces computational load by decomposing the calculations of the traditional self-attention mechanism into two separate axes.

4.3. Adapting YOLOv5

This paper utilizes the improved version of EfficientViT as the backbone network in the YOLOv5 architecture, responsible for extracting feature representations from input images. The network structure is depicted in Figure 6, with the other components of YOLOv5 remaining unchanged.

5. Experimental validation and result analysis

5.1. Dataset and experimental setup

Our experiments focus on fault diagnosis for four states: membrane dry faults, flooding faults, normal states, and unknown states. The dataset consists of 4400 feature curves diagrams, such as Figure 7, representing membrane dry faults, flooding faults, normal states, and unknown states. These curves include the pressure drop at the anode/cathode outlet and inlet, proton exchange membrane (PEM) water content, Membrane Electrode Assembly (MEA) current, and fuel cell current (Redmon and Farhadi, n.d.).

5.2. Experimental equipment

The experimental environment in this paper includes the Windows 10 operating system, NVIDIA GeForce GTX 1050 GPU, i5-7300HQ CPU, and the programming environment consists of Python 3.10, PyTorch 1.12.1, and CUDA 12.2.

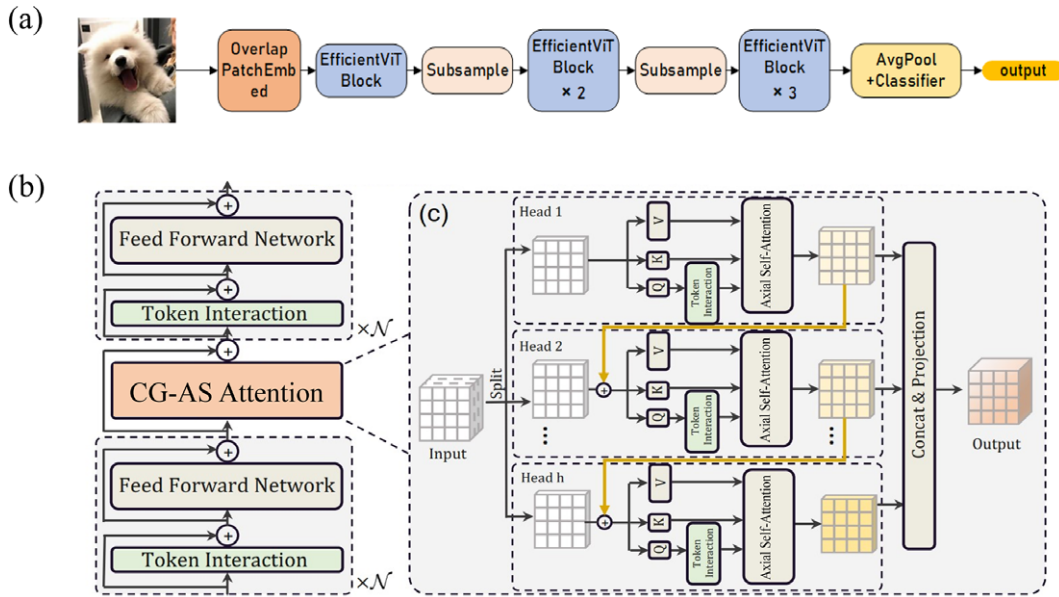


Figure 5. EfficientViT: (a) Flowchart of the EfficientViT, (b) EfficientViT Block, and (c) CG-AS Attention.

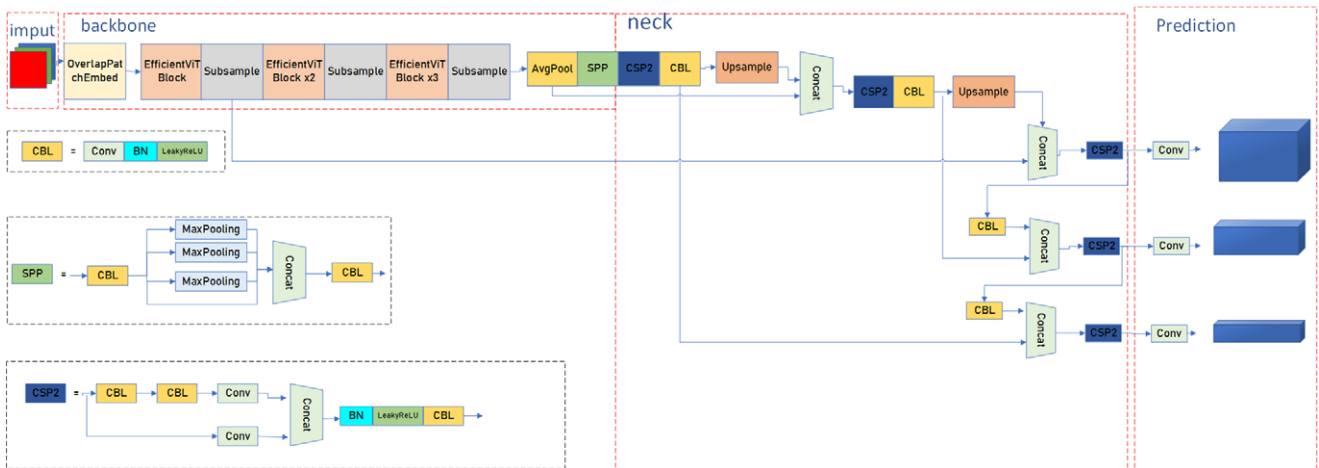


Figure 6. Improved YOLOv5 network model: Utilizing CG-AS Attention improved EfficientViT Network as the Backbone of YOLOv5.

The model training parameters are set as follows: a total of 300 iterations for training, the weight file is efficientViT_m0.pth, the initial learning rate is 0.01, and 4 samples are selected for each training batch.

5.3. Performance metrics

To objectively evaluate the performance of the network model, this study employs Precision, Recall, and mean Average Precision (mAP) as evaluation metrics. Precision refers to the proportion of correctly predicted targets among all targets predicted by the model; recall refers to the proportion of correctly predicted targets among all actual targets; mAP is the average value of prediction accuracy across all categories. (Benjumea et al., n.d.; Li et al., 2023). The PR curves of the YOLOv5s model and YOLOv5-CG-AS are shown in Figure 8 and Figure 9. The PR curve demonstrates the trade-off between precision and recall. By comparing the PR curves of various models, it is possible to intuitively evaluate which model

achieves a better balance between precision and recall. The greater the area under the curve (AUC-PR), the superior the model's performance.

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$mAP = \frac{1}{N} \sum_i^N \frac{TP}{TP + FP} \tag{15}$$

TP represents the number of true positive samples correctly predicted by the model, FP represents the number of false positive samples incorrectly predicted by the model, and FN represents the number of false negative samples incorrectly predicted by the model.

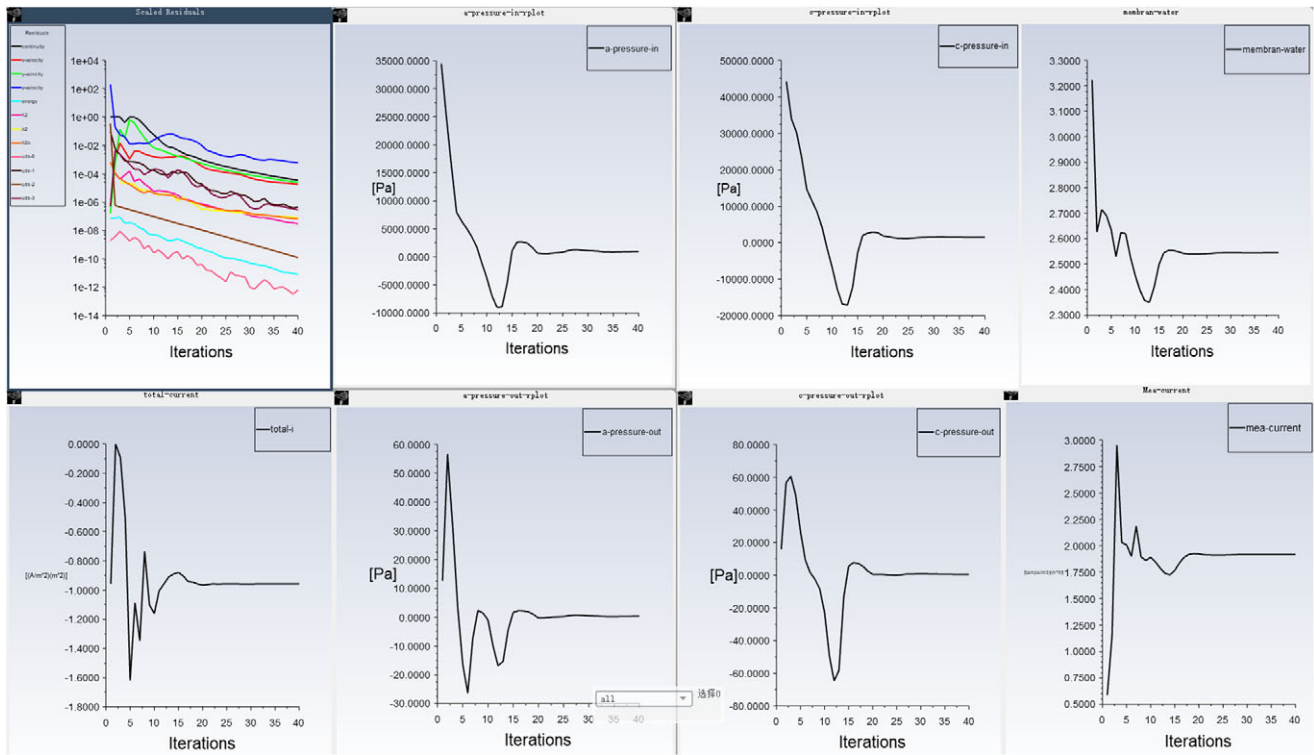


Figure 7. Fuel cell data curve.

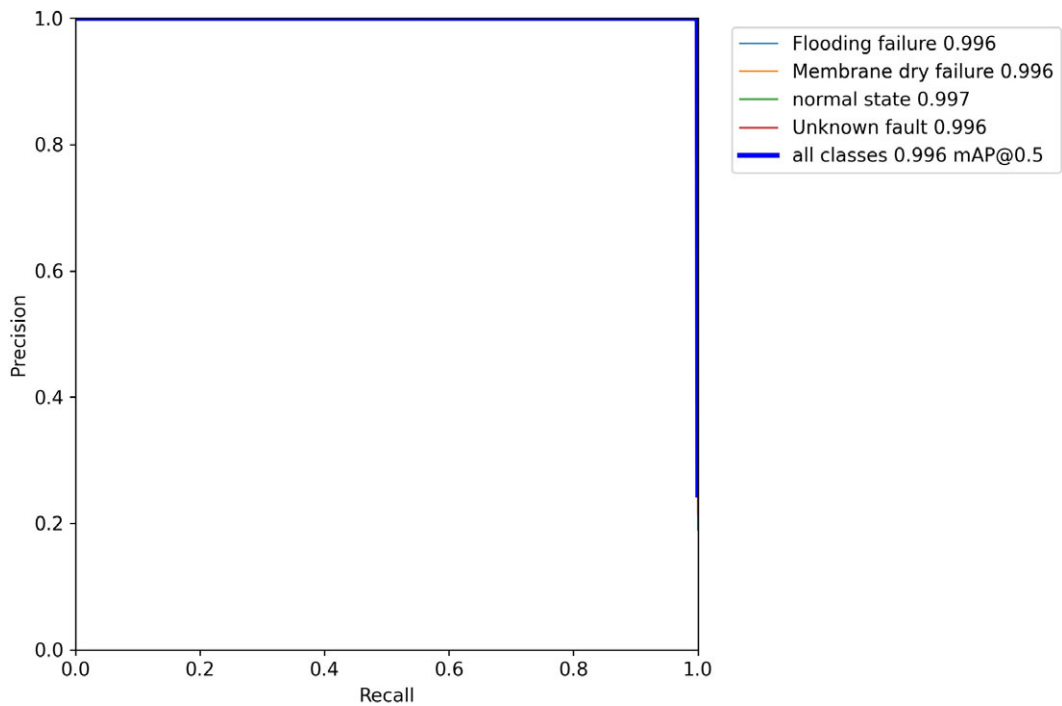


Figure 8. PR curves.

The YOLOv5-CG-AS model has sacrificed a portion of precision to attain a more optimal balance between precision and recall, especially in the context of addressing classification issues within imbalanced datasets. This equilibrium allows the model to not only

detect a greater number of positive class samples but also to minimize the misclassification of negative class samples as positive. This is critically important for applications requiring precise identification, such as medical diagnostics, fraud detection, etc.

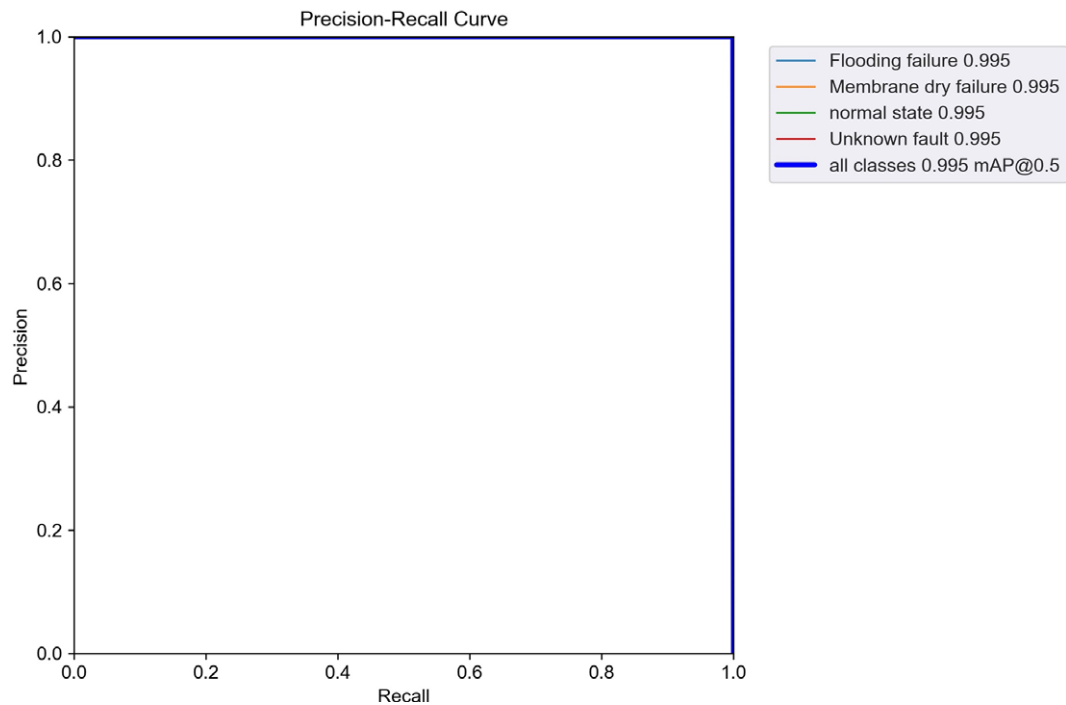


Figure 9. YOLOv5-CG-AS PR curve.

5.4. Experimental results and analysis

5.4.1. Comparison of lightweight models

Because the YOLOv5s model achieved a diagnostic accuracy of 99.6% in the experiment, showing excellent performance with little room for improvement in accuracy, this paper considers lightweight processing for it, as lightweight models are more suitable for embedding in hardware devices.

EfficientViT is used to replace the backbone network of the YOLOv5s model, denoted as YOLOv5- EfficientViT. The EfficientViT model improved with CG-AS Attention is used to replace the backbone network of the YOLOv5s model, denoted as YOLOv5-CG-AS. The popular lightweight network EMO is used to replace the backbone network of the YOLOv5s model, denoted as YOLOv5-EMO, for comparison with the YOLOv5s model (Xu *et al.*, 2020; Xu and Li, 2021). The results are shown in Table 2.

5.4.2. Offline fault diagnosis

Traditional machine learning algorithms such as SVM, KNN, and deep learning algorithms like CNN are widely used in fault diagnosis for bearings, motors, fuel cells, etc., and have shown good diagnostic performance (Xiang *et al.*, 2012; Dang *et al.*, 2020; Kattenborn *et al.*, 2021; Du *et al.*, 2023). Through the comparison of YOLOv5-CG-AS with traditional algorithms, further validation

Table 3. Comparison results of traditional machine learning algorithms

Health status	Diagnostic accuracy/%			
	SVM	CNN	KNN	YOLOv5-CG-AS
Membrane dry failure	86.6	99.8	100	99.5
Flooding failure	88.1	96.7	96.8	99.5
normal state	93.2	100	99.7	99.5
Unknown fault	90.4	98.1	97.8	99.5
Average accuracy	89.6	98.7	98.6	99.5

of the feasibility of this method for fault diagnosis is conducted. The experimental results are shown in Table 3. The results indicate that YOLOv5-CG-AS can be used for offline fault diagnosis, similar to traditional algorithms, and its diagnostic performance is superior to traditional methods.

5.4.3. Online fault diagnosis

In real-life online fault diagnosis, taking the example of an EC-type proton exchange membrane fuel cell stack, multiple sensors are placed at various positions in the battery pack. These sensors

Table 2. Comparison results of lightweight models: compared to the YOLOv5s model, YOLOv5-CG-AS has a slight decrease in mAP by 0.001, but the optimization levels of Parameters, GFLOPs, and model size all exceed 50%. When compared to other lightweight networks, YOLOv5-CG-AS achieves a more balanced optimization across various parameters

Model	Parameters	GFLOPs	FPS	Size/MB	mAP
YOLOv5s	7030417	16.0	41	13.7	0.996
YOLOv5-EfficientViT	3028849(−56.9%)	5.9(−63.1%)	54(+31.7%)	6.51(−52.5%)	0.966(−0.030)
YOLOv5-CG-ASA	2804315(−60.1%)	7.2(−55%)	44(+7.3%)	5.9(−56.9%)	0.995(−0.001)
YOLOv5-EMO	3192821(−54.6%)	25.6(+60%)	37(−9.8%)	6.37(−53.5%)	0.987(−0.009)

transmit real-time data from the battery pack to the Supervisory computer. The YOLOv5-CG-AS algorithm is then employed for fault diagnosis, as shown in Figure 10. In this study, a FLUENT simulation of the PEMFC model is used to obtain real-time curves for various parameters. This simulates the process of sensors obtaining real-time data and transmitting it to the Supervisory computer. The YOLOv5-CG-AS algorithm performs fault diagnosis by capturing real-time images from the Supervisory computer, as shown in Figure 11.

5.4.4. Analysis of experimental results

The YOLOv5-CG-AS algorithm proposed in this paper is compared with the original YOLOv5 model and some popular light-weight algorithms, using criteria such as parameter count, model size, GFLOPs, FPS, and mAP. The YOLOv5-CG-AS algorithm emerges as the optimal algorithm in these comparisons. Compared to machine learning algorithms used in real-life offline fault

diagnosis, the algorithm proposed in this paper achieves a higher diagnostic accuracy. Moreover, this algorithm can be applied to online fault diagnosis, demonstrating fast diagnosis speed and high accuracy. This highlights the superiority of using the YOLOv5-CG-AS algorithm for fuel cell fault diagnosis.

6. Conclusion

In this study, a multiphysics field-coupled PEMFC simulation model in the FLUENT environment is developed to avoid irreversible damage caused by faults in fuel cells. The feasibility of the model is validated through experiments. The YOLOv5-CG-AS algorithm is obtained by replacing the backbone network of the YOLOv5s model with the improved EfficientViT algorithm. This algorithm can be used for both offline and online fault diagnosis of proton exchange membrane fuel cells. In offline fault diagnosis, the algorithm achieves a higher correct diagnosis rate compared to

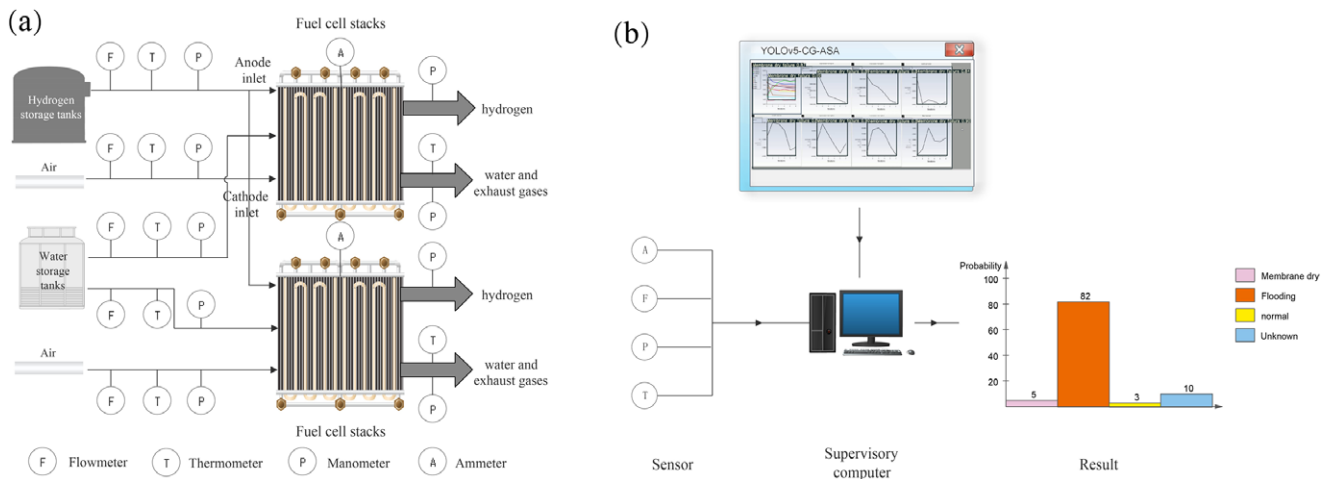


Figure 10. Online fault diagnosis: (a) schematic diagram of the actual PEMFC system and sensor placement configuration and (b) flowchart of online diagnosis.

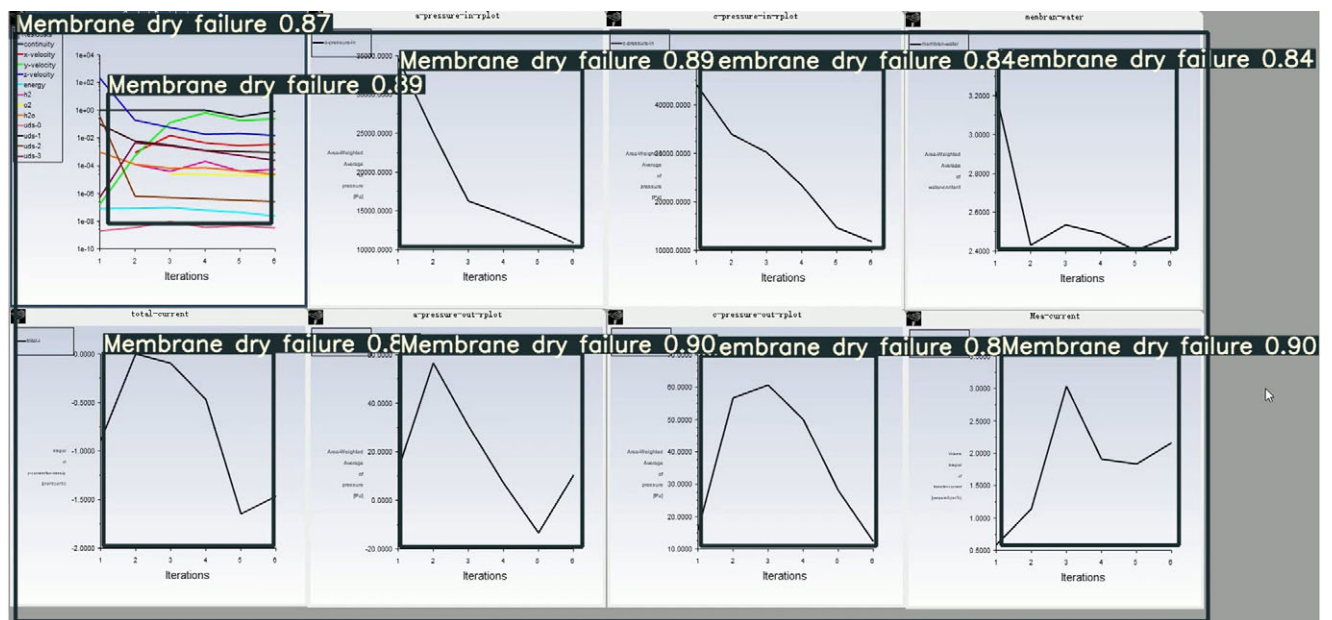


Figure 11. Online fault diagnosis diagram: through online monitoring of 200 faults, it is found that when the average FLUENT simulation is iterated six times, the YOLOv5-CG-ASA algorithm can diagnose the correct fault type with a diagnostic accuracy of 99.5%.

machine learning algorithms already in use. In online fault diagnosis, the algorithm demonstrates fast diagnosis speed and high accuracy.

The YOLOv5-CG-AS algorithm has the advantages of low parameter count, small size, fast response, high accuracy, and can perform fault diagnosis both offline and online. It holds the potential to be embedded in hardware devices and used as a mobile fault diagnosis tool. This has significant implications for advancing fault diagnosis methods for PEMFC and promoting further development in the field. However, the current YOLO algorithm has a significant drawback in the information fusion method at the neck, hindering effective cross-layer integration. Future work will focus on optimizing the model structure and improving the information fusion method at the neck, potentially using Gold-YOLO to enhance performance.

References

- Allam A, Michael M, and Zhang P. Fault diagnosis of proton exchange membrane fuel cells. *2021 5th International Conference on Control and Fault-Tolerant Systems (SysTol)*. IEEE, Saint-Raphael, France, 29 September 2021.
- Benjumea A, Teeti I, Cuzzolin F, et al. YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles. *arXiv preprint arXiv: 2112.11798*, 2021.
- Calasan M, Micev M, Hasanien HM, et al. (2024) PEM fuel cells: Two novel approaches for mathematical modeling and parameter estimation. *Energy* **290**, 130130
- Cui D and Xiang Z (2023) Nonsingular fixed-time fault-tolerant fuzzy control for switched uncertain nonlinear systems. *IEEE Transactions on Fuzzy Systems* **31**(1), 174–183
- Dang H, Ma R, Zhou J, et al. (2020) CNN-based fault diagnosis of proton exchange membrane fuel cell. *Power Electronics Technology* **54**(12), 52–55
- Darzentas D, Wagner H, Craigon P, et al. (2022) Data-inspired co-design for museum and gallery visitor experiences. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* **36**, e3
- Du D, Sheng Y, Zhao H, et al. (2023) Fault diagnosis of proton exchange membrane fuel cell system based on VMD. *Control Engineering of China* **30** (07), 1190–1197
- Gao Y, Liu X and Xiang J (2021) Fault detection in gears using fault samples enlarged by a combination of numerical simulation and a generative adversarial network. *IEEE/ASME Transactions on Mechatronics* **27**(5), 3798–3805
- Guo B, Zhu H, Wei Y, et al. (2014) 3-D steady-state numerical simulation of proton exchange membrane fuel cells. *Computer and Applied Chemistry* **31** (2), 135–139
- Ho J, Kalchbrenner N, Weissenborn D, et al. *Axial attention in multidimensional transformers*. *arXiv preprint arXiv: 1912.12180*, 2019.
- Huang F, Geng L, Liu F, et al. A review of water fault diagnosis of the proton exchange membrane fuel cell. *2023 IEEE 7th Information Technology and Mechatronics Engineering Conference (ITOEC)*. IEEE, Chongqing, China, 15 September 2023.
- Huang S and Xiang Z (2015) Finite-time stabilization of a class of switched stochastic nonlinear systems under arbitrary switching. *International Journal of Robust and Nonlinear Control* **26**(10), 21362152
- Jeong H, Park B, Park S, et al. (2019) Fault detection and identification method using observer-based residuals. *Reliability Engineering and System Safety* **184**, 27–40
- Kattenborn T, Leitloff J, Schiefer F, et al. (2021) Review on convolutional neural networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* **173**, 24–49
- Kim J, Kim N, Park Y, et al. (2022) Object detection and classification based on YOLO-V5 with improved maritime dataset. *Journal of Marine Science and Engineering* **10**(3), 377
- Lee B, Cooper R, Hands D, et al. (2022) Continuous cycles of data-enabled design: Reimagining the IoT development process. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* **36**, e11
- Li K, Wang J, Jalil H, et al. (2023) A fast and lightweight detection algorithm for passion fruit pests based on improved YOLOv5. *Computers and Electronics in Agriculture* **204**, 107534
- Liu G, Hu Y, Chen Z, et al. (2023) Lightweight object detection algorithm for robots with improved YOLOv5. *Engineering Applications of Artificial Intelligence* **123**, 106217
- Liu X, Peng H, Zheng N, et al. EfficientViT: Memory efficient vision transformer with cascaded group attention. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17 June 2023*.
- Luo Y, Kumar A and Xiang J (2022) Machinery fault diagnosis based on domain adaptation to bridge the gap between simulation and measured signals. *IEEE Transactions on Instrumentation and Measurement* **71**, 1–9
- Ma R, Dang H, Zhang Y, et al. (2024) A review on failure mechanism analysis and diagnosis for proton exchange membrane fuel cell systems. *Proceedings of the CSEE* **44**(01), 407–427
- Redmon J, Farhadi A. *Yolov3: An incremental improvement*. *arXiv preprint arXiv: 1804.02767*, 2018.
- Sinha V and Mondal S (2021) Adaptive unknown input observer approach for multi-fault diagnosis of PEM fuel cell system with time-delays. *Journal of Control and Decision* **8**(2), 222–232
- Sun Y, Mao L, Huang W, et al. (2022) Fault diagnosis of proton exchange membrane fuel cell using magnetic field data. *Journal of Mechanical Engineering* **58**(22), 106–114
- Wang J, Yang B, Zeng C, et al. (2021) Recent advances and summarization of fault diagnosis techniques for proton exchange membrane fuel cell systems: A critical overview. *CIESC Journal* **500**, 229932
- Xiang J, Zhong Y and Tang H (2016) A novel personalized diagnosis methodology using numerical simulation and an intelligent method to detect faults in a shaft. *Applied Sciences* **6**(12), 414
- Xiang Z, Qiao C and Mahmoud M (2012) Finite-time analysis and H^∞ control for switched stochastic systems. *Journal of the Franklin Institute* **349**(3), 915–927
- Xu H, Ai Z and Zhang Z (2020) A lightweight target detection network based on channel rescheduling. *Computer and Modernization* (02), 94–98
- Xu Y and Li C (2021) Lightweight object detection network based on YOLO optimization. *Computer Science* **48**(S2), 265–269
- Zhang X and Guo X (2021) Fault diagnosis of proton exchange membrane fuel cell system of tram based on information fusion and deep learning. *International Journal of Hydrogen Energy* **46**(60), 30828–30840
- Zhang X, Jiang Y, Sun T, et al. (2020) Review of flooding and membrane drying fault diagnosis in proton exchange membrane fuel cells. *Journal of Southwest Jiaotong University* **55**(04), 828–838
- Zhang Y, He Y, Tang H, et al. 2024 Adversarial domain adaptation approach for axial piston pump fault diagnosis under small sample condition based on measured and simulated signals. *IEEE Transactions on Instrumentation and Measurement*.
- Zheng Z, Petrone R, Péra MC, et al. (2013) A review on non-model based diagnosis methodologies for PEM fuel cell stacks and systems. *International Journal of Hydrogen Energy* **38**(21), 8914–8926
- Zhong Z, Zhu X and Cao G (2006) Modeling a PEMFC by a support vector machine. *Journal of Power Sources* **160**(01), 293–298
- Zhou S, Ren H and Pei F (2015) Research on PEMFC fault diagnosis method based on COMSOL model. *Chemical Industry and Engineering* **32**(4), 56–62