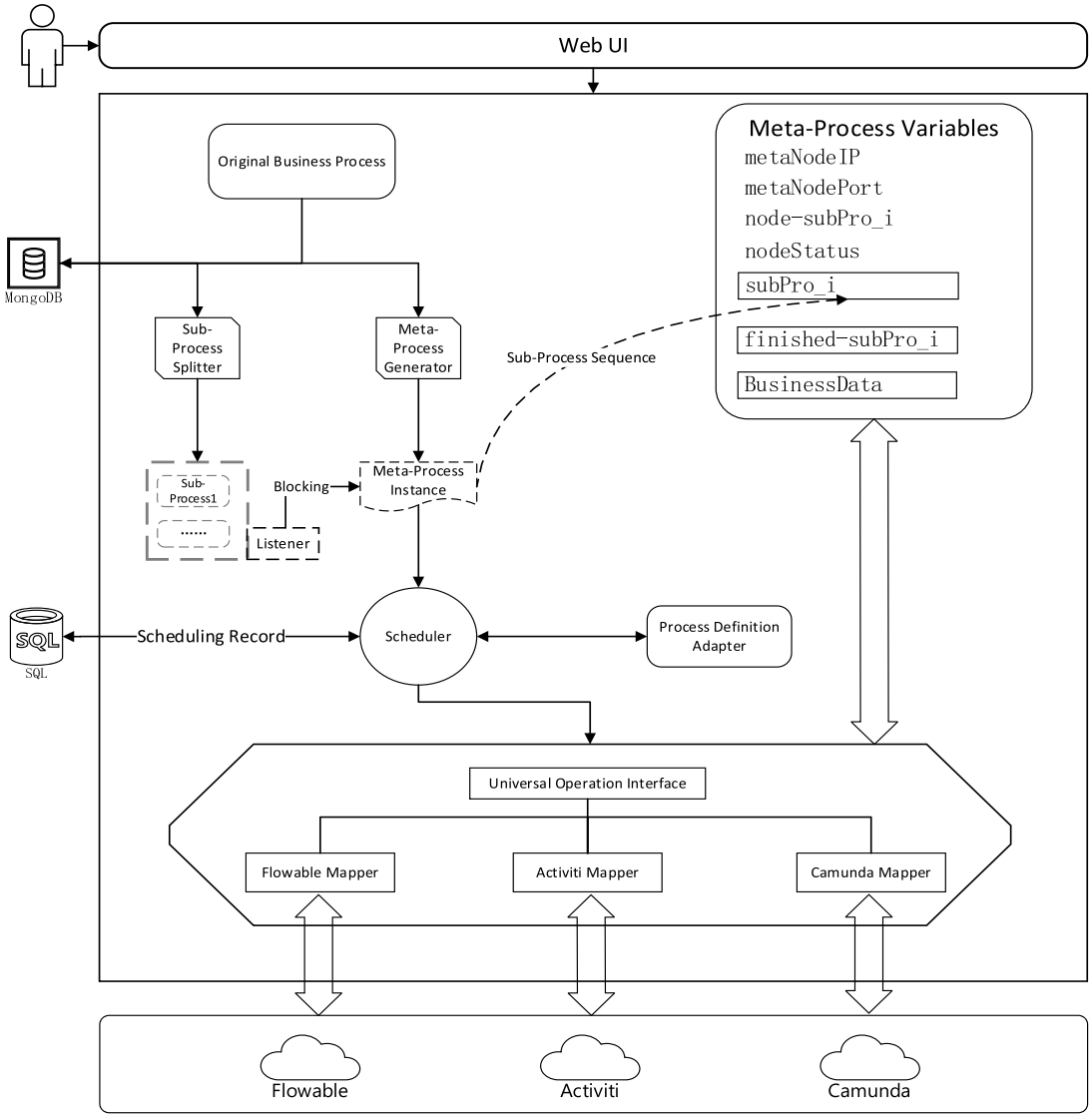


Figure 1. System architecture diagram for heterogeneous engine collaboration

3.1. Related definitions

In order to support distributed workflow scheduling, this paper is based on the "meta-process" method, which splits the original process into fine grains, each service task is split into independent sub-processes, and a listener is added to each sub-process to listen to its completion status. The generated sub-processes are stored as XML strings in meta-process variables for meta-process scheduling. Next, we will introduce the three key concepts of sub-process generation, meta-processes, and meta-process variables in detail.

Definition 1. Subprocess: the subprocess splitter (see **Figure 2**) represents the original process as a directed graph *G(V,E)*, where *V* is the set of nodes and *E* is the set of edges. For each service task node *v_task* in *G*, it is fine-grained partitioned into a set of sub-processes, sub-process *Sub_i* is represented as a directed graph $G\_task\left(V\_task, E\_task\right)$:

$$V\_task = \left\{v\_start, v\_task, v\_end\right\},\ E\_task = \left\{e\_start\_task, e\_task\_end\right\}$$

A listener node *v_listener* is inserted into *G_task*, which is connected to *v_task* and *v_end* to form an extension of *E_task*. The generated collection of sub-processes is stored as XML strings in the original process variables. The model ensures sequential process execution and data consistency.

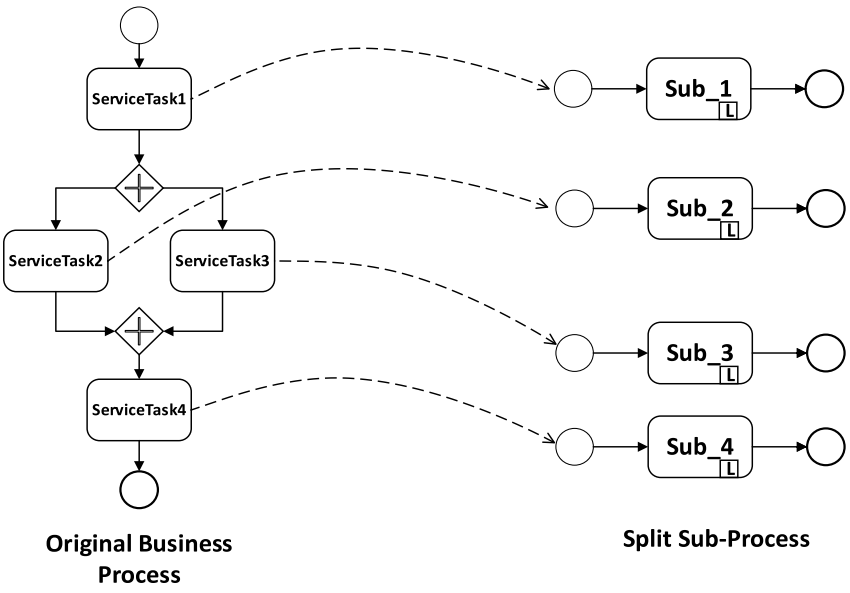


Figure 2. Sub-process splitter

Definition 2. Meta-process: meta-process *M*, as a control process, consists of a set of service tasks $C = \{C_1, C_2, ..., C_n\}$, which is responsible for coordinating and controlling the execution order of sub-process. *M* coordinates the execution order of *Sub_i* by transferring information through process variables. Each service task *Ci* corresponds to a sub-process *Sub_i.* When *M* is started, *Ci* will obtain the relevant information of *Sub_i* from the process variables of the meta-process, including its process model and deployment node information. After the successful deployment of sub-process *Sub_i*, *M* will block at the corresponding user task *Ui* and wait for the completion of *Sub_i.* When the execution of *Sub_i* is completed, the listener will be triggered, and the role of the listener is to complete the user task *Ui*, thus unblocking the meta-process.

In this way, the meta-process can ensure that the execution order of the sub-processes is consistent with the original process. In addition, the meta-process realizes the state synchronization and data transfer with the sub-processes through the transfer of process variables, which ensures the correctness and completeness of the execution of the whole workflow.

Definition 3. Meta-process variables: meta-process variables, including control data and business data, are the medium of communication between engine nodes, and are used to transfer information and state in different parts of the workflow, coordinate and control the execution of the workflow, and accompany the whole life cycle of the workflow. The data structure of meta-process variables is shown in **Figure 3**.

After the meta-process is started, the control task *Ci* reads the process definition of the process of the sub-process *Sub_i* and the deployment node information of the sub-process in the process variable of the meta-process instance. ci carries the information obtained above and deploys it to the engine *node_i* through the common operation interface. At this point the control task *Ci* of the meta-process is completed and blocked at the user task *Ui*. When the sub-process *Sub_i* is completed a completion marker is inserted into the meta-process variable and

triggers the listener *L*. The listener completes the user task *Ui* through the generic operation interface and unblocks the meta-process. Subsequently the sub-process *Sub_i+1* is deployed until the meta-process completes all control tasks. The process variable design is shown in **Figure 3**.

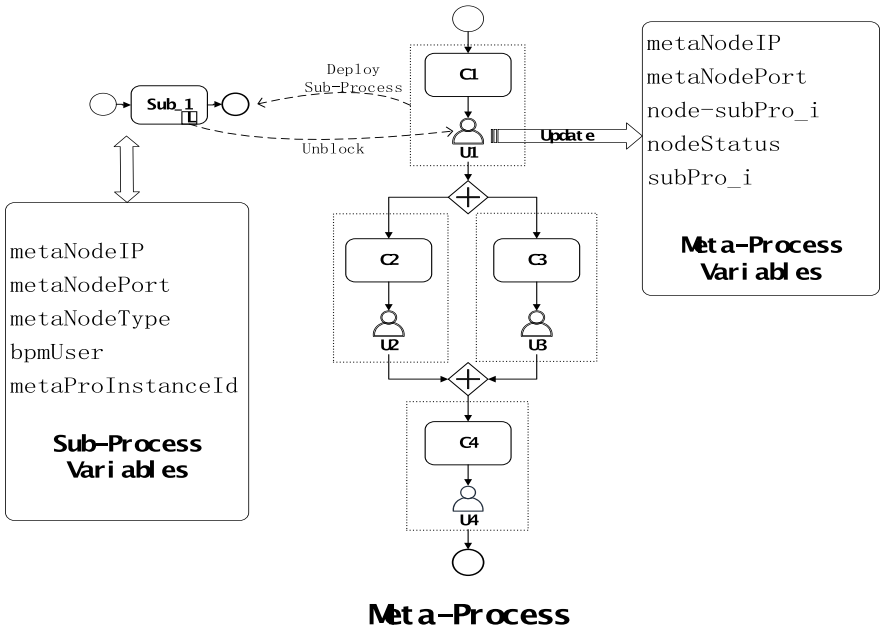


Figure 3. Process Variables design

4. Experimental analysis

4.1. Preparation

Consisting of two Flowable nodes and two Activiti nodes, the heterogeneous environment in the cross-organizational case is simulated, 300 business processes are randomly generated, and one sample is taken as a case sample (see **Figure 4**) for detailed analysis:

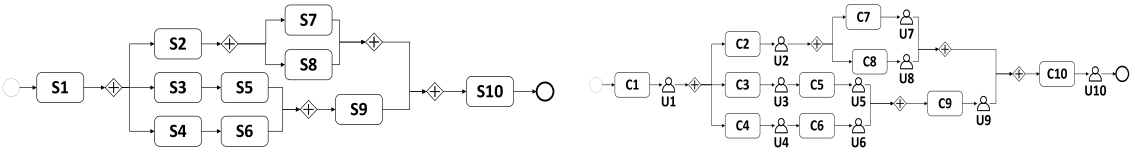


Figure 4. Experimental sample & sample meta-process

4.2. Experimental processes

The 300 business processes are deployed sequentially, and their execution status is observed. The deployment status and execution progress of each business process can be observed in the user portal platform. Take the case sample as an example, it is partitioned into 10 sub-tasks, and the control task in the meta-process will deploy these sub-processes according to the state of the engine node. The deployment of sub-processes is shown in **Table 1**.

Table 1. Results of sub-process deployment

Sub-task ID	Node	Node engine type
subPro_1	Wuhan:10.5.83.189:7001	activiti
subPro_2	Beijing:10.5.83.189:7000	activiti
subPro_3	Wuhan:10.5.83.189:7001	activiti
subPro_4	Wuhan:10.5.83.189:7001	activiti
subPro_5	Wuhan:10.5.83.189:7001	activiti
subPro_6	HangZhou:10.5.83.189:8282	flowable
subPro_7	Wuhan:10.5.83.189:7001	activiti
subPro_8	Beijing:10.5.83.189:7000	activiti
subPro_9	Wuhan:10.5.83.189:7001	activiti
subPro_10	Shanghai:10.5.83.189:8181	flowable

4.3. Experimental results and analysis

The experimental results show that 300 experimental samples were deployed to the four heterogeneous nodes, see **Figure 5**. The experiments demonstrate that the method can effectively improve inter-organizational process execution efficiency, ensure a consistent process execution order, and achieve the collaboration of heterogeneous process engines when dealing with cross-organizational business processes.

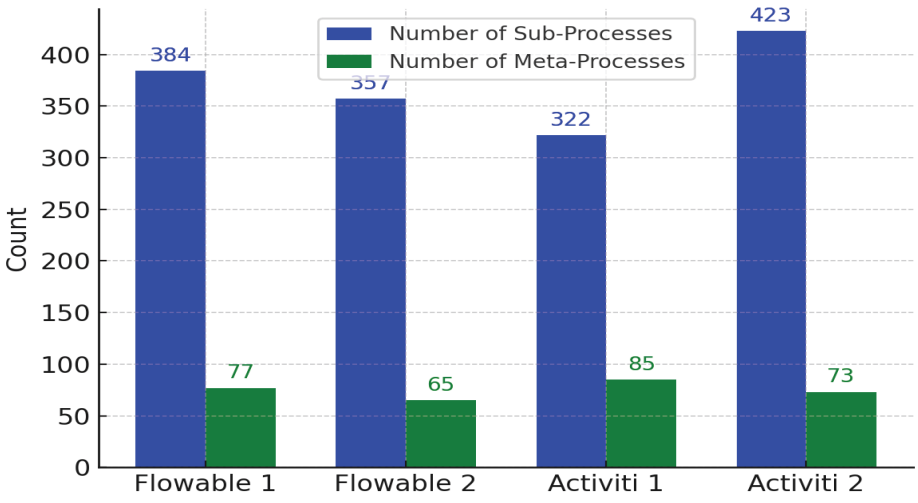


Figure 5. Deployment details of sub-processes and meta-processes for 300 business processes

5. Conclusion

In this paper, we have proposed a collaborative approach for a distributed heterogeneous process engine for cross-organizational purposes. The approach uses the meta-process as a control process and splits the original process into sub-processes. The proposed process model adapter and common operation interface achieve efficient collaboration between different process engines. In the future, we will further optimize the collaboration method of distributed heterogeneous process engines and explore more intelligent technical means, such as artificial

intelligence, to achieve more intelligent and efficient process management.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Xu X, Sheng QZ, Zhang LJ, et al., 2015, From Big Data to Big Service. Computer, 48(7): 80–83.

[2] Aouachria M, Leshob A, Ghomari AR, et al., 2024, A Process Mining Method for Inter-Organizational Business Process Integration. ACM Transactions on Management Information Systems, 15(1): 1–29.

[3] Gonzalez L, Delgado A, 2021, Compliance Requirements Model for Collaborative Business Process and Evaluation with Process Mining, 2021 XLVII Latin American Computing Conference (CLEI), 1–10.

[4] Norta A, Grefen P, Narendra NC, 2014, A Reference Architecture for Managing Dynamic Inter-Organizational Business Processes. Data & Knowledge Engineering, 91: 52–89.

[5] Pena L, Andrade D, Delgado A, et al., 2023, An Approach for Discovering Inter-Organizational Collaborative Business Processes in BPMN 2.0, International Conference on Process Mining, Oct 23, 2023, 487–498.

[6] Shan Z, Chiu DK, Li Q, 2005, Systematic Interaction Management in a Workflow View-Based Business-to-Business Process Engine, Proceedings of the 38th Annual Hawaii International Conference on System Sciences, Jan 6, 2005, 162.

[7] Verginadis Y, Mentzas G, 2008, Agents and Workflow Engines for Inter-Organizational Workflows in E-Government Cases. Business Process Management Journal, 14(2): 188–203.

[8] Osorio AL, Camarinha-Matos LM, 2008, Distributed Process Execution in Collaborative Networks. Robotics and Computer-Integrated Manufacturing, 24(5): 647–655.

[9] Fang L, Yang ZG, Qin S, et al., 2021, Meta-Process: A Novel Approach for Decentralized Execution of Process, 2021 International Conference on Service Science (ICSS), 38–44.

[10] Domingos D, Martinho R, Candido C, 2013, Flexibility in Cross-Organizational WS-BPEL Business Processes. Procedia Technology, 9: 584–595.

[11] Klinger P, Bodendorf F, 2020, Blockchain-Based Cross-Organizational Execution Framework for Dynamic Integration of Process Collaborations, Wirtschaftsinformatik (Zentrale Tracks) Mar 8, 2020, 1802–1817.

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

The Design and Application of a Mobile Sound Source Localization System

Yue Kan[1], Tengfei Zhang[1]*, Fusheng Zha[2]

[1]School of Mechanical and Power Engineering, Henan Polytechnic University, Jiaozuo 454150, Henan, China
[2]State Key Laboratory of Robotics and Systems, Harbin Institute of Technology, Harbin 150001, China

Corresponding author: Tengfei Zhang, fz2758526@gmail.com

Abstract: The mobile sound source localization system is a technology that can track and locate mobile sound sources in real time and has broad application prospects in many fields. This article first provides an overview of the mobile sound source localization system, introducing its concept and composition, as well as its design and application significance. It elaborates on the importance of the mobile sound source localization system from multiple aspects, such as safety, production, and daily life, and deeply explores its design and application strategies. The problems faced by the mobile sound source localization system and its future development direction were pointed out.

Keywords: Mobile sound source; Localization system; Design application

Online publication: October 17, 2025

1. Introduction

In modern society, sound, as an important carrier of information, contains rich information. Abnormal sounds emitted by equipment in industrial production and the determination of the location of sound sources in public security, such as people's cries for help and explosions, are of vital importance. With the increasing mobility of sound sources, fixed sound source localization systems can no longer meet the actual needs, and thus, mobile sound source localization systems have emerged. The mobile sound source localization system can follow the moving sound source for positioning and is no longer limited to fixed positioning. This paper mainly focuses on the research of mobile sound source localization systems. Through the analysis and study of various aspects of mobile sound source localization systems, it provides references for related research and applications.

2. The current design status of the mobile sound source localization system

2.1. The development of localization algorithms

Localization algorithm is the main theoretical basis of the mobile sound source localization system. The

development of a localization algorithm directly affects the positioning accuracy and efficiency of the system. Most early localization algorithms relied on a single parameter for positioning. The Time of Arrival (TOA) algorithm determines the position of the sound source by measuring the absolute time it takes for the sound source signal to reach the sensor. This algorithm has a high requirement for time synchronization and is easily affected by clock errors in actual mobile environments, which limits its application. With the deepening of research, the Time Difference of Arrival (TDOA) algorithm has gradually become mainstream. The TDOA algorithm determines the position of the sound source by using the time difference of the sound source signal reaching different sensors, reducing the requirement for absolute time synchronization and demonstrating better stability in mobile environments. However, when the sound source moves rapidly, the positioning error of the TDOA algorithm will increase due to the dynamic changes in signal propagation time. Therefore, researchers have proposed a dynamic TDOA correction model. By introducing the parameter of the sound source's movement speed, the time difference is compensated in real time, thereby improving the positioning accuracy of high-speed moving sound sources [1].

2.2. Theoretical research on noise and multipath interference processing

In the localization of mobile sound sources, noise and multipath interference are the main factors affecting positioning accuracy. Relevant theoretical research has always been a focus. Regarding environmental noise, in the early days, filtering algorithms were mainly used for suppression. Adaptive filtering algorithms can dynamically adjust filtering parameters based on noise characteristics, thereby removing stationary noise. However, in mobile scenes, noise often has non-stationary characteristics. The effect of traditional filtering algorithms is poor. To overcome the above problems, the noise suppression theory based on wavelet transform emerged. It uses multi-scale analysis to concentrate noise in the high-frequency band for suppression, while the sound source signal is concentrated in the low-frequency band. In this way, noise is effectively removed while maintaining the important components of the sound signal, and it performs well when dealing with impulse noise and burst noise. In addition, noise suppression methods based on sparse representation have emerged. This method regards the sound source signal as a sparse signal and then uses a sparse reconstruction algorithm to recover a clean signal from the noisy signal. This not only improves the accuracy of noise suppression but is also suitable for the localization of mobile sound sources with relatively low signal-to-noise ratios. Multipath interference occurs when the sound source signal, after being reflected and refracted by obstacles during its propagation, reaches the sensor, causing multiple path signals to superimpose and resulting in distortion in the measurement of positioning parameters. The multipath compensation theory based on the propagation model has also been developed. By establishing the propagation model of sound in complex environments, the propagation path and attenuation of multipath signals are predicted, and the measured positioning parameters are compensated and corrected to improve the positioning accuracy. However, this theory has relatively high requirements for environmental modeling. When the environment where the moving sound source is located is constantly changing, the applicability of the model needs to be improved.

2.3. Theoretical research on cooperative localization of mobile platforms

Mobile sound source localization systems generally require installation on mobile platforms, such as robots and unmanned aerial vehicles, and the mobility of mobile platforms also poses new requirements for positioning theory. The motion error of the mobile platform can lead to inaccurate sensor position, thereby causing positioning error. Therefore, the theory of platform motion state estimation and compensation has become the focus of research. The platform positioning theory, based on the inertial Navigation System (INS), measures the

acceleration and angular velocity of the platform to estimate its position and attitude in real time. However, it has cumulative errors, which will increase significantly after long-term operation. Therefore, the integrated navigation theory combining GPS and INS is adopted. The absolute position information of GPS can correct the cumulative errors of INS. INS can make up for the problem of signal loss of GPS in occluded environments, providing stable and reliable location information for mobile platforms, thereby ensuring the positioning reference accuracy of the sensor array [2]. In terms of multi-platform collaborative localization, the distributed positioning theory has become relatively mature. Multiple mobile platforms share the sound source information they measure through wireless communication and jointly solve the sound source localization using distributed estimation algorithms. This not only expands the positioning range but also reduces the uncertainty of single-platform positioning through the data fusion of multiple platforms. However, when the distributed positioning theory encounters data synchronization, how to achieve efficient collaboration while ensuring real-time positioning and avoiding communication delay is also a point that needs to be deeply explored.

3. The design significance of the mobile sound source localization system

3.1. Ensuring public safety

The use of mobile sound source localization systems in public security is of great significance. For example, in the event of natural disasters such as earthquakes and fires, the affected areas are often highly complex environments with limited visibility, so victims' cries for help cannot be heard easily. However, this mobile sound source localization system can promptly locate the sounds made by trapped victims by crossing smoke barriers and other means, and provide centimeter-level position information. It has greatly shortened the rescue time and increased the possibility of a successful rescue. In anti-terrorism and anti-violence, what is encountered are cunning terrorist criminals. Their gunshots, explosion sounds, etc., are all key clues for quickly identifying the criminals. The system can continuously track and locate the sounds, allowing the police to make rapid arrests based on the movement trajectory of the sound source, minimizing the injury and death of innocent people to the greatest extent. It has effectively protected the safety of lives and property. When holding major meetings or other events, we can also promptly hear abnormal sounds from the crowd, such as quarrels and cries for help, and intervene in advance to prevent more serious incidents.

3.2. Improving industrial production efficiency

In the industrial production process, whether the equipment can operate normally is directly related to the level of production efficiency. The mobile sound source localization system can conduct real-time monitoring of mobile production equipment and promptly detect abnormal sounds produced by the equipment, such as high-frequency abnormal sounds caused by bearing wear and periodic noises caused by poor gear meshing. By precisely locating these abnormal sound sources and analyzing their spectra, the specific location and severity of equipment failures can be identified, thereby enabling early prediction of equipment failure conditions. This facilitates targeted maintenance and repair work by maintenance personnel, preventing sudden equipment shutdowns that could lead to production line disruptions. For instance, in an automotive manufacturing workshop, if a robot's mechanical arm joint malfunctions during mobile operations, the system will immediately locate and issue an alarm. Maintenance personnel can handle the issue promptly before it worsens, reducing equipment downtime and enhancing production efficiency. Moreover, this system can also record and analyze the operation of the equipment for a long time, providing data support for the formulation of equipment maintenance plans and improving the

production process [3].

3.3. Improving traffic management

Traffic management is a key part of urban management. The mobile sound source localization system also plays a significant role in traffic management. Taking urban roads as an example, the collision sounds and sudden braking sounds of vehicles produced during traffic accidents are often direct signals of accidents. This system can quickly determine the specific lane and location where the accident occurred. Through the linked monitoring system, the traffic management department can immediately dispatch police forces and rescue resources to the scene, thereby preventing prolonged traffic congestion time due to information lag. The frequent occurrence of illegal honking in cities has a very low manual efficiency in traditional supervision methods. However, this system can accurately determine the location and time of the honking vehicle, and can also be linked with electronic police to record violation information, forming a deterrent, reducing noise pollution, and improving the traffic civilization level of the city. It can also detect abnormal sounds emitted by vehicle malfunctions on highways. For instance, if there are sounds like tire blowouts or abnormal noises from the engine, the system can promptly notify the patrol personnel to handle them and prevent secondary accidents.

3.4. Promoting the development of scientific research

The mobile sound source localization system provides a very powerful tool for related scientific research. In acoustic research, this system can be used to explore the propagation characteristics of sound in various environments, such as reflection and diffraction rules in complex terrains and building complexes, etc., thereby providing a lot of precise experimental data for the development of acoustic theory. Promoting the development of new acoustic materials and sound insulation technologies, in the field of biological research, researchers can use this system to trace the moving sound sources of wild organisms, such as the calls of birds and the sounds of animal activities, thereby understanding their activity range, migration routes, and social behaviors, and providing key information for animal protection and ecological environment research.

4. Design and application strategies of mobile sound source localization systems

4.1. Hardware selection

Hardware is the foundation of a mobile sound source localization system, and the selection of hardware will directly affect the performance of the system. When selecting hardware, the following points should be noted: The microphone array is an important component for collecting sound source signals. Its performance directly affects the quality of the collected signals. Microphones with high sensitivity, low noise, and wide frequency bands should be selected. The size and arrangement of the microphone array should also be determined according to the requirements of the actual application scenario. If a high positioning accuracy is to be achieved, larger-scale uniform linear arrays or circular arrays can be adopted. The signal acquisition card is used to convert the analog signals collected by the microphone array into digital signals. Its sampling rate and resolution must meet the requirements of the system. The higher the sampling rate, the wider the frequency range of the collected signals. The higher the resolution, the higher the quantization accuracy of the signals. The mobile platform is the carrier on which the entire system relies. The mobile platform should be selected based on the application scenarios, such as robots, drones, vehicles, etc. The mobile platform should have good mobility, stability, and battery life to adapt to different terrains and environments. The processor is used to process and analyze the collected signals.

Its computing speed and performance will affect the real-time performance of the system. High-performance embedded processors or digital signal processors should be selected to meet the system's requirements for real-time signal processing.

4.2. Algorithm design

The algorithm is the heart of the mobile sound source localization system. Its quality directly determines the positioning accuracy and reliability of the system. In the design of the algorithm, the following points should be considered: Sound source localization algorithm. Common sound source localization algorithms include the TDOA algorithm, the AOA algorithm, the beamforming algorithm, etc. The TDOA algorithm locates the position of the sound source by measuring the time difference of the sound source signal reaching each microphone, and has a relatively high positioning accuracy. The AOA algorithm determines the position of the sound source by measuring the angle at which the sound source signal reaches the microphone, and is suitable for short-range positioning. The beamforming algorithm forms a beam directed towards the sound source by weighting the output signal of the microphone array, and it has a good anti-interference ability. In practical applications, it is necessary to select the appropriate sound source localization algorithm or combine multiple algorithms based on specific application scenarios and requirements to enhance the system's location performance. Noise suppression algorithms: In actual environments, there are numerous noise interferences, such as environmental noise and equipment noise. These noises can affect the collection and processing of sound source signals, thereby reducing the positioning accuracy of the system. Therefore, it is necessary to design effective noise suppression algorithms, such as adaptive filtering algorithms and wavelet transform algorithms, to remove noise interference and improve the signal-to-noise ratio of the sound source signal. Multipath effect processing algorithm: Multipath effect refers to the situation where the signal emitted by the sound source reaches the microphone array through different paths during transmission, causing reflection, refraction, and other conditions, which in turn affect the accuracy of sound source localization. To address the multipath effect, one can adopt a multipath suppression algorithm relying on the signal model.

4.3. Software development

Software is an important component of the mobile sound source localization system, mainly used to control the entire system, process data, display information, etc. When writing software, the following points should be considered: Operating system. An operating system suitable for the mobile platform should be selected, such as Linux or Android. These systems are stable and expandable, meeting the real-time requirements of the system. The data processing module is used to process and analyze the collected signals, involving functions such as signal preprocessing, feature extraction, and sound source localization. It is necessary to adopt efficient data processing algorithms and programming techniques to accelerate the data processing speed and improve the processing efficiency. The control module is responsible for controlling the mobile platform and hardware devices, enabling the system to move autonomously and switch working modes. A stable and reliable control algorithm needs to be designed to ensure the safety of the system. The display module is used to show the real-time location information of the sound source, signal strength, and other parameters, which is convenient for the operator to understand the working condition of the system. A friendly interface needs to be designed to enhance the usability of the system [4].

4.4. System integration

System integration is the organic combination of various components, such as hardware and software, to form

a complete mobile sound source localization system. When integrating the system, the following points should be taken into consideration: the interface between hardware and software, which should ensure compatibility between the software and hardware interfaces to achieve precise data transmission and effective execution of control commands [5]; the stability and reliability of the system; the system integration needs to be fully tested and debugged to ensure stable and reliable operation in different environments. The power consumption of the system: Mobile sound source localization systems are generally powered by batteries. Therefore, to reduce the power consumption of the system and extend its battery life, the hardware design and software algorithms can be improved to cut down on unnecessary energy consumption.

5. Conclusion

In conclusion, as a new technology, the mobile sound source localization system has a wide range of applications and a promising development prospect. The research on this system is of great significance. The mobile sound source localization system will play a greater role in public safety, industrial production, traffic management, scientific research, and other aspects in the future, making tremendous contributions to the development and progress of society. It will play an increasingly important role in future life. Through the continuous efforts and innovations of scientists, the mobile sound source localization system will be continuously improved and developed. It brings convenience and security to our lives and work.

Funding

This project was supported by the National Natural Science Foundation of China (U2013602).

Disclosure statement

The authors declare no conflict of interest.

References

[1] Tong H, Liu Y, Yi W, et al., 2023, Accurate Measurement of Underwater Sound Sources Based on Beamforming. Metrology Science and Technology, 67(12): 27–33 + 66.

[2] Tan H, Bian P, Xu B, et al., 2023, Influencing Factors of Sound Source Localization Ability in Patients with Cochlear Implant. Chinese Journal of Otolaryngology, 21(06): 904–909.

[3] Liu Q, Yan S, Liu B, et al., 2023, Design and Research of Sound Source Localization Method and Device. Physics and Engineering, 33(06): 60–69.

[4] Wei Z, 2023, Design of PWM Acoustic Source Location and Tracking System Based on K210 Single-Chip Microcomputer Control. Building Materials Technology and Application, (06): 57–60.

[5] Li N, Han D, Zhang H, et al., 2023, Underwater Sound Source Localization Method Based on Similarity Matching of Time-Frequency Distribution Graph. Ship Electronic Engineering, 43(11): 184–188.

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

An Image Manipulation Localization Method Based on Dual-Branch Hybrid Convolution

Chengliang Yan[1], Lei Zhang[2]*, Minhui Chang[2]

[1]School of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, Shanxi, China
[2]School of Mathematics and Information Technology, Yuncheng University, Yuncheng 044000, Shanxi, China

**Author to whom correspondence should be addressed.*

Abstract: In existing image manipulation localization methods, the receptive field of standard convolution is limited, and during feature transfer, it is easy to lose high-frequency information about traces of manipulation. In addition, during feature fusion, the use of fixed sampling kernels makes it difficult to focus on local changes in features, leading to limited localization accuracy. This paper proposes an image manipulation localization method based on dual-branch hybrid convolution. First, a dual-branch hybrid convolution module is designed to expand the receptive field of the model to enhance the feature extraction ability of contextual semantic information, while also enabling the model to focus more on the high-frequency detail features of manipulation traces while localizing the manipulated area. Second, a multi-scale content-aware feature fusion module is used to dynamically generate adaptive sampling kernels for each position in the feature map, enabling the model to focus more on the details of local features while locating the manipulated area. Experimental results on multiple datasets show that this method not only effectively improves the accuracy of image manipulation localization but also enhances the robustness of the model.

Keywords: Image manipulation localization; Content awareness; Dual branch; Hybrid convolution; U-Net

Online publication:

1. Introduction

In recent years, with the rapid development of smart devices, digital images have become widely accessible and easily disseminated. Concurrently, the operational barriers for image editing software have progressively lowered, enabling more individuals to acquire image editing skills. This has made distinguishing manipulated images increasingly challenging. Particularly driven by the recent surge in AIGC technology, image editing tasks can now be accomplished with just a few well-crafted instructions. The emergence of these edited images poses a serious threat to the authenticity and reliability of digital imagery [1]. The illicit applications of manipulated images are extensive, frequently exploited in critical domains such as spreading fake news, fabricating judicial evidence, and infringing intellectual property rights, severely impacting socioeconomic development [2,3]. As a vital method in

image forensics, image tampering localization aims to identify and pinpoint altered regions within images, a task attracting increasing attention from researchers.

Existing image tampering localization methods primarily fall into two categories: those based on traditional feature extraction and those based on deep learning. Traditional feature extraction methods employ manually designed features for extraction, perform statistical analysis on the extracted features, and use this to determine image tampering and its location. Traditional feature extraction methods mainly include methods based on color filter array (CFA) [4,5] consistency detection methods, which analyze the damage caused by tampering operations to the CFA interpolation patterns specific to different camera models to locate tampered regions. Methods based on illumination consistency [6–8] analyze inconsistencies in the direction, intensity, and shadow distribution of illumination across different object surfaces within an image to locate tampered regions. Detection methods based on imaging system noise consistency [9–11] detect tampering by identifying differences in systematic noise distribution (e.g., sensor noise) between tampered regions and the overall image. Methods based on JPEG compression artifact consistency [12–14] identify tampered regions through anomalies in JPEG compression artifacts introduced by tampering (e.g., inconsistent block effects, double compression artifacts). These approaches offer high interpretability but detect only limited types of tampering, with constrained localization accuracy and robustness in complex scenarios.

Leveraging the powerful feature extraction capabilities of deep learning, particularly in recent years, an increasing number of researchers have applied it to the field of image tampering, gradually developing deep learning-based image tampering detection methods. Based on mainstream deep learning network architectures, these methods fall into two categories: those based on convolutional neural networks [15–19] and those based on transformers [20–22]. Convolutional neural network-based methods utilize convolutions to extract local image features, employing upsampling or deconvolution operations to achieve pixel-level localization. Transformer-based methods enhance the model's focus on tampered regions by incorporating attention mechanisms and integrating features from different scales [23,24].

Although existing methods based on convolutional neural networks have significantly improved the accuracy of image tampering localization, some shortcomings remain. Current approaches suffer from limited receptive fields due to the size constraints of neural network convolutional kernels. Furthermore, during feature propagation, high-frequency information related to tampering traces is prone to loss. Cross-scale feature interactions are weak, and feature fusion across different levels is insufficient. Furthermore, existing methods employ fixed sampling kernels during feature fusion, failing to effectively capture local details and global contextual information. This inability to focus on local feature variations limits localization accuracy.

To address the aforementioned issues, this paper proposes an image tampering localization method based on dual-branch hybrid convolutions. First, a dual-branch hybrid convolution module is designed. The spatial domain feature extraction branch employs two dilation convolutions to expand the receptive field, enhancing the capture of multi-scale feature information. The frequency domain detail enhancement branch utilizes WTConv to more effectively improve the model's extraction of frequency feature information. This dual-branch hybrid convolution module enables the model to extract multi-scale feature information from both spatial and frequency domains, enhancing its ability to detect tampering traces and thereby improving localization accuracy. Second, introducing content-aware upscaling, we designed a multi-scale content-aware feature fusion module. This module dynamically generates adaptive sampling kernels for each feature map through multi-scale content-aware upscaling, enabling the fused features to effectively capture contextual information and local detail features at

different scales within the tampering region.

2. Research methods

In image manipulation localization tasks, the location is typically determined based on pixel differences between the manipulated and authentic regions. This is primarily achieved by extracting feature information from the manipulated image, learning edge features, frequency domain features, artifact features, and other characteristic details. Leveraging the powerful feature extraction capabilities of deep learning, the manipulated regions within the image are localized. To enhance the accuracy of manipulation localization, it is necessary to strengthen the model's ability to extract these manipulation-specific features.

2.1. Model architecture

This paper proposes an image tampering localization method based on dual-branch hybrid convolutions. The proposed method adopts RRU-Net [25] as its backbone network, primarily consisting of an encoder and a decoder. The encoder performs feature extraction and downsampling on the input image through dual-branch hybrid convolutions and max-pooling layers, while the decoder fuses features via content-aware upsampling. Additionally, the recurrent residual structure within this network addresses the vanishing gradient problem caused by excessive network depth, enhancing the utilization of contextual spatial information in images. The specific network architecture is illustrated in **Figure 1**.

First, the input image $X \in \mathbb{R}^{H \times W \times C}$ is fed into the encoder on the left. The feature maps obtained from each encoding layer are downsampled and used as input for the next encoding layer. Ultimately, the encoder produces feature outputs at different scales.

$$F_i = D_i(X_{i-1}), i \in \{1,2\ 3\ 4\ 5\} \quad F_i = \{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}\} \tag{1}$$

Here, D_i denotes the encoder. Subsequently, the features extracted by the encoder are fused with those processed through multiscale content-aware upscaling, enabling the fused features to effectively utilize contextual information and variations in local features.

$$U_i = upsample_i(U_{i+1}, F_i), i \in \{1,2,3\} \tag{2}$$

$$U_4 = upsample_4(F_5, F_4) \tag{3}$$

Finally, after a 1×1 convolution, the predicted tampering region mask map $Y = Conv(U_1) \in \mathbb{R}^{H \times W \times 1}$ is obtained.

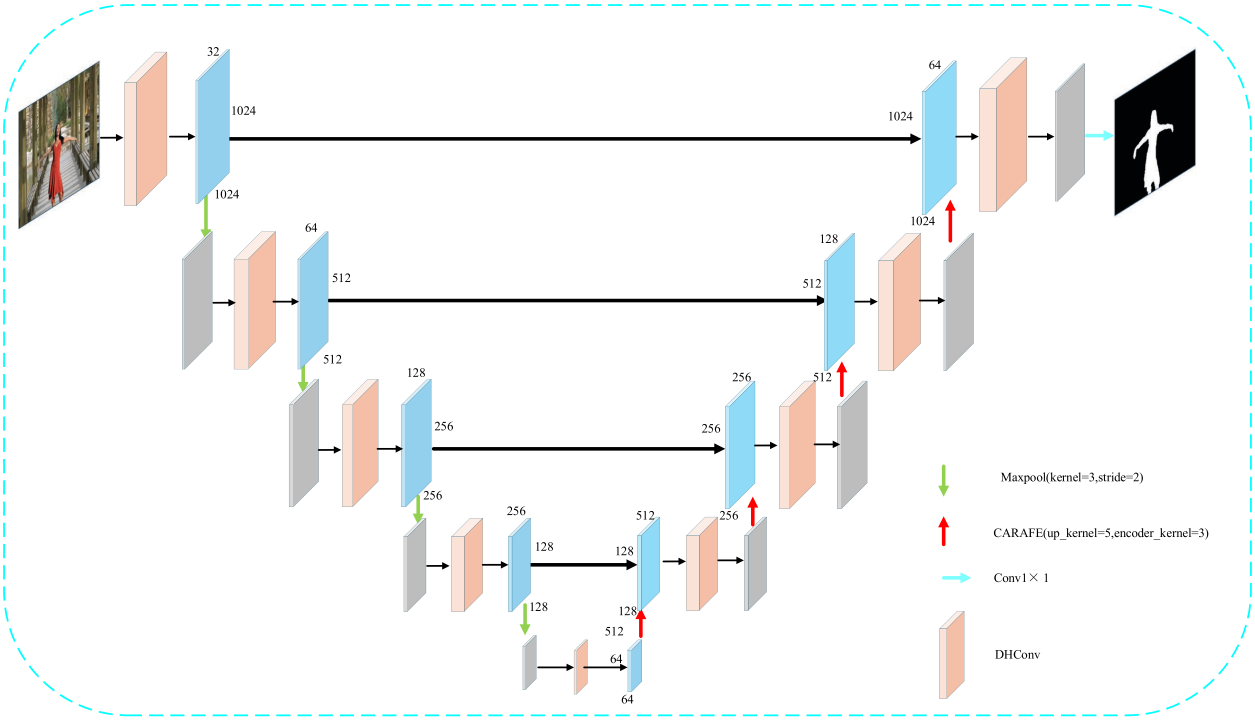


Figure 1. Overall network architecture

2.2. Dual-branch hybrid convolution

Image manipulation operations typically leave traces of human intervention in the altered regions, and these traces constitute crucial features that can be learned by manipulation localization methods. Feature extraction at a single scale captures extremely limited information and fails to fully leverage the complementary nature of multi-scale data. Consequently, in recent years, an increasing number of researchers have turned to multi-scale features for image manipulation localization studies.

In this paper, to leverage frequency-domain information across different scales while addressing limitations such as restricted receptive fields in traditional convolutions and increased computational overhead and model complexity when expanding receptive fields using conventional methods, we designed a dual-branch hybrid convolutional module inspired by Finder *et al.* [26] to enhance the model's feature learning capability. As shown in **Figure 1**, within the encoder, the dual-branch hybrid convolution module extracts image features at different scales. Low-scale features are fed into the high-scale feature extraction process, ultimately yielding a multi-scale feature map $\{F_1, F_2^{\frac{1}{2}}, F_3^{\frac{1}{4}}, F_4^{\frac{1}{8}}, F_5^{\frac{1}{16}}\}$. In the decoder, high-scale feature maps undergo successive upsampling, with the dual-branch hybrid convolution module extracting image features at varying scales. The structure of the dual-branch hybrid convolution module is illustrated in **Figure 2**.

As shown in **Figure 2**, the dual-branch hybrid convolution module is primarily divided into two branches: the spatial receptive field expansion branch, mainly composed of dilated convolutions, extracts spatial domain features; the frequency domain detail enhancement branch, primarily composed of wavelet convolutions, extracts frequency domain features.

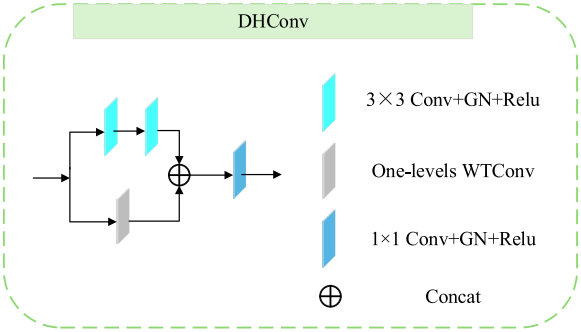


Figure 2. Dual-branch hybrid convolution module

In the spatial receptive field expansion branch, two 3×3 dilated convolutions with a dilation rate of 2 and ReLU activation functions are applied to the input feature map to expand the model's receptive field. This enhances the model's ability to extract contextual semantic information, thereby capturing the spatial feature information X_s of the input features. In the frequency domain detail enhancement branch, WTConv wavelet transformation is employed to separate low-frequency and high-frequency information from the input feature map. This wavelet transform employs four distinct filters: the *LL* filter captures low-frequency information, the *LH* filter captures horizontal information, the *HL* filter captures vertical information, and the *HH* filter captures diagonal information. Among these, *LL* is a low-pass filter, while *LH*, *HL*, and *HH* form a set of high-pass filters. These four filters constitute an orthogonal basis as shown in (4):

$$f_{LL}=\frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix},\ f_{LH}=\frac{1}{2}\begin{bmatrix}1 & -1\\ 1 & -1\end{bmatrix},f_{HL}=\frac{1}{2}\begin{bmatrix}1 & 1\\ -1 & -1\end{bmatrix},f_{HH}=\frac{1}{2}\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix} \tag{4}$$

For each input channel, the convolution output has four channels, each with a resolution equal to half of *X*.

$$[X_{LL},X_{LH},X_{HL},X_{HH}]=Conv5\times5([f_{LL},f_{LH},f_{HL},f_{HH}],X) \tag{5}$$

Among these, convolving the orthogonal basis generated by the four filters with 5×5 produces the low-frequency component X_{LL} of the input *X*, along with the horizontal, vertical, and diagonal high-frequency components X_{LH}, X_{HL}, and X_{HH}.

For the orthogonal basis of equation (4), the inverse wavelet transform (IWT) is obtained through transpose convolution:

$$X_I=transposedConv([f_{LL},f_{LH},f_{HL},f_{HH}],[X_{LL},X_{LH},X_{HL},X_{HH}]) \tag{6}$$

By separating the convolution operation from the frequency components through wavelet transformation, high-frequency and low-frequency information is isolated. The enhanced low-frequency information improves feature extraction. Ultimately, the frequency domain detail enhancement branch extracts the frequency feature information X_F from the input *X*.

$$X_F=Conv(X)+IWT(X_I) \tag{7}$$

The WTConv wavelet transform structure is shown in **Figure 3**.

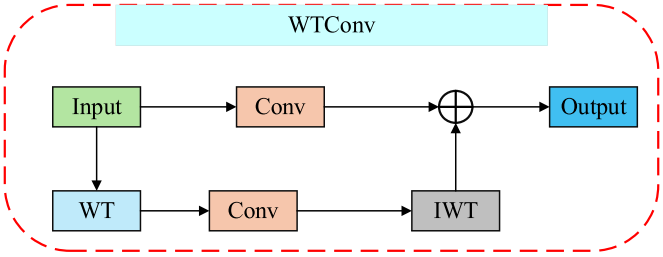


Figure 3. WTConv Architecture

Finally, the spatial feature information X_S extracted by the spatial feature field expansion branch and the frequency feature information X_F extracted by the frequency detail enhancement branch are fused. This fusion undergoes a 1×1 convolution to produce the final output features.

2.3. Multi-scale content-aware feature fusion

In our approach, to fully leverage feature information across different scales and effectively capture both contextual information and local detail features of the tampered region at various scales, we designed a multi-scale content-aware feature fusion module at the decoder end, inspired by Wang *et al.* [27]. As shown in **Figure 1**, at the decoder stage, high-scale features undergo content-aware upsampling and are concatenated with low-scale features. After four rounds of upsampling, the fused features are obtained. These fused features effectively perceive the contextual information and local detail features of the tampered region across multiple scales, enhancing the model's ability to recognize tampering edges and texture details. The network architecture for content-aware upsampling is illustrated in **Figure 4**.

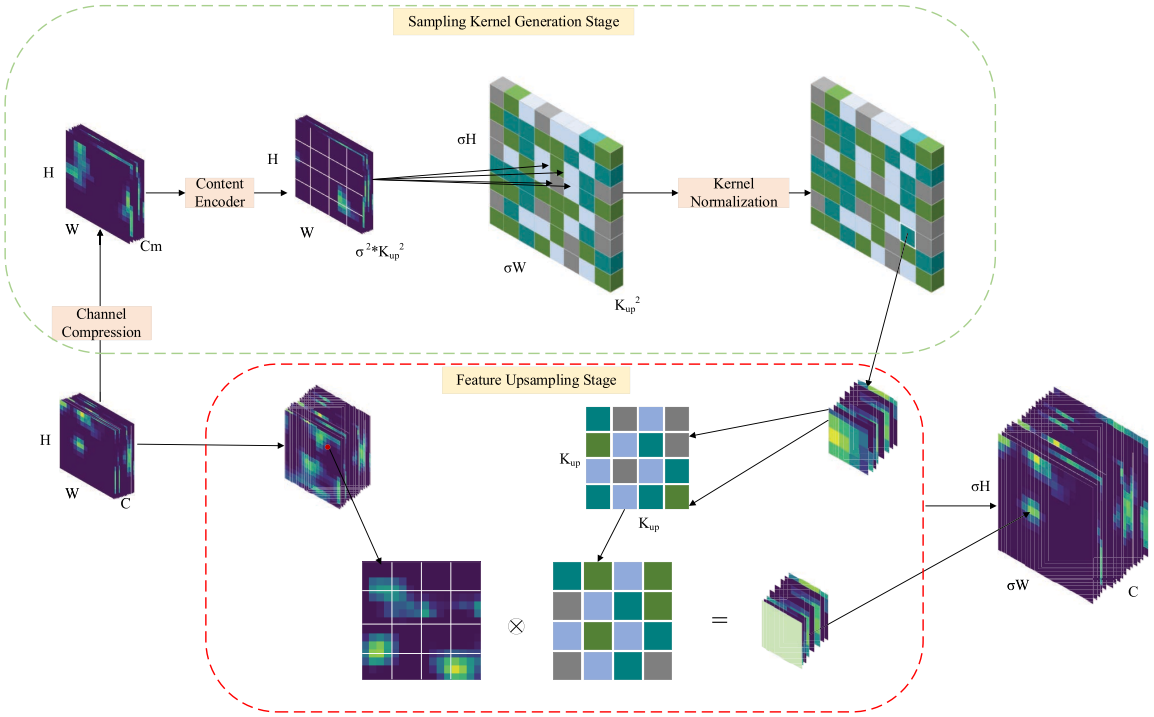


Figure 4. Content-aware upsampling network architecture

As shown in **Figure 4**, the content-aware upscaling process primarily consists of two stages: the sampling

kernel generation stage and the feature upscaling stage. During the sampling kernel generation stage, adaptive sampling kernels are dynamically generated for each position of the target feature based on its specific content. The specific process is as follows: For the input feature map $F \in \mathbb{R}^{H\times W\times C}$, a 1×1 convolution is first applied to perform channel compression on feature F, yielding the compressed feature map $F' \in \mathbb{R}^{H\times W\times C_m}$, where C_m denotes the number of compressed channels. Then, encoding is performed using the content encoder to obtain $F'' \in \mathbb{R}^{H\times W\times(\sigma^2\times k_{up}^2)}$. This encoder employs a 3×3 convolution kernel, where σ denotes the upsampling factor and k_{up} represents the upsampling kernel size. Subsequently, F'' undergoes pixel reordering to produce $F''' \in \mathbb{R}^{\sigma H\times\sigma W\times k_{up}^2}$. Finally, F''' is normalized to generate the sampling kernel F_{kernel}, creating a sampling kernel $k_{up}\times k_{up}$ with k_{up}^2 weights for each position.

During the feature sampling phase, the input features are upsampled based on the generated sampling kernels. First, for each position $p = (i,j)$ in the input feature $F \in \mathbb{R}^{H\times W\times C}$, extract the $k_{up}\times k_{up}$ neighborhood N_p centered at $p = (i,j)$. Then, at the corresponding position $p' = (i',j')$ in the sampling kernel F_{kernel} generated during the kernel sampling phase, the corresponding $k_{up}\times k_{up}$ sampling kernel $K_{p'}$ is extracted. Multiplying the corresponding positions of F and $K_{p'}$ and summing the results yields the feature value at that position for the target feature. Performing this operation for each position produces the output feature map $F_{out} \in \mathbb{R}^{\sigma H\times\sigma W\times C}$.

3. Experimental design and results

3.1. Experimental setup

(1) Dataset: To validate the effectiveness of our method, we conducted experiments on various public datasets, including CASIAv1 [28], CASIAv2 [28], Columbia [29], and NIST16 [30]. To ensure a more scientific and accurate comparison, we split the Columbia and NIST16 datasets into training and testing sets. For CASIA, we used CASIAv2 as the training set and CASIAv1 as the testing set for experiments. The dataset configurations are shown in **Table 1**. All training images underwent data augmentation techniques such as rotation, cropping, and transformation.

Table 1. Training and test set partitioning of the dataset

Datasets	CASIA	Columbia	NIST16
Train	5123	126	404
Test	921	54	160

(2) Experimental environment: The experiment was implemented using the PyTorch deep learning framework, with Python 3.8 as the programming environment. Training was conducted on an NVIDIA GeForce RTX 3090 GPU. During training, the Adam optimizer was employed with an initial learning rate of 1e-4 and 200 epochs. Input image resolutions were adjusted based on resolution characteristics for different training datasets.

(3) Evaluation metrics: This paper primarily employs pixel-level F1 score and area under the curve (AUC) as model evaluation metrics. These represent the two most commonly used core indicators for assessing model performance in image tampering localization tasks, measuring the model's localization accuracy and discrimination capability from different perspectives, respectively. They serve complementary evaluation functions within this research domain, collectively establishing a multidimensional

performance assessment framework.

The Pixel-level F1 Score is the harmonic mean of Precision and Recall, used to evaluate a model's accuracy in locating tampered regions at the pixel level. It is defined as:

$$F_1 = \frac{2Precision * Recall}{Precision + Recall} \tag{8}$$

Among these, accuracy is defined as:

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

The recall rate is defined as:

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

The AUC evaluates a model's overall performance across different resolution thresholds by calculating the area enclosed by the receiver operating characteristic (ROC) curve. The ROC curve is generated by adjusting classification thresholds, with true positive rate (TPR) plotted on the vertical axis and false positive rate (FPR) on the horizontal axis. The closer the AUC value is to 1, the better the model performs in classifying pixels between the tampered region and the true background.

3.2. Ablation experiment

To validate the effectiveness of the proposed dual-branch hybrid convolution and multi-scale content-aware feature fusion module, ablation experiments were designed on the CASIAv1, Columbia, and NIST16 datasets. The experimental design is as follows:

Option 1: Remove all modules simultaneously to validate the baseline model's performance across datasets;

Option 2: Remove the dual-branch hybrid convolution module and use standard convolutions for feature extraction within the network;

Option 3: Remove the multi-scale content-aware feature fusion module and directly up-sample the encoder's output feature map to the decoder for prediction;

Option 4: Utilize all modules simultaneously to validate the model's overall performance across datasets.

As shown in the ablation experiment results in **Table 2**, when all modules are used simultaneously, the model achieves an average F1 score of 0.691 and an average AUC of 0.871 across all datasets. Removing any single module results in a decrease in both the average F1 score and average AUC. The baseline model exhibits the lowest average F1 score and average AUC, validating the effectiveness of the proposed modules. In Scheme 3, removing the multi-scale content-aware feature fusion module resulted in a significant decrease in both F1 and AUC scores, dropping by 4.8 percentage points and 3.7 percentage points, respectively. This indicates that multi-scale content-aware feature fusion effectively captures contextual information and local detail features across different scales, thereby enhancing localization accuracy.

Table 2. Ablation experiment results

Option	CASIAv1		Columbia		NIST16		MEAN	
	F1	AUC	F1	AUC	F1	AUC	F1	AUC
1	0.397	0.728	0.724	0.839	0.850	0.985	0.657	0.851
2	0.411	0.742	0.745	0.869	0.896	0.992	0.684	0.868
3	0.417	0.751	0.640	0.766	0.873	0.985	0.643	0.834
4	0.422	0.751	0.750	0.869	0.900	0.993	0.691	0.871

3.3. Objective evaluation

To objectively evaluate model performance, the proposed method is compared with state-of-the-art approaches, including SPAN [18], GSR-Net [31], DenseFCN [32], LocateNet [33], RGB-N [15], U-Net [24], and RRU-Net [25]. To assess model effectiveness, we recorded the F1 scores of each method across different datasets. Additionally, to visually compare the overall performance of different approaches, we calculated the average F1 scores for each method across various datasets. The results are presented in **Table 3**.

Table 3. Comparison of F1 scores across different methods on each dataset

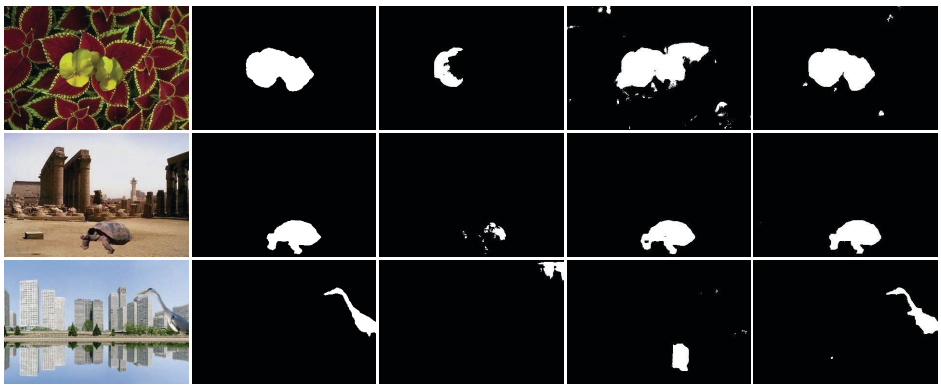
Method	F1 scores (%)			
	CASIAv1	Columbia	NIST6	MEAN
RGB-N [15]	40.8	69.7	72.2	60.9
SPAN [18]	38.2	81.5	58.2	48.0
LocateNet [33]	27.3	41.1	73.8	47.4
GSRNet [31]	34.0	43.3	64.0	47.1
DenseFCN [32]	20.3	25.7	81.2	42.4
U-Net [24]	35.0	51.5	67.2	51.2
RRU-Net [25]	39.7	72.4	85.0	65.7
Ours	42.2	75.0	90.0	69.1

As shown in **Table 3**, the model achieves an average F1 score of 69.1% across the CASIAv1, Columbia, and NIST16 datasets, representing a 3.4 percentage point improvement over RRU-Net. Our method achieves the highest F1 scores on both the CASIAv1 and NIST16 datasets. On the Columbia dataset, it falls below the SPAN method. This discrepancy stems from the Columbia dataset's homogeneous tampering types and high image resolution. During the adjustment of input image resolution, the tampered regions may have been adversely affected.

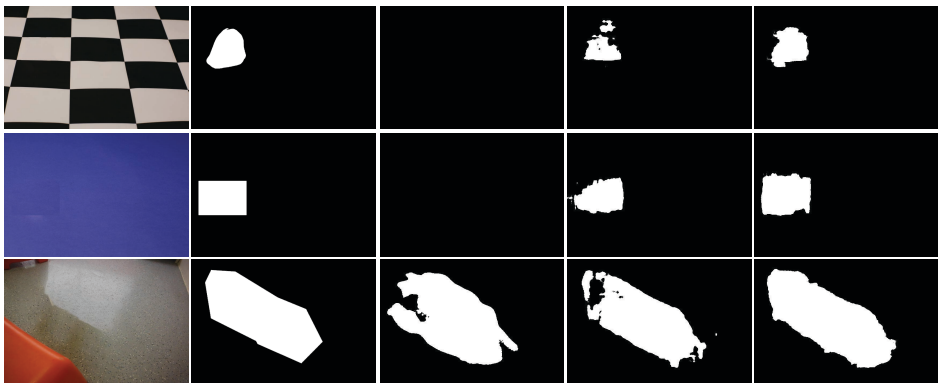
3.4. Subjective evaluation

To enable a more intuitive comparison from a subjective visual perspective, the proposed method is compared with other approaches based on their predicted masks on the test dataset. By contrasting the predicted mask images with the ground truth masks, differences in localization accuracy, boundary clarity, and false detection rate among various methods can be clearly observed. Selected results are shown in **Figure 5**. **Figure 5(a)** shows the comparison of predicted masks from different methods on the CASIA v1 dataset, **Figure 5(b)** shows the

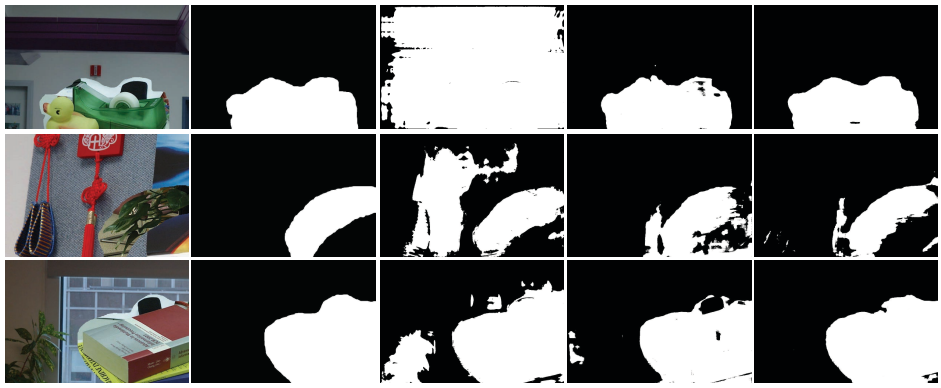
comparison on the NIST16 dataset, and **Figure 5(c)** shows the comparison on the Columbia dataset. In these three figures, Column 1 shows the tampered image. Column 2 displays the true mask image. Column 3 presents the predicted mask from the U-Net method. Column 4 shows the predicted mask from RRU-Net. Column 5 displays the predicted mask from the proposed method.



(a) Comparison of prediction masks across different methods on the CASIA v1 dataset



(b) Comparison of prediction masks across different methods on the NIST16 dataset



(c) Comparison of prediction masks across different methods on the Columbia dataset

Figure 5. Comparison of prediction masks between the proposed method and other methods

Figure 5 demonstrates that compared to other methods, our approach achieves more precise localization of tampered regions, with clearer predicted boundaries and higher overlap with the ground truth mask. Additionally, our method exhibits favorable false positive rates, significantly reducing instances where pixels from genuine

regions are erroneously identified as tampered. This indicates strong capability in distinguishing between authentic and tampered image regions.

3.5. Robustness assessment

In image tampering localization tasks, robustness evaluation is a critical step for verifying model effectiveness and reliability. Real-world images may be subject to various disturbances such as noise, light pollution, compression, and transformations, which significantly impact model performance. Robustness evaluation tests a model's stability under different disturbance conditions and validates its generalization capability.

In this paper, we conduct robustness experiments on the CASIAv1 dataset, applying two types of image distortion: Gaussian Blur and JPEG Compression. The experimental results are shown in **Figure 6**. As illustrated in **Figure 6**, the proposed method demonstrates strong robustness against both types of interference attacks.

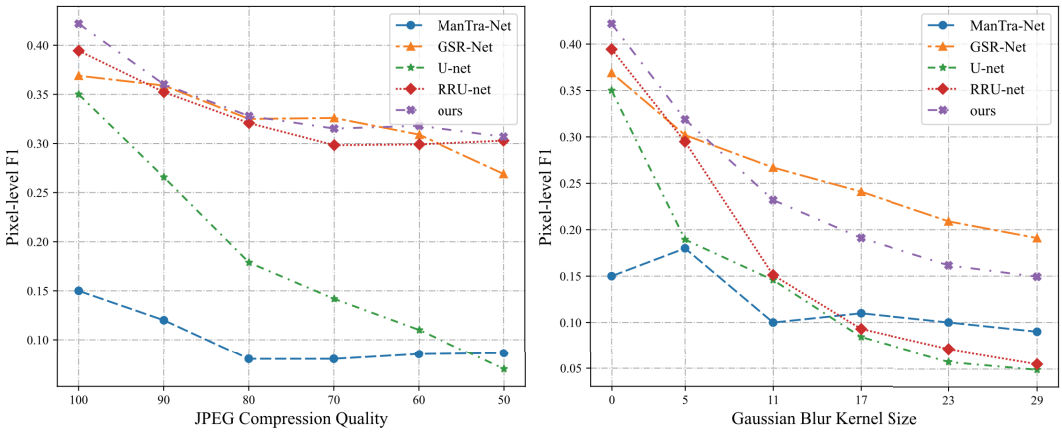


Figure 6. Robustness evaluation of the model on the CASIAv1 dataset

4. Conclusion

This paper proposes an image tampering localization method based on dual-branch hybrid convolutions. First, a dual-branch hybrid convolution module is designed to expand the model's receptive field and enhance its ability to extract contextual semantic information. This enables the model to focus on high-frequency detail features of tampering traces while locating tampered regions. Second, a multi-scale content-aware feature fusion module is employed to dynamically generate adaptive sampling kernels for features at different scales. This enables the model to identify tampering edges and local feature details while locating tampered regions, effectively enhancing robustness and generalization capabilities. However, challenges remain: the proposed method's localization accuracy requires improvement for datasets with limited tampering types and low image resolution. Future research could explore incorporating edge information during feature extraction to further enhance the model's localization precision.

Funding

National Natural Science Foundation of China (61703363); Shanxi Provincial Basic Research Program (202403021221206); Key Project of Shanxi Provincial Strategic Research on Science and Technology

(202304031401011); Funding Project for Scientific Research Innovation Team on Data Mining and Industrial Intelligence Applications (YCXYTD-202402); Yuncheng University Research Project (YQ-2020021)

Disclosure statement

The authors declare no conflict of interest.

References

[1] Jin X, Yu W, Shi W, 2024, Image Manipulation Localization via Dynamic Cross-Modality Fusion and Progressive Integration. Neurocomputing, 610: 128607.

[2] Wei H, Yan C, Li H, 2024, Image Tampering Localization Based on Integrated Multiscale Attention. Journal of Computer-Aided Design & Computer Graphics, 36(08): 1237–1245.

[3] Zeng Z, Tan P, 2025, Image Tampering Detection and Localization Model Based on Multi-Branch HRNet. Modern Electronic Technique, 48(03): 35–42.

[4] Varlamova AA, Kuznetsov AV, 2017, Image Splicing Localization Based on CFA-Artifacts Analysis. Computer Optics, 41(6): 920–930.

[5] Hussien NY, Mahmoud RO, Zayed HH, 2020, Deep Learning on Digital Image Splicing Detection Using CFA Artifacts. International Journal of Sociotechnology and Knowledge Development (IJSKD), 12(2): 31–44.

[6] Vidyadharan DS, Thampi SM, 2018, Evaluating Color and Texture Features for Forgery Localization from Illuminant Maps. Multimedia Tools and Applications, 77: 21131–21161.

[7] Niyishaka P, Bhagvati C, 2021, Image Splicing Detection Technique Based on Illumination-Reflectance Model and LBP. Multimedia Tools and Applications, 80(2): 2161–2175.

[8] Zhe S, Peng S, 2020, Authentication of Splicing Manipulation by Exposing Inconsistency in Color Shift. Multimedia Tools and Applications, 79(11): 8235–8248.

[9] Lyu S, Pan X, Zhang X, 2014, Exposing Region Splicing Forgeries with Blind Local Noise Estimation. International Journal of Computer Vision, 110: 202–221.

[10] Dong J, Chen L, Tian J, et al., 2016, A Novel Image Splicing Detection Method Based on the Inconsistency of Image Noise, 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), IEEE, 560–563.

[11] Zhu N, Li Z, 2018, Blind Image Splicing Detection via Noise Level Function. Signal Processing: Image Communication, 68: 181–192.

[12] Wang SL, Liew AWC, Li SH, et al., 2014, Detection of Shifted Double JPEG Compression by an Adaptive DCT Coefficient Model. EURASIP Journal on Advances in Signal Processing, 2014: 1–17.

[13] Thai TH, Cogranne R, Retraint F, et al., 2016, JPEG Quantization Step Estimation and Its Applications to Digital Image Forensics. IEEE Transactions on Information Forensics and Security, 12(1): 123–133.

[14] Iakovidou C, Zampoglou M, Papadopoulos S, et al., 2018, Content-Aware Detection of JPEG Grid Inconsistencies for Intuitive Image Forensics. Journal of Visual Communication and Image Representation, 54: 155–170.

[15] Zhou P, Han X, Morariu VI, et al., 2018, Learning Rich Features for Image Manipulation Detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1053–1061.

[16] Wu Y, Abd Almageed W, Natarajan P, 2019, Mantra-Net: Manipulation Tracing Network for Detection and Localization of Image Forgeries with Anomalous Features, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9543–9552.

[17] Bayar B, Stamm MC, 2018, Constrained Convolutional Neural Networks: A New Approach Towards General Purpose Image Manipulation Detection. IEEE Transactions on Information Forensics and Security, 13(11): 2691–2706.

[18] Hu X, Zhang Z, Jiang Z, et al., 2020, SPAN: Spatial Pyramid Attention Network for Image Manipulation Localization, Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16, Springer International Publishing, 312–328.

[19] Chen X, Dong C, Ji J, et al., 2021, Image Manipulation Detection by Multi-View Multi-Scale Supervision, Proceedings of the IEEE/CVF International Conference on Computer Vision, 14185–14193.

[20] Hao J, Zhang Z, Yang S, et al., 2021, Transforensics: Image Forgery Localization with Dense Self-Attention, Proceedings of the IEEE/CVF International Conference on Computer Vision, 15055–15064.

[21] Wang J, Wu Z, Chen J, et al., 2022, Objectformer for Image Manipulation Detection and Localization, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2364–2373.

[22] Ma X, Du B, Jiang Z, et al., 2023, IML-ViT: Benchmarking Image Manipulation Localization by Vision Transformer. arXiv. https://arxiv.org/abs/2307.14863

[23] Zeng K, Cheng R, Tan W, et al., 2024, MGQFormer: Mask-Guided Query-Based Transformer for Image Manipulation Localization, Proceedings of the AAAI Conference on Artificial Intelligence, 38(7): 6944–6952.

[24] Ronneberger O, Fischer P, Brox T, 2015, U-Net: Convolutional Networks for Biomedical Image Segmentation, Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18, Springer International Publishing, 234–241.

[25] Bi X, Wei Y, Xiao B, et al., 2019, RRU-Net: The Ringed Residual U-Net for Image Splicing Forgery Detection, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.

[26] Finder SE, Amoyal R, Treister E, et al., 2024, Wavelet Convolutions for Large Receptive Fields, European Conference on Computer Vision, Springer Nature Switzerland, Cham, 363–380.

[27] Wang J, Chen K, Xu R, et al., 2019, Carafe: Content-Aware Reassembly of Features, Proceedings of the IEEE/CVF International Conference on Computer Vision, 3007–3016.

[28] Dong J, Wang W, Tan T, 2013, Casia Image Tampering Detection Evaluation Database, 2013 IEEE China Summit and International Conference on Signal and Information Processing, IEEE, 422–426.

[29] Hsu YF, Chang SF, 2006, Detecting Image Splicing Using Geometry Invariants and Camera Characteristics Consistency, 2006 IEEE International Conference on Multimedia and Expo, IEEE, 549–552.

[30] Guan H, Kozak M, Robertson E, et al., 2019, MFC Datasets: Large-Scale Benchmark Datasets for Media Forensic Challenge Evaluation, 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), IEEE, 63–72

[31] Zhou P, Chen B C, Han X, et al., 2020, Generate, Segment, and Refine: Towards Generic Manipulation Segmentation, Proceedings of the AAAI Conference on Artificial Intelligence, 34(07): 13058–13065.

[32] Zhuang P, Li H, Tan S, et al., 2021, Image Tampering Localization Using a Dense Fully Convolutional Network. IEEE Transactions on Information Forensics and Security, 16: 2986–2999.

[33] Zhuo L, Tan S, Li B, et al., 2022, Self-Adversarial Training Incorporating Forgery Attention for Image Forgery Localization. IEEE Transactions on Information Forensics and Security, 17: 819–834.

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Synthesis and Application of Zero-Dimensional Metal Oxide Composites in Energy Chemistry

Runtian Hu*

University College London (UCL), London WC1E 6BT, UK

**Author to whom correspondence should be addressed.*

Abstract: Against the backdrop of increasingly prominent global energy shortages and environmental issues, the development of efficient energy conversion and storage technologies has become crucial. Zero-dimensional (0D) metal oxide composites exhibit significant application value in the field of energy chemistry due to their unique properties, such as quantum size effect and high specific surface area. From a broad perspective, this paper reviews the main synthesis methods of these composites, including sol-gel method, hydrothermal/solvothermal method, precipitation method, and template method, while analyzing the characteristics of each method. It further discusses their applications in photocatalytic hydrogen production, fuel cells, lithium-ion batteries, and supercapacitors. Additionally, the current challenges, such as material dispersibility and interface bonding, are pointed out, and future development directions are prospected, aiming to provide references for related research.
Keywords: Zero-dimensional metal oxide; Composite material; Synthesis method; Energy chemistry; Energy conversion; Energy storage

Online publication: October 21, 2025

1. Introduction

The rapid development of industrialization has led to the extensive use of fossil energy, resulting in an energy supply-demand imbalance and severe environmental problems such as the greenhouse effect and pollution, which seriously restrict the process of sustainable human development. The research and development of renewable energy as well as efficient energy conversion and storage methods have become key energy research topics worldwide. Nanomaterials demonstrate excellent performance in energy chemical reactions due to their unique properties such as size effect and surface effect. Three-dimensional nanomaterials are important members of the nanomaterial family; when their size is ≤ 100 nm, they are referred to as zero-dimensional nanomaterials, which exhibit a prominent quantum size effect [1].

Zero-dimensional (0D) metal oxides (such as TiO_2, ZnO, etc.) have attracted widespread attention in the energy field due to their chemical stability, abundant reserves, and low cost. However, single 0D metal oxides

usually suffer from drawbacks such as low electrical conductivity and high recombination probability of photogenerated electrons and holes, which limit the improvement of their performance. Composites formed by 0D metal oxides with carbon materials, polymers, and other substances can regulate the material structure and improve performance through the interaction between components. For example, composites with graphene can accelerate electron migration, and those with carbon nanotubes can address the volume expansion issue. Therefore, summarizing and introducing the preparation of 0D metal oxides and their composite structures, as well as their applications in energy, is conducive to understanding the formation of 0D nanomaterials and effectively promoting the sustainable development and application of energy materials from the perspective of preparation.

2. Synthesis methods of zero-dimensional metal oxide composites

The synthesis of zero-dimensional metal oxide composites is the basis for their performance, and different methods affect the particle size, morphology, and interface state of the materials. The following are four common and representative synthesis methods.

2.1. Sol-gel method

The sol-gel method is a classic synthesis technique. It uses metal alkoxides or inorganic salt aqueous solutions as precursors, which undergo hydrolysis and polycondensation to form a sol. The sol then aggregates into a gel, and finally, the composite material is obtained through drying and calcination [2].

In the synthesis process, precursors such as tetrabutyl titanate and zinc nitrate are selected as raw materials. Composite components, such as graphene oxide, are first dispersed and mixed. For example, in the synthesis of TiO_2/graphene composites, graphene oxide is dispersed in a tetrabutyl titanate ethanol solution, then a catalyst is added to initiate hydrolysis and polycondensation to form a sol. The sol is further converted into a gel, and the product is obtained after drying and calcination [3].

The advantages of this method include low reaction temperature, easy control of particle size and morphology, realization of molecular-level mixing of components, simple process, and suitability for the synthesis of various composite materials. However, it has shortcomings such as a long synthesis cycle, easy volume shrinkage and agglomeration during drying and calcination, and high cost of some precursors [4].

2.2. Hydrothermal/solvothermal method

The hydrothermal/solvothermal method is carried out in a closed high-pressure reaction kettle, using water or organic solvents as the medium, and promoting reactions through high temperature and high pressure to prepare composite materials. When water is used as the medium, it is called the hydrothermal method; when organic solvents are used, it is called the solvothermal method.

The synthesis process is as follows: Metal salt precursors, precursors of composite components, and surfactants are dispersed in the medium, then placed into a reaction kettle. The reaction temperature is controlled at 100–300°C, and the reaction time ranges from several hours to dozens of hours. After cooling, centrifugation, and filtration, the product is obtained.

This method can prevent the introduction of environmental impurities, resulting in pure products. The crystallinity, particle size, morphology, etc., of the desired product can be accurately controlled by adjusting reaction conditions, and it is applicable to various composite systems. However, it requires special high-pressure

equipment, which leads to high costs and relatively high risk. In addition, the reaction time is long, and it is not suitable for sensitive composite components.

2.3. Precipitation method

The precipitation method involves adding a precipitant to a metal salt solution to convert metal ions into precipitates, which are then mixed with composite component precursors to obtain composite materials. Among various precipitation techniques, the coprecipitation method is the most widely used.

In the coprecipitation preparation process, solutions of two or more metal ion salts are mixed with a precipitant (such as sodium hydroxide or sodium carbonate). Under specific conditions (e.g., certain pH and temperature), a mixed precipitate is formed, which undergoes washing, drying, and calcination to produce the composite material. For example, to synthesize the Co_3O_4/NiO mixed compound, a mixed solution of cobalt nitrate and nickel nitrate is dropped into a sodium carbonate solution to obtain a mixed carbonate precipitate, which is then calcined to yield the Co_3O_4/NiO mixed compound [5].

This method features a simple process, convenient operation, no need for complex equipment, low cost, and easy industrialization, and can regulate the material composition and particle size. However, the difference in precipitation rates of different metal ions tends to cause uneven component distribution, and the precipitates are prone to agglomeration during drying and calcination, resulting in low crystallinity.

2.4. Template method

The template method regulates the structure and morphology of composite materials using templates with specific structures (hard templates or soft templates). Precursor assembly reactions are guided by the template, and the target material is obtained after template removal.

Hard templates are rigid solid materials (e.g., mesoporous silica, carbon spheres). Precursors are introduced into the template's pores or onto its surface, and the template is removed after the reaction. For instance, using mesoporous silica SBA-15 as a template, tetrabutyl titanate and graphene oxide are injected into its pores; after calcination and template removal, a mesoporous TiO_2/graphene composite is obtained. Soft templates are ordered aggregates (e.g., micelles) formed by surfactants, polymers, etc. They guide precursor reactions, and the product is obtained after template removal.

The template method can precisely regulate the structure, morphology, and pore size of materials, prepare composite materials with special structures, and improve transport performance. However, hard templates are complex to prepare, high in cost, and their removal is likely to damage the structure; soft templates have poor stability, are greatly affected by reaction conditions, and involve complex synthesis processes, which are not conducive to large-scale production [6].

3. Application of zero-dimensional metal oxide composites in energy chemistry

Zero-dimensional metal oxide composites are widely used in the field of energy conversion and storage, as they can enhance energy utilization efficiency and storage performance. Below is an introduction to their applications in four typical areas.

3.1. Photocatalytic energy conversion

Photocatalytic technology converts solar energy into chemical energy, playing a significant role in hydrogen

production and CO_2 reduction. Single 0D metal oxides have limited photocatalytic performance, but their performance is significantly improved after compositing.

Composite materials can expand the light absorption range and facilitate the separation of photogenerated carriers. In photocatalytic hydrogen production, take the TiO_2/graphene composite as an example: TiO_2 absorbs photons to generate electron-hole pairs, while graphene enables rapid electron transfer, reducing the recombination of electron-hole pairs in the composite and thus improving catalytic efficiency. In photocatalytic CO_2 reduction, composite materials can convert CO_2 into organic fuels as a resource. For instance, the carbon quantum dot/ZnO composite uses carbon quantum dots to enhance the light absorption and electron transport of the composite, thereby accelerating the rate of CO_2 conversion into methane and methanol and reducing greenhouse gas content [7].

3.2. Fuel cells

Fuel cells convert chemical energy into electrical energy through electrocatalytic reactions, featuring high efficiency and cleanliness. The oxygen reduction reaction is a crucial step in the cell reaction, and the activity of the catalyst directly affects the efficiency of the fuel cell. 0D metal oxide composites can be used as fuel cell catalysts or catalyst supports to improve the catalytic activity and stability of fuel cell catalysts.

By compositing Pt-based catalysts with 0D metal oxides (such as CeO_2 and Fe_3O_4), the metal oxides can modify the electronic configuration of Pt to enhance catalytic activity, while also dispersing the catalyst, reducing the amount of Pt used, and lowering catalyst costs. As supports, composite supports composed of carbon materials and 0D metal oxides (e.g., carbon nanotube/Fe_3O_4) possess both electrical conductivity and stability, providing good support for catalysts, accelerating the diffusion of reactants and electron transport, and extending the service life of fuel cells [8].

3.3. Lithium-ion batteries

Lithium-ion batteries are commonly used energy storage devices, widely applied in electronic equipment and electric vehicles. The performance of electrode materials determines the battery's capacity and cycle stability. Zero-dimensional metal oxide composites find applications in both cathode and anode materials.

Regarding anode materials, the Si-based anode material composite with zero-dimensional TiO_2 allows TiO_2 to suppress the volume expansion of Si during charging and discharging, thereby enhancing cycling stability; meanwhile, the electrical conductivity of zero-dimensional TiO_2 improves the electron transport property of Si, boosting the rate performance of the battery. For cathode materials, the $LiFePO_4$ cathode material composite with zero-dimensional ZnO enables ZnO to increase the electronic conductivity and ion diffusion rate of the cathode material, which in turn improves battery capacity and charge-discharge efficiency, while also enhancing material stability and extending battery lifespan.

3.4. Supercapacitors

Supercapacitors possess characteristics such as high power density, fast charge-discharge rate, and long cycle life. Electrode materials are a crucial component of supercapacitors, and zero-dimensional metal oxide composites, when used as electrode materials, can enhance capacitive performance and stability.

On the one hand, composites have a high specific surface area, which can provide more active sites for charge storage. Take MnO_2/carbon material composites as an example: zero-dimensional MnO_2 has high pseudocapacitance, while carbon materials improve electrical conductivity and specific surface area; the

synergistic effect of the two can increase the specific capacitance of supercapacitors. On the other hand, the excellent electrical conductivity and structural stability of composites can accelerate the charge transfer rate and reduce structural damage during the charge-discharge process. For RuO_2/graphene composites, graphene's good electrical conductivity promotes charge transfer, and RuO_2 provides high pseudocapacitance, enabling supercapacitors to achieve both high power density and high energy density [9].

4. Current research challenges and future outlook

4.1. Current research challenges

First, materials exhibit poor dispersibility and significant agglomeration. Zero-dimensional metal oxide particles have high surface energy, making them prone to agglomeration during synthesis and application. This reduces their specific surface area and inhibits the exertion of their performance. Although measures such as surface modification and the introduction of dispersants have alleviated the issue to some extent, the effect is not obvious, and controlling their dispersibility remains more difficult.

Second, the interface bonding strength is low. The interface bonding state of different components in composites determines their electron transport and synergistic effects. Weak interface bonding can easily lead to component separation of composites, reducing material stability and performance. At present, the approaches for interface regulation are relatively single, making it difficult to achieve strong and stable interfacial bonding between different components [10].

Third, large-scale preparation and industrial application are difficult. Most of the existing synthesis methods are only suitable for small-batch preparation in laboratories, such as the sol-gel method and template method. Meanwhile, factors such as long synthesis time, high preparation cost, and complex preparation methods make these methods unable to meet the needs of large-scale industrial production. Additionally, the performance stability and consistency of materials are difficult to guarantee in large-scale production, and industrial application is hindered due to limitations in material preparation.

4.2. Future outlook

In terms of material structure design and performance improvement, it is possible to precisely regulate the particle size, morphology, and component distribution of materials to form special structures (core-shell structures, porous structures), thereby further increasing the specific surface area and electron transport performance of materials. Additionally, the mechanism of synergistic effects between components can be further explored, and new composite systems can be developed—such as composites with two-dimensional materials or metal-organic framework materials—to further expand material properties and application fields.

Regarding the innovation of synthesis technologies, first, it is necessary to break through and master synthesis technologies for high-yield, low-cost, and large-scale production. Combining the concept of green chemistry, low-toxicity reaction systems should be developed to reduce energy consumption. Second, greater efforts should be made to explore the automation and intellectualization of the synthesis process to achieve the consistency and stability of products, so as to meet the needs of industrial production. Furthermore, in the promotion of industrial applications, it is essential to advance industry-university-research technology alliances and promptly translate basic laboratory research and development into industrialization. Specific composite materials should be produced according to different energy application needs, such as high-capacity lithium-ion battery electrodes and high-

efficiency fuel cell catalysts. In addition, material performance evaluation standards and material quality standards should be established to ensure material performance and safety, thereby realizing more applications of zero-dimensional metal oxide composites in energy chemistry.

5. Conclusion

The unique structural and performance advantages of zero-dimensional metal oxide composites bring broad application prospects to the field of energy chemistry. Their preparation processes and related application explorations are of positive significance for alleviating the energy crisis and promoting sustainable development. This paper summarizes the sol-gel method, hydrothermal/solvothermal method, precipitation method, and template method, and presents the advantages, disadvantages, and applicable scopes of each synthesis technology. Meanwhile, it points out the applications and values of zero-dimensional metal oxide composites in photocatalytic energy conversion, fuel cells, lithium-ion batteries, and supercapacitors, and clearly identifies the current existing problems and challenges, including the inhomogeneity of such materials, interface connection, and large-scale fabrication. Furthermore, the paper looks forward to future structural design, preparation technology, and commercial applications. It is believed that with the continuous research and expansion of researchers on zero-dimensional metal oxide composites in energy chemistry, these materials will achieve greater economic and social benefits in practical applications, thereby making due efforts for the construction of a clean energy environment and an efficient energy system.

Disclosure statement

The author declares no conflict of interest.

References

[1] Fang D, 2021, Preparation of Low-Dimensional Metal Oxides and Their Performance in Electrocatalytic CO2 Reduction and Water Splitting, dissertation, Tianjin University of Technology.

[2] Hu SS, 2020, Study on Ion Transport and Mechanical Behavior of Low-Dimensional Metal Oxide Electrode Materials, dissertation, Wuhan University.

[3] Qin YX, 2019, Study on Gas Recognition, Property Regulation and Mechanism of Low-Dimensional Metal Oxides, dissertation, Tianjin University.

[4] Wang X, 2022, Preparation of Zero-Dimensional Quantum Dot/Two-Dimensional Nanosheet Composites and Their Photocatalytic and Gas-Sensing Properties, dissertation, Lanzhou University.

[5] Liu MM, Cai C, Zhang ZJ, et al., 2023, Transition Metal Oxides Supported on Nanocarbon Materials as Supercapacitor Electrode Materials. Materials Reports, 33(01): 103–109.

[6] Xu CL, 2022, Preparation of One-Dimensional Metal Oxide/Graphene Nanoscroll Core-Shell Composites and Their Electrochemical Properties, dissertation, Harbin Engineering University.

[7] Wang YB, 2023, Preparation of One-Dimensional Metal Oxide and Hydroxide Nanoarray Materials by Bipolar Binary Liquid Phase Method and Their Performance Study, dissertation, Harbin Engineering University.

[8] Cai J, 2023, Analysis of Physicochemical Properties of Low-Dimensional Metal Oxides. Physics Teaching Reference for Middle Schools, 46(02): 55–56.

[9] Ding LP, 2022, Synthesis of Zero-Dimensional Metal Oxide Composites and Their Application in Energy Chemistry, dissertation, Hefei University of Technology.

[10] Jiang C, 2022, Synthesis and Gas-Sensing Properties of One-Dimensional Metal Oxide MO_x (M=Sn, Zn) Composites, dissertation, Heilongjiang University.

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

SW-YOLO: Lightweight Attitude Estimation Algorithm Based on Weighted Convolution and Star Network

Qian Xu*

School of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, China

**Author to whom correspondence should be addressed.*

Abstract: This paper proposes SW-YOLO (StarNet Weighted-Conv YOLO), a lightweight human pose estimation network for edge devices. Current mainstream pose estimation algorithms are computationally inefficient and have poor feature capture capabilities for complex poses and occlusion scenarios. This work introduces a lightweight backbone architecture that integrates WConv (Weighted Convolution) and StarNet modules to address these issues. Leveraging StarNet's superior capabilities in multi-level feature fusion and long-range dependency modeling, this architecture enhances the model's spatial perception of human joint structures and contextual information integration. These improvements significantly enhance robustness in complex scenarios involving occlusion and deformation. Additionally, the introduction of WConv convolution operations, based on weight recalibration and receptive field optimization, dynamically adjusts feature importance during convolution. This reduces redundant computations while maintaining or enhancing feature representation capabilities at an extremely low computational cost. Consequently, SW-YOLO substantially reduces model complexity and inference latency while preserving high accuracy, significantly outperforming existing lightweight networks.

Keywords: YOLO11-Pose; WConv; StarNet; Lightweight algorithms; Feature fusion

Online publication:

1. Introduction

As a core and highly practical foundational task in computer vision, human pose estimation has demonstrated robust application potential across diverse scenarios in recent years. Technologically, it has become deeply integrated into multiple critical domains, including abnormal behavior analysis in intelligent surveillance systems, gesture recognition for human-machine interaction in consumer electronics, posture correction assessment in medical rehabilitation, and safety monitoring for human-machine collaboration in industrial environments. However, these algorithms are primarily used on edge devices, such as embedded terminals, IoT sensors, and

mobile terminals. These devices typically have limited computational power, memory, and storage capacity, which makes supporting the complex computational demands of traditional, high-precision pose estimation algorithms challenging. This contradiction directly drives the evolution of technical requirements; the traditional pursuit of accuracy alone is no longer sufficient for edge scenarios. The industry now demands more stringent standards for the lightweight nature, real-time inference efficiency, and resource compatibility of pose estimation algorithms. Against this backdrop, the critical challenge—and the core focus of this paper's research—is to design lightweight human pose estimation algorithms that can overcome computational resource limitations on edge devices. These algorithms must simultaneously ensure the accurate detection of key pose information while meeting demands for low computational power and a minimal storage footprint.

Both academia and industry have proposed various lightweight human pose estimation solutions for edge devices [1–4]. Among these, OpenPose uses Part Affinity Fields (PAFs) to achieve high-precision joint association. However, even when adapting its optimized version for edge devices and replacing feature extraction modules with lightweight alternatives to reduce computational overhead, its efficiency remains insufficient for real-time deployment. Furthermore, OpenPose struggles with keypoint feature capture and contextual integration in scenarios involving occlusion and joint deformation [5]. MobileNetV4 is commonly used as a backbone network due to its parameter compression and enhanced local features via grouped convolution, decoupling, and channel attention mechanisms [6]. However, it suffers from weaknesses in multi-level feature fusion and modeling long-range dependencies. It has a poor perception of human joint spatial structures, and the insufficient dynamic adjustment of feature importance during convolution leads to computational redundancy. In summary, existing solutions have yet to address the core challenges of insufficient computational efficiency on edge devices and weak keypoint feature capture under complex poses. This provides direction for improving the SW-YOLO net.

This paper addresses the issues of insufficient computational efficiency in edge devices and reduced accuracy in human pose estimation under complex scenarios by proposing a lightweight pose estimation network, SW-YOLO (StarNet Weighted-Conv YOLO). By incorporating the StarNet module into the backbone network, the model's ability to represent multi-level features is enhanced, significantly improving robustness in challenging scenarios such as occlusion and motion blur [7]. Simultaneously, replacing standard convolutions with weighted convolutions (WConv) reduces computational overhead while preserving sensitivity to key feature points. Experimental results demonstrate that this approach effectively improves the accuracy and real-time performance of human pose estimation on edge devices while maintaining low computational complexity [8].

In summary, the SW-YOLO model proposed in this paper contributes as follows:

(1) The introduction of lightweight StarNet modules constructs an efficient backbone network. Through its sparse connections and parameter sharing mechanism, it significantly reduces model complexity while enhancing multi-scale feature fusion capabilities, effectively mitigating feature discrimination bias in occlusion and deformation scenarios.

(2) Replacing standard convolution with WConv allows for dynamic kernel weight optimization and receptive field adjustment. This reduces computational overhead while enhancing the ability to capture key point details, further improving the balance between model accuracy and efficiency.

2. Related work

YOLO11-Pose performs well on the COCO2017 dataset, striking an optimal balance between accuracy and speed.

However, the model faces computational efficiency challenges on edge devices and is susceptible to occlusion, motion blur, and background interference in complex scenes. These issues lead to degraded performance in detecting human keypoints. To address these issues, this paper presents a lightweight pose estimation model called SW-YOLO (StarNet Weighted-Conv YOLO). The model incorporates the StarNet module to enhance multi-scale feature fusion and spatial structure perception. This effectively suppresses complex background interference while improving keypoint discrimination under occlusion and deformation. The model simultaneously replaces standard convolutions with WConv structures to significantly reduce parameters and computational load while preserving sensitivity to fine features. Experiments demonstrate that SW-YOLO substantially improves robustness and estimation accuracy in complex, real-world scenarios while reducing computational complexity.

The backbone network of YOLO11-Pose was restructured by introducing the StarNet module and the WConv convolution operation to construct the lightweight and efficient SW-YOLO model [9]. The SW-YOLO model leverages StarNet's strengths in multi-level feature fusion and long-range dependency modeling to improve the perception of human keypoints in complex backgrounds. It simultaneously employs WConv weighted convolutions to dynamically adjust receptive fields and recalibrate features, reducing the number of parameters while improving the model's sensitivity to fine-grained features. The proposed model architecture is illustrated in **Figure 1**.

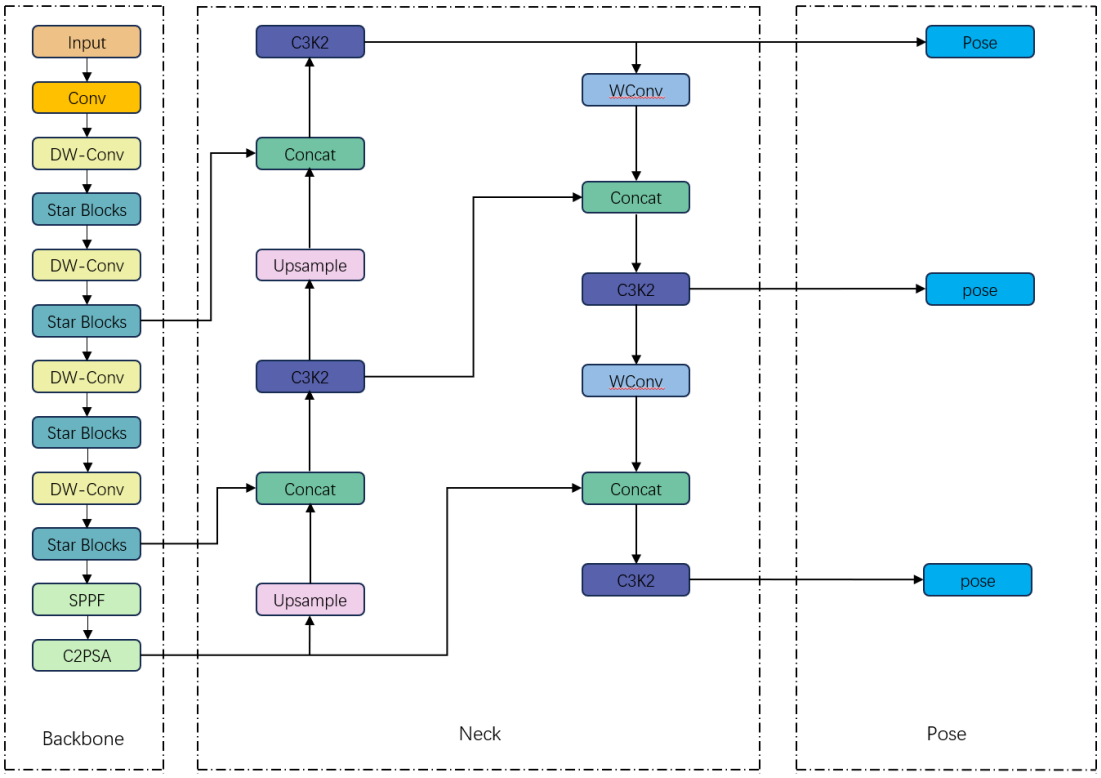


Figure 1. SW-YOLO

2.1. WConv

The WConv module is an efficient convolution operation specifically designed for lightweight neural networks. Its core functionality lies in introducing a dynamic channel attention mechanism that adaptsively recalibrates the importance of each channel in the input feature map through attention weights generated online. This module

enables the network to autonomously amplify information-rich feature channels while suppressing responses from redundant or noisy channels, thereby simulating a "visual focus" effect. This process significantly enhances the model's ability to capture and distinguish critical detail features in complex scenarios, such as occlusion, deformation, and background interference. Crucially, WConv cleverly encapsulates this powerful feature enhancement within minimal theoretical computational overhead. This design significantly boosts standard convolution representation capabilities while perfectly meeting edge devices' stringent efficiency demands, achieving an exceptional balance between accuracy and speed. Consequently, WConv replaces standard convolutions in Yolo11-Pose.

First, perform global information embedding on the input feature map $X \in \mathbb{R}^{C\times H\times W}$ perform global average pooling (GAP). Compress it into a global context vector $z \in \mathbb{R}^{C}$. This vector encodes the global statistical information for each channel.

$$z_c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} x_c\,(i,j) \tag{1}$$

Then, generate weights by feeding the global context vector z into a lightweight gating mechanism, dynamically producing a channel attention weight vector $\alpha \in \mathbb{R}^{C}$.

$$\alpha = \sigma\big(\mathbf{W}_2\delta(\mathbf{W}_1 z)\big) \tag{2}$$

Among these, W_1 and W_2 represent the weights of the fully connected layer, δ denotes the ReLU function, and σ denotes the Sigmoid function. Each element of the weight vector α ranges between 0 and 1, indicating the importance of the corresponding channel.

Next, feature reweighting is performed by multiplying the generated attention weights α with the original input feature map X on a per-channel basis, yielding a calibrated, weighted feature map $\widetilde{X}$.

$$\tilde{x}_c = \alpha_c \cdot x_c \tag{3}$$

Finally, a standard convolution operation (such as a 3×3 convolution) is performed on the weighted feature map $\widetilde{X}$, yielding the output feature map Y.

$$Y = \mathrm{Conv}\big(\widetilde{X}\big) \tag{4}$$

WConv implements a "focus first, compute later" mechanism, significantly enhancing the efficiency of convolution operations and making it highly suitable for lightweight neural network designs.

2.2. StarNet

This paper introduces the StarNet module, an efficient component of the backbone network. The StarNet module uses lightweight techniques, such as separable convolutions, batch convolutions, and parameter sharing, to create a star-shaped topology with shared central nodes and multiple parallel branches. The central nodes perform feature dimension reduction and preliminary fusion with 1×1 batch convolutions, which significantly reduces redundant computations between channels. The peripheral branches use parallel separable convolutions with different receptive fields to extract multi-scale features and capture rich contextual information while avoiding the computational burden of standard dilated convolutions. Finally, an adaptive feature recalibration mechanism enhances important feature channels. This design enables StarNet to strike a remarkable balance. It effectively enhances the model's robustness against occlusion, scale changes, and non-rigid deformations with extremely low parameter counts and computational overhead. It also significantly reduces the model's overall complexity. This

enables high-precision, real-time human pose estimation on edge devices.

StarNet is a 4-layer hierarchical structure that uses convolutional layers for downsampling and modified demo blocks for feature extraction. To meet efficiency requirements, batch normalization replaces layer normalization and is placed after deep convolutions, which can be fused during inference [10,11]. Inspired by MobileNet, a depthwise convolutional layer (DWConv) is added at the end of each block [12]. The channel expansion factor is consistently set to four, doubling the network width at each stage. Within the demo blocks, GELU activation is replaced with ReLU6, following MobileNetv2's design [13]. **Figure 2** illustrates the StarNet framework.

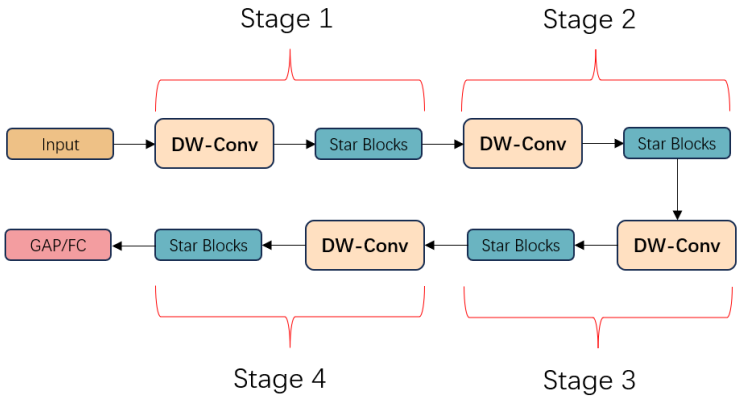


Figure 2. StarNet network

The Star Block efficiently and powerfully constructs lightweight models by stacking multiple fundamental building blocks. Its core "Star Operation" has a mathematical mechanism similar to kernel methods [14–16]. It explicitly constructs higher-order feature interaction terms through systematic, pairwise multiplication operations across different feature channels [17]. This process closely resembles the expansion of polynomial kernel functions. When embedded within neural network architectures and stacked across multiple layers, each layer enables exponential growth in the complexity of feature representations. Consequently, Star Operations achieves ultra-high-dimensional feature representations within an extremely compact feature space, even with just a few stacked layers [18]. This significantly reduces the number of parameters while preserving strong model expressiveness. The architecture is illustrated in **Figure 3**.

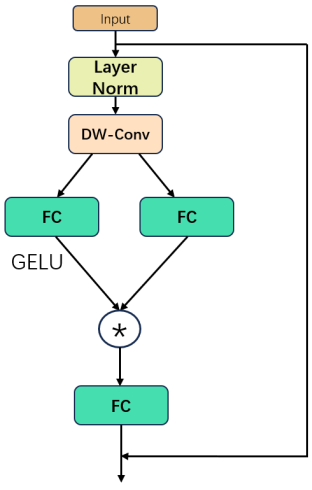


Figure 3. Star block

3. Experimental design and results

3.1. Details of the experiment

(1) Introduction to the dataset

The experiments use the COCO2017 human keypoint dataset, which contains labeled image data covering complex scenes, such as multiple people, occlusion, and diverse actions, with 17 keypoints. The dataset includes 58,945 images, of which 56,599 are in the training set and 2,346 are in the validation set.

(2) Evaluation metrics

Our method was systematically validated and evaluated on the COCO2017 dataset. This work employs Object Keypoint Similarity (OKS) as the core metric for evaluating model performance. OKS comprehensively and accurately reflects the performance of pose estimation models by calculating the normalized distance between predicted and ground-truth keypoints, while accounting for the annotation difficulty and scale variations across different keypoints. Its specific calculation formula is as follows:

$$OKS = \frac{\sum_i \exp(-\frac{d_i^2}{2s^2k_i^2})\delta(v_i>0)}{\sum_i \delta(v_i>0)} \tag{5}$$

Where d_i denotes the Euclidean distance between the detected keypoint and its true label, v_i denotes the visibility flag bit of the true keypoint on the surface, s denotes the scale of the target, and k_i denotes the normalization factor of the i^(th) human keypoint. This indicates that a keypoint is a positive example if its OKS is greater than a certain threshold. AP50 is the prediction accuracy at a 0.5 similarity threshold. AP50-95 is the average precision (AP) calculated at multiple OKS thresholds (from 0.5 to 0.95 in 0.05 increments).

(3) Experimental environment and settings

The experimental environment is as follows: the operating system is Windows 10, and the GPU is an NVIDIA Tesla T4 that utilizes parallel computing to accelerate model training and inference. The optimizer uses the stochastic gradient descent (SGD) algorithm. The initial learning rate is 0.01, the batch size is 32, and there are 100 iteration rounds.

3.2. Ablation experiments

To validate the effectiveness of each proposed improvement module, we conducted systematic ablation experiments on the COCO2017 human keypoint dataset. We compared four configurations: the baseline model, the model with the StarNet module, the model with the WConv module, and the model incorporating both StarNet and WConv modules. The experimental results are shown in **Table 1**.

Table 1. Ablation experiments on the COCO2017 dataset

Methods	Params/M	GFLOPs	AP50	AP50-95
YOLO11n-Pose	StarNet	WConv	Params/M	GFLOPs
√			2.91	7.7
√	√		2.27	6.1
√		√	2.86	7.2
√	√	√	2.23	6.3

Research indicates that using the StarNet backbone network alone reduces the number of parameters by 0.64 million, while lowering AP50 by 0.7% and computational load by 1.6 GFLOPs, demonstrating improved lightweight efficiency and computational effectiveness. Using WConv alone reduces parameters by 0.5M, lowers AP50 by 0.2%, and decreases computational load by 0.5GFLOPS. This minimizes redundant computations while maintaining or even enhancing feature representation capabilities at extremely low computational overhead. Integrating both modules into the YOLOv11n-Pose framework reduces parameters by 0.68M, lowers AP50 by 0.5%, and decreases computational load by 1.4GFLOPS. This demonstrates the superiority of the proposed SW-YOLO model, effectively reducing model parameters while enhancing its ability to capture keypoint details.

3.3. Comparison experiment

To visually validate the superiority of the SW-YOLO model, comparative training experiments were conducted. As shown in **Table 2**, compared to YOLO series pose estimation models, the proposed model achieves optimal performance in both parameter count and computational complexity. While its accuracy is slightly lower than that of YOLO series algorithms, this reduction is an unavoidable consequence of significantly decreasing the number of parameters.

Table 2. Comparison experiments on the COCO2017 dataset

Methods	Params/M	GFLOPs	AP50	AP50-95
Yolov5n-Pose	2.80	8.3	72.6	39.6
Yolov8n-Pose	3.30	9.2	75.6	42.7
Yolov10n-Pose	2.55	7.8	75.3	42.6
Yolov11n-Pose	2.91	7.7	74.2	42.0
Yolov12n-Pose	2.88	7.8	74.9	41.2
Ours (SW-YOLO)	2.23	6.3	73.7	40.8

4. Conclusion

To address the high computational complexity of current pose estimation algorithms and the challenges of deploying them on resource-constrained edge devices, this paper proposes a novel lightweight network architecture, SW-YOLO. Building upon the YOLOv11n-Pose baseline architecture, this model enhances multi-scale feature fusion and spatial structure modeling capabilities by incorporating the StarNet module. It employs WConv to dynamically adjust the receptive field and recalibrate features, significantly reducing model parameters while strengthening the capture of fine-grained details. Experiments demonstrate that SW-YOLO effectively improves pose estimation accuracy in complex scenarios involving occlusion and deformation while maintaining low computational complexity, achieving an optimized balance between precision and efficiency.

Disclosure statement

The author declares no conflict of interest.

References

[1] Ma N, Zhang X, Zheng HT, et al., 2018, ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design, in Ferrari V, Hebert M, Sminchisescu C, Weiss Y, (eds) Computer Vision – ECCV 2018. Lecture Notes in Computer Science, vol 11218, Springer, Cham.

[2] Zhang X, Zhou X, Lin M, et al., 2017, ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. arXiv. https://arxiv.org/abs/1707.01083

[3] Yu C, Xiao B, Gao C, et al., 2021, Lite-HRNet: A Lightweight High-Resolution Network, 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[4] Kreiss S, Bertoni L, Alahi A, 2022, OpenPifPaf: Composite Fields for Semantic Keypoint Detection and Spatio-Temporal Association. IEEE Transactions on Intelligent Transportation Systems, 23(8): 13498–13511.

[5] Cao Z, Hidalgo G, Simon T, et al., 2021, OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(1): 172–186.

[6] Howard AG, Zhu M, Chen B, et al., 2017, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv. https://arxiv.org/abs/1704.04861

[7] Ma X, Dai X, Bai Y, et al., 2024, Rewrite the Stars, 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[8] Cammarasana S, 2025, Optimal Weighted Convolution for Classification and Denosing. arXiv. https://arxiv.org/abs/2505.24558

[9] Khanam R, Hussain M, 2024, YOLOv11: An Overview of the Key Architectural Enhancements. arXiv. https://arxiv.org/abs/2410.17725

[10] Ioffe S, Szegedy C, 2015, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, Proceedings of the 32nd International Conference on Machine Learning, PMLR 37: 448–456.

[11] Ba JL, 2016, Layer Normalization. arXiv. https://arxiv.org/abs/1607.06450

[12] Chollet F, 2016, Xception: Deep Learning with Depthwise Separable Convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1800–1807.

[13] Sandler M, Howard A, Zhu M, et al., 2018, MobileNetV2: Inverted Residuals and Linear Bottlenecks, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[14] Wang J, Sun K, Cheng T, et al., 2021, Deep High-Resolution Representation Learning for Visual Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(10): 3349–3364.

[15] Yuan F, Lin Z, Tian Z, et al., 2025, Bio-Inspired Hybrid Path Planning for Efficient and Smooth Robotic Navigation. Int J Intell Robot Appl.

[16] Liang B, Yuan F, Deng J, et al., 2025, Cs-pbft: A Comprehensive Scoring-Based Practical Byzantine Fault Tolerance Consensus Algorithm. J Supercomput, 81: 859.

[17] Yuan F, Huang X, Jiang H, et al., 2025, An xLSTM–XGBoost Ensemble Model for Forecasting Non-Stationary and Highly Volatile Gasoline Price. Computers, 14: 256.

[18] Zhang K, Yuan F, Jiang Y, et al., 2025, A Particle Swarm Optimization-Guided Ivy Algorithm for Global Optimization Problems. Biomimetics (Basel), 10(5): 342.

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Lightweight Multi-Object Detection for Construction Sites Based on YOLO-World

Bing Chen*

School of Computer Science and Technology, Taiyuan Normal University, Taiyuan 030619, China

**Author to whom correspondence should be addressed.*

Abstract: Addressing the current issues in construction site detection algorithms—such as missed detections, false positives, and high model complexity—caused by occlusions and scale variations in dense environments. This paper proposes a lightweight multi-object detection model for construction sites based on YOLO-World, named the LCS-YOLO model, to achieve a balance between detection efficiency and accuracy. We propose the RGNet (Re-parameterization GhostNet) module, which integrates re-parameterized convolutions and a multi-branch architecture. This approach addresses the issue of information redundancy in intermediate feature maps while enhancing feature extraction and gradient flow capabilities. Combined with the adaptive downsampling module ADown (Adaptive Downsampling), it better captures image features and achieves spatial compression, reducing model complexity while enhancing interaction between images and text. Experiments demonstrate that the LCS-YOLO model outperforms other comparison models in overall performance, achieving a balance between accuracy and efficiency.

Keywords: Dense; LCS-YOLO; RGNet; ADown; Re-parameterized

Online publication: October 17, 2025

1. Introduction

In the construction industry, to ensure worker safety, workers are required to wear basic protective gear such as hard hats and reflective vests. Therefore, ensuring workers wear safety equipment is a key aspect of construction site safety management. If workers lack sufficient safety awareness, manual inspections of protective gear usage become necessary. However, this approach is not only inefficient but also incurs extremely high labor costs. Early methods of using sensors to detect whether workers were wearing safety equipment were limited by high costs and operational inconvenience [1–3].

With the advancement of image processing technologies, particularly the progress in deep learning and convolutional neural networks [4–6], construction site detection has experienced rapid development. The YOLO [7] series, as a representative of object detection, has undergone continuous iterations and integrated new technologies

such as FPN [8], PANet [9], and RepVGG [10] to optimize detection speed and accuracy. Additionally, Yi *et al.* [11] employed YOLOv5 as the baseline model and added an extra detection head to capture small object features, thereby enhancing the recognition capability for small objects. However, this also resulted in an increase in the number of model parameters.

Although the aforementioned research has achieved promising results in detection, for construction site safety equipment inspection, reducing redundant information to enhance feature extraction and minimizing model complexity remain critical factors in detection. This paper proposes the RGNet module, combined with the adaptive downsampling module ADown [12], to introduce the LCS-YOLO model. This model effectively reduces redundant information while enhancing feature extraction and gradient flow capabilities, achieving a balance between accuracy and efficiency.

In summary, the main contributions of the LCS-YOLO model proposed in this paper are as follows:

(1) The RGNet module proposed for backbone networks integrates re-parameterized convolutions and multi-branch structures. This approach addresses information redundancy in intermediate feature mappings within the C2F module while enhancing feature extraction and gradient flow capabilities.

(2) The introduction of the ADown module replaces the original downsampling module, enabling more effective extraction of image features and spatial compression. This approach reduces model complexity while enhancing interaction between images and text.

2. Related work

Dense pedestrian detection typically involves a trade-off between feature extraction requirements and computational resource constraints, particularly in edge devices or real-time applications where model compactness and efficiency are paramount. We found that YOLO-World [13] tends to introduce significant information redundancy when processing image features via C2F. The original downsampling primarily employed convolution operations with a stride of 2, which tended to filter out fine-grained features. This approach may limit generalization capabilities in complex scenarios such as multi-scale objects. Additionally, the model's complexity was relatively high due to the introduction of visual-language modeling, making deployment on construction sites more challenging.

In the LCS-YOLO model, we propose the RGNet module, which integrates re-parameterized convolutions [14] and a multi-branch architecture. This addresses the information redundancy present in the intermediate feature maps of the C2F module, thereby enhancing feature extraction and gradient flow capabilities. The introduction of the ADown module addresses the issue of fixed subsampling in the original model, which tends to filter out fine-grained features. It simultaneously acquires image features and performs spatial compression to reduce computational load, thereby enhancing interaction between images and text. The overall structure of the model is shown in **Figure 1**.

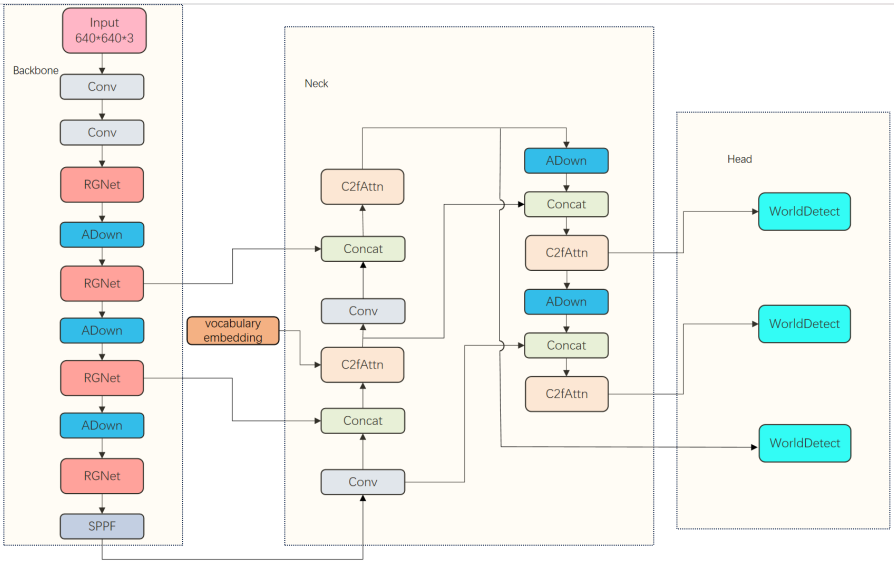


Figure 1. Overall structure

2.1. RGNet model

In the backbone network, when performing feature mapping with feature maps P3, P4, and P5 obtained at different scales, ensuring detection accuracy inevitably leads to redundant computations and underutilization of features. This results in a high computational load for the model, which can cause errors during subsequent object detection tasks, particularly when detecting dense crowds. We propose the RGNet module, which abandons the original Bottleneck module. To compensate for the performance loss resulting from discarding residual blocks, we employ re-parameterized convolutions on the gradient flow branch to enhance feature extraction and gradient propagation capabilities. Reduce the number of parameters and computational complexity while preserving the expressive power of key features, as shown in **Figure 2**.

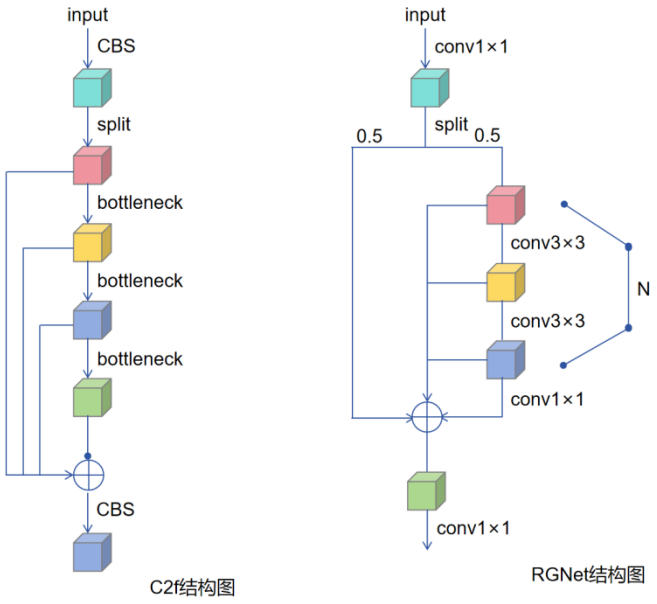


Figure 2. C2f and RGNet structures

RepConv primarily reduces computational complexity by fusing convolution and normalization operations. Standard convolution is expressed as in Equation 1:

$$Y=X\cdot W+b \tag{1}$$

Here, X represents the input feature map, W denotes the convolutional weights, and b is the bias term.

Batch normalization is defined as shown in Equation 2:

$$\begin{aligned}\hat{x}_i &= \gamma\cdot\frac{x_i-\mu}{\sqrt{\sigma^2+\varepsilon}}+\beta \\ &= \frac{\gamma}{\sqrt{\sigma^2+\varepsilon}}\cdot x_i+\left(\beta-\frac{\gamma\cdot\mu}{\sqrt{\sigma^2+\varepsilon}}\right)\end{aligned} \tag{2}$$

The combined convolution formula is shown in Equations 3 and 4:

$$w_{BN}=\frac{\gamma}{\sqrt{\sigma^2+\varepsilon}},b_{BN}=\beta-\frac{\gamma\cdot\mu}{\sqrt{\sigma^2+\varepsilon}} \tag{3}$$

$$\begin{cases}w=w_{BN}\cdot w_{conv}\\ b=w_{BN}\cdot b_{conv}+b_{BN}\end{cases} \tag{4}$$

γ and β represent learnable parameters. μ denotes the mean of the input, σ^2 indicates the variance of the input, and ε serves as a smoothing term to prevent the denominator from becoming zero.

Drawing inspiration from GhostNet, mainstream CNNs exhibit extensive redundancy in their intermediate feature maps. Despite the high correlation among these features, leveraging them can enhance the richness and robustness of feature representations, thereby improving detection performance. RGNet generates partially redundant feature maps through channel splitting and lightweight convolution operations. The main branch extracts features via re-parameterized convolutions, followed by multi-branch lightweight convolutions to extract deep features from redundant information, culminating in feature concatenation. As shown in Equation 5:

$$Y=\mathrm{Conv}_{1\times1}\left(\mathrm{Concat}\left(\mathrm{Conv}_{1\times1},\{\mathrm{Conv}_{3\times3}^{(i)}\}_{i=1}^{n},\mathrm{Conv}_{1\times1}(\cdot)\right)\right) \tag{5}$$

$\{\mathrm{Conv}_{3\times3}^{(i)}\}_{i=1}^{n}$ represents a series of stacked 3×3 convolution operations, serving as the final output after pooling.

The RGNet module enhances feature extraction and gradient diversion through re-parameterized convolutions and multi-branch structures. By employing scaling factors to control module size, it effectively reduces computational load and parameter count while preserving feature expressiveness, making it well-suited for multi-object detection in construction sites.

2.2. ADown

In detection tasks, small-sized targets are often embedded within complex backgrounds, exhibiting distinct scale differences and texture discontinuities. In YOLO-World, downsampling is primarily achieved using convolutions with a stride of 2,although highly efficient, this fixed structural approach inevitably loses some detailed information, particularly the characteristics of small targets or the edge details of objects. ADown employs a dual-branch architecture to simultaneously perform feature extraction and edge information preservation, forming a lightweight downsampling mechanism that integrates convolution and pooling for joint modeling. This addresses

the limitation of fixed downsampling in the original model, which tends to filter out fine-grained features. It achieves spatial compression while capturing image features, thereby reducing computational complexity and facilitating interaction between images and text (**Figure 3**).

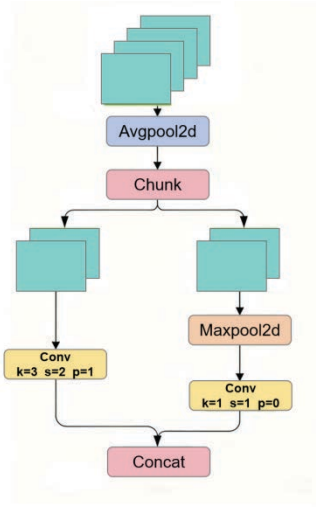


Figure 3. ADown structure

First, the input feature map undergoes average pooling to achieve global compression. Subsequently, it is divided equally along the spatial dimension. One branch performs downsampling and local context modeling through a stride-2 convolution, as shown in Equation 6:

$$F_1=\mathrm{Conv}_{3\times3}^{s=2,p=1}(X_1) \tag{6}$$

The kernel size is 3×3 with 1-based padding.

Another approach first extracts features from edges and regions with strong responses through max pooling operations, followed by downsampling via convolutions with a stride of 1, as shown in Equation 7:

$$F_2=\mathrm{Conv}_{1\times1}^{s=1,p=0}\left(\mathrm{MaxPool}_{2\times2}(X_2)\right) \tag{7}$$

Finally, the outputs from both branches are concatenated using Concat to form a downsampling mechanism that combines multi-scale feature representation with gradient stability.

This module effectively minimizes feature information loss during the downsampling stage while maintaining a lightweight design, enhancing the ability to preserve multiple targets and improving the discrimination accuracy of subsequent detection branches.

3. Experiment

3.1. Configuration and datasets

Our model was developed based on Ultralytics. Experiments were conducted using NVIDIA RTX 5080 GPUs, employing the open-source CLIP text encoder with pre-trained weights to encode input text.

The construction site dataset we employed comprises five categories: nohat, helmet, people, reflect, and workcloth. It consists of 2,204 training images and 551 validation images. During training, the initial learning rate

was set to 0.002, with a weight decay of 0.05. The model was trained for 100 epochs using a batch size of 32.

3.2. Test metrics

To evaluate and improve algorithm metrics, we employ common evaluation indicators [15] such as mean average precision (mAP), number of parameters (Params), and floating-point operations per second (FLOPs). These metrics reflect both model performance and resource consumption.

mAP is a key indicator to evaluate the performance of the object detection model, mAP is mainly divided into mAP@0.5 and mAP@0.5–0.95, mAP@0.5 indicates that the average accuracy of each class is calculated when the IOU threshold is 0.5, and then the average value of all categories is taken, and mAP@0.5–0.95 represents the average mAP value under different IOU thresholds (0.5–0.95, step size is 0.05). Params is expressed as the memory occupied by the model, usually in millions (M). FLOPs are used to measure the computational complexity of a model, usually measured in GFLOPs (billions of floating-point calculations per second).

3.3. Comparative experiments

To further validate the detection performance of our proposed improved algorithm, we conducted comparative experiments using multiple popular object detection models on the construction site dataset. The experimental results are shown in **Table 1**.

Table 1. Comparative experiments

Model	mAP@0.5%	mAP@0.5–0.95%	Params/M	FLOPs/10^9
Yolov5	80.4	55.3	2.65	7.37
Yolov8	83.0	57.8	3.15	8.7
Yolov10 [16]	79.5	54.1	2.30	6.7
Yolov11 [17]	82.8	58.2	2.62	6.6
YOLO-Worldv8	82.4	58.0	3.69	10.0
Ours	83.0	58.9	2.85	8.0

As shown in **Table 1**, compared to YOLO-Worldv8s, the improved algorithm achieves a 22.8% reduction in parameters and a 20% reduction in computational complexity, mAP@0.5% and mAP@0.5–0.95% improved by 0.6% and 0.9%, respectively. Compared to other mainstream YOLO series models, it maintains comparable parameters and FLOPs while achieving a 3.5% improvement over YOLOv10 and a 0.2% improvement over YOLOv11 in mAP@0.5. The improved algorithm proposed in this paper demonstrates enhanced efficiency by achieving greater accuracy while reducing the number of parameters and computational complexity. This performance advantage stems from the judicious introduction of innovative modules and the optimized design of the overall structure.

3.4. Ablation experiment

To further validate the algorithm's effectiveness, we conducted a series of ablation experiments to investigate the specific impact of each module on model performance. The experimental results are shown in **Table 2**.

Table 2. Ablation experiment

Worldv8	RGNet	ADown	mAP@0.5%	mAP@0.5–0.95%	Params/M	FLOPs/10^9
√			82.4	58.0	3.69	10.0
√	√		82.3	57.5	3.27	8.7
√		√	83.9	59.8	3.28	9.3
√	√	√	83.0	58.9	2.85	8.0

As shown in **Table 2**, introducing each module yields varying performance improvements over the baseline YOLO-Worldv8 model. After incorporating RGNet attention, the model's accuracy remains largely unchanged while reducing parameters by 11% and computational complexity by 13%. After introducing the ADown module, the model parameters were reduced by 11%, computational complexity decreased by 7%, and mAP@0.5–0.95% improved by 1.8%. Ultimately, our model achieved a 0.9% increase in mAP@0.5–0.95% while reducing parameters by 22.8% and computational complexity by 20%, striking a balance between accuracy and lightweight performance.

3.5. Visualization experiments

To thoroughly evaluate the adaptability and robustness of the LCS-YOLO model in real construction site scenarios, this paper selected complex background and dense crowd scenes to conduct a visual comparative analysis, as shown in **Figure 4**.

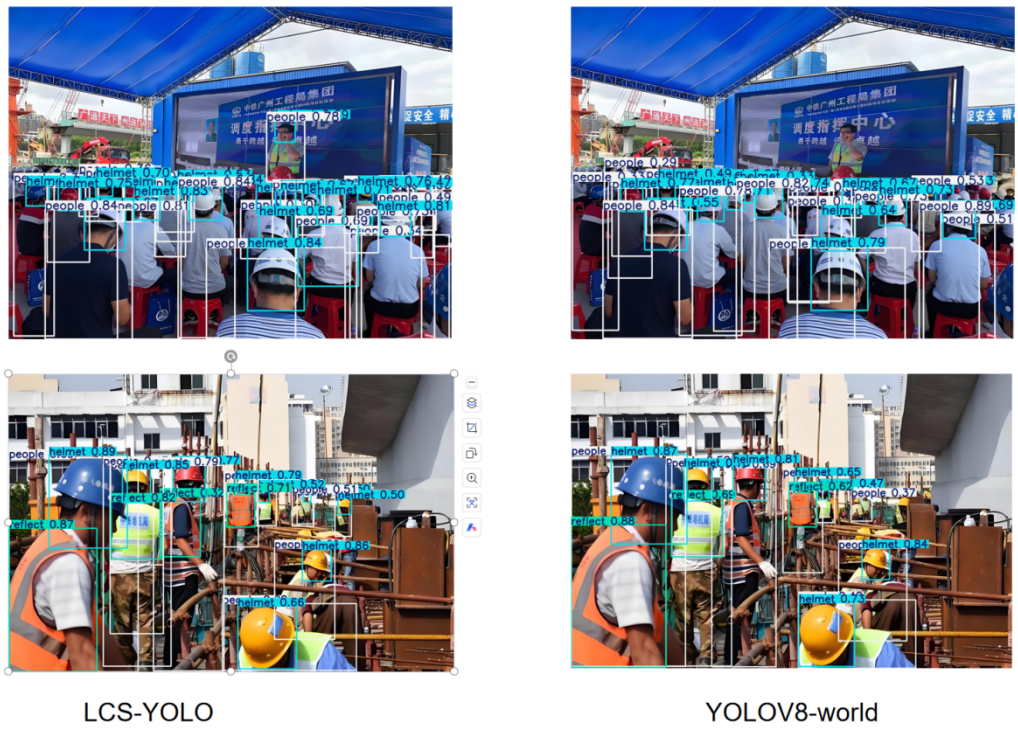


Figure 4. Visual comparison chart

Figure 4 demonstrates the detection performance of LCS-YOLO and YOLOV8-World in complex backgrounds and dense crowd scenes. The figure reveals that YOLOV8-World exhibits missed detections when objects are too small. The results indicate that LCS-YOLO achieves greater lightweight efficiency while demonstrating superior object perception and background suppression capabilities.

4. Conclusion

This paper addresses the information redundancy in the intermediate feature maps of the C2F module within the YOLO-World model and the loss of effective information during fixed-stride dimensionality reduction, which hinders its efficient application in construction site detection. By introducing the RGNet module, which employs re-parameterized convolutions on the gradient flow branch, the model enhances both feature extraction and gradient flow capabilities. While reducing the number of parameters and computational complexity, the model preserves the expressive power of key features. The introduction of the ADown subsampling mechanism employs a dual-branch path design to simultaneously achieve feature extraction and edge information retention. This forms a lightweight subsampling approach that combines convolution and pooling modeling, enhancing the model's perception of multi-scale objects to facilitate interaction between images and text.

The improved model achieves a 7.9% reduction in parameters and a 2.8% decrease in computational complexity while maintaining comparable accuracy. This advancement significantly lowers model complexity without compromising precision, making it better suited for on-site detection requirements.

Disclosure statement

The author declares no conflict of interest.

References

[1] Kelm A, Laußat L, Meins-Becker A, et al., 2013, Mobile Passive Radio Frequency Identification (RFID) Portal for Automated and Rapid Control of Personal Protective Equipment (PPE) on Construction Sites. Automation in Construction, 3638–3652.

[2] Yuan F, Lin Z, Tian Z, et al., 2025, Bio-Inspired Hybrid Path Planning for Efficient and Smooth Robotic Navigation. Int J Intell Robot Appl.

[3] Liang B, Yuan F, Deng J, et al., 2025, Cs-pbft: A Comprehensive Scoring-Based Practical Byzantine Fault Tolerance Consensus Algorithm. J Supercomput, 81: 859.

[4] Yuan F, Huang X, Jiang H, et al., 2025, An xLSTM–XGBoost Ensemble Model for Forecasting Non-Stationary and Highly Volatile Gasoline Price. Computers, 14: 256.

[5] Girshick R, Donahue J, Darrell T, et al., 2014, Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Columbus, 580–587.

[6] Zhang K, Yuan F, Jiang Y, et al., 2025, A Particle Swarm Optimization-Guided Ivy Algorithm for Global Optimization Problems. Biomimetics (Basel), 10(5): 342.

[7] Sohan M, Sai Ram T, Reddy R, et al., 2024, A Review on YOLOv8 and Its Advancements, International Conference on Data Intelligence and Cognitive Informatics, Springer, Singapore, 529–545.

[8] Bochkovskiy A, Wang CY, Liao HYM, 2020, YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv. https://arxiv.org/abs/2004.10934

[9] Liu S, Qi L, Qin HF, et al., 2018, Path Aggregation Network for Instance Segmentation, Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Salt Lake City, 8759–8768.

[10] Ding X, Zhang XY, et al., 2021, RepVGG: Making VGG-Style ConvNets Great Again, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[11] Zhang YJ, Xiao FS, Lu ZM, 2022, Helmet Wearing State Detection Based on Improved YOLOv5s. Sensors, 22(24): 9843.

[12] Fang S, Chen C, Li Z, et al., 2024, YOLO-ADual: A Lightweight Traffic Sign Detection Model for a Mobile Driving System. World Electric Vehicle Journal, 15(7): 323.

[13] Cheng T, Song L, et al., 2024, Yolo-World: Real-Time Open-Vocabulary Object Detection, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 16901–16911.

[14] Ding X, Zhang X, Ma N, et al., 2021, RepVGG: Making VGG-Style ConvNets Great Again, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 13733–13742.

[15] Jiang B, Luo R, Mao J, et al., 2018, Acquisition of Localization Confidence for Accurate Object Detection, Proceedings of the European Conference on Computer Vision (ECCV), 784–799.

[16] Wang A, Chen H, Liu LH, et al., 2024, YOLOv10: Real-Time End-to-End Object Detection, Proceedings of the 38th Annual Conference on Neural Information Processing Systems 2024, NeurIPS, Vancouver, 1–28.

[17] Khanam R, Hussain M, 2024, YOLOv11: An Overview of the Key Architectural Enhancements. arXiv. https://arxiv.org/abs/2410.17725

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Persimmon Fruit Quality Grading Detection Based on an Improved YOLOv8s Lightweight Model

Haogang Wang[1], Yunge Jing[2]*

[1]School of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, Shanxi, China
[2]School of Mathematics and Information Technology, Yuncheng University, Yuncheng 044000, Shanxi, China

**Author to whom correspondence should be addressed.*

Abstract: Addressing challenges in accurately detecting persimmon fruit quality in orchards—such as reliance on manual grading, low efficiency, severe foliage obstruction, and subtle differences between quality grades—this paper proposes a lightweight persimmon detection model based on an improved YOLOv8s architecture. First, the Conv layer in the backbone network is replaced with an ADown module to reduce model complexity. Second, MSFAN is introduced in the Neck layer to fully extract texture features from persimmon images, highlighting differences between quality grades. Finally, the Wise-IoU loss function mitigates the impact of low-quality sample data on grading accuracy. The improved model accurately identifies and separates persimmons of varying quality, effectively addressing quality grading detection in complex backgrounds. This provides a viable technical approach for achieving persimmon quality grading detection.

Keywords: Persimmon quality grading; YOLOv8; Deep learning; Lightweight; Image detection

Online publication:

1. Introduction

Persimmon (*Diospyros kaki*) is a deciduous fruit tree belonging to the genus *Diospyros* within the family Ebenaceae. Native to China, it has been cultivated for over 2,000 years. As one of the world's four ancient fruit trees, persimmon holds significant economic and cultural value in East Asia [1–3]. Currently, persimmon is widely distributed across China, Japan, the Republic of Korea, and other countries, with China leading globally in both cultivation area and production volume.

Machine vision has demonstrated unprecedented development prospects in agricultural applications [4–6]. For a long time, fruit quality grading has primarily relied on manual visual inspection. This traditional method has significant limitations: judgment criteria vary among individuals, detection accuracy is low, and the workload is excessively high [7]. Advancements in image processing and machine learning have enabled researchers to integrate traditional image processing with machine learning for fruit quality assessment [8–12]. However, in natural environments, severe occlusion and overlapping among multiple fruits, coupled with partial coverage by

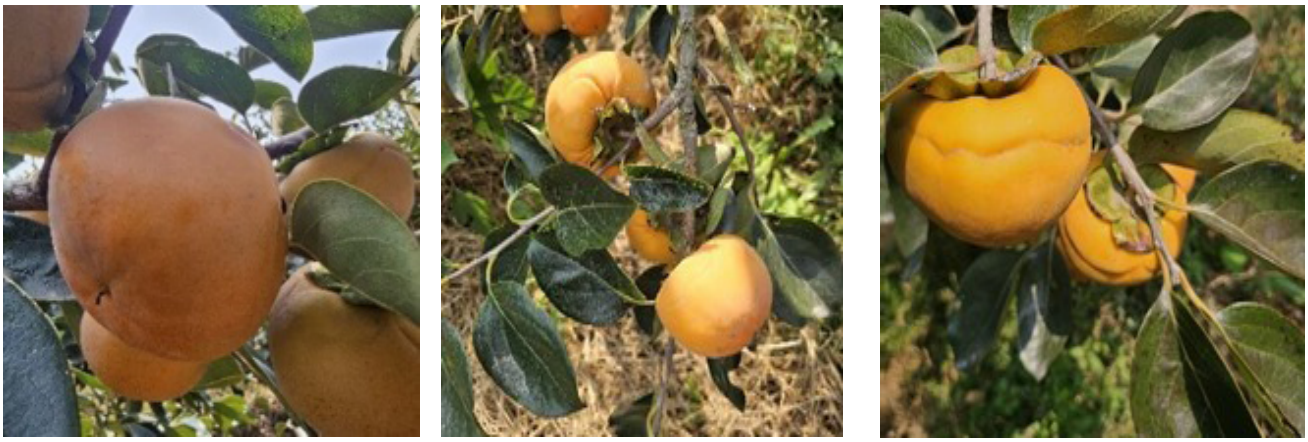
branches and foliage, pose significant challenges for target fruit detection, recognition, and spatial localization. Deep learning technology, with its powerful capability for autonomous image feature extraction, has been widely applied across various branches of computer vision [13]. Integrating object detection algorithms with agricultural product sorting has become a growing trend. Liu [14] proposed a Faster-YOLO-AP algorithm for apple detection, realized the lightweight of the model, and improved the model's mAP@50:95 to 84.12%. Zhang *et al.* [15] proposed a lightweight apple recognition method based on an improved YOLOv7 model. By introducing an ECA module to enhance detection performance, they reduced the number of parameters and computational complexity by 22.93% and 27.41%, respectively, compared to the original model. Qiu *et al.* [16] employed CNN and image segmentation to achieve automated detection and identification of surface damage on fragrant pears, achieving an accuracy rate of 91%. Sun [17] proposed an optimized residual neural network-based grading method for apple appearance quality. Using ResNet-34 as the base network, the standard convolutions within residual units were replaced with grouped convolutions to reduce the model's parameter count and computational complexity.

Although significant progress has been made in fruit quality grading algorithms, practical applications still face challenges. Some fruits exhibit insufficiently distinct grading characteristics, leading to misclassification. Conventional deep learning models struggle to extract shallow-level details, limiting overall detection accuracy. The negative impact of low-quality training samples weakens the model's generalization performance. To address these issues, this paper proposes a fruit persimmon quality grading detection model based on an improved YOLOv8s (YOLOv8s-MSFAN).

2. Materials and methods

2.1. Dataset construction

This study selected persimmons as its research subject, with images captured at orchards in Linyi County, Yuncheng City, Shanxi Province. The collection device was an OPPO smartphone, saving images in JPG format at a resolution of 4000×1800. During data collection, particular emphasis was placed on scene diversity and representativeness, capturing the dataset under varying lighting conditions and shooting angles. A total of 1,200 images were collected. After removing unqualified samples—such as blurred or disordered images—the final dataset comprised 950 images. A portion of the dataset is illustrated in **Figure 1**.



upward shot aerial view front lighting

Figure 1. Example of persimmon fruit image acquisition

To enhance the model's noise resistance, random adjustments were made to contrast and brightness, and Gaussian noise was added, resulting in a total of 1,054 images. The image samples were categorized into two classes: normal and sunken. Manual dataset annotation was performed using the open-source annotation software LabelImg, generating a text file. The dataset was then split in a 7:1:2 ratio, yielding a training set of 730 images, a validation set of 105 images, and a test set of 209 images.

2.2. YOLOv8 network model improvements

To address issues such as weak interference resistance and poor information extraction in the YOLOv8s model, this paper enhances the model's perception and capture capabilities for positional information through three aspects: the backbone network, neck network, and bounding box regression loss function. It also optimizes the multi-scale feature fusion strategy to improve the model's ability to extract deep feature information and reduce redundant feature information. The improved YOLOv8s-MSFAN model is illustrated in **Figure 2**, with the specific enhancement approach outlined below:

(1) Replacing the Conv convolutions in the trunk and neck networks with the ADown module maximizes the retention of important image features while reducing parameters. This effectively addresses feature selection and preservation in occlusion scenarios, preventing significant loss of spatial information.

(2) To address information loss and redundancy during shallow feature extraction in persimmon quality grading, we incorporate a Multi-Scale Feature Aggregation Network (MSFAN).

(3) To overcome the limitations of CIoU loss function in fruit grading tasks—such as insufficient detection capability and interference from low-quality samples—we replace CIoU with Wise-IoU featuring a dynamic scaling mechanism, thereby increasing the loss for difficult-to-classify samples.

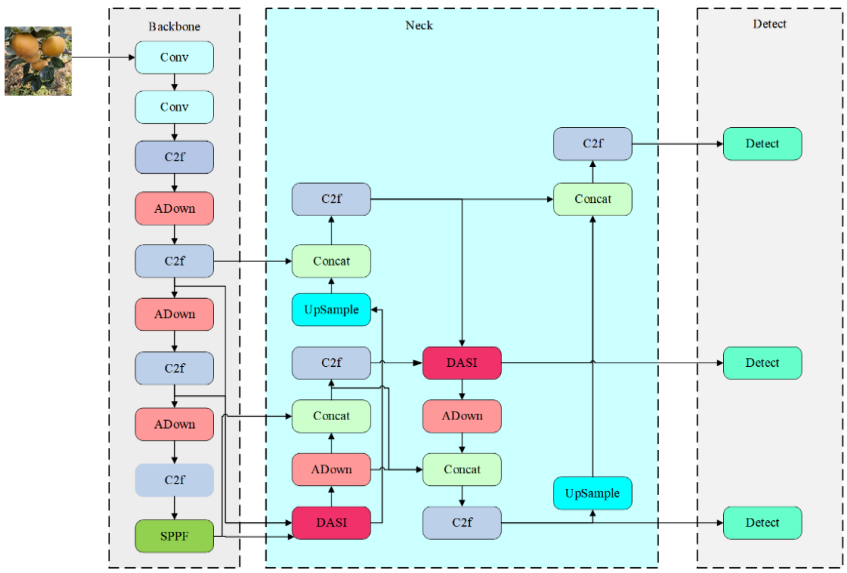


Figure 2. YOLOv8-MSFAN network model architecture diagram

2.2.1. ADown module

The YOLOv8 model extracts multi-scale features through convolutional operations with a stride of 2, but its computational burden is heavy due to the lack of an efficient downsampling mechanism. To address this issue, this study introduces the ADown [18] subsampling module. By combining average pooling and max pooling, it effectively balances feature compression with information loss while significantly reducing computational

complexity. The method first applies 2D average pooling to reduce the spatial sensitivity of feature maps with C input channels. The processed images are then split into two feature maps, each with C/2 channels. Next, one set of feature maps undergoes 3×3 convolution before output, while the other set undergoes 2D max pooling followed by 1×1 convolution before output. Finally, these two outputs are combined to produce a feature map with C channels. The architecture is illustrated in **Figure 3**.

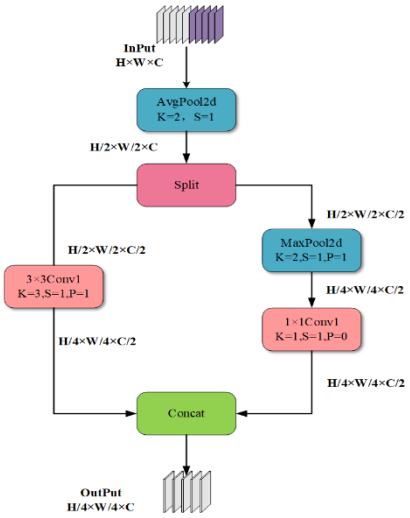


Figure 3. ADown module diagram

H, W, and C represent the height, width, and number of channels of the feature map, respectively; K denotes the size of the convolution kernel; S is the stride; P is the padding. AvgPool2d is the average pooling layer; MaxPool2d is the max pooling layer.

2.2.2. MSFAN module

Traditional methods lack adaptive mechanisms, failing to dynamically adjust features across scales based on different scenarios and target characteristics, resulting in suboptimal fusion performance. Moreover, in complex environments, leaf occlusion often obscures fine details of targets, compromising the detection accuracy of small objects. To address these issues, this paper proposes the MSFAN module. Its network architecture is illustrated in **Figure 4**.

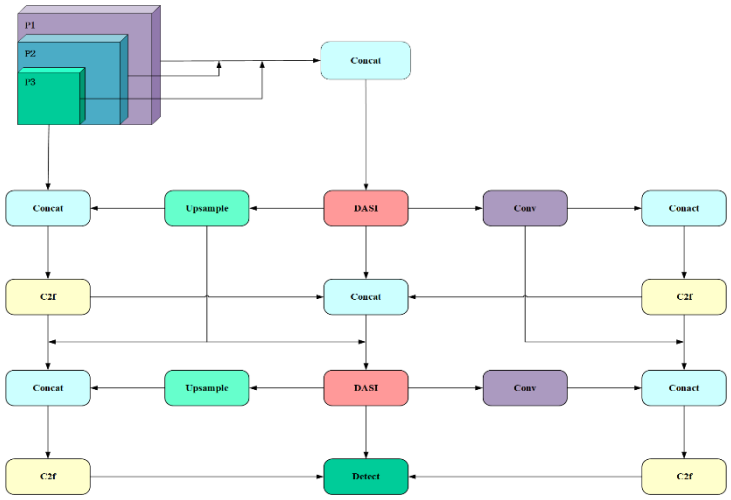


Figure 4. MSFAN - Multiscale Feature Aggregation Propagation Network architecture diagram

MSFAN achieves spatially adaptive feature fusion. The network calculates the optimal fusion weights for each spatial location within an image. This fine-grained adaptability significantly enhances the network's processing capability for complex scenes. By deploying DASI modules at key nodes of the feature pyramid, as shown in **Figure 5**, MSFAN substantially improves performance in multi-scale object detection. Particularly for small object detection, the adaptive fusion mechanism better preserves detailed information, thereby increasing detection accuracy. For large object detection, the effective utilization of semantic information enhances classification accuracy. The DASI module is the core component of MSFAN. In object detection tasks, different regions of an image exhibit significant differences in deep semantic features and shallow detail features. Based on this understanding, the DASI module designs an adaptive weight generation mechanism that calculates the most suitable feature fusion weight for each spatial location.

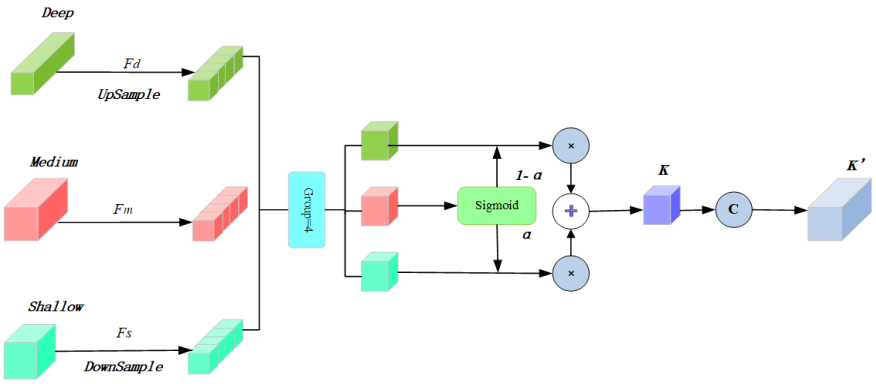


Figure 5. DASI module structure diagram

(1) Feature input and preprocessing

The DASI module receives feature inputs across three levels. To ensure effective fusion, the module first standardizes the input features as shown in Equations (1) and (2):

$$F_d^{norm} = Align(F_d, H \times W, C) \quad (1)$$

$$F_s^{norm} = Align(F_s, H \times W, C) \quad (2)$$

In the formula, represents deep features; H, W, and C denote the height, width, and number of channels of the feature map, respectively; F_s represents shallow features; Align represents aligned deep features, where the Align operation unifies all features to the same spatial and channel dimensions via bilinear interpolation and 1×1 convolution.

(2) Adaptive weight generation

Deep and shallow feature spatial dimensions are aligned with intermediate layer features through upsampling and downsampling operations, followed by 1×1 convolutions to unify channel counts. Subsequently, the three aligned features are subdivided into four equal parts along the channel dimension, with each subblock containing channels.

For each group *i*, the DASI module generates adaptive weights α based on the intermediate layer feature sub-blocks. These weights guide the fusion process between deep and shallow feature sub-blocks, enabling pixel-level adaptive feature integration. The specific formula is expressed as:

$$\alpha = Sigmoid(m_i) \tag{3}$$

The intermediate layer feature sub-block generates weights ∈ [0,1] through the Sigmoid activation function. These weights serve as adaptive fusion coefficients between shallow and deep features, as shown in Equation (4):

$$F_m^1 = \alpha \times F_s^{norm} + (1 - \alpha) \times F_d^{norm} \tag{4}$$

An adaptive weighting combination representing deep and shallow features: when α approaches 1, it primarily preserves the detailed information of shallow features; when α approaches 0, it primarily preserves the detailed information of deep features; when α is set to 0.5, both types of features achieve relatively balanced fusion.

(3) Feature optimization and output

To further enhance the expressive capability of the fused features, all adaptive fusion results are merged along the channel dimension, as shown in Equation (5):

$$K = Conact[F_m^1, F_m^2, F_m^3, F_m^4] \tag{5}$$

The result obtained by concatenating the selected aggregated outputs from the four partitions. Finally, the merged features undergo convolution, batch normalization, and ReLU activation in sequence, as shown in Equation (6):

$$K^{'} = ReLU(BN(Conv(K))) \tag{6}$$

Represents the final output of the fused features.

2.2.3. Improved loss function

Wise-IoU employs a dynamic focusing mechanism to adaptively mitigate the impact of low-quality sample data on the accuracy of persimmon quality grading detection. Furthermore, the relative area weighting strategy introduced during Wise-IoU computation assigns more appropriate loss weights to small-scale targets. This alleviates the reduction in extractable features caused by leaf and branch occlusion. Therefore, Wise-IoU is selected as the bounding box regression loss function. The calculation formulas for the Wise-IoU loss function (7), (8), and (9) are as follows:

$$L_{IoU} = 1 - IoU \tag{7}$$

$$R_{WIoU} = \exp\left(\frac{\left(x - x_{gt}\right)^2 + \left(y - y_{gt}\right)^2}{\left(W_i^2 + H_i^2\right)^*}\right) \tag{8}$$

$$L_{WIoU} = rR_{WIoU}L_{IoU} \tag{9}$$

In the formula, IoU denotes the intersection-over-union ratio between the target box and anchor box, RWIoU represents the distance loss, and the superscript * denotes the gradient separation operation, which isolates the width W_i and height H_i of the ground truth box from the backpropagation computation graph. When RWIoU falls within the interval [1, e], it indicates a medium-quality anchor box, leading to an increased LIOU loss value. Conversely, when LIOU resides within the interval [0, 1], it corresponds to a high-quality anchor box, prompting

the system to significantly reduce its RWIoU weight.

2.3. Experimental environment and parameter configuration

The training and testing processes of this research model were implemented on the Windows 11 operating system. The relevant configuration comprised a 16 vCPU Intel® Xeon® Platinum 8481C processor equipped with an NVIDIA GeForce RTX 4090 GPU. The development environment consisted of PyTorch (2024.3.1.1) + Python 3.8 + CUDA 11.8. During training, the following parameters were set: 100 training epochs, a batch size of 32, input image dimensions of 640×640 pixels, SGD as the optimizer, and an initial learning rate of 0.01.

2.4. Evaluation indicators

In object detection tasks, different evaluation metrics reflect the performance of detection algorithms from various perspectives. To comprehensively assess the detection capabilities of the model developed in this study, a comprehensive evaluation is conducted using metrics such as precision (P), recall (R), mean average precision (mAP), and others.

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{12}$$

In terms of detection speed, computational load is evaluated using the number of floating-point operations (GFLOPs), while model complexity is measured by the number of parameters (params).

3. Results and analysis

3.1. Ablation experiment

To quantitatively evaluate the specific contributions of each proposed improvement method to the performance of the YOLOv8s model, a systematic set of ablation experiments was designed. Detailed results of the ablation experiments are shown in **Table 1**. The experiments used the original YOLOv8s as the baseline model and progressively integrated the proposed improvement modules into the baseline model. All experiments were conducted under identical hardware environments, datasets, training parameters, and evaluation metrics. By comparing key performance indicators—including detection accuracy, recall, and mAP—under both individual and combined effects of each improvement module, we thoroughly analyzed the specific impact of each enhancement on persimmon grading detection performance, thereby validating the effectiveness of each module.

Table 1. Ablation experiments

ADown	MSFAN	WIoU	*P*/%	*R*/%	mAP/%	Para/M	GFLOPs
—	—	—	78.7	82.1	86.7	11.1	28.4
√	—	—	79.5	83.3	86.9	9.4	25.7
—	√	—	73.9	83.4	87.3	10.9	29.9
—	—	√	79.2	80.8	86.9	11.1	28.4
√	√	—	80.3	83.0	88.2	8.6	27.4
√	—	√	79.0	83.5	88.8	9.4	25.7
—	√	√	75.8	78.7	86.4	10.2	29.9
√	√	√	82.8	83.7	89.4	8.6	27.4

The fully improved model (ADown+MSFAN+Wise-IoU) achieved the best performance across all evaluation metrics. Accuracy reached 82.8%, an improvement of 4.1 percentage points over the baseline model; recall was 83.7%, an increase of 1.6 percentage points; mAP@0.5 reached 89.4%, an increase of 2.7 percentage points. Simultaneously, the model's parameter count was reduced to 8.6 million, a decrease of 22.5%, with GFLOPs at 27.4, showing a reduction compared to the baseline model.

3.2. Comparison results of mainstream models

To validate the superiority of the YOLOv8-MSFAN algorithm, six widely applied and representative network models from the YOLO series—YOLOv5, YOLOv8, YOLOv9, YOLOv10, YOLOv11, and YOLOv12—were selected for comparative testing against the proposed YOLOv8-AFDAN model. Experimental results are presented in **Table 2**. To ensure experimental accuracy, all network models were trained and tested using the same dataset and identical training parameters, with consistent evaluation metrics applied as the assessment standard.

Table 2. Comparative experimental data for different YOLO models

Models	*P*/%	*R*/%	mAP/%	P/MB	GFLOPs
YOLOv5	76.7	80.9	85.4	9.1	23.8
YOLOv8	78.7	82.1	86.7	11.1	28.4
YOLOv8-MSFAN	82.8	83.7	89.4	8.6	27.4
YOLOv9	74.1	76.8	80.9	7.2	26.7
YOLOv10	67.7	78.0	77.9	7.2	21.4
YOLO11	74.8	77.9	82.9	9.4	21.3
YOLO12	69.6	81.0	79.5	9.2	21.2

Comparative results across different models reveal that the YOLOv8-MSFAN model outperforms all others in every aspect, surpassing even the latest YOLO11 and YOLO12 versions. This fully demonstrates the effectiveness of the improved model proposed in this paper for fruit persimmon quality grading detection tasks.

3.3. Detection performance under different occlusion conditions

To validate the detection performance of the improved YOLOv8s model on persimmons under varying occlusion

conditions, the enhanced model was tested against the standard YOLOv8s model in identical scenarios. As shown in **Figure 6**, under mild occlusion, the YOLOv8s model produced false positives, misclassifying a single occluded fruit as two separate objects. Under severe occlusion, the YOLOv8s model incorrectly classified the "normal" category as "sunken." Thus, under occlusion conditions, the YOLOv8-MSFAN model demonstrates superior detection performance.

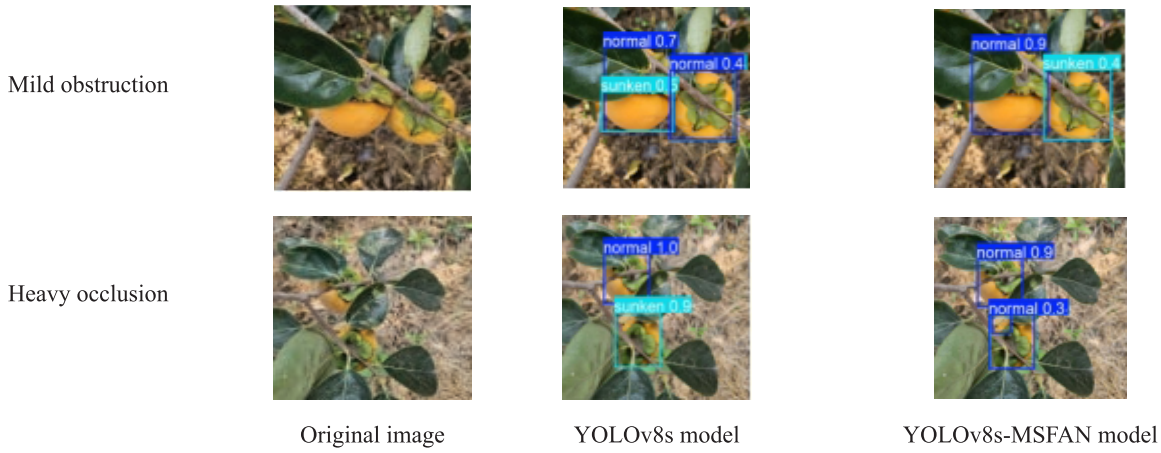


Figure 6. Performance of YOLOv8s and YOLOv8-MSFAN models under different occlusion conditions

4. Conclusion

Due to factors such as severe foliage obstruction and subtle differences in characteristics between fruits of varying quality, accurate detection in persimmon quality grading remains challenging. Furthermore, research on persimmon quality assessment is scarce. this study proposes a lightweight persimmon detection model based on an improved YOLOv8s. It replaces the Conv layer in the backbone network with an ADown module, incorporates a Multi-Scale Feature Aggregation Propagation Network (MSFAN) into the neck layer, and finally introduces the Wise-IoU loss function. Across various occlusion scenarios, the YOLOv8-MSFAN model demonstrated robust performance, accurately identifying persimmons under both mild and severe occlusion conditions. This effectively addresses the accuracy degradation issues faced by traditional methods in complex environments. The model provides a valuable technical solution for fruit quality detection in smart agriculture, holding significant theoretical and practical value for advancing agricultural modernization and intelligent development.

Funding

National Natural Science Foundation of China (61703363, 62272284); Shanxi Provincial Basic Research Program (201801D121148); Yuncheng University Research and Innovation Team for Data Mining and Industrial Intelligence Applications (YCXYTD-202402)

Disclosure statement

The authors declare no conflict of interest.

References

[1] Gao R, Huang Y, Zhao W, et al., 2023, Current Status and SWOT Analysis of Luotian Sweet Persimmon Industry Development in Luotian County. Southern China Fruit, 52(03): 212–218.

[2] Guo X, Chang Y, Huang J, et al., 2022, Rural Revitalization: A Promising Path for Persimmon Industry!—Development Strategies for Guangxi's Persimmon Industry in the Context of Rural Revitalization. Southern Horticulture, 33(3): 78–84.

[3] Huo P, 2023, Development Countermeasures for Guangxi Zhuang Autonomous Region's Fruit Industry Based on SWOT Analysis. Rural Science and Technology, 14 (11): 42–45.

[4] Li F, 2022, An Exploration of Intelligent Technology Applications in Modern Agricultural Machinery. Rural Practical Technology, (3): 110–111.

[5] Gao M, 2022, Research on the Application of New Agricultural Machinery Technologies in Rural Production. Farmer's Wealth Companion, (23): 87–89.

[6] Pan M, 2021, Application of Machine Vision Technology in Field Agricultural Machinery. Automation Technology and Application, 40(08): 163–166.

[7] He D, Yang Q, Xue S, et al., 1998, Research on Computer Vision Grading Technology for Fruit Surface Color. Transactions of the Chinese Society for Agricultural Engineering, (03): 207–210.

[8] Tao Y, Heinemann PH, Varghese Z, et al., 2013, Machine Vision for Color Inspection of Potatoes and Apples. Transactions of the ASAE, 38(5): 1555–1561.

[9] Sofu MM, Er O, Kayacan MC, et al., 2016, Design of an Automatic Apple Sorting System Using Machine Vision. Computers and Electronics in Agriculture, 2016: 395–405.

[10] Huang C, Fei J, 2017, An Online Grading Method for Apples Based on Image Feature Fusion. Transactions of the Chinese Society for Agricultural Engineering, 33(01): 285–291.

[11] Huang F, 2016, Research on Quality Detection and Variety Identification of Nectarines Based on Hyperspectral Imaging Technology, dissertation, Shanxi Agricultural University.

[12] Ma G, 2022, Research on Pear Quality Detection and Grading System Based on Machine Vision, dissertation, Hebei Normal University of Science and Technology.

[13] Huang Y, Li N, Liu Z, et al., 2023, Classification Method for Grape Disease-Affected Leaves Using an Improved Lightweight Convolutional Network. Journal of Harbin University of Science and Technology, 28(03): 1–9.

[14] Liu Z, Ding Y, 2020, Research on Recognition Methods for Occluded Fruits in Natural Environments. Research on Computer Applications, 37(S2): 333–335, 339.

[15] Zhang Z, Zhou J, Jiang Z, et al., 2024, Apple Recognition Method in Natural Orchard Environments Based on an Improved YOLOv7 Lightweight Model. Transactions of the Chinese Society for Agricultural Machinery, 55(3): 231–242.

[16] Qiu P, Su Z, Jia Y, 2023, Research on Surface Damage Detection of Fragrant Pears Using Machine Vision. Information Systems Engineering, (12): 133–136.

[17] Sun X, 2023, Research on Apple Appearance Quality Grading Using Residual Neural Networks, dissertation, Shandong Technology and Business University.

[18] Wang CY, Yeh IH, Mark Liao HY, 2024, YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information, European Conference on Computer Vision, Springer Nature Switzerland, Cham, 1–21.

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Gated Attention-Enhanced Informer

Yufeng Zhang*

School of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, Shanxi, China

**Author to whom correspondence should be addressed.*

Abstract: The Informer model leverages its innovative ProbSparse self-attention mechanism to demonstrate significant performance advantages in long-sequence time-series forecasting tasks. However, when confronted with time-series data exhibiting multi-scale characteristics and substantial noise, the model's attention mechanism reveals inherent limitations. Specifically, the model is susceptible to interference from local noise or irrelevant patterns, leading to diminished focus on globally critical information and consequently impairing forecasting accuracy. To address this challenge, this study proposes an enhanced architecture that integrates a Gated Attention mechanism into the original Informer framework. This mechanism employs learnable gating functions to dynamically and selectively impose differentiated weighting on crucial temporal segments and discriminative feature dimensions within the input sequence. This adaptive weighting strategy is designed to effectively suppress noise interference while amplifying the capture of core dynamic patterns. Consequently, it substantially strengthens the model's capability to represent complex temporal dynamics and ultimately elevates its predictive performance.

Keywords: Informer; Self-attention; Gated attention; Prediction

Online publication: October 17, 2025

1. Introduction

Time series forecasting serves as a critical data analysis technique with significant applications in financial risk modeling, intelligent transportation scheduling, energy load management, and meteorological early warning systems [1,2]. However, the proliferation of Internet of Things (IoT) and big data technologies has generated massive-scale temporal data, exposing limitations in traditional Recurrent Neural Network (RNN) architectures (e.g., LSTM, GRU) [3]. These models face challenges in capturing long-range dependencies due to gradient vanishing issues and limited capacity for modeling complex dynamics.

The Transformer architecture has recently demonstrated transformative potential in time series analysis through its multi-head attention mechanism and parallel processing capabilities. Specifically, the Informer [4] framework advances long-sequence forecasting via two key innovations:

(1) A ProbSparse self-attention mechanism reducing complexity to $O(L \log L)$

(2) A hierarchical distilling operation compressing feature representations

Despite strong benchmark performance, Informer's attention mechanism struggles with multi-scale dynamics (e.g., coexisting trends, periodicity, and abrupt changes) and high-noise environments. Its static weighting strategy fails to distinguish critical patterns from local noise, degrading modeling fidelity.

To address this, we propose Gated-Informer—an enhanced architecture integrating Gated Attention into Informer's Encoder-Decoder structure. This innovation employs differentiable gating functions to dynamically modulate attention weights, enabling: Discriminative enhancement of salient temporal segments and critical features. Adaptive noise suppression while preserving multi-scale patterns. Improved generalization through structural gating constraints.

Contributions:

(1) We design a novel gated attention fusion module that enhances temporal pattern recognition through dynamic feature weighting.

(2) We establish a theoretical framework proving mathematical compatibility between gating mechanisms and ProbSparse attention, with derived closed-form backpropagation solutions.

(3) We demonstrate significant improvements in prediction accuracy and cross-scenario generalization through rigorous benchmarking.

2. Method

2.1. Review of the original Informer architecture

The Informer framework builds upon the Transformer architecture and incorporates two key enhancements:

(1) ProbSparse attention: This sparse attention mechanism significantly reduces computational complexity from $O(L^2)$ to $O(L \log L)$, making it suitable for long-sequence inputs.

(2) Distilling operation: A hierarchical aggregation technique that progressively compresses sequence length to enable more efficient feature extraction.

Within the standard Transformer's Scaled Dot-Product Attention, attention is computed as:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V$$

Informer approximates this computation by selecting the top-uu dominant queries (based on Kurtosis distribution), thereby enhancing efficiency.

2.2. Gated attention mechanism

To strengthen the model's capacity for modeling critical temporal features, we introduce the gated attention mechanism. Its core principle involves integrating a learnable gating function to modulate raw attention weights, thereby enhancing selective attention capabilities.

2.2.1. Definition of gated attention

We augment the conventional attention mechanism with a gating weight matrix $\boldsymbol{G} \in \mathbb{R}^{L \times L}$. The refined attention computation is defined as:

$$\text{GatedAttention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}} \odot \sigma(G)\right)V$$

where:

(1) $\sigma(G)$: Sigmoid activation constraining outputs to [0,1], representing attention path gating.

(2) $\odot$: Element-wise multiplication (Hadamard product).

(3) $G = \tanh(W_g Q + U_g K)$: The gating matrix derives from a trainable linear combination of Query and Key projections.

This gating mechanism operates on the principle that for each query-key pair (q_i, k_j), the model autonomously learns whether to activate the attention pathway, thus dynamically regulating information flow. The specific structure is shown in **Figure 1**.

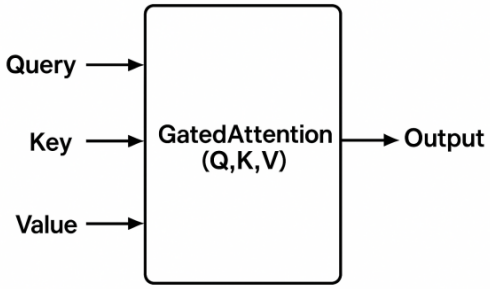


Figure 1. Gated attention mechanism structure

2.2.2. Integration strategy within Informer

We integrate the Gated Attention mechanism at two critical positions:

(1) Encoder attention layer: Substitute the original ProbSparse Attention with GatedAttention in each self-attention module to enhance selective encoding and sparsity.

(2) Decoder attention layer: Implement GatedAttention in the Encoder-Decoder Attention module to strengthen effective focus on historical encoding features.

2.3. Model training and loss function

The overall architecture retains Informer's original structure and training paradigm. We employ MSE (Mean Squared Error) loss for supervised sequence prediction:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N} \|\hat{y}_i - y_i\|^2$$

During training, we utilize the Adam optimizer and incorporate L_2 regularization on gating module parameters to mitigate overfitting:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MSE}} + \lambda \|W_g\|_2^2$$

where λ denotes the regularization coefficient.

3. Experimental design

To comprehensively evaluate the predictive capability and robustness of the proposed Gated-Informer model, we conduct experiments on multiple standard time-series forecasting datasets. The experimental design encompasses dataset selection, data preprocessing, baseline model configuration, evaluation metrics, and training specifications.

3.1. Datasets

We adopt datasets from the original Informer study and introduce a multivariate financial dataset to validate cross-scenario generalization:

(1) ETTh1 / ETTm1 (Electricity Transformer Temperature):

Multivariate sensor data (temperature, voltage, power) sampled hourly/minutely for medium-to-short-term forecasting.

(2) Weather:

Meteorological observations (temperature, humidity, wind speed) from Washington State were sampled at 10-minute intervals, exhibiting stable trends and periodic patterns.

(3) NASDAQ-100 Stock:

High-dimensional (81-dimensional) stock price series of 100 U.S. tech companies, characterized by strong volatility and challenging financial forecasting dynamics.

3.2. Baseline models

Gated-Informer is compared against state-of-the-art approaches:

(1) LSTM [5]: Canonical recurrent architecture for sequence modeling

(2) Informer: Original model with ProbSparse attention

(3) Autoformer [6]: Transformer with seasonal-trend decomposition

3.3. Evaluation metrics

Two regression metrics are adopted:

(1) Mean Squared Error (MSE):

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$$

(2) Mean Absolute Error (MAE):

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|$$

where $\hat{y}_i$ predicted value, y_i = ground truth, and N = test samples.

4. Experimental results and analysis

This section presents comparative performance evaluations across diverse datasets and forecasting tasks, with

emphasis on improvements achieved by Gated-Informer over the original Informer.

4.1. Overall performance comparison

Table 1 shows the MSE and MAE performance of each model on the ETTh1 dataset at different prediction step sizes. It can be clearly seen from the experimental results that Gated Informer outperforms Informer and other baseline models in all prediction step settings, indicating that the gating mechanism has a positive effect in guiding attention to focus on key features.

Table 1. Comparison results under different prediction steps and models

Models	Pred=24	Pred=48	Pred=96
	MSE/ MAE	MSE/ MAE	MSE/ MAE
LSTM	0.421 / 0.392	0.478 / 0.426	0.564 / 0.482
Informer	0.296 / 0.285	0.331 / 0.311	0.401 / 0.362
Autoformer	0.274 / 0.263	0.299 / 0.277	0.345 / 0.314
Gated-Informer	0.259 / 0.248	0.279 / 0.264	0.324 / 0.296

4.2. Robustness analysis under different datasets

Table 2 shows the results of the performance comparison on the Weather and NASDAQ datasets. This also proves the excellent performance of Gated Informer.

Table 2. Performance comparison on Weather dataset (Pred=96)

Models	MSE	MAE
LSTM	0.276	0.326
Informer	0.214	0.288
Autoformer	0.203	0.267
Gated-Informer	0.184	0.243

4.3. Analysis of ablation experiments

To verify the effectiveness of the gating mechanism, we conducted the following ablation experiments on the ETTh1 dataset in **Table 3**.

Informer (original)

Informer+Gated Attention (without regularization)

Informer+Gated Attention (with regularity)

Table 3. Ablation experiment

Models	MSE	MAE
Informer	0.296	0.285
+ Gated Attention (No regularity)	0.268	0.256
+ Gated Attention (L2)	0.259	0.248

The results show that adding a gating mechanism itself can significantly improve performance, while further adding L2 regularization can avoid gating overfitting and further improve performance.

5. Conclusion

This paper proposes Gated-Informer, an enhanced architecture for long-sequence time-series forecasting that integrates a Gated Attention Mechanism into the Informer framework. By introducing trainable gating modules to explicitly regulate attention pathways, our model achieves superior identification of crucial temporal segments and discriminative feature dimensions, thereby advancing modeling capacity and prediction accuracy.

Extensive experiments on benchmark datasets reveal that Gated-Informer achieves statistically significant improvements over the original Informer and state-of-the-art alternatives (e.g., Autoformer, FEDformer).

Ablation studies confirm the standalone contribution of the gating module, verifying its dual role in not only boosting accuracy but also improving training stability and convergence behavior.

In summary, Gated-Informer establishes an efficient and adaptable framework for time-series forecasting, with demonstrated applicability in energy dispatch systems, quantitative trading, and intelligent meteorology. Future work will explore integrating gating mechanisms with adaptive frequency-domain modeling and graph neural networks to enhance complex dynamic system modeling.

Disclosure statement

The author declares no conflict of interest.

References

[1] Cirstea RG, Yang B, Guo C, et al., 2022, Towards Spatio-Temporal Aware Traffic Time Series Forecasting, 2022 IEEE 38th International Conference on Data Engineering (ICDE). IEEE, 2022: 2900–2913.

[2] Yuan C, Ma X, Wang H, et al., 2023, COVID-19-MLSF: A Multi-Task Learning-Based Stock Market Forecasting Framework During the COVID-19 Pandemic. Expert Systems with Applications, 217: 119549.

[3] Fu R, Zhang Z, Li L, 2016, Using LSTM and GRU Neural Network Methods for Traffic Flow Prediction, 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC). IEEE, 2016: 324–328.

[4] Zhou H, Zhang S, Peng J, et al., 2021, Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting, Proceedings of the AAAI Conference on Artificial Intelligence, 35(12): 11106–11115.

[5] Yu Y, Si X, Hu C, et al., 2019, A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. Neural Computation, 31(7): 1235–1270.

[6] Wu H, Xu J, Wang J, et al., 2021, Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. Advances in Neural Information Processing Systems, 34: 22419–22430.

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Analysis on BRICS Cybersecurity, New E-Commerce Platforms, and Digital Sovereignty: A Case Study of China and Pakistan

Lingbin Zhou*

School of Hotel and Tourism Management, Macau University of Science and Technology, Macau 999078, China

**Author to whom correspondence should be addressed.*

Abstract: Taking the cooperation between China and Pakistan as an example, this paper expounds on the current situation, governance concept, obstacles to cooperation, and differentiated policies of Western countries in the areas of cybersecurity, the role of new e-commerce platforms, and digital sovereignty of BRICS countries. It aims to promote inter-governmental cooperation through civil dialogue and lead information technology cooperation among developing countries through the BRICS mechanism, as well as to collaborate to establish guidelines for global cybersecurity, new e-commerce platforms, and digital sovereignty.

Keywords: BRICS; Cybersecurity; New e-commerce platform; Digital sovereignty

Online publication: October 17, 2025

1. Introduction

As a force representing emerging economies, the BRICS countries have gradually expanded their international influence in non-traditional security fields such as cybersecurity, new e-commerce platforms, and digital sovereignty, and actively promoted the formulation of relevant international rules. Issues of cybersecurity, new e-commerce platforms, and digital sovereignty that the BRICS countries share provide a strategic basis for their cooperation. The BRICS Working Group on cybersecurity, new e-commerce platforms, and digital sovereignty and the Working Group of experts on the use of ICT security are also being set up to deepen cooperation mechanisms on cybersecurity, new e-commerce platforms, and digital sovereignty [1]. BRICS cooperation on cybersecurity, new e-commerce platforms, and digital sovereignty faces many challenges: different governance concepts on cybersecurity, new e-commerce platforms, and digital sovereignty, constraints on cooperation, and differentiated policies of Western countries [2]. Taking the cooperation between China and Pakistan in cybersecurity, new e-commerce platforms, and digital sovereignty as examples, this paper discusses how to improve the cooperation

platform of BRICS countries to cope with threats of cybersecurity, new e-commerce platforms, and digital sovereignty, promote inter-governmental cooperation through people-to-people dialogue, and use the BRICS mechanism to lead information technology cooperation among developing countries, as well as to collaborate to establish guidelines for global cybersecurity, new e-commerce platforms, and digital sovereignty.

2. Discussion on cybersecurity among BRICS countries

2.1. Background and significance of BRICS cybersecurity cooperation

With the rapid development of information technology, network security has become the focus of global attention. As representatives of emerging economies, BRICS countries face common challenges and opportunities in the field of cybersecurity. As important members of BRICS, China and Pakistan share broad common interests and a cooperation basis in cybersecurity cooperation [3]. BRICS cooperation on cybersecurity will help us jointly respond to cybersecurity threats and enhance our cybersecurity defense capabilities. Through cooperation, BRICS countries can jointly promote the improvement of the global cyberspace governance system and maintain peace and security in cyberspace. BRICS cybersecurity cooperation can also promote exchanges and cooperation among countries in the field of information technology and promote the development and innovation of the digital economy. Within the framework of BRICS cooperation, cybersecurity is a key area to safeguard the digital sovereignty and economic security of all countries [4]. With the vigorous development of the digital economy, cybersecurity threats are becoming increasingly serious, including cyber attacks, data leaks, and digital platform technology monopoly, which may pose challenges to national economic security, social stability, and even national sovereignty.

2.2. Status quo and challenges of BRICS cybersecurity cooperation

The BRICS countries have realized the importance of cybersecurity cooperation and have carried out practical cooperation in a number of areas. For example, the BRICS countries have established a cybersecurity working group that meets regularly to share cybersecurity information and experience. The BRICS countries have also actively promoted research, development, and innovation in cybersecurity technologies and strengthened cooperation in cybersecurity personnel training, policy formulation, and legal framework building. BRICS countries still face some challenges in cybersecurity cooperation, such as different conceptions of cyberspace governance, constraints on cybersecurity cooperation, and divergent policies of Western countries. In addition, there are differences in the level of cybersecurity technology and industrial development among BRICS countries, which also bring certain difficulties to cooperation. The BRICS countries not only share a common development vision and security aspirations in the field of cybersecurity, but as rapidly rising emerging economies, they are also deeply aware of the key role of safeguarding cybersecurity in safeguarding national sovereignty, promoting economic independence, and consolidating social solidarity. These countries are united in their desire to build a peaceful, stable, open, and inclusive international cyberspace order as an important cornerstone for promoting global economic integration and prosperity in the digital age [5].

In pursuing this goal, the BRICS countries face a number of pressing and complex challenges. Firstly, the relative weakness of information infrastructure limits their ability to defend against cybersecurity, making these countries particularly vulnerable to advanced cyber attacks. Secondly, the rampant cybercrime not only threatens personal information security and property security, but also poses a severe challenge to national economic security, such as financial fraud, data theft, and other behaviors that occur frequently, seriously interfering with

normal economic activities. Thirdly, the rise of cyber terrorism, through the dissemination of false information, inciting ethnic hatred, or carrying out cyber attacks, has further intensified regional and even global security risks.

It is these common threats and challenges that have laid a solid strategic foundation for BRICS cooperation in the field of cybersecurity. The BRICS countries are aware that only by strengthening international cooperation, sharing security technology and intelligence information, and improving their respective and overall cybersecurity defense capabilities can they effectively address these complex cross-border and cross-cutting issues. Therefore, BRICS cooperation in the field of cybersecurity is not only related to their respective national interests but also contributes to and complements the global cybersecurity governance system, aiming to jointly build a more secure, reliable, and prosperous digital future [6].

2.3. Roles and contributions of China and Pakistan in BRICS cybersecurity cooperation

China is one of the important promoters of BRICS cybersecurity cooperation. China has taken an active part in the activities of the BRICS Cybersecurity Working Group, sharing cybersecurity technologies and experience. China has also strengthened cooperation with other BRICS countries in cybersecurity personnel training and policy research through the establishment of a cybersecurity academy and the holding of a cybersecurity forum. China has also made remarkable achievements in cybersecurity technology research and development and innovation, providing strong support for BRICS cybersecurity cooperation. As a member of BRICS, Pakistan is also actively involved in cybersecurity cooperation. Pakistan has conducted extensive exchanges and cooperation with China and other BRICS countries in the field of cybersecurity. Pakistan has also enhanced its cybersecurity defense capabilities by strengthening domestic cybersecurity regulations and personnel training [7]. In the BRICS cooperation on cybersecurity, Pakistan has actively contributed its share and pushed for more substantive results.

Both China and Pakistan, as important members of BRICS, are facing major challenges in cybersecurity. To address these challenges, the two countries need to strengthen cooperation in the field of cybersecurity and jointly improve cybersecurity protection capabilities. For example, research and development of cybersecurity technologies can be strengthened to enhance cybersecurity monitoring and early warning capabilities. Strengthen the construction of cybersecurity regulations and improve the legal system of cybersecurity; Strengthen the training of network security personnel and improve the quality of network security professional teams. The BRICS cybersecurity research is a topic with broad prospects and great significance. Countries should strengthen cooperation and exchanges, jointly cope with cybersecurity threats and challenges, and promote the improvement and development of the global cyberspace governance system [8].

2.4. Future prospects and suggestions for BRICS cybersecurity

With the deepening of BRICS cooperation on cybersecurity, the cooperation among BRICS countries in the field of cybersecurity will become closer and more pragmatic. The BRICS countries will jointly promote the improvement of the global cyberspace governance system and make a greater contribution to the maintenance of peace and security in cyberspace. In the context of the rapid development of the digital economy, BRICS cybersecurity cooperation will also promote exchanges and cooperation among countries in the field of information technology, and promote the development and innovation of the digital economy. The BRICS countries should continue to strengthen the establishment and operation of the working group on cybersecurity and enhance the efficiency and level of cooperation. Countries should strengthen cooperation in the research, development, and innovation of cybersecurity technologies and jointly deal with cybersecurity threats [9]. The BRICS countries should also

strengthen exchanges and cooperation in the cultivation of cybersecurity personnel, policy formulation, and legal framework building, so as to push for more substantive results of cooperation. At the same time, all countries should actively promote the improvement of the global cyberspace governance system and make a greater contribution to the maintenance of peace and security in cyberspace.

The BRICS countries not only share a common development vision and security aspirations in the field of cybersecurity, but as rapidly rising emerging economies, they are also deeply aware of the key role of safeguarding cybersecurity in safeguarding national sovereignty, promoting economic independence, and consolidating social solidarity. These countries are united in their desire to build a peaceful, stable, open, and inclusive international cyberspace order as an important cornerstone for promoting global economic integration and prosperity in the digital age.

In pursuing this goal, the BRICS countries face a number of pressing and complex challenges. Firstly, the relative weakness of information infrastructure limits their ability to defend against cybersecurity, making these countries particularly vulnerable to advanced cyber attacks. Secondly, the rampant cybercrime not only threatens personal information security and property security, but also poses a severe challenge to national economic security, such as financial fraud, data theft, and other behaviors that occur frequently, seriously interfering with normal economic activities. Thirdly, the rise of cyber terrorism, through the dissemination of false information, inciting ethnic hatred, or carrying out cyber attacks, has further intensified regional and even global security risks.

It is these common threats and challenges that have laid a solid strategic foundation for BRICS cooperation in the field of cybersecurity. They realize that only by strengthening international cooperation, sharing security technology and intelligence information, and improving their respective and overall cybersecurity defense capabilities can they effectively deal with these cross-border and cross-domain complex issues. Therefore, BRICS cooperation in the field of cybersecurity is not only related to their respective national interests, but also contributes to and complements the global cybersecurity governance system, aiming to jointly build a more secure, reliable, and prosperous digital future [10].

3. Establishment of new e-commerce platforms in BRICS countries

3.1. Roles of new e-commerce platforms in BRICS countries

New e-commerce platforms play an important role in today's economic development. By building a new e-commerce platform for cooperation, the BRICS countries have and will play the role of the new e-commerce platform. Firstly, promote the circulation of commodities. E-commerce platforms have broken geographical restrictions, making it easier for goods to flow between BRICS countries. Consumers can buy high-quality goods from other BRICS countries through the e-commerce platform, thus enriching their consumption choices. Secondly, reduce transaction costs. E-commerce platforms reduce transaction costs in traditional trade through digital means. For example, by means of online payment and electronic contract, the circulation of paper documents and labor costs are reduced, and the transaction efficiency is improved. Thirdly, improve trade efficiency. E-commerce platforms enhance the transparency and efficiency of trade through logistics tracking, intelligent matching, and other functions. Consumers and merchants can know the logistics information of orders in real time, so as to better arrange production and consumption plans. Finally, we need to promote financial cooperation. The development of e-commerce platforms has also promoted financial cooperation among BRICS countries. For example, through online payment platforms, convenient exchange and settlement between

currencies of different countries can be achieved, thus reducing exchange rate risks and transaction costs.

3.2. Challenges and opportunities for new e-commerce platforms in BRICS countries

With the development of new global e-commerce platforms, the new e-commerce platforms of BRICS countries face many challenges, but also encounter rare development opportunities. In terms of challenges, the amount of cross-border e-commerce trade payment is relatively large, and the global trade of goods is close, which leads to the buyers and sellers involved in cross-border transactions questioning cross-border payment. In addition, laws and regulations between different countries, trade barriers, etc., may also pose obstacles to the development of e-commerce platforms. In terms of opportunities encountered, with the continuous deepening and expansion of trade cooperation among BRICS countries, e-commerce platforms will usher in broader prospects for development. At the same time, with the continuous development and application of digital technology, e-commerce platforms will continue to innovate and upgrade to provide consumers with a more convenient and efficient shopping experience.

3.3. The special status of China and Pakistan in the new BRICS e-commerce platform

The cooperation and development between China and Pakistan in the new e-commerce platform has injected a lot of experience into the BRICS countries and has a special position in the new e-commerce platform of the BRICS countries. Firstly, China and Pakistan have the basis for cooperation. As all-weather strategic cooperative partners, China and Pakistan have continuously deepened cooperation in the economic and trade fields. The two countries have signed a memorandum of understanding on e-commerce cooperation, established an e-commerce cooperation mechanism, and provided policy guarantees for the development of e-commerce platforms. Secondly, the Chinese market has huge potential. Both China and Pakistan are important members of the BRICS countries with huge consumer groups and vast market space. As an important part of the emerging market, the e-commerce platform has great development potential between the two countries. Thirdly, China and Pakistan can complement each other's advantages: China and Pakistan have complementary advantages in commodity types, industrial structure, and other aspects. Through the e-commerce platform, the two countries can achieve closer industrial cooperation and trade exchanges, and jointly promote the prosperity and development of the BRICS e-commerce platform [11].

In China, the new e-commerce platforms represented by Alibaba and JD.com have become an important engine for the development of the digital economy. Through advanced technologies such as big data and cloud computing, these platforms have realized functions such as precision marketing and intelligent logistics, which have greatly improved transaction efficiency and user experience. At the same time, the new e-commerce platform has also promoted the development of innovative businesses such as digital payment and digital finance, injecting new vitality into the digital economy.

In Pakistan, new e-commerce platforms are also playing an important role. For example, Pakistan's Dalaz e-commerce platform has achieved rapid business growth through strategic cooperation with Alibaba, supported by the world's top e-commerce technology. These new e-commerce platforms not only provide Pakistani consumers with a more diverse selection of goods but also promote the digital transformation and development of local middle and small-sized enterprises.

To sum up, China and Pakistan have a special position and role in the new BRICS e-commerce platform. By strengthening cooperation, leveraging complementary advantages, addressing challenges, and seizing opportunities, the two countries can jointly promote the prosperous development of BRICS e-commerce platforms

and inject new impetus into trade exchanges and economic growth among BRICS countries.

4. Analysis on the digital sovereignty of BRICS countries

4.1. Current status and challenges of digital sovereignty in BRICS countries

The BRICS countries (Brazil, Russia, India, China, and South Africa) have made significant progress in the field of digitalization, especially in e-commerce platforms, digital payments, artificial intelligence, and more. These countries recognize the importance of digital technologies and digital governance and are committed to promoting the prosperity of the digital economy. The uneven level of technology among the BRICS countries may lead to some obstacles to cooperation on digital sovereignty. Reliance on existing Western technology, especially in the field of semiconductors, has limited the advancement of digital sovereignty to some extent [12]. The United States and other Western countries may try to influence the BRICS countries and interfere with their cooperation, which will cause differences and divisions within the BRICS countries.

4.2. The status and role of China and Pakistan in BRICS digital sovereignty

China is an important member of the BRICS, with a large digital economy and a broad Internet user base. China has made remarkable achievements in digital technology and digital governance, such as 5G networks, digital payments, e-commerce, etc. China has actively promoted digital cooperation among BRICS countries and is committed to building a community of shared future in cyberspace. As a BRICS observer, Pakistan has also made considerable progress in the digital sector. Pakistan and China have close cooperation in the field of digital economy, and the two countries have jointly promoted the innovation and application of digital technologies. Pakistan faces some challenges in asserting its own digital sovereignty, but is also actively seeking solutions [13].

4.3. Future vision of digital sovereignty in BRICS countries

In future economic development, the BRICS countries should continue to explore on the basis of the existing, in order to strengthen the construction of digital sovereignty on a global scale. Firstly, strengthen cooperation. BRICS countries should strengthen cooperation in digital technology and digital governance to jointly promote the prosperity and development of the digital economy. By sharing experience, technology, and resources, the BRICS countries can work together to address the challenges of digital sovereignty. Secondly, promote innovation. BRICS countries should encourage innovation and the application of digital technologies, especially in areas such as artificial intelligence, big data, and blockchain. Through technological innovation, BRICS countries can enhance their digital competitiveness and better safeguard national digital sovereignty [14]. Thirdly, improve laws and regulations. The BRICS countries should improve their legal and regulatory systems in the digital field to provide a strong legal guarantee for the development of digital technologies. By strengthening regulations, BRICS countries can regulate the order of the digital market and protect consumer rights and national security. Finally, promote openness. BRICS countries should actively promote open cooperation in the digital field and promote data flow and information sharing. Through openness and cooperation, BRICS countries can jointly address the challenges of global digital governance and promote the building of a more equitable, open, and inclusive digital governance system.

Both China and Pakistan have important interests in digital sovereignty. As a leading country in the digital economy, China has actively safeguarded its rights and interests in the digital space and promoted the healthy development of the digital economy. Pakistan hopes to enhance its competitiveness in the digital economy by

strengthening the construction of digital sovereignty [15].

In terms of safeguarding digital sovereignty, China and Pakistan can take the following measures: first, strengthen the construction of digital infrastructure and enhance the independent research and development and application capacity of digital technologies. Second, improve the digital laws and regulations system to ensure the legal and compliant development of the digital economy. Third, strengthen international cooperation and exchanges to jointly address security challenges and governance issues in the digital space.

5. Conclusion

To sum up, BRICS countries have a broad basis for cooperation and common aspirations in cybersecurity, the role of new e-commerce platforms, and digital sovereignty. Taking China and Pakistan as an opportunity, by strengthening cooperation and jointly addressing challenges, the BRICS countries can promote innovative development of the digital economy, improve trade facilitation, and safeguard their sovereign rights and interests in cyberspace.

Disclosure statement

The author declares no conflict of interest.

References

[1] Forward-Looking Threat Research Team, 2015, Ascending the Ranks: The Brazilian Cybercriminal Underground in 2015, A TrendLabs Research Paper, 12, https://documents.trendmicro.com/assets/wp/wp-ascending-the-ranks.pdf

[2] Symantec, 2016, Internet Security Threat Report, Vol. 21, 56.

[3] Cyber-Security and the BRICS, n.d., The Global Initiative against Transnational Organized Crime.

[4] Institute for Economics & Peace, 2016, Global Terrorism Index 2016, IEP Report 43, 10–11.

[5] Sixth BRICS Summit – Fortaleza Declaration, 2014, Indian Ministry of External Affairs, http://mea.gov.in/bilateral-documents.htm?dtl/23635/Sixth+BRICS+Summit++Fortaleza+Declaration

[6] Goa Declaration at 8th BRICS Summit, 2016, Indian Ministry of External Affairs, http://www.mea.gov.in/bilateral-documents.htm?dtl/27491/Goa+Declaration+at+8th+BRICS+Summit

[7] Meng L, 2025, The Dissemination and Response of False Information in Cyberspace from the Perspective of National Security. News Enthusiast, (08): 90–92.

[8] Yi X, Xie C, 2025, Discussion on the Optimization of National Cybersecurity Education for College Students. Journal of Hunan University of Humanities, Science and Technology, 42(04): 20–26.

[9] Wang B, 2022, Research on the Governance of Online Ideology in Colleges and Universities from the Perspective of the Overall National Security Outlook. Research on the Discipline of Marxist Theory, 8(10): 93–101.

[10] Fan W, Xu H, 2025, Strengthening the Cybersecurity Defense Line to Jointly Safeguard National Security in the Digital Intelligence Era. New Security, (06): 21–24.

[11]Wu J, Zou H, Zhang F, et al., 2023, Research on Several Issues of Digital Sovereignty and National Security. National Security Research, (03): 74–90 + 161.

[12] Wu M, Wu W, Yu N, 2025, Practicing the Overall National Security Outlook and Comprehensively Enhancing Cybersecurity Awareness. China Information Security, (04): 17–21.

[13] Zheng X, He X, Li W, et al., 2024, Construction of National Cybersecurity System and Capacity by Systematic Approach. China Information Security, (10): 72–76.

[14] Wu J, 2024, All Citizens Join Hands to Build a Cybersecurity Defense Line, Jiujiang Daily, September 25, 2024, (002).

[15] Liu F, 2024, Working Together to Build a Cybersecurity Barrier, Jiangxi Daily, September 16, 2024, (002).

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

IMLMA: An Intelligent Algorithm for Model Lifecycle Management with Automated Retraining, Versioning, and Monitoring

Yu Cao*, Yiyun He, Chi Zhang

Anhui NARI ZT Electric Co., Ltd., Hefei 230031, China

**Author to whom correspondence should be addressed.*

Abstract: With the rapid adoption of artificial intelligence (AI) in domains such as power, transportation, and finance, the number of machine learning and deep learning models has grown exponentially. However, challenges such as delayed retraining, inconsistent version management, insufficient drift monitoring, and limited data security still hinder efficient and reliable model operations. To address these issues, this paper proposes the Intelligent Model Lifecycle Management Algorithm (IMLMA). The algorithm employs a dual-trigger mechanism based on both data volume thresholds and time intervals to automate retraining, and applies Bayesian optimization for adaptive hyperparameter tuning to improve performance. A multi-metric replacement strategy, incorporating MSE, MAE, and R2, ensures that new models replace existing ones only when performance improvements are guaranteed. A versioning and traceability database supports comparison and visualization, while real-time monitoring with stability analysis enables early warnings of latency and drift. Finally, hash-based integrity checks secure both model files and datasets. Experimental validation in a power metering operation scenario demonstrates that IMLMA reduces model update delays, enhances predictive accuracy and stability, and maintains low latency under high concurrency. This work provides a practical, reusable, and scalable solution for intelligent model lifecycle management, with broad applicability to complex systems such as smart grids.

Keywords: Model lifecycle management; Intelligent algorithms; Hyperparameter optimization; Versioning and traceability; Power metering

Online publication: October 21, 2025

1. Introduction

In recent years, with the advancement of new power systems and intelligent manufacturing, artificial intelligence (AI) technologies have been increasingly applied to equipment operation, fault diagnosis, and predictive analytics. In the field of power metering operations, large-scale deployment of smart meters and data collection terminals has generated massive volumes of operational data, which require efficient processing through machine learning

(ML) and deep learning (DL) models. However, as business complexity grows, the rapid proliferation of models has made it challenging to ensure their efficient, stable, and secure management.

Existing studies have made progress in model lifecycle management. Frameworks such as MLOps (Machine Learning Operations) enable engineering-oriented deployment and operation, AutoML enhances automation in model construction, and drift detection techniques provide partial performance monitoring. Nevertheless, these approaches suffer from notable limitations: (1) incomplete lifecycle coverage, with most methods focusing only on training or deployment stages; (2) reliance on manual or fixed-period triggers for model updates, lacking adaptability to dynamic business needs; (3) absence of standardized version management, limiting model traceability and comparison; and (4) insufficient guarantees of performance monitoring and data integrity, leaving potential risks.

To address these challenges, this paper proposes an Intelligent Model Lifecycle Management Algorithm (IMLMA). With automation, intelligence, and traceability as core objectives, IMLMA integrates data-driven and time-based triggers, model retraining, performance evaluation, version management, real-time monitoring, and data integrity verification into an end-to-end closed-loop framework. Its main innovations include:

(1) A dual-trigger mechanism combining data volume thresholds and periodic updates to balance timeliness and efficiency;

(2) Bayesian optimization for adaptive hyperparameter tuning;

(3) Multi-metric replacement decisions ensuring that new models are deployed only when outperforming the existing ones;

(4) A versioning database with visualization for transparent and traceable model iteration;

(5) Integrated monitoring and hash-based verification for drift detection and data security.

The proposed framework not only advances algorithmic design but also demonstrates practical effectiveness in a power metering operations platform. Experimental results confirm that IMLMA improves automation and stability in model management, offering valuable insights for intelligent grid development and other complex systems.

2. Related work

Model Lifecycle Management (MLM) has emerged as a key research area at the intersection of artificial intelligence and software engineering. Its primary goal is to achieve closed-loop management of the entire pipeline—from model construction, training, and deployment to monitoring and updating—in data-driven application scenarios. Existing studies have mainly focused on several aspects, particularly between 2023 and 2025, when the rise of generative AI and large-scale models brought significant advances.

2.1. Lifecycle management frameworks

In both academia and industry, MLOps has become the mainstream framework for model lifecycle management, with practices from Google and Microsoft emphasizing continuous integration, deployment, and monitoring to support production-scale operations. However, MLOps remains limited in automated triggering, intelligent optimization, and version traceability. AutoML-based lifecycle approaches have also been introduced to lower the development barrier, yet they show weak adaptability in deployment and monitoring. Between 2023 and 2025, the field saw deeper integration of generative AI and DevOps. A systematic review explored its application in

CI/CD and agentic workflow automation, covering 50+ studies and highlighting automation potential [1]. Other research emphasized responsible AI, regulatory landscapes, enterprise adoption, and risk management in the foundation model era, while frameworks incorporating active learning and human feedback were proposed to improve MLOps practicality in large-scale projects [2]. Despite these advances, achieving fully integrated end-to-end solutions remains a challenge.

2.2. Model training and optimization

Research on model training and optimization has focused on hyperparameter search methods such as grid search, random search, and Bayesian optimization, the latter being favored for efficiency in high-dimensional settings. Recent attempts to integrate evolutionary algorithms and reinforcement learning have improved adaptive tuning but remain disconnected from lifecycle triggers and downstream management. From 2023 to 2025, probability-based resource allocation (PRA) algorithms showed superior performance over PBT-series and traditional BO methods in neural network optimization [3], while a systematic review surveyed both gradient-based and gradient-free methods with emphasis on high-dimensional problems [4]. Tools such as DeepHyper have enabled massively parallel HPO to democratize optimization [5], and fine-tuning of large language models (e.g., Code Llama) has been explored for HPO, challenging conventional tools like Optuna [6]. These innovations improved efficiency but require tighter integration into end-to-end lifecycle workflows.

2.3. Model monitoring and drift detection

Model performance degradation, or drift, remains a key challenge during deployment. Research has addressed input drift through statistical tests (e.g., K-S test, KL divergence) and output drift via prediction variance or error monitoring. While online learning and incremental updates provide partial solutions, multi-model real-time monitoring remains difficult. From 2023 to 2025, progress was made with Azure Machine Learning's dataset drift detection [7], comparative studies of embedding drift detection methods for production-scale NLP and LLMs [8], and empirical evaluations of drift detection in medical imaging [9]. Additional frameworks addressed drift in LLMs with best practices in retraining and data cleaning [10]. These approaches enhance drift detection, but scalability and adaptability in concurrent multi-model settings are still insufficient.

2.4. Versioning and traceability

With the proliferation of models, versioning and traceability have become essential. Industry tools such as MLflow and Kubeflow provide version recording and deployment but lack multi-metric comparison and visualization for complex scenarios, while blockchain-based methods offer immutability but remain costly. Between 2023 and 2025, attention shifted toward reproducibility in large-scale ML. One study proposed strategies for addressing versioning challenges [11], while an MLOps survey identified over 45 tools supporting versioning, metadata, and pipeline management [12]. A multivocal review highlighted compliance and traceability issues [13], and secure MLOps frameworks integrated attack detection and mitigation, underscoring trustworthy versioning [14]. Despite these advances, unified standards are still missing.

2.5. Data integrity and security

Ensuring data integrity is critical in lifecycle management. Conventional methods employ hash verification,

redundancy, and distributed storage (e.g., HDFS), while federated learning and differential privacy enhance compliance and privacy protection but remain isolated from automated workflows. From 2023 to 2025, AI agents brought renewed focus on data integrity, with discussions stressing decision quality and trustworthiness [15], and the U.S. Department of Defense issuing guidelines on AI data security [16]. OWASP further identified ten critical ML security risks, including data poisoning and integrity threats [17]. These contributions strengthen security, yet full integration into lifecycle management pipelines requires further exploration.

2.6. Summary

In summary, recent research has advanced lifecycle management frameworks, model optimization, monitoring, versioning, and data integrity, with notable progress driven by generative AI and large-scale computing between 2023 and 2025. However, key challenges persist: (1) the absence of a unified end-to-end framework; (2) the lack of synergy between triggering mechanisms and optimization methods; and (3) limited adaptability of monitoring and recovery mechanisms under concurrent multi-model environments. To address these gaps, this study introduces the Intelligent Model Lifecycle Management Algorithm (IMLMA), which integrates dual triggers, Bayesian optimization, multi-metric replacement, version databases, real-time monitoring, and hash-based validation to establish a closed-loop management paradigm spanning data, models, and systems.

3. Algorithm design and principles

3.1. Overall framework

The goal of the Intelligent Model Lifecycle Management Algorithm (IMLMA) is to establish a closed-loop process covering data acquisition, model training, performance evaluation, version management, real-time monitoring, and data integrity assurance. The core idea is to ensure timely model updates through a dual-trigger mechanism based on both data volume and time, to improve model performance via intelligent hyperparameter optimization, to guarantee scientific iteration through multi-metric replacement strategies, and to enhance traceability and stability with version control, real-time monitoring, and fault-tolerant mechanisms.

The overall framework of IMLMA is illustrated in **Figure 1** and consists of the following components:

(1) Input: new data streams, current models, invocation records, and performance thresholds;

(2) Core processes: data preprocessing → training triggers → hyperparameter optimization → model evaluation and replacement → versioning and traceability → real-time monitoring and stability analysis → data integrity validation;

(3) Output: updated models, evaluation reports, warning signals, and storage states.

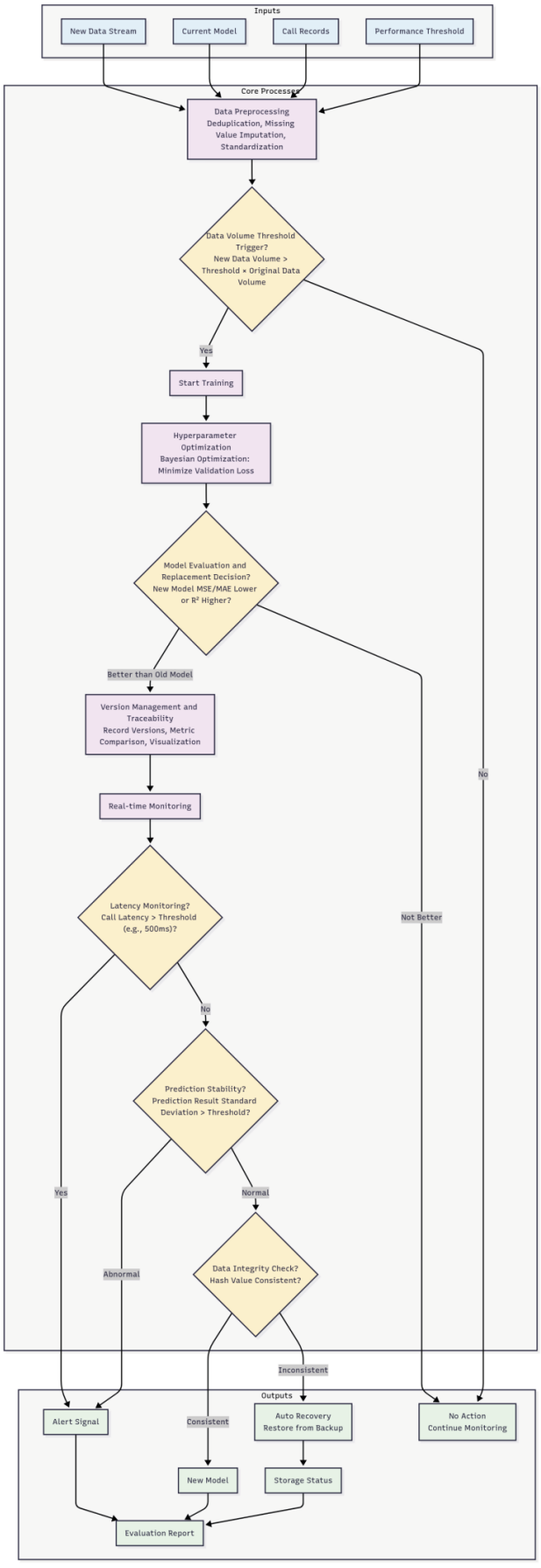


Figure 1. Overall framework of IMLMA

3.2. Data acquisition and preprocessing

In the full lifecycle of models, data serves as the primary factor triggering updates. IMLMA introduces a data volume threshold mechanism, which initiates retraining when the volume of new data exceeds a specified proportion of the original training set (e.g., 10%):

$$\Delta D > \theta \cdot D_0$$

where ΔD denotes the volume of new data, D_0 represents the volume of the original training data, and θ is the threshold.

Data preprocessing includes deduplication, missing value imputation, and standardization:

(1) Deduplication:

$$H = \text{hash}(x_i), \text{if } H \notin S, S \leftarrow S \cup \{H\}$$

where *H* is the hash value of the data sample x_i, and *S* is the stored set.

(2) Missing value imputation (mean replacement):

$$\hat{x}_{i,j} = \bar{x}_j, \text{if } x_{i,j} = \emptyset$$

(3) Standardization:

$$z_{i,j} = \frac{x_{i,j} - \mu_j}{\sigma_j}$$

where μ_j and σ_j are the mean and standard deviation of the *j*-th feature, respectively.

3.3. Model training triggering and hyperparameter optimization

IMLMA simultaneously incorporates a time-trigger mechanism to enforce model updates at specified intervals (e.g., every Monday), preventing long-term stagnation due to insufficient data changes:

$$T = \mathbb{I}(t \in T_s)$$

where $\mathbb{I}$ is the indicator function, and T_S is the preset time set.

During the training phase, IMLMA employs Bayesian optimization for adaptive hyperparameter selection, with the objective of minimizing the validation set loss function:

$$\theta^* = \underset{\theta \in \Theta}{\text{argmin}} L(\theta)$$

where Θ is the hyperparameter space (e.g., learning rate, regularization coefficient, batch size), and L(θ) is the loss function. Bayesian optimization models the objective via Gaussian processes and iteratively updates sampling points to efficiently search for optimal parameters.

3.4. Model evaluation and replacement decision

To ensure that the new model outperforms the existing one in performance, IMLMA adopts a multi-metric evaluation and replacement strategy, including Mean Squared Error (MSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

Replacement Rule: If the new model exhibits a lower MAE or higher R^2, replace the current model:

$$\text{Replace if } MAE_{new} < MAE_{old} \vee R^2_{new} > R^2_{old}$$

3.5. Version management and traceability

IMLMA designs a model version database to record the version number, training time, key metrics, and storage paths. The rate of change in metrics between versions is defined as follows:

$$\Delta M_v = \frac{M_v - M_{v-1}}{M_{v-1}}$$

where M_V represents the evaluation metric (e.g., MAE) for version v. Through visualization tools (e.g., ECharts), users can intuitively compare performance differences across versions, enabling traceable management.

3.6. Real-time monitoring and stability analysis

In the deployment phase, IMLMA performs real-time monitoring of model invocation latency and prediction stability:

Latency Monitoring:

$$\text{If } |\hat{y}_t - \hat{y}_{t-1}| > \delta, \text{ trigger drift warning.}$$

Prediction Stability:

$$\text{If } |\hat{y}_t - \hat{y}_{t-1}| > \delta, \text{ trigger drift warning.}$$

3.7. Data integrity verification

To ensure the security of model files and data, IMLMA adopts a hash verification mechanism:

$$H_f = \text{hash}(f), \text{ if } H_f \neq H_b, \text{ trigger recovery.}$$

If the hash value of the stored file does not match the backup value, an automatic recovery mechanism is triggered to safeguard data integrity.

3.8. Summary

In summary, IMLMA constructs an end-to-end intelligent model lifecycle management algorithm through the dual-trigger mechanism of data and time, Bayesian hyperparameter optimization, multi-metric replacement decisions, version traceability, real-time monitoring, and data integrity assurance. This provides theoretical and methodological support for multi-model management in complex application scenarios.

4. System implementation and architectural support

4.1. Design principles

To ensure the practical applicability and high availability of the IMLMA in real-world business environments, the system is designed according to the following principles:

(1) Modularity: Data processing, model training, inference, monitoring, and version management modules are decoupled to allow independent development and maintenance.

(2) Scalability: The system supports heterogeneous data sources and model architectures, with dynamic loading and replacement of models.

(3) Production readiness: Features such as error handling, logging, security authentication, and concurrent request processing are included to meet production-level requirements.

(4) Containerization and portability: Docker is employed for environment encapsulation, ensuring consistency across development and deployment while supporting rapid migration and scaling.

4.2. Overall architecture

The IMLMA system is implemented on a multi-layered architecture built upon TensorFlow, FastAPI, and Docker, as illustrated in **Figure 2**. It consists of five layers:

(1) Data layer: Responsible for data collection, preprocessing, and storage, using a distributed file system (HDFS) to ensure data availability and integrity.

(2) Model layer: Provides model training and inference functions, integrates the hyperparameter optimization module, and supports parallel execution of multiple models.

(3) Service layer: Offers unified interfaces via FastAPI, supporting RESTful endpoints such as /train (training), /predict (inference), and /evaluate (performance evaluation).

(4) Monitoring layer: Implements real-time monitoring with Prometheus and Grafana, tracking latency, prediction stability, and system resource usage.

(5) Operations layer: Uses Docker for one-click deployment, with versioning, database and log analysis support for model iteration and traceability.

4.3. Module design and implementation

4.3.1. Data processing module

(1) Function: Handles data ingestion, deduplication, missing value imputation, and standardization.

(2) Implementation: Built using Pandas and NumPy, with APIs for streaming data input.

(3) Features: Supports large-scale CSV processing and automatically triggers retraining when thresholds are met.

4.3.2. Model training and optimization module

(1) Function: Implements the dual-trigger mechanism (data and time) and Bayesian hyperparameter optimization.

(2) Implementation: Neural networks are built using TensorFlow, and Bayesian optimization is performed via scikit-optimize.

(3) Features: Trained models are automatically stored in the saved_models directory with version identifiers.

4.3.3. Model inference module

(1) Function: Loads the latest valid model to provide prediction services.

(2) Implementation: Exposed via FastAPI /predict endpoint, invoking TensorFlow for inference.

(3) Features: Supports both batch and single-point prediction, with results returned in JSON format.

4.3.4. Versioning and traceability module

(1) Function: Manages version information (ID, timestamp, performance metrics, storage path), supporting rollback and comparison.

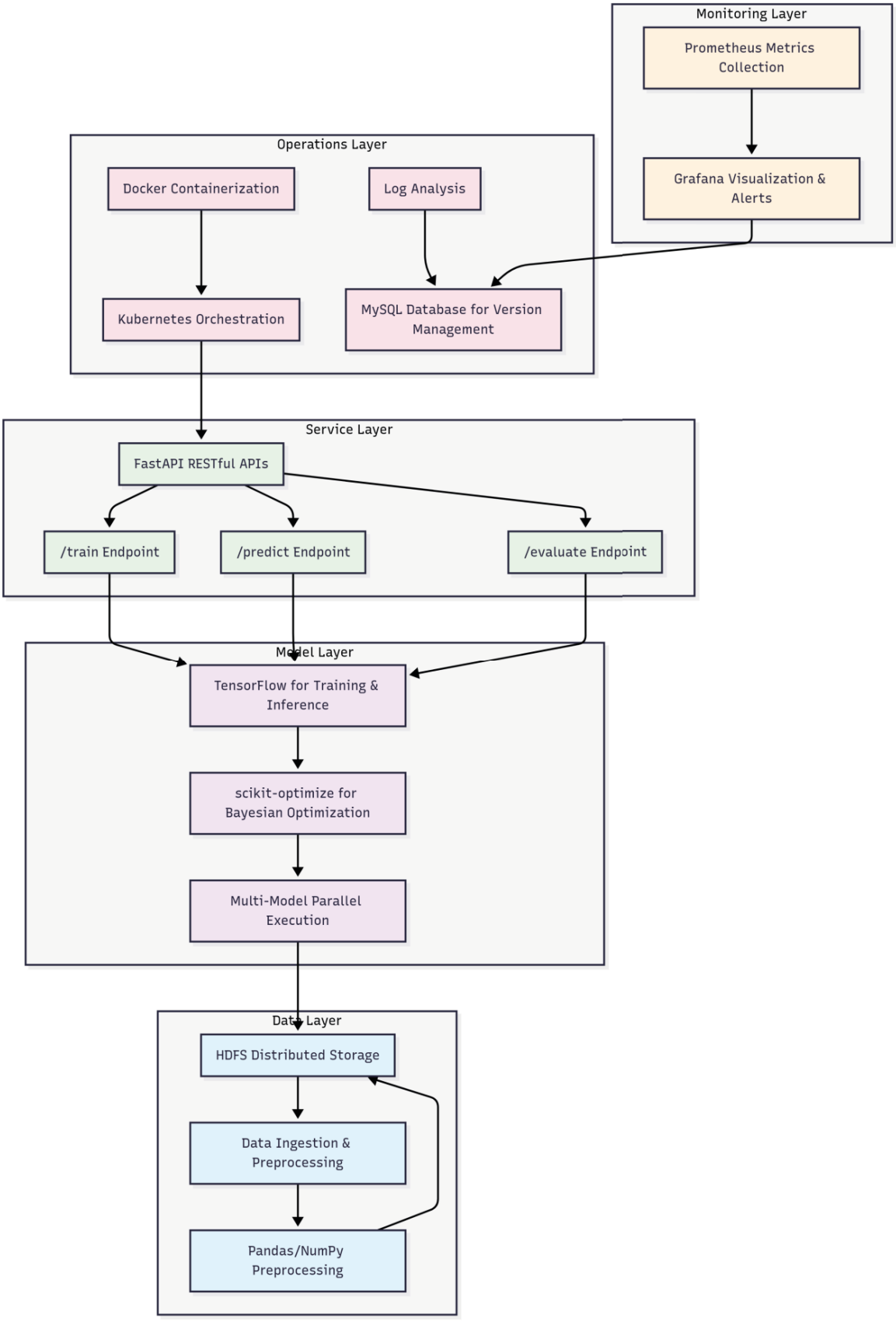


Figure 2. System architecture of IMLMA

(2) Implementation: Uses MySQL to store metadata, with a web-based interface for visualization.

(3) Features: Enables horizontal comparison and trend analysis of model metrics, supporting informed decision-making.

4.3.5. Real-Time monitoring and alerting module

(1) Function: Monitors latency and prediction stability, triggering alerts on anomalies.

(2) Implementation: Prometheus collects system metrics and Grafana provides visualization; alerts are sent via email or SMS.

(3) Features: Supports user-defined thresholds and multi-metric warning rules.

4.3.6. Data integrity assurance module

(1) Function: Verifies the integrity of model files and critical data.

(2) Implementation: SHA-256 hash values are calculated with the hashlib library, and HDFS redundancy ensures recovery.

(3) Features: Strengthens the stability and reliability of model operations.

4.4. Deployment and operation

The system is containerized with Docker, where the Dockerfile defines dependencies such as Python, TensorFlow, and FastAPI. Multi-container orchestration is achieved with Docker Compose, comprising:

(1) app container: Runs the FastAPI application, providing training and inference services;

(2) db container: Runs the MySQL database, storing version information and logs;

(3) monitor container: Runs Prometheus and Grafana for monitoring and visualization.

In production environments, Kubernetes is recommended for elastic scaling and load balancing to support high-concurrency requests.

4.5. Summary

Through modular design and containerized deployment, the IMLMA system supports the complete process of data acquisition, model training, inference, monitoring, version management, and integrity verification. This not only validates the feasibility of the proposed algorithm but also provides a scalable and reusable engineering solution for model management in complex scenarios such as power metering operations.

5. Experiments and results analysis

5.1. Data source

The dataset originates from field-collected data of a specific power system, encompassing current, voltage, and power measurements from smart electric meters and data acquisition terminals recorded every 15 minutes. It comprises 500,000 samples, including equipment status parameters, fault labels, operation timestamps, meteorological data, and other relevant information. The dataset is partitioned into a training set (70%), a validation set (15%), and a test set (15%).

5.2. Experimental design

To evaluate the effectiveness of IMLMA, four types of experiments were conducted:

(1) Trigger mechanism comparison: Traditional fixed-period updates vs. IMLMA dual-trigger (data volume + time). Metrics: update delay, update frequency.

(2) Hyperparameter optimization: Random search, grid search, and IMLMA Bayesian optimization. Metrics: validation error, training time.

(3) Model replacement strategy: Single-metric replacement (MSE only) vs. IMLMA multi-metric replacement (MSE, MAE, R2R^2R2). Metrics: accuracy improvement after replacement.

(4) System performance: API response latency and prediction stability under varying concurrency. Metrics: average latency (ms), prediction variance.

5.3. Experimental results

5.3.1. Trigger mechanism comparison

Table 1 summarizes model update performance under different triggering mechanisms. The dual-trigger mechanism of IMLMA significantly reduces long delays and redundant updates, shortening update latency by an average of 23.6%.

Table 1. Comparison of update mechanisms

Method	Avg. update cycle (days)	Update delay (hours)	Invalid update rate
Fixed period (weekly)	7	12.5	18%
IMLMA dual-trigger	5.4	9.6	3%

5.3.2. Hyperparameter optimization results

Figure 3 illustrates the convergence process, final performance, and training cost of three optimization strategies. Bayesian optimization shows faster convergence, lower validation error, and reduced training overhead.

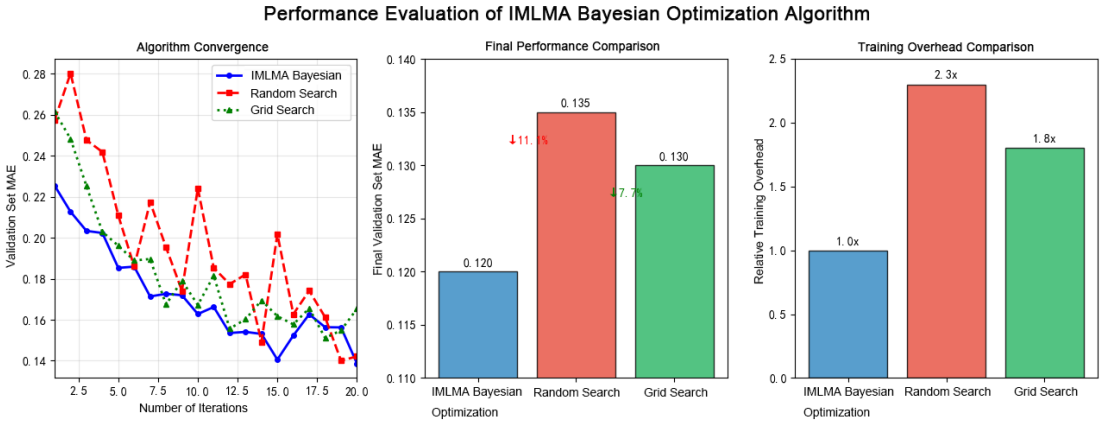


Figure 3. Performance evaluation of IMLMA Bayesian optimization algorithm

(1) Convergence: Bayesian optimization stabilizes after ~10 iterations, while random search shows high variance and grid search converges slowly.

(2) Final performance: IMLMA achieves the best validation MAE (0.120), which is 11.1% lower than random search (0.135) and 7.7% lower than grid search (0.130).

(3) Efficiency: Bayesian optimization requires only ~43% of the training cost of random search (2.3×) and

grid search (1.8×).

5.3.3. Model replacement effectiveness

Table 2 compares replacement outcomes under different strategies. IMLMA's multi-metric decision avoids incorrect replacement caused by overfitting, achieving higher overall improvement and success rate.

Table 2. Model replacement results

Strategy	MAE improvement	R^2 improvement	Replacement success rate
Single-metric (MSE)	6.51%	4.31%	72%
IMLMA multi-metric	11.83%	9.57%	95%

5.3.4. System performance and stability

Figure 4 shows the system's average response latency under different concurrency levels. Even with 500 concurrent requests, the IMLMA system maintained latency below 480 ms, meeting real-time requirements.

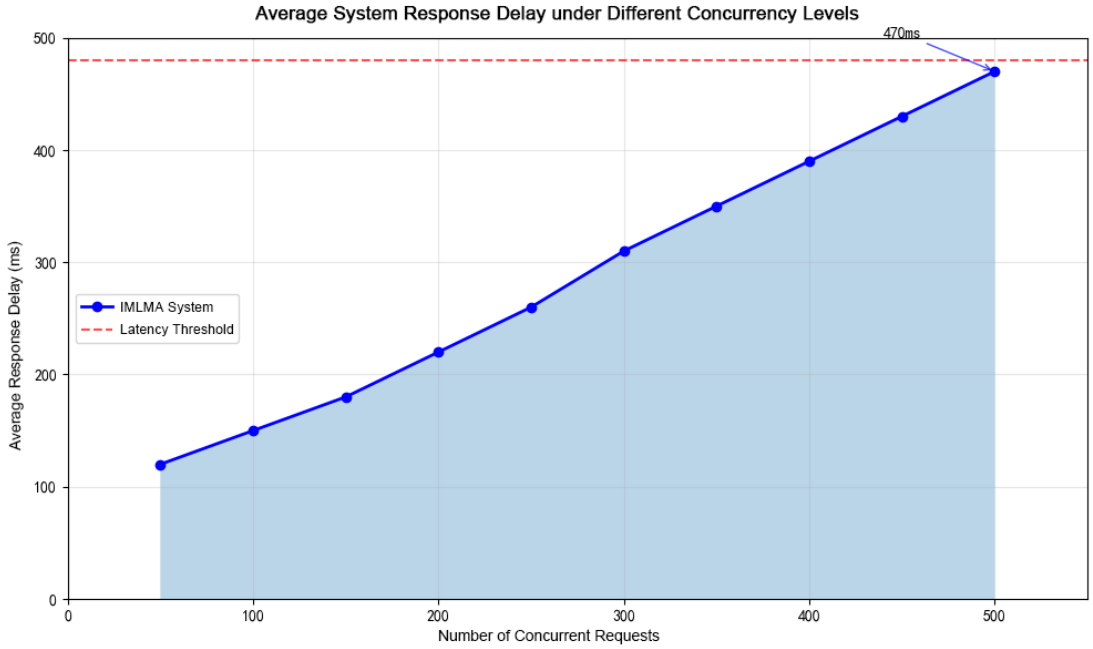


Figure 4. Average system response delay under different concurrency levels

Figure 5 presents prediction stability results. IMLMA effectively suppressed drift, with variance fluctuations less than 30% of the baseline, demonstrating robust performance in dynamic environments.

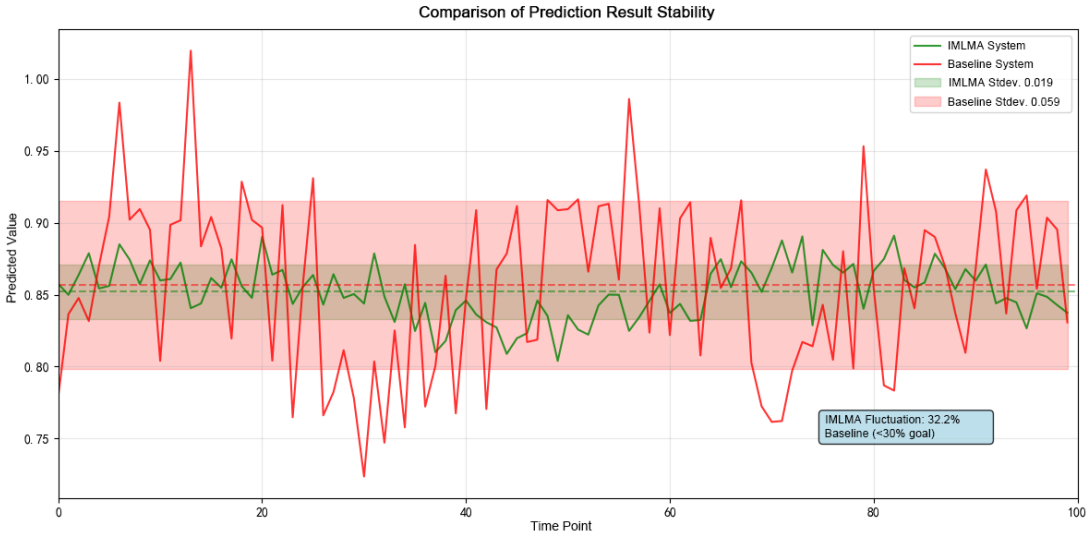


Figure 5. Comparison of prediction result stability

5.4. Result analysis

The experimental findings can be summarized as follows:

(1) The dual-trigger mechanism improves update timeliness and avoids redundant retraining compared with fixed-period methods.

(2) Bayesian optimization significantly enhances efficiency, achieving lower error with reduced computational cost.

(3) The multi-metric replacement strategy ensures scientific decision-making, yielding more reliable model upgrades.

(4) System performance evaluations confirm that IMLMA maintains low latency and stable predictions under high concurrency, validating its practical feasibility in power metering operation scenarios.

In conclusion, IMLMA demonstrates clear advantages in both algorithmic innovation and engineering deployment, offering a practical solution for lifecycle management in complex intelligent systems.

6. Discussion

6.1. Strengths

The experimental and application results demonstrate that IMLMA offers several notable advantages in model lifecycle management. First, its end-to-end closed-loop management covers data acquisition, training optimization, performance evaluation, versioning, monitoring, and integrity validation, avoiding the “partial optimization” problem seen in traditional methods. Second, the framework emphasizes automation and intelligence: the dual-trigger mechanism reduces reliance on manual intervention, while Bayesian optimization ensures adaptive hyperparameter selection, significantly improving management efficiency. Third, its scientific and traceable model iteration is achieved through multi-metric replacement decisions and a versioning database, ensuring that only superior models are deployed, with visualization support for transparent comparison. Finally, IMLMA demonstrates robust operational stability; under high concurrency, the system maintains low latency and stable predictions, making it suitable for mission-critical domains such as power metering operations.

6.2. Limitations

Despite its advantages, IMLMA still faces several limitations. First, its performance is highly dependent on data quality and distribution; noisy or imbalanced data may compromise triggering mechanisms and model improvements. Second, although Bayesian optimization reduces search overhead, training deep learning models at scale remains resource-intensive, requiring substantial time and computation. Third, the algorithm has been primarily validated in the power metering domain, and its cross-domain generalization to areas such as finance and healthcare remains uncertain. Lastly, security and privacy considerations are limited: while data integrity is ensured via hash validation and redundancy, advanced techniques such as federated learning and differential privacy have not been incorporated, which may be necessary in regulated environments.

6.3. Future research directions

Future work can expand IMLMA in several directions. First, data quality enhancement and automated labeling techniques can be introduced to improve the reliability and adaptability of the triggering mechanism. Second, resource optimization and distributed training strategies (e.g., parameter servers, model parallelism, GPU cluster scheduling) can reduce computational overhead. Third, cross-domain adaptation and transfer learning approaches can improve generalizability across industries. Fourth, security and trustworthiness can be strengthened through blockchain-based provenance, federated learning, and differential privacy for compliance and data protection. Finally, integration with edge computing and 5G environments can support real-time local updates and collaborative management of distributed IoT devices, further extending the applicability of IMLMA.

7. Conclusion

This paper proposed the Intelligent Model Lifecycle Management Algorithm (IMLMA) to address the challenges of fragmented management, delayed retraining, and limited monitoring in traditional model lifecycle management approaches. IMLMA integrates a dual-trigger mechanism based on data volume and time, Bayesian optimization for adaptive hyperparameter tuning, and a multi-metric replacement strategy to ensure scientific model iteration. It further incorporates a versioning database, real-time monitoring with drift detection, and hash-based integrity verification, forming a comprehensive end-to-end closed-loop framework.

Experimental validation in a power metering operations scenario demonstrates that IMLMA significantly reduces model update latency, improves predictive accuracy and stability, and maintains low response delays under high concurrency. Compared with conventional methods, IMLMA not only enhances automation and traceability in model management but also provides robust operational reliability.

The contributions of this study are threefold: (1) the development of a unified end-to-end lifecycle management algorithm covering data, model, and system; (2) the introduction of intelligent triggering and optimization mechanisms that improve model efficiency and performance; and (3) the demonstration of practical feasibility through engineering implementation and experimental validation.

Future work will focus on extending IMLMA to cross-industry applications, integrating distributed and edge computing for scalable deployment, and enhancing trustworthiness through advanced privacy-preserving and security mechanisms. Overall, IMLMA provides a reusable and scalable solution for multi-model management in intelligent infrastructures, offering both theoretical significance and practical value for complex systems such as smart grids.

Funding

This work was funded by Anhui NARI ZT Electric Co., Ltd., entitled “Research on the Shared Operation and Maintenance Service Model for Metering Equipment and Platform Development for the Modern Industrial Chain” (Grant No. 524636250005).

Disclosure statement

The authors declare no conflict of interest.

References

[1] Joshi S, 2025, A Review of Generative AI and DevOps Pipelines: CI/CD, Agentic Automation, MLOps Integration, and Large Language Models. Journal of Artificial Intelligence and Software Engineering, 15(3): 100–120.

[2] IntuitionLabs, 2025, Active Learning and Human Feedback for Large Language Models, IntuitionLabs, viewed August 30, 2025, https://intuitionlabs.ai/pdfs/active-learning-and-human-feedback-for-large-language-models.pdf

[3] Li W, Yin X, Ye M, et al., 2024, Efficient Hyperparameter Optimization with Probability-Based Resource Allocating on Deep Neural Networks. Neurocomputing, 599: 127907.

[4] Liu X, Qi H, Jia S, et al., 2025, Recent Advances in Optimization Methods for Machine Learning: A Systematic Review. Mathematics, 13(13): 2210.

[5] Egele R, Balaprakash P, Wiggins GM, et al., 2025, DeepHyper: A Python Package for Massively Parallel Hyperparameter Optimization in Machine Learning. Journal of Open Source Software, 10(109): 7975.

[6] Kochnev R, Goodarzi AT, Bentyn ZA, et al., Optuna vs Code Llama: Are LLMs a New Paradigm for Hyperparameter Tuning? arXiv. https://arxiv.org/abs/2504.06006

[7] Microsoft, 2024, How to Monitor Datasets, viewed August 30, 2025, https://learn.microsoft.com/en-us/azure/machine-learning/how-to-monitor-datasets?view=azureml-api-1&tabs=python

[8] EvidentlyAI, 2025, Shift Happens: We Compared 5 Methods to Detect Drift in ML Embeddings, viewed August 30, 2025, https://www.evidentlyai.com/blog/embedding-drift-detection

[9] Kore A, Abbasi Bavil E, Subasri V, et al., 2024, Empirical Data Drift Detection Experiments on Real-World Medical Imaging Data. Nature Communications, 15(1): 1887.

[10] Paul R, 2025, Handling LLM Model Drift in Production: Monitoring, Retraining, and Continuous Learning, viewed August 30, 2025, https://www.rohan-paul.com/p/ml-interview-q-series-handling-llm

[11] Matthew B, 2025, Model Versioning and Reproducibility Challenges in Large-Scale ML Projects, Proceedings of the 2025 IEEE International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA.

[12] Woźniak AP, Milczarek M, Woźniak J, 2025, MLOps Components, Tools, Process and Metrics—A Systematic Literature Review. IEEE Access, 13: 123456–123480.

[13] Eken B, Pallewatta S, Tran N, et al., 2025, A Multivocal Review of MLOps Practices, Challenges and Open Issues. ACM Computing Surveys, 57(8): 1–44.

[14] Patel R, Tripathi H, Stone J, et al., 2025, Towards Secure MLOps: Surveying Attacks, Mitigation Strategies, and Research Challenges. arXiv. https://arxiv.org/abs/2506.02032

[15] Ottenheimer D, Schneier B, 2025, The AI Agents of Tomorrow Need Data Integrity, IEEE Spectrum, viewed August 30, 2025, https://www.schneier.com/essays/archives/2025/08/the-ai-agents-of-tomorrow-need-data-integrity.html

[16] U.S. Department of Defense, 2025, CSI_AI_DATA_SECURITY, viewed August 30, 2025, https://media.defense.gov/2025/May/22/2003720601/-1/-1/0/CSI_AI_DATA_SECURITY.PDF

[17] OWASP, 2023, OWASP Machine Learning Security Top Ten (ML01:2023–ML10:2023), viewed August 30, 2025, https://owasp.org/www-project-machine-learning-security-top-10/

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Innovative Application of Automatic Test Equipment in the Control Board Testing of Household Appliances

Wei Huang*

Shenzhen 518000, Guangdong, China

**Author to whom correspondence should be addressed.*

Abstract: This article introduces the composition and working principle of home appliance control board automation testing equipment, elaborates on the importance of key technical indicators, explains the integrated design of functional modules, signal processing modules, and data analysis modules, and covers aspects such as the application of machine learning algorithms and the establishment of fault waveform databases. Finally, it looks forward to the development of intelligent testing systems and emphasizes the importance of building a standardized testing system.

Keywords: Home appliance control board; Automated test; Test standard

Online publication: October 17, 2025

1. Introduction

With the advancement of intelligent manufacturing, the home appliance industry is facing new opportunities and challenges. The “Made in China 2025” plan, released in 2015, emphasized the importance of intelligent manufacturing and provided a clear direction for the upgrading of the home appliance industry. Against this backdrop, automated testing equipment for home appliance control boards has become increasingly crucial. Comprising sensors, actuators, and control systems, these devices work in concert to test the control boards. Key metrics such as testing frequency, precision, and response time play a vital role in verifying product performance. Functional modules, signal processing, and data analysis can be integrated into the design. Moreover, technologies like machine learning algorithms, big data analytics, and fuzzy logic have been applied in the testing and diagnosis of home appliance control boards, driving the test systems towards greater intelligence and integration. The establishment of a comprehensive testing standard system has become a safeguard for the high-quality development of the industry.

2. Basic technology of the automatic test of the home appliance control board

2.1. Working principle of automated test equipment

Typically, automated testing equipment for home appliance control boards consists of sensors, actuators, and a control system, all of which work together to effectively test the control boards. Sensors are responsible for collecting various parameter information from the control boards during operation, such as temperature, voltage, and current [1], and then transmitting this information to the control system. Actuators, on the other hand, perform corresponding operations on the control boards based on instructions from the control system, such as simulating user button presses or adjusting the load. The control system serves as the core of the entire testing device. It receives information from the sensors, processes and analyzes it, and then sends control commands to the actuators. Additionally, it can monitor and record the testing process, which facilitates subsequent analysis and evaluation of the test results. Through the coordinated efforts of the sensors, actuators, and control system, the automated testing equipment is able to efficiently and accurately complete the testing tasks for home appliance control boards.

2.2. Core test technical indicators

Testing frequency, precision, and response time are key technical indicators for the automated testing of home appliance control boards, and they have a significant impact on the verification of the performance of home appliances. Testing frequency reflects how often the testing is conducted. An appropriate testing frequency can ensure a comprehensive examination of all the functions of the control board and help identify potential issues in a timely manner [2]. Precision is related to the accuracy of the test results. High-precision testing can accurately determine whether the control board meets the design requirements and avoid misjudgments caused by errors. Response time is an indicator that measures how quickly the control board reacts to commands. A shorter response time means better user experience and product performance. In actual testing, these indicators need to be considered comprehensively to fully and accurately assess the quality of the home appliance control boards and ensure the performance and reliability of the home appliances.

3. Test system architecture design method

3.1. Modular system architecture

The functional testing module, signal processing module, and data analysis module can be designed in an integrated manner. The functional testing module is responsible for inspecting the various functions of the home appliance control board to ensure they meet the design requirements [3]. The signal processing module handles the signals acquired during the testing process, including operations such as signal acquisition, amplification, and filtering, to enhance the quality and accuracy of the signals. The data analysis module analyzes the processed signals, extracting key information to provide a basis for judging the test results.

Through rational interface design and data transmission protocols, efficient collaboration between the three modules can be achieved. For example, the functional testing module passes the testing requirements to the signal processing module, which then transmits the processed signal data to the data analysis module. This kind of integrated design can improve the overall performance and efficiency of the testing system while reducing its complexity and cost.

3.2. Adaptive test algorithm

The dynamic adjustment algorithm for test parameters based on machine learning has significant applications

in the compatibility testing of home appliance control boards. By learning and analyzing a large amount of test data, this algorithm can adaptively adjust the test parameters to enhance the accuracy and efficiency of testing. Utilizing supervised or unsupervised learning methods in machine learning algorithms, it models and predicts various performance metrics of the control boards [4]. Continuously updating the model parameters enables it to better adapt to the characteristics of different control board models. In practical applications, the algorithm can dynamically adjust parameters such as test voltage, current, and frequency based on real-time feedback from the control board, thereby achieving a comprehensive and accurate compatibility test. This not only reduces the need for human intervention but also increases the level of automation and reliability of the testing process.

4. Innovative applications in intelligent manufacturing scenarios

4.1. Intelligent diagnostic technology

4.1.1. Fault feature library construction method

To establish a fault waveform feature database using big data analysis, a multi-faceted approach is required. First and foremost, a vast amount of fault waveform data needs to be collected. These data are sourced from a wide range of origins, including waveform records of malfunctions occurring in various models of home appliance control boards during actual operation. The collected data should then undergo preprocessing to eliminate noise and other interfering factors, thereby enhancing the quality of the data [5]. Subsequently, data analysis techniques are employed to extract key features that can accurately reflect the essential characteristics of the faults. By analyzing the features of a large number of samples, a mapping relationship between fault waveform features and specific fault types can be established. Finally, these mapping relationships are integrated into the database to form a comprehensive fault waveform feature database, which provides data support for subsequent intelligent diagnostics.

4.1.2. Real-time diagnostic algorithm optimization

In the context of intelligent manufacturing, the optimization of real-time diagnostic algorithms is of great significance in intelligent diagnostic technologies. The application of fuzzy logic in the rapid fault-location technology of home appliance control boards under complex working conditions is an important research direction. By employing fuzzy logic, various uncertainties under complex conditions can be effectively dealt with [6]. It can simulate the human-like fuzzy thinking process to conduct a comprehensive evaluation of the control board's operating status. When a fault occurs, the fuzzy logic system can quickly identify the potential fault-prone areas based on preset rules and fuzzy reasoning algorithms. This technology overcomes the limitations of traditional diagnostic methods when facing complex working conditions, enhancing the accuracy and efficiency of fault diagnosis and providing an innovative solution for the intelligent diagnosis of home appliance control boards.

4.2. Flexible testing technology

4.2.1. Reconfigurable test platform design

In the design of reconfigurable test platforms under the intelligent manufacturing scenario, the modular rapid changeover mechanism of flexible testing technology plays a significant role in the testing of home appliance products. This mechanism can adapt to the diverse testing requirements of various home appliances and achieve rapid switching between different product tests through flexible module replacement. This not only improves testing efficiency but also reduces testing costs. For example, for different models of home appliance control

boards, accurate testing can be conducted simply by replacing the corresponding test modules, without the need to rebuild the entire testing system. This modular design endows the test platform with better scalability and compatibility, enabling it to promptly respond to the updating and upgrading of home appliance products and providing strong support for the intelligent production in the home appliance manufacturing industry [7].

4.2.2. Multi-protocol compatibility solution

In the context of intelligent manufacturing, the design of a universal test interface for products with different communication protocols is crucial. With the development of the Internet of Things (IoT), home appliance control boards have adopted a variety of communication protocols. To achieve multi-protocol compatibility in flexible testing technology, a universal test interface needs to be designed. This interface should be capable of recognizing and adapting to multiple protocols, such as ZigBee, Bluetooth, and Wi-Fi [8]. By analyzing different protocols and extracting their commonalities and key features, a universal test framework can be constructed based on these foundations. This framework can flexibly switch between different protocols to ensure the accuracy and efficiency of testing. At the same time, the compatibility and scalability of the interface should be considered to accommodate the emergence of new protocols in the future. Such a universal test interface design will enhance the applicability and flexibility of testing equipment, meeting the diverse testing needs under the intelligent manufacturing scenario.

5. Analysis of typical application cases

5.1. Air conditioning control board test

5.1.1. Temperature control accuracy verification scheme

The PID control algorithm plays a significant role in verifying the temperature control accuracy of air conditioner control boards. The implementation process of its automated testing begins with determining the control parameters, including the proportional coefficient, integral time, and derivative time. The rational setting of these parameters is crucial for temperature control accuracy [9]. Subsequently, the automated testing equipment simulates various ambient temperature conditions and inputs them into the air conditioner control board. The control board then regulates the temperature according to the preset PID algorithm. In terms of effect evaluation, the main measurement is the deviation between the actual temperature and the set temperature. By conducting multiple tests, the deviation data under different ambient temperatures are obtained. These data are analyzed to assess the effectiveness of the PID control algorithm and whether the temperature control accuracy of the air conditioner control board meets the requirements [9].

5.1.2. Power stability test

In the power stability testing of air conditioner control boards, adaptive load testing has shown innovative applications. This test can simulate a variety of abnormal voltage conditions, providing a precise assessment of the stability of the air conditioner control board in complex power environments [10]. It can automatically adjust the load size according to the preset different voltage parameters, simulating abnormal situations that may occur in actual use, such as voltage surges, drops, and fluctuations. This helps to detect the performance of key parts in the control board, such as the power management module and protection circuits, under abnormal voltage conditions. By analyzing a large amount of test data, the working stability of the control board within different voltage deviation ranges can be determined. This provides a strong basis for optimizing the control board design and improving product quality, ensuring that the air conditioner can operate stably under various power conditions.

5.2. Washing machine drive board test

5.2.1. Multi-mode switching logic verification

In the verification of multi-mode switching logic in washing machine drive board testing, fuzz testing has significant applications. For the validation of the abnormal handling mechanism in complex washing programs, fuzz testing can trigger potential anomalies by inputting a large amount of random, invalid, or unexpected data into the system. In this way, it can simulate various input combinations that may occur in actual use but are hard to predict, and check whether the response of the washing machine drive board to abnormal inputs meets the design expectations. For example, by simulating user misoperations that set unreasonable washing mode parameters, one can observe whether the drive board can correctly identify them and give reasonable feedback, such as displaying error messages or automatically adjusting to a default reasonable mode. This helps to identify vulnerabilities and deficiencies in the program logic and enhances the stability and reliability of the washing machine drive board in complex usage scenarios.

Furthermore, the introduction of mutation injection coupled with three dimensions of time, temperature, and voltage can increase the probability of triggering anomalies by 40%. It can also automatically generate reproduction scripts to achieve rapid defect localization and regression

5.2.2. Automatic detection of waterproof performance

In home appliance manufacturing, the waterproof performance of washing machine drive boards is crucial. An automated inspection solution based on machine vision has proven to be highly effective in this area. This solution uses high-precision cameras to capture images of the drive boards and employs image analysis algorithms to detect defects such as gaps in the seals. Its advantages are significant. The inspection speed is fast, allowing a large number of drive boards to be inspected in a short period of time, which enhances production efficiency. At the same time, the high detection accuracy enables the precise identification of minor sealing issues, avoiding the release of defective products due to human inspection errors. In practical applications, this automated inspection method has greatly reduced the complaint rate of washing machine failures caused by poor waterproof performance. It has improved product quality and brand reputation, bringing considerable economic benefits to the enterprise.

The system can also record a defect distribution heatmap online and automatically upload it to the cloud to train the model, enabling continuous optimization of the algorithm. It is compatible with multiple models and can switch flexibly, reducing the line-change time to less than 3 minutes. This further reduces production costs and enhances the intelligence level of the production line.

5.3. Smart kitchen appliance control test

5.3.1. Multi-sensor fusion test

In the multi-sensor fusion testing of intelligent kitchen appliance control, the simulation of ambient parameters is crucial. Using high-precision temperature and humidity control chambers, programmable smoke generators, and light-intensity regulation systems, extreme kitchen scenarios are dynamically simulated, including temperature ranges from -10℃ to 60℃, relative humidity from 20% to 90% RH, smoke concentration from 0 to 5,000 ppm, and illuminance from 0 to 100 kLux. At the same time, composite interferences such as VOC, CO, and PM2.5 are introduced to verify the robustness of multi-source data fusion algorithms for MEMS temperature and humidity sensors, optical smoke sensors, infrared human-body sensors, and gas-sensor arrays under drift, cross-sensitivity, and hysteresis effects.

The test platform can set gradient or abrupt-change curves, record sensor response times, linkage delays,

and false-alarm rates, and automatically generate a three-dimensional evaluation matrix of environment–event–action. This provides closed-loop data for cloud-based OTA strategies, edge-AI threshold adaptation, and failure-prediction models. Ultimately, it ensures that kitchen appliances achieve millisecond-level precise coordination in real-world high-risk situations such as stir-frying, dry-burning, and gas leaks, significantly enhancing user safety and experience retention.

5.3.2. Voice interaction function verification

To verify the voice-interaction function of intelligent kitchen appliances, it is necessary to construct high-fidelity acoustic scenarios. The background noise library covers a wide range of sound levels from 20 to 80 dB, including sizzling sounds from frying and stir-frying, popping sounds from deep-frying, exhaust fan noises, and multi-source interferences such as living room TV sounds and baby crying. The dialect speech dataset includes 12 accents such as Cantonese, Sichuanese, and Hokkien, with 8,000 variant commands, and introduces continuous speech flow, inverted sentence structures, ellipsis, and colloquial expressions.

During testing, a movable artificial head and an 8-microphone circular array are used to repeatedly trigger compound semantics such as “turn the left burner to low heat” and “turn off the fire in 3 minutes” at distances of 1–5 meters and azimuths of 0–90 degrees. The wake-up rate, false wake-up rate, command delay, and rejection rate are recorded in real time. Meanwhile, a signal-to-noise ratio gradient from -5 dB to +20 dB is injected to verify the noise reduction algorithm and the directionality of Beamforming. A scenario performance heatmap is generated to drive continuous model iteration, ensuring zero false triggers and a highly robust interaction experience in noisy kitchen environments.

6. Conclusion

Automated testing equipment has shown significant advantages in the testing of home appliance control boards. By accurately simulating various working conditions, it can efficiently detect the performance of control boards, greatly enhancing the quality of home appliances. At the same time, it provides rapid feedback on test results, effectively shortening the R&D cycle and accelerating the product’s time to market.

Looking ahead to the AIoT era, intelligent testing systems will evolve towards greater intelligence and integration. By leveraging IoT technology to achieve interconnectivity among devices and utilizing AI algorithms to enhance testing accuracy and efficiency, these systems will continue to advance.

To better promote the development of the industry, it is essential to establish a comprehensive testing standard system. This system should cover various aspects such as testing methods, indicators, and procedures, ensuring the accuracy and comparability of test results. In this way, it can provide a solid guarantee for the high-quality development of the home appliance industry.

Disclosure statement

The author declares no conflict of interest.

References

[1] Vadan AM, 2024, Alternative Manual and Automatic Testing Methods of IoT Systems for Smart Homes, Doctoral

dissertation, Technical University of Cluj-Napoca.

[2] Bianchini A, Ceruti A, D'Anniballe A, et al., 2022, Inventive Redesign for Automatic Assembly in the Household Appliances Industry. Assembly Automation, 42(5): 638–652.

[3] Ameen AAM. IoT Circuit Board and Application for Smart Home Systems. 2023.

[4] Irawan Y, Wahyuni R, 2021, Electronic Equipment Control System for Households by Using Android Based on IoT (Internet of Things), Journal of Physics: Conference Series. IOP Publishing, 1783(1): 012094.

[5] Casinillo RML, So ALA, Mandaya MV, et al., 2024, Development of Arduino-Based High Heat Detector Temperature Control Prototype for Household Appliances. IAES International Journal of Robotics and Automation (IJRA), 13(2): 140.

[6] Yaldaie A, Porras J, Drogehorn O, 2024, Innovative Home Automation with Raspberry Pi: A Comprehensive Approach to Managing Smart Devices. Asian Journal of Computer Science and Technology, 13(1): 27–40.

[7] Motta LL, Ferreira LCBC, Cabral TW, et al., 2023, General Overview and Proof of Concept of a Smart Home Energy Management System Architecture. Electronics, 12(21): 4453.

[8] Gheorghe AC, Andrei H, Diaconu E, et al., 2025, Advances in Reducing Household Electrical and Electronic Equipment Energy Consumption in Standby Mode: A Review of Emerging Strategies, Policies, and Technologies. Energies (19961073), 18(4).

[9] Chaudhary P, Sharma J, Gupta RK, et al., 2023, Design and Implement a Mesh Network Communication for Smart Home Automation, 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). IEEE, 523–528.

[10] Balasingam S, Zapiee MK, Mohana D, 2022, Smart Home Automation System Using IoT. International Journal of Recent Technology and Applied Science (IJORTAS), 4(1): 44–53.

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

The Application of Artificial Intelligence Technology in Assisting R&D Project Initiation

Zhenhuan Liu*

Golden Gate University, San Francisco, CA, 94105, USA

**Author to whom correspondence should be addressed.*

Abstract: This paper reviews the latest advancements in artificial intelligence-assisted R&D project initiation, aiming to provide intelligent solutions for R&D management. It thoroughly examines the value of artificial intelligence technologies in four core areas: intelligent requirement analysis, technical feasibility assessment, market prospect forecasting, and automated risk identification. Furthermore, it proposes three forward-looking trends—enhanced intelligence, the establishment of industry standards, and deeper human-machine collaboration. These insights are expected to improve project approval success rates and shorten initiation timelines, driving a paradigm shift in R&D management from experience-based to data-driven decision-making. The review highlights how artificial intelligence, through machine learning, natural language processing, and data mining, effectively addresses chronic challenges in traditional initiation processes such as inefficiency, delayed decisions, and resource misallocation. It also identifies critical hurdles, including data quality, model interpretability, and organizational transformation, offering a vital reference framework for the future of intelligent R&D development.

Keywords: Artificial intelligence; Project initiation; Management

Online publication: October 21, 2025

1. Introduction

R&D project initiation is a critical step toward project success, shaping its direction, objectives, technological approach, and resource allocation. Scientific and well-structured initiation enables organizations to clearly define R&D goals, market positioning, and expected outcomes, ensuring alignment with corporate strategy and market demands—thereby minimizing investment risks and enhancing both efficiency and success rates. However, traditional R&D initiation processes face numerous challenges, including low efficiency, reliance on subjective judgment, imprecise market analysis, and suboptimal resource allocation. These issues often lead to project delays, budget overruns, or even outright failure.

With the rapid advancement of artificial intelligence (AI) technology, these longstanding challenges are now

becoming addressable. Currently, more than 78% of technology-driven enterprises have deeply integrated AI into their R&D initiation frameworks—a trend that continues to accelerate. AI empowers smarter, data-driven decision-making by forecasting market trends and technological feasibility based on historical data, significantly reducing uncertainty. It can automatically analyze patents, academic literature, and market reports to support technology pathway selection, identify potential risks, and optimize resource distribution, thereby increasing the precision of R&D investments. Furthermore, AI enables dynamic resource management, allowing real-time adjustments to budgets and staffing based on project progress, maximizing R&D efficiency. As a result, AI is emerging as an essential tool in modern R&D management, driving more effective and intelligent innovation strategies [1].

2. The role of AI in project initiation

2.1. Enhancing R&D project initiation efficiency

The rapid advancement of AI technology has brought unprecedented convenience and innovative support to R&D project initiation. As a core driver of digital transformation, AI empowers the project initiation process through three key technological pillars. First, machine learning (ML) enhances the scientific rigor and foresight of R&D decision-making by building high-precision predictive models, combined with clustering analysis and intelligent classification algorithms. Second, natural language processing (NLP) enables smart parsing of vast volumes of textual requirements, automatically extracting key technical indicators and supporting real-time expert consultations through intelligent Q&A systems. Third, data mining leverages advanced techniques—such as association rule analysis, time-series pattern recognition, anomaly detection, and customer segmentation—to uncover strategically valuable insights from complex datasets. The integrated application of these AI technologies not only significantly shortens the traditional R&D project initiation timeline but also dramatically improves success rates through data-driven intelligent analysis. With continuous breakthroughs in cutting-edge technologies like deep learning, AI is reshaping the entire R&D management workflow, emerging as a core competitive advantage and a pivotal enabler of innovation. For instance, Mooghal *et al.* demonstrated how AI can greatly enhance the efficiency of Ki-67 (a prognostic biomarker) assessment in molecular subtyping of breast cancer, serving as a powerful assistant for pathologists and potentially transforming breast cancer diagnosis and treatment by enabling faster, more accurate Ki-67 analysis. This substantially accelerates the efficiency of R&D project initiation [2].

2.2. Intelligent demand analysis

In the process of initiating R&D projects, accurately capturing market and user needs is crucial. AI technologies enable in-depth mining and analysis of vast amounts of textual data, user feedback, and market research, facilitating intelligent demand analysis. By leveraging natural language processing, AI can perform semantic analysis on textual information from sources such as social media, industry forums, and customer reviews, identifying user pain points, expectations, and latent needs. AI also enables segmentation and clustering of diverse user demands. Based on multidimensional data—including age, gender, location, and consumption behavior—users can be grouped into distinct segments, allowing for a deeper understanding of each group's unique requirements. This empowers companies to precisely target different user segments during new product development, enhancing market fit and product relevance. Furthermore, AI supports real-time monitoring of evolving market demands. As technology advances and society changes, customer needs continuously shift. By

continuously gathering and analyzing relevant data, AI can promptly detect early signals of changing demand patterns. For instance, Ding *et al.* suggest that applying AI in the food market enables better alignment with customer expectations [3].

2.3. Assessment of technical feasibility

Assessing technical feasibility is a critical step in initiating R&D projects. Artificial intelligence can evaluate a project's technical viability by comprehensively analyzing factors such as current technological capabilities, trends in technological advancement, and the competencies of the research team. By leveraging machine learning algorithms, AI can process vast amounts of scientific literature and patent data to understand the present state and cutting-edge developments in relevant fields. It can also simulate and anticipate potential obstacles and challenges during the R&D process. Through mathematical modeling and simulation systems, AI can replicate research and development scenarios, forecasting the likelihood of success and associated risks for various technical approaches. Furthermore, AI can assess whether a team's technical expertise is sufficient to support project execution. By examining team members' educational backgrounds, professional experience, publications, and patents, AI can quantitatively evaluate the team's technical proficiency. For instance, Claudino *et al.* applied AI techniques to analyze athletic performance and injury risk, while also examining the specific AI methods adopted across different sports disciplines [4].

2.4. Market outlook forecasting

Accurately forecasting market potential is crucial for initiating R&D projects. Artificial intelligence can analyze diverse information—including market data, industry trends, and consumer behavior—to predict a project's market viability. AI also evaluates competitive dynamics, assessing a project's position within the market landscape. By examining competitors' product features, market share, and marketing strategies, AI identifies a project's strengths and weaknesses. Furthermore, AI can uncover latent market demand and emerging opportunities by analyzing societal trends, technological advancements, and shifts in consumer lifestyles. For instance, Yu *et al.* leveraged AI techniques to explore the fermented beverage market, significantly enhancing product competitiveness [5].

2.5. Automated risk identification

In the process of initiating R&D projects, various risks may arise, including technological, market, and financial risks. Artificial intelligence can monitor project progress and the resolution of technical challenges to promptly identify technological risks. By analyzing shifts in market demand, competitor activities, and changes in policies and regulations, AI can predict the likelihood and impact of market risks. Through assessment of budget execution, funding sources, and expenditure patterns, AI is capable of detecting financial risks. Companies can leverage these early warnings to optimize financial management and ensure smooth project execution. AI can also classify and prioritize risks, enabling project teams to focus on critical issues and enhance the likelihood of project success. Additionally, AI supports the identification of food safety risks. For instance, as noted by Mu *et al.*, AI can detect bacteria and fungi in food, thereby improving food safety [6].

3. Challenges ahead

3.1. Data quality constraints

In terms of accuracy, 10–20% of clinical trial data in the pharmaceutical industry contains recording errors.

Regarding completeness, critical data is often lost in chip testing due to equipment failures. On timeliness, rapid iteration in the software industry renders outdated data misleading for decision-making. These issues directly impair the training performance and predictive accuracy of AI models, forming a key bottleneck that limits the quality of R&D project initiation.

3.2. Model interpretability

Current AI models, especially deep learning systems, suffer from the "black box" problem, making it difficult to clearly explain the rationale behind predictions such as drug development success rates. This lack of interpretability makes risk assessment challenging for decision-makers, significantly restricting the deeper integration of AI into R&D decision processes.

3.3. Organizational resistance to change

The implementation of AI faces three major organizational barriers: operationally, many employees resist automated project approval workflows; in terms of talent, there is a shortage of data scientists, and R&D staff require new skill training; culturally, a conservative mindset leads employees to rely more on traditional experience. Additionally, departments may resist AI adoption out of concern that it will diminish their decision-making authority. These factors collectively hinder the effective application of AI in R&D management.

4. Future development trends

4.1. Advancement in intelligence

AI in R&D project initiation is poised for a transformative leap—from basic data analysis to autonomously identifying complex patterns. Progress in natural language interaction will enable more efficient human-machine communication, significantly reducing coordination costs during project initiation. By 2030, intelligent AI systems are expected to boost R&D success rates by 30%.

4.2. Establishment of industry standards

There is an urgent need to establish unified standards for AI applications in R&D. Addressing collaboration barriers caused by inconsistencies in current evaluation methods, future standards will streamline critical processes such as data quality control and model assessment—establishing clear requirements for data sources and defining unified metrics like accuracy rates.

4.3. Deepening human-AI collaboration

AI will evolve into an intelligent partner for human researchers. In the initial project planning phase, AI can uncover global research data to suggest innovative directions, while researchers apply domain expertise to refine these proposals. During execution, AI continuously monitors progress and flags potential risks, enabling researchers to adapt strategies in real time.

5. Conclusion and prospects

AI-assisted project initiation is gradually transforming traditional R&D management. Through intelligent analysis, forecasting, and decision optimization, it significantly enhances both efficiency and success rates in research

and development. Despite ongoing challenges related to data quality, model interpretability, and organizational adoption, continuous technological advancements, standardized frameworks, and deeper human-AI collaboration are paving the way for broader AI integration. In the future, AI will further drive R&D decision-making toward data-driven, intelligently collaborative models, offering stronger support for technological innovation and industrial upgrading.

Disclosure statement

The author declares no conflict of interest.

References

[1] Hamamoto R, 2021, Application of Artificial Intelligence for Medical Research. Biomolecules, 11(1): 90.

[2] Mooghal M, Anjum S, Khan W, et al., 2024, Artificial Intelligence-Powered Optimization of KI-67 Assessment in Breast Cancer: Enhancing Precision and Workflow Efficiency. A Literature Review. J Pak Med Assoc, 74(4 (Supple-4)): S109–S116.

[3] Ding H, Tian J, Yu W, et al., 2023, The Application of Artificial Intelligence and Big Data in the Food Industry. Foods, 12(24): 4511.

[4] Claudino JG, Capanema DO, de Souza TV, et al., 2019, Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: A Systematic Review. Sports Med Open, 5(1): 28.

[5] Yu H, Liu S, Qin H, et al., 2024, Artificial Intelligence-Based Approaches for Traditional Fermented Alcoholic Beverages' Development: Review and Prospect. Crit Rev Food Sci Nutr, 64(10): 2879–2889.

[6] Mu W, Kleter GA, Bouzembrak Y, et al., 2024, Making Food Systems More Resilient to Food Safety Risks by Including Artificial Intelligence, Big Data, and Internet of Things into Food Safety Early Warning and Emerging Risk Identification Tools. Compr Rev Food Sci Food Saf, 23(1): e13296.

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Quantum-Secure OTN Framework Integrating QKD-PQC Technologies

Wenliang Zhang[1,2], Jiao Zhao[2]*, Bao Tang[2]*, Wei Huang[1,2], Binbin Xu[1,2], Miao Li[1,2], Linfeng Wang[1,2], Bo Liu[1,2], Gongchong Zhong[1,2]

[1]Xintong Digital Intelligence Quantum Technology Co., Ltd., Beijing 100176, China
[2]CAS Quantum Network Co., Ltd., Shanghai 201315, China

**Authors to whom correspondence should be addressed.*

Abstract: The Optical Transport Network (OTN) is a protocol for sending network messaging over optical fiber networks. Intelligent optical networks provide an ideal solution for high-bandwidth services. Currently, data encryption schemes for OTN typically rely on mathematical problems such as elliptic curve cryptography or discrete logarithms, which are vulnerable to attacks by quantum computers. This paper investigates a quantum-secure OTN Framework that integrates Quantum Key Distribution (QKD) and Post-Quantum Cryptography (PQC) technologies, enabling OTN leased lines to resist quantum attacks. This framework can provide users with highly secure quantum-encrypted OTN leased lines services.

Keywords: OTN leased lines; QKD; Quantum key; Quantum Security Service Platform

Online publication: October 17, 2025

1. Introduction

The Optical Transport Network (OTN) is a transport network that implements the transmission, multiplexing, routing, and monitoring of service signals within the optical domain. It ensures the performance and survivability of service signals [1]. The high-quality OTN-based leased lines have become a key area of competition among operators due to their large capacity, long distance, low latency, and high availability [2].

Currently, operator-provided OTN leased lines typically employ non-encrypted transmission. To meet the practical requirements of enterprises for secure communications, OTN supports encryption technologies to mitigate the risks of data leakage and theft.

Traditional encryption solutions are utilized for achieving customer data protection, including elliptic curve cryptography, large-number factorization mathematical models, and AES block cipher algorithms based on public-key negotiation are employed to encrypt Optical Channel Payload Unit (OPUk) payloads between OTN devices (**Figure 1**).

However, these encryption assumptions are vulnerable to quantum computing threats and incapable of resisting quantum computational attacks using quantum algorithms such as Shor's and Grover's algorithms [3]. With the rapid advancement of quantum computing, addressing threats posed by quantum computational attacks has become an urgent necessity for business applications in highly secure industries, including government affairs, finance, electricity, and data centers.

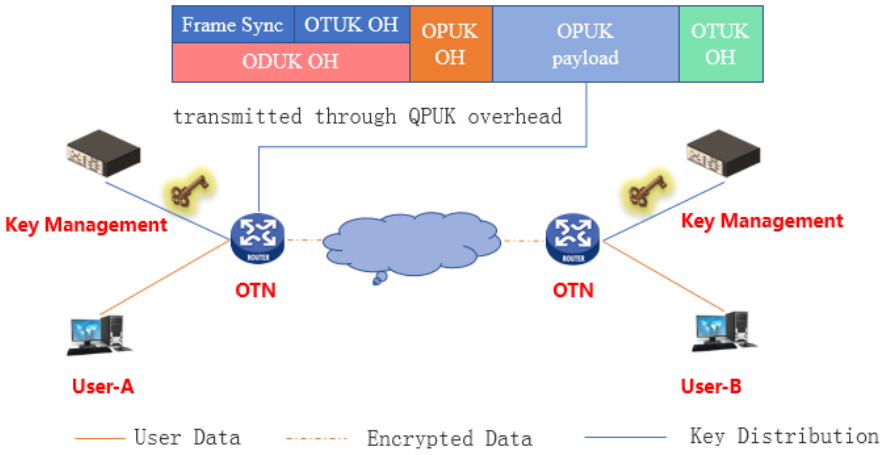


Figure 1. Quantum-secure communication between OTN equipment

To defend against threats posed by quantum computing, two main approaches have gained international recognition: Quantum Key Distribution (QKD), which leverages the principles of quantum physics, and Post-Quantum Cryptography (PQC), which is based on computational complexity.

QKD refers to the method and process by which two communicating parties generate and distribute cryptographic keys with information-theoretic security through the transmission of quantum states. Its security is guaranteed by fundamental principles of quantum mechanics, notably the indivisibility of single photons and the no-cloning theorem of quantum states [4]. In parallel, software systems and protocol standards supporting QKD have been developed. For example, the Chinese standard YD/T 4303-2023 specifies technical protocols, functional performance requirements, and testing methods for quantum-secure communication gateway and terminal devices based on the IPsec protocol [5].

Another quantum-resistant security technique is Post-Quantum Cryptography (PQC). It builds on well-studied hardness assumptions in mathematics, including lattice-based, code-based, and multivariate polynomial cryptography. These hardness assumptions are widely believed to remain computationally infeasible even for the quantum computer. PQC is mandated by the U.S. National Security Agency (NSA) in its Commercial National Security Algorithm Suite 2.0 (CNSA 2.0) [6]. The U.S. National Institute of Standards and Technology (NIST) has forecasted that quantum computers may pose a realistic threat to existing classical encryption algorithms as early as 2030 [7].

This paper presents a quantum-secure OTN Framework that merges PQC for initialization and QKD for encrypted transmission.

(1) Initialization phase (PQC-based): Legacy OTN devices are provisioned with initial keys using Post Quantum Cryptography (PQC) algorithms. This scheme enables secure device onboarding, authentication, authorization, and quantum key injection.

(2) Encrypted transmission phase (QKD-based): QKD technology is used to establish quantum-secure OTN transmission leased lines. By interfacing with the QKD network, OTN devices can securely and cost-effectively obtain quantum-generated keys. These keys are then used in symmetric encryption algorithms to encrypt the payloads of OPUk. This approach enables OTN leased lines to be resistant to quantum attacks.

2. Challenges on the quantum-secure OTN framework

2.1. Unified interface definition

Traditional OTN equipment manufacturers often use proprietary protocols in OTN networking, resulting in interoperability challenges across devices from different manufacturers. Hardware products from different manufacturers may also vary in their implementation. The key to this technical solution is addressing how to adapt to heterogeneous OTN equipment models, standardize interface types, and unify transmission protocols to enable secure and reliable quantum key distribution across diverse OTN infrastructures.

2.2. Quantum key distribution integrating QKD-PQC

QKD networks offer a secure solution for key distribution between cities. However, within urban areas—such as between an OTN aggregation point and a nearby OTN terminal—deploying a full QKD network for key distribution poses challenges in terms of cost-efficiency and operational convenience. The core challenge for the feasibility of this technical solution lies in how to distribute quantum keys to OTN devices securely and economically, while establishing end-to-end encrypted OTN leased lines. Integrating QKD and PQC technologies plays a crucial role in addressing the "last mile" problem in quantum key delivery and accelerating the large-scale commercialization of quantum-secure communication technologies [8].

3. Methods

3.1. Architecture of quantum-secure OTN framework

Figure 2 shows the overall network architecture, which is divided into three layers: the quantum layer, the key management layer, and the service layer.

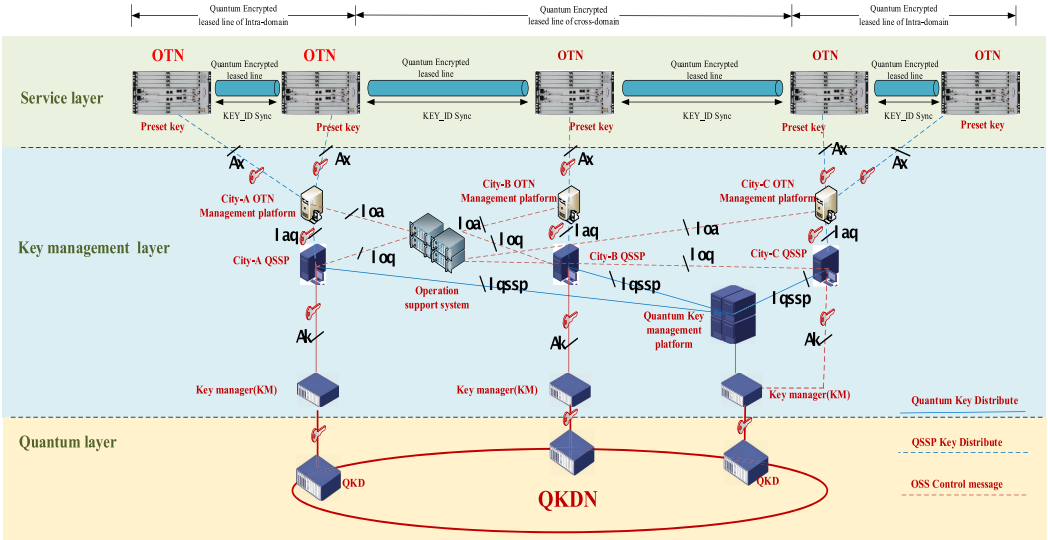


Figure 2. Architecture of quantum-secure OTN framework

3.1.1. Quantum layer

This layer leverages an integrated ground-satellite quantum key distribution (QKD) network to facilitate quantum key agreement and distribution between OTN devices across city-level, provincial, and cross-provincial domains.

3.1.2. Key management layer

This layer is organized into two levels—provincial and municipal (**Table 1**).

Table 1. Levels of key management layer

Locations	Platform	Shortcuts
Provinces	Operation support system	Responsible for user accounts opening, management, etc., and distribute control information to the OTN management platform and the QSSP.
Provinces	Quantum key management platform	Manage the QSSP of each city, and synchronize quantum keys among the QSSP through QKDN.
Cities	QSSP	Distribute quantum keys among OTN equipment through the OTN management platform.
Cities	OTN management platform	Management authentication and authentication for OTN equipment in each city, distribute quantum keys from the QSSP to OTN equipment.

3.1.3. Service layer

OTN-CPE devices are deployed at the user end. User-side OTN devices obtain quantum session keys through the municipal-level quantum security service platform. During the initialization phase, PQC algorithms are employed to securely inject pre-distributed quantum keys into the secure cryptographic modules of the OTN devices via the municipal platform. During the secure communication phase, the OTN devices establish quantum-encrypted leased lines using quantum session keys distributed by the QKD network.

3.2. Unified interface

The unified interface is defined as follows (**Table 2**).

Table 2. Unified interfaces

NUM	Interfaces	Shortcuts
1	AK	Reference point between the APP and the KSA of the QKDN, whose main functions include key request and provision between the APP and the KM, etc. [9]
2	Iqssp	Reference point between the QSSP and the OTN management platform. Be responsible for the operation maintenance, authentication, and authorization, distributed quantum keys from the QSSP to OTN equipment.
3	Ioq	Reference point between the operation management platform and the QSSP, provides authentication, and the management of encrypted services, such as leased lines establishment and deletion, etc.
4	Ioa	Reference point between the operation management platform and the OTN management platform. provides authentication and management of encrypted services, such as establishment, deletion, and billing, etc.
5	Iaq	Reference point between the OTN management platform and the QSSP, and the logical interface between the OTN CPE and the QSSP. Provide authentication management and distribution of quantum keys, etc.
6	Ax	Reference point between the OTN equipment and the OTN management platform. Exchange information between OTN equipment based on the specific protocol of the OTN manufacturer.

3.3. Methods of the initialization of OTN CPE

The initialization of OTN CPE mainly includes authentication and authorization between OTN equipment and municipal quantum security service platforms, as well as the injection of quantum keys into OTN CPE. The transmission protocol from the provincial OTN management platform to the OTN equipment of the user layer is private for each equipment manufacturer (**Figure 3**).

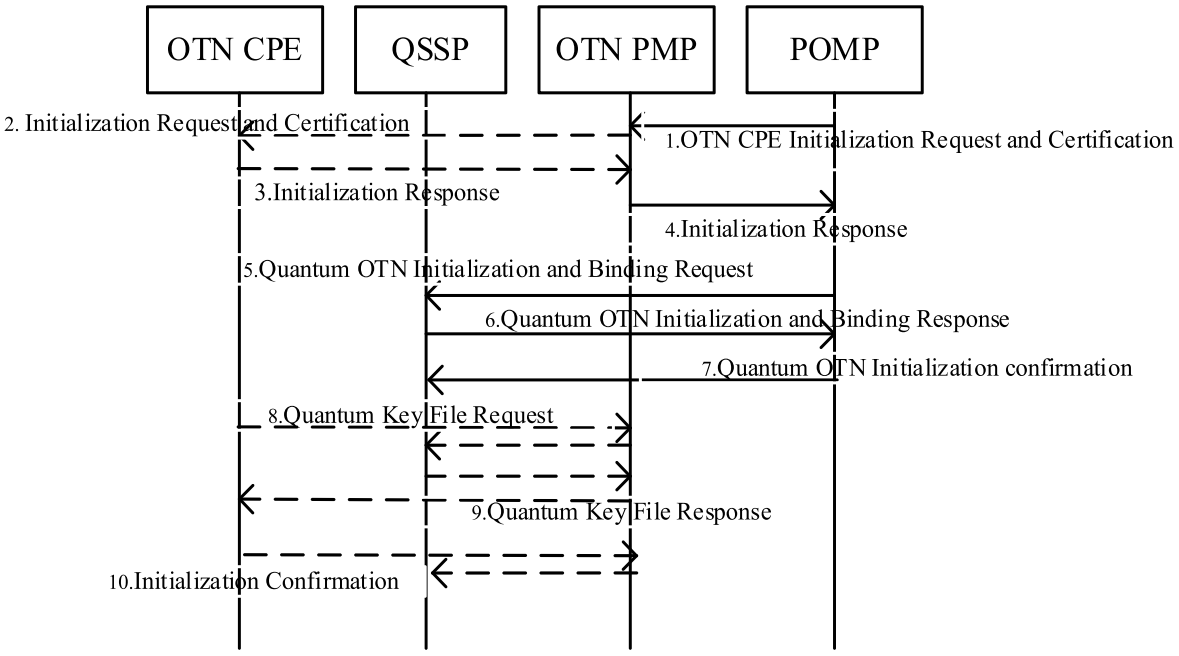


Figure 3. Initialization process of OTN CPE

(1) The provincial operation management platform (POMP) initiates the message "OTN CPE Initialization Request and Certification" to the OTN provincial management platform (OTN PMP). This message carries the unique identifier of the OTN CPE, which requires quantum encryption.

(2) The OTN PMP forwards the message "Initialization Request and Certification" to the OTN CPE according to the unique identifier.

(3) The OTN CPE replies with the message "Initialization Response" to the OTN PMP, and completes authentication and registration within the OTN PMP.

(4) The OTN PMP forwards the "OTN CPE Initialization Response" to complete authentication and registration within the POMP.

(5) The POMP initiates the message "Quantum OTN Initialization and Binding Request" to QSSP, carrying the OTN CPE identification information required for quantum encryption. QSSP prepares the quantum key file if it receives the message successfully.

(6) The municipal QSSP responds to the POMP with a message "Quantum OTN Initialization and Binding Response," when OTN CPE identity authentication and OTN CPE binding is finished.

(7) The POMP responds to QSSP with "Quantum OTN Initialization confirmation," and records the results of identity authentication and binding.

(8) The OTN CPE initiates a "Quantum Key File Request" to the municipal QSSP, which is forwarded by the OTN PMP, carrying the unique identifier of the OTN CPE.

(9) The municipal QSSP responds to the OTN CPE with "Quantum Key File Response," which is forwarded by the OTN PMP.

(10) The OTN CPE initiates a message "Initialization Confirmation" to the municipal QSSP, which is forwarded by the OTN PMP, carrying the Initialization results.

3.4. Method of quantum OTN leased line

The establishment of a quantum OTN leased line primarily involves mutual identity authentication, service authentication, and confirmation between the OTN equipment and the OTN PMP, POMP, and municipal QSSP. The two QSSPs can synchronize relevant business information and obtain quantum session keys from the QKD network. The quantum session keys are encrypted and distributed to two OTN CPEs, which require quantum encryption through the OTN PMP. The transmission protocol from the OTN PMP to the OTN equipment of the user layer is private for each equipment manufacturer (**Figure 4**).

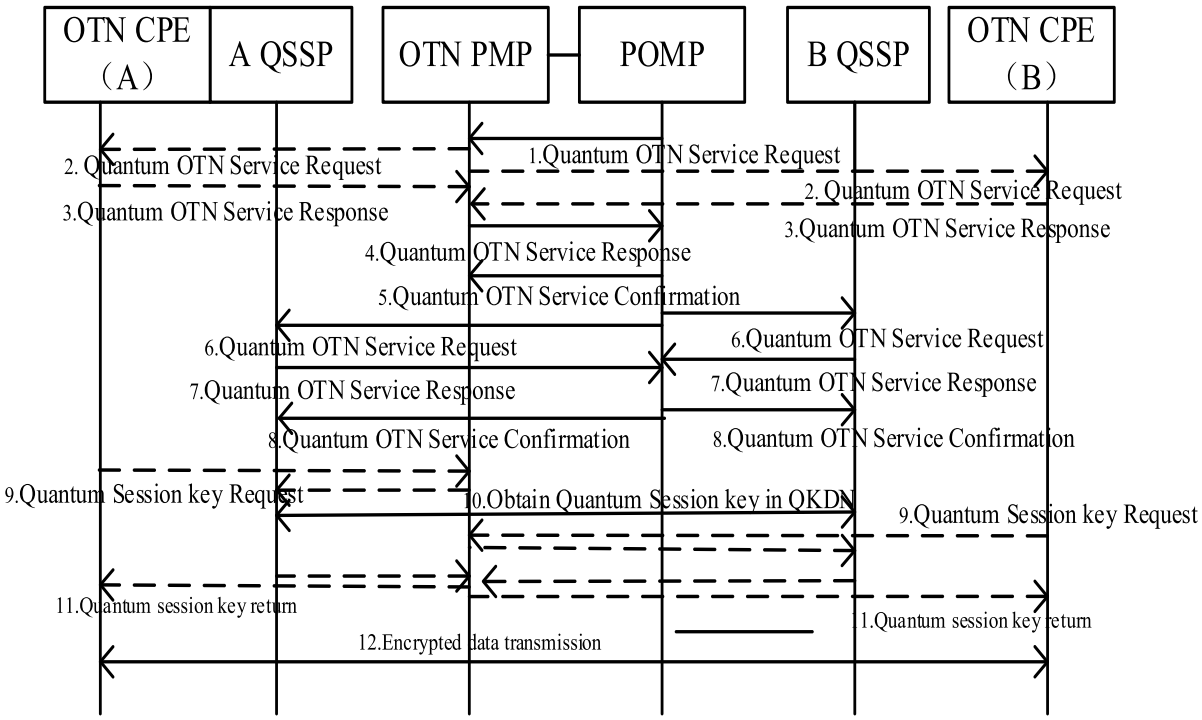


Figure 4. Quantum OTN dedicated line establishment

(1) The POMP initiates the message "Quantum OTN Service Request," with QOS and authentication information to the OTN PMP, carrying the unique identifier of two OTN CPEs to establish a quantum OTN dedicated line.

(2) The OTN PMP forwards the "Quantum OTN Service Request" message to the OTN CPEs at both ends, notifying both OTN CPEs ready for service.

(3) The OTN CPE replies with the message "Quantum OTN Service Response" to the OTN PMP and completes the identity authentication process.

(4) The OTN PMP forwards the message "Quantum OTN Service Response" to the POMP, and completes the identity authentication of the OTN CPE to the POMP.

(5) The POMP initiates the "Quantum OTN Service Confirmation" message to the OTN PMP, verifies the

QOS information, and the Token information corresponding to the OTN CPEs.

(6) The POMP initiates the message “Quantum OTN Service Request” to the municipal QSSP, which the OTN CPE belongs to.

(7) The municipal QSSP replies with the message “Quantum OTN Service Response” to the OTN PMP and completes the identity authentication process.

(8) The POMP returns “Quantum OTN Service Confirmation” message to the municipal QSSP, and completes bidirectional quantum OTN service authentication and confirmation.

(9) The OTN CPE (A) initiates the message “Quantum Session key Request” to the home QSSP, which is forwarded by the OTN PMP.

(10) After the identity authentication is performed by the municipal QSSP, relevant management information is synchronized between the two QSSPs, and quantum session keys are obtained from the QKD network.

(11) The QSSP distributes the quantum session keys, which are encrypted to two OTN CPEs, and this distribution is forwarded by the OTN PMP.

(12) After obtaining the quantum session keys, the OTN CPE encrypts and transmits the OPUk payload using the quantum session keys..

4. Retrospect and prospects

This paper presents a Quantum-Secure OTN Framework integrating QKD-PQC technology, enabling OTN devices to obtain quantum keys from the QKD network safely and at low cost. Then, quantum keys are used to encrypt the payload of OTN OPUk frames, making the OTN secure communication achieve the security capability of resisting quantum attacks. On the one hand, the unfiled interface definitions have been unified to address the issue of interface types and transmission protocol adaptation for different models of OTN devices from various manufacturers. On the other hand, by integrating QKD and PQC technologies, it simplifies the processes of key issuance and subsequent usage and management, breaks through the “last mile” problem of key transmission, and thus promotes the implementation of the OTN quantum encryption technology solution.

One of the practical goals of quantum encryption technology is to achieve large-scale commercial deployment. Therefore, by leveraging the existing classic OTN optical fiber network infrastructure, co-fiber transmission of OKD and classic signals is an important approach to achieving large-scale deployment of OKD. Due to the weak strength and susceptibility of quantum signals, as well as the single function of the OTN terminal equipment on the user side. The co-fiber transmission technology of QKD and OTN still faces challenges. Research on the co-fiber transmission technology of QKD and OTN networks needs to be continued [10].

Funding

National Development and Reform Commission (NDRC) New-Generation Information Infrastructure Construction Project: National Wide-Area Quantum Secure Communication Backbone Network Construction Project (0747-2260SCCSHV90(001))

Disclosure statement

The authors declare no conflict of interest.

References

[1] Fang C, 2012, Research on Key Technologies of Optical Transport Network (OTN), dissertation, Beijing University of Posts and Telecommunications.

[2] Zang Y, Yin Z, Wang L, 2023, Discussion on Interoperability Application of OTN-CPE Multi-Vendors. Telecommunications Design Technology, 2023(4).

[3] Shor PW, 1999, Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. SIAM J. Sci. Statist. Comput.

[4] Tang B, Ying J, Li J, 2025, The Technology and Networking Strategy of the Quantum Secure Communication Bearer Network with Space-Integrated-Ground. Communications Technology, 58(3): 232–240.

[5] YD/T 4303-2023 Technical Specifications for Quantum Secure Communication Application Equipment Based on IPsec Protocol, 2023, https://www.chinesestandard.net/PDF/English.aspx/YDT4303-2023

[6] National Security Agency, 2022, Announcing the Commercial National Security Algorithms.

[7] Chen L, Jordan S, Moody D, et al., 2016, Report on Post-Quantum Cryptography. National Institute of Standards and Technology. https://nvlpubs.nist.gov/nistpubs/ir/2016/nist.ir.8105.pdf

[8] Wang LJ, Zhang KY, 2020, Experimental Authentication of Quantum Key Distribution with Post-Quantum Cryptography. arXiv. https://arxiv.org/abs/2009.04662

[9] YDT 4301-2023 Quantum Secure Communication Network Architecture, 2023, https://www.gb-gbt.com/PDF/Related.aspx/YDT4301-2023

[10] Gong J, 2022, Research on Quantum Key Distribution System for Classical-Quantum Co-Fiber Transmission, dissertation, University of Science and Technology of China.

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

CW-HRNet: Constrained Deformable Sampling and Wavelet-Guided Enhancement for Lightweight Crack Segmentation

Dewang Ma*

College of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, China

**Author to whom correspondence should be addressed.*

Abstract: This paper presents CW-HRNet, a high-resolution, lightweight crack segmentation network designed to address challenges in complex scenes with slender, deformable, and blurred crack structures. The model incorporates two key modules: Constrained Deformable Convolution (CDC), which stabilizes geometric alignment by applying a tanh limiter and learnable scaling factor to the predicted offsets, and the Wavelet Frequency Enhancement Module (WFEM), which decomposes features using Haar wavelets to preserve low-frequency structures while enhancing high-frequency boundaries and textures. Evaluations on the CrackSeg9k benchmark demonstrate CW-HRNet's superior performance, achieving 82.39% mIoU with only 7.49M parameters and 10.34 GFLOPs, outperforming HrSegNet-B48 by 1.83% in segmentation accuracy with minimal complexity overhead. The model also shows strong cross-dataset generalization, achieving 60.01% mIoU and 66.22% F1 on Asphalt3k without fine-tuning. These results highlight CW-HRNet's favorable accuracy-efficiency trade-off for real-world crack segmentation tasks.

Keywords: Crack segmentation; Lightweight semantic segmentation; Deformable convolution; Wavelet transform; Road infrastructure

Online publication: October 17, 2025

1. Introduction

Crack detection is a critical task in road infrastructure maintenance, as road cracks are among the most common and hazardous pavement defects [1]. Accurate segmentation directly impacts the development of maintenance strategies and the long-term safety of transportation systems. Traditional manual inspection, while reliable to some extent, suffers from high labor costs, low efficiency, and strong subjectivity, making it increasingly unsuitable for modern, intelligent maintenance workflows [2].

Before the advent of deep learning, crack segmentation primarily relied on handcrafted features and classical machine learning methods. Typical approaches included edge operators, texture statistics, and descriptors such

as HOG or LBP, combined with classifiers like SVM, random forest, or K-means clustering [3]. While effective under low-noise and homogeneous conditions, these methods lacked robustness and generalizability in the face of complex materials, lighting conditions, and diverse crack shapes, due to their limited representational capacity.

With the rise of deep learning, convolutional neural networks have become the dominant paradigm for crack detection and segmentation. End-to-end pixel-wise training significantly improves robustness under complex imaging conditions [4]. Contextual modeling via feature pyramids and multi-scale fusion/attention strategies enhances the network's adaptability to cluttered backgrounds [5]. Architectures such as UNet, which utilize encoder-decoder frameworks with skip connections, effectively integrate fine-grained details and high-level semantics, thereby improving crack boundary sharpness and structural continuity [6]. Recent works further optimize the consistency and efficiency of feature learning by introducing dense feature aggregation and multi-level skip connections [7].

Despite these advances, two major challenges remain: First, in terms of geometric alignment, traditional convolutions with fixed sampling locations struggle to adapt to the irregular, slender, and bifurcated nature of cracks. While deformable convolutions introduce learnable offsets for enhanced shape modeling, unconstrained offsets may lead to instability, such as excessive deformation or irrelevant region sampling, resulting in distorted features and training difficulty. To address this, we propose Constrained Deformable Convolution (CDC), which introduces a tanh-based offset limiter and a learnable scaling factor to adaptively control sampling magnitudes, thereby achieving stable and precise alignment of complex crack structures [8]. Second, in the frequency domain, repeated downsampling in CNNs tends to erase high-frequency details, leading to blurred boundaries and poor texture representation. Purely spatial convolutions also struggle to jointly capture low-frequency global topology and high-frequency fine edges. To bridge this gap, we design the Wavelet Frequency Enhancement Module (WFEM), which decomposes feature maps into low-frequency (LL) and high-frequency (LH/HL/HH) subbands via Haar wavelets. Each subband undergoes lightweight convolutional projection and cross-subband residual modeling for information interaction, followed by an inverse wavelet transform to reconstruct the full-resolution feature map. This enables the network to preserve global topology while enhancing boundary and texture fidelity.

In summary, this paper addresses two critical limitations—unstable geometric alignment and loss of high-frequency details—through the integration of CDC and WFEM in a high-resolution, lightweight architecture, providing a principled foundation for the design and evaluation of the proposed network in subsequent sections.

2. Related work

2.1. Constrained Deformable Convolution

Deformable convolution augments standard convolution by introducing learnable spatial offsets that relax the constraint of a fixed sampling grid, thereby increasing the network's capacity to model geometric variation. By predicting offsets with an auxiliary branch and applying bilinear interpolation at displaced locations, the effective receptive field becomes content-adaptive rather than purely grid-bound. This property has made deformable operators highly effective in dense prediction tasks—particularly object detection and semantic/instance segmentation—where target shapes may be elongated, multi-scale, or discontinuous. In essence, the kernel is no longer tied to a rigid lattice; it can bend toward informative structures and away from distractors, improving coverage of thin filaments, junctions, and tortuous boundaries.

In crack segmentation, this adaptivity is especially valuable: cracks are typically slender, irregular, and

branched, with low contrast and significant texture interference from surrounding materials. Vanilla deformable convolutions, however, come with a notable caveat. When the offset field is unconstrained, the model may push sampling points far beyond the neighborhood where the underlying features are reliable. Such excessive displacements or drifts into irrelevant regions lead to distorted local evidence, noisy gradients, and training instability. The risk is amplified along crack boundaries, where subtle misplacement can blur edges or fragment topology.

To counter these effects, we introduce CDC, which explicitly regularizes the offset magnitude during generation. Concretely, CDC applies a tanh-based limiter followed by a learnable scaling factor s, mapping raw offsets into a bounded range that still permits meaningful deformation. The tanh operation suppresses extreme values symmetrically, while s adapts the allowable offset scale to local statistics and task difficulty. This adaptive upper bound curbs erratic sampling without reverting to a rigid grid, striking a practical balance between flexibility and stability.

The resulting offset field is smoother, better conditioned, and less prone to outliers, which in turn improves boundary alignment and feature fidelity near thin, branching structures. Empirically, CDC stabilizes optimization, reduces artifacts linked to offset explosion, and yields cleaner gradients for the backbone. In downstream decoding, features produced by CDC exhibit crisper edges and more coherent topology, enabling the overall network to maintain high-resolution detail while remaining robust to geometric variability and background clutter.

2.2. Wavelet transform and frequency-domain feature fusion

The wavelet transform offers a distinctive balance between spatial locality and frequency resolution, enabling simultaneous representation of structural context and detailed variations. Recently, wavelet-based approaches have gained attention in deep learning as both an alternative to conventional downsampling/upsampling layers and as a feature enhancement tool [9]. Compared to max-pooling or strided convolutions, which often discard fine details, wavelet decomposition retains richer structural cues by explicitly separating low-frequency and high-frequency components. This property is particularly beneficial for segmentation tasks, where subtle edge continuity and texture fidelity are critical for accurate predictions. By preserving edges and textures through multiple downsampling stages, wavelet-based modules allow networks to sustain fine-grained representation capacity that standard CNN architectures often fail to maintain [10]. Typical strategies in prior work have leveraged the discrete wavelet transform (DWT) to replace pooling operations, and its inverse counterpart (IDWT) to replace upsampling layers. Another common design decomposes feature maps into a low-frequency subband (LL) that captures overall structure and several high-frequency subbands (LH, HL, HH) that emphasize edges and fine patterns. These components are then recombined in various ways to enrich feature hierarchies and strengthen boundary localization. Despite their promise, most existing implementations suffer from two notable limitations. First, decomposition is often restricted to low-resolution stages, thereby neglecting early and mid-level features where much of the fine-grained information resides. Second, subbands are frequently handled in a simplistic manner, such as direct concatenation, without explicit modeling of inter-subband dependencies. As a result, these approaches may underutilize complementary relationships between frequency components, leading to boundary blurring, detail loss, and limited robustness.

To address these shortcomings, we propose the WFEM. In WFEM, input features are decomposed using fixed Haar wavelets into one LL and three HF subbands. Each subband is then projected through lightweight 1×1 convolutions with normalization and nonlinearity, ensuring compactness and recalibration of channel responses.

The processed subbands are concatenated along the channel dimension, followed by cross-subband residual modeling, which explicitly enables information flow and interaction across LL and HF components. Finally, the refined features are reconstructed using IDWT to restore spatial resolution. This design simultaneously preserves low-frequency topological connectivity and enhances high-frequency boundary sharpness and texture richness, effectively overcoming the deficiencies of pure decomposition or naive concatenation strategies.

3. Methodology

3.1. Overall network architecture

CW-HRNet follows a dual-path encoder–multi-scale fusion–progressive decoder design to preserve high-resolution representations while effectively integrating global semantics with fine crack details, as illustrated in **Figure 1**. The input is first processed by shallow convolutions and normalization to obtain unified-scale features. In the dual-path encoder, the high-resolution branch stacks multiple CDC layers, which progressively apply constrained geometric adaptation to sampling positions. This stabilizes the alignment of elongated, branched, and tortuous crack boundaries while retaining shallow textures and fine-grained structures. In parallel, the low-resolution branch enlarges the receptive field to capture global context and complements the high-resolution pathway through cross-scale fusion and semantic interaction.

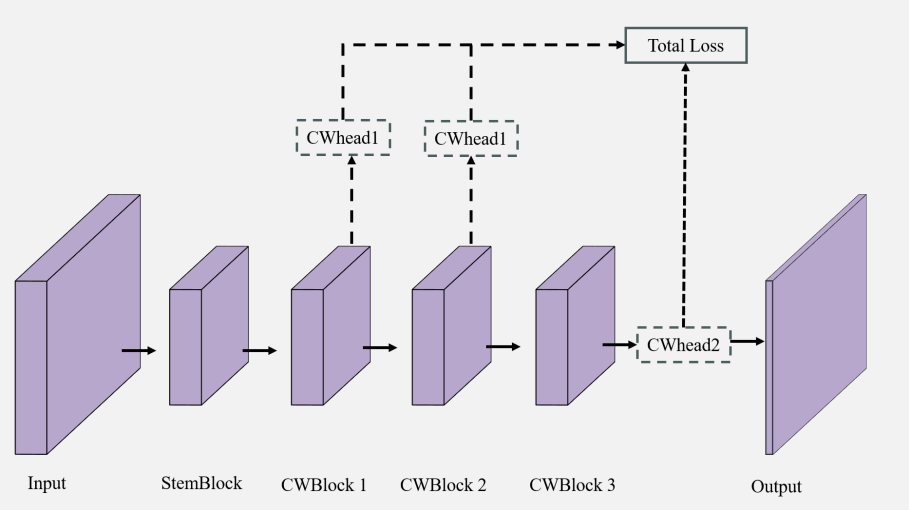


Figure 1. Overall architecture of CW-HRNet

To compensate for high-frequency loss caused by downsampling, the high-resolution branch incorporates the WFEM after CDC. Specifically, Haar wavelets are employed to decompose features into LL/LH/HL/HH subbands. Each subband undergoes lightweight projection and cross-subband residual modeling to enable effective information interaction, followed by inverse DWT reconstruction. This design achieves "topology preservation in low frequency and boundary strengthening in high frequency", thereby enhancing feature representation in the frequency domain.

During the decoding stage, cascaded convolutions and progressive upsampling gradually restore spatial resolution, while skip connections mitigate detail degradation. A final 1×1 convolution maps features into pixel-wise crack probability maps, optimized jointly with a standard binary classification loss.

In summary, CW-HRNet introduces targeted modifications to the classical high-resolution framework along two complementary dimensions: CDC constrains offset drift and improves boundary alignment stability, while WFEM explicitly models cross-subband dependencies to recover high-frequency details. Their synergy enables the network to achieve a superior balance between accuracy and efficiency, while maintaining robustness in complex crack segmentation scenarios.

3.2. Deformable convolution enhancement module

In real-world scenarios, cracks often appear irregular, slender, and branched. Standard convolutions, constrained by fixed sampling locations, struggle to adapt to such complex geometries. Although the original deformable convolution introduces learnable offsets, the absence of proper constraints may lead to excessive deformations and boundary drift, causing training instability. To address this, we propose the CDC module, as illustrated in **Figure 2**, the offset generation stage replaces conventional convolutions with an OffsetConvBlock, which progressively extracts geometric cues and enhances the discriminability and stability of offset prediction. Furthermore, we design an Offset Regulation Module (ORM), which imposes adaptive constraints on the offset magnitude by combining a tanh-based limiter with a learnable scaling factor. This mechanism suppresses structural distortion at the source by preventing extreme sampling.

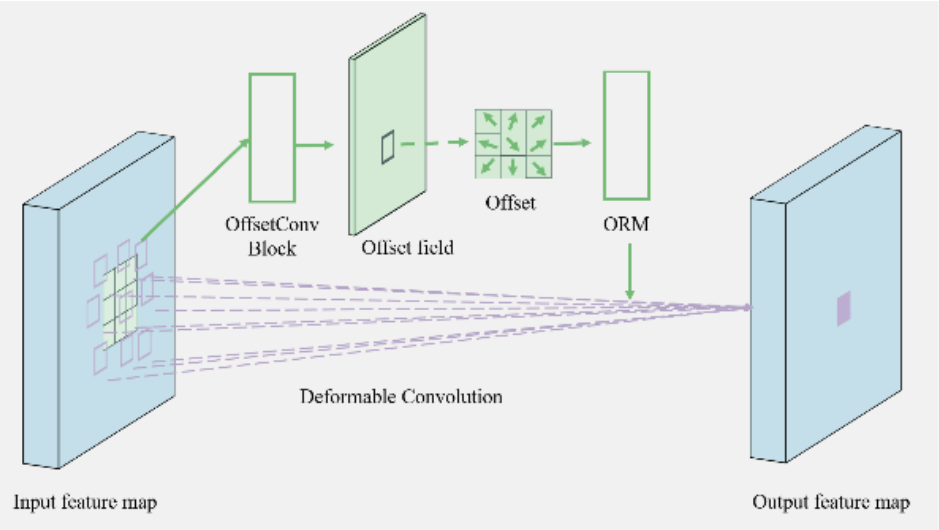


Figure 2. Structure of the constrained deformable convolution module

In the CDC, each convolutional sampling point k is associated with a learnable offset. The ORM constraint is formulated as:

$$\Delta p_k^*(x) = s \cdot \tanh(\Delta p_k(x)) \tag{1}$$

where tanh(·) compresses offsets into [−1,1] to suppress extreme values, and s is a learnable scaling factor that adaptively adjusts the offset magnitude. Based on the constrained offsets Δp∗(x), the deformable convolution can be expressed as:

$$y(p_0) = \sum_{k=1}^{K} w_k \cdot x\left(p_0 + p_k + \Delta p_k^*(x)\right) \tag{2}$$

where p_0 denotes the convolution kernel center, $p_k \in \mathbb{Z}^2$ represents the regular sampling grid, and x(·) indicates bilinear interpolation sampling [11].

Compared with the unconstrained DCN, the proposed CDC maintains geometric flexibility while significantly improving the stability of offset prediction and boundary alignment accuracy. When stacked in multiple layers within the high-resolution branch, CDC provides structurally coherent and edge-preserving geometric representations, thereby supplying the decoder with more reliable features and ultimately improving both segmentation accuracy and robustness [12].

3.3. High–low frequency feature decoupling and fusion module

Cracks often exhibit slender, tortuous, or even branched patterns, with a high degree of similarity to background textures. This leads to significant differences in the statistical distribution of high-frequency details and low-frequency structures. Conventional convolutional networks tend to lose high-frequency information after multiple downsampling operations, while low-frequency global semantics are insufficiently captured due to limited receptive fields [13]. As a result, boundaries become blurred, fine-grained textures are lost, and local topology is often disrupted. To overcome these limitations, we propose the WFEM, which achieves joint enhancement of global and local features through frequency-domain decomposition and cross-subband modeling, as illustrated in **Figure 3**.

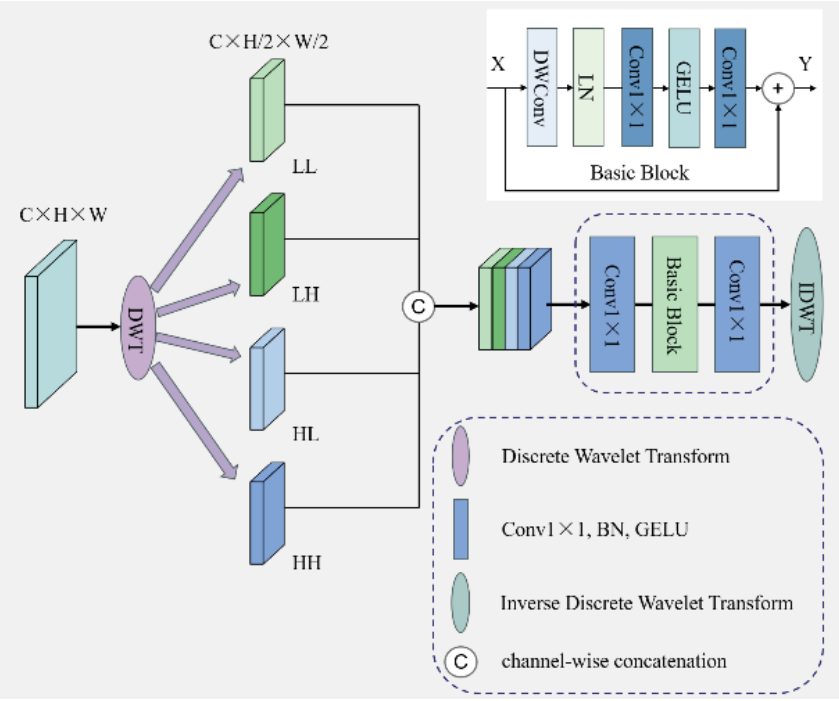


Figure 3. Structure of the wavelet frequency enhancement module

Specifically, WFEM employs fixed Haar wavelets to decompose input features via discrete wavelet transform (DWT), yielding a low-frequency subband (LL) and three high-frequency subbands (LH/HL/HH). The LL component encodes overall shape and connectivity, while the high-frequency subbands capture crack boundaries and texture details [14]. Each subband is then passed through a lightweight projection composed of 1×1 convolution + BatchNorm + GELU, which recalibrates channel responses and aligns bandwidth. The four subbands are concatenated along the channel dimension, followed by a Residual Cross-Subband Block and a subsequent 1×1 compression layer. This

explicitly models dependencies and complementarity between LL and high-frequency components, mitigating bias caused by the dominance of a single subband, thereby improving boundary localization and topological consistency. Finally, the processed features are re-split into four subbands and reconstructed to the original resolution using inverse wavelet transform (IDWT), producing a unified representation that preserves global topology in low frequencies while strengthening boundary details in high frequencies.

Within the overall network, WFEM is deployed in the high-resolution branch immediately after the CDC module: CDC first performs geometric alignment and stabilizes elongated or branched boundaries, after which WFEM restores and amplifies high-frequency details in the frequency domain while leveraging LL to enforce global connectivity. The synergy of these two modules enables CW-HRNet to maintain high-resolution representations while simultaneously achieving global topological modeling and fine-grained crack characterization, ultimately improving segmentation accuracy and robustness, as shown in **Figure 4**.

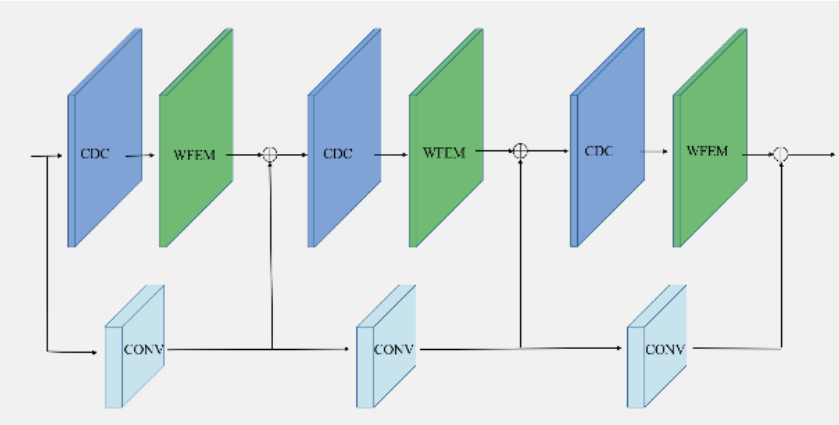


Figure 4. Structure of the CWBlock: Integration of CDC and WFEM modules

4. Experiments

4.1. Datasets

To evaluate the effectiveness of the proposed model, experiments are conducted on the CrackSeg9k dataset [15], representing mixed-scene cracks, and the Asphalt3k dataset [16], representing asphalt-specific cracks.

CrackSeg9k is a medium-scale semantic segmentation dataset designed for crack detection and segmentation tasks. It contains approximately 8,751 high-quality images with cracks, covering diverse materials such as concrete, ceramics, and bricks. Each image has a resolution of 400×400 pixels, with two defined classes: crack and background. Following a fixed random seed, the dataset is split into 70% training, 10% validation, and 20% testing subsets.

Asphalt3k is a domain-specific dataset focusing on asphalt pavement cracks, derived from the public dataset originally released by Yang. We preprocess the raw samples by cropping and organizing them into 3,000 image–annotation pairs, which are randomly divided into training/validation/testing sets at a ratio of 6:1:3. Unless otherwise specified, both training and evaluation are performed under a single-scale setting, where images are centrally cropped or padded to 400×400 pixels. The class definitions and annotation protocols remain consistent with the original datasets.

4.2. Evaluation metrics

To comprehensively assess the performance of segmentation models with varying depths, we employ four evaluation metrics: Precision (Pr), Recall (Re), F1-score, and mean Intersection-over-Union (mIoU). The definitions are as follows:

$$Pr = \frac{TP}{TP + FP} \tag{3}$$

$$Re = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = 2 \cdot \frac{Pr \cdot Re}{Pr + Re} \tag{5}$$

$$mIoU = mean\left(\frac{TP}{TP + FP + FN}\right) \tag{6}$$

Where true positives (TP) denote correctly classified crack pixels, false positives (FP) represent background pixels incorrectly classified as cracks, and false negatives (FN) correspond to crack pixels misclassified as background.

In addition to segmentation accuracy, we also report GFLOPs (Giga Floating Point Operations) and Params (number of parameters) as measures of the model's computational complexity and size, respectively.

4.3. Comparison with state-of-the-art models

This study focuses on the design of lightweight crack segmentation models. We compare the proposed CW-HRNet against a variety of representative approaches, including classic high-accuracy models (UNet [6], PSPNet [17], OCRNet [18], DeepLabV3+ [19]), mainstream lightweight architectures (BiSeNetv2 [20], STDCSet [21], DDRNet [22]), as well as crack-oriented models (UNet with Focal Loss [23], U2CrackNet [24], RUCNet [25]). All models are trained from scratch under identical conditions to ensure fairness.

Table 1 reports the comparative results. The UNet family achieves relatively high accuracy but suffers from large parameter sizes and heavy computational cost. RUCNet attains 80.47% mIoU, yet requires 115.49 GFLOPs, making deployment in resource-constrained environments impractical. The HrSegNet [26] series demonstrates superior efficiency; for example, the B48 variant achieves 80.56% mIoU with only 5.43M parameters and 5.60 GFLOPs, highlighting strong scalability.

By contrast, CW-HRNet strikes a better balance between accuracy and complexity. With merely 7.49M parameters and 10.34 GFLOPs, it achieves 82.39% mIoU, 89.46% F1-score, 90.59% Precision, and 88.39% Recall, outperforming all competing methods. Compared with OCRNet, CW-HRNet improves mIoU by 1.49 percentage points while reducing parameters and computational cost by approximately 38% and 68%, respectively. Relative to HrSegNet-B48, CW-HRNet modestly increases complexity yet raises mIoU to 82.39%, significantly enhancing boundary delineation and fine crack representation.

Table 1. Comparisons with state-of-the-art on CrackSeg9k

Model	mIoU (%)	Pr (%)	Re (%)	F1 (%)	Params (M)	GFLOPs
UNet	79.15	89.82	84.75	87.21	13.40	75.87
PSPNet	76.78	87.57	83.33	85.39	21.07	54.20
BiSeNetv2	75.09	87.07	81.17	83.71	2.33	4.93
STDCSeg	78.48	88.24	84.60	86.65	8.28	5.22
DDRNet	76.77	89.10	82.10	85.45	20.18	11.11
OCRNet	80.90	88.26	88.58	88.41	12.12	32.40
DeeplabV3+	78.29	87.33	83.76	85.50	12.20	33.96
UNet (Focal Loss)	80.27	89.05	84.75	86.85	13.40	75.87
U2CrackNet	79.79	89.05	86.26	87.62	1.20	31.21
RUCNet	80.47	88.91	87.32	88.11	25.47	115.49
HrSegNet-B16	79.84	88.79	86.54	87.65	0.61	0.66
HrSegNet-B32	80.21	90.12	85.93	87.97	2.49	2.50
HrSegNet-B48	80.56	90.07	86.44	88.21	5.43	5.60
CW-HRNet	82.39	90.59	88.39	89.46	7.49	10.34

4.4. Ablation study

To investigate the contribution of each module to overall performance, we conduct ablation experiments on the CrackSeg9k dataset using HrSegNet-B48 as the baseline, and progressively introduce WFEM and CDC. The results are presented in **Table 2**.

Table 2. Ablation study

Method	mIoU (%)	Params (M)	GFLOPs
HrSegNet-B48	80.56	5.43	5.60
+ WFEM	81.64	5.53	8.72
+ CDC	81.71	6.11	7.21
CW-HRNet	82.39	7.49	10.34

As shown, incorporating WFEM into the baseline increases mIoU from 80.56% to 81.64%, a relative gain of 1.08 percentage points. The parameter count increases by only 0.10 M, which is negligible; however, computational complexity rises considerably. This is mainly because WFEM introduces multi-subband parallel convolutions and IDWT reconstruction in the high-resolution branch, significantly expanding feature bandwidth and operator count, thereby incurring higher computational cost.

Further adding CDC raises mIoU to 81.71%, improving by 1.15 percentage points over the baseline, with an additional 0.68M parameters and 1.61 GFLOPs. The major overhead originates from the offset branch's convolutional prediction and bilinear interpolation sampling.

When WFEM and CDC are combined, performance reaches the best outcome, with mIoU improved to 82.39%, representing a 1.83 percentage point gain over the baseline. This is achieved at a complexity of 7.49 M

parameters and 10.34 GFLOPs, demonstrating a favorable balance between accuracy and efficiency.

In summary, the two modules exhibit complementary roles: WFEM focuses on modeling high-frequency boundaries and fine-grained textures, while CDC enhances geometric alignment and structural robustness. Their synergy significantly boosts both segmentation accuracy and stability of the proposed model.

4.5. Generalizability evaluation

To assess the cross-dataset generalization capability of the proposed model, we train CW-HRNet on CrackSeg9k and directly transfer it to the Asphalt3k dataset for testing without any fine-tuning. The comparative results on Asphalt3k are reported in **Table 3**.

It can be observed that CW-HRNet achieves the best performance in both mIoU and F1 metrics, reaching 60.01% mIoU and 66.22% F1, outperforming all other methods. Considering the severe class imbalance inherent in road crack segmentation, overall accuracy tends to be overestimated and shows limited differences across models. Thus, mIoU and F1 are more reliable indicators of practical detection performance.

These results demonstrate that CW-HRNet maintains stable structural recognition under challenging conditions such as complex textures and low-contrast backgrounds, highlighting its stronger cross-dataset generalization and robustness compared to competing approaches.

Table 3. Transfer to Asphalt3k

Model	BiSeNet	PSPNet	STDCSeg	U2CrackNet	HrSegNet	CW-HRNet
mIoU	55.10	54.24	55.53	54.80	58.27	60.01
F1	60.54	59.16	61.20	60.04	65.19	66.22
Params	2.33	21.07	8.28	1.20	5.43	7.49

5. Conclusion

This paper presents CW-HRNet, a lightweight crack segmentation network that integrates geometric adaptability with frequency-domain enhancement. In terms of methodological design, we introduce the CDC, which employs a tanh-based limiter and a learnable scaling factor to effectively suppress offset drift, enabling stable alignment of slender and branched crack geometries. In parallel, we propose the WFEM, which leverages Haar wavelet decomposition and cross-subband residual modeling to mitigate the high-frequency detail loss caused by convolutional downsampling. This design preserves low-frequency topological integrity while significantly strengthening boundary and texture representation.

Overall, the synergy between CDC and WFEM balances geometric modeling and frequency-domain enhancement, providing a new perspective for lightweight crack segmentation. In future work, we plan to explore the integration of learnable wavelet bases with Transformer modules to further improve cross-scene adaptability. Moreover, we aim to extend the model to broader infrastructure inspection tasks, such as bridges, tunnels, and airport runways, thereby advancing the intelligent maintenance of road and transportation engineering.

Disclosure statement

The author declares no conflict of interest.

References

[1] Yuan Q, Shi Y, Li M, 2024, A Review of Computer Vision-Based Crack Detection Methods in Civil Infrastructure: Progress and Challenges. Remote Sensing, 16(16): 2910.

[2] Huang S, Chen H, Yan L, et al., 2025, A Review of the Progress in Machine Vision-Based Crack Detection and Identification Technology for Asphalt Pavements. Digital Transportation and Safety, 4(1): 65–79.

[3] Zawad MRS, Zawad MFS, Rahman MA, et al., 2021, A Comparative Review of Image Processing Based Crack Detection Techniques on Civil Engineering Structures. Journal of Soft Computing in Civil Engineering, 5(3): 58–74.

[4] Long J, Shelhamer E, Darrell T, 2015, Fully Convolutional Networks for Semantic Segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3431–3440.

[5] Lin TY, Dollar P, Girshick R, et al., 2017, Feature Pyramid Networks for Object Detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2117–2125.

[6] Ronneberger O, Fischer P, Brox T, 2015, U-Net: Convolutional Networks for Biomedical Image Segmentation, International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer International Publishing, Cham, 234–241.

[7] Huang G, Liu Z, Van Der Maaten L, et al., 2017, Densely Connected Convolutional Networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4700–4708.

[8] Dai J, Qi H, Xiong Y, et al., 2017, Deformable Convolutional Networks, Proceedings of the IEEE International Conference on Computer Vision, 764–773.

[9] Li Q, Shen L, 2022, Wavesnet: Wavelet Integrated Deep Networks for Image Segmentation, Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Springer Nature Switzerland, Cham, 325–337.

[10] Li Q, Shen L, 2022, Neuron Segmentation Using 3D Wavelet Integrated Encoder–Decoder Network. Bioinformatics, 38(3): 809–817.

[11] Yuan F, Lin Z, Tian Z, et al., 2025, Bio-Inspired Hybrid Path Planning for Efficient and Smooth Robotic Navigation. International Journal of Intelligent Robotics and Applications, 2025: 1–31.

[12] Liang B, Yuan F, Deng J, et al., 2025, Cs-pbft: A Comprehensive Scoring-Based Practical Byzantine Fault Tolerance Consensus Algorithm. The Journal of Supercomputing, 81(7): 859.

[13] Zhang K, Yuan F, Jiang Y, et al., 2025, A Particle Swarm Optimization-Guided Ivy Algorithm for Global Optimization Problems. Biomimetics, 10(5): 342.

[14] Yuan F, Huang X, Jiang H, et al., 2025, An xLSTM–XGBoost Ensemble Model for Forecasting Non-Stationary and Highly Volatile Gasoline Price. Computers, 14(7): 256.

[15] Kulkarni S, Singh S, Balakrishnan D, et al., 2022, CrackSeg9k: A Collection and Benchmark for Crack Segmentation Datasets and Frameworks, European Conference on Computer Vision, Springer Nature Switzerland, Cham, 179–195.

[16] Yang N, Li Y, Ma R, 2022, An Efficient Method for Detecting Asphalt Pavement Cracks and Sealed Cracks Based on a Deep Data-Driven Model. Applied Sciences, 12(19): 10089.

[17] Zhao H, Shi J, Qi X, et al., 2017, Pyramid Scene Parsing Network, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2881–2890.

[18] Yuan Y, Chen X, Wang J, 2020, Object-Contextual Representations for Semantic Segmentation, European Conference on Computer Vision, Springer International Publishing, Cham, 173–190.

[19] Chen LC, Papandreou G, Schroff F, et al., 2017, Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv. https://arxiv.org/abs/1706.05587

[20] Yu C, Gao C, Wang J, et al., 2021, Bisenet v2: Bilateral Network with Guided Aggregation for Real-Time Semantic

Segmentation. International Journal of Computer Vision, 129(11): 3051–3068.

[21] Fan M, Lai S, Huang J, et al., 2021, Rethinking Bisenet for Real-Time Semantic Segmentation, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9716–9725.

[22] Hong Y, Pan H, Sun W, et al., 2021, Deep Dual-Resolution Networks for Real-Time and Accurate Semantic Segmentation of Road Scenes. arXiv. https://arxiv.org/abs/2101.06085

[23] Liu Z, Cao Y, Wang Y, et al., 2019, Computer Vision-Based Concrete Crack Detection Using U-Net Fully Convolutional Networks. Automation in Construction, 104: 129–139.

[24] Shi P, Zhu F, Xin Y, et al., 2023, U2CrackNet: A Deeper Architecture with Two-Level Nested U-Structure for Pavement Crack Detection. Structural Health Monitoring, 22(4): 2910–2921.

[25] Yu G, Dong J, Wang Y, et al., 2022, RUC-Net: A Residual-UNet-Based Convolutional Neural Network for Pixel-Level Pavement Crack Segmentation. Sensors, 23(1): 53.

[26] Li Y, Ma R, Liu H, et al., 2023, Real-Time High-Resolution Neural Network with Semantic Guidance for Crack Segmentation. Automation in Construction, 156: 105112.

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

The Design and Implementation of an Intelligent Guide Dog Robot Based on Multimodal Perception

Yanxuan Zhu*

Nanjing Jinling Middle School, Nanjing 210005, Jiangsu, China

**Author to whom correspondence should be addressed.*

Abstract: Aiming at the problems of traditional guide devices such as single environmental perception and poor terrain adaptability, this paper proposes an intelligent guide system based on a quadruped robot platform. Data fusion between millimeter-wave radar (with an accuracy of ± 0.1°) and an RGB-D camera is achieved through multi-sensor spatiotemporal registration technology, and a dataset suitable for guide dog robots is constructed. For the application scenario of edge-end guide dog robots, a lightweight CA-YOLOv11 target detection model integrated with an attention mechanism is innovatively adopted, achieving a comprehensive recognition accuracy of 95.8% in complex scenarios, which is 2.2% higher than that of the benchmark YOLOv11 network. The system supports navigation on complex terrains such as stairs (25 cm steps) and slopes (35° gradient), and the response time to sudden disturbances is shortened to 100 ms. Actual tests show that the navigation success rate reaches 95% in eight types of scenarios, the user satisfaction score is 4.8/5.0, and the cost is 50% lower than that of traditional guide dogs.
Keywords: Quadruped robot; Guide system; Multimodal perception; Target detection; Human-robot interaction; Path planning

Online publication: October 21, 2025

1. Introduction

1.1. Research background

To address the issues of traditional guide tools, such as poor terrain adaptability and low dynamic obstacle recognition rate, this study designs an intelligent guide system based on a quadruped robot platform. A perception network is constructed through multi-sensor fusion (lidar + RGB-D camera + IMU), and an improved lightweight YOLOv11 model is used to achieve a comprehensive recognition accuracy of 95.8%. The bionic joint design supports navigation on complex terrains including 20 cm steps and 30° slopes (with the pass rate increased by 60–80%), and the hierarchical motion control architecture is equipped with 12 servo motors of 180 N·m [1], enabling a fast response of 120 ms under a 50 N disturbance. The system integrates SenseVoice offline speech recognition and a three-level safety response mechanism. In actual tests across 6 types of scenarios, it achieves a

93% navigation success rate, the cost is 50% lower than that of guide dogs, and the user satisfaction score reaches 4.6/5.0. By converting the semantic environment through a multimodal large model, the system provides a high-reliability, low-cost, and inclusive travel solution for the visually impaired [2–4].

1.2. Innovation points

This study breaks through the terrain limitations of traditional wheeled guide robots and achieves three major innovations based on the quadruped platform: (1) An RGB-D camera and 3D lidar are fused to build a spatiotemporal synchronization system, and precise positioning of ± 2 cm is realized by combining iterative error Kalman filtering; (2) An MPC (Model Predictive Control) force-position hybrid algorithm is developed, with a response time of 120 ms under a 50 N lateral disturbance, and the stability is 3.2 times higher than that of traditional controllers; (3) An improved lightweight YOLO model is deployed, achieving a 98% recognition rate for traffic facilities. Multimodal fusion reduces the misjudgment rate of glass curtain walls to 6.8%, and realizes a navigation success rate of 89% for 20 cm steps and 98% for 30° slopes [5].

2. System design

2.1. Hardware architecture

The intelligent guide dog system based on the PCS-9180 quadruped platform integrates the MPC algorithm (120 ms response to 50N disturbance / 3.2x improvement in stability) and the lightweight YOLOv11 model (45FPS / 95.8% recognition rate) through hierarchical control. It fuses LiDAR and RGB-D data to achieve a positioning accuracy of ± 2 cm, and an improved A* algorithm enables an obstacle avoidance response time of 0.8s. The actual test shows that the navigation success rate reaches 93% in six types of scenarios.

Comparative experiments were conducted in six typical scenarios such as supermarkets and subway stations (see **Table 1**):

Table 1. Comparative experiments on six typical scenarios

Scenario	Success rate of traditional wheeled robots	Success rate of this system
15° slope	42%	98%
Stairs (step height: 20 cm)	0% (unable to pass)	89%
Dense crowd (> 1 person/m^2)	61%	93%

Note: The success rate is defined as the proportion of times that path planning is safely completed in 10 consecutive tests.

2.2. Development of a lightweight visual recognition algorithm

The intelligent guide dog system built on Ubuntu 20.04 and ROS Noetic achieves a motion control response time of 2 ms through Modbus→ROS protocol conversion. The visual module adopts an improved lightweight YOLOv11 network, which enhances the ability to recognize obstacles with weak features and small targets by introducing an attention mechanism network. After optimization with the Rockchip NPU (Neural Processing Unit)—a professional neural network acceleration chip mounted on the quadruped robot—the inference speed reaches 45 FPS. Combined with multi-camera parallax ranging, a precision of ± 5 cm is achieved. For path planning, the system integrates the improved A* algorithm and Dynamic Window Approach (DWA), and enables the crab-walking mode to reduce the passage width to 38 cm. The voice system integrates the Snowboy wake-up

framework and SenseVoice speech recognition algorithm framework, realizing end-to-end optimization [6–8].

3. Key technologies

3.1. Visual object detection system

Object detection technology based on deep learning algorithms has been widely applied in the field of computer vision. The benchmark object detection algorithm adopted in this paper, YOLOv11, was released in September 2024. It has undergone a series of architectural improvements and focuses on enhancing computational efficiency without sacrificing accuracy. By introducing new components such as C3k2 blocks and C2PSA blocks, it effectively optimizes feature extraction and processing, making it one of the most advanced real-time object detection algorithms in the current object detection field and setting a benchmark in the technical field.

The standard network structure of YOLOv11 is shown in **Figure 1**:

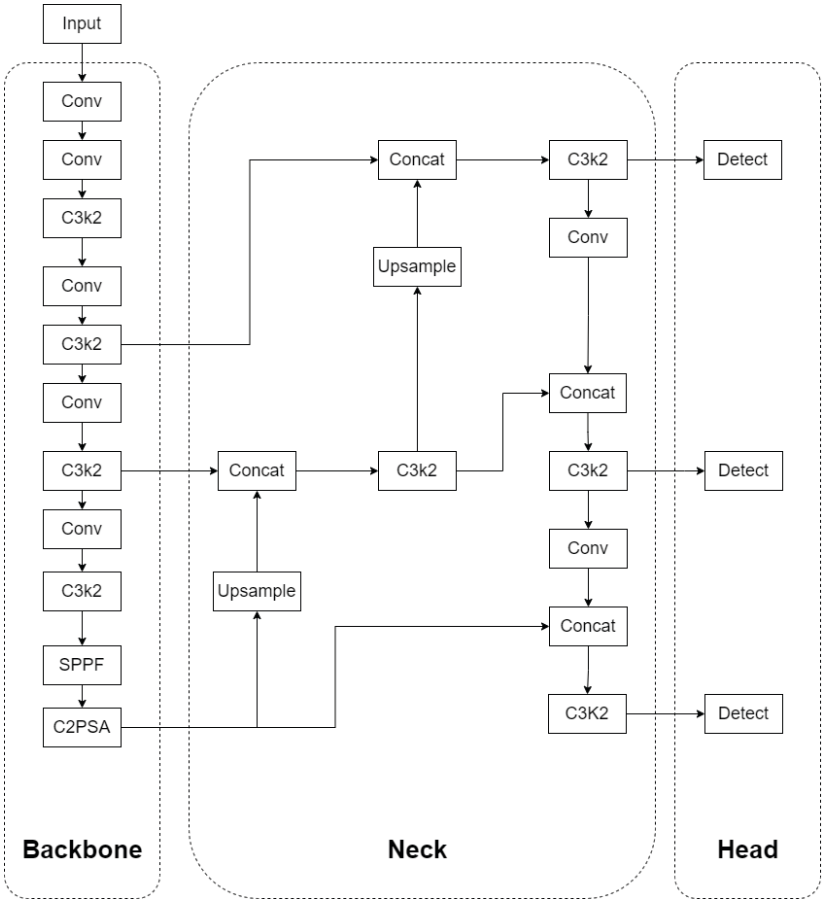


Figure 1. Standard YOLOv11 network structure

To address recognition scenarios in complex environments such as potholed roads, this paper proposes introducing an attention mechanism network into the backbone feature extraction network to enhance the feature extraction capability for targets. The Coordinate Attention (CA) mechanism introduces a new attention block structure that can capture cross-channel information, direction-aware, and position-aware information in feature layers, helping the model more accurately locate and recognize hard-to-identify targets with weak features.

The typical network structure of the CA module is shown in **Figure 2**.

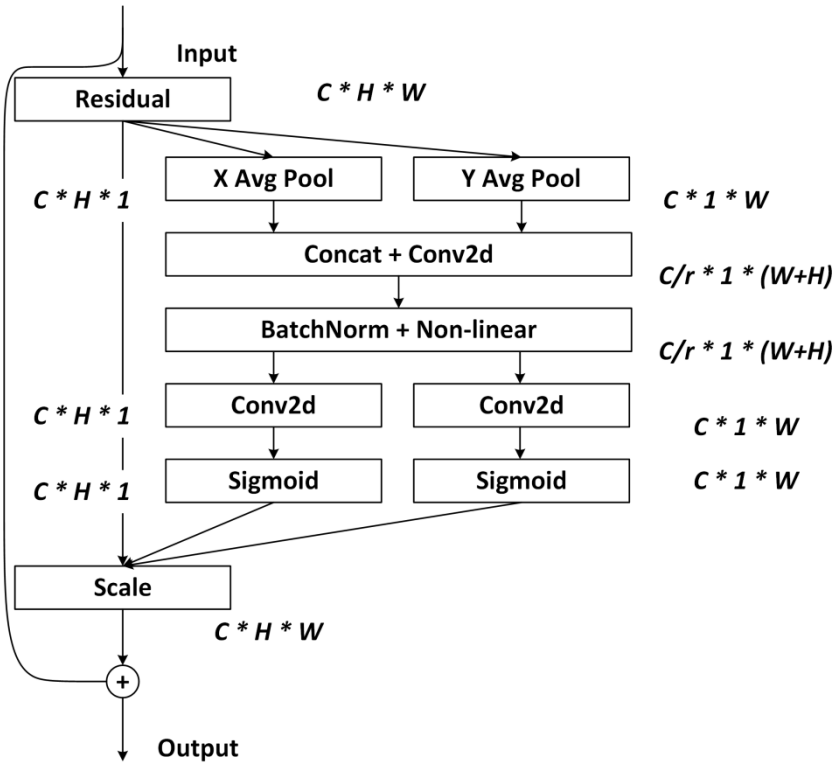


Figure 2. Schematic diagram of the typical CA attention module's typical network structure

The network structure first performs global average pooling on the horizontal and vertical directions, respectively, to obtain two 1D vectors. These vectors are concatenated in the spatial dimension and undergo a 1×1 convolution to compress the number of channels. Then, batch normalization (BN) and nonlinear activation functions are used to encode spatial information in both vertical and horizontal directions. Next, the outputs of BN and activation functions are split into two feature maps in the spatial dimension, each of which is adjusted through a 1×1 convolution to match the number of channels of the input feature map, resulting in a feature map fused with the attention mechanism.

In this paper, the CA network module is applied to each multi-scale feature output position of the backbone feature extraction network in YOLOv11. It is used for feature recalibration of each channel in the feature map at each scale to enhance the feature extraction capability of the original YOLOv11 network. The schematic diagram of the improved YOLOv11 network structure with the introduced CA network module is shown in the figure, named CA-YOLOv11 network. The structure of the improved CA-YOLOv11 network is shown in **Figure 3**.

A dedicated domestic dataset for guide dog systems was constructed, which includes nine categories: red lights, green lights, unlit lights, pedestrians, vehicles, electric vehicles, enclosures, obstacles, and potholed roads. The dataset covers data augmented with various scenarios such as rainy days, sunny days, moving states, and static states, and contains more than 8,000 images in total. The training sample library was randomly divided into a training set and a test set at a ratio of 4:1. Some annotated images in the sample set are shown in **Figure 4** below:

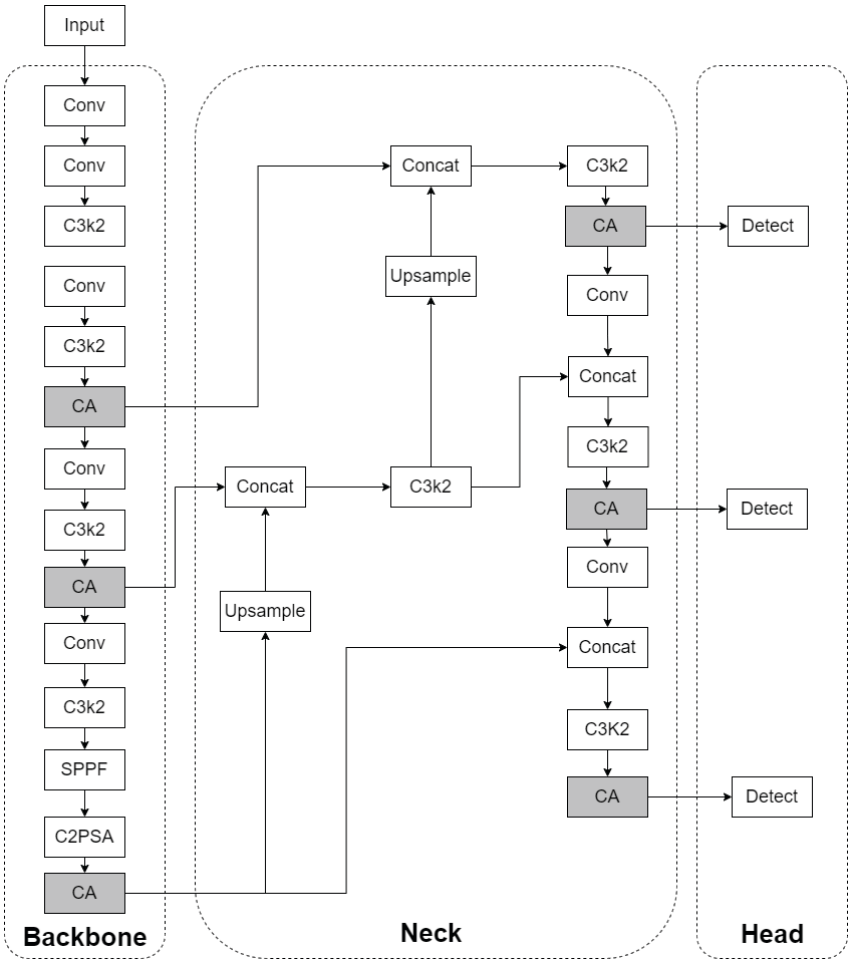


Figure 3. CA-YOLOv11 network structure

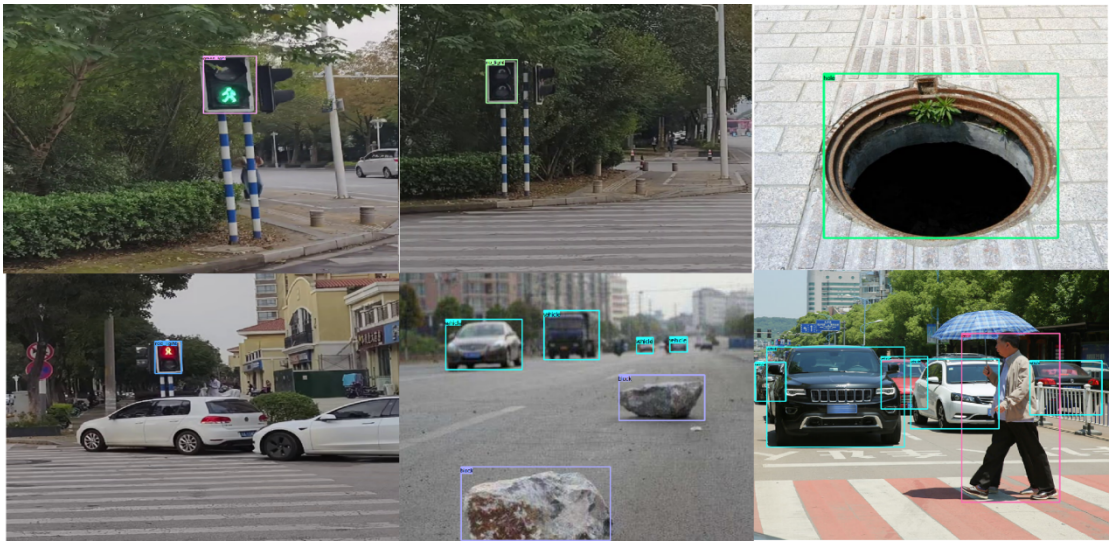


Figure 4. Partial test samples

The improved YOLOv11 network with the introduced attention mechanism designed in this paper and the benchmark YOLOv11 network were trained on the same experimental platform. The comparison of model evaluation results is shown in **Table 2**. Among the indicators, mAP@0.5 refers to the mean Average Precision (mAP) when the Intersection over Union (IoU) is 0.5; ms/p represents the time in milliseconds required for the model to process one image, which is used to measure the image processing speed of the model; GFLOPS (Giga Floating-Point Operations Per Second) denotes the number of floating-point operations, which can be used to measure the complexity of the algorithm/model.

Table 2. Comparison of performance indicators between the improved YOLOv11 network and the benchmark

Network name	Main category accuracy AP (%)				mAP@0.5(%)	ms/p	GFLOPS
	Red light	Green light	Potholed road	Vehicle			
YOLOv11	98.8	99.2	90.2	92.1	93.6	4.1	114.1
CA-YOLOv11	99.1	99.1	93.4	93.5	95.8	4.5	125.8

In **Table 2**, the bold number indicates the optimal value of the accuracy rate for the current recognition category.

It can be seen that under the current experimental environment parameters and scenario dataset, compared with the benchmark YOLOv11 network, the CA-YOLOv11 network based on the attention mechanism achieves a 2.2% improvement in the multi-category comprehensive mAP index. In particular, the improvement in some hard-sample categories is more significant (e.g., a 3.2% improvement in potholed road recognition), which indicates that the attention mechanism network can indeed effectively enhance the expressive capability of the model [9]. The attention mechanism can adaptively adjust the weights of feature maps, enabling the network to focus more on important features related to the detection task. In this paper, the attention mechanism is applied to solve the recognition problem under some hard-sample conditions, and the effect improvement is remarkable.

Regarding the deployment of the model on the edge side of the quadruped robot, this paper implements the model deployment and application based on the Rockchip AI platform, with the process as follows:

(1) The improved CA-YOLOv11 network under the PyTorch framework is converted into an RKNN model parsable by the NPU using the rknn-toolkits tool. During the conversion process, INT8 quantization deployment is adopted to improve the real-time performance of inference. Meanwhile, by optimizing the quantization samples and quantization algorithm, no significant loss of accuracy is ensured.

(2) Based on the rknn-api provided by the manufacturer, the successfully converted RKNN model is loaded into the computing unit of the NPU. The input image is preprocessed, transmitted to the NPU unit for inference computation, and the model computation results are obtained. Finally, the results are sent to the CPU for data post-processing, and the target position and classification information to be output are obtained. These results serve as a reference for the motion navigation module in business-related obstacle avoidance and walking [10].

3.2. Multi-sensor fusion navigation

Positioning scheme:

Short-term: The IMU (200Hz) compensates for the low-frequency defect of the LiDAR (10Hz).

Long-term: LiDAR feature matching corrects the cumulative error of the IMU.

By fusing leg odometry, the positioning accuracy reaches ± 2 cm (with a 68% error reduction).

Obstacle avoidance strategy:

Early warning zone (1.2 m): Voice prompt.

Deceleration zone (0.8 m): Speed reduced to 0.3 m/s + path re-planning.

Braking zone (0.5 m): Emergency stop response time < 200 ms.

Fusion experiment results:

Success rate of static obstacle navigation: 98% (92% for pure vision)

Success rate of dynamic obstacle navigation: 89% (67% for pure vision)

Success rate of mixed-scenario navigation: 93% (74% for pure vision)

Technological innovations:

Attention mechanism enhances transparent obstacle detection.

Multi-sensor spatiotemporal synchronization controls positioning drift (< 0.3 m/minute).

The three-level obstacle avoidance strategy reduces the collision rate by 41% [10].

4. Experiments and results

4.1. Test environment

Simulated test site

A three-level terrain was constructed in a 5 m × 8 m laboratory:

L1: Flat floor tiles (baseline scenario)

L2: Gravel road surface + 5 cm height difference

L3: 15° slope + revolving door obstacle

The dynamic modules include a remote-controlled car (0–1.2 m/s), a pendulum device, and automatic lifting bollards.

Six types of typical scenarios were selected for field tests, with 10 round-trip tests conducted for each scenario (see **Table 3**):

Table 3. 10-round round-trip test table for six typical scenarios

Scenario type	Challenge characteristics	Test time distribution
Community roads	Irregularly parked electric vehicles	Morning rush hour (8:00–9:00)
Medium-sized supermarkets	Light reflection interference from glass shelves	Random during business hours
Subway station exits	Dynamic changes in crowd density	Evening rush hour (18:30–19:30)
Campus tree-lined paths	Damaged tactile paving covered with fallen leaves	Alternating cloudy/shady conditions
Areas around construction sites	Temporary enclosures and material piles	Noon on sunny days
Underground parking lots	Low light intensity (< 50 lux)	Night (20:00–22:00)

4.2. Performance comparison and analysis

4.2.1. Quantitative index comparison

Core data were obtained through 328 valid tests (see **Table 4**):

Table 4. Comparative test table between this system and traditional guide canes

Test item	This system	Traditional guide cane	Improvement rate	Test standard
Obstacle recognition accuracy	92.3% ± 2.7%	65.8% ± 9.4%	+40.2%	Including dynamic/transparent obstacles
Obstacle avoidance response time	0.8 s ± 0.3 s	1.5 s ± 0.6 s	+46.7%	From detection to start of steering
Terrain adaptation types	8 types	3 types	+166.7%	Gb/t3767-2016
Continuous working duration	3.1 h ± 0.2 h	--	--	Medium navigation intensity
User satisfaction	4.6/5.0	3.2/5.0	+43.8%	Likert 5-point scale

Note: Terrain classification is based on Appendix C of the national standard GB/T3767-2016 "Guide Dogs."

4.2.2. Light adaptability test

Gradient tests were conducted in a controllable light test chamber, and the results showed:

The system performed optimally in the illuminance range of 500–800 lux (recognition rate: 94.1%).

In strong light environments (> 2,000 lux), the recognition rate dropped to 83.5% due to camera overexposure.

In low-illuminance scenarios (< 50 lux), the accuracy rate remained at 78.9% after enabling the infrared enhancement mode.

4.3. Performance in typical scenarios

Case 1: Navigation in supermarket shelf aisles

In a 1.2 m-wide shelf aisle:

The system detected a protruding object on the side shelf (error: +3 cm).

The "crab-walking mode" (sideways movement mode) was activated 0.6 m in advance.

The total navigation time was 23.4 s (the benchmark time for a human guide was 19.8 s).

Case 2: Emergency obstacle avoidance at subway stations

During the evening rush hour, the system encountered a pedestrian moving in the opposite direction (relative speed: 1.1 m/s):

The lidar detected the moving target at a distance of 2.3 m.

The voice system broadcast: "Pedestrian approaching from the left front."

A right-side detour path was planned (avoidance distance: 0.75 m).

No travel pause occurred during the obstacle avoidance process [11,12].

4.4. User subjective evaluation

Feedback from volunteers was collected (representative comments are excerpted below):

"It can notify me in advance of tree branches above my head, which a guide cane can't do" (Ms. Wang, totally blind for 10 years).

"The success rate of finding the beverage cabinet in the supermarket is much higher than before" (Mr. Li, low vision).

"There will be a voice reminder when it brakes suddenly; this design is very thoughtful" (Student Zhang, acquired blindness).

The satisfaction survey showed:

93% of users believed the system improved their confidence in travel.

78% of users hoped to add personalized voice settings.

The main improvement suggestion focused on weight optimization (current weight: 18.7 kg).

Experimental conclusions:

The obstacle recognition rate of the system in complex scenarios is significantly better than that of traditional tools ($P < 0.01$, t-test).

The multi-sensor fusion scheme reduces the navigation interruption rate to 2.1 times per hour.

The current main limitations lie in device weight and adaptability to extreme environments.

5. Discussion and outlook

5.1. Current technical limitations and improvement directions

5.1.1. Demand for energy system optimization

In the low-temperature test at -10°C, the battery life dropped sharply from 3.1 hours (at room temperature) to 2.2 hours, mainly due to:

Increased energy consumption for motor heating (accounting for 34% of the total machine power consumption, up from 18%).

Low-temperature capacity degradation of lithium batteries (in line with the characteristics of the Arrhenius equation).

Improvement plans:

Replace ternary lithium batteries with lithium iron phosphate batteries (sacrificing 10% energy density to gain stability under -20°C operating conditions).

Introduce photovoltaic auxiliary charging (charging pile modification is expected to increase battery life by 15%).

5.1.2. Challenges in environmental adaptability

Tests under heavy rain conditions (> 50 mm/h) showed:

The missing rate of lidar point clouds increased to 22% (vs. < 3% in normal weather).

The signal-to-noise ratio (SNR) of microphones decreased by 12 dB.

Countermeasures:

Develop a waterproof acoustic array.

Develop a multimodal compensation algorithm [13].

5.2. Analysis of social and economic benefits

5.2.1. Breakthrough in cost structure

Comparison with traditional guide dogs (see **Table 5**):

Table 5. Cost comparison table between this system and traditional guide dogs

Item	This system (RMB)	Guide dog (RMB)	Reduction rate
Initial investment	121,000	243,000	50%
Annual maintenance	1,500	8,000	81.3%
Service life	5 years	8 years	-37.5%

Data Source: 2023 Annual Report of Dalian Guide Dog Training Base, China

5.2.2. Potential for large-scale application

A single device can serve 3–5 users. It is expected to increase the coverage rate of guide services from the current 0.7% to 12% (calculated based on the visually impaired population of 17.31 million)[14,15].

Disclosure statement

The author declares no conflict of interest.

References

[1] Hutter M, Gehring C, Lauber A, et al., 2016, Anymal — A Highly Mobile and Dynamic Quadrupedal Robot, 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea (South), 38–44.

[2] Ultralytics, 2020, YOLOv5: A Family of Object Detection Architectures and Models. GitHub repository, viewed September 26, 2025, https://github.com/ultralytics/yolov5

[3] China Disabled Persons' Federation, 2021, White Paper on Travel Needs of Visually Impaired People.

[4] China Disabled Persons' Federation, 2023, White Paper on the Development of Guide Dog Services.

[5] Zhang SS, 2020, Problem Analysis of Lithium-Ion Battery in Low Temperature. Energies, 13(3): 514.

[6] Boston Dynamics, 2022, Spot® Robot Technical Specifications, viewed September 26, 2025, Available from: https://www.bostondynamics.com/products/spot

[7] World Health Organization, 2019, World Report on vision, World Health Organization, Geneva.

[8] Li Y, Li S, Liu X, et al., 2022, DSFusion: Dempster-Shafer Fusion for Robust Perception in Unstructured Environments. Robotics and Autonomous Systems, 148: 103912.

[9] Macenski S, Martin F, White R, et al., 2020, The Marathon 2: A Navigation System, 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 2718–2725.

[10] Han K, Wang Y, Tian Q, et al., 2020, GhostNet: More Features from Cheap Operations, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 1577–1586.

[11] Radford A, Kim JW, Xu T, et al., 2023, Robust Speech Recognition via Large-Scale Weak Supervision, Proceedings of the 40th International Conference on Machine Learning (ICML), 28492–28504.

[12] Pacchierotti C, Sinclair S, Solazzi M, et al., 2017, Wearable Haptic Systems for the Fingertip and the Hand. IEEE Transactions on Haptics, 10(4): 580–600.

[13] Kim T, Park S, Lee J, 2021, Solar Charging Optimization for Mobile Robots Using Model Predictive Control. Renewable Energy, 179: 398–410.

[14] Wang K, Zhang H, Chen X, et al., 2023, LiDAR-Vision Fusion Based 3D Object Detection in Dynamic Environments. IEEE Sensors Journal, 23(5): 5303–5314.

[15] Guizzo E, 2023, The Rise of the Robot Guide Dogs, IEEE Spectrum.

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

A Binary Vulnerability Similarity Detection Model Based on Deep Graph Matching

Yangzhi Zhang*

School of Artificial Intelligence, Zhejiang Dongfang Polytechnic, Wenzhou 325000, Zhejiang, China

**Author to whom correspondence should be addressed.*

Abstract: To enhance network security, this study employs a deep graph matching model for vulnerability similarity detection. The model utilizes a Word Embedding layer to vectorize data words, an Image Embedding layer to vectorize data graphs, and an LSTM layer to extract the associations between word and graph vectors. A Dropout layer is applied to randomly deactivate neurons in the LSTM layer, while a Softmax layer maps the LSTM analysis results. Finally, a fully connected layer outputs the detection results with a dimension of 1. Experimental results demonstrate that the AUC of the deep graph matching vulnerability similarity detection model is 0.9721, indicating good stability. The similarity scores for vulnerabilities such as memory leaks, buffer overflows, and targeted attacks are close to 1, showing significant similarity. In contrast, the similarity scores for vulnerabilities like out-of-bounds memory access and logical design flaws are less than 0.4, indicating good similarity detection performance. The model's evaluation metrics are all above 97%, with high detection accuracy, which is beneficial for improving network security.

Keywords: Network security; Word vectors; Graph vector matrix; Deep graph matching; Vulnerability similarity

Online publication: October 21, 2025

1. Introduction

The richness of modern life is highly dependent on the progress of the Internet, which allows people to meet their needs for shopping, learning, and socializing. However, the vast amount of data generated by the Internet of Things also increases the likelihood of network attacks. The main cause of network attacks is software vulnerabilities, which can lead to the theft of a large amount of information from computers, system slowdowns, and even complete system failures [1]. In general, software vulnerabilities are analyzed using both source code and binary code. However, the majority of software is released in binary form. When two or more binary files have vulnerabilities that are similar in functionality and code logic, this is referred to as binary vulnerability similarity. The presence of similar binary vulnerabilities can rapidly spread network attacks and increase the difficulty of defense. Therefore, detecting the similarity of binary vulnerabilities is of great significance.

The main idea of deep graph matching technology is to detect the similarity of two or more binary vulnerabilities by matching the network feature graphs and network feature words. This study is an innovation in vulnerability detection technology, integrating deep graph matching with vulnerability similarity detection. It plays a key role in the detection process and lays a solid foundation for future developments in this field.

2. Vulnerability similarity detection model construction

2.1. Word vectorization

The prerequisite for inputting network data samples into the detection model is to complete the vectorization conversion of network information feature data. This vectorization conversion is achieved through word embedding, which maps words into a real-valued vector space, transforming words expressed in natural language into vectors or matrices that computers can understand. Similarly, transforming images into vectors or matrices is known as image embedding. The Continuous Bag-of-Words (CBOW) model is an important component of the embedding model, capable of generating word vectors by predicting the input itself based on the semantic features of preceding words and the features of following words. In this study, the CBOW model is selected to vectorize the features of network data samples. The first step is to segment the network data samples into multiple words based on instructions and features, and then train the CBOW model to obtain the word and image vector matrices [2]. The vector conversion process typically uses a corpus with dimensions ranging from 130 to 220. Since the vocabulary of the network data samples to be converted is smaller than that of ordinary text classification, this study converts the first eight non-numeric features of the network data samples into 20-dimensional vectors, with the remaining numeric features placed after the 20-dimensional feature vectors. Viewing the sample as a code segment, each code segment contains 50 instructions, which can be converted into a 176-dimensional vector. Therefore, the vector dimension of a code segment is 50×176. Finally, the vectorized network data features are input into the trained vulnerability similarity detection model to complete the vulnerability similarity detection.

2.2. Vulnerability similarity detection model

Long Short-Term Memory (LSTM) is an important deep learning method. The advantage of LSTM is its ability to extract contextual information from sequential data. The binary vulnerability similarity detection based on deep graph matching involves feature extraction of the embedded words and images to analyze their intrinsic associations and similarities. The structure of the vulnerability similarity detection model based on LSTM is shown in **Figure 1**.

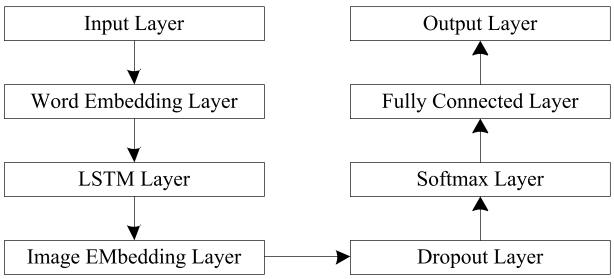


Figure 1. Structure of the vulnerability similarity detection model based on LSTM

After the network information data is input, the Word Embedding layer is responsible for vectorizing the data. The LSTM layer is in charge of mining and extracting the contextual associations of the quantized network

information word vector matrix, and it also performs feature extraction on the word vectors from the Word Embedding layer and the embedded images to analyze their intrinsic associations and similarities. The Dropout layer then randomly deactivates neurons in the results analyzed by the LSTM layer to reduce the probability of overfitting. To transform the raw output of the neural network into a probability distribution problem, the Softmax layer maps the output results analyzed by the LSTM layer to the interval (0,1). Finally, the fully connected layer completes the detection result output with a dimension of 1.

3. Vulnerability similarity detection based on deep graph matching

3.1. Word embedding

Word embedding can express network information data samples as real-valued vectors, and these real-valued vectors can significantly increase the similarity between word vectors of contextually similar words. This similarity can greatly reduce the distance between two word vectors in the vector space [3]. In the Word Embedding layer, assume that a word in a data sample is *w*, and there are *c* words before and after it. The word vectors of these *c* words are input into the CBOW model for training and the output of the word vector for *w*. After the CBOW model is trained, the output word vectors have clear relevance and analogy. Moreover, the word vectors can clearly express the meaning of the word itself as well as the intrinsic relationships and similarities between words.

3.2. Graph embedding

Assume that a 3-layer control flow graph *g* is input into the detection model, and all vertices *v* within the graph are related to a feature vector k_i. Under training conditions, the features of all vertices will be updated to form a new feature vector $\vec{g}$. Through iterative aggregation, the embedded graph is obtained $\vec{g}$. When the number of iterations is *t*, the new feature $\mu_i^{(t)}$ is obtained by combining the vertex features and the graph structure features. The *t*+1-th iterative update formula for the updated feature vector μ_i of all vertices is:

$$\mu_i^{(t+1)} = F\left(k_{\upsilon_i}, \sum_{j \in N_{(\upsilon_i)}} \mu_j^{(t)}\right), \forall \upsilon_i \in V \tag{1}$$

Here, the *i*-th vertex, the set of all vertices, and the set of neighbors of the vertex are denoted by v_i, V, and $\backslash N_{(\upsilon_i)}$, respectively. F represents the nonlinear mapping, and the mapping formula is:

$$F\left(k_{\upsilon_i}, \sum_{j \in N_{(\upsilon_i)}} \mu_j^{(t)}\right) = tanh\left(W_1 k_{\upsilon_i} + \sigma\left(\sum_{j \in N_{(\upsilon_i)}} \mu_j^{(t)}\right)\right) \tag{2}$$

Here, F represents the matrix of dimension k_i (vertex feature vector) × dimension of the graph embedding vector. The hyperbolic tangent function and the nonlinear function are denoted by *tanh* and σ, respectively. The formula for the nonlinear function of the *n*-layer fully connected neural network is:

$$\sigma(y) = P_1 \times ReLU\left(P_2 \times \cdots ReLU\left(P_n \times y\right)\right) \tag{3}$$

Here, P_1, P_2, and P_n are the hyperparameter matrices of dimensions *n* (graph embedding vector dimension) × graph embedding vector dimension for the 1st, 2nd, and *n*-th layers of the neural network graph embedding vectors, respectively. $ReLU(\cdot)$ represents the output value of the rectified linear unit function, and it satisfies $ReLU(\cdot) = max\{0, m\}$. After *T* iterations, the embedded graph $\vec{g}$ is:

$$\vec{g} = W_2 \sum_{v_i \in V} \mu_i^{(T)} \tag{4}$$

Here, W_2 represents the matrix of dimension × graph embedding vector dimension × embedding vector dimension, used to transform the final graph embedding vector.

3.3. LSTM

Long Short-Term Memory (LSTM) networks are an extension of the standard Recurrent Neural Network (RNN) and are highly effective in detecting sequential feature relationships. RNN can dynamically simulate the input, output, and hidden states of network nodes. Suppose the input, output, and hidden state are represented by x_t, y_t, and h_t, respectively. Since the current network information is based on the previous network information, the formula for the hidden state of a standard RNN is:

$$h_t = f_h\left(x_t, h_{t-1}\right) \tag{5}$$

Here, f_h represents the state transition function of the network node, and h_{t-1} represents the hidden state of the network node at the previous moment. The formula for the network node output state is:

$$y_t = f_0\left(h_t\right) \tag{6}$$

Here, f_0 represents the output function of the network node. The input state x_t of the network node can be regarded as the sequence element of the RNN network. By combining the current input state x_t with the output h_{t-1} of the previous hidden layer state, the hidden layer state output h_t can be obtained. However, the RNN network can only preserve short-term input sequence information of the network nodes. Therefore, it is necessary to build upon the RNN network with LSTM to preserve long-term input sequence information of the network nodes. LSTM also has the capability to extract contextual information from sequential data, which enables effective feature extraction of the embedded words and images, and analysis of their intrinsic associations and similarities [4]. The LSTM network is a standard cell structure, where all cells update the current cell state c_t and hidden state h_t based on the output of the previous moment, and then output the updated states. Therefore, the LSTM network can effectively preserve the current cell state and the previous cell states. The forget gate f_t, input gate i_t, and output gate o_t contribute to the memory-preserving function of the LSTM network.

The detection of binary vulnerability similarity first requires converting binary code into a sequence and inputting the result into the LSTM network. The forget gate of the LSTM network can preserve long-term memory of the sequence cell state. In other words, the forget gate selects information relevant to the similarity detection of the current word feature vector and the corresponding graph feature vector from the previous cell state for selective memory. The current input state x_t can be regarded as the code feature being analyzed, while the output h_{t-1} of the previous hidden layer state can be seen as a previously saved code sequence fragment. Since not all word and graph code features are useful for the entire detection, the forget gate needs to set the corresponding memory unit value to 0 to forget the word and graph code features. However, if the detection model is detecting a vulnerability pattern with a loop, the forget gate needs to set the corresponding memory unit value close to 1 to promote the memory of the word and graph code features, thereby achieving vulnerability detection. The forget gate is responsible for forgetting code features unrelated to vulnerability detection, while the input gate judges the new word and graph code features and their value, storing them in long-term memory. During this period, the vulnerability similarity detection model can form a feature representation that meets the requirements

of vulnerability patterns and good refinement based on the sequence of word and graph code features [5,6]. The candidate state vector containing high-value information is formed by compressing the network node input state x_t through the tanh function. Then, the nonlinear function outputs the code features x_t of the word and graph being analyzed and the previously saved code sequence fragment x_t, forming a valve vector that can determine the importance of all information in the candidate memory. In the LSTM network, the output gate is responsible for selecting the sequence information that can represent the time step from long-term memory as the feature vector of the vulnerability word and graph for output. The valve vector formed by the nonlinear function output of the input gate analyzes the outputtable vulnerability feature vectors in long-term memory, and then the tanh function activates the output of the vulnerability word and graph feature vectors. The final output feature vector formula is:

$$h_t = o_t \times tanh\left(c_t\right) \tag{7}$$

Assume that a binary function graph embedding is transformed into a basic block feature sequence (BB_1,BB_2,…,BB_T), which is input into the LSTM network and processed through three gates to form a new hidden state. The formula for calculating the binary function graph embedding vector is:

$$V_{func} = h_T \tag{8}$$

Here, h_T is the hidden state vector generated by the LSTM network during feature sequence processing, which also represents the semantic features of the binary function graph embedding.

Assume that the word embedding vector and the graph embedding vector obtained through the binary function graph embedding vector calculation method are V_{word} and V_{image}, respectively. The detection result is determined by the vulnerability similarity score, which is calculated using the following formula:

$$Similarity\left(word, image\right) = cos\left(\theta\right) = \frac{V_{word} \cdot V_{image}}{\left\|V_{word}\right\|\left\|V_{image}\right\|} = \frac{\sum_{i=1}^{n} V_{word}^{i} V_{image}^{i}}{\sqrt{\sum_{i=1}^{n}\left(V_{word}^{i}\right)^2}\sqrt{\sum_{i=1}^{n}\left(V_{image}^{i}\right)^2}} \tag{9}$$

Here, $cos(\theta)$ represents the dot product of the word embedding vector V_{word} and the graph embedding vector V_{image} . $\left\|V_{word}\right\|$ and $\left\|V_{image}\right\|$ are the magnitudes of the word embedding vector V_{word} and the graph embedding vector V_{image}, respectively. The closer the final vulnerability similarity score is to 0, the more dissimilar the two or more vulnerabilities are. Conversely, the closer the vulnerability similarity score is to 1, the higher the similarity between multiple vulnerabilities.

4. Experimental analysis

4.1. Vulnerability similarity detection

Before training the detection model, the corresponding training parameters need to be set. The learning rate and the number of epochs are set to 0.0001 and 50, respectively. The training set consists of 3-layer control flow graphs compiled from two sources: the same source function and different source functions. One graph is selected from each of the two sources of 3-layer feature control flow graphs, denoted as g_1 and g_2. Here, g and g_1 are images generated by compiling the same source function, while g and g_2 are images generated by compiling different source functions. The training samples ⟨ g, g_1⟩ and ⟨ g, g_2⟩ are labeled as +1 and -1, respectively.

Due to the uncertainty of images across different epochs, the training data varies from epoch to epoch. To enhance the comparability of the experiments, all the data trained in different epochs are re-randomly input into

the detection model for training. The graph embedding dimension is set to 128, the embedding depth to 2, and the number of iterations to 5. The SVM detection model, KNN detection model, and GGNN detection model are selected as the benchmarks for comparison with the LSTM-based vulnerability similarity detection model used in this paper. The ROC curves of different methods are compared, as shown in **Figure 2**. The ROC curves of the test set and the training set for the LSTM vulnerability similarity detection model almost overlap, and it is evident that the LSTM model outperforms the other detection models. The AUC of the LSTM vulnerability similarity detection model is 0.9721. In contrast, after training with the same data, the AUC of the SVM detection model drops from 0.878 in the test set to 0.725, the KNN detection model drops from 0.776 in the test set to 0.602, and the GGNN detection model drops from 0.648 to 0.524. It is clear that the vulnerability similarity detection model proposed in this paper has the most stable performance.

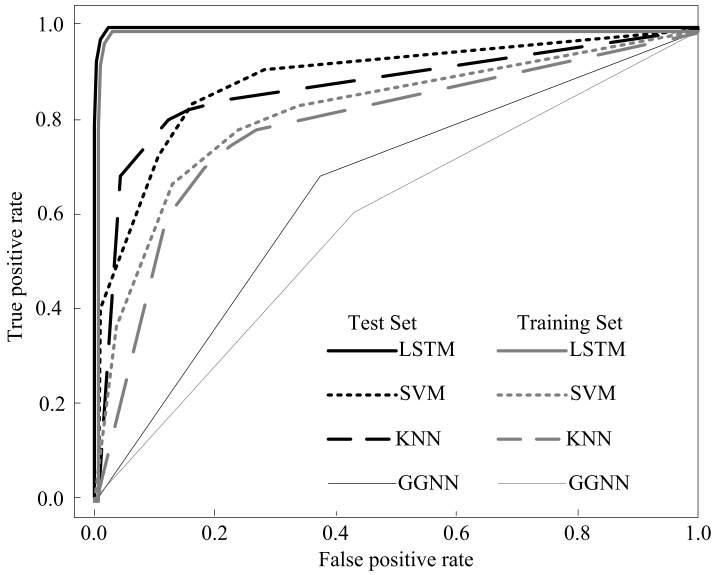


Figure 2. Comparison of ROC curves for different methods

4.2. Detection effect

The binary vulnerability similarity detection model based on deep graph matching proposed in this paper is used to detect the similarity of vulnerabilities that have been determined to be similar. By comparing the detection results with the actual situation, the detection effect of the vulnerability similarity detection model in this paper is determined. The detection effect of the model is shown in **Table 1**.

The closer the similarity score is to 0, the more dissimilar the two or more vulnerabilities are. The closer the similarity score is to 1, the higher the similarity between multiple vulnerabilities. For the three types of vulnerabilities—memory leaks, buffer overflows, and targeted attacks—the similarity scores obtained using the detection model in this paper are close to 1, indicating a clear similarity between the vulnerabilities. For the two types of vulnerabilities—out-of-bounds memory access and logical design flaws—the similarity scores are less than 0.4, indicating no similarity between the vulnerabilities. These results are consistent with the actual situations of the five types of vulnerabilities. Therefore, the binary vulnerability similarity detection model based on deep graph matching can effectively detect the similarity of vulnerabilities and enhance network security.

Table 1. Detection effect of the model

Program vulnerability categories	Actual situation	Similarity detection results	
		Vulnerability similarity scores	Similarity determination
Memory leak	Vulnerability is the same	0.98	Similarity vulnerability
Buffer overflow	Vulnerability is the same	0.99	Similarity vulnerability
Out-of-bounds memory access	Different functionality	0.28	Non-similarity vulnerability
Logical design flaw	Function after patching	0.37	Non-similarity vulnerability
Targeted attack	Vulnerability is the same	0.95	Similarity vulnerability

4.3. Evaluation metrics

To verify that the binary vulnerability similarity detection model based on deep graph matching proposed in this paper can effectively improve the detection accuracy, this section will analyze the SVM detection model, the KNN detection model, the GGNN detection model, and the detection model proposed in this paper. Four commonly used evaluation metrics—accuracy, recall, precision, and F1-score—will be used to assess the proposed detection model.

Assume that the data features in the dataset are divided into two classes: positive and negative. *TP* (True Positive) represents the number of positive instances correctly predicted as positive by the algorithm. *TN* (True Negative) represents the number of negative instances correctly predicted as negative. *FP* (False Positive) represents the number of negative instances incorrectly predicted as positive. *FN* (False Negative) represents the number of positive instances incorrectly predicted as negative.

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{10}$$

$$recall = \frac{TP}{TP + FN} \tag{11}$$

$$precision = \frac{TP}{TP + FP} \tag{12}$$

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{13}$$

The evaluation results for the SVM, KNN, and GGNN detection models are all below 85% in terms of accuracy, recall, precision, and F1-score, indicating that their overall detection accuracy is not satisfactory (**Table 2**). In contrast, the binary vulnerability similarity detection model based on deep graph matching proposed in this paper achieves evaluation metrics above 97%, demonstrating its ability to accurately detect the similarity of binary vulnerabilities.

Table 2. Evaluation results for different detection models

Detection model name	Accuracy (%)	Recall (%)	Precision (%)	F1 (%)
SVM detection model	81.39	79.47	82.95	78.48
KNN detection model	83.68	79.76	81.54	80.45
GGNN detection model	82.54	84.35	83.41	83.49
Binary vulnerability similarity detection model based on deep graph matching	97.33	98.52	97.46	97.86

5. Conclusion

This paper introduces deep graph matching technology into program vulnerability detection. By using deep learning to mine the word vector features and graph vector features of vulnerabilities, similarity detection is performed on vulnerabilities within different binary programs. The LSTM network is primarily used to extract the feature vectors of word embeddings and graph embeddings, analyzing their intrinsic associations and similarities. The results show that the AUC of the LSTM-based vulnerability similarity detection model in this paper is 0.9721. In contrast, the AUC of the SVM detection model drops to 0.725 after training, the KNN detection model drops to 0.602, and the GGNN detection model drops to 0.524. The evaluation metrics of the SVM, KNN, and GGNN detection models are all below 85%, while the evaluation metrics of the vulnerability similarity detection model in this paper are all above 97%. The binary vulnerability similarity detection model based on deep graph matching not only has stable performance but also can accurately detect the similarity of binary vulnerabilities, effectively safeguarding network security.

Funding

Special Project Funded by Tsinghua University Press: "Engineering Drawing and CAD" Course Construction and Textbook Development

Disclosure statement

The author declares no conflict of interest.

References

[1] Yang S, Xu Z, Xiao Y, et al., 2023, Towards Practical Binary Code Similarity Detection: Vulnerability Verification via Patch Semantic Analysis. ACM Transactions on Software Engineering and Methodology, 32(6): 1–29.

[2] Li L, Ding S H H, Tian Y, et al., 2023, VulANalyzeR: Explainable Binary Vulnerability Detection with Multi-Task Learning and Attentional Graph Convolution. ACM Transactions on Privacy and Security, 26(3): 1–25.

[3] Zhu Y, Lin G, Song L, et al., 2023, The Application of Neural Network for Software Vulnerability Detection: A Review. Neural Computing and Applications, 35(2): 1279–1301.

[4] Wen X C, Gao C, Ye J, et al., 2023, Meta-Path Based Attentional Graph Learning Model for Vulnerability Detection. IEEE Transactions on Software Engineering, 50(3): 360–375.

[5] Tang M, Tang W, Gui Q, et al., 2024, A Vulnerability Detection Algorithm Based on Residual Graph Attention Networks for Source Code Imbalance (RGAN). Expert Systems with Applications, 238: 122216.

[6] Yan X, Sun M, Han Y, et al., 2023, Camouflaged Object Segmentation Based on Matching–Recognition–Refinement Network. IEEE Transactions on Neural Networks and Learning Systems, 35(11): 15993–16007.

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

An Intelligent Control Method Based on the Artificial Neural Network Model

Liangkai Zhou[1]*, Dan Han[1], Qinzhe Wang[2], Nv Yang[3]

[1]Sanya Institute of Technology, Sanya 572022, Hainan, China
[2]China National Software, Beijing 102200, China
[3]Yum China Catering (Shenzhen) Co., Ltd., Haikou 570203, Hainan, China

**Author to whom correspondence should be addressed.*

Abstract: The topology structure of the artificial neural network is an intelligent control model, which is used for the intelligent vehicle control system and household sweeping robot. When setting the intelligent control system, the connection point of each network is regarded as a neuron in the nervous system, and each connection point has input and output functions. Only when the input of nodes reaches a certain threshold can the output function of nodes be stimulated. Using the networking mode of the artificial neural network model, the mobile node can output in multiple directions. If the input direction of a certain path is the same as that of other nodes, it can choose to avoid and choose another path. The weighted value of each path between nodes is different, which means that the influence of the front node on the current node varies. The control method based on the artificial neural network model can be applied to vehicle control, household sweeping robots, and other fields, and a relatively optimized scheme can be obtained from the aspect of time and energy consumption.
Keywords: Artificial neural network; Model; Control method; Optimization scheme

Online publication: October 21, 2025

1. Parameters and methods

1.1. Experimental parameters

The experimental platform is MATLAB R2016 simulation software. It is assumed that the neurons in the artificial neural network model have two sets of parameters: input and output, and the input end has 60 random numbers. When the threshold value of the output terminal is 0.53 (that is, when the threshold value of the input exceeds 0.53), the neuron can have an output. There are three curves, respectively representing three samples: theory, practice and test. Enter yes in the theoretical sample SamIn = sort(59*rand(1,SamNum)+1); The output result is SamOut = 0.5447*SamIn.^0.1489. There is randomness at the input end in the actual sample TargetIn=1:60. Output has Targ etOut=[0.53173198482933,0.599828865,0.644564773...]. For the test sample, the value of the input end is found by training, and the value of the output end is calculated by the hidden layer, which is similar to the structure of

dendrite-axon-synapses in neurons, and the formula is .

1.2. Experimental methods

1.2.1. Test preparation

The network networking mode is divided into two types: one is the ordinary networking mode to calculate the energy consumption and the time required for operation of the network; the other is to test the data according to the intelligent control method of the artificial neural network model.

1.2.2. Calculation method

The output range radius of the node in the input layer is set to R = 1 km, and the output range after being weighted by the input end and the hidden layer is a circle with a diameter of 2R, with the time of 20 minutes as the horizontal coordinate and the output direction as the vertical coordinate.

1.2.3. Determination method

The output direction of the network and the number of output nodes are measured by the theoretical value measurement and the actual data input. The measurement process is determined by three measurements to determine the same value. Finally, the theoretical and actual values are compared to draw a graph.

1.3. Data processing

The algorithm data processing of the experimental model is set and processed by MATLAB R2016a. Different input data plus different weights make the output angles and directions different. Different output values are marked with different colors to get different experimental result sets, so that comparative research can be conducted.

2. Results and analysis

2.1. Experimental method

As shown in **Figure 1**, the output direction of the network and the number of nodes that can be output increase continuously with the change of time. However, when the time is 12 minutes, the number of nodes decreases slightly after reaching the number, and then rises continuously with time. It can be seen from **Figure 2** that the two curves change with the change of obstacle time, and as in **Figure 1**, the output direction and output network node increase continuously after the node reaches the threshold. **Figures 3** and **4** show the training and testing and error.

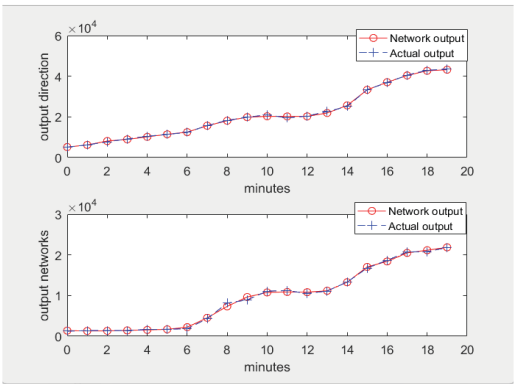


Figure 1. Output direction and output network in a period of time

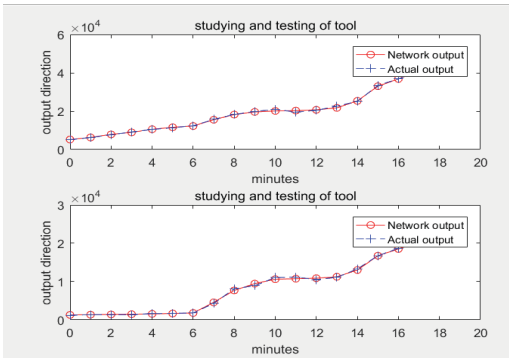


Figure 2. Output direction and output network in a period of time

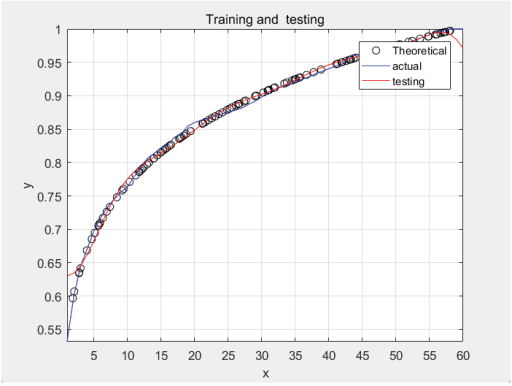


Figure 3. Training and testing

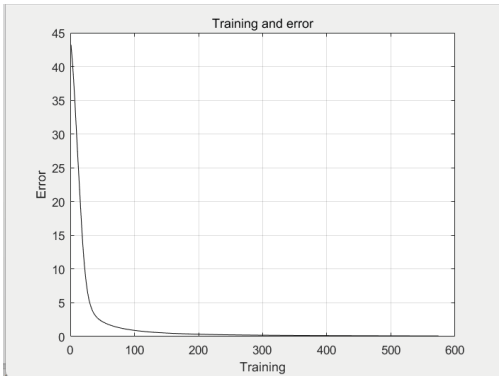


Figure 4. Test sum and error

2.2. Impact of networking on energy consumption of mobile nodes

The network model of an artificial neural network can effectively control the transmission efficiency of the network, reduce the congestion caused by the input and output of the network, decrease the transmission cost of the network, and thus lower the total energy consumption of the network [1].

2.3. Effect on the transmission efficiency of mobile devices

In the networking model of an artificial neural network, the output direction of nodes can be multiple, and each

node can also receive input from other nodes. When the input received by the node from other nodes reaches a certain threshold value, the node outputs only when each output direction of the node also generates different weights for the preceding node. Therefore, the node can choose the direction of output with smaller values through the weight selection of the normalized function, thus improving the output efficiency of the network [2].

3. Discussion

In this paper, we compare the results between the traditional mobile node networking model and the artificial neural network networking model. In the networking mode of the artificial neural network model, the transmission direction of the network nodes of the artificial neural network model is found through the comparison of the networking systems of different networking models. The intelligent algorithm is used to select different output directions and transmission paths to reduce the collision of the transmission network nodes during the transmission process and improve the fan-in and fan-out rate of the network. Therefore, to improve the transmission performance of the network and the utilization rate of energy consumption, the experimental results show that the intelligent control method using the artificial neural network model is more effective in improving the performance of the network.

The transmission efficiency is also an advantage of the intelligent control method of an artificial neural network. The intelligent control method is used to select the optimal path and improve the transmission efficiency and the optimal direction selection in the aspects of computing the multi-dimensional network, complex path selection, and network intrusion nodes [3].

4. Conclusion

According to the research results, the nodes of the artificial neural network model can receive input from other nodes in different directions, and only choose output when the input value reaches the threshold value. Different network nodes can choose the output direction and output path of different weights through the intelligent control algorithm of the artificial neural model, and the output direction will produce different weights for the input of the pre-node. The input path selection of different nodes is weighted by the hidden network layer. The intelligent control method of the artificial neural network model has great application to household sweeping robots and vehicle intelligent control systems [4].

Disclosure statement

The authors declare no conflict of interest.

References

[1] Li J, 2023, Real-Time Performance Prediction of TBM Based on Dual Neural Network. Journal of Mechanical Engineering, 2023: 3–5.

[2] Wang F, 2023, Research on Optical Performance Monitoring of Optical Communication System Based on Artificial Neural Network, dissertation, Suzhou University, 1–8.

[3] Ruan E, 2022, Research on Engine Performance and Emission Prediction Based on Artificial Neural Network,

dissertation, Jilin University, 2–4.

[4] Tian N, Lan H, 2020, Performance Comparison of Artificial Neural Network and Decision Tree Model in Landslide Susceptibility Analysis. Journal of Geo-Information Science, 22(12): 2304–2309.

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Research on Optimization Strategies for Signal Integrity of High-Speed Digital Circuits in Electronic Information Engineering

Yiming Li*

Empyrean Technology Co., Ltd., Chengdu 610054, Sichuan, China

**Author to whom correspondence should be addressed.*

Abstract: With the rapid development of electronic information engineering, high-speed digital circuits have been increasingly widely applied in various fields. In high-speed digital circuits, signal integrity is prone to interference from various external factors, leading to issues such as signal distortion or degradation of system performance. Based on this, this paper conducts research on the optimization strategies for signal integrity of high-speed digital circuits in electronic information engineering. It deeply analyzes the importance of high-speed digital circuits, elaborates on the challenges they face and the specific manifestations of signal integrity issues, and proposes a series of optimization strategies in electronic information engineering. The aim is to improve the signal integrity of high-speed digital circuits and provide theoretical support and practical guidance for the development of related fields.
Keywords: High-speed digital circuits; Signal integrity; Electronic information engineering; Optimization strategies

Online publication: October 21, 2025

1. Introduction

In the digital era, the technology of electronic information engineering is developing with each passing day. As a core component of electronic information systems, the performance of high-speed digital circuits directly affects the operating efficiency and reliability of the entire system. With the development of electronic devices towards miniaturization and high performance, higher requirements have been put forward for high-speed digital circuits [1]. As a key issue in the design of high-speed digital circuits, signal integrity has attracted widespread attention. If the signal cannot maintain its integrity during transmission, it will lead to problems such as signal distortion and increased bit error rate, which seriously affect the performance of electronic systems. Therefore, researching the optimization strategies for signal integrity of high-speed digital circuits in electronic information engineering has important practical significance.

2. The importance of high-speed digital circuits

In today's digital era, high-speed digital circuits have become a core pillar in the field of electronic information engineering, and their significance is fully demonstrated in various key domains such as communications, computing, and consumer electronics.

In the communications sector, high-speed digital circuits serve as the central hub for enabling high-speed data transmission. With the rapid development of next-generation communication technologies like 5G and 6G, user demands for data transmission rates and quality have grown exponentially. High-speed digital circuits, leveraging their superior performance, can significantly enhance communication speeds, pushing data transmission to new heights and making high-definition video calls and rapid transmission of massive data a reality [2]. Additionally, they effectively reduce bit error rates, ensuring accurate and reliable information transmission, thus providing a solid foundation for the stable operation of communication networks. Whether in mobile communication base stations or satellite communication systems, high-speed digital circuits are indispensable components that directly determine the overall performance of communication systems and user experience [3].

In the computing field, high-speed digital circuits form the critical cornerstone for building high-performance computer systems. With the rise of emerging technologies such as artificial intelligence and big data processing, the volume of data that computers need to handle has exploded, placing extremely high demands on computing speed and system architecture. The application of high-speed digital circuits can significantly boost the operational speed of computers, drastically reducing the processing time for complex computing tasks [4]. Furthermore, they optimize system architectures, enabling more efficient collaboration between computer components, thereby enhancing the overall performance and stability of computers. This meets the stringent computing power requirements of application scenarios such as scientific research, industrial design, and financial analysis [5].

In the consumer electronics sector, the application of high-speed digital circuits is equally ubiquitous and crucial. From high-definition televisions and smartphones to solid-state drives (SSDs) and gaming consoles, various consumer electronic products rely on high-speed digital circuits to achieve functions such as high-definition video transmission and fast data storage. Taking smartphones as an example, the smooth realization of their functions, such as photo taking, video recording, and game running, is inseparable from the rapid processing and transmission of image data by high-speed digital circuits. Thanks to high-speed digital circuits, solid-state drives can achieve fast reading and writing of data, which greatly improves the performance of storage devices and brings users a more convenient and efficient experience [6].

3. Signal integrity issues in high-speed digital circuits

3.1. Reflection problems

When a signal travels along a transmission line, reflection occurs once it encounters a point of impedance discontinuity, such as the end of the transmission line, a connector, or the interface between circuit components. This impedance mismatch forces part of the signal energy to return to the source, forming a reflected signal [7]. For example, in a typical digital circuit, after a signal is output from the driver, it travels through a section of transmission line to the receiver. If the characteristic impedance of the transmission line does not match the output impedance of the driver or the input impedance of the receiver, reflection will occur when the signal reaches the end of the transmission line. The superposition of the reflected signal and the original signal causes overshoot and undershoot in the signal waveform [8]. Overshoot may exceed the voltage tolerance of the chip, damaging it; undershoot, on the other hand, may cause the signal level to drop below the chip's recognition threshold, resulting

in data errors. Furthermore, the reflected signal will reflect back and forth in the transmission line, creating a ringing phenomenon, which further exacerbates signal distortion and seriously affects signal integrity and system reliability.

3.2. Crosstalk phenomenon

The impact of crosstalk should not be underestimated. It can directly cause distortion of adjacent signal waveforms, blurring the originally well-defined signal edges, attenuating signal amplitude, and even, in some extreme cases, inverting the signal phase. These changes not only increase the difficulty of signal interpretation but also may lead to data errors, thereby affecting the performance and reliability of the entire electronic system. In high-speed digital circuit applications, crosstalk is particularly problematic [9]. For instance, in high-speed parallel buses, multiple signal lines are closely arranged, resulting in strong electromagnetic coupling between adjacent lines, which easily causes crosstalk. When the signal on one line changes rapidly, the electromagnetic field it generates affects the signal on adjacent lines, leading to issues such as false transitions in adjacent signals and impairing the correct transmission of data. Moreover, as circuit integration and signal transmission rates increase, the distance between signal transmission lines becomes shorter, making crosstalk problems more severe.

4. Factors affecting signal integrity in high-speed digital circuits

4.1. Mismatch in characteristic impedance of transmission lines

Mismatch in the characteristic impedance of transmission lines is one of the main causes of reflection phenomena. When a signal propagates along a transmission line, if the characteristic impedance of the line changes suddenly at a certain point, reflection will occur at that point. For example, changes in line width, openings, and connected devices can all lead to impedance mismatch. A reduction in line width increases the resistance of the transmission line and alters the values of capacitance and inductance, resulting in an increase in characteristic impedance. For openings and connected devices, their special structural properties introduce additional parasitic capacitance and inductance, causing changes in the characteristic impedance of the transmission line. Such impedance mismatch leads to signal reflection during transmission, which affects signal integrity [10].

4.2. Coupling between transmission lines

Coupling between transmission lines causes crosstalk, interfering with the transmission of adjacent signals. This coupling includes capacitive coupling and inductive coupling. Parasitic capacitance between transmission lines gives rise to capacitive coupling: when the signal voltage in one line changes, an induced current is generated, which spreads to other lines through parasitic capacitance and affects them. Additionally, magnetic fields generated by current fluctuations in a transmission line induce voltages in other lines, thereby influencing their signals. For instance, in a multi-layer PCB structure, lines on different layers and adjacent lines on the same layer interact with each other. Improper design can lead to crosstalk [11].

4.3. Power supply noise and thermal noise

Power supply noise, thermal noise, and other forms of noise also interfere with signal transmission. Noise from the power supply system stems from unstable power sources or interference from other circuit components, causing fluctuations in power lines and affecting the performance of connected digital devices. Thermal noise, caused by the thermal motion of electrons in electronic components, is an unordered noise that accumulates in signals,

reducing the signal-to-noise ratio and impairing transmission quality. For example, in a complex electronic system, multiple power units operating simultaneously can cause significant fluctuations in power supply voltage. Such power supply noise may propagate through power lines to high-frequency digital devices and interfere with signals [12].

5. Optimization strategies for signal integrity of high-speed digital circuits in electronic information engineering

5.1. Wiring optimization strategies

Wiring is a crucial link in high-speed digital circuit design, and its strategies largely affect signal integrity. An appropriate wiring layout can prevent impedance mismatch in transmission lines, suppress coupling effects, and reduce the impact of reflection and series signal quality. First, shorten the length of transmission lines. Information may have corresponding delays during transmission, and the delay is proportional to the cable length. The longer the cable, the more interfered signals there will be, and the more serious the echo and other conditions will be [13]. Therefore, one of the measures to ensure the design quality of electronic circuit boards is to optimize the circuit settings. During specific wiring, various devices and components should be arranged densely to minimize the obstruction of transmission lines to signal transmission. Special attention should be paid to some signals, especially important high-frequency signals such as clock frequency and high-speed data transmission. The length of the cables they use should be noted to reduce the threat of information transmission delay and interference and enhance their reliability. Second, avoid unnecessary bends and crossings. When the transmission line is bent, the signal transmission path is lengthened, and it may also lead to a certain increase in resistance, which may cause signal reflection. In addition, when two transmission lines cross, mutual interference may occur. We should try to keep the transmission line straight and avoid too many bends. If bending is necessary, the bending radius should be as large as possible to reduce the impact on the signal. In addition, the position of transmission lines should be reasonably arranged to ensure that no transmission line crosses another. For example, in the production process of multi-layer printed boards, different types of information can be arranged in each layer, and then information transmission can be realized by means of appropriate inter-layer connection methods, so as to minimize the probability of transmission lines crossing each other, thus realizing signal transmission and reducing the probability of signal reflection and interference. Third, increase the width of transmission lines. Widening the line can effectively reduce line resistance, thereby reducing losses in the information transmission process. At the same time, increasing the line width can also reduce a certain line impedance, facilitate the impedance requirements of other electrical components, and avoid signal reflection. In addition, widening the line width can also enhance the current-carrying capacity of the line and avoid interference caused by excessive current. For example, when dealing with power signals with large current or high-frequency data signals, the width of the transmission channel for such signals should be appropriately increased. However, it should be noted that increasing the line width will occupy more PCB space, so a certain balance should be struck between signal requirements and space layout.

5.2. Termination matching technology

In practical operation, commonly used termination matching technologies include parallel termination, series termination, and Thevenin termination, which play an important role in ensuring the quality of signal technology. For different application scenarios, appropriate termination matching technologies should be carefully selected according to specific situations. First, parallel termination. The so-called parallel termination is a way to achieve matching by connecting a resistor in series at the end of the line, which is equal to the characteristic impedance of

the line. It avoids signal echo and eliminates it through the resistor. Usually, the parallel method can effectively reduce the harm of the reflected echo to the signal and enhance the signal quality. In high-frequency digital systems, if there is impedance mismatch at both the source and load ends, the parallel method can well avoid the problems. For example, at the end of some high-speed data lines, a resistor with the same characteristic impedance as the transmission line is connected to the line to avoid signal reflection that prevents transmission. However, it often causes a certain system power loss because the resistor consumes part of the signal energy [14]. Second, series termination. Series termination is to add a resistor between the signal source and the transmission circuit to make the output impedance of the signal source suitable for the impedance corresponding to the line. When the signal is transmitted from the signal source to this resistor, its output impedance will be equal to the corresponding impedance of the output circuit, reducing reflection. This method is suitable for occasions where the signal source has a higher impedance and the receiver has a lower impedance. For example, in the clock transmission process, due to the high output impedance of the clock, series-parallel matching can greatly reduce reflection to ensure the consistency and stability of the clock output. Its advantage is that there is no need to increase additional power loss because the resistor does not consume energy during signal transmission. Third, Thevenin termination. Thevenin termination is a high-end connection method with two resistance values. One is connected in series at the end of the line and the power connection end, and the other is connected in series at the end of the line and connected to the ground. By setting these two resistance values, the resistance at the end of the line can be controlled to meet the needs of different resistance values, and the DC bias potential in signal transmission can be controlled. This termination connection method is suitable for electronic equipment with certain requirements for DC bias potential [15]. For example, in many systems where analog signals and digital signals are shared, Thevenin termination connection is adopted, which can not only ensure the quality of signal transmission but also meet the DC bias potential requirements of analog signals. However, its termination connection is more complex and requires calculating the resistance value.

6. Conclusion

To sum up, the signal integrity of high-speed digital circuits plays a pivotal role in electronic information engineering. With the continuous advancement of science and technology, electronic information systems will have increasingly higher performance requirements for high-speed digital circuits. To ensure signal integrity, effective optimization strategies should be adopted, such as wiring optimization strategies and termination matching technologies. These measures can improve the signal integrity of high-speed digital circuits and enhance the performance and reliability of electronic systems. In future work, engineers and researchers in related fields need to further strengthen research on the signal integrity of high-speed digital circuits and continuously explore new optimization strategies and technical means to meet the evolving needs of the electronic information engineering field.

Disclosure statement

The author declares no conflict of interest.

References

[1] Zhang J, 2024, R-HS2M Reversed Treated Copper Foil for High-Speed Digital Circuit HSD, Proceedings of the 25th

China Copper Clad Laminate Technology Symposium, Jiujiang Defu Technology Co., Ltd., 478–484.

[2] Zhu RX, 2024, Optimization Strategies for Signal Integrity of High-Speed Digital Circuits in Electronic Information Engineering. China Broadband, 20(07): 55–57.

[3] Hu Y, Huang JJ, Yao JJ, 2022, Architecture Design of Calibration Device for High-Speed Digital Channel in Integrated Circuit Test System. Ship Electronic Engineering, 42(02): 161–164.

[4] Huang FD, 2024, Research on High-Speed Signal Acquisition and Processing of FMCW Laser Detection System, dissertation, Beijing University of Posts and Telecommunications.

[5] Yang BS, Ma XF, Xu S J, et al., 2024, Analysis of Optimization Algorithms for Layout and Wiring of High-Speed Digital Circuits. Application of ICs, 41(05): 114–115.

[6] Liu YH, 2024, Design of a Four-Eye Vision Hardware Platform Based on FPGA, dissertation, East China Normal University.

[7] Wang CZ, Su CY, Li Z, et al., 2022, Simulation Analysis and Optimization of SDIO Board-Level Signal Integrity Based on Sigrity. Computer Measurement & Control, 30(03): 204–210 + 221.

[8] Guo CW, 2024, Design and Implementation of Mainboard Based on Domestic High-Performance C86 CPU, dissertation, University of Electronic Science and Technology of China.

[9] Liu HD, 2024, Research on Adaptive Downhole High-Speed Communication System, dissertation, University of Electronic Science and Technology of China.

[10] Wu PY, Gong YB, Zhang L, 2023, Principles of Printed Circuit Board Design and Electromagnetic Compatibility Measures. Electronics Quality, (11): 62–67.

[11] Zhao RJ, 2023, Analysis of Radiation Characteristics of Microstrip Lines in High-Speed Circuits Based on Analytical Approximation, dissertation, Xidian University.

[12] Huang H, 2023, Research on New Co-Simulation Methods for Signal Integrity and Power Integrity in High-Speed Digital Circuits, dissertation, Zhejiang University.

[13] Yuan Q, Liu XM, Liang XM, 2022, Implementation of Digital AGC Functional Circuit with Large Dynamic Range and High-Speed Response. Research & Progress of Solid State Electronics, 42(06): 467–472.

[14] Zhang ZH, Zhao JZ, Zhou YM, 2021, Design of Slave D-PHY Digital Circuit Based on MIPI Specification. Application of Electronic Technology, 47(11): 33–38.

[15] Wang J, Chen HM, Zhang HZ, et al., 2022, Design of Digital Down-Conversion Circuit in High-Speed TIADC Acquisition System. Microelectronics, 52(03): 418–424.

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Research on the Intelligent Evaluation of University Bursary Based on Blockchain

Juan Li*

Gannan University of Science and Technology, Ganzhou 341000, Jiangxi, China

**Author to whom correspondence should be addressed.*

Abstract: Aiming at the problems of easy falsification of information and inaccurate evaluation results in the existing university bursary evaluation, a bursary evaluation model (XGBoost Model based on Blockchain, XMB) combining machine learning and blockchain was designed. The relevant basic data of the bursary evaluation is stored on the chain to solve the problem of easy falsification of data in the evaluation process. At the same time, the evaluation results of the student bursary are uploaded to the chain to realize the traceability of historical data. In addition, the improved XGBoost algorithm is used to intelligently analyze and evaluate the basic data of students, and objectively give the student a bursary grade, which realizes the intelligence and scientific nature of the evaluation process and ensures the accuracy of the evaluation results. The experimental results prove that the model proposed in this paper has an accuracy rate of about 6% higher than that of the traditional XGBoost model, which has higher evaluation accuracy, throughput, and time efficiency. The method proposed in this paper is suitable for the evaluation of scholarships and bursaries in the student management system of colleges and universities.
Keywords: Blockchain; Bursary evaluation; Machine learning; XGBoost model

Online publication: October 21, 2025

1. Introduction

The management of university bursaries is an important task in the management of students in colleges and universities in our country, and it plays an important role in reflecting the country's care for students in financial difficulties and achieving the fair development of higher education. However, due to the continuous expansion of enrolment in colleges and universities across the country in recent years, the large number of students, and uneven regional economic development, there are some difficulties and problems in the management of university bursaries. For example, it is difficult to determine the qualifications of university bursaries, the assessment of university bursaries is inaccurate and unobjective, and the information on university bursaries has been tampered with.

In order to solve the above problems, researchers have applied machine learning technology to the

management of university scholarships. Lu *et al.* [1] combined the XGBoost model and principal component analysis to establish a classification and prediction method for poor students in colleges and universities with high accuracy. Li [2] used the K-nearest neighbor algorithm to increase the accuracy of predicting the real poor students. However, the above scheme does not involve a data protection mechanism, and lacks data security protection and personal data privacy protection. At the same time, some researchers have proposed applying blockchain technology to the management of university scholarships and bursaries to solve the above-mentioned related problems. Blockchain is a technical architecture that integrates cryptography, game theory, computer science, and other disciplines. It includes specific technologies such as distributed storage, consensus algorithms, asymmetric keys, peer-to-peer networks, and smart contracts. The combination of these technologies enables the blockchain to have the characteristics of decentralization, a high degree of autonomy, safety, reliability, and non-tamperability, giving the blockchain unique functions that are different from traditional systems. In the management of university scholarships and bursaries, Ding *et al.* [3] built a blockchain scholarship management platform to record student-related evaluation data, and trigger smart contracts to generate evaluation results when the scholarship conditions are met; Yan [4] applied blockchain technology to create a bursary data sharing platform, which improved the credibility of data interaction and avoided the problem of data tampering and forgery in the bursary management process. However, the above scheme lacks data intelligent optimization processing and fails to solve problems such as inaccurate evaluation results.

In order to solve the above problems, this paper proposes a research plan for the intelligent evaluation model of university bursaries based on blockchain:

(1) Combined with the use of the machine learning XGBoost algorithm, intelligent model training, and poverty level result prediction on student-related data.

(2) Store student data and assessment results on the blockchain, and design related smart contracts to ensure that the data can be traced and cannot be tampered with, etc.

2. Intelligent evaluation model

Aiming at the difficulties in identifying the qualifications of university bursaries and the different evaluation reference standards, this paper will use the improved XGBoost model to predict and analyze the grade evaluation of poor students.

2.1. Introduction of XGBoost

XGBoost (eXtreme Gradient Boosting) is an integrated learning model based on the gradient boosting algorithm proposed by Chen *et al.* [5]. The model idea is to build an integrated model with strong learning ability by using a combination of multiple weak models. The model is continuously upgraded and updated. Each iteration continues to generate a new model on the basis of the generated model, so that the model is constantly approximating the sample distribution. XGBoost combines the first-order derivative with the second-order derivative, and the algorithm uses the tree model complexity as a regular term in the objective function to avoid overfitting [6].

The objective function of the XGBoost model is:

$$Obj^{(t)} = \sum_{i=1}^{n} l(y_i \hat{y}^{(t-1)} + f_t(x_i)) + \Omega(f_t) + C \quad (1)$$

Where $l(\bullet)$ is the loss function, $\Omega(f_t)$ is the regular term, used to control the complexity of the model and

prevent over-fitting, and C is the inherent normal error. The regular term is defined as:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{2}$$

Among them, T is the number of leaf nodes, ω is the node weight, γ and λ are values between 0–1, which are used for the objective function to control the leaf nodes and the weight of each node.

Then use Taylor expansion to approximate the original goal. The second-order Taylor expansion is:

$$f(x+\Delta x) \simeq f(x) + f'(x)\Delta x + \frac{1}{2} f''(x)\Delta x^2 \tag{3}$$

And define:

$$g(i) = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}^{(t-1)}) \tag{4}$$

$$h(i) = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}^{(t-1)}) \tag{5}$$

So as to sort out the substitution and get a new objective function:

$$Obj^{(t)} = \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{6}$$

Taking the partial derivative of , get:

$$\omega_j = -\frac{\sum_{i \in I, j} g_i}{\sum_{i \in I, j} h_i + \lambda} \tag{7}$$

Finally got:

$$Obj^{(t)} = -\frac{1}{2} \sum_{j=1}^{T} \frac{(\sum_{i \in I, j} g_i)^2}{\sum_{i \in I, j} h_i + \lambda} + \gamma T \tag{8}$$

Use the objective function to find the tree with the best structure. The smaller the value, the better the structure of the tree. Each time we try to add a split to an existing leaf node, we calculate the difference between the structure score before the branch and the structure score after the branch, which is called the gain calculation:

$$Gain = \frac{1}{2} \left[\frac{(\sum_{i \in I, L} g_i)^2}{\sum_{i \in I, L} h_i + \lambda} + \frac{(\sum_{i \in I, R} g_i)^2}{\sum_{i \in I, R} h_i + \lambda} - \frac{(\sum_{i \in I, j} g_i)^2}{\sum_{i \in I, j} h_i + \lambda} \right] - \gamma \tag{9}$$

2.2. XGBoost model optimization

The XGBoost algorithm can improve and optimize the model through objective function optimization and parameter adjustment.

This paper improves on the original XGBoost model, uses the third-order approximation of the loss function to improve the accuracy of the model, and performs a third-order Taylor expansion of the objective function,

assuming:

$$p(i) = \partial^3_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}) \tag{10}$$

Take the derivative of equal to 0 and get a new result:

$$\omega_j = -\frac{\sum_{i \in I,j} h_i + \lambda \pm \sqrt{\left(\sum_{i \in I,j} h_i + \lambda - 2\sum_{i \in I,j} g_i \sum_{i \in I,j} p_i\right)}}{\sum_{i \in I,j} p_i} \tag{11}$$

And suppose:

$$\Delta = \sqrt{\left(\sum_{i \in I,j} h_i + \lambda - 2\sum_{i \in I,j} g_i \sum_{i \in I,j} p_i\right)} \tag{12}$$

And then get the new objective function:

$$Obj^{(t)} \approx \frac{\left(-\sum_{j=1}^{T} g_i(h_i + \lambda) \pm \Delta\right) p_i - \frac{1}{2}\left(\sum_{j=1}^{T}(h_i + \lambda) \pm \Delta\right)^2 - \frac{1}{6}\left(\sum_{j=1}^{T}(h_i + \lambda) \pm \Delta\right)^3}{p_i^2} + \gamma T \tag{13}$$

After adopting the third-order Taylor expansion, the complexity of the model is increased, and the prediction and classification ability of the model is also enhanced.

2.3. XGBoost modeling

The XGBoost model establishment process is divided into six stages: reading data, setting parameters, training model, predicting results, saving model, and model calling. Among them, parameter setting and optimization are the key steps.

(1) Reading data: XGBoost can load training data in a variety of data formats, such as text data in libsvm format, two-dimensional numpy arrays, and XGBoost binary cache files. The data selected in this paper are from a college student's campus card consumption data, family economic surveys, etc. Before loading the data, it needs to be cleaned, summarized, feature selection, data discretization, missing values and outlier processing, etc. In addition, it is necessary to fully consider the data indicators (characteristics) that can reflect poor students in these data. Here, we select the total number of personal consumption, total consumption amount, average consumption amount per time, average daily consumption amount, etc.

(2) Setting parameters: The parameter settings are shown in **Table 1**. This is a gradual process and requires constant debugging to find the optimal tree structure.

(3) Training model: After parameter setting and optimization are completed, the core of modeling is model training. Read the relevant data of the students, divide the training set and the test set according to the ratio of 7:3, and use the train function class to train 70% of the data to obtain a poor student grade prediction model.

(4) Predicting results: After the model is trained, you can use the trained model to predict the test data, that is, use the predict function to predict.

(5) Saving model: Save the model by using the pickle function of the library to save and call the model in Python, and complete the model on-chain operation in the smart contract layer.

(6) Model calling: After the model is saved and loaded on the chain, the data can be reimported for predictive analysis when in use.

Table 1. Parameter list

Parameter settings	Specific description
N_estimators=200	The number of weak classifiers in the ensemble algorithm, and the experiment is set to 200
Eta=0.1	The learning rate in the integration, the default is 0.3, the value range is [0,1]
Max_depth=6	The maximum tree depth of the weak classifier, the default is 6
Objective =multi:softmax	Specify the learning objective function and learning task, and actually solve the minimum value of the generalization error
Gamma=20	The descent of the objective function required for further branching on the leaf nodes of the tree, range [0,+∞]
booster=gbtree	It stands for weak evaluator, gbtree stands for gradient boosting tree
Subsample=1	The proportion of the sample drawn during random sampling, the range is (0,1], the default is 1

3. Blockchain-based bursary evaluation model

This article built a university bursary evaluation system based on Hyperledger Fabric [7] to meet the decentralized management of university bursary management and improve management efficiency, and to ensure that student user-related information is not leaked or tampered with.

3.1. Model architecture design

The blockchain-based bursary assessment model design is shown in **Figure 1**, which is divided into data layer, smart contract layer, and service layer. In the data layer, the blockchain uses HDFS (Hadoop distributed file system) [8]. As the underlying platform for storing massive amounts of data, HDFS can store massive amounts of structured and unstructured data, and is suitable for university student bursary data application scenarios. In the smart contract layer, functions such as chaining and calling of the XGBoost model are implemented. When the service layer sends a contract call request to the smart contract layer, the smart contract layer verifies the permissions and interacts with the data layer. After completing the data-related operations, the smart contract layer will return the processing result.

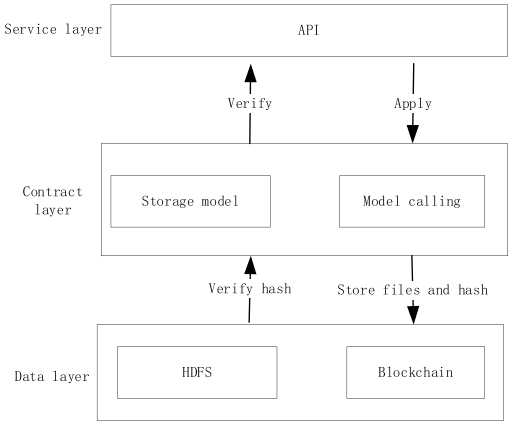


Figure 1. Blockchain-based bursary evaluation model

3.2. Smart contract design

The essence of a smart contract is that a piece of code is event-driven, using the agreement and user interface to complete the automatic execution of the contract. After the agreement is formulated and deployed, it can realize self-execution and verification without any peripherals or human intervention. In the Hyperledger Fabric platform, smart contracts are called chaincodes, which refer to application codes written in programming languages that provide state processing logic for distributed ledgers. In the smart contract layer of the system, two main functions are designed: XGBoost model store to blockchain and XGBoost model calling.

The model-on-chain contract deploys the bursary evaluation model that has achieved good results in training and testing on the blockchain. The smart contract on the chain of the model is as follows:

Algorithm 1: Deploy the model to the blockchain

Input: 1)ModelInfomation;2)TrainData;3)TestData

Output: 1)ResponseResult

```
1.InputData:PackageModel(Parameters,TrainData,TestData);
2.ArgsValidation:CheckArgs(InputData);
3.Function ChaincodeInvoke(Operation,InputData);
4.  if Operation=Storage;
5.  if (Verify(Args) ==True;
6.          StorageModel();
7.          Return Success;
8.  else;
9.          Return Error;
10.         end if
11.  end if
```

After the XGBoost model is successfully connected to the chain, the service layer can call the model by calling the model contract. The model calls the smart contract as follows:

Algorithm 2: Invoke the model from the blockchain

Input: 1)UserId;2)Password;3)Function(Args);4)StudentData

Output: 1) valuationResult

```
1. InputData: UserId, Password, Fun(Args), StundentData
2. InformationValidation:CheckArgs(InputData);
3. Function ChaincodeInvoke(Operation,InputData);
4. if Operation=ModelInvoke;
5.  if Model==Null;
6.  Return True;
7.  else;
8.  Return ModelInvokeResult;
9.  Return EvaluationResult;
10. end if
11. end if
```

4. Analysis of experimental results

The simulation experiment built a blockchain-based intelligent evaluation system for university scholarships. The experimental environment was Hyperledger Fabric version 2.1.1, Docker 19.2.1 version, Java 1.7 version, and Hadoop CDH3. Four hosts were configured, and 100 Docker nodes were set up. The experimental machine was configured as a CentOS 7.1 system, an Intel i7-5700k processor, 16 GB of memory, and JMeter 3.3 was used for stress testing.

4.1. Improved XGBoost model comparison

The basic XGBoost algorithm uses the second-order approximation method of the loss function to solve the objective function. This paper uses the basic XGBoost algorithm to solve the Taylor third-order expansion. As shown in **Figure 2**, the experiment shows that the accuracy rate [9] of the basic XGBoost model is between 70% and 80%, while in the improved XGBoost model, the accuracy rate reaches more than 80% in different test data scales. With the increase in test data, the accuracy rate of the methods shows an overall upward trend. The reason is that the improved XGBoost algorithm is based on the Taylor second-order expansion of the basic XGBoost algorithm by performing a third-order expansion to approximate the target item. Through more iterative updates, its accuracy is improved, and a better model effect is achieved.

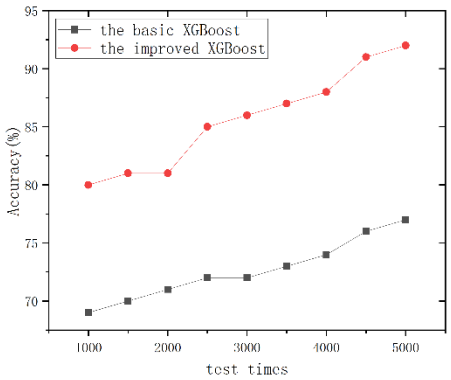


Figure 2. Comparison of XGBoost

4.2. Performance analysis

Insufficient performance is one of the challenges facing the blockchain. Throughput transactions per second (TPS) and response time are two key performance indicators that measure the blockchain system. Throughput refers to the number of transactions completed within a fixed time, and response time refers to the processing time to complete the transaction. In the TPS test, the original fabric model was compared with the model in this paper (XMB), and 50 and 100 nodes were set up for analysis and comparison, as shown in **Figures 3** and **4**. From the experimental data, when the number of nodes is 50, the number of XMB transactions per second reaches more than 500; when the number of nodes is 100, the number of XMB transactions per second reaches more than 700. No matter whether the node is 50 or 100, the TPS of XMB is higher than that of the native fabric system. The reason is that the model in this paper uses HDFS to store data in the data layer. Its advantage is that it supports massive data storage and has a large throughput; the state database that the native Fabric system can support is LevelDB and CouchDB, which use key-value pairs as storage methods. These two databases support complex queries, but their throughput performance is not as good as HDFS.

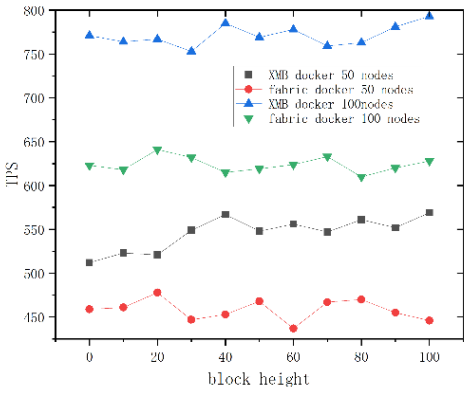


Figure 3. Comparison of TPS

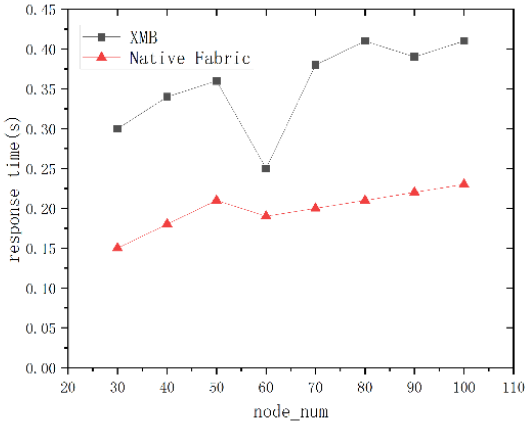


Figure 4. Response time comparison

In terms of response time, XMB has a slightly longer response time than native fabrics, because HDFS has the disadvantage of not being able to access data with lower latency [10]. But overall, XMB improves transaction processing efficiency and has less impact on time performance.

5. Conclusion

Aiming at the problems of traditional bursary evaluation, such as difficult qualification and inaccurate evaluation of university bursaries, this paper proposed a bursary evaluation model based on blockchain and machine learning algorithms, which realizes intelligent and scientific bursary evaluation. The non-tamperable and traceable basic data explores the innovative application of blockchain technology in the teaching field. Through in-depth analysis, it can be found that blockchain and artificial intelligence have a natural intergrowth and symbiosis [11]. Artificial intelligence provides optimization strategies for the core technology of blockchain, and blockchain can provide trust mechanism guarantees and infrastructure for artificial intelligence. In the future scenarios of blockchain digital infrastructure construction, most of which are faced with vertical structure scenarios such as government affairs, medical care, Internet of Things, education, etc., which need to be integrated into artificial intelligence, big data, and other fields, so artificial intelligence and blockchain are studied. The cross-integration of the company has a wide range of application prospects. The next step is to consider applying the model to other areas of university teaching and improving the data processing and optimization capabilities of the model.

Disclosure statement

The author declares no conflict of interest.

References

[1] Lu GM, Zhang Y, Zhou ZM, 2019, Research on Classification and Prediction of Poor Students Based on Machine Learning. Computer Applications and Software, 36(1): 316–319.

[2] Li MJ, 2017, Research on the Identification Method of Poverty Grants Based on Data Mining, Master's thesis,

Central China Normal University.

[3] Ding ML, Pang ST, Shang GY et al., 2020, A Blockchain-Based Method, Equipment and Medium for Evaluation and Scholarship Management, Innovation: CN111310200A.

[4] Yan XY, 2020, A Financial Aid Management Method and System Based on Blockchain Technology, Innovation: CN111797162A.

[5] Chen TQ, Guestrin C, 2016, XGBoost: A Scalable Tree Boosting System, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

[6] Torlay L, Perrone-Bertolotti M, Thomas E, et al., 2017, Machine Learning-XGBoost Analysis of Language Networks to Classify Patients with Epilepsy. Brain Informatics, 4(3): 159–169.

[7] Androulaki E, Barger A, Bortnikov V, et al., 2018, Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains, Proceedings of the 13th EuroSys Conference, 1–15.

[8] Bao GB, Yu CJ, Zhao H, 2013, Researching on the Placement of Data Replicas in the System of HDFS Cloud Storage Cluster, China Intelligent Automation Conference Proceedings 255, 259–269.

[9] Zhou JD, Yang Y, Zhang MJ, et al., 2016, Constructing ECOC Based on Confusion Matrix for Multiclass Learning Problems. Science China Information Sciences, 59(1): 135–148.

[10] Wu XY, Wang SH, Zhang YD, 2017, Summary of the Theory and Application of K-Nearest Neighbor Algorithm. Computer Engineering and Applications, 53(21): 1–7.

[11] Liu Y, Yu FR, Li X, 2020, Blockchain and Machine Learning for Communications and Networking Systems. IEEE Communications Surveys & Tutorials, 22(2): 1392–1431.

Journal of Electronic Research and Application, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Research on Low-Energy Information Transmission Based on Wireless Network

Liangkai Zhou[1]*, Dan Han[1], Nv Yang[2], Qinzhe Wang[3]

[1]Sanya Institute of Technology, Sanya 572022, Hainan, China
[2]Yum China Catering (Shenzhen) Co., Ltd., Haikou 570203, Hainan, China
[3]China National Software, Beijing 100031, China

**Author to whom correspondence should be addressed.*

Abstract: In this paper, the topological structure of the vehicle wireless network M2M (Machine to Machine) is used as the experimental research model, and four kinds of light coefficients are set as factors affecting the experimental results, namely, light intensity factor ∈ and α, to represent the light intensity coefficient and influence factor. The remaining energy consumption of mobile terminal equipment was measured respectively, the distance parameter from device to device, the maximum transmission energy consumption, and the correlation coefficient between environmental parameters and energy consumption parameters was analyzed. This paper discusses the impact of different topological structures on the environment, energy saving and emission reduction in the relatively flat terrain area, based on the planning scheme of parking area within the coverage range of base station signal, the transmission capability of vehicles as mobile device nodes within the coverage range of base station signal, and the signal coverage range of base station under different light intensity. As the distance between the base station and the vehicle mobile device node changes, the maximum transmission energy consumption of the mobile device node is obtained. Based on the above factors, the optimal performance optimization parking scheme and the optimal energy consumption optimization transmission scheme are obtained.

Keywords: Wireless network; Transmission capacity; Coverage area; Available energy consumption; Optimization scheme

Online publication: October 21, 2025

1. Introduction

The maximum energy consumption available for communication between two mobile device nodes (D2D) increases with the increase of the distance between the device node and the base station (dist). There are two factors that restrict the transmission energy consumption between the mobile device node and the base station, namely, ∈ (the maximum limit of the transmission process obstacle coefficient) and dist (the distance between the mobile device and the base station). However, the influence coefficient ∈ of the transmission process obstacle changes with the change of light intensity α. When the light coefficient α is constant, the signal transmission ability

gradually weakens with the increase in distance. In a certain area, that is, when the transmission process barrier coefficient is fixed, the better the light intensity is, the better the transmission capacity is. Among the distance base station dist and the transmission process impact factor coefficients, the most important constraints are as follows: From the above, it can be seen that at least two aspects of influence should be considered in energy consumption optimization [1].

2. Parameters and methods

2.1. Experimental parameters

The relevant parameters of the experimental data were adopted from a shared car parking space in Haikou City, Hainan Province. The data for the experiment include the base station (base station of existing shared car parking space supplier), mobile vehicle equipment (shared car Qoros 3, Qoros 5), simulation software used in the experiment (MATLAB R2016a simulation software), and the algorithm is D2D communication technology (Device-to-Device) Communication. The result set (Maximum allowed transmit power, dBm) is C1, C2, C3, C4, C5, C6, respectively. The distance between the mobile communication equipment and the base station is set as dist = 20 m, dist = 80 m, dist = 150 m. The light influencing factors are $\in = 0$ and $\in = 0.5$.

2.2. Experimental methods

2.2.1. Test preparations

On August 1, the mobile devices within the coverage of the base station will check the available initial energy consumption quantitative detection before the test, and then determine the remaining energy consumption of the mobile devices after several rounds of communication, so as to know the energy consumed in the communication process.

2.2.2. Calculation method

On August 1, August 2, and August 3, a total of three days of testing, the measurement time of 11:50, 12:10, and 12:30, three times of average energy consumption is the coefficient value 0, 18:50, 19:10, and 19:30, three times of average energy consumption coefficient 0.5, to determine the remaining energy consumption of mobile equipment.

2.2.3. Determination method

Measurement of light intensity, distance from base station, and energy consumption: After the initial energy is determined, the communication energy consumption is tested three times a week during consecutive 3d sunny days with excellent light intensity, and the results are averaged over three days. Similarly, the communication energy consumption was tested three times in the period of poor lighting conditions, and the results were averaged over three days.

2.3. Data processing

Simulation software MATLAB R2016a was used to analyze and process experimental data. Different values were set to obtain experimental result sets to form curves. Different curves could be calibrated by symbols and compared and analyzed between different marks of the same color.

3. Results and analysis

3.1. Influence of different obstacle coefficients on experimental results

As can be seen from **Figure 1**, when the distance from the base station is constant, the greater the obstacle coefficient, the smaller the maximum transmitted power of the mobile device node. It can be seen that when the distance from the base station is constant, the obstacle factor ∈ has a significant impact on the transmission power cost.

As can be seen from **Figure 2**, with the increase of obstacle coefficient ∈ (cost factor ∈) of each curve, the maximum distance between the mobile device and the base station also becomes smaller; when ∈ = 1, the value of each curve tends to zero, that is, the maximum transmission distance from the base station is zero.

As can be seen from **Figure 3**, as the distance between the mobile device and the base station increases, the influence factor of barriers in the transmission process has the greatest impact on energy consumption. The transmission power of mobile devices also increases gradually. When the distance from the nearest base station is 250 m, the transmission power values under the three curves (that is, different influence factors) are equal.

As can be seen from **Figure 4**, a scheme with ∈ = Area RL and a coverage range of 45 ± 0.5 should be adopted to optimize the energy consumption of signal transmission between the mobile device and the base station.

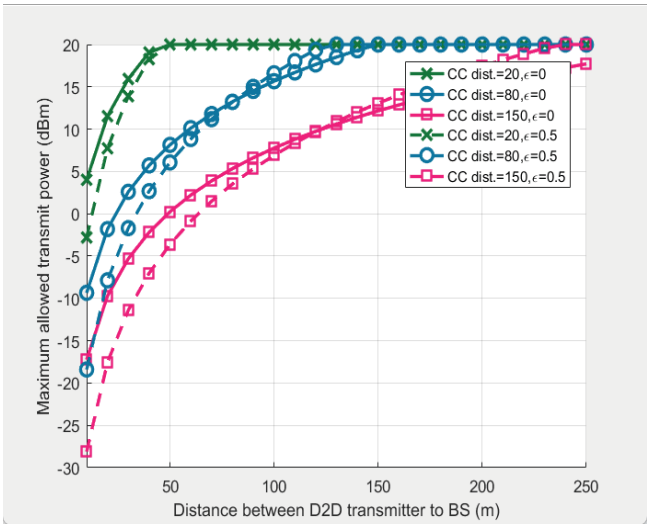


Figure 1. Influence of different barrier coefficients on the maximum transmission power

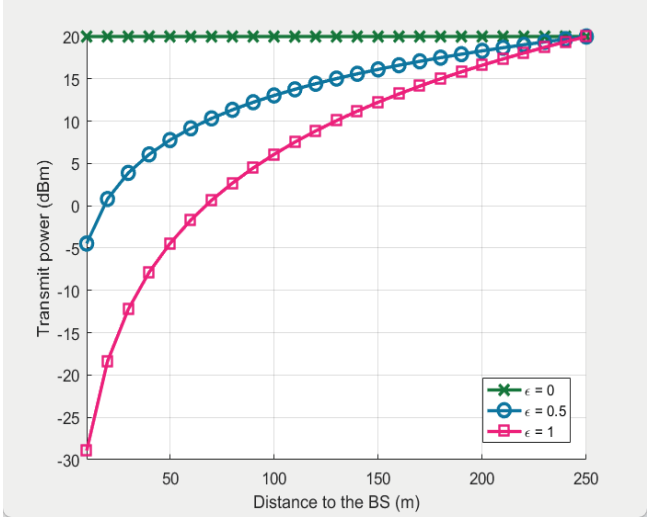


Figure 2. Influence of different light intensities on maximum transmission power

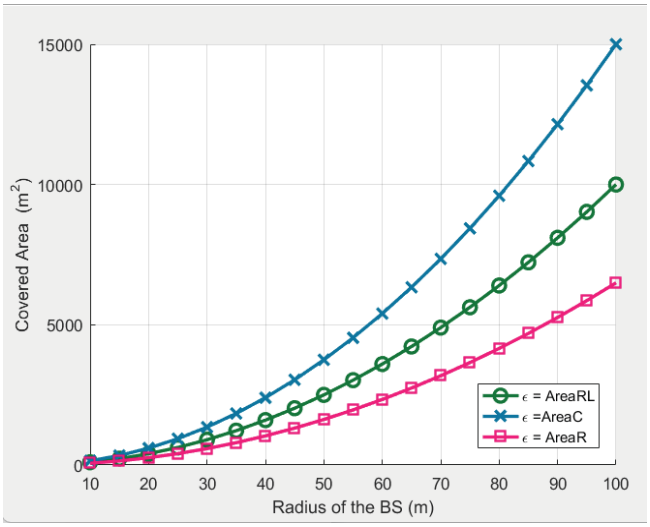


Figure 3. Different influence factors for maximum transmission power

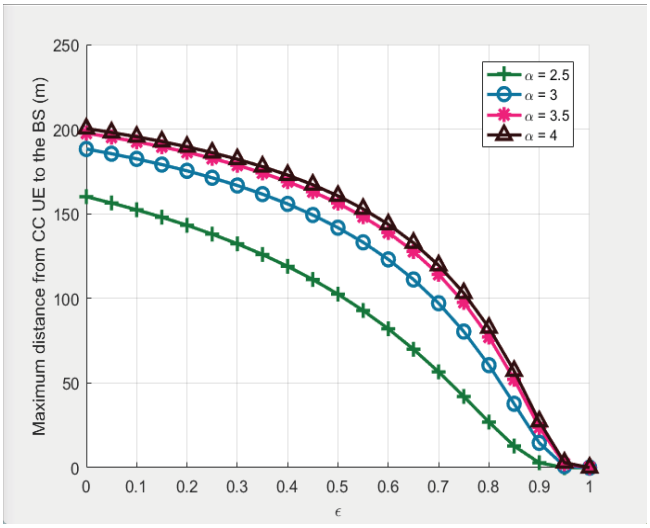


Figure 4. Effect of different base station layouts on signal coverage

3.2. Influence of different weather factors on the energy consumption of mobile devices

The obstacle factor is one of the factors that affect the energy consumption of mobile devices in the communication process of signal transmission of mobile devices. Under the influence of different obstacle factors, the transmission rate, energy consumption, output power, and input power of the three types of mobile devices are significantly different, and the obstacle factors caused by different factors are significantly different.

3.3. Influence of different lighting on the energy consumption of mobile devices

As the light intensity decreases, the energy consumption of the influencing factors in the transmission process will gradually increase. In the same type of mobile devices, different light intensity has a significant impact on the information transmission process of mobile devices; that is, the stronger the light, the faster the transmission rate, and the lower the energy consumption in the transmission process. On the contrary, the light is weak, the

transmission rate is low, and the energy consumption in the transmission process is large.

4. Discussion

In this study, the influence of different light intensities measured under different weather conditions on the transmission rate and energy consumption of different types of mobile devices during the transmission process was determined. The average light intensity in Hainan in summer is more than 1,800 $\mu mol \cdot m^2 \cdot s^{-1}$. It can be seen from the results that although different equipment types have differences in transmission rate and energy consumption, the transmission power is minimum and the energy consumption in the transmission process is maximum when the weather conditions are extremely poor, that is, the weather lighting conditions are the worst. When the light is not strong, the transmission rate is large, and the energy consumption in the transmission process is small. When the weather is clear and sunny, the transmission rate is the fastest and the energy consumption is the lowest, so the light intensity is the main factor affecting the transmission rate and energy consumption between mobile devices [2–5].

Topography is also one of the main factors affecting the transmission rate and energy consumption between mobile devices. During the transmission process of the same device, when mobile device nodes are in the low-lying area, the transmission rate between mobile device nodes is smaller than that of flat land, and the transmission energy consumption also increases, which is also one of the main reasons affecting the transmission rate and transmission energy consumption of nodes in the network [6–8].

5. Conclusion

The results of this study show that different types of mobile device nodes have certain differences in transmission speed and energy consumption. The information transmission rate and energy consumption of the same type of mobile device nodes are mainly affected by the light intensity, weather conditions, and terrain of the region, and secondly, the layout of the base station and the network coverage model are also related. That is, the terrain is relatively flat and the light intensity is good, and the network coverage is carried out in a rectangular way [9,10].

Disclosure statement

The authors declare no conflict of interest.

References

[1] Anon., 2018, Performance Analysis of Cooperative NOMA-Based Intelligent Mobile Edge Computing System. China Communications, 08: 46–52.

[2] Ciociola A, Giordano D, Vassio L, et al., 2023, Data-Driven Scalability and Profitability Analysis in Free-Floating Electric Car Sharing Systems. Information Sciences, 621: 545–561.

[3] Luo X, Ouyang J, 2020, A Low-Cost Switchable Circularly-Polarized Transmit 1024-Element Phased Array for Ka-Band SATCOM. Proceedings of the 5th International Conference on Communication, Image and Signal Processing (CCISP 2020), 1(1): 33–36.

[4] Zhang H, Zhang L, Guo Y, et al., 2022, Data Transmission Mechanism of Vehicle Networking Based on Fuzzy Comprehensive Evaluation. Open Mathematics, 1(22): 1909–1923.

[5] Ju X, Li W, 2022, Research on Real-time Road Condition Display System Based on Vehicle Network. SID Symposium Digest of Technical Papers, S1(53): 1053–1056.

[6] Hu Y, Zhang W, Zhou Z, et al., 2024, Research on Low-Energy Multipath Routing Algorithm Wireless Sensor Network. Computer and Digital Engineering, 52(02): 349–352.

[7] Ding X, Zhang W, Xu S, 2022, Research on Low Energy Consumption Routing Optimization Method in F Optic Wireless Integrated Network. Laser Journal, 43(12): 128–132.

[8] Yi Y, 2022, Research on the Localization and Time Synchronization Mechanism of Wireless Sensor Networks under Low Energy Consumption. Journal of Hunan Post and Telecommunications Vocational College, 21(03): 21–24.

[9] Yan Y, 2021, Research on Energy-Efficient Data Collection Mechanism in Wireless Sensor Networks, dissertation, Nanjing University of Posts and Telecommunications.

[10] Liu P, 2021, Design of Low-Power and High-Reliability Wireless Sensor Network Routing Protocol, dissertation, Nanjing University Aeronautics and Astronautics.

Integrated Services Platform of International Scientific Cooperation

Innoscience Research (Malaysia), which is global market oriented, was founded in 2016.Innoscience Research focuses on services based on scientific research. By cooperating with universities and scientific institutes all over the world, it performs medical researches to benefit human beings and promotes the interdisciplinary and international exchanges among researchers.

Innoscience Research covers biology, chemistry, physics and many other disciplines. It mainly focuses on the improvement of human health. It aims to promote the cooperation, exploration and exchange among researchers from different countries. By establishing platforms, Innoscience integrates the demands from different fields to realize the combination of clinical research and basic research and to accelerate and deepen the international scientific cooperation.

Cooperation Mode

Clinical Workers

In-service Doctors

Foreign Researchers

Hospital

University

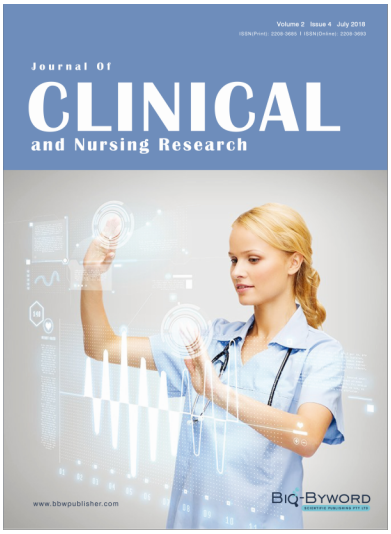
Scientific institutions

Our Journals

The *Journal of Architectural Research and Development* is an international peer-reviewed and open access journal which is devoted to establish a bridge between theory and practice in the fields of architectural and design research, urban planning and built environment research.

Topics covered but not limited to:

- Architectural design
- Architectural technology, including new technologies and energy saving technologies
- Architectural practice
- Urban planning
- Impacts of architecture on environment



Journal of Clinical and Nursing Research (JCNR) is an international, peer reviewed and open access journal that seeks to promote the development and exchange of knowledge which is directly relevant to all clinical and nursing research and practice. Articles which explore the meaning, prevention, treatment, outcome and impact of a high standard clinical and nursing practice and discipline are encouraged to be submitted as original article, review, case report, short communication and letters.

Topics covered by not limited to:

- Development of clinical and nursing research, evaluation, evidence-based practice and scientific enquiry
- Patients and family experiences of health care
- Clinical and nursing research to enhance patient safety and reduce harm to patients
- Ethics
- Clinical and Nursing history
- Medicine

Journal of Electronic Research and Application is an international, peer-reviewed and open access journal which publishes original articles, reviews, short communications, case studies and letters in the field of electronic research and application.

Topics covered but not limited to:

- Automation
- Circuit Analysis and Application
- Electric and Electronic Measurement Systems
- Electrical Engineering
- Electronic Materials
- Electronics and Communications Engineering
- Power Systems and Power Electronics
- Signal Processing
- Telecommunications Engineering
- Wireless and Mobile Communication