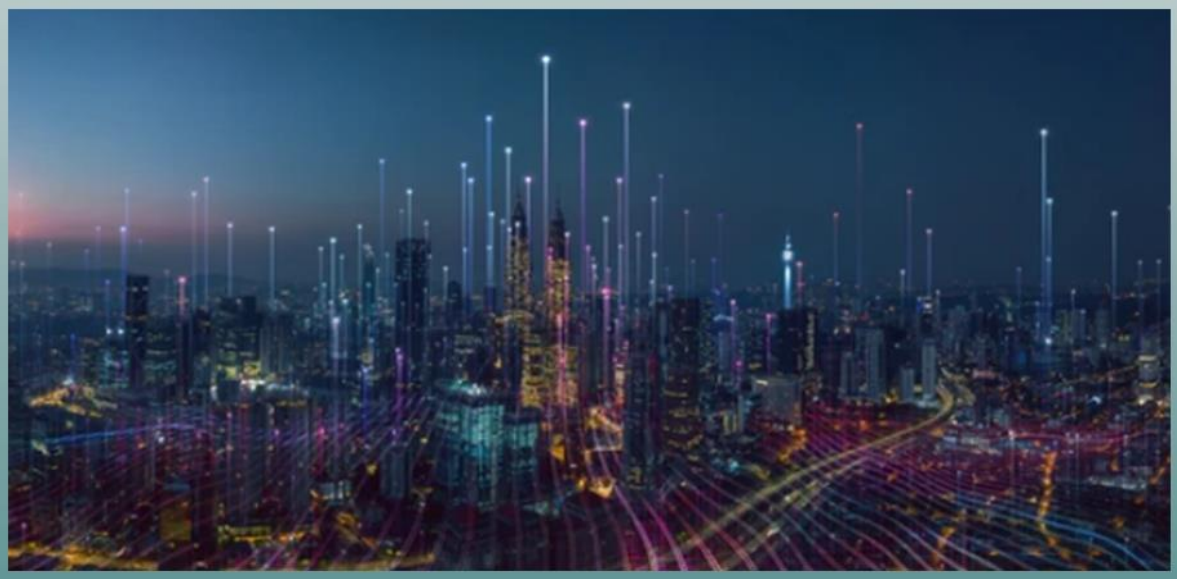
Volume 7, Issue 4, 2025

Print ISSN: 2663-1938
Online ISSN: 2663-1946

JOURNAL OF COMPUTER SCIENCE AND ELECTRICAL ENGINEERING



Journal of Computer Science and Electrical Engineering

Volume 7, Issue 4, 2025

Published by Upubscience Publisher

Journal of Computer Science and Electrical Engineering
Print ISSN: 2663-1938 Online ISSN: 2663-1946
Email: info@upubscience.com
Website: http://www.upubscience.com/

Table of Content

DESIGN AND IMPLEMENTATION OF LIGHTWEIGHT ADSORPTION UAVS 1-5
JiaYi Zeng*, ZhengYu Zhong, SiQi Wang, BoYu Xu, RunLin Chen

COGNITIVE COLLABORATION-BASED DECISION-MAKING FRAMEWORK FOR MANNED/UNMANNED AERIAL VEHICLE SYSTEMS 6-16
YiDuo Wang*, HeLin Wang, EnNing Liu, WenXuan Liu, QiLian Ge

DROUGHT RECOGNITION FOR THE SOYBEAN PLANT BASED ON LIGHTWEIGHT DEEP LEARNING MODEL 17-24
WenLiang Hou*, ZhiPeng He, JieWen Gu, JiaLin Zhang, JiaoYing Li

EXPLORING THE PATH OF VALUE NETWORK EMPOWERING THE TRANSFORMATION OF UNIVERSITY TEACHING EVALUATION 25-28
Chu Sun

FOREST FIRE POINT RECOGNITION BASED ON SUPER-RESOLUTION TECHNIQUES 29-36
FuLin Li, WenFa Xu*, Zhen Min, XuePeng Wu, ChangYu Xiang, YuDong Liu

DESIGN OF A MICROCONTROLLER-BASED SMART BLIND CANE 37-42
WenHao He*, GuangBo Su, ZhiJie Wang, ShuMing Li

OPTIMIZATION STRATEGIES FOR WIRELESS COMMUNICATION NETWORKS BASED ON MACHINE LEARNING 43-46
ZiXuan Wu

SELECTION STRATEGY FOR FULL-SPECTRUM VERSUS LOCAL-SPECTRUM ENERGY DETECTION UNDER CONSTANT FALSE ALARM RATE CONDITIONS 47-59
YuXin Cheng

ENHANCING THE YOLOv11 MODEL FOR TEACHING BEHAVIOR RECOGNITION 60-65
Yao Tian, Cheng Peng*

TFE-NET: A TINY-AWARE FEATURE ENHANCEMENT NETWORK FOR COLLABORATIVE OPTIMIZATION IN SMALL OBJECT DETECTION 66-73
Qiang Zeng*, YiDan Chen

Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3061

DESIGN AND IMPLEMENTATION OF LIGHTWEIGHT ADSORPTION UAVS

JiaYi Zeng*, ZhengYu Zhong, SiQi Wang, BoYu Xu, RunLin Chen
School of Aviation, Beijing Institute of Technology, Zhuhai 519000, Guangdong, China.
Corresponding Author: JiaYi Zeng, Email:18219292674@163.com

Abstract: The increasing demand for intelligent monitoring in engineering and construction has highlighted several limitations within current Unmanned Aerial Vehicle (UAV) technologies. These challenges include inadequate flight endurance, restricted payload capacity, and a reliance on continuous manual control—particularly evident in tasks requiring reliable adhesion and autonomous operation in the assessment of vertical structures and high-altitude hazardous environments. To mitigate these issues, the present project proposes the development of a lightweight adsorption-capable UAV, designed to achieve attitude reversal during flight and to adhere to both horizontal and vertical surfaces of structures for the purposes of site surveying, monitoring, and data collection. This research details the design and production of a functional prototype of the adsorption UAV, which addresses the aforementioned deficiencies by enhancing endurance, increasing payload capability, and reducing dependence on manual intervention. This innovative solution offers significant advancements in automated monitoring capabilities within complex operational environments.
Keywords: Adsorption; Unmanned aerial vehicles; Flight control; Lightweight design

1 INTRODUCTION

The rapid advancement of drone technology has catalyzed its deployment in engineering construction and restoration, resulting in a multi-dimensional operational framework. In the construction domain, drones serve various functions, including site surveys, 3D modeling, and real-time monitoring. Furthermore, they are equipped with infrared sensors to facilitate nighttime inspections and can undertake material transportation tasks, thereby replacing traditional high-risk manual operations. In challenging terrain scenarios, such as water conservancy and ecological restoration, drones harness their high-precision aerial surveying capabilities to tackle the limitations inherent in conventional auditing and monitoring practices, which often suffer from difficulties in evidence collection and incomplete coverage. For instance, in a water environment remediation initiative, drones enabled accurate auditing of a significant investment project valued at 3.466 billion yuan. Nonetheless, the current technological landscape encounters two principal bottlenecks. Firstly, there are inherent limitations regarding flight endurance and payload capacity. Mainstream industrial-grade drones typically exhibit flight endurance ranging from 30 to 60 minutes, with individual units costing between 100,000 and 300,000 yuan. This presents challenges in meeting the continuous operational demands of extensive projects. Furthermore, the lack of breakthroughs in battery technology necessitates frequent replacements, while environmental factors such as wind and precipitation further impede operational efficiency. Secondly, there exists a pronounced reliance on human operators. Although some applications have achieved levels of semi-automation, complex tasks continue to necessitate real-time control by skilled drone pilots. Data indicate that in circumstances such as high-altitude curtain wall cleaning, human operational errors account for approximately 35% of high-altitude accidents. Compounding this issue is the shortage of qualified drone pilots, which escalates labor costs. Collectively, these challenges not only complicate project management but also constrain the scalable application of drones within long-term engineering projects.

In recent years, extensive research on drone - related technologies has been carried out globally. The research scope encompasses various aspects such as drone technology itself, its applications in different fields, and relevant algorithms. Regarding drone technology, Vergouw et al. provided a comprehensive overview of drone technology [1], covering types, payloads, applications, frequency spectrum issues, and future developments. They also explored the opportunities and threats of drone use from ethical and legal perspectives. Budiharto et al. focused on mapping and 3D modelling using quadrotor drones and GIS software [2], which is of great significance for applications like surveying and 3D reconstruction. In terms of sensor technologies on drones, Guo et al. reviewed the progress, challenges [3], and trends of vision sensing technologies in automatic/intelligent robotic welding. This has implications for drones equipped with vision - based sensors for tasks such as inspection and mapping. Deliry and Avdan studied the accuracy of unmanned aerial systems photogrammetry and structure from motion in surveying and mapping [4], which is crucial for applications requiring high - precision mapping, like topographic surveying.For the flight and control aspects of drones, Oosedo et al. investigated the flight control systems of a quad tilt rotor unmanned aerial vehicle for large attitude changes [5]. This research is beneficial for improving the stability and maneuverability of drones in complex flight conditions.Regarding algorithms, Grisetti et al. introduced a tutorial on graph - based Simultaneous Localization and Mapping (SLAM) [6]. SLAM technology is fundamental for drones to navigate and map their environment autonomously. Chetverikov et al. proposed the trimmed Iterative Closest Point(ICP) algorithm [7], which is useful for tasks such as point - cloud registration in drone - based 3D modelling. In the security and communication fields of

drones. However, challenges still remain, such as further improving the accuracy and reliability of sensors, optimizing flight control algorithms for more complex environments, and strengthening the security of drone communication. Future research could focus on integrating multiple technologies to enhance the overall performance of drones and exploring more innovative application scenarios.

This paper presents a novel lightweight adsorption UAV system design premised on a comprehensive analysis of existing research by scholars both domestically and internationally. The proposed system integrates advanced map-assisted positioning via SLAM technology to enhance the precision of locating adsorption points. The UAV is equipped with intelligent suction cups and a negative-pressure adsorption mechanism that interfaces with a ground control station and flight control operation code. This integration facilitates adjustments in flight attitude, enabling the UAV to execute horizontal and vertical adhesions on various engineering and construction surfaces. The overarching objective of this research is to bolster productivity in engineering construction while simultaneously advancing sustainable development, thereby advocating for environmentally responsible practices within the industry.

2 CONSTRUCTION OF THE MODEL STRUCTURAL FRAMEWORK

2.1 Design Principles for Lightweight Adsorption-Based Unmanned Structures

The paper presents a comprehensive design framework for a lightweight adsorption unmanned aerial vehicle [8], addressing critical factors such as weight, performance, and endurance in airframe design. The overall structural framework of the UAV is meticulously developed to optimize these parameters. The UAV system comprises several integral components, including a ground control station, a visual sensing unit, a flight control system, and an advanced computational algorithm. The visual sensor serves a pivotal role in determining the UAV's position during operational tasks. Communication between the UAV and the ground control station is facilitated through the MAVLink protocol [9], allowing for seamless data exchange. Furthermore, the flight control system is engineered to autonomously manage attitude stabilization and trajectory adjustments, thereby enhancing the UAV's operational efficacy.

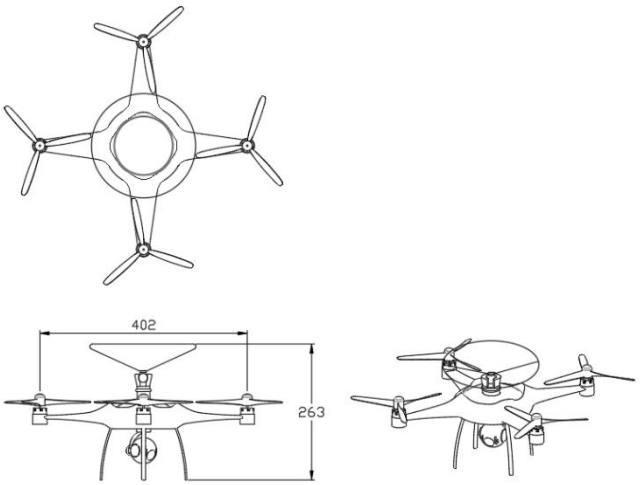


Figure 1 Overall Structural Design of the Unmanned Aerial Vehicle

The design of the unmanned aerial vehicle structure was conducted utilizing SolidWorks, and the resulting three-dimensional model is illustrated in Figure 1. The UAV structure comprises five integral components: the visual sensing system, flight control system, basic fuselage, suction cup control device, and electronic control system. Additionally, a servo-driven tilting mechanism has been developed at the terminal end of the suction cup robotic arm. This suction cup system is interconnected with the flight control system to receive PWM signals, which facilitate the adjustment of the arm's angle of rotation.

2.2 Adsorption of Unmanned Aerial Vehicles under Diverse Complex Conditions

In this study, we examine the characteristics of unmanned aerial vehicle adsorption systems by synthesizing existing literature and analyzing a variety of empirical data. The focus is placed on key factors influencing the adsorption process, including the adsorption mode and the implementation of adaptive control algorithms. The UAV is equipped with an intelligent suction cup designed to accurately detect the adsorption state, allowing for the effective planning of the UAV's movement trajectory while minimizing exposure to unfavorable adsorption conditions. Utilizing flow field theory as a foundational framework, we conduct a comprehensive modeling analysis of the adsorption system's flow dynamics. Moreover, we investigate the parameters that impact adsorption performance through the lens of structural optimization. The objective is to develop an integrated strategy aimed at enhancing the overall efficiency and efficacy

of UAV adsorption performance. This research contributes to the advancement of UAV applications where precise adhesion capabilities are critical.

3 ALGORITHMS AND CONTROL

3.1 Design Principles for Lightweight Adsorption-Based Unmanned Structures

In the context of UAV localization, representative feature points, such as corner points and edge points, are meticulously extracted from captured images. Subsequent to this extraction, feature point matching is conducted against a pre - established map to ascertain the UAV's position and orientation. Specific markers or distinctive feature patterns are strategically positioned on the target surfaces, such as building walls. The UAV employs onboard cameras to acquire images, thereby computing its own position and orientation relative to these targets by analyzing the positional variations of the identified feature points within the images. To enhance the robustness of feature extraction across varying lighting conditions and viewing angles, algorithms such as Scale - Invariant Feature Transform (SIFT) and Speeded - Up Robust Features (SURF) are utilized. During the matching process[10-11], the similarity between feature points is evaluated using descriptor - based matching techniques, facilitating the identification of corresponding feature point pairs. The UAV's position and orientation are subsequently derived through triangulation methods, leveraging the geometric relationships among the matched point pairs. While in flight, the UAV employs LIDAR technology to systematically scan and analyze its surrounding environment, integrating this capability with Simultaneous Localization and Mapping (SLAM) methodologies[12-13]. In LIDAR-based SLAM systems, the laser beam emitted by the LIDAR system creates reflections on the surface of an object, and the time-of-flight of these reflections yields distance measurements to construct a point cloud representation of the environment. Concurrently, the ICP algorithm is implemented to match the real - time acquired laser point cloud data and image features with previously mapped information [14]. Through iterative optimization of these matching outcomes, the UAV's position within the map can be accurately determined, thus facilitating precise localization.

3.2 Adaptive Control (MRAC) Algorithms and Optimisation

In the domain of high - precision positioning for unmanned aerial vehicles, the performance of flight and adsorption can be significantly influenced by various environmental factors, such as airflow dynamics and variations in surface roughness. To address these challenges, an adaptive control algorithm has been proposed, which involves the development of a reference model that characterizes the desired dynamic behavior of the UAV. This model serves as a benchmark against which the actual system's performance is evaluated. By continuously monitoring the discrepancy between the actual outputs and those of the reference model, the control parameters can be dynamically adjusted, thereby ensuring that the UAV maintains a high degree of accuracy in locating and adhering to the target surface, even in the presence of environmental fluctuations and system uncertainties. The proposed control architecture comprises a feed - forward controller coupled with a feedback controller, allowing for real - time adjustment based on the adaptive rate. The reference model functions as an idealized control system, delineating the desired performance metrics essential for optimizing the system's characteristics, such as overshoot and damping time. The adaptive rate plays a crucial role in minimizing the error between the actual outputs and the prescribed outputs of the reference model. This mechanism facilitates the modification of controller parameters or the generation of auxiliary inputs, thereby enhancing the robustness and resilience of the UAV's positioning and adsorption capabilities under varying operational conditions.

3.3 Seal Cell Flow Field Modelling

The reduction of leakage and the enhancement of adsorption efficacy can be achieved by optimizing the sealing design of suction cups. This optimization allows for more effective control of negative pressure within the system, thereby amplifying the adsorption force while minimizing the potential for air escape. Consequently, the modeling of the flow field associated with the sealing unit establishes a theoretical framework for improving adsorption capacity. Given that the slit between the sealing unit and the adsorption wall is notably narrow, it can be approximated as a planar surface. Furthermore, the airflow in this region can be characterized as laminar, attributable to the minimal height of the slit and the resultant low Reynolds number. It is also reasonable to assume that the partial velocities of the airflow along the Z - axis can be considered negligible.

$$q_{V1} = \frac{lh^3}{12\mu} \cdot \frac{p_0 - p_1}{x_0 - x_1} = \frac{lh^3}{12\mu} \cdot \frac{\Delta p}{B} \tag{1}$$

In the formula (1), q_{v_1} represents the volumetric flow rate, which is the volume of fluid passing through a certain cross - section per unit time and is a key indicator for measuring the gas leakage amount. l is the length, often referring to the characteristic length related to the sealing unit. h indicates the height, generally being the gap height between the sealing unit and the adsorption wall. μ is the dynamic viscosity, reflecting the fluid's ability to resist flow deformation. p_0 and p_1 are the pressures at different positions respectively, and their difference Δp, namely the pressure difference, serves as the driving force for gas flow. x_0 and x_1 are position coordinates. B represents a characteristic quantity related to flow in different contexts, such as the gap width. This formula is used to quantitatively describe the gas

leakage dynamics of the sealing unit. By correlating various parameters, it calculates the volumetric flow rate of gas leakage, which is of great significance for analyzing the performance of the sealing system and optimizing the sealing design.

The dynamics of gas leakage from a sealing unit can be quantitatively described through an established mathematical framework. According to Equation 1, the gas leakage—representative of power loss—exhibits a cubic dependency on the gap height, the seal circumference, and the pressure differential that generates the adsorption force. Conversely, the leakage is inversely related to the seal width. To further investigate the influence of specific design parameters on gas leakage, we analyze the effects of varying gas heights, specifically at 0.4 mm and 1 mm. This analysis facilitates an understanding of the relationship between gas leakage and the width of the sealing ring. It is observed that an increase in gap height significantly impacts gas leakage. Conversely, at a smaller gap height of 0.4 mm, variations in seal width and pressure differential exert a minimal influence on leakage rates. Moreover, the design constraints imposed by the overall dimensions of unmanned aerial vehicles restrict the possibility of arbitrarily increasing seal width to mitigate gas leakage. Such an increase could result in a disproportionate escalation of UAV weight, which does not necessarily correlate with enhanced adsorption safety. Thus, it becomes imperative to minimize the leakage gap during the design phase of the sealing unit. In scenarios where the seal width is predetermined, an increase in adsorption necessitates a corresponding rise in air leakage. Therefore, it is essential to upgrade centrifugal pumps to facilitate efficient gas removal, thereby sustaining a low negative pressure environment. By reducing the leakage gap within the sealing mechanism, gas leakage diminishes significantly, thereby enhancing the adsorption capacity of the negative pressure adsorption system.

In addition, the research on the flow field and sealing performance also refers to the work of Chen et al.[15], who conducted in - depth analysis on the flow field and sealing performance of compliant foil face gas seals, providing valuable insights for optimizing the sealing design of the adsorption system.

4 CONCLUSION

This project involves the design and preliminary production of a lightweight adsorption unmanned aerial vehicle prototype, addressing the prevalent challenges of limited endurance, inadequate load capacity, and restricted maneuverability observed in existing UAVs utilized within the engineering and construction sectors through a comprehensive and innovative design approach. In the realm of adsorption technology, advanced high-precision positioning techniques are employed, which include feature point matching, map-assisted positioning, and the optimization of adaptive control algorithms. These methodologies facilitate the precise positioning and adsorption of the UAV onto targeted surfaces in complex environments. Additionally, a thorough modeling analysis of the flow field associated with the intelligent suction cup sealing unit elucidates the impact of various structural parameters on adsorption performance, thereby providing a theoretical foundation for the design of the sealing unit and contributing to enhancements in both adsorption stability and efficiency. The lightweight adsorptive UAV exhibits considerable application potential within the engineering construction domain, enabling rapid site surveys, comprehensive multi-angle monitoring, and prompt data collection that supports engineering endeavors while ensuring the safety and quality of construction activities. Furthermore, this technology mitigates the need for personnel to enter hazardous work zones, thereby enhancing operational safety. From an environmental perspective, it aids in the timely identification of construction impacts on surroundings, advocates for sustainable construction practices, and further promotes the sustainable development of the engineering construction industry. Despite the advancements achieved, several areas warrant further improvement. Regarding endurance, while the adsorption design contributes to reduced energy consumption, further exploration into alternative energy sources or the optimization of energy management systems is necessary to extend the UAV's operational duration. To enhance load capacity, focus on material selection and structural optimization is required to increase load-bearing capabilities without compromising the UAV's lightweight design. Additionally, although multiple strategies have been implemented to address environmental interference, further enhancement is needed in the stability and reliability of UAVs under extreme weather conditions and strong electromagnetic interference. AS technology advances and research deepens, the lightweight adsorption UAV is anticipated to be applied across a broader spectrum of fields, including bridge inspections, high-altitude building maintenance, and logistics distribution. Through ongoing technological innovation and optimization, this UAV is poised to deliver greater efficiency and convenience across various industries, emerging as a pivotal force in driving industry progress.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This research is funded by the Guangdong Province College Students' Innovation and Entrepreneurship Training Program. Project Title: Design and Implementation of Lightweight Adsorption UAVs, Project Number: XJDC201.

REFERENCES

[1] Vergouw B, Nagel H, Bondt G, et al. Drone technology: Types, payloads, applications, frequency spectrum issues and future developments. The future of drone use: Opportunities and threats from ethical and legal perspectives, 2016: 21-45.
[2] Budiharto W, Irwansyah E, Suroso J S, et al. Mapping and 3D modelling using quadrotor drone and GIS software. Journal of Big Data, 2021, 8: 1-12.
[3] Guo Q, Yang Z, Xu J, et al. Progress, challenges and trends on vision sensing technologies in automatic/intelligent robotic welding: State-of-the-art review. Robotics and computer-integrated manufacturing, 2024, 89: 102767.
[4] Deliry S I, Avdan U. Accuracy of unmanned aerial systems photogrammetry and structure from motion in surveying and mapping: a review. Journal of the Indian Society of Remote Sensing, 2021, 49(8): 1997-2017.
[5] Oosedo A, Abiko S, Narasaki S, et al. Flight control systems of a quad tilt rotor unmanned aerial vehicle for a large attitude change//2015 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2015: 2326-2331.
[6] Grisetti G, Kümmerle R, Stachniss C, et al. A tutorial on graph-based SLAM. IEEE Intelligent Transportation Systems Magazine, 2010, 2(4): 31-43.
[7] Chetverikov D, Svirko D, Stepanov D, et al. The trimmed iterative closest point algorithm//2002 International Conference on Pattern Recognition. IEEE, 2002, 3: 545-548.
[8] Tian J M. Structural Design and Research of a Vacuum Adsorption-Based Wall Cleaning Robot. Southwest Jiaotong University, 2013.
[9] Lai Q S. Design of an Unmanned Aerial Vehicle System Based on the MAVLink Protocol. Hangzhou Dianzi University, 2017.
[10] Qin C. Drone Image Matching Point Cloud Filtering Processing and 3D Reconstruction. Southwest Jiaotong University, 2015.
[11] Dong W, Mao G, Zhang H D, et al. Research on Small Deformation Initial Value Estimation Based on SIFT Feature Matching. Laser Technology, 2025: 1-12. http://kns.cnki.net/kcms/detail/51.1125.TN.20250407.1242.002.html.
[12] Wang Z Z, Zhang S, Ning C, et al. An Image Feature Point Matching Algorithm Based on Improved SURF. Hebei Industrial Science and Technology, 2024, 41(06): 418-425.
[13] Hu Q S, Li J W, Zhang Y S, et al. IMU and LiDAR Fusion SLAM Technology for Unmanned Mining Vehicles. Mining and Industrial Automation, 2024, 50(10): 21-28. DOI: 10.13272/j.issn.1671-251x.18209.
[14] Dai J L, Chen Z Y, Ye X Z. Application of the ICP Algorithm in Point Cloud Registration . Journal of Image and Graphics, 2007(03): 517-521.
[15] Chen Y, Wang Q, Peng X, et al. Flow field and sealing performance analysis of compliant foil face gas seal. Advances in Mechanical Engineering, 2022, 14(6): 1687132221108488.

Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3062

COGNITIVE COLLABORATION-BASED DECISION-MAKING FRAMEWORK FOR MANNED/UNMANNED AERIAL VEHICLE SYSTEMS

YiDuo Wang*, HeLin Wang, EnNing Liu, WenXuan Liu, QiLian Ge
School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, Liaoning, China.
Corresponding Author: YiDuo Wang, Email: yw542@student.le.ac.uk

Abstract: This paper investigates a cognitive collaboration-based decision-making framework for manned/unmanned systems, aiming to address the limitations of traditional methods in situation assessment, task allocation, and path planning. Firstly, a dual layer coupled situation and threat assessment model is constructed using Dynamic Intuitionistic Fuzzy Cognitive Maps (DIFCM) and Genetic Algorithms (GA), achieving collaborative optimization of global situation inference and local threat quantification. Secondly, an adaptive task allocation mechanism is designed by integrating an improved Contract Net Protocol with UAV intelligent emotional modes, effectively balancing task execution efficiency and resource utilization. Finally, an emotion-driven improved A* algorithm is introduced, enhancing the adaptability and safety of path planning in dynamic threat environments through dynamic threat avoidance radii and cognitive load feedback mechanisms. Simulation experiments demonstrate that the proposed algorithm improves replanning response time by 23% compared to traditional path planning algorithms under sudden threats, while reducing path threat costs by 58%. The research outcomes provide new theoretical exploration and practical references for manned/unmanned collaborative tasking and decision-making in intelligent mission scenario, aiming to advance the theoretical and practical development of deep human-machine intelligence integration.
Keywords: Manned/Unmanned collaborative combat; DIFCM; Manned aircraft cognitive load; Improved A* algorithm; UAV intelligent emotional mode; Genetic algorithm

1 INTRODUCTION

Future strategic scenarios are becoming increasingly intricate, making independent operations by either humans or machines insufficient. Effective collaboration between humans and unmanned systems, combining human judgment with robotic accuracy, is key to advanced intelligent strategies. Programs like "Loyal Wingman" and Su-57–"Okhotnik" validate its effectiveness [1]. This highlights the need for a full-process framework covering "situational awareness → threat assessment → mission decision → path planning." Key challenges include task allocation, interaction mechanisms, and performance evaluation, while current research overlooks UAV attributes, environmental factors, and cross-level cognitive interaction. Seamlessly integrating human cognition with machine autonomy remains crucial for enhancing collaborative work effectiveness.

Research on core technical systems for collaborative decision-making includes several key studies. Fu reviews U.S. programs like the Software-Enabled Control Plan and the Common Architecture for Manned-Unmanned Systems [2], which facilitate mission planning, decision-making, and standardized communication between manned aircraft and UAVs. In contrast, domestic research remains largely theoretical. Wu models collaborative air strategic scenarios as a partially observable Markov decision process (POMDP), employing distributed training to address environmental non-stationarity and centralized training to mitigate computational challenges [3], suitable for heterogeneous multi-aircraft formations. Xie introduces a direction-finding cross-target localization model, proposing variable-curvature Dubins curves for low-speed maneuvering targets and an optimal control model with penalty function-based trajectory planning for high-speed targets [4], achieving localization accuracy improvements of 36.9% and 23.5% in two scenarios. In collaborative task decision-making, Zhong presents a hybrid fuzzy cognitive map (HFCM) decision method [5], establishing an interactive decision framework with intervention strategies derived from RBFCM and IFCM. Xue builds a decision requirement reasoning model using fuzzy grey cognitive maps (FGCM) and particle swarm optimization (PSO) to refine weight learning [6], enabling rapid task selection. Liu proposes an ACO-A* hybrid path planning algorithm with a dynamic grid environment and k-means clustering for improved UAV collaborate maneuver decision-making [7]. Lastly, Gu outlines international advancements in manned/unmanned teaming [1], proposing integration architectures and future research directions in collaborative control.

Existing research often focuses on isolated decision-making aspects, lacking a comprehensive framework integrating "situational awareness → threat assessment → task allocation." This study optimizes decision-making through cognitive collaboration, advancing human-machine integration [8]. By combining dynamic intuitionistic fuzzy cognitive maps, genetic algorithms, and emotional adaptive mechanisms, it constructs a full-process model while incorporating a cognitive load quantification module. A data-knowledge dual-driven approach refines weight learning, and an emotional state transition matrix enables UAVs to adapt autonomously. Unlike prior work, this framework couples human cognition with machine autonomy, validated through simulation. It enhances adaptability via bidirectional human-machine interaction and improves coherence across decision-making stages, addressing challenges like fragmented decision logic and response delays.

2 COLLABORATIVE DECISIONING MODEL

2.1 Collaborative Decision-Making Mechanism

The core of the manned/unmanned collaborative system is a "human cognition-led, machine intelligence-enhanced" decision-making framework. By integrating manned aircraft control with UAV autonomy, complementary collaborative effectiveness is achieved. The heterogeneous nature of these systems provides significant advantages but also introduces challenges: manned aircraft, with greater payload and tactical flexibility, can perform critical tasks, but their survival requirements and performance differences complicate decision-making. The dynamic authority allocation mechanism optimizes decision-making and task execution by assessing operator workload and task attributes in real time. Figure 1 illustrates a typical scenario: the manned aircraft, as mission commander, sets objectives and strategies, while the UAV performs reconnaissance and tasking based on real-time data, feeding task status back. Dynamic authority allocation and closed-loop control enable efficient task handling in complex environments [9].

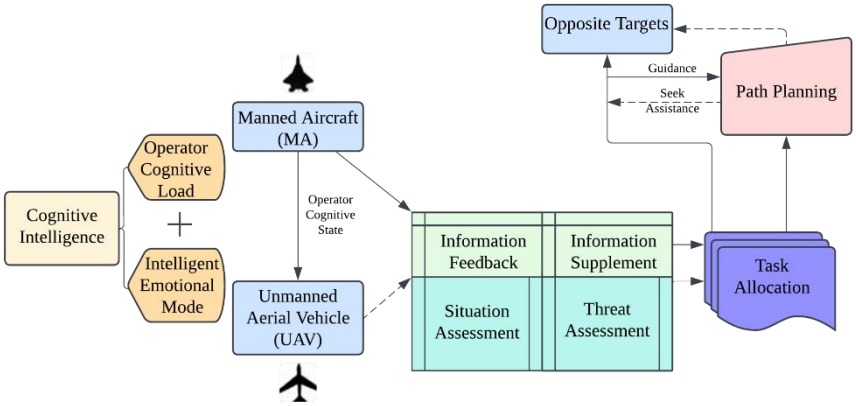


Figure 1 Task Scenario of Manned/Unmanned Collaborative Decision-Making

2.2 Situation and Threat Assessment Model

The modeling of task situation and threat assessment is the core cognitive layer of the human-machine collaborative decision-making system. Its essence lies in constructing a hybrid reasoning framework that integrates human experiential knowledge with machine computational capabilities. Based on intuitionistic fuzzy cognitive map theory, this paper proposes a dual-layer coupled assessment architecture: the first layer employs Dynamic Intuitionistic Fuzzy Cognitive Maps (DIFCM) to achieve global situation evolution inference, while the second layer utilizes a Genetic Algorithm-enhanced Threat Assessment Network (GATAN) to precisely quantify local threats. The two layers form a synergistic enhancement effect through shared key state variables and weight adjustment mechanisms, reducing model complexity while improving adaptability to dynamic mission scenarios.

2.2.1 DIFCM

As a key tool for complex system modeling, cognitive map theory has been widely applied in social network analysis and economic system simulation since Axelrod's foundational work [10]. However, traditional cognitive maps face two major limitations when addressing high-dimensional uncertainties in modern systems like highly dynamic mission perception networks: (1) reliance on binary causal reasoning based on Boolean logic [11], which restricts the representation of fuzzy relationships, and (2) the absence of dynamic topology optimization, leading to rigid knowledge structures. To overcome these challenges, Iván S introduced Fuzzy Cognitive Maps (FCMs) [12], achieving three breakthroughs: (I) integrating fuzzy membership functions for continuous-value causal reasoning, (II) developing concept node activation models using nonlinear transfer functions, and (III) optimizing weight matrices to enable dynamic topology evolution. Mathematically, FCMs can be modeled as weighted directed graphs $FCM=(G,E,W)$, where the concept node set $C=\{C_i\}_{i=1}^{n}$ represents key system elements (e.g., threat levels), and the directed edge set $E\subseteq C\times C$ describes causal influence paths between nodes. In the weight matrix:

$$W=\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n_c} \\ w_{21} & w_{22} & \cdots & w_{2n_c} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n_c1} & w_{n_c2} & \cdots & w_{n_cn_c} \end{bmatrix}, \tag{1}$$

$w_{ij}(j=1,2,...,n)$ indicates the degree and direction of the influence of concept node C_i on C_j with $w_{ij} \in [-1,1]$. When there is no causal relationship between C_i and C_j, or when $i=j$, $w_{ij}=0$ holds. By introducing intuitionistic fuzziness to node information and adjacency weight matrices and replacing the weighted summation and threshold functions in traditional FCM models with intuitionistic fuzzy ordered weighted averaging (IFOWA) operators, the Intuitionistic Fuzzy Cognitive Map (IFCM) model $IFOWA=(C,w_i,OWA_e)$ is obtained. The values of node state information and inter-node association information lie within $[0,1]$. Defining the relevant weight vector of the IFOWA operator as $e=(e_1,e_2,...,e_n)^T$, where $e_j \in [0,1], j=1,2,...,n$, the IFCM reasoning process is as follows. Let the influence of node $C_i(i=1,2,...,n)$ on node C_j through directed arcs at time t be:

When $i=j$,

$$\begin{aligned} r_{ij}(t) &=< \mu_{r_{ij}(t)}, \nu_{r_{ij}(t)} >= C_i(t) \otimes w_{ij}(t) =< \mu_{C_i(t)}, \nu_{C_i(t)} > \otimes < \mu_{w_{ij}(t)}, \nu_{w_{ij}(t)} > \\ &=< \mu_{C_i(t)}\mu_{w_{ij}(t)}, \nu_{C_i(t)} + \nu_{w_{ij}(t)} - \nu_{C_i(t)}\nu_{w_{ij}(t)} > \end{aligned} \tag{2}$$

When $i \neq j$, $r_{ij}(t)=C_j(t)$, and the value of node $C_j(t)$ at time $t+1$ is

$$\begin{aligned} c_j(t+1) &= IFOWA[r_{1j}(t), r_{2j}(t), ..., r_{nj}(t)] = e_1 r_{\sigma(1)j}(t) \oplus e_2 r_{\sigma(2)j}(t) \oplus ... \oplus e_n r_{\sigma(n)j}(t) \\ &= \left\langle 1 - \prod_{k=1}^{n} (1-\mu_{r_{\sigma(k)j}(t)})^{e_{kj}}, \prod_{k=1}^{n} \nu_{r_{\sigma(k)j}(t)}^{e_{kj}} \right\rangle = \left[1 - \prod_{k=1}^{n} (1-\mu_{r_{\sigma(k)j}(t)})^{e_{kj}}, 1 - \prod_{k=1}^{n} \nu_{r_{\sigma(k)j}(t)}^{e_{kj}} \right] \end{aligned} \tag{3}$$

In engineering implementation, IFCM functions as a state propagation system based on fuzzy logic, following the steps of initializing concept node states, iteratively updating via weighted causal networks, and stabilizing outputs. Initially, environmental states (e.g., threat levels, resource distribution) are loaded, and states are propagated using intuitionistic fuzzy operators until fluctuations fall within a predefined threshold, leading to stable decision outputs. Unlike traditional cognitive maps, which offer static conclusions, the DIFCM model incorporates time-series variables and inter-node associations, enhancing real-time adaptability and memory.

2.2.2 Genetic Algorithm

Evolutionary algorithms, inspired by biological evolution, are population-based optimization methods that efficiently address complex scenario decision-making by simulating natural selection and genetic variation. Their core mechanism involves dynamically evolving populations within the solution space, iteratively refining solutions through selection, recombination, and mutation as shown in Figure 2.

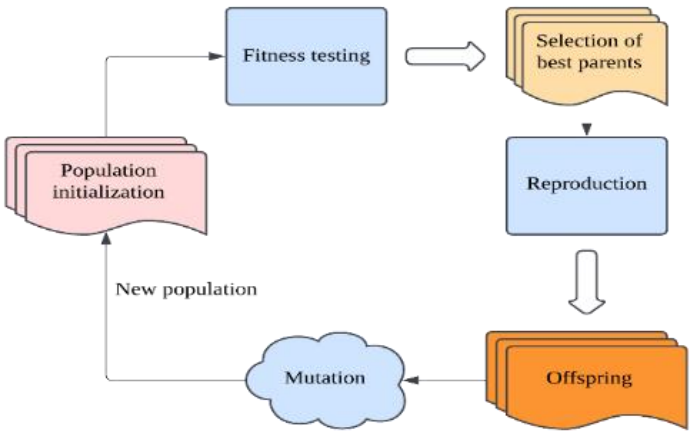


Figure 2 Evolutionary Learning Algorithm

Current mainstream branches of evolutionary algorithms include Genetic Algorithms (GA), Ant Colony Optimization (ACO), Asexual Reproduction Optimization (ARO), and the Jaya algorithm. Each variant exhibits differentiated characteristics in the task of Fuzzy Cognitive Map (FCM) weight learning as shown in Table 1:

Table 1 Property of Evolution Algorithms

Algorithm	Core Operation	FCM Features	Limitations
GA	Selection/Crossover/Mutation	Strong global search capability	Convergence speed sensitive to parameters
Ant Colony Optimization	Pheromone deposition/Path selection	Distributed collaborative optimization	High memory usage
Asexual Reproduction Optimization	Cloning/Gene fragment recombination	High local search efficiency	Prone to premature convergence
Jaya Algorithm	Favorable-avoidance strategy	No need to adjust control parameters	Significant performance degradation in high -dimensional problems

Genetic algorithms, with their robust operator system and global convergence, are ideal for FCM weight matrix learning. Their evaluation includes: 1) encoding scenario element correlations through chromosomes; 2) customizing fitness functions for tasking effectiveness; and 3) utilizing parallel computing for timely joint decision-making. This study uses an adaptive genetic algorithm, with pseudocode in Algorithm 1.

Algorithm 1 Genetic Algorithm

Algorithm 1 Genetic Algorithm

Initialize population $P_0 = \{w_i^{(0)}, w_j^{(0)}\}_{i=1}^{N}$

repeat

  repeat

    w_i ← Arithmetic Crossover: $w_i^{'} = \alpha w_i + (1-\alpha) w_i$, where $\alpha \sim U(0,1)$

    w_j ← Multi-Point Crossover: $w_j^{'} = \Phi(w_j, w_i, k)$, k is random cut point

    $w_i^{'}$ ← Gaussian Mutation: $w_i^{''} = w_i^{'} + \mathrm{N}(0, \sigma^2)$

    $w_j^{'}$ ← Polynomial Mutation: $w_j^{''} = w_j^{'} + \Delta(q)$, q is mutation intensity

    Compute Fitness $f(w_i^{''}) = 1/(1+\sum d_{ij})$

  until generate offspring population Q_t satisfy $|Q_t| = N$

  Select parents $P_{t+1} = \arg\max_{w \in P_t \cup Q_t} \{f(w)\}$

until $\max f(w) > \theta$ **or** $t > T_{\max}$

As shown in Figure 3, the construction of DIFCM is based on six core situational elements ($(C_i, i = 1,2,...,6)$) and three environmental control variables. The causal relationships between nodes are characterized by an interval intuitionistic fuzzy weight matrix $W = [w_{ij}]_{6\times 6}$, where each weight $w_{ij} = <[\mu_{ij}^L, \mu_{ij}^U], [\nu_{ij}^L, \nu_{ij}^U]>$ is determined by expert groups using an improved Delphi method [13].

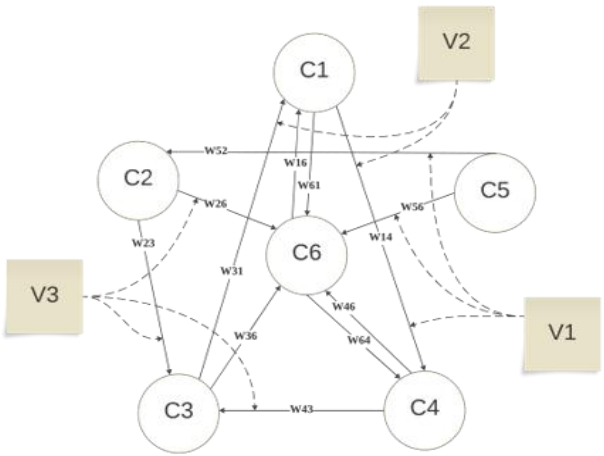


Figure 3 Conceptual Nodes of the Situation and Threat Assessment Model

The threat assessment network focuses on eight-dimensional elements, and its weight matrix construction abandons traditional expert weighting methods in favor of data-driven optimization using an improved genetic algorithm. The chromosome encoding scheme is designed as follows: for 11 sets of weight parameters, the membership degree μ_{ij} and non-membership degree ν_{ij} are concatenated to form a 22-dimensional real-valued vector:

$$chromosome = [\mu_{18}, \nu_{18}, \mu_{21}, \nu_{21}, \mu_{28}, \nu_{28}, ...\mu_{78}, \nu_{78}, \mu_{87}, \nu_{87}] \ . \tag{4}$$

The population size is set as $P = 55$ to ensure search space coverage. The fitness function is defined as:

$$Fitness = \frac{1}{1+\frac{1}{T}\sum_{t=1}^{T}\left|N_8^{sim}(t) - N_8^{ref}(t)\right|}, \tag{5}$$

where N_8 represents the threat level synthesis node, and its reference value N_8^{ref} is calibrated based on expert evaluation results from historical missions. The genetic operations adopt a hierarchical strategy:

(1) Selection Phase: Roulette wheel selection is used, where the probability of each chromosome being selected is proportional to its fitness:

$$P_i = \frac{Fitness_i}{\sum_{j=1}^{P} Fitness_i}. \tag{6}$$

(2) Crossover Phase: Directed arithmetic crossover is performed on selected parent individuals, with new weights calculated as:

$$\mu_{ij}^{'} = \alpha\mu_{ij}^{(p)} + (1-\alpha)\mu_{ij}^{(q)}, \alpha \sim U(0.2, 0.8). \tag{7}$$

(3) Mutation Phase: Cauchy-Gaussian hybrid mutation is applied to 10% of the individuals:

$$\mu_{ij}^{''} = \mu_{ij}^{'} + \delta(\lambda C(0,1)) + (1-\lambda)\mathrm{N}(0,1)). \tag{8}$$

Here, λ controls the mutation intensity distribution and δ is the step size coefficient. The algorithm termination conditions are set as either a fitness improvement of less than 0.1% for 100 consecutive generations or reaching the upper limit of 300 iterations.

Table 2 Part of conceptual nodes of the model

Conceptual Nodes	Implication
C_1	Destructive Capability of Opposing Equipment Systems
C_2	Cognitive Load of Manned Aircraft Operators
C_3	UAV Equipment Systems
C_4	Manned and Unmanned Rear Support Systems
C_5	Manned and Unmanned Information Systems
C_6	Advantage/Disadvantage Level of Our Situation Status
$^*V_1$	Atmospheric Density
$^*V_2$	Opposing Support
$^*V_3$	Manned Aircraft Operator Intervention

Note: * is environmental control variable.

As shown in Table 2, manned aircraft operators influence the assessment process through two interfaces: Environmental Control Variable Adjustment, where weather changes affect value ranges and atmospheric density compensation, and DIFCM, which generates situational index vectors that integrate cognitive load in threat assessment. The cognitive foundation module uses interval intuitionistic fuzzy numbers to represent expert knowledge and scenario uncertainty. The hybrid learning framework combines genetic optimization and rule-based reasoning, reducing subjective reliance while ensuring interpretability. A bidirectional mapping mechanism guides human cognition to correct UAV assessment deviations.

2.3 Task Allocation and Path Planning Model

In manned/unmanned collaborative tasking systems, task allocation and path planning are key decision-making elements that ensure operational effectiveness. Balancing task efficiency, resource use, and adaptability, this paper presents a cognitive intelligence-based framework using an improved Contract Net Protocol (CNP) and adaptive A* algorithm for efficient resource scheduling and safe path generation in dynamic mission environments.

2.3.1 Contract Net Protocol

The task allocation problem is a multi-constraint, multi-objective optimization challenge, complicated by scenario uncertainties and the collaboration of heterogeneous platforms as shown in Figure 4:

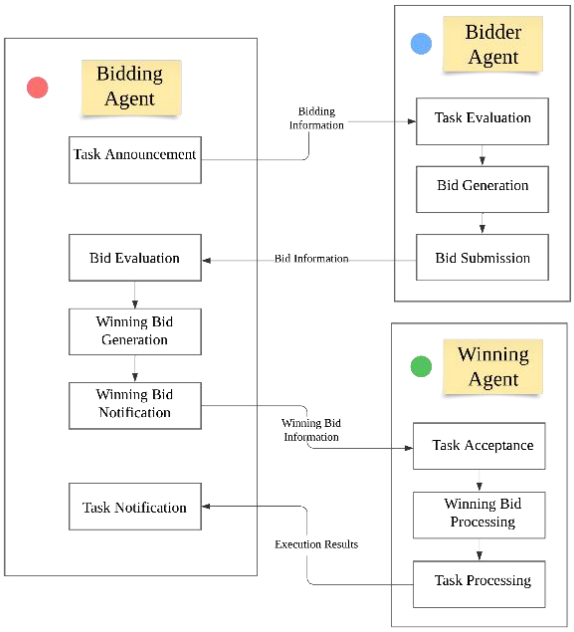


Figure 4 Flowchart of the Enhanced Contract Net Protocol Algorithm

Traditional CNP simulates market bidding but neglects human cognitive load and UAV autonomy. To overcome this, the cognitive intelligence-based collaborative decision-making model includes the cognitive state of manned aircraft operators in task allocation. The task allocation matrix is defined as $X_{N\times M}$, where the element $x_{ij}=1$ represents the assignment of target to UAV i. The task effectiveness function $U(X)$ is composed of a profit function $E(X)$, a cost function $C(X)$, and a profit adjustment function $D(X)$:

$$U(X)=E(X)-C(X)+D(X). \tag{9}$$

The profit function $E(X)$ integrates target value, damage probability, and operator reliability factors, with its mathematical model expressed as:

$$E_u=\frac{V_{TARGET_j}\cdot P_k\cdot e^{-\delta t}}{1+e^{-h\lambda}}, \tag{10}$$

where V_{TARGET_j} represents the strategic value of the target, δ is the operator error rate, P is the UAV's probability of target destruction, and t is the current time. The cost function V_{TARGET_j} quantifies UAV loss risk and target threat intensity:

$$C_u=k_1\cdot UAV_{V_i}\left(1-\prod_{j=1}^{m}(1-P_j)\right)+k_2\cdot threat_{ij}, \tag{11}$$

where UAV_{V_i} is the asset value of UAV i, P_j is the damage probability when executing a task on target j, and the weight coefficients k_1 and k_2 satisfy $k_1+k_2=1$. To enhance the adaptability of task allocation, UAV intelligent emotions are introduced into the profit adjustment function, with values dependent on three emotional modes (fear, relaxation, aggression) and task profit intervals as shown in Table 3. The preference coefficient α is calibrated through regression analysis of historical mission data, ensuring decision logic aligns with actual mission requirements.

Table 3 Impact of Different Emotional Modes on UAV Task Selection

Emotional Mode	Task Preference	Profit Adjustment Value
Fear	Low risk, low reward	Low
Relaxation	Medium-low risk, medium reward	Average
Aggression	High risk, high reward	High

2.3.2 Self-adaptive A star algorithm

A star algorithm, as a typical representative of heuristic search algorithms, holds significant theoretical value and engineering significance in the field of grid-based environment path planning [14]. By constructing a cost evaluation function that balances actual path costs and heuristic estimates, the algorithm achieves efficient search while ensuring global optimality. Its core mathematical model can be expressed as:

$$f(n) = g(n) + h(n), \tag{12}$$

where n represents the current expanded node, $g(n)$ is defined as the cumulative actual path cost from the start point to node n, and $h(n)$ is the heuristic function used to estimate the minimum expected cost from node n to the target point. This study adopts the Euclidean distance:

$$h(n) = \sqrt{(x_n - x_{goal})^2 + (y_n - y_{goal})^2} \tag{13}$$

as the heuristic function. This function satisfies the admissibility condition (i.e., it does not exceed the true path cost), thereby ensuring the algorithm's optimality. Core algorithm as shown in Algorithm 2:

Algorithm 2 Classic A* Path Planning Algorithm

Algorithm 2 Classic A* Path Planning Algorithm

Algorithm A_star(Start point s_{start}, Target point s_{goal}, Cost function $c: S \times S \to \mathbb{R}^+$)
Initialize OpenList $O = \varnothing$, CloseList $C = \varnothing$
Set $g(s_{start}) = 0$
O.insert$(s_{start}, f(s_{start}))$
while $O \neq \varnothing$ **do**
  $s_{current} = O$.extract_min()
  if $s_{current} == s_{goal}$ **then**
    return $reconstruct\_path(s_{current})$
  C.add($s_{current}$)
  for $s_{neigh} \in Neighbors(s_{current})$ **do**
    if $s_{neigh} \in C$ **then continue**
    $g_{tent} = g(s_{current}) + c(s_{current}, s_{neigh})$
    if $g_{tent} < s_{neigh}$ **then**
      $g(s_{neigh}) = g_{tent}$
      $f(s_{neigh}) = g(s_{neigh}) + h(s_{neigh})$
      $Parent(s_{neigh}) = s_{current}$
      if $s_{neigh} \notin O$ **then**
        O.insert$(s_{neigh}, f(s_{neigh}))$
return *Failure*

Theoretically, the A* algorithm is guaranteed to be both complete and optimal: it will return a feasible path if a solution exists, and it ensures global optimality when the heuristic function satisfies admissibility [15]. To enhance the algorithm's adaptability in dynamic scenario environments, this study introduces UAV intelligent emotion mode settings as part of the collaborative decision-making model.

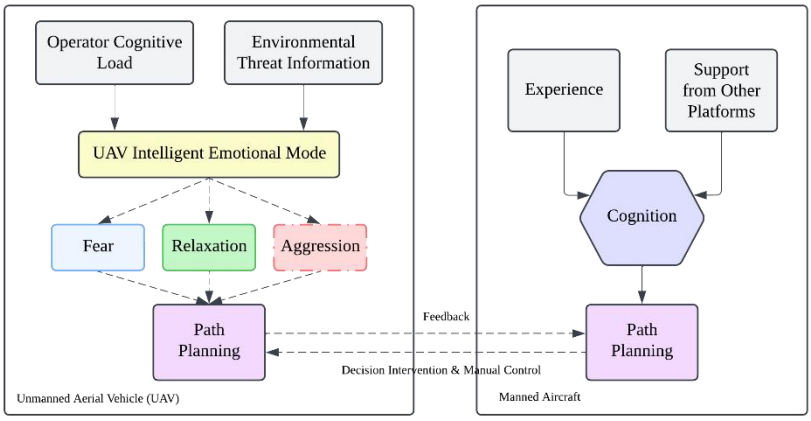


Figure 5 Path Planning Model

Of which basic structure is shown in Figure 5 the path planning problem can be formalized as a network graph search problem, where the node set $S=\{s_1,s_2,...,s_m\}$ represents the discretized scenario space, and the set of all feasible paths from the start point to the target point is $E=\{e_k \mid e_k=(s_{i_1},s_{i_2},...,s_{im})\}$. Defining the path $e_k \in E$ with adjacent node pairs (s_i,s_j), their connecting edges are denoted as $V(s_i,s_j)$, and the flight cost as d_{ij}. The UAV trajectory planning problem can then be modeled as:

$$\min f(e_k)=\sum_{(s_i,s_j\in E)} d_{ij} \text{ , s.t. } s_i \in E, s_j \in E, e_k \in E. \tag{14}$$

The heuristic function $h(s_i)$ uses Euclidean distance to estimate the remaining flight distance, ensuring algorithm convergence and optimality. To address dynamic threat environments, an emotion-dependent threat avoidance radius is designed as:

$$D=\mathrm{R}\cdot \mathrm{e}^{(1-1/R)\cdot(1-C_w)}, \tag{15}$$

where R is the threat influence radius, and the UAV intelligent emotional mode C_w is determined by parameters α and β as shown in Table 4.

Table 4 Determination Range of UAV Intelligent Emotional Modes

Value of C_w	Emotional mode
$C_w<\alpha$	Fear
$\alpha\le C_w\le\beta$	Relaxation
$C_w>\beta$	Aggression

Human-machine collaboration mechanisms are integrated throughout the task allocation and path planning process. Operators can intervene through two types of interfaces: at the task allocation level, dynamically adjusting emotional mode thresholds or manually specifying high-value targets; at the path planning level, setting temporary no-fly zones or modifying threat avoidance parameters. When path risk exceeds the operator's preset threshold, an alarm signal is triggered, suggesting task termination or replanning as shown in Figure 6. This bidirectional information flow design ensures the stability and real-time performance of the human-machine decision-making loop.

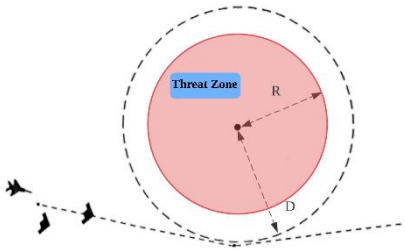


Figure 6 Schematic Diagram of Threat Avoidance Radius

The model improves traditional methods with three innovations: 1) modeling cognitive load and emotions as endogenous decision variables, removing the need for human experience; 2) designing an emotion-threat-coupled path cost function to balance safety and economy; and 3) creating a task-path joint optimization framework to reduce risks from local optimization. Its value lies in advancing collaborative decision-making from static rule-driven to dynamic cognition-guided evolution, enabling deeper human-machine intelligence integration.

3 RESULTS AND ANALYSIS

To validate the effectiveness of the proposed manned/unmanned collaborative decision-making framework, a simulation experimental system was constructed based on the MATLAB platform. The experimental design follows a progressive logic of "modular verification-scenario simulation-comprehensive evaluation," with a focus on analyzing the performance of three core modules: threat assessment, task allocation, and path planning, in comparison with traditional methods. In the verification of the situation and threat assessment module, a genetic algorithm was used to optimize the 11 sets of weights in the Intuitionistic Fuzzy Cognitive Map (IFCM). The population size was set to 200, with a crossover probability of 0.8 and a mutation probability of 0.15. After 300 iterations, the convergence curve is shown in Figure 7. Experimental data indicate that the optimal fitness value stabilized at 0.812 (Best curve), and the population mean converged to 0.813, demonstrating the algorithm's strong global search capability and stability. The optimized membership and non-membership heat matrix reveals that the membership values of key threat nodes (e.g., target radar detection accuracy N_8) are concentrated in the [0.6, 0.85] range. In sudden threat response tests, the model's re-evaluation time was 0.87 seconds, meeting the real-time requirements of dynamic battlefields as shown in Figure 8.

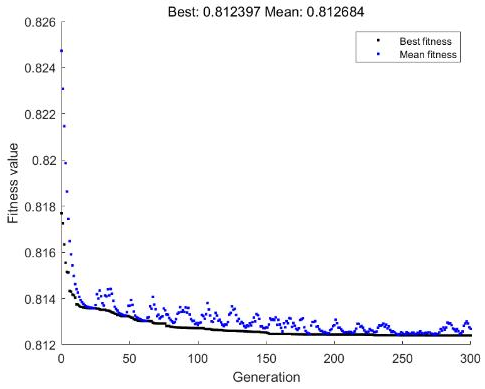


Figure 7 IFCM Genetic Algorithm Optimization

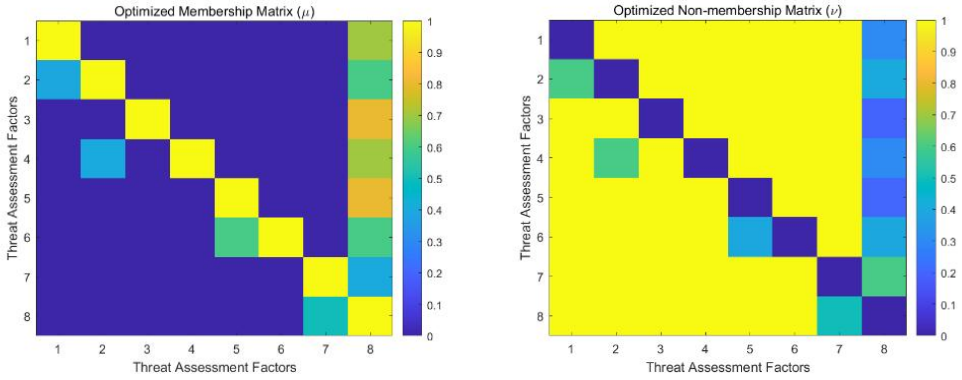


Figure 8 Optimized Threat Assessment Membership and Non-Membership Matrix

Based on the improved Contract Net Protocol (CNP) task allocation model, a tasking scenario was constructed with four heterogeneous UAVs targeting six opposite objectives. The heatmap of bid values optimized by Particle Swarm Optimization (PSO) shows that UAV4, equipped with electronic countermeasure devices, achieved bid values of 0.51 and 0.47 for high-value targets T1 and T4, respectively. UAV1, in aggressive mode, increased its bid intensity for high-risk target T1 by 26.8%. Compared to UAV2 with the same emotional mode, UAV3, when assigned to accept signals with higher operator cognitive load, saw its intelligent emotional mode's influence on bidding relatively suppressed, enhancing the safety and effectiveness of human-machine collaboration strategies as shown in Figure 9.

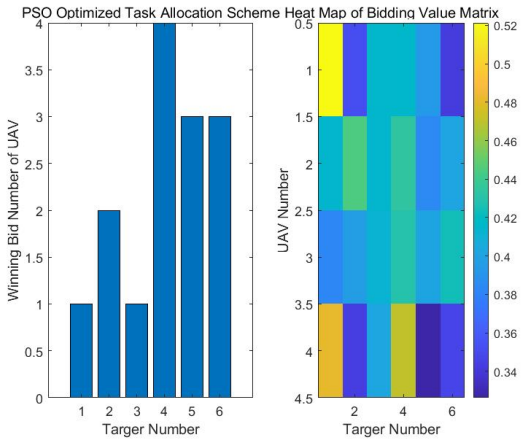


Figure 9 Heatmap of Task Allocation Scheme and Bid Value Matrix Optimized by PSO

In an 800×600 grid tasking environment with six fixed threat sources, the improved A* algorithm was used for path planning. A comparison of routes under three emotional modes shows that the fear mode (C=0.15) had an average threat cost of 0.17, a 58% reduction compared to the aggressive mode, but with a 22.3% increase in path length. The aggressive mode (C=0.85) resulted in a total threat cost of 165, while the relaxed mode (C=0.5) achieved the best balance, with standard deviations of path length and threat cost being only 43% and 52% of the other modes, respectively. The algorithm's average planning time was 0.64 seconds, and the sudden threat replanning response time was 1.43 seconds, a 63% improvement over the traditional RRT algorithm. Path planning situation map as shown in Figure 10.

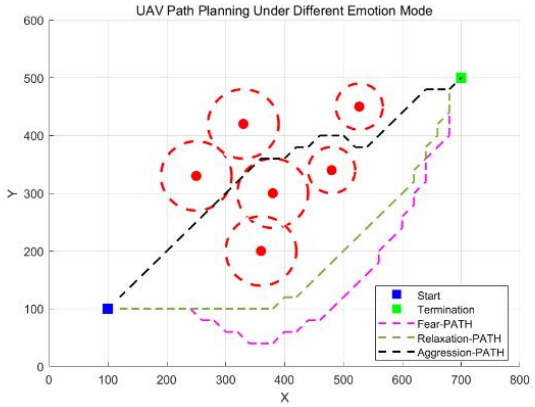


Figure 10 UAV Path Planning Situation Map Under Different Emotion Modes

Overall the collaborative decision-making framework proposed in this paper has been validated at the algorithmic level. Its core value lies in revealing the intrinsic relationships of threat elements through weight optimization heatmaps, quantifying UAV capability differences using bid value distribution characteristics, and demonstrating the effectiveness of the emotional adaptive mechanism through path comparison experiments. Future work should further develop hardware-in-the-loop simulations and incorporate reinforcement learning strategies to optimize algorithm convergence, advancing theoretical research toward practical applications.

4 CONCLUSIONS AND OUTLOOKS

This paper addresses decision-making optimization in manned/unmanned collaborative tasking by proposing a cognitive collaboration-based decision-making framework. Through the deep integration of Dynamic Intuitionistic Fuzzy Cognitive Maps (DIFCM), Genetic Algorithms (GA), and an improved A* algorithm, a full-process model covering "situation assessment, threat analysis, task allocation, and path planning" is constructed. Experimental results show that the framework surpasses traditional methods in threat assessment accuracy, task allocation efficiency, and path planning safety. Specifically, DIFCM, optimized by genetic algorithms, enhances the real-time performance and accuracy of threat assessment. The improved Contract Net Protocol, incorporating UAV intelligent emotional modes, enables adaptive task allocation. Meanwhile, the emotion-driven A* algorithm demonstrates superior path planning in dynamic threat environments. The key innovation lies in coupling human cognitive characteristics with machine autonomous decision-making, reducing dependence on human experience and offering a novel methodological perspective for manned/unmanned collaborative mission.

Future research should integrate advanced reinforcement learning strategies to improve algorithm convergence and adaptability. Additionally, hardware-in-the-loop simulations are necessary to validate the framework's practical applicability, ensuring its robustness in real-world tasking scenarios.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Gu H, Xu C. Human-Machine Collaborative Teamwork Technology. Command Information Systems and Technology, 2017, 8(06): 33-41.
[2] Fu J, Zhang D, Wang M, et al. UAV Trajectory Planning for Target Localization Accuracy Enhancement. Acta Armamentarii, 2023, 44(11): 3394-3406.
[3] Wu Y. Research on Active Perception Methods for UAV Swarm Target Search. Shanghai Institute of Technology, 2020.
[4] Xie Y, Lu Y, Guan C, et al. Key Issues in Cooperative Aerial Tasking and Multi-Agent Reinforcement Learning. Aircraft Design, 2023, 43(01): 6-10.
[5] Zhong Y, Yao P, Zhang J, et al. Interactive Cooperative Decision-Making in Human-Machine Systems Based on HFCM. Systems Engineering—Theory & Practice, 2021, 41(10): 2748-2760.
[6] Xue Z, Zhang Y, Chen J. Research on Hybrid Active Interactive Decision-Making for Unmanned-Manned Aircraft. Aeronautical Science and Technology, 2022, 33(05): 44-52.
[7] Liu B, Wei X, Qu H, et al. Collaborative Decision-Making in Human-Machine Aerial Systems. Advances in Aeronautical Engineering, 2023, 14(06): 63-72.
[8] Chen J, Liang J, Cheng L, et al. Multi-UAV Cooperative Decision-Making Modeling Based on FCM. Acta Aeronautica et Astronautica Sinica, 2022, 43(07): 377-394.
[9] Luo Y, Ding D, Tan M, et al. A Review of Autonomous Decision-Making Methods for Unmanned Aerial Vehicles. Acta Aeronautica et Astronautica Sinica, 2025: 1-30.
[10] Robert Axelrod. Structure of Decision. Princeton University Press, 2015.
[11] Thong N T, Smarandache F, Hoa N D, et al. A novel dynamic multi-criteria decision making method based on generalized dynamic interval-valued neuromorphic set. Symmetry, 2020, 12.
[12] Iván S, Anne C, Anna D, et al. Fuzzy cognitive mapping in participatory research and decision making: a practice review. Archives of Public Health, 2024, 82(1): 76-76.
[13] Zhida S. Use of Delphi in health sciences research: A narrative review. Medicine, 2023, 102(7): e32829-e32829.
[14] Guo Z. Efficient path planning method based on A-STAR//Department of Civil & Environmental Engineering, Brunel University London, Department of Computer and Information Sciences, University of Strathclyde. Proceedings of the 2023 International Conference on Mechatronics and Smart Systems. Ulster College Shaanxi University of Science and Technology, 2023: 196-205.
[15] Erke S, Bin D, Yiming N, et al. An improved A-Star based path planning algorithm for autonomous land vehicles. International Journal of Advanced Robotic Systems, 2020, 17(5).

Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3063

DROUGHT RECOGNITION FOR THE SOYBEAN PLANT BASED ON LIGHTWEIGHT DEEP LEARNING MODEL

WenLiang Hou*, ZhiPeng He, JieWen Gu, JiaLin Zhang, JiaoYing Li
College of Information and Electrical Engineering, Heilongjiang Bayi Agricultural University, Daqing 163319, Heilongjiang, China.
Corresponding Author: WenLiang Hou, Email: houwenliang1024@gmail.com

Abstract: To address the issues of low soybean self-sufficiency and drought - related production constraints, multispectral imaging can non - invasively detect crop characteristics, while deep convolutional networks identify drought from multispectral images. However, large models face mobility limitations due to high computational requirements.This study presents a mobile detection approach integrating ReliefF feature screening and lightweight convolutional neural networks, and develops the "Early Acknowledgment for Soybean Drought" App. It embeds a lightweight model that optimizes 37 - dimensional soybean canopy multispectral features via ReliefF and uses a three - layer 1D convolutional network for drought identification.The model achieves 96.88% classification accuracy on the self - built dataset, with an inference time of 18 ms, a size under 30 MB, and less than 60 MB memory usage on mobile. The APP integrates the multispectral camera SDK and PyTorch inference engine, enabling real - time spectral analysis. Field tests show its one - button operation, low learning curve for farmers, and significant water - saving and yield - increasing effects, offering a lightweight, high - precision mobile solution for soybean drought management and promoting smart agriculture development.
Keywords: Soybean drought; Multispectral image; Feature screening; Recognition model; Mobile app

1 INTRODUCTION

Food security is the strategic cornerstone of national security, although China has realized the absolute security of staple food, but the soybean self-sufficiency rate continues to go down to less than 20%[1], superimposed on the growth of consumer demand and international trade friction intensified, the dependence on imports climbed to the first place in the world. For soybean growth is highly dependent on the characteristics of water, drought stress has become a key adversity factor restricting its production, there is an urgent need to break through the drought-resistant varieties through accurate phenotypic detection technology and water-saving irrigation bottlenecks, in order to enhance the production capacity of soybeans to provide scientific support. As an emerging non-destructive testing method, multispectral imaging technology can simultaneously analyze the external morphology (color, texture) and internal physicochemical properties (water, nutrients) of the crop through the synergistic perception of multiple wavelengths[2], such as visible light and near-infrared light, to realize the multidimensional characterization of the canopy structure and growth status. With the breakthrough of artificial intelligence technology, the high-throughput phenotyping system integrating machine learning and image processing algorithms is becoming a cutting-edge paradigm for precision agriculture research, providing innovative solutions for crop physiological monitoring and stress diagnosis[3].
Gao Shijiao et al innovatively integrated Gaussian filtering and multi-threshold segmentation strategy to construct an efficient extraction model for soybean canopy near-infrared band[4], and the effective segmentation rate was increased to 97.81%; Fu Hongyu et al developed a dynamic monitoring system for physical and chemical traits of Ramie flax by integrating RFE feature selection and multi-temporal remote sensing data[5], and the accuracy of the SVR-LAI model in estimating leaf area index reached R^2=0.737; Fan Xuexing et al proposed a chlorophyll inversion model based on multispectral imaging and PSO-SVR optimization algorithm[6], and realized the prediction of chlorophyll content through the nonlinear mapping of spectral reflectance and SPAD value. The accuracy was improved to R^2=0.91; Han WT et al innovatively adopted the XGBoost-GRA bimodal feature preference strategy combined with RF machine learning model to achieve an inversion of salinity in 0-20 cm soils with an accuracy of R^2=0.820[7], and the accuracy of spatial distribution map was improved to RPI=0.820. The spatial distribution map accuracy is improved to RPIQ=2.273; Huang Linsheng et al fused phase correlation alignment and UNet semantic segmentation network to overcome the problem of multispectral image channel bias[8], so that the lettuce canopy segmentation accuracy reaches 99.19%; Pang Qi integrated hyperspectral imaging and YOLOv3 deep learning framework to break through the bottleneck of hidden defect detection in apples[9], and realized 100% of the F1 scores vs. 68 fps real-time processing performance, which is significantly better than the traditional threshold segmentation method.
In summary, chemometrics methods were crucial in the early development of multispectral image analysis. However, their limited feature extraction and low model complexity severely restrict classification accuracy in complex scenarios. In contrast, deep learning-based methods have emerged as a breakthrough. By constructing deep neural networks with self-learning features, these methods reduce data preprocessing and enable end-to-end analysis. Deep convolutional networks capture multispectral image features at multiple scales, but increased network depth raises model complexity and computational costs, preventing large models from running on mobile devices.This study addresses the inefficiencies of traditional soybean drought phenology detection and the limited field applicability of large devices. By

integrating multispectral image feature extraction with deep learning, we developed a portable mobile drought detection system. The system includes canopy segmentation, spectral analysis, and a drought identification model, culminating in the lightweight "Early Acknowedgment for Soybean Drought" App. This enables non-destructive, real-time detection of soybean canopy drought traits, providing a high-throughput tool for breeding and irrigation management. The system's technology roadmap is shown in Figure 1.

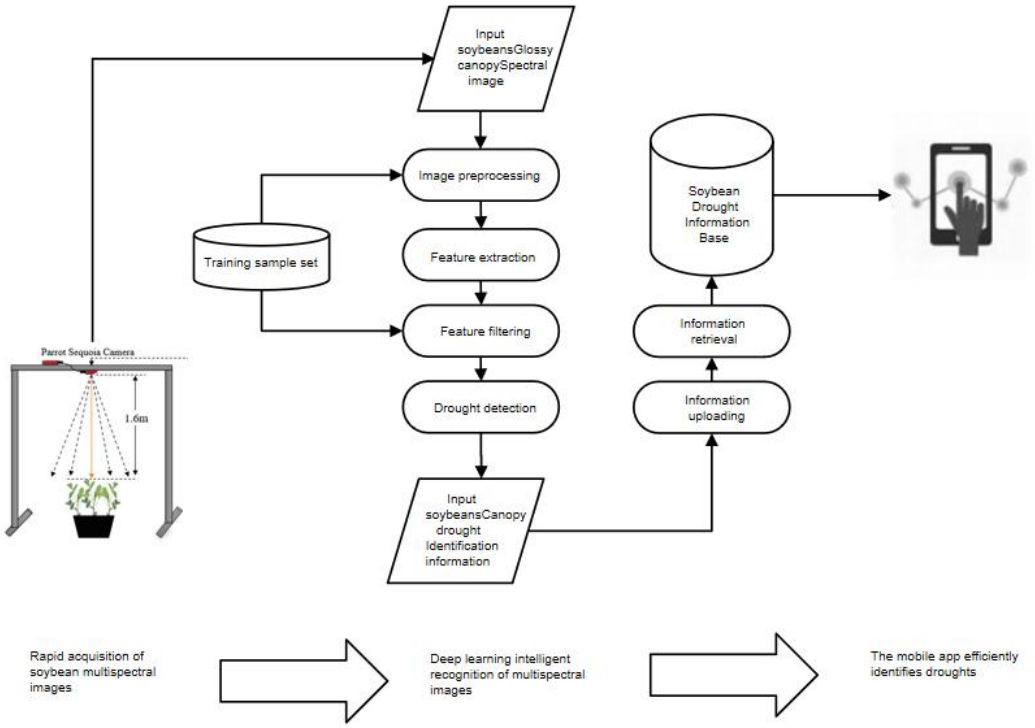


Figure 1 System Technology Roadmap

2 MULTISPECTRAL IMAGE FEATURE SCREENING OF SOYBEAN CANOPY BASED ON RELIEFF ALGORITHM

2.1 Relevant Features Algorithm Principle

As a classical feature evaluation method, Relief algorithm assigns differentiated weight values by analyzing the correlation between each feature and the target category[10]. When the weight evaluation value of a feature is lower than a preset threshold value, the system will automatically reject the feature. The algorithm is concise, effective, and addresses a series of computational and model performance issues derived from the high dimensionality of multispectral features, so the ReliefF algorithm is used as the multispectral feature screening method in this paper. The specific steps are as follows.

Firstly, a sample x_i is randomly selected among the soybean multispectral image samples, and then the distances of the k nearest-neighbor samples in the same class as x_i are computed for $\sum_{j=1}^{k} S(f_i, x_i, H_j)$, and the distances of the k nearest-neighbor samples in a different class from x_i are computed for $\sum_{j=1}^{k} D(f_i, x_i, M_j(A))$, respectively. According to the size of the class spacing and intra-class distance, the size of the weights is adjusted, and the weights are updated by iterating m times, and the feature selection is carried out by the final weights. The formula for calculating the weights is as follows:

$$W^{i+1}(f_l) = W^{i}(f_l) - \frac{\sum_{j=1}^{k} S(f_i, x_i, H_j)}{mk} + \sum_{C \neq class(x_i)} \frac{\frac{p(A)}{1 - P(class(x_i))} \sum_{j=1}^{k} D(f_i, x_i, M_j(A))}{mk} \tag{1}$$

Where $W^{i}(f_l)$ is the weight of the l feature f in the first i sample; $H_j(j = 1, 2, \cdots, k)$ is the j th sample among the k nearest neighbor samples in the same class as x_i ; $p(A)$ is the ratio of samples belonging to the class A

in the training samples; $P(class(x_i))$ is the ratio of samples in the same class to the total samples, where $class(x_i)$ is the label of x_i ; $M_j(A)(j=1,2,\cdots,k)$ is the j th sample among the k nearest neighbor samples in the different class from x_i . The distance formula is

$$D_f(x_1,x_2)=\frac{\left|x_{1f}-x_{2f}\right|}{\max(f)-\min(f)} \tag{2}$$

where $D_f(x_1,x_2)$ is the normalized distance between the samples x_1 and x_2 on the f th feature, x_{1f} and x_{2f} are the f th features of the samples x_1 and x_2 , respectively, and $\max(f)$ and $\min(f)$ are the maximum and minimum values corresponding to the feature f in all the samples, respectively.

2.2 Soybean Canopy Drought Recognition Based on 3-Layer Convolution-Relevant Features

In this paper, the multispectral feature dataset of soybean canopy preferred by ReliefF algorithm is used as the input vector to build a multispectral soybean drought convolutional classification model for model simulation[11], and the architecture of the multispectral soybean drought convolutional classification model constructed in this paper is shown in Figure 2.

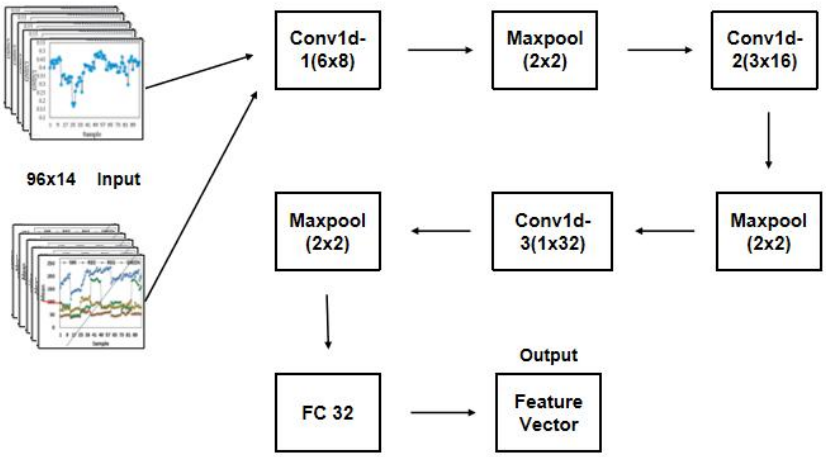


Figure 2 CNN Soybean Drought Classification Model Architecture

Convolutional neural network has powerful feature extraction ability[12], so this paper constructs a soybean canopy drought identification model based on 3-layer convolutional-Relevant Features. This multispectral soybean drought convolutional neural network contains 7 layers, among which there are 3 convolutional layers, 2 pooling layers, 1 fully connected layer and 1 output layer. In this case, since the multispectral soybean canopy in this study is structured data, unlike unstructured image data, the multispectral data features were extracted by a one-dimensional convolutional kernel. The convolutional layers are Conv1d-1, Conv1d-2 and Conv1d-3, with the size of the convolutional kernel of 6, 3 and 1, and the number of convolutional windows of 8, 16 and 32, respectively.The size of the pooling layer is 2, and the number of neurons in the fully connected layer is 32, and the number of the output layer is 1.The principle of the convolutional neural network model constructed in this study and the computational process are as follows:

The whole Convolutional-Relevant Features model consists of three Block modules, where each Block module consists of a convolutional layer, downsampling, and pooling layer. Take the first Block module as an example, the specific process is, first of all, the size of the data set for the 14x1 sample data input to the first convolutional layer, the use of eight 1x6 convolution kernel size, convolution step default 1, feature data by the first layer of the convolution of the size of the shortening of the 9, which is a one-dimensional convolutional operation formula is as follows

$$y_j^l=f\left(\sum_{i=1}^{N} w_{ij}^l * x_i^{l-1}+b_j^l\right) \tag{3}$$

Where y_j^l represents the feature map of the j th convolutional output of the l layer: f represents the nonlinear activation function. $*$represents the convolution operation, N is the number of kernels in the $l-1$ layer, x_i^{l-1} is the feature map of i in the $l-1$ layer, w_{ij}^l represents the weights, and b_j^l is the bias of the j convolution kernel in the l

layer[13]. After completing the convolution operation, the ReLU activation function is used to increase the nonlinear mapping ability of the network. In this paper, Max Pooling with a window size of 2 is used for window sliding calculation for feature compression and information filtering of data. The Max Pooling process used in this paper is shown in the following equation:

$$p_i^{l+1}(j) = \max_{(j-1)w+1 \le l \le jw} \left\{ q_i^l(t) \right\} \tag{4}$$

Where $q_i^l(t)$ is the value of the t neuron corresponding to the i feature quantity in the l layer, w is the pooling layer width, and $p_i^{l+1}(j)$ is the value of the $l+1$ neuron. In order to avoid overfitting, the data are downsampled, the neurons with a random inactivation ratio of 0.2 are randomly inactivated, and the length of the feature map is reduced to 4 and the number of feature channels is changed to 8 after maximum pooling.

3 IMPLEMENTATION OF APP FOR SOYBEAN DROUGHT IDENTIFICATION SYSTEM BASED ON ANDROID TERMINAL

The realization of the overall framework of the APP system of soybean drought identification system based on Android is shown in Figure 3[14]. In the home page, users can log in through the cell phone authentication code and enter the main interface after successful login, which realizes the switching of real-time monitoring, historical data, farmland management and other functions through ViewPage control. In the real-time monitoring module, the SDK of Sequoia multispectral camera is called to obtain multi-band data such as red edge, near infrared, etc. The raw spectral stream is converted to NDVI pseudo-color map and displayed on ImageView control through JNI technology, and at the same time, an asynchronous thread is created to call the PyTorch-trained Conv-ReliefF model to perform the analysis of drought level. The multispectral image data is stored to an Oracle database through a RESTful API and the raw image files are synchronized using MinIO. The server side is built based on Eclipse Java, providing interfaces such as /api/upload to receive images and return model inference results, combined with Redis caching of high-frequency query data to improve response efficiency. The system realizes accurate and safe agricultural drought monitoring through multi-spectral data fusion and HTTPS encrypted communication, and supports the expansion needs of smart farmland management scenarios.

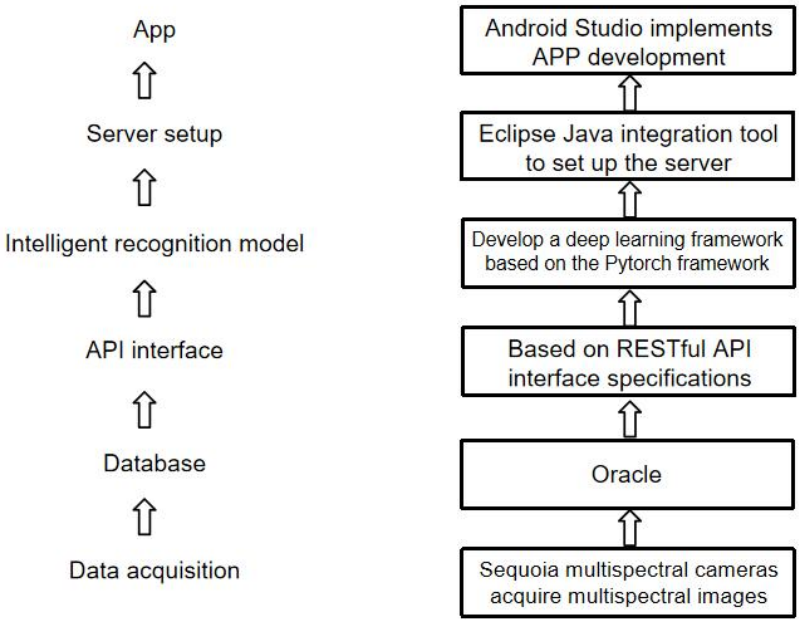


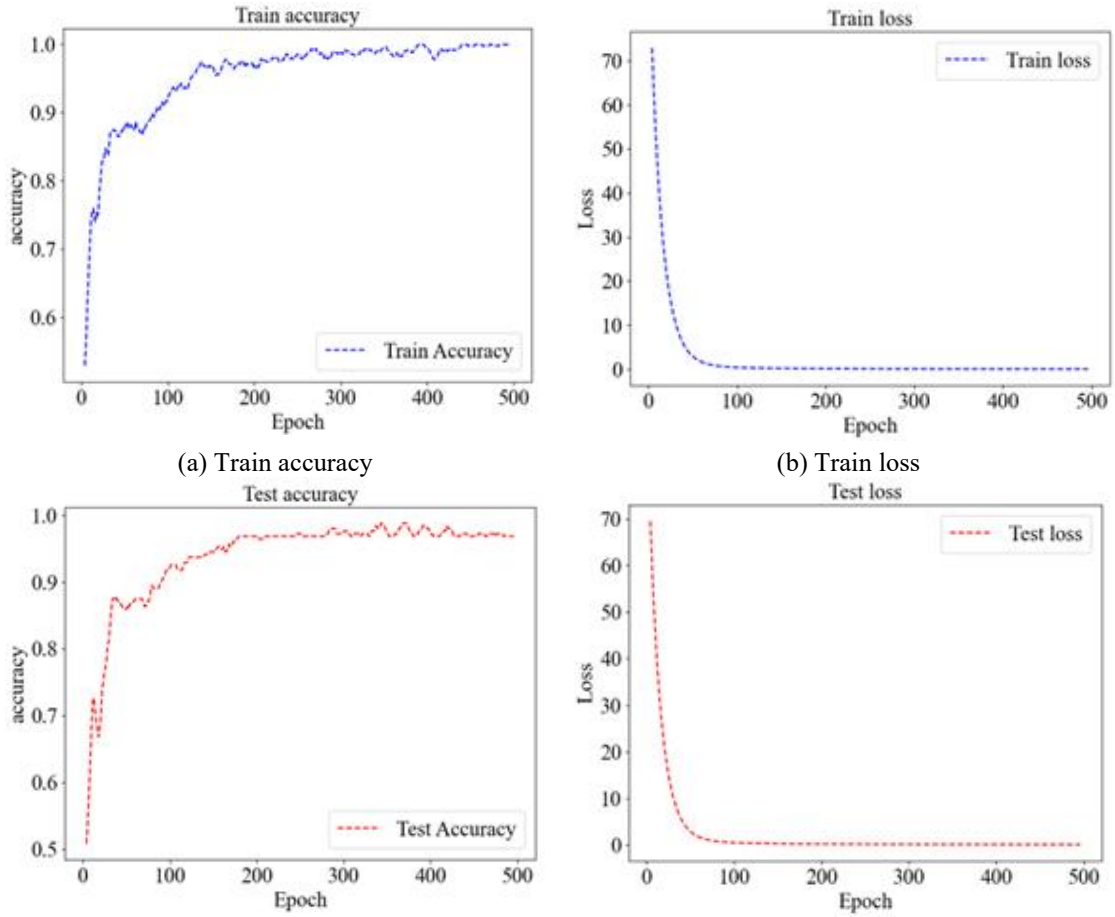
Figure 3 Overall System Framework of the Mobile Soybean Drought Identification System

4 RESULTS AND ANALYSIS

4.1 Conv-ReliefF Method Training Results

In this paper, the experiments used Parrot Sequoia multispectral camera to collect soybean canopy multispectral data and carry out model training. In the data preprocessing stage, the acquired canopy feature data were divided into a training set and a test set at a ratio of 2:1, in which the training samples totaled 32. In order to eliminate the differences in feature scales, the training set data were normalized using the standardized (Standard-Score) method. For model construction, a one-dimensional convolutional neural network architecture is selected, and the Adam optimizer is used

for model optimization (the initial learning rate is set to 0.01), and the L2 regularization strategy is introduced (the coefficient is set to 0.6) to improve the model generalization ability and training stability. During the network training process, binary cross entropy is chosen as the loss function to monitor the model performance, and accuracy is used as the evaluation index to prevent overfitting phenomenon. After 500 rounds of iterative training, the performance evaluation results of the model on the training set and test set are shown in Fig., which specifically demonstrates the trend of the accuracy rate and the loss value.



(a) Train accuracy (b) Train loss

(c) Test accuracy (d) Test loss

Figure 4 Conv-ReliefF Method Training Results

Observing Figure 4 above, it can be seen that as the number of iterations increases, the overall stabilization is 96.875% at 426 iterations. The model training time is short, only 0.28 s. To evaluate the performance of the trained recognition model, 32 groups of test samples are input into the Conv-ReliefF soybean drought recognition model. To evaluate the performance of the trained identification model, 32 test samples were fed into the Conv-ReliefF soybean drought identification model. The results showed that the model took 0.018 seconds to test, which is more efficient than traditional machine learning models. Among the 32 groups of samples, the model correctly recognizes 31 groups and only 1 group is misclassified, with a prediction rate of 0.96875, which shows good prediction results in the soybean drought classification task.

Focusing on the accurate identification of soybean drought, which is a key issue in agricultural production, this paper constructs soybean drought identification models based on the principles of BP, RBF, RF, SVM and Convolutional neural networks, such as BP, RBF, RF, SVM, OD_Conv and Conv-ReliefF,The accuracy of each model and its equilibrium is shown in Table 1. The study divided the training set and test set in the ratio of 2:1, adjusted the parameters for training in several rounds, and evaluated the training effect with the help of accuracy, running time, and balanced accuracy, and the accuracy of each model reached 100% on the training set. In order to verify the actual performance of the models, 32 sets of test samples were inputted into the trained models and evaluated comprehensively by using the indicators of Accuracy, Precision, F1-score, Recall, etc. The results show that the Conv-ReliefF model has a classification accuracy of 96.88% and a running time of 0.28 seconds, which is higher in recognition efficiency and better in comprehensive performance than the other models, and provides an efficient and effective way to recognize soybean drought. Compared with other models, the Conv-ReliefF model provides an efficient and reliable solution for soybean drought identification. For space reasons, this paper only shows the confusion matrix results of Conv-ReliefF model, as shown in Figure 5.

Table 1 Equilibium Accuracy Table for Each Model

Model	Equilibium accuracy
BP	90.63%
RBF	87.50%
RF	90.63%
SVM	93.75%
OD Conv	87.50%
ReliefF_Conv	96.88%

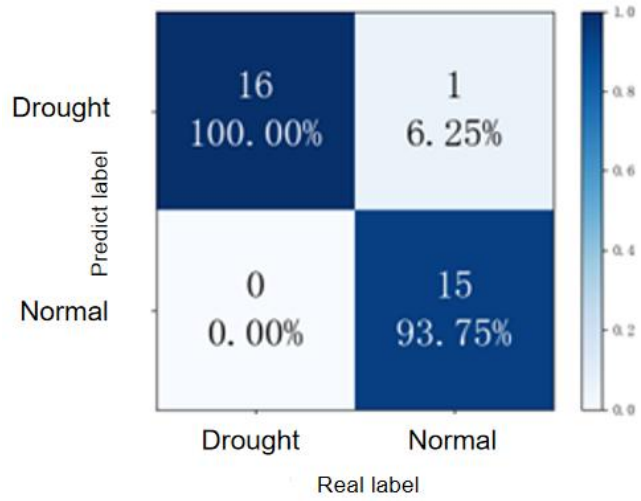


Figure 5 ReliefF-Conv Confusion Matrix Map

4.2Analysis of Application Results Based on Android Conv-ReliefF Soybean Drought Identification System APP

In the operation of the drought recognition module, the original canopy images of the four channels are opened first, and by clicking the Canopy Segmentation button, the system will automatically obtain the target canopy areas of the four channels and calculate the gray value of the canopy in each channel. The segmented image will be displayed in the QLabel control on the right, and the result of the gray value calculation will be presented in the QTextEdit label below. In the Feature Calculation section, click on the Extract All Canopy Features button, the program will calculate the original 37 dimensional feature data and generate an Excel data sheet to be saved to the current path of the program. Users can select the desired features by checking the box, or use shortcut operations such as select all, unselect all, or directly select the model training features to complete the feature selection quickly. The selected feature data is used as input for prediction by the trained model, and the prediction result will be displayed in the model recognition result area below. The specific operation process is as follows: first, select the four channel images to perform automatic crown segmentation, after segmentation, the system automatically calculates the gray value; then click on the training features button to complete the automatic selection of features; and finally click on the trained model, you can get the results of soybean drought recognition in the output box, the recognition effect is shown in Figure 6.

Figure 6 App Specific Implementation Map

Drought identification is realized on a tablet, and using the Conv-ReliefF method, the average computing time for identification is about 280ms, and the accuracy rate is about 96.88%, compared with the traditional machine learning model, the soybean drought classification convolutional neural network established in this paper performs optimally in terms of classification performance, which improves the operational efficiency and the accuracy of model identification. Combined with the application of Android APP, the Conv-ReliefF model shows high efficiency and practicability in actual agricultural production, and the portable device is convenient for farmers to quickly obtain the soybean drought conditions, take timely irrigation measures, effectively reduce the losses caused by drought, and enhance the intelligent level of agricultural production management. It further verifies its superiority in the field of soybean drought identification. Field measurements show that the system supports one-button operation, with a learning cost of less than 10 minutes for farmers, water savings of 15%-20% per hectare, and yield estimation error of <5%, which provides a lightweight, high-precision mobile solution for drought-resistant soybean breeding and precision irrigation, and significantly improves the intelligent monitoring level of smart agriculture.

5 CONCLUSIONS

Aiming at the difficulties of low efficiency of soybean drought phenotype detection and poor applicability in the field, this study innovatively integrates ReliefF feature screening algorithm and lightweight convolutional neural network (Conv-ReliefF) to construct a mobile detection system based on multispectral analysis. By optimizing the screening of 37-dimensional features in the soybean canopy layer through the ReliefF algorithm, combined with a three-layer one-dimensional convolutional network (convolutional kernel 6-3-1), the model achieves a classification accuracy of 96.88% on the self-constructed dataset, with an inference time of only 18 ms for the test set and a training time of 0.28 sec, which significantly outperforms the traditional models such as SVMs, RFs, and so on, in terms of comprehensive performance. The system adopts L2 regularization (coefficient 0.6) to improve the generalization ability, and optimizes the model size through convolution kernel time domain compression (time width=5, step size=2), which is 1/5 of the traditional PC algorithm.The results show that the Conv-ReliefF model size is optimized to be less than 30 MB, and the memory consumption of the mobile terminal is less than 60 MB, which verifies the dual advantages of model lightweighting and high accuracy. The results show that the Conv-ReliefF model is optimized to be less than 30 MB in size and less than 60 MB in memory consumption on the mobile terminal, which verifies the advantages of both lightweight and high accuracy of the model, and is able to meet the demand for real-time detection of soybean drought phenotypes in the field. The "Early Acknowedgement for Soybean Drought" App developed in this paper integrates multispectral data analysis, canopy segmentation (accuracy >97%) and drought diagnosis, and the response time of the whole process is <300 ms, which provides a highly efficient and portable mobile solution for intelligent monitoring of soybean cultivation and precise irrigation, and significantly improves agricultural management. It provides an efficient and portable mobile solution for intelligent monitoring of soybean planting and precise irrigation, and significantly improves the intelligent level of agricultural management.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This research is funded by the National Innovation Training Program for College Students in Heilongjiang Province, Project Name: Design and Implementation of "Early Acknowedgment for Soybean Drought" App, Project No. 202410223078.

REFERENCES

[1] Li Hefeng. Steady increase of domestic soybean production capacity and self-sufficiency. Economic Daily News, 2024-11-22(006). DOI: 10.28425/n.cnki.njjrb.2024.008771.
[2] Shen Panpan. Research on the calculation method of soybean canopy wilt index based on multispectral image processing. Heilongjiang Bayi Nongken University, 2023. DOI:10.27122/d.cnki.ghlnu.2023.000122.
[3] Haoxin Song. Research on citrus information extraction from UAV multispectral images based on deep learning. Guilin University of Technology, 2023. DOI: 10.27050/d.cnki.ghlgc.2023.001051.
[4] Gao Shijiao, Guan Haiou, Ma Xiaodan, et al. A multispectral image extraction method for soybean canopy. Spectroscopy and Spectral Analysis, 2022, 42(11): 3568-3574.
[5] Fu Hongyu, Wang Wei, Lu Jianning, et al. Estimation of physical and chemical traits of ramie based on multi-spectral remote sensing by unmanned aircraft and machine learning. Journal of Agricultural Machinery, 2023, 54(05): 194-200+347.
[6] Fan Xuexing, Zhang Huichun, Zou Yiping, et al. Inversion of plant chlorophyll content based on multispectral imaging and machine learning. Forestry Science, 2023, 59(07): 78-88.
[7] Han Wenting, Cui Jiawei, Cui Xin, et al. Research on salinity estimation of agricultural soil based on feature optimization and machine learning. Journal of Agricultural Machinery, 2023, 54(03): 328-337.
[8] Huang Linsheng, Shao Song, Lu Xianju, et al. Multispectral image segmentation and alignment of lettuce based on convolutional neural network. Journal of Agricultural Machinery, 2021, 52(09): 186-194.
[9] Pang Q. Research on the detection of obvious/hidden defects in apple epidermis based on deep learning and spectral imaging . Shanghai Ocean University, 2022. DOI:10.27314/d.cnki.gsscu.2022.000147.
[10] Bedi S R, Singh A. A Feature Selection Based Relief Algorithm with Fuzzy Logic for Software Effort Estimation.Research Cell: An International Journal of Engineering Sciences, 2018, 30(SP): 1-5.
[11] Jiang C, Sun X, Dai Y, et al. EEG Emotion Recognition Employing RGPCN-BiGRUAM: ReliefF-Based Graph Pooling Convolutional Network and BiGRU Attention Mechanism. Electronics, 2024, 13(13): 2530-2530.
[12] Mahbod A, Saeidi N, Hatamikia S, et al. Evaluating pre-trained convolutional neural networks and foundation models as feature extractors for content-based medical image retrieval. Engineering Applications of Artificial Intelligence, 2025: 150110571-110571.
[13] Yifan D, Chuanbo W, Zheng W. A motor bearing fault diagnosis method based on multi-source data and onedimensional lightweight convolution neural network. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 2023, 237(2): 272-283.
[14] Carmona A M ,Sautua J F, Pérez-Hernández O, et al. AgroDecisor EFC: First Android™ app decision support tool for timing fungicide applications for management of late-season soybean diseases. Computers and Electronics in Agriculture, 2018: 144310-313.

Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3064

EXPLORING THE PATH OF VALUE NETWORK EMPOWERING THE TRANSFORMATION OF UNIVERSITY TEACHING EVALUATION

Chu Sun
School of Management, Guangdong University of Education, Guangzhou 510303, Guangdong, China.
Corresponding Email: sunchu20222022@163.com

Abstract: Aiming at the problems of single subject, unreasonable standards and inefficient feedback in Guangzhou colleges and universities teaching quality evaluation, this study introduces the value network innovation theory of Internet enterprises. The logical approach includes analyzing the logic of value network construction, interpreting existing problems, reconstructing the evaluation system, exploring resource pool construction and coordination mechanism; the path analysis emphasizes the value network construction to improve quality, user value development as the orientation, and value network realization as the goal, and promotes reform in stages; the countermeasures and suggestions involve strengthening policy guidance, promoting industry-university-research cooperation, strengthening technological innovation, establishing incentive mechanisms, and promoting experience, which will help Guangzhou build a strong education city and provide reference for regional education development.
Keywords: Value network; University teaching evaluation; Path exploration

1 INTRODUCTION

In response to the current problems in the teaching quality evaluation of colleges and universities, such as the single evaluation subject, the lack of rationality in the evaluation standards, the teacher-centered evaluation, the neglect of student needs, the low effectiveness of teaching quality evaluation, the unreasonable feedback mechanism, and the poor timeliness of evaluation result feedback, this paper aims to break the boundaries of traditional educational research, introduce the Internet enterprise value network innovation theory into the research on teaching quality evaluation and resource pool construction in colleges and universities, solve the existing problems in teaching evaluation and resource allocation in Guangzhou colleges and universities, and explore the path of building an education-strong city that fits the local industrial characteristics.

2 LOGICAL APPROACH

2.1 Analysis of the Logic of Teaching Evaluation and Resource Pool Construction in Guangzhou Universities Based on Value Network Theory

The actor network theory is introduced to regard the teaching quality evaluation and resource pool construction of Guangzhou universities as a dynamic value network system [1]. The role positioning of universities, teachers, students, enterprises, social institutions and other subjects in the network [2], as well as the flow and interaction of teaching resources, evaluation data, policy support and other elements are analyzed to explore the formation, construction and transformation status of the value network [3]. Combined with the experience of Internet companies in achieving product innovation through value networks, the uniqueness and potential advantages of value network construction in the education field of Guangzhou universities are analyzed to clarify the key role of digital transformation in value network optimization.

2.2 Interpretation of the Existing Problems of Teaching Quality Evaluation in Guangzhou Universities from the Perspective of Value Network

Re-examine the existing problems of teaching evaluation in Guangzhou universities from the perspective of value network. Analyze the problems such as the lack of network nodes caused by the single evaluation subject, the poor value transmission caused by the imperfect evaluation index system, the inefficient network interaction caused by the traditional evaluation method, and the shallow value mining caused by the insufficient data application [4]. Learn from the ideas of Internet companies to solve the coordination problems in the construction of value network, find out the bottleneck factors that restrict the digital transformation of teaching quality evaluation in Guangzhou universities and the coordinated development of value network, and highlight the necessity of reform.

2.3 Reconstruction of the Teaching Quality Evaluation System of Guangzhou Universities with Value Network Realization as the Core

Reference to the core position of value network realization in product innovation of Internet enterprises, and construct a

multi-cooperative teaching quality evaluation system of Guangzhou universities. Expand the evaluation subjects, include enterprises, industry associations, alumni, etc. into the evaluation network [5], and form a multi-node collaborative evaluation model; optimize the evaluation indicators, and design a dynamically adjusted indicator system around the underlying platform support (such as digital teaching facilities, evaluation systems), user value development (student ability improvement, teacher professional development), and value network realization (resource sharing and synergy between various subjects [5, 6]; innovate the evaluation method, use big data, artificial intelligence and other technologies to realize the real-time collection and intelligent analysis of evaluation data, promote the networking and intelligence of the evaluation process, and make the teaching quality evaluation an important link in the interaction and value creation of various subjects in the value network.

2.4 Construction Path of Guangzhou College Education Resource Pool under the Internet Enterprise Resource Integration Model

Drawing on the experience of Internet enterprises in integrating resources to build underlying platforms to support product innovation, explore the construction path of Guangzhou college education resource pool. Use digital technology to integrate course resources, teaching staff, scientific research results, training bases, etc., to create an underlying platform for sharing college education resources [7]. Establish a sharing mechanism for resource pools, clarify the rights and obligations of each subject in resource contribution, acquisition and use; design incentive mechanisms to give policy preferences, honorary awards, etc. to subjects who actively participate in resource co-construction and sharing; improve security mechanisms to ensure copyright protection and data security of resources [8]. Through the construction of resource pools, promote the complementary advantages among Guangzhou colleges and universities, form a resource synergy effect similar to that in the value network of Internet enterprises, and provide strong support for improving teaching quality.

2.5 The Value Network Linkage Mechanism of the Collaborative Development of Teaching Quality Evaluation and Resource Pool

In-depth research on the intrinsic connection between teaching quality evaluation and educational resource pool, and the construction of a collaborative development mechanism based on value network. With the realization of value network as the goal, the construction of resource pool is optimized through feedback from teaching quality evaluation, such as adjusting the direction of resource allocation and optimizing the resource allocation structure according to the evaluation results; adjusting the teaching quality evaluation standards and methods according to the dynamic update of the resource pool, such as using new resources to enrich the evaluation dimensions and improve the evaluation methods. Promote the flow of information, resource sharing and collaborative interaction between the two, and form a virtuous cycle similar to the mutual promotion and spiral rise of various elements in product innovation of Internet companies [9], promote the overall improvement of the education and teaching level of Guangzhou universities, and help build a strong city of education.

3 PATH ANALYSIS

3.1 Value Network Construction is the Key Path to Improve the Teaching Quality of Guangzhou Universities

Drawing on the successful experience of Internet companies in achieving product innovation through value networks, the teaching quality evaluation and resource pool construction of Guangzhou universities are incorporated into the value network system [10]. Each subject and element is like a node and connection in the value network. Only through effective collaboration and interaction can we achieve the optimal allocation of teaching resources, accurate evaluation of teaching quality and maximum creation of educational value, thereby improving the teaching quality of universities and promoting the construction of a strong education city.

3.2 User Value Development is the Core Orientation of Teaching Quality Evaluation and Resource Pool Construction

Similar to the fact that Internet companies take user value development as an important factor in product innovation, in the field of higher education in Guangzhou, whether it is teaching quality evaluation or resource pool construction, user value development such as student learning growth and teacher professional development should be taken as the core orientation [11]. Taking meeting user needs as the starting point and foothold [12], optimize the evaluation system and resource allocation to ensure that teaching activities can effectively enhance user value and enhance the attractiveness and competitiveness of higher education.

3.3 The Realization of Value Network is the Core Goal of the Coordinated Development of Teaching Quality Evaluation and Resource Pool

The core position of value network realization in the coordinated development of teaching quality evaluation and resource pool is emphasized, just like the key role of value network realization in product innovation of Internet

enterprises. By building a multi-faceted and collaborative value network, we can promote the deep integration and collaborative progress between teaching quality evaluation and resource pool, realize resource sharing, information exchange and collaborative innovation among various subjects [13], promote the continuous optimization and upgrading of Guangzhou higher education value network, and help achieve the goal of strengthening education city.

3.4 The Innovation Stage Theory can Provide Stage Guidance for the Education Reform of Guangzhou Universities

Referring to the theory that Chinese Internet companies' innovation has gone through three stages: investment, application, and diffusion, the reform process of Guangzhou universities' teaching quality evaluation and resource pool construction is matched with it [14]. In the investment stage, increase policy, financial and technical input; in the application stage, actively promote new evaluation systems and resource sharing models; in the diffusion stage, realize the wide dissemination and radiation of education reform results in the region [15], and promote the construction of Guangzhou universities into a strong city of education in stages and steps.

4 COUNTERMEASURES AND SUGGESTIONS

4.1 Strengthen Policy Guidance and Build an Educational Value Network Ecology

The Guangzhou Municipal Government should introduce special policies to encourage universities, enterprises, social institutions and other entities to participate in the construction of the educational value network. Set up special funds to support the construction of the educational value network platform, the development of the teaching quality evaluation system and the construction of the educational resource pool. Formulate relevant standards and specifications to promote information sharing and collaborative cooperation among various entities, create a good educational value network ecological environment, and provide policy guarantees and financial support for the teaching quality evaluation and resource pool construction of Guangzhou universities.

4.2 Promote Industry-University-Research Cooperation and Expand Value Network Nodes

Learn from the model of Internet companies and multi-party cooperation to expand the value network, and strengthen the cooperation between Guangzhou universities and local technology companies, industry associations, and scientific research institutions. Encourage enterprises to participate in the evaluation of university teaching quality, provide industry needs and practice standards; promote universities and enterprises to jointly build training bases and jointly carry out scientific research projects, and introduce enterprise resources into the education resource pool. Through industry-university-research cooperation, expand the nodes of the education value network, enrich the evaluation subjects and education resources, and improve the fit between university education and industry needs.

4.3 Strengthen Technological Innovation and Optimize Value Network Connections

Encourage universities and technology companies to cooperate in digital technology innovation and develop intelligent platforms suitable for teaching quality evaluation and resource pool management. Use technologies such as big data, artificial intelligence, and blockchain to achieve real-time collection, intelligent analysis, and precise matching of teaching data, and optimize the connection and interaction between nodes in the value network. For example, use blockchain technology to protect the copyright and data security of educational resources, and use artificial intelligence algorithms to achieve automation and intelligence in teaching quality evaluation, thereby improving the operating efficiency and synergy of the education value network.

4.4 Establish an Incentive Mechanism to Promote Value Network Collaboration

Reference to the method of Internet companies to incentivize value network subject collaboration, establish an incentive mechanism for Guangzhou colleges and universities' education value network. Reward the subjects who actively participate in teaching quality evaluation, contribute high-quality educational resources, and promote resource sharing and collaborative cooperation, such as honorary awards, financial subsidies, and policy preferences. Establish teaching quality evaluation innovation awards, education resource co-construction and sharing awards, etc., to stimulate the enthusiasm and initiative of various subjects to participate in the construction of value networks, promote the collaborative progress of various nodes in the education value network, and promote the coordinated development of teaching quality evaluation and resource pool construction.

4.5 Focus on Experience Promotion and Realize Value Network Diffusion

Summarize the successful experience and typical cases of Guangzhou universities in teaching quality evaluation and resource pool construction, and strengthen exchanges and cooperation with universities in other cities in the Guangdong-Hong Kong-Macao Greater Bay Area through holding seminars, forums, and achievement exhibitions. Promote the promotion and application of educational reform results of Guangzhou universities in the region, promote cross-regional coordination and diffusion of educational value networks, enhance the influence and radiation of

Guangzhou universities in the field of education in the Greater Bay Area, jointly promote high-quality development of regional education, and help build Guangzhou into a strong educational city.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This study was supported by the “Research on the reform of evidence-based teaching quality evaluation of Artificial Intelligence empowerment” (Project No. 2022jxgg26), Guangdong Higher Education Teaching Reform Project “Research on the reform of AI enabling evidence-based teaching quality evaluation and the construction of evaluation resource pool” and The Fourteenth Five-Year Plan for the Development of Philosophy and Social Science in Guangzhou Project “Study on Accelerating the Construction of a Strong Education City in Guangzhou—Teaching Quality Evaluation and Resource Pool Construction in Universities” (Project No.2023GZQN49), youth philosophy and Social Sciences in Guangdong project “Artificial Intelligence Enabled Evidence-Based Evaluation of Teaching Effectiveness in Universities in the Greater Bay Area” (Project No.GD23YJY06). Research Platforms and Projects of Guangdong Provincial Department of Education - Young Innovative Talents Project "Research on the Mechanism of Value Co-creation of E-commerce Digital Brand Community Empowered by Gamification in the Context of High Quality Development" (Project No.2023WQNCX052). Research on e-commerce talent training model based on digital industry-education integration (Project No. 2023YC008).

REFERENCES

[1] Quinn Ú, McCusker P, Gallagher P. Social enterprise development in Ireland: making a case for actor-network-theory. Journal of Business and Socio-Economic Development, 2025. DOI: 10.1108/jbsed-06-2024-0061.

[2] Appleton J J, Christenson S L, Furlong M J. Student engagement with school: Critical conceptual and methodological issues of the construct. Psychology in the Schools, 2008, 45(5): 369–386. DOI: 10.1002/pits.20303.

[3] Li T, He P, Peng L. Measuring high school student engagement in science learning: an adaptation and validation study. International Journal of Science Education, 2023, 46(6): 524–547. DOI: 10.1080/09500693.2023.2248668.

[4] Boston M D, Candela A G. The Instructional Quality Assessment as a tool for reflecting on instructional practice. ZDM, 2018, 50(3): 427–444. DOI: 10.1007/s11858-018-0916-6.

[5] Hein A M, Jankovic M, Feng W, et al. Stakeholder power in industrial symbioses: A stakeholder value network approach. Journal of Cleaner Production, 2017, 148: 923–933. DOI: 10.1016/j.jclepro.2017.01.136.

[6] Howson K, Johnston H, Cole M, et al. Unpaid labour and territorial extraction in digital value networks. Global Networks, 2022, 23(4): 732–754. DOI: 10.1111/glob.12407.

[7] Flanders N A. Intent, action and feedback: A preparation for teaching. Journal of Teacher Education, 1963, 14(3): 251–260. DOI: 10.1177/002248716301400305.

[8] Harrison R, Meyer L, Rawstorne P, et al. Evaluating and enhancing quality in higher education teaching practice: a meta- review. Studies in Higher Education, 2022, 47(1): 80–96. DOI: 10.1080/03075079.2020.1730315.

[9] Dang P, Geng L, Niu Z, et al. A value-based network analysis for stakeholder engagement through prefabricated construction life cycle: evidence from China. Journal of Civil Engineering and Management, 2024, 30(1): 49–66. DOI: 10.3846/jcem.2024.20726.

[10] Blumenstock J E, Chi G, Tan X. Migration and the value of social networks. Review of Economic Studies, 2023, 92(1): 97–128. DOI: 10.1093/restud/rdad113.

[11] Shi Y, Zhao Y. Digital green value co-creation behavior, digital green network embedding and digital green innovation performance: moderating effects of digital green network fragmentation. Humanities and Social Sciences Communications, 2024, 11: 228. DOI: 10.1057/s41599-024-02691-5.

[12] Martinez Lunde I. Learning analytics as modes of anticipation: enacting time in actor-networks. Scandinavian Journal of Educational Research, 2022, 68(2): 218–232. DOI: 10.1080/00313831.2022.2123851.

[13] Goodchild B, Ferrari E. Intermediaries and mediators: an actor-network understanding of online property platforms. Housing Studies, 2021, 39(1): 102–123. DOI: 10.1080/02673037.2021.2015297.

[14] Conti E, Farsari I. Disconnection in nature-based tourism experiences: an actor-network theory approach. Annals of Leisure Research, 2022, 27(4): 525–542. DOI: 10.1080/11745398.2022.2150665.

[15] Taraza E, Anastasiadou S, Papademetriou C, et al. Evaluation of quality and equality in education using the European foundation for quality management excellence model—A literature review. Sustainability, 2024, 16(3): 960. DOI: 10.3390/su16030960.

Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3065

FOREST FIRE POINT RECOGNITION BASED ON SUPER-RESOLUTION TECHNIQUES

FuLin Li, WenFa Xu*, Zhen Min, XuePeng Wu, ChangYu Xiang, YuDong Liu
School of Information Science and Engineering, Wuchang Shouyi University, Wuhan 430064, Hubei, China.
Corresponding Author: WenFa Xu, Email: xuwenfa@wsyu.edu.cn

Abstract: Forest fires, as a global ecological disaster, pose a serious threat to the stability of environmental systems, biodiversity security, and socio-economic development. In response to the shortcomings of traditional fire point detection methods in small target recognition accuracy, model robustness, and real-time performance, this paper proposes an intelligent forest fire point detection model based on an improved YOLOv8n framework. This model integrates the Fast Super Resolution Convolutional Neural Network feature enhancement module, Enhanced Squeeze Excitation attention mechanism, and an improved Minimum Point Distance Intersection over Union bounding box regression algorithm, aiming to improve the detection ability of early fire points and overall system performance. Through systematic experiments on a self built multi scene forest fire image dataset, the results showed that compared with Faster R-CNN, YOLOv5s, and the standard YOLOv8n model, the proposed method performed well in comprehensive detection performance, with mAP reaching 84.7%, Precision reaching 84.2%, Recall reaching 80.4%, and also possessing high real-time processing capabilities. This study not only provides effective technical support for intelligent monitoring of forest fires, but also proposes a multi module collaborative optimization framework, which provides new theoretical references and practical paths for research and application in the field of small target detection.
Keywords: Forest fire point recognition; Super-resolution; YOLOv8n; Object detection; Deep learning

1 INTRODUCTION

Forests, as one of the most important terrestrial ecosystems on Earth, play an irreplaceable role in global carbon cycling, climate regulation, and biodiversity conservation [1]. However, due to the impact of global climate change, forest fires have been frequent and expanding in scale in recent years. The "Black Summer" forest fires in Australia in 2020 burned over 24 million hectares of forest land, causing nearly 3 billion animals to die or be displaced; In 2021, a forest fire broke out in Liangshan Prefecture, Sichuan Province, China, causing dozens of casualties and damage to tens of thousands of acres of forest land. According to a report released by the United Nations Environment Programme in 2022, the direct economic losses caused by forest fires worldwide exceed $50 billion annually, and ecological losses are difficult to estimate.

How to achieve early monitoring and rapid response to forest fires has gradually become an important research direction for modern intelligent forestry and emergency response. Early intelligent fire monitoring methods mainly relied on the color features, temperature information, and smoke dynamics of remote sensing images for fire point extraction. Algorithms included color thresholding, background subtraction, and thermal infrared recognition [2-3], but these methods generally had problems such as high false alarm rates and strong dependence on the environment [4].

With the rapid development of deep learning technology, object detection algorithms based on convolutional neural networks have been widely explored and applied in the field of forest fire point recognition. Early research mainly used Faster R-CNN [5] and YOLO series models, among which YOLO demonstrated significant engineering applicability in multi-source heterogeneous data scenarios of unmanned aerial vehicle aerial video and fixed monitoring equipment due to its real-time advantage of single-stage detection [6]. In response to the technical difficulties of small scale and weak features of forest fire point targets, researchers have introduced super-resolution reconstruction algorithms such as FSRCNN and EDSR to construct a preprocessing module, effectively improving the spatial resolution of input images and enhancing the model's ability to extract features from small-scale fire points. It is worth noting that in recent years, visual Transformer architecture has shown breakthrough progress in the field of fire detection. Swin Transformer has achieved global context modeling through a hierarchical window attention mechanism, while DETR has significantly improved algorithm robustness under complex background interference based on an end-to-end detection framework.

Although forest fire point recognition technology based on deep learning has made significant progress in image understanding and object detection, it still faces many challenges in practical applications. Firstly, forest fire points often exhibit extremely small early scales, irregular shapes, and blurred edges, which are easily disturbed by complex backgrounds, making it difficult for detection models to accurately capture their features and prone to missed and false detections. Secondly, due to the suddenness and danger of forest fires, there are safety hazards in the process of obtaining real fire image data, and early fire point samples are extremely scarce, resulting in uneven sample distribution in most datasets, especially the lack of diverse small fire point instances, which affects the training quality and generalization ability of the model. In addition, although lightweight object detectors such as YOLOv5 have been applied in fire point recognition, their actual deployment on edge devices such as drones and front-end monitoring terminals is still limited by factors such as computing power, power consumption, and transmission delay, making it difficult to balance high accuracy and real-time performance.

Therefore, improving the ability to detect small targets, building high-quality multi scene datasets, and implementing deployment solutions that balance lightweight and accuracy will be key directions for future fire point recognition research. In order to solve the above problems, this paper proposes a forest fire point recognition method based on YOLOv8 object recognition algorithm and fused with image super-resolution. This method improves the detection accuracy of fire points while being lightweight, providing new ideas for the field of fire point recognition.

2 REAL TIME YOLO OBJECT DETECTION WITH FSRCNN FUSION

YOLOv8 is the latest generation object detection model in the YOLO series launched by the Ultralytics team in 2023. As an unofficial continuation and architectural innovation version of YOLOv5, YOLOv8 has been systematically optimized in model design, training strategy, and inference efficiency. YOLOv8 supports tasks including object detection, image segmentation, and pose estimation, making it an ideal choice for achieving tasks such as object detection. This article has made the following improvements to the YOLOv8 model. The introduction of FSRCNN (Fast Super Solution Convolutional Neural Network) proposed by Dong et al. on ECCV in 2016 [7] for image super-resolution reconstruction has improved the recognition accuracy of small targets; To enhance the feature representation of fire points, we drew inspiration from the eSE attention mechanism module proposed by Gao et al. in 2021 [8] and added a feature optimization module that can explicitly model channel relationships; At the same time, the MPDloU bounding box regression loss function proposed by Li Yang et al. in 2022 [9] was introduced to improve YOLOv8, thereby enhancing the model's ability to detect low resolution fire points and achieving higher accuracy and stronger robustness in forest fire point recognition.

2.1 Hyperfractionation Module

As an improved version of SRCNN, the FSRCNN model not only achieves better reconstruction quality, but also significantly improves computational efficiency. The core innovation of FSRCNN lies in abandoning the strategy of SRCNN first interpolating and then convolving, and instead directly extracting features from low resolution images, and achieving resolution improvement at the end of the network through deconvolution operation, effectively reducing the redundancy of previous calculations. In addition, FSRCNN introduces a five stage structure of "feature extraction reduction nonlinear mapping expansion deconvolution", which significantly reduces computational complexity and improves nonlinear expression ability by adding a 1x1 convolution bottleneck module in the middle layer. The grid structure diagram of the FSRCNN super-resolution model is shown in Figure 1.

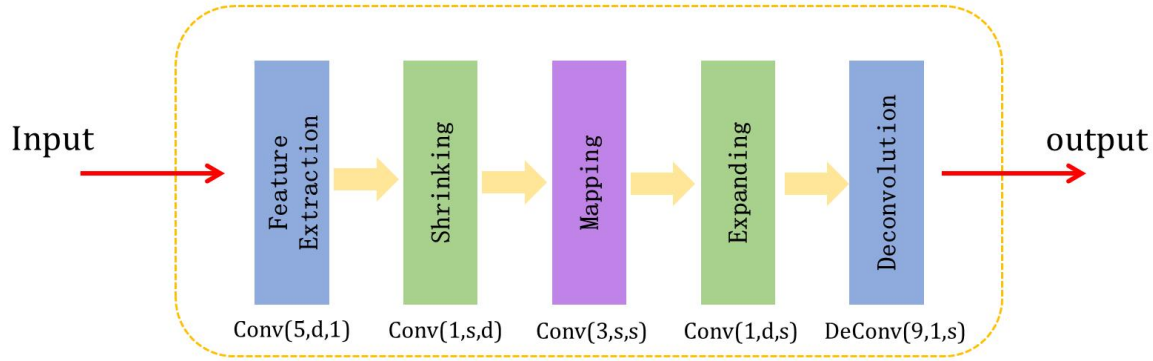


Figure 1 FSRCNN Grid Structure Diagram
Note: Feature Extraction is the feature extraction layer, Shrinking is the feature compression layer, Mapping is the non-linear mapping layer, Expanding is the channel expansion layer, and Deconvolution is the deconvolution/upsampling layer

FSRCNN has lower latency while maintaining good image reconstruction quality, making it particularly suitable for real-time visual applications that are sensitive to computing resources. In subsequent tasks such as object detection and small object recognition, FSRCNN is also widely used in the image preprocessing stage to improve image clarity and detail expression ability, thereby enhancing the detection model's ability to recognize fine-grained objects.
The core idea of FSRCNN is to directly learn end-to-end from low resolution images, using shallow convolutional networks to learn the mapping relationship from low resolution to high-resolution images, avoiding the multi-stage image reconstruction process in traditional super-resolution methods. This model adopts feature mapping reduction processing to reduce computational complexity, and then upsampling to restore high-resolution images, significantly improving computational efficiency and processing speed. Its mathematical expression is:

$$I_{SR} = \int (I_{LR}; \theta) \quad (1)$$

Among them θ is the grid parameter, $\int(\cdot)$ is the network, I_{LR} is the input image, I_{SR} is the input image.
Compared to other models, FSRCNN adopts a more efficient design, including simplified feature extraction and upsampling processes, which significantly accelerates inference speed while maintaining high image reconstruction

quality.

According to Dongetal's research "Accelerating the Super Solution Neural Network"[10], the speed improvement of FSRCNN compared to SRCNN is mainly due to the following reasons: (1) reducing the dimensionality and computational complexity of the input image through feature dimensionality reduction; (2) By optimizing the deconvolution operation in the upsampling stage, the computational cost has been reduced; (3) Through end-to-end training, the redundant calculation process of the model has been reduced. The experimental results in the literature indicate that FSRCNN improves speed several times compared to SRCNN and maintains a high level of performance in image quality.

Therefore, integrating the FSRCNN super-resolution model into the YOLO object recognition model can effectively improve the quality of the input image, enabling the model to perform accurate recognition even in the face of low resolution. At the same time, the lightweight design of FSRCNN can also enable the fusion model to maintain high accuracy while having good response speed.

2.2 eSE Attention Module

ESE attention mechanism (Efficient Squeeze and Excitation Attention Mechanism) is a technique for optimizing the channel attention mechanism in convolutional neural networks, aiming to improve the expression ability and efficiency of the model. The eSE attention mechanism models the dependency relationships between channels to adaptively adjust the weights of different channels, enhancing the network's attention to useful features while reducing the impact of unimportant information. Compared with traditional SE mechanisms, eSE significantly improves computational efficiency and reduces computational complexity while maintaining performance through further optimization of computation and parameter quantity.

The core idea of eSE attention mechanism is to guide the network to focus on important features and suppress irrelevant or redundant information by weighting the channels of the feature map. The biggest improvement of the eSE mechanism lies in its optimization of the traditional SE mechanism. Specifically, in the Excitation stage, eSE uses a more efficient operation, avoiding the computational complexity of the fully connected layer in the traditional SE mechanism.

The specific principle of eSE attention mechanism is shown in Figure 2. The input image is first fed into an average pooling Favg, then into a 1x1 convolutional layer Wc, and finally, output feature maps with different weights are obtained through an h-sigmoid activation function.

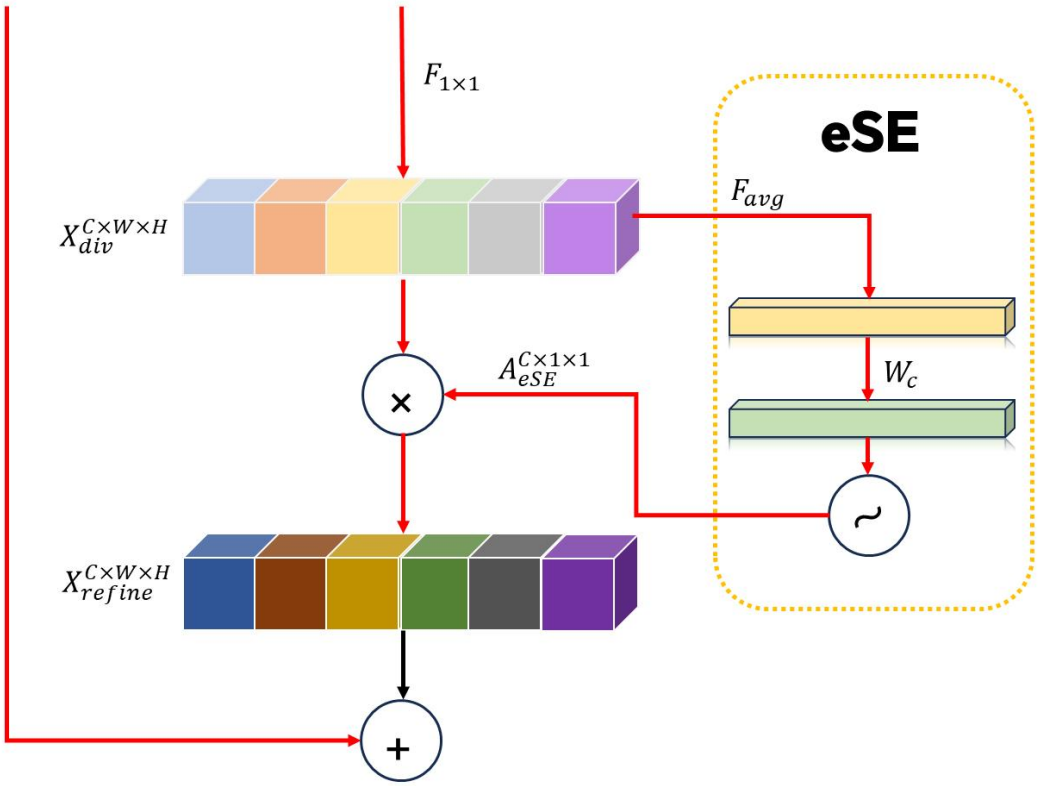


Figure 2 eSE Principle Diagram of Attention Mechanism

In addition, the eSE attention mechanism optimizes the structure of traditional SE modules by enhancing the modeling of inter channel dependencies. The specific mathematical process can be expressed as follows:

2.2.1 Input feature map:

Given an input feature map $X \in R^{C\times W\times H}$, where is the number of channels and is the spatial dimension。

2.2.2 Global average pooling：

Compress the space of each channel and generate channel descriptors：

$$E_{avg}(X_c) = \frac{1}{W\times H}\sum_{i=1}^{W}\sum_{j=1}^{H} X_c(i,j)\ ,\ \ c = 1,2,\ldots,C \quad (2)$$

Output vector： $\boldsymbol{z \in R^{C\times 1\times 1}}$。

2.2.3 Channel weight learning：

Through an uncompressed **1x1** convolutional layer $\boldsymbol{W_c}$（Preserve channel dimension）and h-sigmoid activation function generates attention weights：

$$A_{eSE} = \sigma_h(W_c z),\ \ W_c \in R^{C\times C},\ \ \sigma_h(\cdot) = h - sigmoid \quad (3)$$

Among them, $\boldsymbol{\sigma_h}$ is the improved sigmoid function (such as Hard Sigmoid) enhances nonlinearity and computational efficiency.

2.2.4 Feature recalibration：

Multiply the weight $\boldsymbol{A_{eSE}}$ with the input feature map channel by channel to obtain the refined output：

$$X_{refine} = X \odot A_{eSE} \quad (4)$$

⊙ represents channel wise multiplication.

In summary, the eSE attention mechanism, as an improved channel attention module, enhances the important features of fire points and suppresses redundant information by dynamically recalibrating the fire point images, significantly reducing the computational complexity of the model while maintaining high performance.

2.3 MPD-IoU Bounding Box Regression Loss Function

MPD IoU is a loss function used for bounding box regression in object detection tasks, designed to improve the fit between detection boxes and real boxes, while enhancing the robustness of the model to difficult to locate samples. Figure 3 shows the schematic diagram of the principle of MPD IoU.

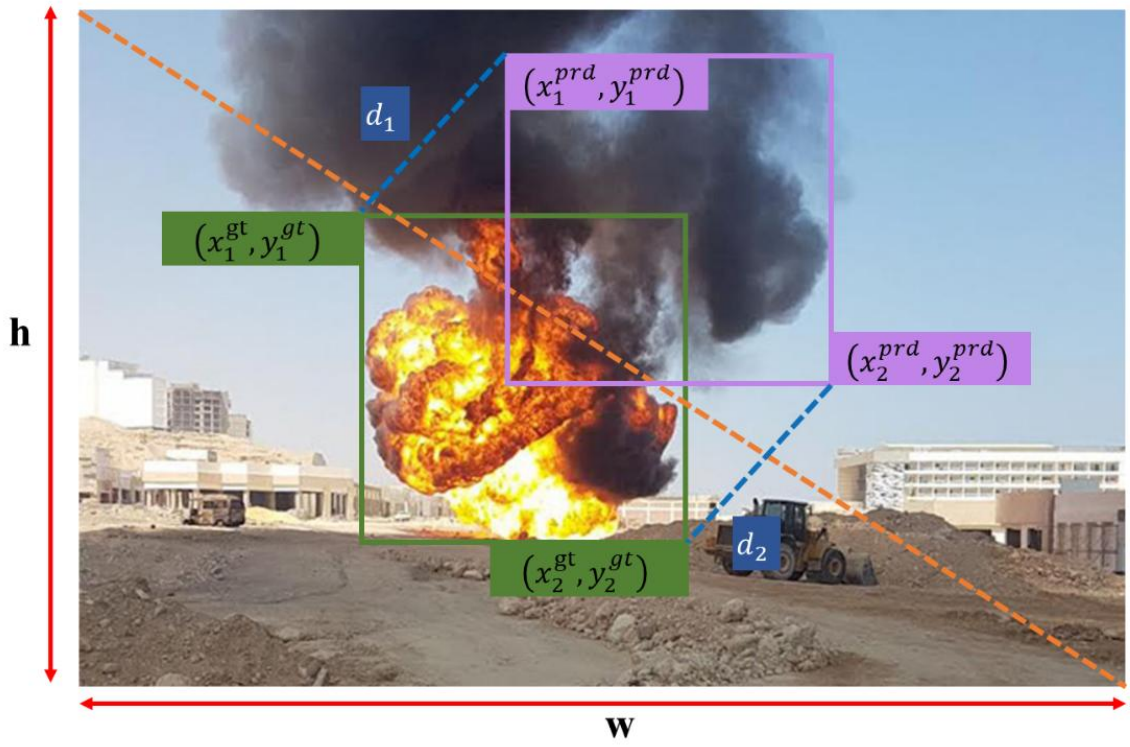


Figure 3 Schematic Diagram of MPD IoU Principle

Note: The purple box represents the prediction box, and the green box represents the target box. (x_1^{gt}, y_1^{gt}) represents the upper left corner coordinate of the annotation box; (x_2^{gt}, y_2^{gt}) represents the coordinates of the upper right corner of the annotation box; (x_2^{gt}, y_2^{gt}) represents the coordinates of the upper right corner of the annotation box; (x_2^{gt}, y_2^{gt}) represents the coordinates of the upper right corner of the annotation box; $\overline{d_1}$ represents the minimum point distance between the annotation box and the prediction box, which is the offset between the diagonals of the two boxes (shown by the blue dashed line in the figure).

Unlike traditional IoU loss functions, MPD IoU integrates multidimensional penalty factors such as center offset and aspect ratio consistency, and combines weighted average IoU with sample dynamic penalty mechanism to significantly

improve the boundary fitting ability for high aspect ratio targets and fuzzy boundary targets. In the detection of small and irregular targets, this loss function has higher gradient sensitivity, effectively alleviating the problem of gradient vanishing during the training process. Meanwhile, MPD loU optimizes bounding box regression by directly minimizing the offset distance between the predicted box, annotated box, and the four corner points of the annotated box, rather than relying on traditional LOU or geometric constraints. The loss function can be expressed as:

$$\zeta_{MPDloU} = 1 - loU + \lambda \cdot \overline{d_1} \quad (5)$$

Among them, loU represents the intersection ratio; $\overline{d_1}$ represents the normalized value of the distance between diagonal points (key innovation point); λ represents balanced weight

Introducing MPD IoU into the fire point detection framework effectively alleviates the gradient degradation and localization error problems of traditional IoU loss in small object detection. The fusion of dynamic penalty and multi-dimensional geometric modeling mechanism significantly improves the fitting accuracy and stability of the model to the bounding box of fire points, especially suitable for complex remote sensing scenes with sparse fire point samples, significant scale differences, and strong noise interference.

In summary, the introduction of MPD IoU can further improve the localization accuracy of regression branches in the fire point detection framework, enhance the model's adaptability to high deformation and multi-scale targets in fire scenes, and achieve more reliable flame and smoke detection results.

3 EXPERIMENT AND RESULT ANALYSIS

3.1 Dataset Establishment

Due to the limited availability of credible open source fire datasets and the lack of annotation in most existing datasets, it is difficult to directly use them for training. Therefore, we constructed a fire dataset. This dataset covers scenarios such as forest fires, indoor fires, traffic fires, building fires, outdoor fires, and candle flames. As shown in Table 1, the dataset contains 10295 images, divided in a 14:3:3 ratio. The training set contains 7206 images, while the validation set and test set each contain 1544 images. The specific distribution is shown in Table 1.

Table 1 Distribution of Fire Point Dataset

category	Data set	Training set	Validation set	Test set
Quantity (in sheets)	10,295	7,206	1,544	1,544

3.2 Experimental Environment

This research experiment was conducted in the following hardware and software environments: the hardware platform used high-performance computing devices equipped with NVIDIA RTX 3090 graphics cards, providing powerful parallel computing capabilities for deep learning tasks. In terms of software environment, the experiment was built on the Windows 10 operating system, using Python 3.8 as the programming language environment, and implementing algorithm design and model training using the PyTorch 2.1.0 deep learning framework. YOLOv8n was used as the core training tool for the object detection model, with an input image size of 512 × 512, a model training period of 300, a batch size of 16, an initial learning rate of 0.01, and a cosine annealing strategy to dynamically adjust the learning rate.

3.3 Evaluation

Adopt the following four indicators：

$$Presicion = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$mAP = \frac{\sum P_A}{N_c} \quad (8)$$

$$mloU = \frac{1}{N}\sum_{i=1}^{N} IoU_i \quad (9)$$

TP, FP, and FN are used to correctly detect, incorrectly detect, and miss detection targets, respectively. They represent the similarity measure between a single predicted result and the true annotation. The average value of all categories is used to evaluate the efficiency of the network. At the same time, the number of image frames processed per second is selected to evaluate the effectiveness of the network in the application scenario of fire point image recognition.

3.4 Ablation experiment

In order to verify the impact of each module on overall performance, this paper conducted a systematic ablation experiment and set the following model combinations：

1. YOLOv8n
2. YOLOv8n+FSRCNN
3. YOLOv8n+FSRCNN+eSE
4. YOLOv8n+FSRCNN+eSE+MPDloU

Table 2 shows the performance of each model on the validation set:

Table 2 Data of ablation experiment results

Number	FSRCNN	eSE	MPDIoU	mAP	Precision	Recall	MIoU	FPS/ms
1	×	×	×	0.78	0.81	0.77	0.72	9.2
2	√	×	×	0.80	0.83	0.79	0.71	12.4
3	√	√	×	0.83	0.84	0.81	0.79	12.7
4	√	√	√	0.84	0.84	0.80	0.81	13.1

Note: √ indicates the introduction of the module, x indicates the non introduction of the module

According to the data in the table above, the introduction of each module has a significant improvement in model performance. Specifically, number 1 only introduces the FSRCNN module, which increases mAP and Precision to 0.78 and 0.81 respectively compared to the baseline, but slightly lowers Recall and MIoU. Number 2 added eSE module on this basis, increasing mAP and Precision to 0.80 and 0.83 respectively, and Recall also slightly increased to 0.79. Number 3 introduces the MPDIoU module on the original basis, further increasing mAP to 0.83 and significantly improving all indicators. In the end, Number 4 integrated all three modules and achieved the best overall performance, with mAP reaching 0.84, Precision and Recall being 0.84 and 0.80, respectively, and MIoU improving to 0.81 while maintaining a high inference speed.

It can be seen that introducing the FSRCNN module alone can effectively improve detection accuracy, but there is a certain trade-off between recall and inference speed. After further introducing the eSE module, the overall performance of the model steadily improved, indicating that the attention mechanism is helpful for feature extraction. After combining the MPDIoU module, not only is the detection accuracy further improved, but the positioning accuracy is also significantly enhanced. When integrating all modules, the model achieved the best performance in all indicators, verifying the effectiveness and superiority of the proposed multi module fusion strategy in improving detection accuracy and maintaining high inference speed.

3.5 Visual Comparison of Model Parameters

To further verify the accuracy improvement effect of the model in this article, a visual parameter comparison was also conducted. The result is shown in Figure 4. The green box represents the prediction results of the YOLOv8n model, the red box represents the prediction results of the improved model in this paper, and the blue arrow indicates the differences in the block diagrams between the previous and subsequent models.

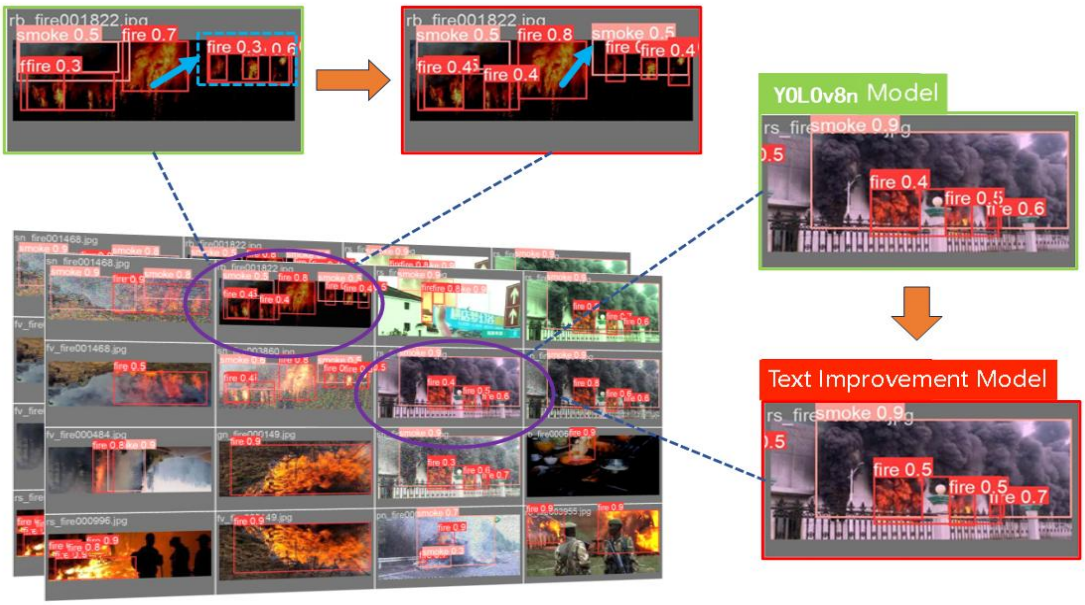


Figure 4 Comparison of Detection Performance between YOLOv8n Model and Text Model

As shown in the figure, YOLOv8n network has the problem of false detection when detecting flames with overlapping and unclear features, and has a high false detection rate for smoke. The improved model can effectively detect the above-mentioned false alarms of flames and smoke, and the confidence level is also higher than that of the YOLOv8n model. The qualitative results further indicate that the detection performance of the improved model is superior to that of the original model, and it has universality on the new dataset.

3.6 Model Comparison Experiment

Compare the performance of the improved model proposed in this article with the following classic object detection models on the same dataset and training configuration. The detailed results are shown in Table 3:

Table 3 Comparison of Fire Point Detection Results of Different Models

Model	mAP	Precision	Recall	FPS
FasterR-CNN	0.76	0.80	0.74	5.2
YOLOv5s	0.77	0.82	0.78	14.5
YOLOv8n	0.78	0.81	0.77	15.8
Model in this article	0.84	0.84	0.80	13.1

From the table, it can be seen that Faster R-CNN performs relatively stably on mAP and Precision, but has low recall and inference speed of only 5.2 FPS, resulting in low inference efficiency. YOLOv5s is slightly better than Faster R-CNN in terms of detection accuracy and inference speed, with better overall performance. YOLOv8n has further improved its inference speed to 15.8 FPS, but the mAP improvement compared to YOLOv5s is only 0.1, while Precision and Recall are slightly lower than YOLOv5s. In contrast, the model proposed in this study achieved the highest values in the three key performance indicators of mAP, Precision, and Recall, and the inference speed also reached 13.1 FPS, which is at a moderate level.

From the results in Table 3, it can be seen that traditional Faster R-CNN is inferior to lightweight detection models in terms of accuracy and inference speed. The YOLO series models exhibit a good speed accuracy balance in fire point detection tasks, especially YOLOv5s and YOLOv8n, which have significant advantages in inference speed. However, the improved model proposed in this study outperforms the comparative model in terms of detection accuracy and maintains a high level of inference speed, verifying that the proposed method has better comprehensive performance and application potential in forest fire point detection tasks.

4 CONCLUSION

This article proposes a novel detection method based on YOLOv8n fusion FSRCNN super-resolution module, eSE attention mechanism, and MPDIoU regression loss to address the challenges of small target recognition, fuzzy features, and high real-time requirements in forest fire point detection tasks. Through sufficient experimental verification on a self built multi scene fire point dataset, the method proposed in this paper outperforms existing mainstream detection frameworks in terms of accuracy, recall, and comprehensive detection performance, and effectively captures fine-grained fire point features while maintaining high inference speed. The ablation experiment showed that the FSRCNN module significantly improved the resolution of small targets, the eSE module enhanced the feature expression ability, and the MPDIoU improved the localization accuracy. The comprehensive comparison results have verified the superiority and applicability of the proposed method in early identification of forest fire points. Future work will further explore lightweight super-resolution methods and self supervised learning mechanisms to enhance detection performance in small sample environments.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This article is funded by the 2024 College Student Innovation and Entrepreneurship Training Program (Project Number: 202412309002)

REFERENCES

[1] Pan YD, Birdsey R, Fang JY, et al. A large and persistent carbon sink in the world's forests. Science, 2011, 333(6045): 988–993.
[2] Li J, Hu F, Du L. Forest Fire Detection Based on Image Processing Technology: A Review. Fire Technology, 2020, 56(5): 1945-1972.
[3] Yuan F, Wan W, Wei H, et al. Review of remote sensing methods for forest fire detection. Remote Sensing, 2018, 10(5): 763.
[4] Koltunov A, Ustin S L. Early Fire Detection Using Non-Linear Multispectral Analysis: International Journal of Wildland Fire, 2007, 16(1): 1-10.
[5] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks: IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
[6] Yuan Z, Zhang Y, Liu Y, et al. A Real-Time Forest Fire Detection Algorithm Based on UAV Video: Remote Sensing, 2021, 13(12): 2305.
[7] Dong C, Loy C C, He K, et al. Accelerating the Super-Resolution Convolutional Neural Network: IEEE Transactions on Image Processing, 2016, 25(12): 5534-5545.

[8] Gao Z, He W, Luo H, et al. ESNet: Efficient Squeeze-and-Excitation Network for Channel Attention: IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(6): 2567-2579.
[9] Li Y, Huang T, Liu Y, et al. Bounding box regression method based on improved dynamic IoU. Journal of Software, 2022, 33(11): 4667–4680.
[10] Dong C, Loy C C, He K, et al. Accelerating the Super-Resolution Convolutional Neural Network: IEEE Transactions on Image Processing, 2016, 25(12): 5534-5545.

Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3066

DESIGN OF A MICROCONTROLLER-BASED SMART BLIND CANE

WenHao He[1*], GuangBo Su[2], ZhiJie Wang[2], ShuMing Li[2]
[1]*School of Business, Guilin University of Electronic Technology, Guilin 541004, Guangxi, China.*
[2]*School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, Guangxi, China.*
Corresponding Author: WenHao He, Email: 13377013467@163.com

Abstract: Individuals with visual impairment encounter substantial difficulties in navigating daily activities due to their inability to perceive their immediate surroundings, particularly during pedestrian crossings at intersections. To address these mobility constraints, this research proposes the development of an intelligent assistive cane incorporating a single-chip microcontroller system. The cane uses a modular design, mainly integrating an STM32F103C8T6 minimum system module, an ultrasonic ranging module, an OpenARTmini vision sensor module, and a voice synthesis module. The intelligent guide cane integrates an ultrasonic ranging module with a visual sensor module, enabling the recognition of traffic light signals and traffic signs while achieving three-dimensional environmental perception. This design overcomes the limitations of ecological sensing imposed by traditional single-sensor systems. Additionally, it incorporates a composite feedback mode consisting of buzzer alarms, vibration feedback, and speech synthesis, which facilitates differentiated information transmission tailored to diverse scenario requirements. By meeting the daily needs of blind individuals through context-specific information delivery, the system enhances their independent mobility and promotes social participation.
Keywords: Blind mobility assistance; Smart navigation cane; Modular design; STM32F103C8T6 minimum system

1 INTRODUCTION

According to the World Health Organization (WHO), visual impairment is one of the most common problems affecting approximately one sixth of the world's population [1]. There are approximately 40 million to 45 million blind people globally, and the number of blind people is still on the rise each year. One of the most crucial senses in human existence is sight [2]. However, visually impaired people have this handicap. They face numerous difficulties in their daily lives, among which the problem of getting around is particularly challenging.

Scholars such as Sharma T and Nalwa T designed a smart cane equipped with infrared and ultrasonic sensors, which can effectively detect obstacles [3]. Pariti J, Tibdewal V, and other scholars developed a smart mobile cane that notifies users through different parts of the hand to convey obstacle-related information [4]. Aravinth T S created a WiFi- and Bluetooth-based smart cane, optimizing its usage methods to enhance accessibility for blind individuals [5]. Soares J M S, Guerra C S D, and other scholars designed an electronic laser-assisted smart cane capable of identifying obstacles at the user's head height, which holds significant importance for the development of environmental perception technology in smart guide canes [6]. Tang J, Sun M, and other scholars developed a visual odometry-assisted cane that directly detects obstacles and attracts the attention of surrounding people, providing multiple layers of protection for the visually impaired [7]. Scholars, including Abu-Abdoun D I and Alsyouf I, innovatively designed a GPS-enabled cane by integrating the specific needs of blind individuals, further facilitating their daily mobility [8]. These research findings provide valuable references for the design of the smart blind cane proposed in this paper. Nevertheless, existing smart blind canes demonstrate limitations in environmental perception: they depend exclusively on a single sensor, exhibit inadequate modular integration, and fail to recognize traffic light colors or issue corresponding alerts when visually impaired individuals encounter traffic lights during their mobility.

This paper uses the STM32F103C8T6 as the main control chip, integrating multiple modules to achieve low-cost dual-mode detection of obstacles and traffic lights. By combining a three-level alarm mechanism with preset interaction parameters, it breaks through the limitations of traditional devices in complex intersection scenarios and enhances the safety of blind individuals when traveling.

2 OVERALL SYSTEM DESIGN

2.1 Design Scheme

The blind guidance system is composed of multiple modules, including a power supply circuit, STM32F103C8T6 minimum system module, HC-SR04 ultrasonic ranging module, OpenART mini vision sensor module, SYN6288 Chinese speech synthesis module, buzzer and vibration motor module, as well as an OLED display module.. This intelligent white cane design prioritizes safe mobility for the visually impaired as its central objective, focusing on two core functionalities: obstacle distance detection and traffic light color recognition at intersections. Through differentiated communication protocols coordinating with the STM32F103C8T6 microcontroller, these modules

collectively form an integrated guidance system that combines environmental perception, signal processing, and multimodal interaction. This architecture ensures comprehensive functional integration and stable system operation, as shown in Figure 1.

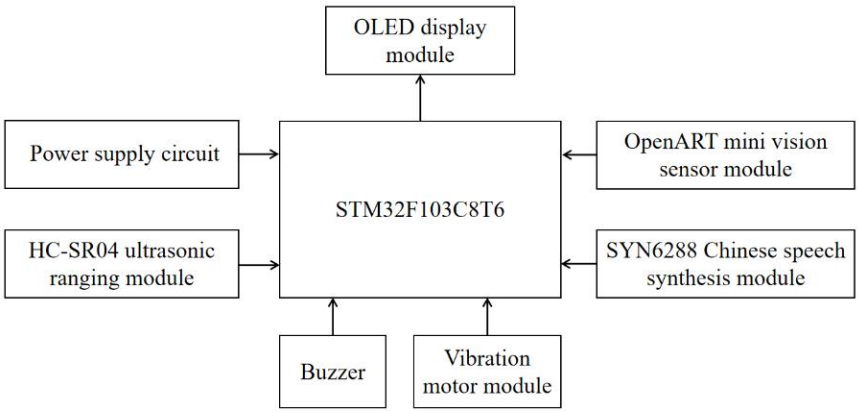


Figure 1 Overall Framework Diagram of the System

2.2 System Hardware Design

2.2.1 Main control chip

The design employs the STM32F103C8T6 microcontroller as the central control unit. It has strong computing capabilities and abundant interfaces, which can meet the requirements of the Intelligent Blind Stick and facilitate the efficient control of each module. The STM32F103C8T6 has a total of five pins connected to the positive pole of the power supply and four ground pins, which supply power to different internal modules respectively. It can control modules such as HC-SR04, SYN6288, and OpenART mini in real time. For HC-SR04, the specific pins of the single-chip microcomputer output control signals for distance measurement. For example, PA0 provides a trigger signal to Trig, and PA1 receives the Echo signal, and the distance is calculated through a formula. For SYN6288, the single-chip microcomputer sends commands and text through the UART interface, sets the baud rate, etc., and the synthesized voice of the module is amplified and played. For OpenART mini, the single-chip microcomputer transmits data through asynchronous serial communication. The module is programmed with OPENMV IDE and the PYTHON language. It collects images of traffic lights, makes judgments, and then informs the single-chip microcomputer of the color situation through the serial port. Then the system reminds the user by controlling the voice module, as shown in Figure 2.

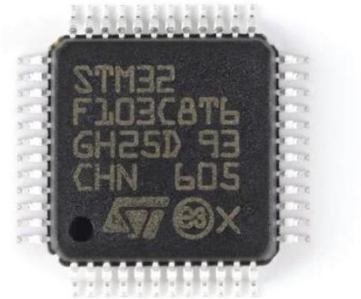


Figure 2 STM32F103C8T6 Chip

2.2.2 Ultrasonic ranging module

The HC-SR04 ultrasonic ranging module is an important part of the Intelligent Blind Stick. It is designed to accurately measure the distance between the blind stick and the obstacles in front. With excellent performance, its range covers a wide interval from as close as 2 centimeters to as far as 400 centimeters, and within this range, its ranging accuracy can reach an amazing 3 millimeters. The module is ingeniously constructed, consisting of three parts: an ultrasonic transmitter, a receiver, and a control circuit. Each part undertakes a crucial task, jointly ensuring the accuracy and stability of ranging. In the intelligent blind guidance cane, this module is connected to the microcontroller, responsible for emitting and receiving ultrasonic waves, and transmits the measured data to the microcontroller for processing, as shown in Figure 3.

Figure 3 HC-SR04 Ultrasonic Ranging Module

2.2.3 SYN6288 Chinese speech synthesis module

The human-machine interaction system operates through the SYN6288 Chinese speech synthesis driver software. It is a speech synthesis module that is used to play Chinese speech, and can support words synthesized using English letters [9]. This driver communicates with the microcontroller via asynchronous serial interfaces, processing text encoded in GB2312, GBK, and Unicode standards. Universal Asynchronous Receiver Transmitter is a serial communication protocol that helps in communicating data between devices [10]. This module also has abundant control functions, such as synthesis, stop, pause, continue, and baud rate adjustment. Through speech synthesis algorithms, it converts text into natural, intelligible audio output, achieving real-time text-to-speech conversion. The full-duplex communication module connects to the main chip through TXD and RXD pins. When receiving command frames from the main controller, it processes structured data packets containing frame headers, data length indicators, command codes, text content, and XOR checksums. For instance, commands like "Front red light, do not advance" trigger corresponding voice alerts through internal processing. Based on visual sensor input regarding traffic light status, the system activates specific voice prompts: "Red or yellow light ahead, stop" or "Green light ahead, proceed." The module supports parameter adjustments, including volume levels, speech rate modulation, and background music integration, ensuring optimized auditory reception of traffic signals for visually impaired users. This implementation demonstrates dynamic audio feedback driven by environmental sensing and standardized communication protocols, as shown in Figure 4.

Figure 4 SYN6288 Chinese Speech Synthesis Module

2.2.4 OpenART mini vision sensor module

The OpenART mini vision sensor module employs the high-performance MIMXRT1064 processor and integrates the OpenMV machine vision library. It supports data storage via Type-C interface or SD card, while its built-in image sensor captures 240×320 resolution images. By defining the value ranges for the L, a, and b components corresponding to red, yellow, and green colors in the Lab color space, the module achieves accurate identification of traffic light color zones. Operating continuously at intersections, the module transmits recognition results through a serial port at a 115200 baud rate to the microcontroller. This provides critical input for the SNT6288 speech synthesis module to determine traffic signal states. This innovation effectively resolves the critical challenge of blind individuals being unable to discern traffic lights at crossings. The system converts visual recognition outcomes into audible voice commands, enabling safe and precise navigation through intersections. This technological advancement significantly enhances the adaptive capabilities of visually impaired individuals to surrounding traffic conditions during independent travel, as shown in Figure 5.

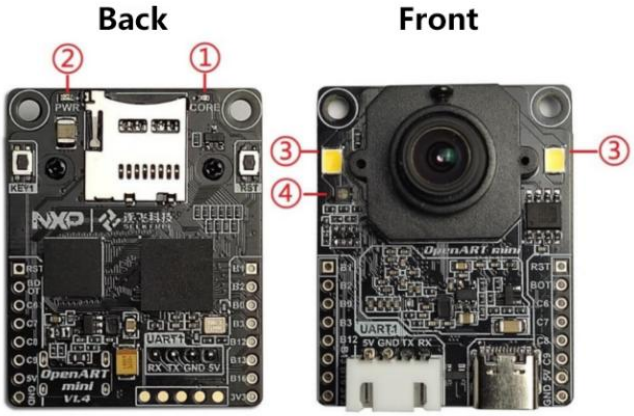


Figure 5 Indicator Lights of the OpenART mini Intelligent Vision Sensor Module

2.2.5 OLED display module

The OLED display module features a 0.96-inch monochromatic screen with 128 × 64 resolution, utilizing the I2C communication protocol to interface with the microcontroller. Driven by the SSD1306 controller chip, it supports 256-level brightness adjustment and operates as a self-emissive device requiring no backlight, making it ideal for low-power applications. In this system, font generation software converts ultrasonic ranging results and traffic light status data into character codes for real-time visualization. This provides graphical feedback for visually impaired users, such as dynamically updated obstacle distance measurements. Combined with voice prompts and vibration alerts, it establishes a multimodal feedback mechanism to enhance environmental perception. The display's visual supplementation proves particularly effective in quiet environments, offering complementary data through graphical cues. Additional configurability includes volume, speech rate, and background music adjustments to facilitate auditory reception of traffic signals. This integration of visual and auditory modalities enhances the system's practical applicability, as shown in Figure 6.

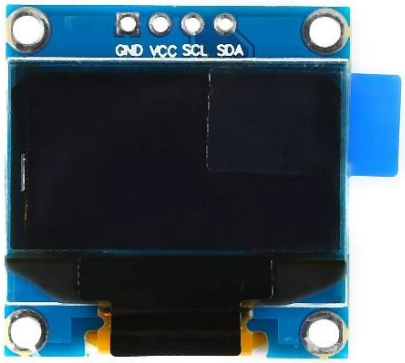


Figure 6 Overall Framework Diagram of the System

2.3 System Software Design

The system program is developed using the C language within the Keil 5 MDK integrated development environment. Its functional architecture comprises multiple core modules, with the main program module serving as the control hub. This hub assumes critical responsibilities, including system initialization configuration, task timing scheduling, and resource coordination and management. Another key component is the HC-SR04 ultrasonic ranging module subroutine, which operates based on the principle of ultrasonic ranging. It transmits ultrasonic waves, receives reflected waves from the surrounding environment to detect road conditions, and identifies the presence of obstacles. When an object is detected within a 50cm range, the system triggers corresponding voice prompts according to the current environmental status, thereby establishing a closed control loop from environmental perception to information feedback. Additionally, the software includes the OpenART mini visual sensor subroutine, which is capable of detecting traffic lights. Upon detecting a red or yellow light, the system invokes the SYN6288 Chinese voice synthesis subroutine to provide prompts such as "Red/yellow light ahead, do not proceed." When a green light is detected, it prompts "Green light ahead, proceed." The software also features an OLED display subroutine, which shows prompt information such as distance and time on the OLED screen, enabling users to easily monitor the current environment and system status. Furthermore, the vibration and buzzer control subroutines deliver dual tactile and auditory alerts, ensuring users effectively receive system notifications. Finally, the software can rapidly generate target code, which is concise and understandable, facilitating maintenance and future upgrades.

3 RESULT

3.1 Distance Measurement Test

This study conducted obstacle detection accuracy tests on the intelligent navigation cane using four standard distances, with the test results shown in Table 1.

Table 1 Precision Test of HC-SR04 Ultrasonic Ranging Module

Distance (cm)	Accuracy rate
25	98%
50	97%
75	93%
100	91%

The results indicate that the system achieves over 95% accuracy in short-distance obstacle detection, with minor deviations observed as distance increases. These deviations are primarily attributed to ultrasonic signal attenuation at longer ranges. Within the 100 cm operational range, assisted obstacle avoidance can be effectively achieved, demonstrating compliance with the design specifications. The observed performance degradation beyond proximity aligns with inherent physical limitations of ultrasonic wave propagation, particularly energy dissipation effects over extended distances.

3.2 Traffic Light Recognition Test

This study evaluated the recognition performance of the OpenART mini module for three traffic light colors under a standardized testing distance of 30 cm, with experimental outcomes shown in Table 2.

Table 2 Recognition Accuracy Table

Colour	Accuracy rate
Red light	90%
Yellow light	93%
Green light	96%

The experimental results demonstrate that the green light recognition achieved the highest accuracy rate, followed by yellow and red lights. The study attributes this performance hierarchy to green light's superior distinctiveness in the Lab color space compared to other colors, coupled with the module's stable recognition of high-luminance colors. Red light detection showed relative susceptibility to ambient light interference due to lower brightness levels, potentially leading to threshold misjudgments. While red light recognition accuracy remained above 90% despite illumination challenges, and yellow light exhibited intermediate performance, the system's overall capability meets the operational requirements for traffic signal identification in close-range scenarios as specified for navigation cane applications.

4 CONCLUSIONS AND OUTLOOKS

The intelligent cane system distinguishes itself from existing commercial alternatives by integrating ultrasonic distance sensors, vision sensors, and voice synthesizers, enabling enhanced environmental exploration, obstacle avoidance, and traffic light recognition for visually impaired users. Through optimized hardware circuit design and refined software logic architecture, this system demonstrates robust adaptability to complex environments while minimizing system crashes during operation, thereby improving travel safety.

However, this system still has some shortcomings, including a relatively limited visual recognition range and non-expandable hardware components in the main control module. In the future, improving the algorithm could enable it to operate better under varying light conditions, thereby further enhancing the adaptability of the intelligent cane system.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This work was financially supported by the 2024 Guangxi Zhuang Autonomous Region Student Entrepreneurship and Innovation Program: Smart Navigation, Cane-Guided Path: Design of an Intelligent Guide Cane Based on Real-Time

Traffic Feedback (202410595028) fund.

REFERENCES

[1] Masal K M, Bhatlawande S, Shingade S D. Development of a visual to audio and tactile substitution system for mobility and orientation of visually impaired people: a review. Multimedia Tools and Applications, 2024, 83(7): 20387-20427.
[2] Singh H, Kumar H, Sabarivelan S, et al. IoT based Smart Assistance for Visually Impaired People. 2023 9th International Conference on Electrical Energy Systems (ICEES), Chennai, India. 2023, 118-122.
[3] Sharma T, Nalwa T, Choudhury T, et al. Smart cane: Better walking experience for blind people. 2017 3rd International Conference on Computational Intelligence and Networks (CINE), Odisha, India. 2017, 22-26.
[4] Pariti J, Tibdewal V, Oh T. Intelligent mobility cane-lessons learned from evaluation of obstacle notification system using a haptic approach. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA. 2020, 1-8.
[5] Aravinth T S. WiFi and Bluetooth based smart stick for guiding blind people. 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India. 2020, 317-320.
[6] Soares J M S, Guerra C S D, da Silva M M, et al. Smart Cane: Laser guide as an inclusion tool for the visually impaired. Proceedings of the XX Brazilian Symposium on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA. 2021, 1-10.
[7] Tang J, Sun M, Zhu L, et al. Design and optimization of an assistive cane with visual odometry for blind people to detect obstacles with hollow section. IEEE Sensors Journal, 2021, 21(21): 24759-24770.
[8] Abu-Abdoun D I, Alsyouf I, Mushtaha E, et al. Developing and designing an innovative assistive product for visually impaired people: smart cane. 2022 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates. 2022, 1-6.
[9] Sung W T, Chen G R, Hsiao S J. Voice Guidance System for Color Recognition Based on IoT. Computer Systems Science & Engineering, 2023, 45(1): 839-855.
[10] Persiya Gnana Golda D, Deborah S P, Suji Sharon J, et al. Verification of UART using AHB VIP with Maximum Coverage//2024 7th International Conference on Devices, Circuits and Systems (ICDCS), Coimbatore, India. 2024, 234-238.

Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3067

OPTIMIZATION STRATEGIES FOR WIRELESS COMMUNICATION NETWORKS BASED ON MACHINE LEARNING

ZiXuan Wu
School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China.
Corresponding Email: 1594243499@qq.com

Abstract: With the rapid development of wireless communication technology, higher demands have been placed on network performance and resource utilization efficiency. This paper deeply explores the application of machine learning in wireless communication network optimization, analyzes the challenges faced by current wireless communication networks, such as scarce spectrum resources, unbalanced network load, and difficult energy consumption management. From a technical perspective, it elaborates on the application methods of machine learning in key fields such as spectrum management, mobility management, and energy consumption management, including the detailed implementation of machine learning-based spectrum allocation algorithms, mobility prediction models, and energy consumption optimization strategies. The paper also delves into the underlying principles, parameter tuning, and performance evaluation of relevant algorithms. Through theoretical analysis and practical case verification, this paper demonstrates how machine learning technologies effectively enhance the performance of wireless communication networks and achieve efficient resource utilization, providing theoretical support and practical guidance for the development of future wireless communication networks.
Keywords: Wireless communication; Machine learning; Network optimization; Spectrum management; Mobility management; Deep learning; Reinforcement learning

1 INTRODUCTION

Wireless communication technology has become an indispensable part of modern society. From mobile communications to emerging fields such as the Internet of Things (IoT) and vehicle networks, the application scope of wireless communication continues to expand. However, with the growth of user numbers and the diversification of service demands, wireless communication networks face numerous challenges. Traditional network optimization methods gradually reveal limitations in coping with complex and dynamic network environments, while the development of machine learning technologies provides new ideas and methods for wireless communication network optimization[1]. Machine learning can automatically learn patterns and rules from large amounts of data, adapt to dynamic network environments, and achieve network performance optimization and efficient resource utilization. In recent years, deep learning and reinforcement learning algorithms have shown significant advantages in handling complex wireless communication scenarios, making them key research directions in this field[2].

2 CHALLENGES FACED BY WIRELESS COMMUNICATION NETWORKS

2.1 Scarce Spectrum Resources

Wireless spectrum is a limited natural resource. With the explosive growth of wireless communication services (e.g., 5G/6G networks, millimeter-wave communication), spectrum resources have become increasingly scarce[3]. Different communication systems and services fiercely compete for spectrum. The static spectrum allocation policy in traditional networks leads to low utilization, with some bands remaining idle while others suffer from congestion. For example, in densely populated urban areas, numerous mobile base stations, Wi-Fi hotspots, and IoT devices simultaneously compete for limited spectrum resources, leading to severe signal interference and degraded communication quality. To address this, dynamic spectrum access (DSA) techniques have emerged, but their implementation requires accurate spectrum sensing and intelligent allocation algorithms[4].

2.2 Unbalanced Network Load

User distribution and service demands are uneven in time and space, leading to unbalanced network loads in wireless communication networks. Network congestion occurs in certain regions or periods with excessively high loads, while network resources remain underutilized in other regions or periods. For instance, in densely populated places such as shopping malls, stadiums, or during large-scale online events, network traffic surges, overloading base stations and deteriorating the user experience. Traditional load balancing methods, such as fixed routing or manual resource adjustment, are insufficient to handle real-time and dynamic load changes. Therefore, intelligent load prediction and autonomous adjustment mechanisms are urgently needed.

2.3 Difficult Energy Consumption Management

The energy consumption problem in wireless communication networks has become increasingly prominent. With the increase in the number of base stations (especially small cells in 5G networks) and the improvement of equipment power, network energy consumption continues to rise. On one hand, high energy consumption costs impose economic pressures on operators; on the other hand, substantial energy consumption is inconsistent with the requirements of sustainable development. Conventional energy-saving strategies, like static power control or periodic device hibernation, lack adaptability to dynamic traffic changes. Machine learning can analyze traffic patterns and environmental factors to achieve more precise energy management.

3 APPLICATIONS OF MACHINE LEARNING IN WIRELESS COMMUNICATION NETWORK OPTIMIZATION

3.1 Spectrum Management

3.1.1 Spectrum sensing with deep learning

Deep learning-based spectrum sensing utilizes convolutional neural networks (CNNs) or recurrent neural networks (RNNs) to analyze time-frequency domain data collected by spectrum sensors. CNNs, with their hierarchical feature extraction capabilities, can effectively identify spectral occupancy patterns. For example, 1D-CNNs can process time-series signal samples, while 2D-CNNs are suitable for spectrogram analysis. By training on a large dataset of signal waveforms, CNNs can distinguish between occupied and idle bands with high accuracy, reducing false alarm and missed detection rates.

3.1.2 Reinforcement learning for spectrum allocation

Reinforcement learning (RL) algorithms, such as Deep Q-Network (DQN) and Proximal Policy Optimization (PPO), are widely used for dynamic spectrum allocation. In RL-based systems, the agent (e.g., a base station or a user device) interacts with the environment (spectrum status) by taking actions (allocating spectrum bands) and receiving rewards (e.g., throughput improvement or interference reduction). For instance, DQN combines deep neural networks with Q-learning to approximate the optimal action-value function, enabling the agent to learn long-term optimal strategies in complex multi-user scenarios. Parameter tuning, such as adjusting the learning rate and discount factor, is crucial for algorithm convergence.

3.2 Mobility Management

3.2.1 Movement prediction with LSTM networks

Long Short-Term Memory (LSTM) networks, a variant of RNNs, are effective for predicting user movement trajectories due to their ability to handle sequential data with long-term dependencies[5]. LSTM models take historical location data, time stamps, and environmental features (e.g., road maps, user behavior patterns) as inputs. By learning temporal correlations, LSTM can predict future positions with high precision. For example, in a vehicular network, an LSTM-based model can predict the driving route of vehicles, allowing base stations to pre-allocate resources and initiate handovers in advance, reducing packet loss and handover latency.

3.2.2 Handover decision with deep reinforcement learning

Deep reinforcement learning (DRL) can optimize handover decisions by considering multiple factors, such as signal strength, available bandwidth, and user QoS requirements. For example, a DRL agent can learn to select the optimal target base station by maximizing cumulative rewards related to connection stability and throughput. In a multi-cell network, DRL algorithms can balance the load among cells while ensuring seamless handovers, improving overall network efficiency.

3.3 Energy Consumption Management

3.3.1 Traffic prediction and power control

Machine learning algorithms, such as Gaussian Process Regression (GPR) and Long Short-Term Memory networks, can predict network traffic patterns based on historical data, time-of-day, and user behavior. GPR provides probabilistic predictions, enabling operators to estimate the uncertainty of traffic variations. Based on traffic forecasts, power control algorithms adjust the transmission power of base stations and devices. For example, when traffic is low, base stations can reduce power or enter sleep mode, while during peak hours, power is increased to maintain service quality.

3.3.2 Joint optimization with deep learning

Deep learning can jointly optimize multiple network parameters (e.g., power, spectrum, and user scheduling) to minimize energy consumption[6]. Autoencoder-based models can compress network state information for efficient decision-making, while deep neural networks can learn complex mappings between network conditions and optimal control strategies. For instance, a deep neural network can be trained to optimize the trade-off between energy consumption and network throughput by adjusting the transmit power and resource allocation in real-time.

4 CASE STUDIES

4.1 Spectrum Optimization for a City's Mobile Network

A mobile network operator in a city adopted a machine learning-based spectrum optimization scheme. By deploying spectrum sensing devices, large amounts of spectrum data were collected, and a 1D-CNN was used for spectrum occupancy detection. The CNN model achieved an accuracy of 92% in identifying idle bands. Subsequently, a DQN-based algorithm allocated spectrum resources among users, considering factors such as interference and data rate requirements. After optimization, the spectrum utilization rate of the city's mobile network increased by 25%, network congestion was significantly alleviated, and the average download rate of users increased by 35%.

4.2 Mobility Management for a Large Enterprise Campus Wireless Network

In a large enterprise campus with frequent personnel and equipment movements, an LSTM-based mobility prediction model was deployed. The model integrated historical location data from Wi-Fi access points, Bluetooth beacons, and employee work schedules. By training on three months of data, the LSTM model achieved a prediction accuracy of 88% for user movement within 15 minutes. The wireless network system used the prediction results to pre-configure resources for target access points, reducing the handover failure rate from 12% to 3% and decreasing communication interruption time by 85%.

4.3 Energy Consumption Optimization for Base Stations in a Region

Mobile communication base stations in a region adopted a machine learning-based energy consumption optimization system. A GPR model was used to predict traffic demand, with an average mean absolute error of 10% compared to actual traffic[7].Based on traffic forecasts, a deep neural network optimized base station power and resource allocation. The system achieved a 20% reduction in overall energy consumption while maintaining a 95% user satisfaction rate for network quality.

5 CONCLUSION

Machine learning technologies have demonstrated great potential in wireless communication network optimization. Through applications in key fields such as spectrum management, mobility management, and energy consumption management, they effectively address numerous challenges faced by wireless communication networks and improve network performance and resource utilization efficiency[8]. However, challenges remain, such as the high computational complexity of deep learning models and the requirement for large-scale labeled data. Future research should focus on developing lightweight machine learning algorithms, enhancing privacy-preserving techniques, and exploring the integration of edge computing to enable real-time optimization[9]. With the continuous development and improvement of machine learning algorithms and the sustained evolution of wireless communication technologies, machine learning will play a more critical role in future wireless communication networks.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Yu H, Zhao Y, Mo L. Fuzzy Reliability Assessment of Safety Instrumented Systems Accounting for Common Cause Failure. IEEE Access, 2020, 8: 135371-135382. DOI: 10.1109/ACCESS.2020.3010878.

[2] Zeng P, Zhang Z, Lu R, et al. Efficient Policy-Hiding and Large Universe Attribute-Based Encryption With Public Traceability for Internet of Medical Things. IEEE Internet of Things Journal, 2021, 8(13): 10963-10972. DOI: 10.1109/JIOT.2021.3051362.

[3] Phuong N H. FuzzRESS: A fuzzy rule-based expert system shell combining positive and negative knowledge for consultation of Vietnamese traditional medicine. 2016 Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS), 2016: 1-6. DOI: 10.1109/NAFIPS.2016.7851624.

[4] Tham M L, Iqbal A, Chang Y C. Deep reinforcement learning for resource allocation in 5G networks and beyond. 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2019: 1852-1855.

[5] Piester D, Bauch A, Peik E, et al. An Uncertainty Study on Traceable Frequency and Time with Disciplined Oscillators for Metrology and Financial Sectors. 2019 Joint Conference of the IEEE International Frequency Control Symposium and European Frequency and Time Forum (EFTF/IFC), 2019: 1-4. DOI: 10.1109/FCS.2019.8856142.

[6] Dangi V. Future frontiers: Artificial intelligence's influence on cybersecurity dynamics. Journal of Computer Science and Electrical Engineering, 2024, 6(4): 7–13. DOI: 10.61784/jcsee3016.

[7] Ahmed R H, Hussain M, Abbas H, et al. Enhancing autonomous vehicle security through advanced artificial intelligence techniques. Journal of Computer Science and Electrical Engineering, 2024, 6(4): 1–6. DOI: 10.61784/jcsee3017.

[8] Zhou Z, Abawajy J. Reinforcement learning-based edge server placement in the intelligent Internet of Vehicles environment. IEEE Transactions on Intelligent Transportation Systems, 2025(99): 1–11. DOI: 10.1109/TITS.2025.3557259.

[9] Zhou Z, Shojafar M, Abawajy J, et al. IADE: An Improved Differential Evolution Algorithm to Preserve Sustainability in a 6G Network. IEEE Transactions on Green Communications and Networking, 2021, 5(4): 1747–1760. DOI: 10.1109/TGCN.2021.3111909.

Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3068

SELECTION STRATEGY FOR FULL-SPECTRUM VERSUS LOCAL-SPECTRUM ENERGY DETECTION UNDER CONSTANT FALSE ALARM RATE CONDITIONS

YuXin Cheng
School of Intelligent Equipment, Shandong University of Science and Technology, Taian 271000, Shandong, China.
Corresponding Email: 15166744447@163.com

Abstract: In information science and electrical engineering, energy detection is a foundational tool for dynamic spectrum access and environmental awareness, playing a critical role in modern wireless communication systems. While Local Spectrum Energy Detection (LSED) is commonly regarded as superior to Full Spectrum Energy Detection (FSED), its performance advantage is not guaranteed across all real-world conditions. In this study, we derive closed-form expressions for the false alarm and detection probabilities of both FSED and LSED under the constant false alarm rate (CFAR) criterion. To address varying signal distributions, we develop an analytical framework based on the SNR ratio and categorize the frequency band selection problem into two typical cases: complete and partial capture of signal energy. Monte Carlo simulations validate our theoretical findings and reveal the specific conditions under which LSED provides measurable benefits over FSED, such as in narrowband signal environments or when prior knowledge of the signal's spectral characteristics is available. Furthermore, we propose a novel frequency-band energy ratio metric, which quantifies the relative energy distribution across different bands, enabling adaptive and resource-efficient detection method selection in complex electromagnetic environments. This metric facilitates intelligent switching between FSED and LSED based on real-time spectrum conditions. This work offers theoretical insights and practical guidance for signal detection, spectrum sensing, and data-driven decision-making in electronic communication and IoT-enabled systems, contributing to more efficient spectrum utilization and improved reliability in next-generation wireless networks.
Keywords: Constant false alarm rate; Energy detection; Full spectrum; Local spectrum

1 INTRODUCTION

Energy detection technology is widely used in the field of cognitive radio due to its reliable performance and relatively simple implementation [1]. Energy detection can be divided into time-domain energy detection (TDED) and frequency-domain energy detection (FDED) [2-4]. TDED uses the sum of the squares of the time-domain amplitudes of the sensing signals as the test statistic, while FDED uses the accumulation of the power spectrum within the sensing bandwidth as the test statistic.

FDED can be classified into two categories based on the spectrum of interest. When the focus is on the entire power spectrum, and the sum of the power spectra of all spectral lines within the detection length is used as the test statistic, it is referred to as Full Spectrum Energy Detection (FSED) [5]. In contrast, if only a specific sub-band is considered, and the sum of the power spectra of spectral lines from the start to the end of the band is used as the test statistic, it is referred to as Local Spectrum Energy Detection (LSED) [6].

LSED is widely favored in signal detection applications. Some authors proposed a two-level fuzzy logic-based detection system for local spectrum sensing [7]. A strategy using a simple recursive estimator was introduced to enhance the reliability of local spectrum sensing in cognitive radio networks [8]. A spectrum sensing technology designed for reliable signal detection in very low SNR environments was explored [9]. A local spectrum sensing model was established in cognitive wireless sensor networks to improve achievable throughput and enhance robustness across different environments [10].

It is generally believed that Local Spectrum Energy Detection (LSED) can achieve better detection performance than Full Spectrum Energy Detection (FSED) by focusing on frequency bands with more concentrated signal energy; however, this assumption has not been fully validated. LSED targets a selected sub-band, where the number of spectral lines and the signal-to-noise ratio (SNR) vary across different frequency regions, resulting in varying detection performance. This paper investigates the SNR relationship between LSED and FSED, with particular emphasis on how the symmetry of signal energy distribution influences detection efficiency. A comparative analysis of the detection performance of LSED and FSED is conducted, and the Excellent Energy Ratio (EER) is introduced based on the derived SNR relationship. The EER serves as a robust criterion for selecting LSED and offers practical guidance for energy detection method selection in real-world applications.

This paper analyzes the detection performance of both methods under the Constant False Alarm Rate (CFAR) condition [11], where the false alarm probability remains constant, and the detection performance is determined by their respective detection probabilities. Recently, CFAR technology has made significant breakthroughs. A deep learning framework was developed to train neural networks that approximate CFAR [12]. A novel constant false positive detection method based on the Gaussian Mixture Model (GMM), using Gabor wavelets to determine the number of

GMM components, was proposed [13]. Bayesian optimization was integrated into the Maximum Constant False Alarm Rate (GO-CFAR) for the prediction of maximum frequency hopping (FH) [14]. Future research will explore further advancements in CFAR technologies.

Additionally, we conducted Monte Carlo simulation experiments, using sinusoidal signals to simulate narrowband signals and Ricker wavelets to simulate broadband signals [15-16]. We then compared the actual and theoretical probabilities. The results confirm that the theoretical analysis is correct.

The remainder of this paper is organized as follows: Section 2 analyzes the mathematical model of signal detection in the frequency domain and derives the formulas for the detection probabilities of FSED and LSED. Section 3 evaluates the detection performance of both methods under various conditions and outlines the method selection strategy. Section 4 presents simulation experiments using narrowband and broadband signals to validate the theoretical derivation in Section 3. Section 5 discusses the relevant aspects of the research process. Finally, Section 6 concludes the paper.

2 THEORETICAL MODEL

The frequency-domain energy detection model is the core component of energy detection technology, primarily used to analyze the energy distribution of signals in the frequency domain. This model can be divided into two methods: Full Spectrum Energy Detection (FSED) and Local Spectrum Energy Detection (LSED). FSED focuses on the entire spectrum, determining the presence of a signal by summing the power spectra of all frequency components. On the other hand, LSED concentrates on a specific frequency band of interest, analyzing only the frequency components within a certain range. This approach is equivalent to applying a band-pass filter to the signal before performing energy detection. In the signal detection process, the received signal can be considered as a binary hypothesis problem, where the signal either exists or does not exist. To implement frequency-domain energy detection, the received signal needs to be transformed from the time domain to the frequency domain using a discrete Fourier transform. After the transformation, the spectrum is divided into positive and negative frequency parts. However, to simplify the analysis, only the positive frequency part of the spectrum is typically considered. By calculating the power spectrum using the periodogram method, the energy distribution at each frequency point can be obtained, providing a basis for further detection.

2.1 Full Spectrum Energy Detection Model

FSED takes the sum of all spectral lines in the power spectrum as the test statistic, that is:

$$T_{FD} = \sum_{k=0}^{N/2-1} P(k) \tag{1}$$

If the signal-to-noise ratio is $\gamma = \frac{2}{N^2\sigma_z^2} \sum_{k=0}^{N/2-1} \left(X_R^2(k) + X_I^2(k) \right)$, then, for sufficiently large detection length N, the central limit theorem indicates that the N independent random variables follow a Gaussian distribution:

$$T_{FD} \sim \begin{cases} \mathcal{N}\left(N\sigma_z^2/2, N\sigma_z^4/2 \right) & H_0 \\ \mathcal{N}\left((1+\gamma)N\sigma_z^2/2, (1+2\gamma)N\sigma_z^4/2 \right) & H_1 \end{cases} \tag{2}$$

Let the decision threshold be η_{FD} . According to the Neyman-Pearson criterion, the following expression can be derived:

$$P_f = \Pr\left(T_{FD} \geq \eta_{FD} \mid H_0 \right) = Q\left(\frac{\eta_{FD} - \frac{N}{2}\sigma_z^2}{\sqrt{\frac{N}{2}}\sigma_z^2} \right), \tag{3}$$

$$\eta_{FD} = \left(Q^{-1}(P_f)\sqrt{N/2} + N/2 \right)\sigma_z^2, \tag{4}$$

$$P_d = \Pr\left(T_{FD} > \eta_{FD} \mid H_1 \right) = Q\left(\frac{Q^{-1}(P_f) - \gamma\sqrt{\frac{N}{2}}}{\sqrt{1+2\gamma}} \right), \tag{5}$$

where P_f is the false alarm probability, P_d is the detection probability, and Q is the right-tail probability function of the standard normal distribution.

2.2 Local Spectrum Energy Detection Model

If the frequency band of interest consists of d spectral lines P(i), P(i+1) …… P(i+d-1), then the detection statistic can be defined as:

$$T_{LS}(i,d)=\sum_{k=i}^{i+d-1}P(k) \tag{6}$$

where i represents the number of the spectral line at the beginning of the frequency band, and d represents the number of spectral lines. In the simulation experiment, if the selected frequency band range is 20-50Hz, i is 20 and d is 30. Let the local spectral SNR of this band be $\gamma_i'=\frac{1}{dN\sigma_z^2}\sum_{k=i}^{i+d-1}\left(X_R^2(k)+X_I^2(k)\right)$, which reflects the ratio of effective signal to noise energy in the band of interest. LSED can be represented by an approximate Gaussian distribution. According to the central limit theorem, we can get:

$$T_{LS}\sim\begin{cases}\mathcal{N}\left(d\sigma_z^2,d\sigma_z^4\right) & H_0\\ \mathcal{N}\left((1+\gamma')d\sigma_z^2,(1+2\gamma')d\sigma_z^4\right) & H_1\end{cases} \tag{7}$$

Set the detection threshold as η_{LS}, then:

$$P_f=\Pr\left(T_{LS}\geq\eta_{LS}\right)=Q\left(\frac{\eta_{LS}-d\sigma_z^2}{\sqrt{d}\sigma_z^2}\right), \tag{8}$$

$$\eta_{LS}=\sqrt{d}\sigma_z^2Q^{-1}(P_f)+d\sigma_z^2\ , \tag{9}$$

$$P_d=\Pr\left(T_{LS}>\eta_{LS}\mid H_1\right)=Q\left(\frac{Q^{-1}(P_f)-\gamma'\sqrt{d}}{\sqrt{1+2\gamma'}}\right). \tag{10}$$

3 SELECTION STRATEGY ANALYSIS

Through the derivation above, under the Constant False Alarm Rate (CFAR) condition, the detection performance of both methods depends on their respective detection rates. The detection rates of both methods are described using Q functions, so the performance comparison essentially boils down to comparing the Q function values under identical conditions. Since the Q function is a monotonically decreasing function, the larger the argument, the smaller its value. Let the inverse of the Q function be denoted as $A=Q^{-1}(P_f)$, where A is also a decreasing function. Since a high false alarm rate is generally considered meaningless, the range of P_f is constrained to (0, 0.5], which implies that A > 0.

We define the function $f(A,N,\gamma)$ as follows:

$$f(A,N,\gamma)=\frac{A-\gamma\sqrt{N/2}}{\sqrt{1+2\gamma}}\ . \tag{11}$$

Let the function $f(A,2d,\gamma')$ be defined as follows:

$$f(A,2d,\gamma')=\frac{A-\gamma'\sqrt{d}}{\sqrt{1+2\gamma'}}\ . \tag{12}$$

The two functions are the independent variable functions in equations (9) and (14). The smaller their values, the greater the corresponding detection probability. Moreover, $1\leq d<N/2$, with only the spectrum of the positive frequency component being considered.

Let the ratio of the signal-to-noise ratio of the two methods be a, that is:

$$\gamma'=a\gamma \tag{13}$$

Analyze separately the cases where the frequency band contains a portion of the signal energy and the cases where the frequency band contains the entire signal energy.

3.1 The Frequency Band Contains a Portion of the Signal Energy

3.1.1 The SNR of LSED is lower than that of FSED

In this case, a < 1. First, let's analyze the monotonicity of the function $f(A,N,\gamma)$. The partial derivative of $f(A,N,\gamma)$ with respect to N is:

$$\frac{\partial f(A,N,\gamma)}{\partial N}=\frac{-\gamma}{2\sqrt{2}\sqrt{1+2\gamma}\sqrt{N}}<0\ . \tag{14}$$

When γ and A are constant, $f(A,N,\gamma)$ decreases as N increases. The partial derivative of $f(A,N,\gamma)$ with respect to γ is given by:

$$\frac{\partial f(A,N,\gamma)}{\partial \gamma}=\frac{-A-\sqrt{N/2}\ -\gamma\sqrt{N/2}}{(1+2\gamma)\sqrt{1+2\gamma}}<0\ . \tag{15}$$

When N and A are constant, $f(A,N,\gamma)$ decreases as γ increases.

Based on the analysis above, since d < N/2, we have $f(A,N,\gamma)<f(A,2d,\gamma)$. Furthermore, since $\gamma'<\gamma$, it follows that $f(A,2d,\gamma)<f(A,2d,\gamma')$. Therefore, $f(A,N,\gamma)<f(A,2d,\gamma')$.

Given that the Q function is a decreasing function, the detection probability of FSED is higher, making it preferable to choose FSED.

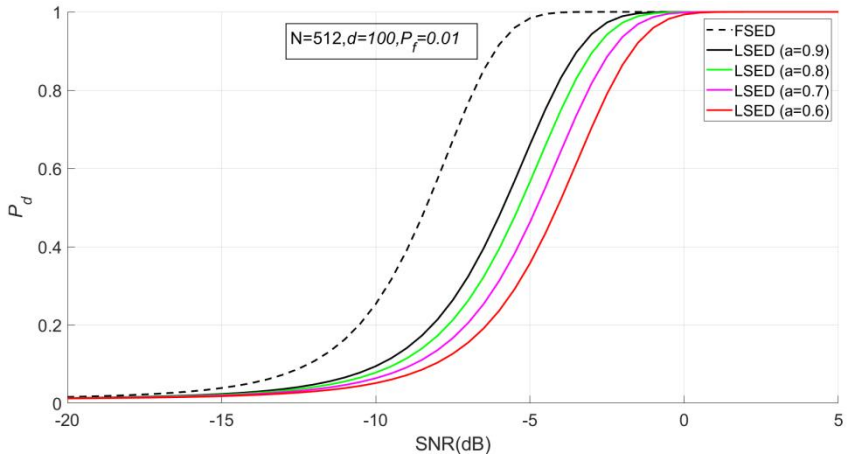


Figure 1 Graph of the Detection Probabilities for FSED and LSED When a < 1

As shown in Figure 1, when a < 1, the P_d of FSED is always higher than that of LSED.

3.1.2 The SNR of LSED is greater than or equal to FSED

In this case, $a \geq 1$. Substituting equation (13) into equation (12) yields:

$$f(A,2d,\gamma')=\frac{A-a\gamma\sqrt{d}}{\sqrt{1+2a\gamma}}\ . \tag{16}$$

The partial derivative of $f(A,2d,\gamma')$ with respect to a is:

$$\frac{\partial f(A,2d,\gamma')}{\partial a}=\frac{-\gamma'\left(\sqrt{d}+2A\right)}{(1+2a\gamma')\sqrt{1+2a\gamma'}}<0. \tag{17}$$

Therefore, $f(A,2d,\gamma')$ is a decreasing function of a. The relationship between $f(A,N,\gamma)$ and $f(A,2d,\gamma')$ is shown in the figure below:

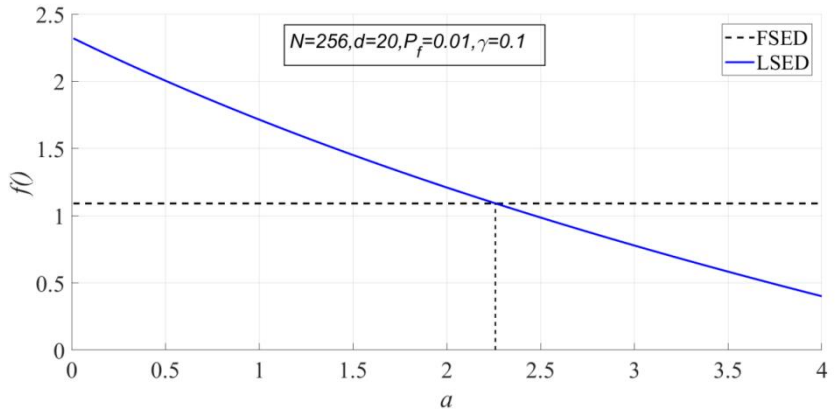


Figure 2 Graph of the Independent Variable Functions for FSED and LSED When $a \geq 1$

In Figure 2, when γ, N, and P_f are fixed, the value of $f(A,N,\gamma)$ remains constant and does not vary with a, resulting in a horizontal line. Meanwhile, $f(A,2d,\gamma')$ decreases as a increases. The two functions intersect at a specific point. To the left of this intersection, the detection performance of LSED is worse than that of FSED, while to the right, LSED outperforms FSED. The intersection point can be determined by solving equation (18):

$$\frac{A-a\gamma\sqrt{d}}{\sqrt{1+2a\gamma}}=\frac{A-\gamma\sqrt{N/2}}{\sqrt{1+2\gamma}}. \tag{18}$$

Let $T=\frac{A-\gamma\sqrt{N/2}}{\sqrt{1+2\gamma}}$, then equation (18) can be simplified as follows:

$$\gamma^2 d a^2-(2A\gamma\sqrt{d}+2\gamma T^2)a+A^2-T^2=0 \tag{19}$$

The discriminant of the root of the unary quadratic equation with respect to a is:

$$\Delta=4\left[T^2+2A\sqrt{d}+d\right]\gamma^2T^2\geq 0 \tag{20}$$

If $\Delta=0$, equation (18) has only one solution, which is T = 0, and $\gamma=\sqrt{2/NA}$.

At this point, $a=\sqrt{\frac{N}{2d}}$. That is, when $a<\sqrt{\frac{N}{2d}}$, LSED performs worse than FSED; when $a>\sqrt{\frac{N}{2d}}$, LSED outperforms FSED.

If $\Delta>0$, two solutions of equation (19) can be obtained according to the quadratic formula, respectively:

$$a_1=\frac{A\sqrt{d}+T^2+|T|\sqrt{T^2+2A\sqrt{d}+d}}{\gamma d},\ a_2=\frac{A\sqrt{d}+T^2-|T|\sqrt{T^2+2A\sqrt{d}+d}}{\gamma d} \tag{21}$$

Obviously, $a_1>a_2$. From Figure 2, it can be seen that there is only one intersection point between the two curves, so there cannot be two solutions. Therefore, only one solution is valid. The following is an analysis of the cases where T > 0 and T < 0, respectively.

3.1.2.1 When T > 0, that is, $\gamma<\sqrt{2/N}A$

In this case, both sides of equation (18) must be greater than 0, i.e., $A-a\gamma\sqrt{d}>0$ and $a<\frac{A}{\gamma\sqrt{d}}$. Substituting a_1 into equation (18), we obtain:

$$a_1-\frac{A}{\gamma\sqrt{d}}=\frac{T^2+|T|\sqrt{T^2+2A\sqrt{d}+d}}{\gamma d}>0 \tag{22}$$

That is, $a_1>\frac{A}{\gamma\sqrt{d}}$, and it is evident that a_1 does not meet the requirements. Substituting a_2 into equation (18) to obtain:

$$a_2-\frac{A}{\gamma\sqrt{d}}=\frac{T^2-|T|\sqrt{T^2+2A\sqrt{d}+d}}{\gamma d}<0 \tag{23}$$

That is, $a_2 < \frac{A}{\gamma\sqrt{d}}$, and it is evident that a_2 meets the requirements. Therefore, when $\gamma < \sqrt{2/NA}$, if $a < a_2$, then $f(A,N,\gamma) < f(A,2d,\gamma')$, and selecting FSED is preferable. If $a > a_2$, then $f(A,N,\gamma) > f(A,2d,\gamma')$, and selecting LSED is preferable.

Let N = 256, d = 20, P_f = 0.01, and γ = 0.1. In this case, $\sqrt{2/NA}$ = 0.2056, and $\gamma < \sqrt{2/NA}$. The intersection occurs at a_2 = 2.2626.

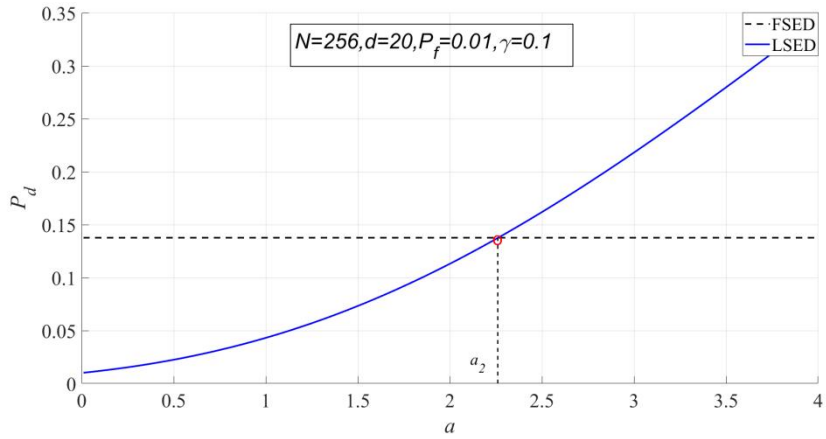


Figure 3 Graph of the Detection Probabilities for FSED and LSED When $\gamma < \sqrt{2/NA}$

It can be seen from Figure 3 that when $a > a_2$, the detection probability of LSED is higher than that of FSED. When $a < a_2$, the detection probability of LSED is lower than that of FSED.

3.1.2.2 When T < 0, that is, $\gamma > \sqrt{2/NA}$

At this point, $A - a\gamma\sqrt{d} < 0$ and $a > \frac{A}{\gamma\sqrt{d}}$. As previously deduced, $a_1 - \frac{A}{\gamma\sqrt{d}} > 0$ and $a_2 - \frac{A}{\gamma\sqrt{d}} < 0$. It is evident that a_1 meets the requirements, while a_2 does not. Therefore, when $\gamma > \sqrt{2/NA}$, if $a < a_1$, $f(A,N,\gamma) < f(A,2d,\gamma')$, selecting FSED is preferable; if $a > a_1$, $f(A,N,\gamma) > f(A,2d,\gamma')$, selecting LSED is preferable.

Let N = 256, d = 20, P_f = 0.01, and γ = 0.3. In this case, $\sqrt{2/NA}$ = 0.2056, and $\gamma > \sqrt{2/NA}$. The intersection occurs at a_1 = 2.7593.

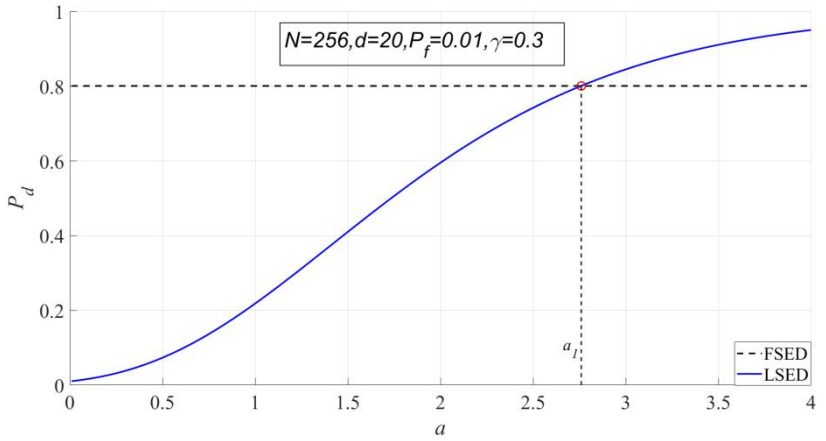


Figure 4 Graph of the Detection Probabilities for FSED and LSED When $\gamma > \sqrt{2/NA}$

It can be observed from Figure 4 that when $a > a_1$, the detection probability of LSED is higher than that of FSED. Conversely, when $a < a_1$, the detection probability of LSED is lower than that of FSED.

Based on the above analysis, the selection strategy for LSED and FSED when $a \geq 1$ is illustrated in the figure 5 below:

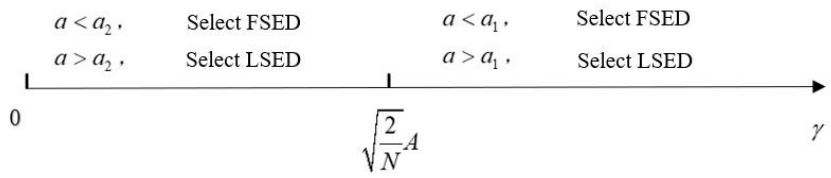


Figure 5 Selection Strategy for LSED and FSED When $a \geq 1$

3.2 The Frequency Band Contains All of the Signal Energy

Let the proportion of the energy contained in the selected frequency band to the total energy of the signal be defined as the frequency band energy ratio (ER), i.e.,

$$ER = \frac{\sum_{k=i}^{i+d-1}\left(X_R^2(k)+X_I^2(k)\right)}{\sum_{k=0}^{N/2-1}\left(X_R^2(k)+X_I^2(k)\right)} \tag{24}$$

The ratio of the signal-to-noise ratio \(a \) for the two methods is then given by:

$$a = \frac{\gamma'}{\gamma} = \frac{N}{2d}\frac{\sum_{k=i}^{i+d-1}\left(X_R^2(k)+X_I^2(k)\right)}{\sum_{k=0}^{N/2-1}\left(X_R^2(k)+X_I^2(k)\right)} \tag{25}$$

If ER = 1, it indicates that the entire spectral energy is concentrated in the frequency band of interest. In this case, $a = \frac{N}{2d}$, that is:

$$\gamma' = \frac{N}{2d}\gamma. \tag{26}$$

Then, equation (12) becomes:

$$f(A,2d,\gamma') = \frac{A-\frac{N}{2\sqrt{d}}\gamma}{\sqrt{1+\frac{N}{d}\gamma}} \tag{27}$$

Since $d < \frac{N}{2}$, it follows that $\frac{A-\sqrt{\frac{N}{2}}\gamma}{\sqrt{1+2\gamma}} > \frac{A-\frac{N}{2\sqrt{d}}\gamma}{\sqrt{1+\frac{N}{d}\gamma}}$. Therefore, $f(A,N,\gamma) > f(A,2d,\gamma')$. Under these conditions, the detection probability of LSED consistently exceeds that of FSED, making LSED the preferred option. Let P_f = 0.01, N = 512, and consider d = 10, 80, and 200 for the respective cases.

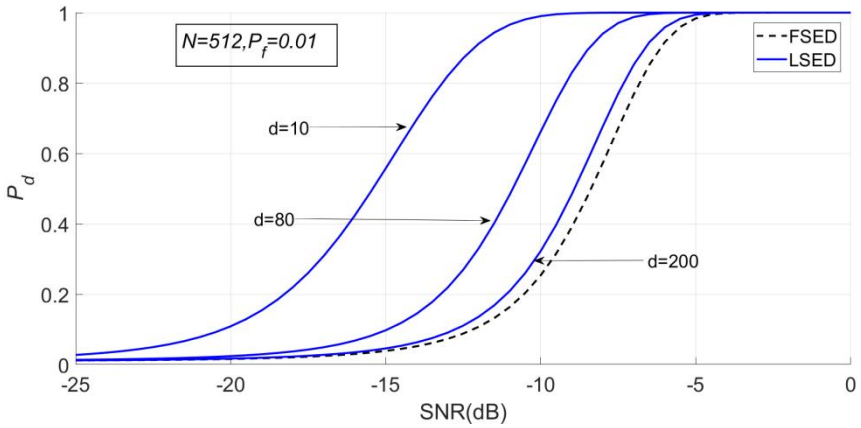


Figure 6 Graph of the Detection Probabilities for FSED and LSED When $\gamma' = \frac{N}{2d}\gamma$.

It can be observed from Figure 6 that as the parameter d increases, the performance of LSED becomes increasingly similar to that of FSED. This trend indicates that the difference in detection probability between the two methods narrows as d grows larger. However, despite this convergence, the detection probability of LSED remains consistently higher than that of FSED across all values of d. This superiority in performance demonstrates that LSED is better suited for scenarios where the frequency band encompasses the entire signal energy. Therefore, when the selected frequency band contains the complete energy of the signal, LSED emerges as the preferred option due to its superior detection capability.

4 SIMULATION EXPERIMENT

This section aims to validate the theoretical analysis results from Section 3 through Monte Carlo simulation experiments. These experiments are designed to test the performance of Full Spectrum Energy Detection (FSED) and Local Spectrum Energy Detection (LSED) under different conditions. To ensure comprehensive validation, two types of signals are used: sinusoidal signals to simulate narrowband signals and Ricker wavelets to simulate broadband signals. The experiments are conducted under controlled parameters, with the false alarm probability P_f set to 0.05, the detection length N fixed at 2048, and at least 10,000 independent Monte Carlo simulation tests performed to ensure statistical reliability.

4.1 The Frequency Band Contains All of the Signal Energy

In this experiment, sine signals are used to simulate narrowband signals, with Gaussian white noise added to generate the experimental data. The sine signal is configured with an amplitude of 0.5, a frequency of 30 Hz, and a phase of 0. The sampling frequency is set to 1024 Hz, and the sampling duration is 2 seconds, resulting in a frequency resolution of 0.5 Hz. The detection length N is set to 2048, ensuring sufficient data points for accurate analysis. To simplify the analysis, the spectrum used in this experiment is a unilateral spectrum, meaning only the positive frequency components are considered. This setup ensures that the selected frequency band contains the entire energy of the narrowband signal, allowing for a clear comparison between FSED and LSED.
The detection probabilities of the two detection methods after simulations with different values of d are as follows:

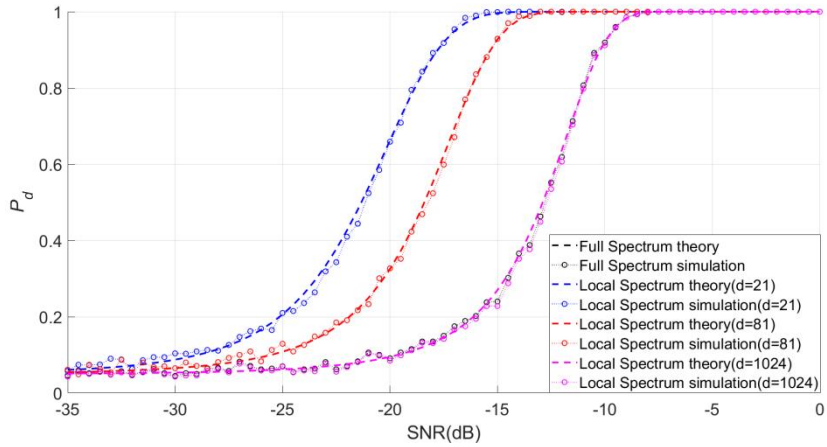


Figure 7 The Actual and Theoretical Detection Probabilities of Narrowband Signals Obtained from FSED and LSED Simulations under Different d

It can be observed from Figure 7 that the detection performance of LSED is always better than that of FSED. As d increases, LSED approaches FSED. When $\frac{N}{2d}=1$, i.e., $d=\frac{N}{2}$, the detection performance of the two methods becomes identical. As previously mentioned, since $d<\frac{N}{2}$, when the frequency band contains all the signal energy, LSED is always superior to FSED. This aligns with the theoretical analysis presented in Section 3.

4.2 The Frequency Band Contains a Portion of the Signal Energy

LSED is commonly used for microseismic monitoring, and the Ricker wavelet is frequently employed to simulate earthquake or microseismic signals in underground vibration simulation experiments. Therefore, the Ricker wavelet is chosen to simulate broadband signals in this study. The Ricker wavelet is particularly suitable for this purpose because it closely resembles the characteristics of real seismic signals, making it an ideal choice for testing the performance of energy detection methods in broadband scenarios. The time-domain expression of the Ricker wavelet is provided to describe its mathematical form, which helps in understanding its behavior in the frequency domain. This allows the selected frequency band to include only a portion of the signal energy, enabling the analysis of LSED's performance when the signal energy is not fully captured within the detection band.

$$s(t)=\left(1-2\pi^2 f_M^2 t^2\right)e^{-\pi^2 f_M^2 t^2} \quad , \tag{28}$$

Where t is the time and f_M is the peak frequency. The Fourier transform expression of the Ricker wavelet is given by:

$$S(f)=\frac{2f^2}{\sqrt{\pi}f_M^3}e^{-\frac{f^2}{f_M^3}} \quad . \tag{29}$$

The Ricker wavelet used in this experiment has a central frequency of 50 Hz, a duration of 2 seconds, a sampling frequency of 1024 Hz, a frequency resolution of 0.5 Hz, and a detection length N of 2048. In this case, the energy occupied by the local frequency band is easy to control. Let the frequency domain signal-to-noise ratio of FSED be γ , and the frequency domain signal-to-noise ratio of LSED be γ'. Let $\gamma'=a\cdot\gamma$, and analyze the detection rate.

Let N = 2048, P_f = 0.05, and $A=Q^{-1}(P_f)$. Then, $\sqrt{2/N}A=0.05$. When $\gamma=0.05$, the SNR is -13. The SNR refers to the time-domain signal-to-noise ratio of the global signal, which can be converted into noise variance to calculate the frequency-domain signal-to-noise ratio. SNR is proportional to γ .

The selected frequency band ranges are 65-90 Hz, 30-80 Hz, and 130-300 Hz. The corresponding values of d are 26, 51, and 171, and the values of a are 9.10, 4.52, and 1.51, respectively. The detection rates obtained through simulation are as follows:

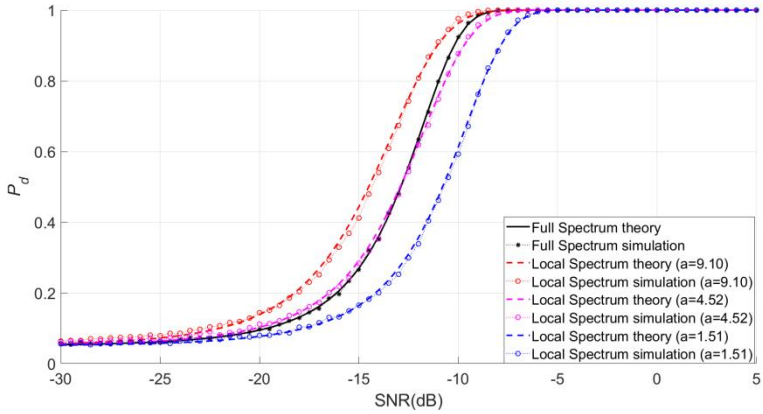


Figure 8 The Actual and Theoretical Detection Probabilities of Ricker Wavelet Obtained from FSED and LSED Simulations under Different a

Taking -13 dB as the dividing line, we use -15 dB and -11 dB as examples. When the SNR is -15 dB, γ = 0.03, the theoretical P_d of FSED is 0.2697, and the actual P_d is 0.2624. When the SNR is -11 dB, γ = 0.08, the theoretical P_d of FSED is 0.7976, and the actual P_d is 0.7942. The detection rates of a_1, a_2, and LSED in both cases are as follows:

Table 1 a_1 and P_d obtained in different frequency bands when $\gamma > \sqrt{2/NA}$

d	a	a1	Pd(theory)	Pd(simulation)
26	9.10	7.08	0.9042	0.9105
51	4.52	4.86	0.7579	0.7564
171	1.51	2.53	0.4716	0.4667

Table 2 a_1 and P_d obtained in different frequency bands when $\gamma < \sqrt{2/NA}$

d	a	a2	Pd(theory)	Pd(simulation)
26	9.10	5.72	0.4439	0.4228
51	4.52	4.21	0.2907	0.2783
171	1.51	2.38	0.1645	0.1610

From Table 1 and Table 2, when $\gamma > \sqrt{2/NA}$, in the first band where $a > a_1$, the detection rate of LSED is higher than that of FSED; in the second band where $a < a_1$, the detection rate of LSED is lower than that of FSED; in the third band where $a < a_1$, the detection rate of LSED remains lower than that of FSED.

When $\gamma < \sqrt{2/NA}$, in the first band where $a > a_2$, the detection rate of LSED is higher than that of FSED; in the second band where $a > a_2$, the detection rate of LSED is also higher than that of FSED; in the third band where $a < a_2$, the detection rate of LSED is lower than that of FSED. This is consistent with the theoretical analysis presented in Section 3.

4.3 Practical Application

To facilitate method selection in practical applications, we can establish clear guidelines by analyzing the frequency band energy ratio (ER). Specifically, when the energy ratio exceeds a certain threshold, referred to as the Excellent Energy Ratio (EER), it ensures that LSED outperforms FSED in terms of detection probability. This approach simplifies decision-making by allowing practitioners to focus on the energy concentration within the selected frequency band, ensuring optimal performance in real-world scenarios.

According to equations (24) and (25), the relationship between a and ER is as follows:

$$a = \frac{N}{2d} \cdot ER \tag{30}$$

It can be seen that a is proportional to ER. Therefore, we can calculate the maximum values of a_1 and a_2 within a certain range of signal-to-noise ratio, and determine the corresponding frequency band energy ratio, which we refer to as the EER. That is:

$$EER = \frac{2d}{N} \cdot \max(\max(a_1), \max(a_2)) \tag{31}$$

When ER is greater than EER, a must be greater than a_1 and a_2, and LSED must be better.

Let the false alarm probability P_f be 0.05, the detection length N be 2048, and the spectral line length d be 100. In Figure 8, when SNR > -5 dB, the detection rate is approximately 1, so only the case where SNR is in the range of -30 to -5 dB is analyzed below. In this interval, the maximum value of a_2 is 3.19, and the maximum value of a_1 is 4.67, with corresponding energy ratios of 31.15% and 45.61%. That is, when the frequency band energy ratio exceeds 45.61%, selecting LSED is definitely better than FSED. Other cases corresponding to the selected frequency bands can also be analyzed using this method.

The frequency bands selected are 125-224Hz, 3-102Hz, and 90-189Hz, with d = 100. The corresponding frequency band energy ratios are 30.46%, 45.63%, and 67.51%, respectively. The simulation results are as follows:

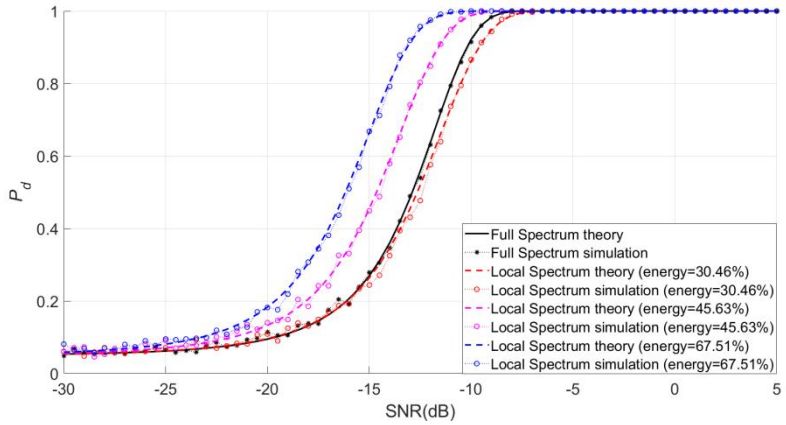


Figure 9 Detection Probabilities of LSED and FSED under Different Energy Ratios When d = 100

It can be seen from Figure 9 that when the energy ratio is 45.63% and 67.51%, i.e., ER > EER, the detection rate of LSED is higher than that of FSED. This indicates that LSED performs better when the frequency band captures a significant portion of the signal's energy. For example, at an SNR of -10 dB, LSED's detection probability is at least 7% higher than FSED, with the performance gain increasing as the energy ratio rises. When the energy ratio is 30.46%, i.e., ER < EER, the detection rate of LSED may be higher or lower than FSED, depending on the SNR conditions. Under low-SNR conditions, LSED outperforms FSED only if its SNR exceeds a specific threshold. In high-SNR scenarios, LSED becomes advantageous when the SNR ratio surpasses another threshold. This behavior aligns with the theoretical analysis, confirming that ER > EER ensures LSED's superiority.

5 DICUSSION

5.1 a_2 Can be Substituted in Practical Application

In the third part of the theoretical analysis, the calculation of a_2 is relatively complex. Through formula derivation, it is found that $a_2 < \sqrt{\frac{N}{2d}}$. Therefore, in practice, $\sqrt{\frac{N}{2d}}$ can be used as a substitute for a_2. Specifically, if $a > \sqrt{\frac{N}{2d}}$, the performance of LSED is superior to FSED; if $a < \sqrt{\frac{N}{2d}}$, the performance of FSED is better than LSED. Furthermore, when $\gamma < \sqrt{2/N}\,A$, indicating a low signal-to-noise ratio, it is more convenient to use $\sqrt{\frac{N}{2d}}$ under such conditions.

5.2 The Effect of N and d on EER

Next, the influence of detection length N and frequency band length d on the EER is analyzed. If the false alarm probability P_f is set to 0.05, the color chart of EER under different values of N and d is as follows:

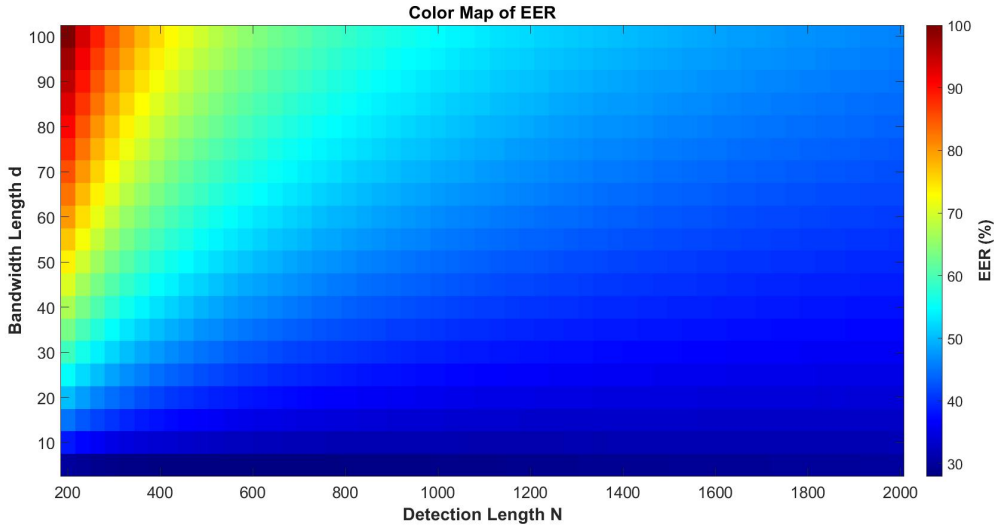


Figure 10 Color Map of EER under Different d and N

It can be observed from Figure 10 that the EER increases as d increases and decreases as N increases. Specifically, when d is large and N is small, the EER value is higher; conversely, when d is small and N is large, the EER value is lower.

5.3 a Indicates the Concentration Degree of the Signal Energy

In the simulation conducted in Section 4.3, the frequency band ranges are set to 125-224 Hz, 3-102 Hz, and 90-189 Hz, with the number of spectral lines, denoted as d, fixed at 100. The corresponding a values for these bands are 3.02, 4.79, and 6.81, respectively, while the energy proportions of the frequency bands are calculated to be 30.46%, 45.63%, and 67.51%. From these results, a clear trend emerges: for the same value of d, as the parameter a increases, the energy proportion of the frequency band also increases. This observation suggests that the frequency band is more concentrated in regions where the signal energy is higher. Consequently, a can be regarded as an indicator of the concentration of signal energy within the frequency band.

This relationship between a and the energy proportion highlights the utility of a as a metric for evaluating how effectively a chosen frequency band captures the dominant components of the signal's energy. When a is higher, it implies that the selected frequency band contains a larger share of the total signal energy, making it more suitable for applications where high detection performance is critical. Conversely, lower values of a indicate that the frequency band captures less of the signal's energy, potentially leading to suboptimal detection performance. This insight provides a practical guideline for selecting frequency bands in real-world applications, particularly when aiming to maximize the detection probability while minimizing resource usage. By leveraging a as an indicator, practitioners can make informed decisions about which frequency bands to prioritize based on their energy concentration characteristics.

6 CONCLUSIONS

This study establishes an optimal selection strategy between Local Spectrum Energy Detection (LSED) and Full Spectrum Energy Detection (FSED) by theoretically deriving the relationship between their signal-to-noise ratios (SNRs) and validating it through 10^4 Monte Carlo simulation experiments. The results indicate that, in full-band scenarios, the detection probability of LSED consistently exceeds that of FSED, and the performance gain is proportional to the bandwidth d, which represents the number of spectral lines analyzed within the frequency range (see Figure 6). Specifically, as the bandwidth d increases, the performance gap between LSED and FSED narrows, but LSED remains superior regardless of the specific value of d. In local-band scenarios, under low-SNR conditions, LSED outperforms FSED when its SNR exceeds that of FSED by a factor of a_2 . Under high-SNR conditions, however, LSED becomes advantageous only when the SNR ratio exceeds a higher threshold, denoted as a_1 , where $a_1 > a_2$.

The main contributions of this study are as follows: (1) A theoretical framework for detection method selection based on the SNR ratio a is established. This framework provides a systematic approach to determine the conditions under which LSED or FSED should be preferred, depending on the SNR ratio and the proportion of signal energy captured within the selected frequency band. (2) An efficient decision rule is proposed that relies solely on the frequency band energy ratio (ER), simplifying the selection process for practical applications. The results demonstrate that when the energy ratio exceeds the Excellent Energy Ratio (EER), the detection probability of LSED is at least 7% higher than that of FSED under an SNR of -10 dB. Furthermore, the performance gain increases proportionally with the energy

ratio, highlighting the importance of selecting frequency bands that capture a significant portion of the signal's energy (see Figure 9).

It is worth noting that the current analysis assumes additive white Gaussian noise, which is a common assumption in many signal processing applications. However, real-world scenarios often involve non-Gaussian noise environments, such as impulsive noise or cluttered backgrounds, which may affect the performance of both LSED and FSED. Future work will explore the applicability of the proposed framework under non-Gaussian noise environments and further evaluate its real-time performance. Additionally, the computational complexity and implementation feasibility of the proposed methods will be assessed to ensure their practical deployment in cognitive radio systems, seismic monitoring, and other relevant fields. These extensions aim to enhance the robustness and versatility of the energy detection techniques in diverse operational settings.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Yucek T, Arslan H. A survey of pectrum sensing algorithms for cognitive radio applications. IEEE Communications Surveys & Tutorials 2009, 11, 116-130. DOI: https://doi.org/10.1109/SURV.2009.090109.

[2] Urkowitz H. Energy detection of unknown deterministic signals. Proceedings of the IEEE, 1967, 55, 523-531. DOI: https://doi.org/10.1109/PROC.1967.5573.

[3] Dikmese S, Ilyas Z, Sofotasios P C, et al. Sparse frequency domain spectrum sensing and sharing based on cyclic preffx autocorrelation. IEEE Journal on Selected Areas in Communications 2017, 35, 159-172. DOI: https://doi.org/10.1109/JSAC.2016.2633058.

[4] Hou C, Liu G, Tian Q, et al. Multisignal modulation classiffcation using sliding window detection and complex convolutional network in frequency domain. IEEE Internet of Things Journal, 2022, 9, 19438-19449. DOI: https://doi.org/10.1109/JIOT.2022.3167107.

[5] Li H, Hu Y, Wang S. A novel blind signal detector based on the entropy of the power spectrum subband energy ratio. Entropy, 2021, 23, 448. DOI: https://doi.org/10.3390/e23040448.

[6] Li H, Hu Y, Wang S. Signal detection based on power-spectrum sub-band energy ratio. Electronics, 2021, 10, 64. DOI: https://doi.org/10.3390/electronics10010064.

[7] Ejaz W, ul Hasan N, Azam M A, et al. Improved local spectrum sensing for cognitive radio networks. EURASIP Journal on Advances in Signal Processing 2012, 2012, 242. DOI: https: //doi.org/10.1186/1687-6180-2012-242.

[8] Adardour H E, Meliani M, Hachemi M H. Improved local spectrum sensing in cluttered environment using a simple recursive estimator. Computers & Electrical Engineering 2017, 61, 208-222. DOI: https://doi.org/10.1016/j.compeleceng.2016.11.037.

[9] Quan Z, Zhang W, Shellhammer S J, et al. Optimal spectral feature detection for spectrum sensing at very low SNR. IEEE Transactions on Communications 2011, 59, 201-212. DOI: https://doi.org/10.1109/TCOMM.2010.112310.090306.

[10] Zhang B, Wu J, Su M, et al. An efffcient cooperative spectrum sensing for cognitive wireless sensor networks. IEEE Access, 2023, 11, 132544-132556. DOI: https://doi.org/10.1109/ACCESS.2023.3336654.

[11] Besson O. Impact of covariance mismatched training samples on constant false alarm rate detectors. IEEE Transactions on Signal Processing, 2021, 69, 755-765. DOI: https://doi.org/10.1109/ TSP.2021.3050567.

[12] Diskin T, Beer Y, Okun U, et al. CFARnet: deep learning for target detection with constant false alarm rate. SIGNAL PROCESSING, 2024, 223, 109543. DOI: https://doi.org/10.1016/j. sigpro.2024.109543.

[13] Li K, Zhang P, Yang Z. Semiparametric constant false alarm rate method for radar and sonar images. Electronics Letters, 2024, 60, e13146. DOI: https://doi.org/10.1049/ell2.13146.

[14] Zhong C, Wu C, Li X, et al. A novel frequency hopping prediction model based on TCN-GRU. IEICE Transactions on Fundamentals, 2024, E107-A, 1577-1581. DOI: https: //doi.org/10.1587/transfun.2023EAL2095.

[15] Zhou Z, Huang L, Christensen M G, et al. Robust spectral analysis of multi-channel sinusoidal signals in impulsive noise environments. IEEE Transactions on Signal Processing, 2022, 70, 919-935. DOI: https://doi.org/10.1109/TSP.2021.3101989.

[16] Ravve I, Koren Z. Analytical hilbert-transform attributes of ricker and gabor wavelets. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61, 1-16. DOI: https://doi.org/10.1109/TGRS.20 40723.3309248.

Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3070

ENHANCING THE YOLOv11 MODEL FOR TEACHING BEHAVIOR RECOGNITION

Yao Tian, Cheng Peng*
School of Computer Science and Technology, Xinjiang Normal University, Urumqi 830054, Xinjiang, China.
Corresponding Author: Cheng Peng, Email: pcxjnu@163.com

Abstract: Traditional methods of teaching behavior recording suffer from inefficiency, long data mining times, and large computational workloads for statistical analysis. Large models and artificial intelligence offer new technical solutions that can significantly improve teaching quality and optimize the teaching process. This study, based on the improved YOLOv11 model, presents a fine-grained teaching behavior recognition technology aimed at addressing the challenges in smart classroom environments. In response to the complexity of classroom environments and the high similarity of teaching behaviors, an improved YOLOv11 algorithm is proposed. The algorithm introduces the MSCB (Multi-Scale Context Block) and SCSA (Spatial-Channel Self-Attention) modules to enhance the robustness of the model's recognition capabilities. Experimental results show that the improved model performs better in teacher behavior detection, with higher accuracy and efficiency, offering a new approach to teaching behavior recognition.
Keywords: Teaching behaviors; Object detection; Fine-grained recognition; YOLOv11; Intelligent teaching

1 INTRODUCTION

Teaching behavior is a crucial means by which teachers convey instructional information and organize classroom activities. By demonstrating various teaching behaviors, teachers can capture students' attention, enhance the effectiveness of their verbal communication, stimulate students' motivation to learn, support their understanding of the content, and improve overall classroom teaching effectiveness[1]. According to constructivist learning theory[2], knowledge is constructed through the interaction between individuals and their environment. Identifying teaching behaviors allows researchers to explore how teachers promote active learning and knowledge construction through their actions. However, both pre-service and in-service teachers often exhibit problematic teaching behaviors, whether in teaching competitions or regular classroom settings[3]. To address this issue, scholars have employed various methods to collect and analyze different types of teaching behaviors.

In recent years, research on teaching behavior recognition has generally fallen into two categories. The first category involves behavior recognition methods based on video image data. Zhao Gang[4] and others proposed a teacher set recognition and extraction algorithm, introducing a behavior recognition network based on three-dimensional bilinear pooling capable of identifying eight types of teaching behaviors. Guo Junqi[5] designed a 3D convolutional neural network tailored to classroom scenarios for recognizing teaching behaviors, achieving high recognition efficiency on a self-constructed dataset. Ding Ning[6], building upon existing coding systems, constructed a high-quality image dataset of teacher body movements and used the VGG16 network model for image recognition.

The second category comprises methods based on teacher skeletal information. Wang Tao[7], grounded in the cognition of body movement characteristics, proposed a teaching activity analysis model based on these features. Zheng Yuhuang[8] introduced a teaching behavior evaluation method based on posture recognition, obtaining teacher pose information through the HRNet deep learning network. Pang Shiyan[9] and others used the human pose estimation algorithm OpenPose to extract coordinate information. However, methods based on skeletal data are easily affected by the way key points are extracted, and the quality of the data obtained can significantly impact the accuracy of final behavior classification.

Although there have been technological breakthroughs in automated classroom teaching evaluation within smart classroom environments, research on automated recognition of teaching behaviors still requires continuous improvement and optimization[10]. This line of research not only advances theoretical development but also promotes teachers' professional growth[11]. Based on this, this study adopts a video image data-based approach and proposes an efficient and accurate method for recognizing teaching behaviors using an improved YOLOv11 model. The aim is to provide a new technical tool for analyzing teaching behaviors, thereby supporting teachers' professional development and enhancing teaching quality.

2 CONSTRUCTION OF TEACHING BEHAVIOR DATASET

2.1 Necessity of Constructing a Teaching Behavior Dataset

Although deep learning technology has demonstrated capabilities surpassing traditional algorithms in most fields, its success still fundamentally relies on the availability of sufficient data. In the field of teaching behavior recognition, current research mainly focuses on static single-frame images, making the acquisition of a large volume of high-quality image data a critical prerequisite. While there are some publicly available action recognition datasets, such as

Kinetics[12], HMDB51[13], and UCF101[14], there remains a lack of large-scale public datasets specifically targeted at teaching behavior recognition. In response to this issue, this study constructs a teaching behavior dataset based on two scenarios: teaching competitions and real classroom environments.

2.2 Classification of Teaching Behaviors

To conduct in-depth research on teaching behaviors, many scholars have started building their own datasets. Pang Shiyan[15] and others, from the perspective of teachers' non-verbal behaviors, constructed a dataset that includes adaptive behaviors, directive behaviors, intentional behaviors, instrumental behaviors, explanatory behaviors, and evaluative behaviors. Liu Qingtang[16] and colleagues built a dataset using the S-T analysis method, but only categorized behaviors into the two coarse-grained categories of teacher behaviors and student behaviors. These classifications lack sufficient granularity and provide limited guidance for evaluation. Therefore, drawing on the S-T[17] and TBAS[18] teaching behavior analysis methods, as well as the aforementioned studies, this paper selects a set of teaching behaviors for study that are visually distinguishable, pedagogically meaningful, and frequently observed in classroom instruction. The selected categories, as listed in Table 1, include "writing," "teaching," "point the board," "show things," "gesture," and "guide students," comprising a total of 1,871 images.

Table 1 Self-Constructed Teaching Behavior Dataset

Behavior Category	Action Classification	Action Description	Quantity
Writing	Writing on the blackboard	The teacher emphasizes key teaching points by writing on the blackboard	331
Teaching	Lecturing while facing students	The teacher explains content verbally without using any teaching aids	303
Point the Board.	Pointing to the board while explaining	The teacher explains key content by pointing to written material on the blackboard	291
Show Things	Holding and explaining with props	The teacher uses physical objects to visually aid understanding of the content	345
Gesture	Gesturing with both hands while explaining	The teacher uses body language and speech to convey instructional content	287
Guide Students	Interacting with students	The teacher asks students questions or comments on their responses	314

2.3 Dataset Construction Based on Teaching Competitions and Real Classrooms

The construction of the dataset mainly involves four steps: data collection, data filtering and cleaning, data preprocessing, and manual annotation. To enhance the diversity of the dataset, classroom videos were sourced from two main categories: one part consists of real classroom recordings taken in physical classrooms, and the other includes publicly available teaching competition videos found online. In the real classroom recordings, cameras were positioned either at the back of the classroom or mounted on the ceiling in the middle of the room, facing the front of the classroom to ensure the teacher was captured in the frame. In contrast, the teaching competition videos are entirely teacher-centered, with only the teacher appearing in the frame, along with elements such as the podium, teacher's desk, blackboard, and multimedia screen.

During the data preprocessing stage, video frames were extracted at regular intervals—one frame every 30 seconds—resulting in a total of 13,000 images. These images were then filtered to remove blurry, duplicate, or otherwise unusable ones. To ensure a balanced number of images for each category of teaching behavior, additional images that met the criteria were selected from the publicly available SCB-Dataset3[19] to supplement the dataset. Ultimately, 1,871 high-quality images were selected. Sample images after filtering are shown in Figure 1. Finally, these images were manually annotated using the LabelImg software.

For the experiment, the dataset was randomly divided into a training set and a test set at a ratio of 8:2. To ensure stability during the training process, data augmentation techniques were applied to the training set. These techniques included mixed enhancement methods such as random adjustments of color (brightness, contrast, saturation), the addition of Gaussian noise, and the application of Gaussian blur. The augmented images are also shown in Figure 1.

Figure 1 Dataset Samples and Augmented Images

3 IMPROVED YOLOv11 ALGORITHM IMPLEMENTATION

3.1 Overview of the YOLOv11 Model

YOLOv11 is the latest object detection model released by the Ultralytics team in 2024[20]. The architecture of YOLOv11 has undergone multiple optimizations aimed at enhancing feature extraction performance and overall efficiency. First, the model replaces the original C2f module with the C3k2 module, significantly improving module adaptability. Then, a new C2PSA module is added after the SPPF module. This module integrates an extended C2f structure with the PSA attention mechanism, effectively enhancing the extraction of key features. In the Neck part, the Concat module fuses multi-level feature maps along the channel dimension, enabling multi-scale feature fusion, which enriches feature representation and improves detection accuracy. These enhancements allow YOLOv11 to maintain high detection efficiency while significantly reducing the number of model parameters and computational load.

3.2 Algorithm Improvement

In the study of teaching behavior, issues such as large variations in object scale and strong background interference can lead to unstable detection performance when using YOLOv11 directly. At the same time, adopting a larger model would significantly increase computational load and parameter cost. Therefore, to improve both the accuracy and efficiency of teaching behavior detection while maintaining a lightweight structure, this study proposes an improved model architecture based on YOLOv11. The structure of the improved model is shown in Figure 2. Each component of the enhanced network is explained in detail in the following two subsections.

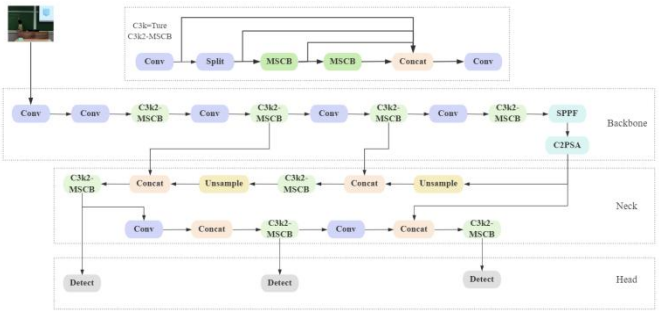


Figure 2 Improved YOLOv11 Model Diagram

3.2.1 MSCB module

In teaching behavior detection, it is necessary to process complex, dynamic, and multimodal behavioral data. To enhance the model's capability in recognizing these diverse behaviors and better adapt to the variability of teaching environments as well as the dynamic nature of behaviors, we introduce the Multi-Scale Convolution Block (MSCB), whose structure is shown in Figure 3.

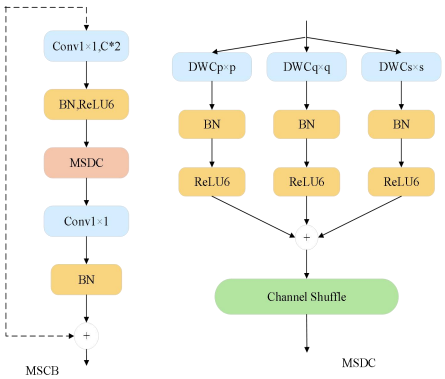


Figure 3 MSCB and MSDC Model Structures

The MSCB module integrates depthwise separable convolution with multi-scale convolution. Its core mechanism, the Multi-scale Depth-wise Convolution (MSDC), is an improved convolutional neural network architecture designed to enhance multi-scale feature extraction capability in convolution operations[21].

The core idea is to perform convolutional operations at multiple scales to capture image features at different hierarchical levels while maintaining the computational efficiency of Depth-wise Convolution. In MSCB, we follow the Inverted Residual Block (IRB) design of MobileNetV2 [22], performing depth-wise convolution at multiple scales and using channel shuffle [23] to shuffle channels between groups.Specifically:First expand the number of channels using a pointwise (1×1) convolutional layer PWC1(·).Followed by batch normalization layer BN(·).Then ReLU6 activation layer R6(·) [24]Next apply multi-scale depth-wise convolution MSDC(·) to capture multi-scale and multi-resolution context.Since depth-wise convolution ignores relationships between channels, channel shuffle operation is used to

integrate inter-channel relationships.Then:Apply another pointwise convolution PWC2(·).Followed by BN(·) to transform back to the original channels.This also encodes correlations between channels.

3.2.2 SCSA module

The SCSA module is a novel co-attention mechanism proposed by combining spatial attention and channel attention [25]. The design of SCSA consists of two main components: Shared Multi-semantic Spatial Attention (SMSA) and Progressive Channel-wise Self-Attention (PCSA). The SCSA mechanism aims to effectively integrate the advantages of both channel and spatial attention while fully utilizing multi-semantic information to enhance performance in visual tasks [26].

As shown in Figure 4, SMSA and PCSA are used in series to achieve spatial-channel co-attention based on dimension decoupling, lightweight multi-semantic guidance, and semantic discrepancy mitigation[27]. The SMSA extracts multi-level spatial information through multi-scale depth-shared 1D convolutions, providing multi-semantic spatial priors for channel attention, which helps enhance the representation of different semantic information[28].

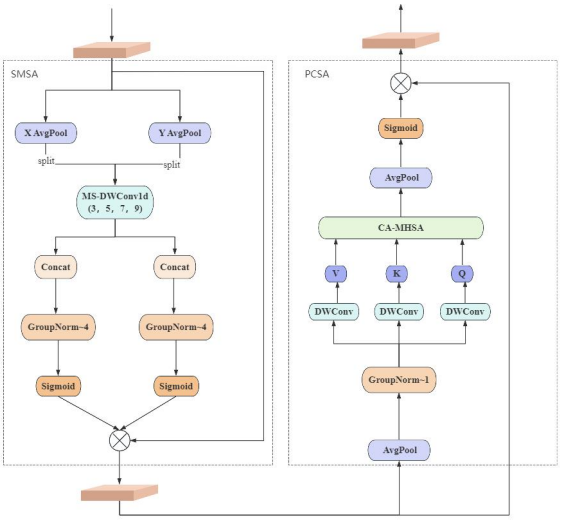


Figure 4 SMSA and PCSA Model Structures

Using a progressive compression strategy, discriminative spatial information is injected into PCSA to effectively guide channel re-calibration. This compression strategy reduces computational complexity while preserving critical spatial structural information, enabling channel attention to leverage more spatial priors during computation[29]. The PCSA module employs an input-aware self-attention mechanism that effectively computes inter-channel similarity, thereby alleviating semantic discrepancies among different sub-features within SMSA[30].

4 EXPERIMENTS

4.1 Experimental Environment and Parameters

Experiments were conducted on an Ubuntu 22.04 system using an NVIDIA RTX 3090 GPU with 24GB of memory. The model training involved 200 epochs with a batch size of 16, and the input image size was set to 640x640.

4.2 Evaluation Metrics

The following metrics are used to evaluate the performance of the model. P represents Precision, and R represents Recall, calculated as follows:

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

mAP50 refers to the mean Average Precision when the IoU threshold is 0.5. The mAP value can be calculated using the following formulas:

$$mAP = \frac{1}{C}\sum_{i=1}^{e} AP_i$$

$$AP = \int_0^1 (Pr\,(Re))d(Re)$$

mAP50-95 indicates the average of mAP values calculated at IoU thresholds ranging from 0.5 to 0.95 in increments of 0.05.

Parameter count measures the scale and complexity of the model and is calculated by summing the number of weight parameters across all layers.
GFLOPs (Giga Floating Point Operations per Second) are used to evaluate the computational complexity and runtime efficiency of the model.

4.3 Ablation Study

To validate the effectiveness of each improvement strategy, we conducted systematic ablation experiments based on YOLOv11. The experimental results are presented in Table 2.

Table 2 Ablation Study

Models	Params/M	GFlOPs	P/%	R/%	mAP50/%	mAP50-95/%
YOLOv11	2.6	6.3	84	82	86.6	50
YOLOv11＋MSCB	2.4	6.2	85	78	86.9	51
YOLOv11＋SCSA	2.5	6.3	83	81	87	50
YOLOv11＋MSCB＋SCSA	2.3	6.2	86	80	88	50

4.4 Comparative Experiments

To verify the effectiveness of our proposed improved network for teaching behavior detection, we performed comparative experiments with other classic deep learning detection networks on our self-built dataset. All comparative experiments employed identical training hyperparameter settings. The results are shown in Table 3.
As demonstrated in Table 3, our proposed improved model outperforms other detection networks in teaching behavior detection performance. Considering both average precision and processing speed for teaching behavior detection, our method exhibits superior overall performance compared to other detection networks, making it more suitable for practical deployment in teaching behavior detection applications.

Table 3 Comparative Experiment of Algorithms

Models	Params/M	GFlOPs	P/%	R/%	mAP50/%	mAP50-95/%
YOLOv8	2.7	6.8	79	80	85	50
YOLOv9	6.2	22.1	81	80	85	49
YOLOv10	2.7	8.2	84	77	83	47
YOLOv11	2.6	6.3	84	82	86.6	50
Ours	2.3	6.2	86	80	88	50

5 CONCLUSION AND OUTLOOK

This study successfully improved the YOLOv11 model for efficient and accurate teaching behavior recognition. The introduction of the MSCB and SCSA modules enhanced the model's ability to handle complex, dynamic, and multi-modal behavior data. The experimental results validate the effectiveness of the proposed model in teaching behavior detection, showing improved precision while maintaining model efficiency, making it suitable for practical classroom deployment. Future developments in this field can expect more accurate and intelligent behavior recognition systems, leveraging advancements in deep learning and computational power.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Xu Lan, Deng Yingfeng. Research on the Path of Empowering High-Quality Development of Vocational Education Through the “Three Education” Reform—Based on the Background of Industrial Digital Transformation. Vocational Education Forum, 2022(7).
[2] Ren Jiemin. Positioning of Teachers' Roles in Multimedia Teaching Environment Under Constructivist Learning Theory. Curriculum Education Research, 2018(33): 193–194.
[3] Zhang Zhe, Chen Xiaohui, Qin Pengxi, et al. Meta-analysis of Factors Influencing Teachers' Use of Intelligent Technology in Teaching. Modern Distance Education, 2019(2).
[4] Zhao Gang, Zhu Wenjuan, Hu Biling, et al. A simple teacher behavior recognition method for massive teaching videos based on teacher set. Applied Intelligence, 2021, 51(12).
[5] Guo J, Lü J, Wang R, et al. Deep learning model-driven teacher-student classroom behavior recognition. Journal of Beijing Normal University (Natural Science Edition), 2021, 57(06): 905–912.
[6] Ding N. Intelligent analysis and recognition of teacher body movements in secondary school classroom videos. Master's Thesis, Central China Normal University, 2020.

[7] Wang T. Research on classroom teaching behavior analysis methods based on human motion detection. Master's Thesis, Chang'an University, 2020.
[8] Zheng Y. A posture recognition-based method for teacher teaching behavior evaluation. Software Engineering, 2021, 24(04): 6–9.
[9] P Shiyan, Z Anran, L Shuhui, Z Zhiqi. Automatic recognition of teachers' nonverbal behavior based on dilated convolution. 2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE), 2022: 162–167.
[10] Ma X. Research and application of teacher behavior recognition for smart classrooms. Master's Thesis, Yunnan Normal University, 2023.
[11] Liu Y. Research on the evaluation of teaching effectiveness for primary and secondary school teachers in digital education environments. Master's Thesis, Northwest Normal University, 2023.
[12] Carreira J, Zisserman A. Quo vadis, action recognition? A new model and the Kinetics dataset. Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, 2017: 4724–4733.
[13] Kuehne H, Jhuang H, Garrote E, et al. HMDB: A large video database for human motion recognition. In Proc. ICCV, 2011: 2556–2563.
[14] Soomro K, Zamir A, Shah M. UCF101: A dataset of 101 human actions classes from videos in the wild. Computer Science, 2012.
[15] Pang S, Hao J, Hu H, et al. Teacher behavior recognition method based on spatiotemporal graph convolutional neural networks. Journal of Central China Normal University (Natural Science Edition), 2023, 57(05): 715–723.
[16] Liu Q, He H, Wu L, et al. Classroom teaching behavior analysis method based on artificial intelligence and its application. China Educational Technology, 2019(09): 13–21.
[17] Fu D, Zhang H, et al. Educational information processing. Beijing: Beijing Normal University Press, 2011.
[18] Mu S, Zuo P. Research on classroom teaching behavior analysis methods in an information-based teaching environment. Educational Technology Research, 2015, 36(09): 62–69.
[19] Yang F, Wang T. SCB-dataset: A dataset for detecting student classroom behavior, 2023.
[20] Jegham N, Chan Y, Marwan A, et al. Evaluating the evolution of YOLO (You Only Look Once) models: A comprehensive benchmark study of YOLO11 and its predecessors. arXiv preprint arXiv:2411.00201, 2024.
[21] Rahman M, Marculescu R. Medical image segmentation via cascaded attention decoding. IEEE/CVF Winter Conference on Applications of Computer Vision, 2023: 6222–6231.
[22] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L. Mobilenetv2: Inverted residuals and linear bottlenecks. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510–4520.
[23] Zhang X, Zhou X, Lin M, Sun J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6848–6856.
[24] Krizhevsky A, Hinton G. Convolutional deep belief networks on CIFAR-10. Unpublished manuscript, 2010, 40(7): 1–9.
[25] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Advances in Neural Information Processing Systems, 2017, 30.
[26] Ma N, Zhang X, Zheng H, Sun J. Shufflenet v2: Practical guidelines for efficient CNN architecture design. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 116–131.
[27] Han K, Wang Y, Tian Q, et al. Ghostnet: More features from cheap operations. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580–1589.
[28] Chen J, Kao S, He H, et al. Run, don’t walk: Chasing higher FLOPs for faster neural networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021–12031.
[29] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132–7141.
[30] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713–13722.

Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3071

TFE-NET: A TINY-AWARE FEATURE ENHANCEMENT NETWORK FOR COLLABORATIVE OPTIMIZATION IN SMALL OBJECT DETECTION

Qiang Zeng*, YiDan Chen
School of Computer Science and Artificial Intelligence, Beijing Technology and Business University, Beijing 102488, China.
Corresponding Author: Qiang Zeng, Email: 18178945645@163.com

Abstract: In object detection tasks, the sparse distribution, weak saliency and context-dependent nature of small objects pose three major challenges for perception systems. Although end-to-end detection architectures like the YOLO series have achieved a good balance between and speed accuracy in recent years, their utilization of shallow-layer features is low, resulting in performance bottlenecks in micro-object recognition. To address this, this paper proposes a small-object perception-enhanced detection framework, TFE-Net (Tiny-aware Feature Enhancement Network). By constructing a shallow high-resolution feature pathway and a multi-scale fine-grained semantic interaction module, it achieves a lightweight improvement of the YOLOv8s model structure. While maintaining the original model's computational complexity, this method significantly enhances the spatial perception and discrimination accuracy for extremely small objects. Experiments were conducted on the VisDrone dataset. Results show that the improved model boosts the mAP@0.5 from 0.386 to 0.421, with noticeable improvements in PR curves across all categories. This confirms the proposed strategy's ability to perceptually reconstruct and detect small objects in complex scenarios.
Keywords: Small object detection; Feature enhancement network; YOLOv8; Multi-scale fusion; Weak saliency awareness

1 INTRODUCTION

In today's rapidly advancing digital and intelligent era, small object detection (SOD), a key branch of computer vision, is growing in importance. It's widely used in intelligent transportation systems for real-time vehicle and pedestrian monitoring, in public safety for threat identification, in precision medicine for detecting cells and minor lesions, and in military surveillance for tracking distant small targets. Yet, SOD faces tough technical challenges. Small targets, occupying minimal pixel areas in images or videos, are hard to capture and recognize. Their lack of distinctiveness makes them hard to spot against complex backgrounds, causing mainstream detection models to suffer from severe perceptual degradation and fail to accurately locate and identify these targets. Moreover, the weak feature representation of small targets further increases detection difficulty.

Nowadays, the field of small object detection (SOD) faces numerous technical challenges, primarily caused by the following key factors. First, data sparsity is a significant issue. Limited samples of small objects and high annotation costs lead to class-imbalanced long-tailed distributions in training datasets, which restricts the effective learning ability of models. Second, during the process of multi-layer downsampling in convolutional neural networks, the fine-grained features unique to small objects are prone to being lost. This results in missing representations in high-level feature expressions, thereby affecting the final detection accuracy. Furthermore, the insufficiency of existing models in spatial context modeling makes it difficult to effectively capture the key discriminative information of small objects in complex backgrounds, thus weakening the ability to distinguish targets from backgrounds. These factors collectively result in "perception blind spots" in specific scales and regions, which seriously limits the performance and scope of application of SOD. The three core challenges in the SOD field are specifically as follows: Inadequate feature representation. Due to their limited size and sparse spatial distribution, the features of small objects gradually degenerate in deep networks, which impacts the extraction and recognition of detailed information; Perceptual scale imbalance. Traditional object detection frameworks focus on the detection performance of medium and large objects and lack dedicated path designs for small objects. This leads to unreasonable allocation of computing resources and attention mechanisms, causing the detection performance of small objects to be far below expectations; Missing contextual information. Accurate recognition of small objects is highly dependent on local contextual clues. However, most existing models lose a significant amount of spatial detail information during high-level feature abstraction. Due to the lack of effective context retention and fusion mechanisms, the detection difficulty is further increased, particularly in complex background scenarios. To address the above challenges, researchers have proposed a variety of innovative models and technical solutions aimed at improving the accuracy and reliability of SOD.

RetinaNet, a refined model based on Feature Pyramid Network (FPN), is widely used in small object detection. Its essence lies in constructing a top-down feature pyramid to enhance the perception of small objects by fusing multi-scale features. For instance, Tian et al. proposed a small object detection algorithm based on an improved RetinaNet model[1]. This algorithm addressed the low accuracy of traditional object detection algorithms when dealing with objects in horizontal and aerial images. Ahmad et al. introduced a detector integrated with RetinaNet to enhance low-level

semantic information and high-level spatial resolution[2], thereby effectively improving the superiority of small object detection in aerial images. Ahmad et al. adopted an anchor optimization method to improve the baseline framework's accuracy, thereby enhancing the detection of extremely small objects[3]. Nevertheless, RetinaNet has limitations, such as over-reliance on high-level features causing detail loss, and a linear feature fusion method lacking non-linear feature reshaping, which affect the detection of tiny objects.

DETR, based on the Transformer architecture, has pioneered a new paradigm for object detection. Its core is using self-attention mechanisms to model global spatial dependencies and enhance contextual information perception. For example, Dubey et al. proposed a normalized inductive bias for object detection with data fusion[4], improving DETR's accuracy in detecting small objects. Dai et al. replaced the cross-attention module with a dynamic attention module based on ROI[5], achieving faster convergence with fewer training epochs. Cao et al. introduced a new decoder layer to improve localization accuracy, especially for small objects[6]. However, DETR has limitations in handling local micro features. Specifically, the self-attention mechanism tends to focus on global features and overlook local details of small objects, leading to suboptimal performance in extremely small object detection tasks.

The YOLO series holds a significant position in object detection, with YOLOv5 and YOLOv8 being representative. YOLOv5 offers more efficient feature extraction and a flexible network structure, improving detection accuracy while maintaining real-time performance. YOLOv8 further introduces technologies like Decoupled Head and Task-Aligned Assigner to enhance training efficiency and detection accuracy. For example, Wang et al. introduced a query-based model with a new pipeline to address remote detection challenges in driving scenarios[7]. Sun et al. combined optical flow with background suppression images as auxiliary inputs[8], significantly improving the detection of infrared moving small objects. Shen et al. incorporated deformable convolution modules and a dynamic non-monotonic focusing mechanism into the backbone network[9], addressing object detection challenges in complex remote sensing image tasks. Despite the YOLO series' excellent real-time performance and detection accuracy, its lack of a dedicated perception path for shallow high-frequency information limits the model's spatial resolution for tiny objects, making it difficult to adapt to high-density small object distributions in complex scenes.

Currently, traditional perception models, constrained by the single paradigm of global representation of deep features, fail to leverage the descriptive advantages of shallow features for local fine-grained information, leading to significant capability gaps in small target perception. Most models also lack effective feature interaction mechanisms, making it hard to achieve fine perception and precise localization of small targets under controllable computational complexity, which further increases the difficulty of SOD. Against this backdrop, SOD research is showing significant paradigm evolution. On the one hand, the hierarchical semantic fusion paradigm is gaining attention. By enhancing shallow feature representation and introducing cross-scale feature interaction, it significantly improves the model's perception granularity of small targets. On the other hand, the cross-domain feature reconstruction direction is emerging. By leveraging the complementarity of multimodal feature spaces to fill representation gaps, it provides a new technical path to maximize perceptual integrity.

The SOD problem is essentially a representation deficiency caused by the imbalanced distribution of small-sample categories in high-dimensional feature spaces. Its core lies in constructing effective high-resolution descriptors within a limited feature perception domain. Starting from the dual perspectives of computational graph spatial modeling and information flow optimization, this paper proposes an architectural adjustment strategy with weakly-supervised feature enhancement capabilities to maximize information integrity in perceptual scenarios. In summary, this paper makes the following technical contributions to the field of SOD: (1) Proposes TFE-Net (Tiny-aware Feature Enhancement Network), an optimized network model based on the YOLOv8s framework. TFE-Net incorporates a dedicated small target perception branch, integrating shallow feature enhancement, multi-scale fine-grained perception fusion, and local non-linear feature reshaping strategies. This innovative architecture effectively expands the feature representation space while maintaining the model's original computational efficiency, significantly improving the model's ability to identify extremely small targets in complex scenes. By enhancing shallow features to capture local fine-grained information of small targets, fusing multi-scale features to enhance the model's perception of different-sized small targets, and reshaping local non-linear features to further improve feature representation, the model can more accurately identify extremely small targets that occupy minimal pixel areas in images. (2) In architectural design, TFE-Net balances practicality and computational efficiency. Through feature path reparameterization and perceptual redundancy compression, TFE-Net not only enhances SOD performance but also ensures efficient inference in practical deployment. Feature path reparameterization optimizes the feature extraction process for more efficient computational resource utilization, while perceptual redundancy compression reduces unnecessary computational overhead and speeds up inference. This balance enables TFE-Net to quickly and accurately process complex visual scenes in practical applications, meeting the demands of real-time applications.

2 METHOD

2.1 Overall Architecture Design

2.1.1 Core design philosophy

The core of TFE-Net's design is to boost feature perception and discrimination for tiny object detection. On the one hand, We introduce a Shallow Perception Enhancement Pathway (SPEP) to strengthen shallow feature extraction and highlight local details of small objects. On the other hand, A Cross-Scale Fine-Grained Aggregation Unit (CFAU) is

used to promote interaction and fusion of features at different scales, preserving crucial small object information. Also, Feature Flow Diversity Pathways are built to enrich feature propagation routes for better mining and utilization of small object features. Overall, TFE-Net inherits YOLOv8s' strong feature extraction ability and effectively alleviates feature degradation and insufficient representation of small objects. It offers a better feature foundation for tiny object detection and enhances detection performance in complex scenarios.

2.2 Core Module Design

TFE-Net has three key innovations: (1)Local Feature Perception Enhancement. TFE-Net strengthens expression of local fine-grained information by adding a small-object detection head and a dedicated shallow high-resolution feature pathway. The detection head is designed for small objects to capture their features precisely. The shallow high-resolution feature pathway preserves local details. This design improves the model's handling of local small-object features and boosts recognition accuracy. (2)Nonlinear Cross-Scale Feature Interaction. TFE-Net introduces a CFAU module for dynamic nonlinear interaction between features of different scales. Through nonlinear transformation and fusion, the CFAU module captures complex relationships between multi-scale features, enhancing context perception of small objects. (3) Feature Reconstruction and Alignment Optimization. TFE-Net uses feature space reparameterization via a selective feature enhancement module (SFEM) to compress redundant information. SFEM selectively enhances important features and suppresses redundant ones, optimizing representation for efficient object detection. This strategy also ensures feature consistency across levels and scales, improving overall model performance.

2.2.1 Shallow Perception Enhancement Pathway(SPSE)

The Shallow Perception Enhancement Pathway (SPEP) is a key component of TFE-Net for boosting small-object detection performance. Building on YOLOv8s' original three-layer detection heads, SPEP adds a new small-object detection head. This enables the model to better utilize the rich local detail information in shallow feature maps, which is crucial for small-object detection. Moreover, the Upsample-Concat-C2f module in SPEP plays a significant role in enhancing feature-map resolution. It strengthens spatial details through up-sampling and reconstructs high-dimensional features effectively via channel concatenation and convolution fusion. Which is,

$$F_{enhanced} = C2f(Concat(Upsample(F_{deep}), F_{shallow})) \quad (1)$$

Where, F_{deep} represents the deep feature map, $F_{shallow}$ represents the shallow feature map, and $C2f$ denotes the feature reconstruction module.

Specifically, in the neck network, deep-layer feature maps are first upsampled to achieve high-resolution feature maps. Then, the Concat module merges these upsampled features with early-stage features from the backbone network. Finally, the $C2f$ module performs multiple convolutions and skip connections to extract and fuse multi-level features, generating richer representations. This fusion strategy effectively combines multi-level feature information, enhancing feature expression. Consequently, the model's perception and detection accuracy for small objects are improved, as shown in Figure 1.

Figure 1 Shallow Perception Enhancement Pathway Diagram

Overall, the SPEP pathway effectively boosts the expressive density of the feature space. By enhancing local perception, it eases the issue of microscopic feature degradation and enables strong responsiveness reconstruction of small-object features.

2.2.2 Cross-Scale Fine-Grained Aggregation Unit(CFAU)

The Cross-Scale Fine-Grained Aggregation Unit (CFAU) is a key module in TFE-Net for boosting feature fusion and small-object detection. It uses multi-path nonlinear interaction strategies to achieve dynamic weighting and context enhancement of features at different scales. Specifically, CFAU has three core mechanisms:

(1) Multi-Scale Adaptive Fusion: CFAU dynamically adjusts the weights of features at different scales, allowing effective fusion based on their importance. This adaptive weighting strategy improves the flexibility and adaptability of feature fusion.

(2) Spatial-Aware Enhancement: CFAU emphasizes enhancing spatial information in local regions of feature maps. It uses spatial attention mechanisms to highlight areas of small objects and reduce background noise interference.

(3) Channel Recalibration: CFAU dynamically adjusts the weights of feature channels. It enhances important channels and suppresses unimportant ones, further optimizing feature representation and improving discriminability.

In summary, CFAU combines these three core mechanisms through multi-path nonlinear interactions. Which is,

$$F_{\mathrm{CFAU}} = \sum_{i=1}^{N} \alpha_i \cdot \mathcal{A}(F_i) \quad (2)$$

Where, $\mathcal{A}$ denotes the adaptive attention module, α_i signifies dynamic weights, and F_i represents feature maps of different scales.

CFAU enhances the model's local context perception and expands the spatial coverage and diversity of feature representations through non-linear cross-scale feature fusion.

2.2.3 Selective Feature Enhancement Module(SFEM)

The Selective Feature Enhancement Module (SFEM) in TFE-Net optimizes feature representation and boosts model performance. It uses a Convolution-BatchNorm-Activation (CBS) pattern with a SiLU activation function to enhance non-linear feature expression. Which is

$$F_{improve} = CBS(Concat(C2f(F_{deep}), F_{enhanced})) \tag{3}$$

Here, F_{deep} represents the deep feature map, $F_{enhanced}$ denotes the shallow feature map obtained from the Shallow Perception Enhancement Pathway (SPEP), and CBS represents the feature re-extraction module, which is:

$$CBS = SiLu(BatchNorm(Conv(F_i)) \tag{4}$$

Here, F_i represents the input feature maps of different scales to the module.

Specifically, the feature map first undergoes a convolution operation to extract higher-level features. Then, BatchNorm and the SiLU activation function are applied to enhance feature expression through the CBS module. Next, the processed feature map is concatenated with the one from the first part to integrate features from different sources. Finally, the $C2f$ module is used again to further fuse and enhance the concatenated features, extracting more robust and discriminative representations. As shown in Figure 2, SFEM's feature-channel selective enhancement mechanism dynamically adjusts channel weights, suppressing irrelevant features and boosting key-feature expression.

Figure 2 Feature Selective Enhancement Module Diagram

SFEM enables dynamic selection and enhancement of local features in the feature space. Through nonlinear mapping and channel recalibration, it improves the discriminative power and information density of feature distributions. This design boosts feature discrimination and enhances the model's detection accuracy and robustness for small objects.

2.3 Improved Architecture Overall Process

The improved architecture presented in this paper enhances YOLOv8s in multiple ways to boost tiny-object detection. The Backbone retains YOLOv8s' original structure for feature extraction. The Neck adds a Shallow Perception Enhancement Pathway (SPEP), which, with the Upsample-Concat-C2f module, introduces shallow high-resolution features, crucial for small-object detection. It also incorporates a Cross-Scale Fine-Grained Aggregation Unit (CFAU) for adaptive weighting and dynamic nonlinear interaction of multi-scale features, thereby strengthening contextual perception of small objects. The Head retains the original Anchor-Free decoupled head and adds a new small-object detection head to enhance the model's ability to detect tiny objects. Moreover, the architecture uses a Selective Feature Enhancement Module (SFEM) to boost feature distinctiveness and refine predictions. In summary, this architecture, as shown in Figure 3, optimizes every component to fully leverage multi-level features, significantly improving tiny-object detection performance.

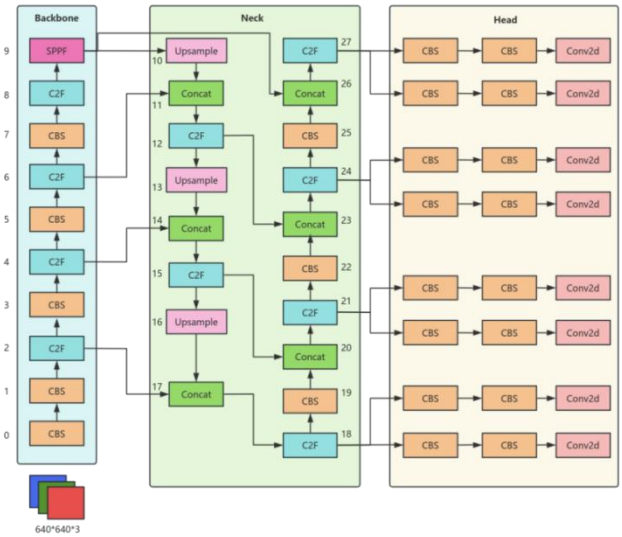


Figure 3 Improved Structure Diagram

3 EXPERIMENTAL DESIGN AND RESULT ANALYSIS

3.1 Experimental Environment and Settings

The experiments were conducted on a server with an NVIDIA RTX 4090 GPU, which excels at parallel computing and handles large-scale deep learning tasks efficiently. The software framework uses PyTorch for YOLOv8s and builds on the Ultralytics team's official implementation. Training hyperparameters are set as follows: SGD optimizer, initial learning rate (lr0) of 0.01, final learning rate (lrf) of 0.01, batch size of 32, image size of 640×640, and 300 training epochs. A linear annealing strategy was adopted to ensure stable model convergence and prevent loss of small-object features. The loss function combines YOLOv8's CIoU Loss, DFL, and Cross-Entropy Loss to enhance small-object localization accuracy.

3.2 Dataset Selection and Feature Analysis

3.2.1 Overview of the VisDrone dataset

The VisDrone dataset, collected by Tianjin University's Machine Learning and Data Mining Laboratory (AISKYEYE Team), is a benchmark for small object detection in low-altitude UAV scenarios[10]. It features complex real-world scenes with challenges like occlusion, multi-scale changes, complex backgrounds, and varying illumination, as shown in Figure 4. These characteristics make it a key benchmark for assessing and improving small object detection models. Additionally, its diversity in weather, lighting, and urban-rural backgrounds enables models trained on it to adapt better to various practical scenarios, thus enhancing their generalization and robustness.

Figure 4 Dataset Scenario Diagram

3.2.2 Categories and sample distribution

The VisDrone dataset comprises 10 common urban-scene object categories, such as pedestrian, car, bus, and bicycle. These categories, typical in urban environments, are applicable across various scenarios. The dataset exhibits a long-tail distribution, where some classes have many samples and others few. In VisDrone, common classes like vehicles and pedestrians have abundant samples, offering rich training data. This helps models learn their features well during training. But less common classes, such as special devices or specific animals, have few samples, causing the long-tail distribution. Due to this distribution, VisDrone is great for testing model robustness and generalization. It can effectively evaluate how well models handle class-imbalance issues.

3.2.3 Challenges

In the VisDrone dataset, objects are typically small, occupying less than 2% of the image on average, which is challenging for detection. The scenes often include multiple object categories that are densely packed, partially occluded, and frequently overlapping, further complicating recognition and differentiation. Additionally, the uneven background in the dataset significantly interferes with local feature perception, demanding greater robustness and adaptability from models during detection. These characteristics make the VisDrone dataset an ideal platform for evaluating object detection models.

3.3 Result Evaluation Metrics and Assessment Protocols

The key metrics for this experiment include mAP@0.5, mAP@0.5:0.95, Precision, and Recall. The mAP@0.5 measures average precision at an IoU threshold of 0.5, a critical indicator of object detection performance. mAP@0.5:0.95 evaluates overall model performance across multiple IoU thresholds. Precision assesses prediction accuracy, i.e., the proportion of correct predictions among all predicted objects. Recall measures the model's ability to recall objects, i.e., the proportion of correct predictions among all actual objects. These complementary metrics offer a comprehensive view of model performance.

Regarding evaluation protocols, all metrics are computed on the validation set to ensure reproducibility. Experimental settings are kept consistent, with only the addition of the small-object detection head being compared. This approach accurately reflects the performance contribution of TFE-Net.

3.4 Quantitative Result Analysis

The following is a quantitative analysis of the experimental results based on the aforementioned experimental setup.

Table 1 Basic Model Performance Table

Class	mAP@0.5	Class	MAP@0.5
pedestrian	0.408	truck	0.357
people	0.322	tricycle	0.267
bicycle	0.130	awning-tricycle	0.153
car	0.786	bus	0.553
van	0.446	motor	0.436
all-classes	0.386		

Table 1 shows the original model's varying performance across categories. Small objects like bicycles and tricycles have lower average precision, indicating challenges in detecting small targets due to their subtle features and smaller pixel presence in images. Additionally, the overall recall is low, suggesting frequent missed detections, especially for small objects. This implies the model's insufficient shallow-feature expression, as it fails to fully capture and utilize local details of small targets, affecting detection comprehensiveness.

Table 2 TFE-Net Model Performance Table

Class	mAP@0.5	Class	MAP@0.5
pedestrian	0.451↑ (+4.3%)	truck	0.394↑ (+3.7%)
people	0.355↑ (+3.3%)	tricycle	0.316↑ (+4.9%)
bicycle	0.158↑ (+2.8%)	awning-tricycle	0.168↑ (+1.5%)
car	0.807↑ (+5.1%)	bus	0.621↑ (+6.8%)
van	0.469↑ (+2.3%)	motor	0.471↑ (+3.5%)
all-classes	0.421 ↑ (+3.5%)		

Table 2 shows TFE-Net detection performance across categories, with significant improvements, especially for "hard-to-detect" objects like pedestrians and bicycles. Compared to the original model, the mAP@0.5 for pedestrian detection increased from 0.408 to 0.451, and for bicycle detection, it rose from 0.130 to 0.158. Overall, the mAP@0.5 improved by 3.5%, indicating enhanced accuracy in detecting small and confusing objects. The results demonstrate that TFE-Net enhances shallow perception paths, enabling better re-expression and re-localization of tiny targets. This constructs more layered and discriminative feature maps and highlights the model's robustness and reliability in small object detection.

3.5 Visualization and Phenomenon Attribution

3.5.1 Analysis of small target enhancement effects

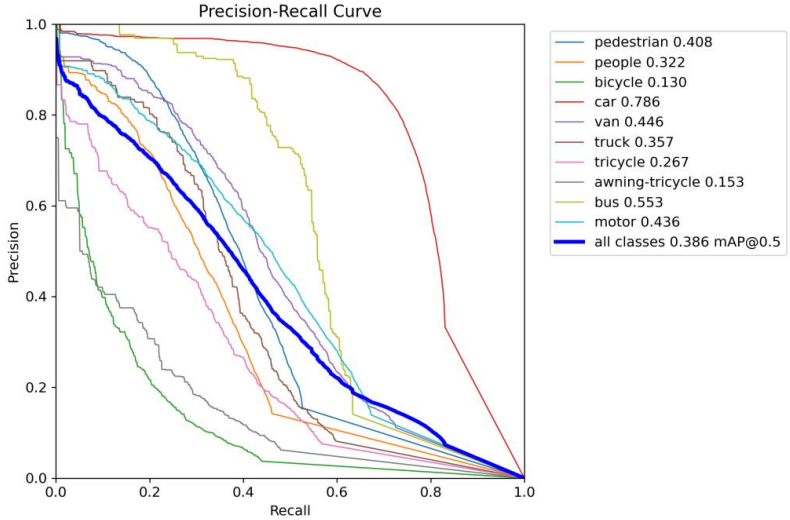


Figure 5 Basic Model PR Curve

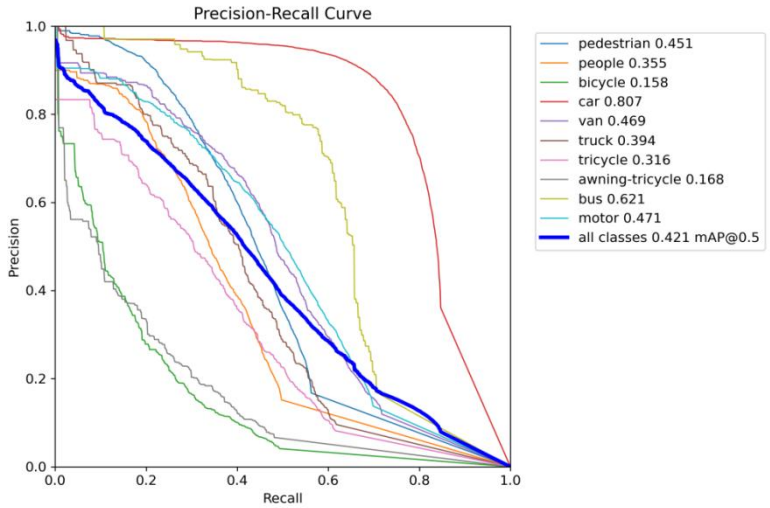


Figure 6 TFE-Net Model PR Curve

By comparing the PR curve trajectories in Figures 5 and 6, it can be observed that the latter achieves a Pareto frontier breakthrough in precision-recall co-optimization, indicating that the improved model has higher precision. For the pedestrian category, the PR curve is more stable in the high recall rate region, indicating more accurate detection of pedestrians. The curve for the bicycle category also shows a significant upward shift, indicating a more substantial enhancement of the model's ability to detect small objects. In terms of specific data, the TFE-Net model shows varying degrees of improvement in the mAP@0.5 values across all categories. For the pedestrian category, it improves from 0.408 to 0.451, and for the bicycle category, it increases from 0.130 to 0.158. This suggests a significant improvement in the model's performance when dealing with small and easily confusable objects. Moreover, the overall mAP@0.5 across all categories improves from 0.386 to 0.421, indicating an enhancement in the model's overall detection performance. Additionally, the regions where the original model suffered from severe missed detections are effectively captured by the improved model. This demonstrates that the improved model has enhanced its ability to express shallow features and can better capture the local details of small objects, thereby reducing missed detections. This is also consistent with the improvement in the PR curve, indicating a significant improvement in the model's ability to detect small objects in complex scenes.

3.5.2 Attribution analysis of improvement effects

The improved model in this study has achieved remarkable performance gains in small object detection, and the main reasons can be attributed to the implementation of the following key strategies. Firstly, a specially designed small object detection head (i.e., a shallow pathway) has been introduced, which effectively compensates for the feature compression and detail loss of small objects caused by the traditional downsampling process. This detection head directly utilizes high-resolution feature maps from shallow layers for processing, thereby significantly reducing feature loss and substantially enhancing the model's ability to represent the features of small objects. Secondly, the feature fusion structure of the model has been optimized, which significantly improves the cross-scale semantic consistency. The optimized feature fusion mechanism can more efficiently integrate feature information from different scales, enabling the model to perform more accurately when detecting objects of varying scales. Finally, high-resolution feature maps have been made to participate directly in the detection process, which has greatly improved the model's ability to recall tiny objects. High-resolution feature maps contain richer local detail information, which helps the model to more accurately locate and identify small objects, thereby effectively reducing the occurrence of missed detections. These design improvements work together in synergy to collectively drive the significant enhancement of the TFE-Net model's detection performance in small object detection tasks.

4 CONCLUSION

This paper presents TFE-Net (Tiny-aware Feature Enhancement Network), a lightweight framework for enhancing shallow perception in YOLOv8s to address performance degradation in small object detection. By integrating shallow high-resolution feature pathways and local perception paths, TFE-Net significantly improves detection of tiny objects. Its architecture focuses on minor structural modifications to build a shallow perception framework with high feature sensitivity and discrimination, offering an optimized, cost-effective solution for small object detection. From a theoretical perspective, TFE-Net's design emphasizes weak feature reconstruction and diverse feature paths, providing a new framework for identifying small objects in complex settings. In practical terms, it efficiently enhances YOLOv8's architecture, ensuring compatibility with mainstream inference frameworks and showcasing strong transferability and deployability. This makes it suitable for various scenarios like traffic monitoring, UAV security, and industrial defect detection.

However, TFE-Net has certain limitations. It hasn't optimized loss function reconstruction or dynamic label allocation, and its local robustness can be further enhanced. It also mainly depends on RGB images and lacks multimodal data

fusion capabilities. Future research will focus on several key areas. We will integrate infrared, depth, and radar data for multimodal enhancement to enrich semantic representation. Unsupervised and self-supervised learning methods will be explored to reduce reliance on labeled data and improve adaptability in data-sparse areas. Architecture optimization via Neural Architecture Search (NAS) will be conducted to achieve high-performance, low-power detection. We will also apply end-to-end DETR structures to small object tasks, promoting the shift from "Anchor to Attention." Additionally, cross-scale dynamic perception and local refinement will be emphasized. By utilizing lightweight convolutional modules and high-resolution feature pathways, we can enhance shallow features and boost local fine-grained feature expression for better small object detail capture. Efficient perception path designs, such as incorporating GhostNet and MobileNet modules along with feature sparsity compression, will continue to be developed. These approaches balance efficiency and effectiveness by reducing computational complexity while maintaining perception performance. Overall, small object detection is moving towards multi-layer, cross-scale, non-linear feature reconstruction. This evolution offers new breakthroughs for micro-perception in complex environments. Feature space density reconstruction and perception path diversity expansion are becoming crucial in SOD research. These improvements will enhance the model's robustness, adaptability, and efficiency, unlocking more potential in complex scenarios. We believe TFE-Net will play a key role in advancing the field of small object detection.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Tian H, Zheng Y, Jin Z. Improved RetinaNet model for the application of small target detection in the aerial images//IOP Conference Series: Earth and Environmental Science. IOP Publishing, 2020, 585(1): 012142.

[2] Ahmed M, Wang Y, Maher A, et al. Fused RetinaNet for small target detection in aerial images. International Journal of Remote Sensing, 2022, 43(8): 2813-2836.

[3] Ahmad M, Abdullah M, Han D. Small object detection in aerial imagery using RetinaNet with anchor optimization//2020 International conference on electronics, information, and communication (ICEIC). IEEE, 2020: 1-3.

[4] Dubey S, Olimov F, Rafique M A, et al. Improving small objects detection using transformer. Journal of Visual Communication and Image Representation, 2022, 89: 103620.

[5] Dai X, Chen Y, Yang J, et al. Dynamic detr: End-to-end object detection with dynamic attention//Proceedings of the IEEE/CVF international conference on computer vision. IEEE, 2021, 10: 2988-2997.

[6] Cao X, Yuan P, Feng B, et al. Cf-detr: Coarse-to-fine transformers for end-to-end object detection//Proceedings of the AAAI conference on artificial intelligence. AAAI, 2022, 36(1): 185-193.

[7] Wang H, Liu C, Cai Y, et al. YOLOv8-QSD: An improved small object detection algorithm for autonomous vehicles based on YOLOv8. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-1.

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

[9] Shen L, Lang B, Song Z. DS-YOLOv8-based object detection method for remote sensing images. IEEE Access, 2023, 11: 125122-125137.

[10] Du D, Zhu P, Wen L, et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results//Proceedings of the IEEE/CVF international conference on computer vision workshops. IEEE, 2019, 1: 0-0.