Publisher: State and Provincial Joint Engineering Lab. of Advanced Network Monitoring and Control (ANMC)

Cooperate:

Xi'an Technological University (CHINA)
West Virginia University (USA)
Huddersfield University of UK (UK)
Missouri Western State University (USA)
James Cook University of Australia
National University of Singapore (Singapore)

Approval:

Library of Congress of the United States

Shaanxi provincial Bureau of press, Publication, Radio and Television

Address:

4525 Downs Drive, St. Joseph, MO64507, USA

No. 2 XueFu Road, WeiYang District, Xi’an, 710021, China

Telephone: +1-816-2715618 (USA) +86-29-86173290 (CHINA)

Website: www.ijanmc.org

E-mail: ijanmc@ijanmc.org

xxwlcn@163.com

ISSN: 2470-8038

Print No. (China): 61-94101

Publication Date: June 28, 2024

Editor in Chief

Associate Editor-in-Chief

Peter the Great St. Petersburg Polytechnic University, Russia

Administrator

Dr. & Prof. George Yang
Department of Engineering Technology
Missouri Western State University, St. Joseph, MO 64507, USA

Professor Zhongsheng Wang
Xi'an Technological University, China
State and Provincial Joint Engineering Lab. of Advanced Network and Monitoring Control, China

Associate Editors

Prof. Yuri Shebzukhov
International Relations Department, Belarusian State University of Transport, Republic of Belarus.

Dr. & Prof. Changyuan Yu
Dept. of Electrical and Computer Engineering, National Univ. of Singapore (NUS)

Dr. Omar Zia
Professor and Director of Graduate Program
Department of Electrical and Computer Engineering Technology
Southern Polytechnic State University
Marietta, Ga 30060, USA

Dr. Baolong Liu
School of Computer Science and Engineering
Xi'an Technological University, CHINA

Language Editor

Would you like to be an Associate Editor? Simply send a request together with your Curriculum Vitae to xxwlcn@163.com. We will have a team of existing editors or at least three experts in your field to review your request and make a decision as soon as we can. The criteria to be an associate editor are: 1. must have advanced degree; 2. must be a leader or have outstanding achievements in the specific research field; 3. must be recommended by the review team.

Table of Contents

Research on Smart City Platform Construction Technology for Digital Twins

Jianxiong Zhang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: agonytriumph@163.com

Shiqian Wang
School of Computer Science & Engineering
Xi'an Technological University
Xi'an, China
E-mail: 1178208937@qq.com

Wuqi Gao
School of Computer Science & Engineering
Xi'an Technological University
Xi'an, China
E-mail: gaowuqi@126.com

***Abstract*—Urban digital twin is a key step to build a smart city, digital twin is an important application scenario for smart city platform, and the relationship between the two are both current research hotspots. In this paper, we will start from the demand of digital twin on smart city platform, study the architecture method and key technology of smart city platform, in this paper's platform construction method compared to the traditional construction method, reduces the difficulty of digital twin smart city construction, and also reduces the coupling degree between smart city platform modules, and use a smart city platform for engineering verification. Finally compared with the traditional smart platform construction techniques, the techniques in this paper are better than the traditional ones in terms of coupling, difficulty and cost. Through engineering verification and experimental results show that this paper on the digital twin-oriented smart city construction technology, the coupling degree of each module is the lowest, and in the development efficiency experiments, this paper by comparing with the traditional technology, the experimental development cycle compared to the traditional technology can be shortened by 61.7% of the development cycle, greatly reducing the development cost and improving the construction efficiency.**

Keywords-Digital Twin; Smart City; Engineering Verification

I. INTRODUCTION

As a leading technology in the construction of new smart cities, digital twin technology has become the core of building smart city platforms with its characteristics such as real-time monitoring and dynamic simulation. This paper aims to explore the key technologies and applications of digital twin technology in the construction of smart city platforms. Through IoT sensing, information modeling, ubiquitous network and other technologies, digital twin technology realizes real-virtual connection and mapping, providing the possibility of global spatio-temporal data fusion for urban operations. In this context, the paper will study the five-dimensional structural model of digital twins in detail, deeply analyze the basic principles of digital twin driving, and explore the application of the digital twin standard system framework. Through the research of the paper, it aims to provide technical support for the digital twin application of smart city platforms and promote the intelligent development of urban management, public services and other fields. Therefore, the construction of urban information models with digital twin technology as the core is the foundation of new smart cities, and provides a new idea for realizing three-dimensional visual monitoring of smart cities [1].

The digital twin technology in this article mainly focuses on three aspects: first, GIS data processing and unity 3D is used to realize three-

DOI: 10.2478/ijanmc-2024-0011

dimensional modeling of highways, green spaces, water systems, and real-life buildings in the urban environment; secondly, geographical information based on GIS (Geographical Information System) is used system) data to visually present a digital twin of the city that integrates sky, earth, and air, thereby providing spatio-temporal data services and other related application support for smart applications in various departments of the city, and assisting the construction of new infrastructure in the park; Finally, edge computing and terminal sensors and cloud computing are used to provide data support for the entire smart city platform [2]-[6].

II. Definition of Digital Twin

The concept of Digital Twins originated in the field of industrial manufacturing, and by combining with a new generation of information technology such as 5G communication, Internet of Things, cloud computing, big data, artificial intelligence, etc., it has realized the actual landing and transformation of Digital Twin theories in many industrial fields, and gradually extended to the application fields of smart city, smart transportation, smart water conservancy, etc. [3]. The digital twin is a digital model of existing or to-be physical entity objects, which perceives, diagnoses, and predicts the state of physical entity objects in real time through actual measurement, simulation, and data analysis, regulates the behavior of physical entity objects through optimization and instruction, and evolves itself through mutual learning among related digital models, while improving the decision-making of stakeholders during the life cycle of physical entity objects. The construction of digital twin smart cities will lead to disruptive innovations in city only management and services. The digital twin smart city is used in scenarios where buildings, roads, water systems, green spaces and other infrastructures in the city have corresponding virtual impacts in the digital world, where attribute information and dynamic change information are visible, trajectories can be traced, and statuses can be investigated; the virtual and real are synchronized to operate, and the scenarios are blended; the past can be traced, and early warnings for safety accidents and expectations of external development trends can be carried out; and the virtual services are realistic, and simulated decision-making can be simulated.

III. Smart City Platform Architecture

In 2023, the "First Edition of Digital Twin Industrial Software White Paper" was released. Many related digital twin manufacturers have also announced their own digital twin platform technology architecture. The divisions are basically similar, and they are basically divided into physical entities, virtual models, application services, and twin data.

It is traditionally believed that the industrial Internet platform deployment architecture is divided into three levels: end, edge and cloud. If this architecture is applied to the construction of smart city digital twins, the production cost and implementation cycle will be very large. Regarding this problem, the architecture proposed in this article as shown in Figure 1, it is divided into three parts: the first part: the digital layer, which refers to converting the city into the corresponding digital model. This article proposes the construction of a converged digital body based on Unity 3D and Web platforms; the second part: the network layer, this article divides it into edge network and central network. The edge network performs field-level calculations on edge devices and finally aggregates the processed data to the central network, thereby reducing the pressure on the central network and improving the efficiency of network transmission; Section Three parts: Device layer, this article will be divided into soft gateway, hard gateway, and IoT devices[7][8].

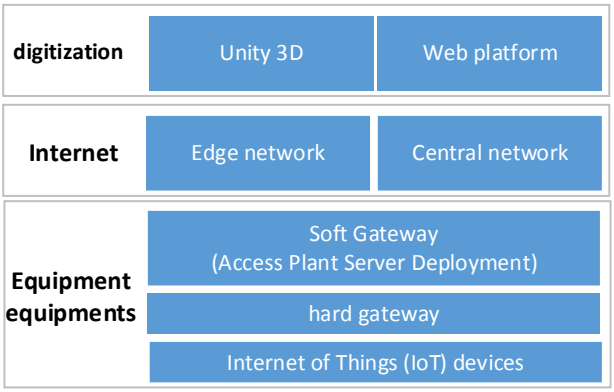


Figure 1. Platform Deployment Architecture Diagram

A. Mathematical layer analysis

The digitalization layer is subdivided into Unity 3D modeling and Web platform display integration, which improves the difficulty and efficiency of digitalization. In this paper, in order to verify the feasibility of the platform, a city's smart digital twin platform is used for verification, which is generally divided into three steps: ① Pruning and calibration of the city's GIS data to ensure that the GIS data and the real city scene correspond to each other, and the whole process is completed using QGIS software; ② The processed GIS data is imported into Unity 3D to make a real-life three-dimensional model of the target area; ③ Construct the Web platform, integrate the city digital model with the Web platform, and construct the three-dimensional display platform; Figure 2 shows the digital layer architecture diagram.

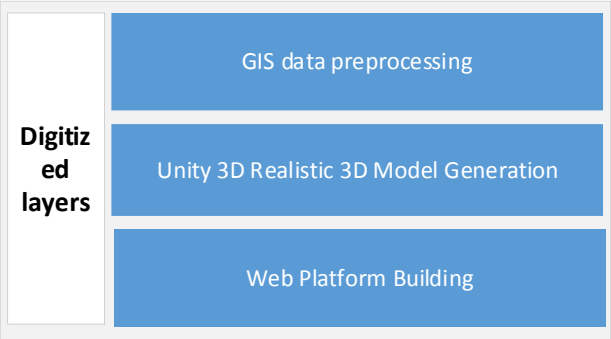


Figure 2. Mathematical Layer Architecture Diagram

B. Network Layer Analysis

For the five-dimensional model proposed in the Digital Twin 2023 whitepaper, which over-centralizes data, this model facilitates data management, but greatly increases the coupling of data and increases the intensity of central data processing. [9]

In this paper, we propose the edge network plus center network, as shown in Figure 3 the edge network is responsible for processing the data on the edge side and sending it to the IoT server through MQTT protocol, and the center network is responsible for the network service of the whole platform. The IoT server and the back-end server in Fig. 4 are in the same layer and they exchange information. This process will be explained in detail in this paper through engineering.

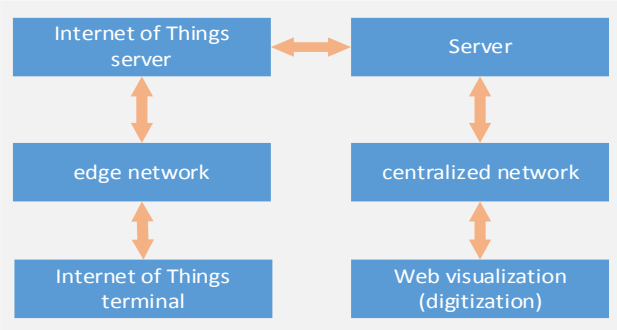


Figure 3. Network layer architecture diagram

C. Device level analysis

The device layer is subdivided into IoT devices, hard gateways, and soft gateways. IoT devices are considered to be at a level that does not have independent collection and communication capabilities. However, as Microcontroller Unit (MCU) technology becomes increasingly mature, more and more hardware is equipped with MCUs, and the time is divided into devices leaving the factory. It can be divided into front-mounted and rear-mounted. The rear-mounted MCU also needs to be installed with sensors or flow meters. It looks like a cheaper customized communication box. The existence of the rear-mounted MCU will weaken or cancel the hard gateway. Hard gateway is the most traditional end-side technology. It is generally believed that soft gateway can replace hard gateway in function and is more economical, because the difference between the two is that the deployment of the collection and forwarding program is changed from the box to the LAN server. However, in this article, hard gateways and soft gateways are used in parallel to ensure normal communication on different types of networks. Figure 4 shows the device layer network topology.

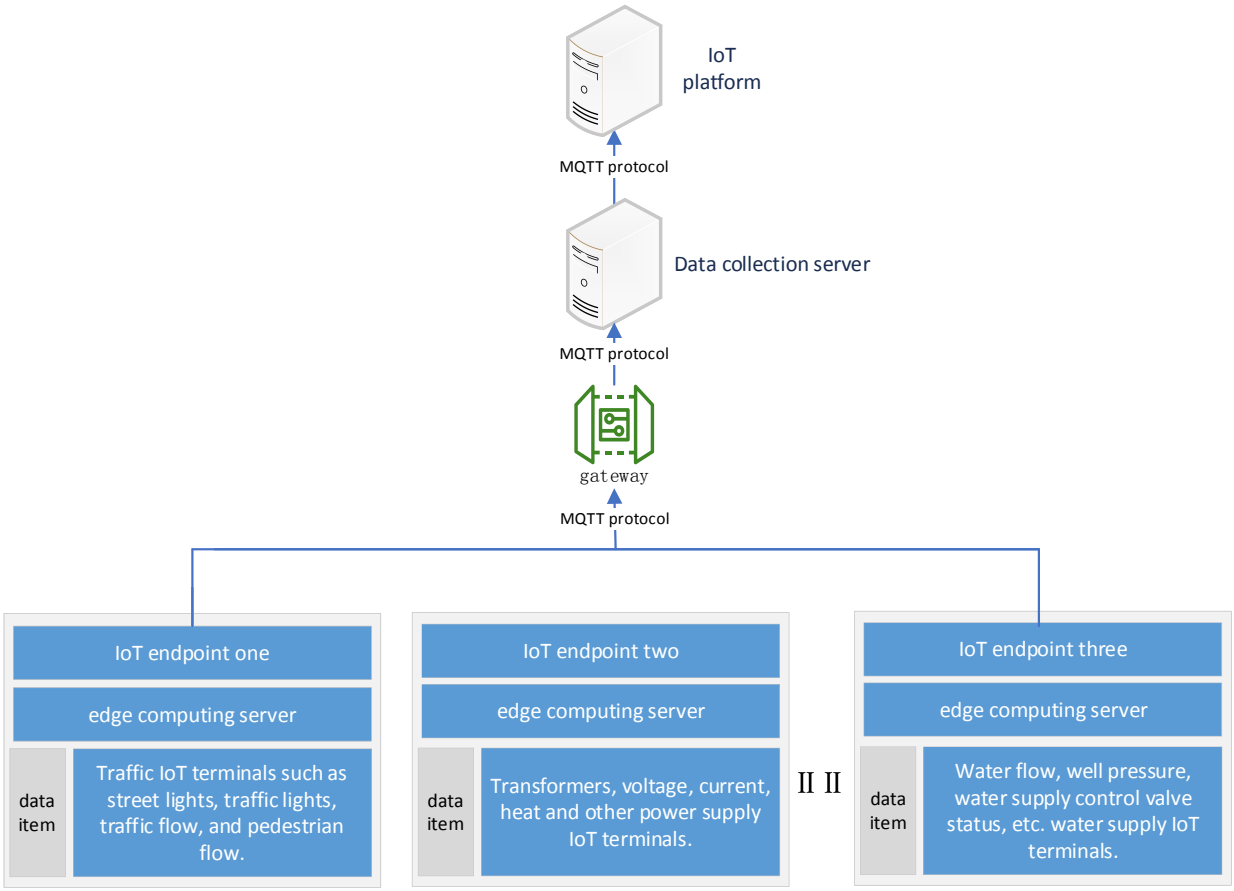


Figure 4. Device layer network topology diagram

IV. ENGINEERING VERIFICATION

As digital city construction continues to advance, local governments and urban district authorities are paying more and more attention to geographical information data. They are using geographical information, satellite remote sensing, smart terminals and other data to assist urban construction and planning, and governance [10]. The application of geographical information data can effectively solve the problems of poor information flow and lack of professional data support in urban construction in the new era. Through the sharing of resources in various fields, a new urban information system and service method can be formed to provide various services. Provide a basis for regional informatization and standardization construction work. This verification will be completed strictly in accordance with the above-mentioned smart city platform architecture for digital twins.

A. Engineering design

Currently, with its distributed system architecture, data sharing services, efficient computing load, independent client and other advantages, map display based on WebGIS technology has become the mainstream architecture model of GIS. The project adopts a B/S architecture and uses a traditional browser as the client to facilitate client access and data sharing. This project refers to the above-mentioned architecture and is also divided into a digital layer and a network layer. The device layer realizes loose coupling between modules, and each module can be used independently.

1) Digital layer. As shown in Figure 5 and Figure 6, QGIS performs pruning operations and verification on urban data, and the GIS data generates a three-dimensional real-life model in Unity 3D.

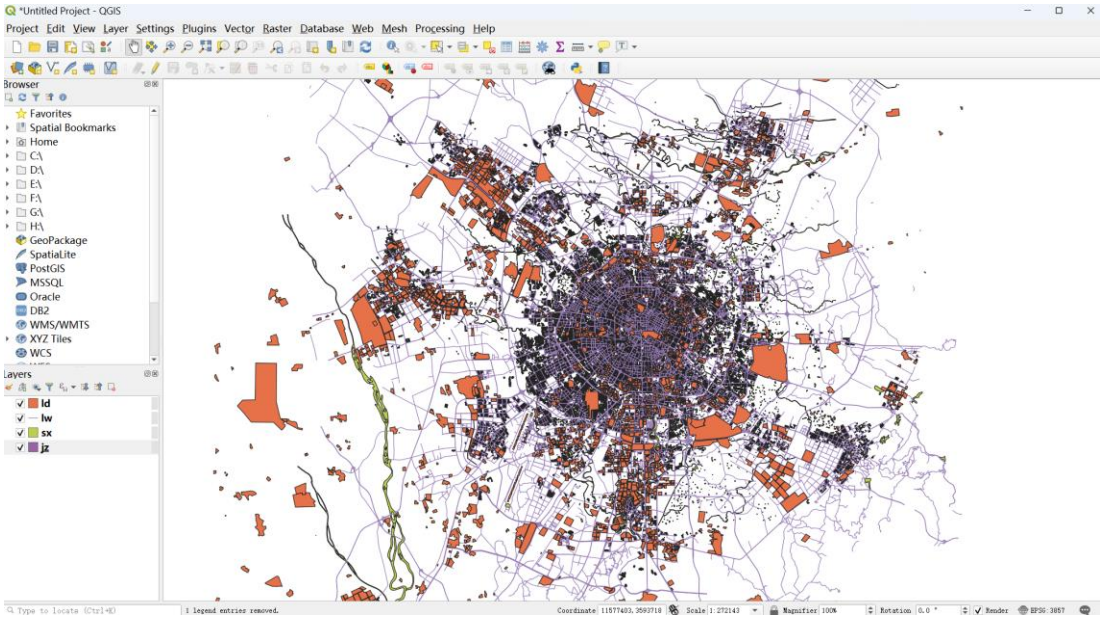


Figure 5. GIS data preprocessing diagram

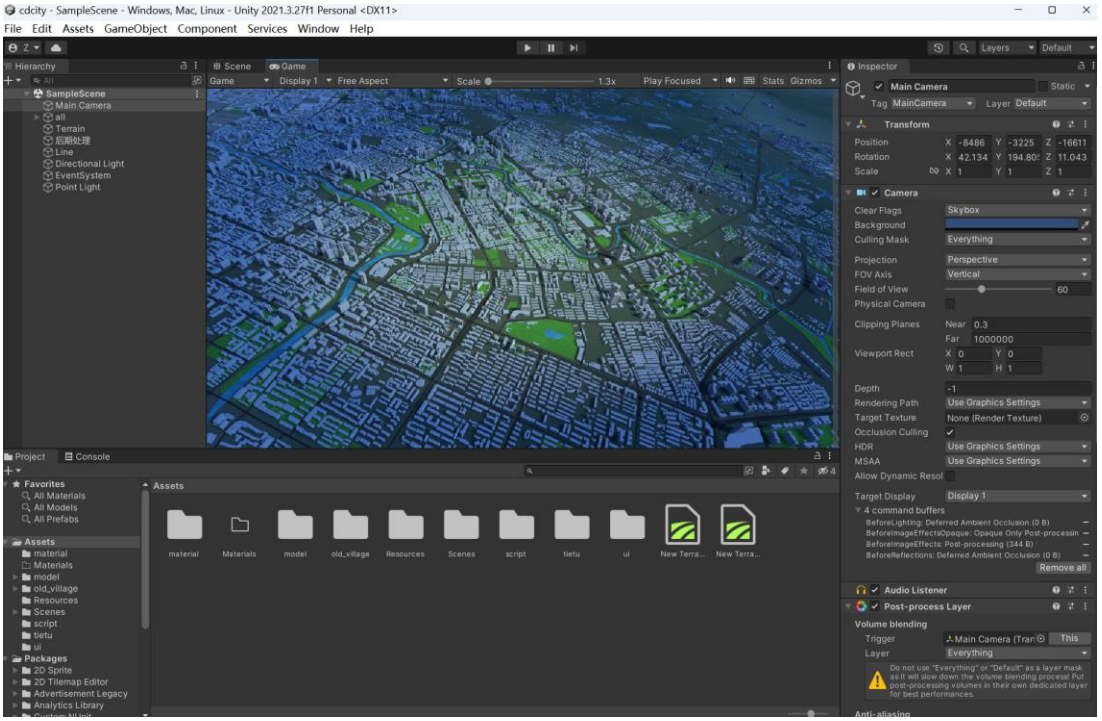


Figure 6. Unity 3D generates three-dimensional real-life imagesimages

2) The network layer. According to the framework of this article, is divided into edge network and central network. The edge network is responsible for the connection between IoT terminals and IoT servers, while the central network is responsible for connecting our back-end servers, which is responsible for our entire system. Intercommunication of data between various modules. Figure 7 is the visual interface of our IoT server, in which IoT terminals can be managed through this interface, such as deleting, adding, and modifying, and we can also observe

the status of terminal devices in real time, such as data measured by sensors and relays. Switch, steering gear rotation angle, etc.

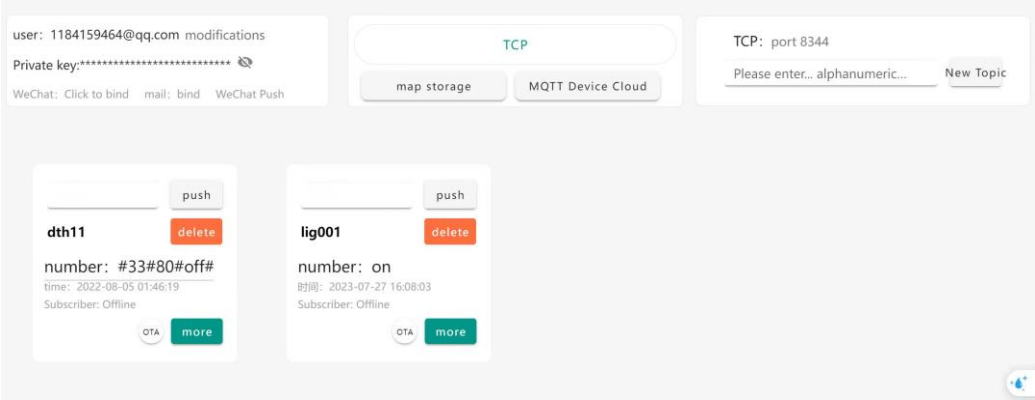


Figure 7. IoT server visual interface

3) Device layer. In this project we use the NodeMCU development board as the city's IoT terminal, in the project can be used to collect data and intelligent control of multiple NodeMCU development version, because the NodeCMU inherited the ESP8266 wifi module, it can be easily accessed to the local area network or the Internet, and you can pre-process the collected data on the board. It can also pre-process the collected data on the board and finally transmit it to the IoT cloud server through the edge network. The whole process of engineering device layer design and realization is strictly in accordance with the network topology diagram in Figure 5. As Figure 8 shows the NodeMCU pinout diagram.

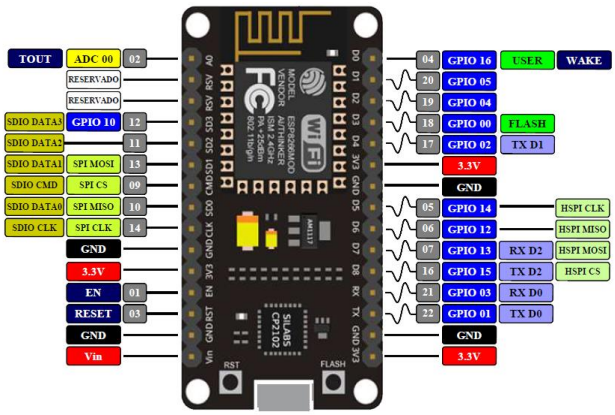


Figure 8. NodeMCU Development Board Pinout

B. Project summary

1) Overall project design. This project is based on the above-mentioned architectural design, and separates each module to reduce the coupling of the project. The entire model can be changed at will like building blocks, and the modules will not be affected. At the digitalization layer, we used GIS data preprocessing and imported it into Unity 3D to quickly obtain our three-dimensional real-life city, including models of the city's road network, water system, grassland, buildings, etc. By packaging it into a webGL file, we integrated the model into the web platform for easy display and can also be used across platforms and devices. At the network layer, the data collected by the IoT terminal will be pre-processed by the terminal's onboard MCU and sent back to the IoT server through the edge network. Finally, the back-end server can request data from the IoT server through API. The relevant data will be analyzed on the cloud computing platform and finally displayed on the web platform through API. At the equipment layer, the IoT terminals mainly collect various data of the city, and there is also a management platform that sends instructions to our IoT terminals to start and stop some equipment in the city. Through the above, our smart city can form a closed loop. This architecture also provides theoretical support for building loosely coupled smart cities, greatly reducing the production difficulty, cost and cycle. Figure 9 is the overall architecture diagram of this project.

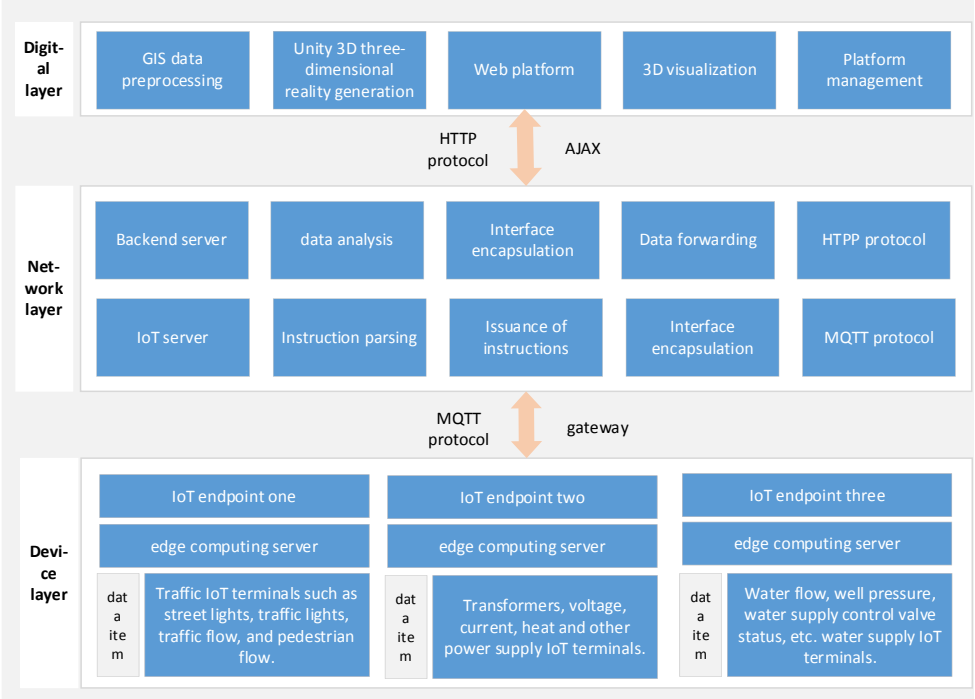


Figure 9. General Architecture Diagram of a Smart City

Figure 10 is a web platform interface diagram that integrates a three-dimensional city model. The model behind the interface can be operated with the mouse. For example, the day and night of the entire city will be dynamically displayed on the platform. When night comes, the model behind the platform will also be displayed. After entering the night state, the command that it is night will be sent to our terminal equipment, and the terminal equipment will control the public lights in the entire city to be on.

Figure 10. Digital twin platform for a smart city

2) Validation results. We have verified the above architecture through projects. As shown in Table 1, the coupling degree of each module, whether it works independently, and whether it is

cross-platform have been verified through the above project.

TABLE I. COUPLING, INDEPENDENT WORKING, CROSS-PLATFORM VERIFICATION TABLE

module	*Coupling*	*working independently*	*cross-platform*
Unity 3D	Low coupling	Yes	Yes
Web platform	Low coupling	Yes	Yes
IoT terminal	Low coupling	Yes	Yes
backend server	Low coupling	Yes	Yes

There are many software designed in this project because it has gone through many processes, such as: GIS data preprocessing, GIS data reality model generation, 3D reality model generation, Web platform construction, data interaction, C# script writing, IoT terminal programming, and hardware simulation design. Wait for multiple processes. Table 2 will list the software designed in the above process and its functions.

TABLE II. SOFTWARE USED IN ENGINEERING AND ITS FUNCTIONS

software	*effect*
Visual Studio Code	Develop web platform and server and write corresponding code.
QGIS	Preprocess GIS data, such as data pruning.
Unity 3D	Generate 3D reality models.
Visual Studio 2022	Write C# scripts for dynamic interaction in Unity 3D.
Arduino IDE	Hardware program writing and burning.
Postman	Data interaction API interface test.

For the platform in this paper to build the feasibility of technology and development cycle, in this paper will be used and the traditional technology compared to develop the same effect in time and coupling to do further experiments, respectively, using the two architectures to build a complete digital twin case study, to see the actual completion of the time (the smallest unit of time for the hour) and coupling degree. For the traditional architecture in the number of chemical layer is used completely unity for development, the network layer uses a central network to build, and unity to build the model to do interaction, for the device layer, the use of serial communication or the use of bluetooth, etc., direct or indirect communication methods. Finally, according to the actual experimental comparison, the construction efficiency is improved by about 61.7%. Specific data as Table III.

TABLE III. COMPARISON OF ARCHITECTURE DEVELOPMENT CYCLE TIMES AND THEIR COUPLING DATA TABLE

Construction Methods	*traditionally constructed*	*this paper constructs*
digitalization layer	24h(high coupling)	18h(Low coupling)
network layer	48h(high coupling)	40h(Low coupling)
device layer	10h(high coupling)	8h(Low coupling)

V. CONCLUSIONS

Through engineering verification, this architecture is feasible and loosely coupled in actual applications. Each module of the architecture can be organized in any way, and each module can run independently. For example, if the real-life model built by our Unity 3D is not integrated into The web platform can also run and communicate with normal data. If you need to access other services later, such as AI functions, you only need to connect the interface to the back-end server at the network layer. A digital twin cannot yet reach the stage of shared intelligence. Multiple digital twins can be constructed, and they can share wisdom and evolve together.

The above project demonstration video has been uploaded to open source platforms and self-media platforms, and has also received suggestions and discussions from digital twin researchers and enthusiasts, as well as very strong interest in this architecture. At present, a rough estimate of the number of people interested in this project and architecture has There are about 20,000 people, of which graduate students and doctoral students can account for about 50%, and the rest are related workers and undergraduate students. Since the original project is relatively large from design to model, the total size is about 2GB, and it has not yet been open sourced on Github. Later, I simplified the interface and rebuilt a simplified version of this project based on the architecture of this article. It is currently open source. Because it is still relatively large, readers need to use the git lft tool when downloading this

open source project to download large files from GitHub document. Open source URL: https://github.com/zjxWeb/digitalTwins-short.

The architecture of this article is still in its infancy, and there is still a long way to go to achieve the goal of digital twin co-intelligence. In the future, more new technologies should be added, the architecture should be kept low-coupled, and a universal digital twin should be built. Twin architecture. Figure 10 is a preliminary idea for further improvement of the architecture. Engineering verification has not yet been carried out. However, this architectural idea is also constructed in strict compliance with the architecture of this article. It only adds and deletes modules based on the loose coupling of this article's architecture. , achieve rapid construction.

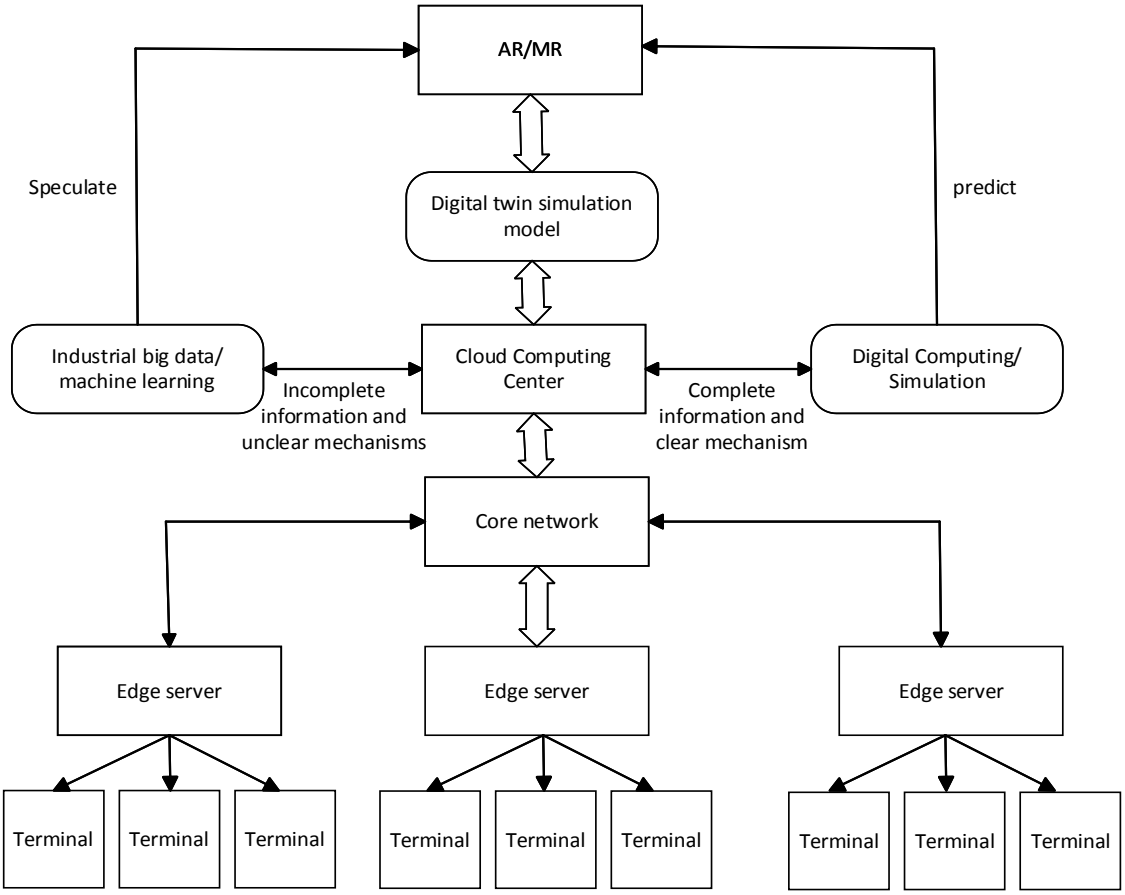


Figure 11. Architecture upgrade flow chart

For Figure 11, the entire display and interaction part has been upgraded during the process. With the rapid development of augmented display technologies such as AR and MR, there is an increasing trend that AR/MR glasses may become an integral part of our daily life, work, and study. A second screen will greatly facilitate all aspects of our lives. At the same time, this flow chart will judge the data in the cloud computing center. When our data has incomplete information and unclear mechanisms, we need to make inferences through industrial big data and machine learning. Finally, the inference results will be transmitted to our display end. It is AR/MR; similarly, when the data in cloud computing has complete information and clear mechanisms, it is necessary to predict future data through digital calculation and simulation, and finally the prediction results are transmitted to our display, which is AR/MR; later The latter has higher accuracy than the former because we make predictions based on complete information and clear mechanisms.

REFERENCES

[1] Wang Yujie, Shi Kaimin, Lou Shujian. Research on comprehensive design and application of digital twin Sanmenxia Water Conservancy Project [J]. Water Conservancy Informatization, 2022, 2022(06):1-6+14.

[2] Xu Yuanxiao, Lou Huan, Chen Shu. Practice of digital twin technology in smart parks [J]. Communications and Information Technology, 2023 (06):41-44.

[3] Wang Hao. Visualization of underwater acoustic information processing platform based on digital twin technology [J]. Electroacoustic Technology, 2022, 46(05):25-28.

[4] Shi Yingna. Research on geographical information data processing methods in digital city construction [J]. Technological Innovation and Productivity, 2022, 2022(11):71-73+77.

[5] Tang Dandan. Design of map data organization model based on improved usability of GIS map vectorization results [J]. Intelligent Buildings and Smart Cities, 2024 (01):36-38.DOI:10.13655/j.cnki.ibci.2024.01.009.

[6] Han Wenlong. Research and development of management system for housing construction on randomly occupied farmland based on WebGIS [D]. 2023. DOI:10.26976/d.cnki.gchau. 2022. 001402.

[7] Wang Jinghong, Chen Yu, Hu Jianqiang. Partial offloading analysis based on reinforcement learning in mobile edge networks [J]. Journal of Minnan Normal University (Natural Science Edition), 2023, 36(04):62-72.DOI:10.16007/j.cnki.issn2095-7122.2023.04.006.

[8] Yuan Qiufeng. Intelligent desk lamp design based on NodeMCU and WxBit [J]. Journal of Ningde Normal University (Natural Science Edition) [10], 2023, 35(04):382-386.DOI:10.15911/j.cnki.35-1311/n.2023.04.012.

[9] Tao Fei. Digital Twin Industrial Software White Paper [R]. 2023.

[10] Zhang Lunyan. Research on industrial Internet platform construction technology for digital twins [J]. Intelligent Manufacturing, 2022.

Uncovering Cybersecurity Vulnerabilities: A Kali Linux Investigative Exploration Perspective

Zarif Bin Akhtar
MPhil Research Postgraduate Student, Master of Philosophy (MPhil) in Machine Learning and Machine Intelligence, Department of Engineering, University of Cambridge, United Kingdom
E-mail: zarifbinakhtarg@gmail.com ; zarifbinakhtar@ieee.org

Ahmed Tajbiul Rawol
Bachelor of Science (B.Sc.) in Computer Science & Software Engineering (CSSE), Faculty of Science and Technology, Department of Computer Science, American International University-Bangladesh (AIUB), Dhaka, Bangladesh
E-mail: tajbiulrawol@gmail.com

***Abstract*—This research exploration presents a comprehensive methodology for conducting penetration testing for networking security protocols and vulnerabilities on the Wi-Fi networks using Kali Linux, an open-source penetration testing platform. The methodology also encompasses four main stages which are Preparation, Information Gathering, Simulated Attack, Reporting. In the Preparation Stage, the scope of the penetration test is defined, authorization is obtained, and within the testing environment the experimentation is prepared. The Information Gathering Stage involves scanning for associated nearby wireless access points (APs), identifying encryption modes, and assessing network coverage. The Simulated Attack Phase verifies the types of vulnerabilities through password cracking, infrastructure penetration tests, and client-side attacks. Finally, the Reporting Phase entails compiling of a very detailed test report with results visualized, findings and recommendations with directions. Experimental results validate the overall effectiveness of the methodology in identifying and mitigating Wi-Fi network vulnerabilities. Through systematic testing and analysis, Kali Linux facilitates proactive security measures to enhance Wi-Fi network defenses.**

Keywords-Artificial Intelligence (AI); Cybersecurity; Kali Linux; Linux Distros; Linux OS; Machine Learning; Network Security; Privacy; Security

I. INTRODUCTION

In an era where connectivity is ubiquitous, the security of the wireless networks, particularly Wi-Fi networks, is of paramount importance.

As the prevalence of Wi-Fi continues to expand, so too do the vulnerabilities inherent in these networks, making them prime targets for malicious actors seeking to exploit security loopholes for unauthorized access or data interception. To address these concerns and fortify the defenses of Wi-Fi networks, comprehensive penetration testing methodologies and tools are indispensable [43]. This research delves into the realm of Wi-Fi network security evaluation and enhancement, focusing on the utilization of Kali Linux, an open-source platform renowned for its robust penetration testing capabilities. Through a detailed exploration of associated Wi-Fi network vulnerabilities and common encryption methods, this investigation also aims to illuminate the intricate process of conducting penetration tests using Kali Linux to identify various types of weaknesses, simulate attacks, and fortify the security posture of Wi-Fi networks. The discourse unfolds by examining the distinctive features of Kali Linux tailored for professional penetration testing and security auditing, juxtaposed with insights into its applicability for users with varying levels of Linux proficiency. Emphasizing the importance of understanding the distinctions of Kali Linux and its targeted usage for security professionals, the discourse navigates through the delineation of Wi-Fi penetration testing stages, from preparation and information gathering to simulated attacks and reporting.

The research also delves deeper into practical experiments conducted within a controlled environment, leveraging Kali Linux to simulate real-world scenarios and assess the effectiveness of various penetration testing techniques. Insights gleaned from these experiments underscore the vulnerabilities inherent in Wi-Fi networks and

offer actionable recommendations to bolster their security posture, ranging from password hygiene and encryption protocols to network configuration best practices. Through meticulous analysis and experimentation, this exploration endeavors to provide a very comprehensive understanding of Wi-Fi network security evaluation and enhancement using Kali Linux.

By shedding light on the intricacies of penetration testing methodologies and the efficacy of Kali Linux as a versatile toolset, this research aims to empower security professionals and network administrators in fortifying the resilience of Wi-Fi networks against evolving threats and vulnerabilities as well.

II. METHODS AND EXPERIMENTAL ANALYSIS

The methodology for conducting penetration testing on Wi-Fi networks using Kali Linux involves a systematic step-by-step approach comprising of several distinct stages. Firstly, in the Preparation Stage, the scope of the penetration test is defined to establish the boundaries and objectives of the assessment. This includes obtaining authorization from stakeholders and ensuring legal compliance for conducting the test. Additionally, thorough planning is undertaken to select target Wi-Fi networks, identify potential attack vectors, and evaluate the workload required for the test. Adequate readiness of the testing environment, including the availability of necessary hardware and software resources, is also ensured during this stage.

Following the Preparation Stage, the Information Gathering Stage is initiated. This phase involves setting the wireless network card to monitoring mode to capture network traffic effectively. Various tools available in Kali Linux, such as Airodump-ng and Kismet, are utilized to scan for nearby wireless access points (APs) and connected clients.

Essential information about target APs, including physical addresses, encryption modes, signal strength, and associated clients, is recorded. Additionally, network scanning techniques are employed to identify potential vulnerabilities, enumerate network devices, and map the network topology. The gathered data is then used to assess the network coverage and identify potential attack surfaces within the scope of the penetration test.

Subsequently, the Simulated Attack Phase is conducted to verify potential vulnerabilities identified during the Information Gathering Stage. This involves analyzing the encryption mode of target Wi-Fi networks to determine appropriate password cracking methods.

Tools such as Reaver, Aircrack-ng, and Crunch in Kali Linux are utilized to crack Wi-Fi passwords based on encryption protocols such as WEP, WPA, and WPS. Additionally, penetration tests are performed on target infrastructure, including port scanning, service enumeration, and vulnerability exploitation.

Pseudo-APs are also established to simulate Wi-Fi phishing attacks, and penetration tests are conducted on client devices connected to these pseudo-APs.

Finally, in the Reporting Phase, a comprehensive test report is compiled detailing the findings, vulnerabilities, and recommendations identified during the penetration test. The report includes documentation of the penetration testing procedures, technical methodologies, and results of the assessment. Actionable recommendations for enhancing the security posture of Wi-Fi networks based on the identified vulnerabilities are also provided. The test report is presented to stakeholders, including network administrators and decision-makers, to facilitate informed decision-making and remediation efforts.

The experimental setup involves configuring a representative network topology comprising wireless routers, physical hosts, virtual attack machines, and mobile devices. VMware virtualization technology is utilized to deploy Kali Linux as the attack platform on a virtual machine, with compatible wireless network cards employed to facilitate packet capture and network monitoring.

Pseudo-APs are constructed using wireless network cards, and mobile intelligent terminals are utilized as client devices to assess the effectiveness of simulated attacks on client-side vulnerabilities.

Through this methodology, the penetration testing process using Kali Linux for Wi-Fi network security evaluation is conducted systematically, allowing for the identification of vulnerabilities, simulation of attacks, and formulation of actionable recommendations to enhance network security.

III. Kali Linux Breakdowns: The Cybersecurity Perspective

A. Background Research and Iterative Exploration for associated Available Knowledge

Kali Linux is a specialized Linux distribution renowned for its focus on digital forensics and penetration testing, developed and maintained by Offensive Security Ltd.

It was initially released in March 2013, with subsequent versions enhancing its features and capabilities. The distribution is based on the Debian Testing branch, importing most of its packages from Debian repositories [1-13].

With the availability of approximately 600 penetration-testing programs, Kali Linux offers a comprehensive suite of tools for cybersecurity professionals and enthusiasts alike. These tools range from graphical management tools like Armitage to powerful utilities such as Nmap, Wireshark, Metasploit, John the Ripper, and sqlmap, among others. The inclusion of these tools has made Kali Linux a go-to-choice for security researchers and practitioners [14-24].

Kali Linux was developed by Khaled Baoween (Kali), Mati Aharoni, and Devon Kearns through the rewrite of BackTrack, their previous information security testing Linux distribution based on Knoppix. Its popularity surged when featured in multiple episodes of the TV series Mr. Robot, showcasing its capabilities and tools like Bluesniff, John the Ripper, and Metasploit Framework [25-30].

The distribution's version history reflects its evolution, including notable changes such as the switch from GNOME to Xfce as the default user interface with version 2019.4 and the transition from Bash to ZSH as the default shell with version 2020.3. Kali Linux has specific hardware requirements for installation, including a minimum of 20GB hard disk space, 2GB RAM, and an Intel Core i3 or AMD E1 processor for optimal performance [31-36].

Supported across various platforms including x86, ARM, and Android devices, Kali Linux aims for broader compatibility. The Kali NetHunter project, dedicated to porting Kali Linux to specific Android devices, exemplifies this commitment. Additionally, Kali Linux is available on Windows 10 through the Windows Subsystem for Linux (WSL), extending its reach to a wider user base. In comparison to other Linux distributions focused on penetration testing and cybersecurity such as Parrot OS, BlackArch, and Wifislax, Kali Linux stands out for its comprehensive toolset and features tailored towards cybersecurity professionals.

The distribution includes popular security tools like Aircrack-ng, Burp Suite, Metasploit framework, Nmap, Wireshark, and many more, making it a preferred choice in the cybersecurity community [37-42].

Offensive Security provides extensive resources for Kali Linux users, including the book ***"Kali Linux Revealed,"*** which is available for free download. This resource, combined with the distribution's active community and continuous development efforts, solidifies Kali Linux's position as a leading platform for digital forensics and penetration testing.

As per the data provided from 2023-2024 of the Kali Linux distributions, formerly known as BackTrack Linux, is a Debian-based open-source distribution designed for advanced Penetration Testing and Security Auditing purposes. Its primary objective is to streamline the process for users by providing a comprehensive set of tools, configurations, and automations, allowing them to focus on the task at hand without distractions.

This distribution is extensively customized with industry-specific modifications and includes over a variety of 500-600 tools targeted towards various Information Security tasks such as Penetration Testing, Security Research, Computer Forensics, Reverse Engineering, Vulnerability Management,

and Red Team Testing. It serves as a versatile solution accessible to both information security professionals and hobbyists. The key features of Kali Linux include various types of toolsets.

Comprehensive Toolset: Kali Linux boasts a wide range of penetration testing tools carefully curated to provide users with the necessary functionalities for various security tasks. Tools are regularly reviewed and updated to ensure efficiency and effectiveness.

Free and Open Source: Continuing the tradition of its predecessor BackTrack, Kali Linux is completely free to use and will remain so indefinitely. It is distributed under an open-source license, allowing users to access and modify the source code according to their needs.

Transparent Development Model: The development tree of Kali Linux is openly available, reflecting its commitment to the open-source ethos. This allows users to track changes, contribute to development, and customize packages as required.

Filesystem Hierarchy Standard (FHS) Compliance: Kali Linux adheres to the FHS, simplifying navigation for Linux users by organizing binaries, support files, libraries, and other resources in a standardized manner.

Extensive Wireless Device Support: Kali Linux is designed to support a wide array of wireless interfaces, ensuring compatibility with numerous USB and other wireless devices, which is essential for wireless security assessments.

Custom Kernel with Injection Support: The distribution comes with a custom kernel patched for wireless injection, catering to the needs of penetration testers who frequently engage in wireless assessments.

Secure Environment and Package Signing: The development team operates in a secure environment, with strict protocols for package management and repository interactions. Every package is signed by individual developers and repositories, ensuring authenticity and integrity.

Multilingual Support: Kali Linux offers true multilingual support, enabling users to operate in their native language and access tools in various languages, enhancing accessibility and usability.

Customizability: Acknowledging diverse user preferences, Kali Linux allows for extensive customization, empowering users to tailor the distribution according to their specific requirements, down to the kernel level.

ARM Support: Recognizing the prevalence of ARM-based single-board systems like Raspberry Pi and BeagleBone Black, Kali Linux provides robust support for ARM architectures, ensuring compatibility and functionality across a wide range of ARM devices.

Kali Linux caters to the specialized needs of penetration testing professionals, offering a comprehensive toolkit, robust security features, and flexibility for customization, all within a user-friendly Debian-based environment.

B. Enhancing Web Penetration Testing's with Kali Linux

In the realm of network security assessment, the importance of web penetration testing cannot be overstated. This process, crucial for identifying vulnerabilities and fortifying network security infrastructure, is central to ensuring the integrity and resilience of digital systems [43]. Kali Linux, a sophisticated Linux distribution derived from Back Track Linux, emerges as a formidable tool in this arena, offering advanced features tailored specifically for web penetration testing.

Advanced Features and Capabilities: Kali Linux represents a significant evolution from its predecessor, Back Track Linux, driven by the imperative to combat the escalating threats posed by cyber-attacks. Central to its appeal is the integration of updated tools sourced from Debian repositories, ensuring users have access to the latest security fixes and enhancements. Furthermore, its filesystem architecture is meticulously designed to facilitate seamless execution of security processes, empowering users to run tools from any location within the system. The distribution's emphasis on customization, unattended installation, and flexible desktop environments further enhances its utility as a comprehensive security assessment tool.

Efficiency in Reconnaissance: At the heart of effective web penetration testing lies the reconnaissance phase, where gathering extensive information about the target environment is paramount. Here, Kali Linux shines with its specialized set of Information Gathering tools, meticulously crafted to surpass other distributions in terms of efficiency and effectiveness. These tools encompass a diverse range of functionalities, including ICMP reconnaissance, ping and traceroute commands, DNS-based information gathering, and the utilization of specialized tools like Fierce and Maltego.

Specialized Tools like Fierce and Maltego: Fierce, an integral component of Kali Linux, emerges as a significant asset in the reconnaissance phase. Leveraging DNS mechanisms, Fierce efficiently extracts crucial information about the target by checking DNS servers for zone transfers and employing brute force techniques when zone transfers are restricted. Similarly, Maltego, another powerful tool embedded in Kali Linux, revolutionizes the information gathering process by harnessing publicly available data on the internet and presenting it in a visually intuitive graph format, thereby enhancing the intelligence-gathering process.

IV. KALI LINUX CYBERSECURITY FORENSICS: AN INVESTIGATIVE ANALYSIS

Kali Linux is a Debian-based Linux operating system renowned for its specialized focus on penetration testing, computer forensics, and security auditing. Formerly known as BackTrack Linux, it was redeveloped by Mati Aharoni and Devon Kearns from Offensive Security, with its first version released in March 2013.

Since then, Kali Linux has become the preferred operating system for penetration testers and security professionals due to its comprehensive suite of applications and tools tailored for various information security tasks. With the various types of penetration testing tools available, Kali Linux offers a wide range of utilities spanning vulnerability analysis, information gathering, web application testing, exploitation, wireless attacks, stress testing, sniffing, spoofing, forensics, password cracking, and more. Its popularity stems from its ability to cater to diverse cybersecurity needs while providing a free and open-source platform accessible to all cybersecurity professionals, ethical hackers, and penetration testers.

The history of Kali Linux distros traces back to its predecessors, such as Whoppix and BackTrack, which laid the foundation for its development as a security-focused operating system.

Initially running on Slackware and later leveraging Ubuntu, Kali Linux emerged as a Debian-based distribution in 2013, offering a stable engine beneath its hood. Offensive Security, the cybersecurity training organization behind Kali Linux, continues to update and enhance the distribution with the latest versions of security and penetration testing tools.

In terms of the cybersecurity perspective, Kali Linux plays a pivotal role for professionals across various domains. Its extensive repository of pre-installed cybersecurity tools alleviates the challenges of cost and time associated with acquiring and downloading individual tools. Security professionals can utilize Kali Linux live or install it on various systems, including virtual machines and Raspberry Pi devices. Furthermore, Kali Linux is instrumental in network security auditing, enabling network architects to assess the efficiency and security of their networks.

Key features of Kali Linux include its open-source nature, customizable repository, wide range of penetration testing tools, support for diverse hardware and wireless devices, multilingual support, and flexibility for customization according to individual or organizational preferences. Its versatility and adaptability make it a valuable asset for cybersecurity professionals seeking to perform comprehensive security audits, penetration testing, and cybersecurity research.

For those looking to get started with Kali Linux, various installation methods are available, including using virtualization software like VirtualBox or booting from a live USB drive. Once installed, users gain access to a plethora of popular penetration testing tools, including Nmap, Metasploit, John the Ripper, Netcat, and

Wireshark, among others. These tools empower security professionals to identify, exploit, and validate vulnerabilities within systems, ensuring robust cybersecurity defenses.

Kali Linux stands as a powerhouse in the realm of cybersecurity, offering a versatile and feature-rich platform for conducting comprehensive security assessments and penetration testing. Its evolution from its predecessors, coupled with its extensive repository of tools and user-friendly installation options, makes it an indispensable tool for security professionals seeking to enhance their cybersecurity posture and mitigate potential threats effectively.

A. Kali Linux Cybersecurity Application Tools

In the ever-evolving landscape of cybersecurity, staying ahead of cyber threats requires the utilization of advanced tools and techniques by ethical hackers and penetration testers. Kali Linux, a Debian-derived open-source distribution, has emerged as a go-to platform for cybersecurity professionals, offering a comprehensive suite of over various types of network tools designed for penetration testing and security auditing. As cybercrime continues to pose a significant threat to IT systems, the need for effective penetration testing tools becomes increasingly crucial in identifying and mitigating vulnerabilities. In the timeline years of 2020-2024, ethical hackers and penetration testers have access to a diverse range of Kali Linux tools tailored to various cybersecurity tasks. These tools play a vital role in evaluating the effectiveness of an organization's cyber defenses by simulating cyberattacks and identifying exploitable vulnerabilities in networks, user security, and web applications. By launching simulated cyberattacks with the host's knowledge, ethical hackers can pinpoint weak spots in network infrastructure and guide efforts to bolster security measures.

One notable Kali Linux tool for Wi-Fi network security testing is Fluxion, which specializes in MITM (Man-In-The-Middle) WPA attacks. Fluxion enables penetration testers to scan wireless networks for security flaws without resorting to time-consuming brute force cracking attempts. Similarly, John the Ripper, a multi-platform cryptography testing tool, facilitates password strength testing through brute force attacks, making it ideal for assessing an organization's password security.

Lynis stands out as one of the most comprehensive tools available for cybersecurity compliance, system auditing, and vulnerability scanning. With its ability to run up to 300 security tests on remote hosts and provide detailed output reports, Lynis serves as an effective platform for penetration testing and system hardening.

The Metasploit Framework, a Ruby-based platform, empowers ethical hackers to develop, test, and execute exploits against remote hosts, making it a potent tool for penetration testing. With features like network enumeration, vulnerability exploitation, and data collection, Metasploit Framework enhances the capabilities of cybersecurity professionals in identifying and addressing security vulnerabilities.

Nikto, a web server scanning tool, enables penetration testers to discover security vulnerabilities and related flaws in web applications. By scanning multiple ports, identifying default file names, and detecting insecure file patterns, Nikto complements other vulnerability scanners and provides valuable insights into web application security.

Nmap, a renowned network mapper tool, allows penetration testers to discover active hosts within a network and gather information related to penetration testing. With features like host discovery, port scanning, and OS detection, Nmap enhances the network reconnaissance capabilities of cybersecurity professionals.

Skipfish, similar to WPScan but applicable to various web applications, acts as an effective auditing tool for crawling web-based data. With its automated learning capabilities and low false positive ratio, Skipfish facilitates quick insight into the security posture of web applications.

The Social Engineering Toolkit (SET) is an indispensable tool for launching social engineering attacks, including Wi-Fi AP-based attacks, SMS and email attacks, web-based attacks, and the creation of malicious payloads. SET empowers

ethical hackers and penetration testers to simulate social engineering attacks and assess an organization's susceptibility to social engineering tactics.

Kali Linux offers a very robust arsenal of penetration testing tools that empower ethical hackers and penetration testers to assess and mitigate cybersecurity risks effectively.

As cyber threats continue to evolve, the utilization of advanced penetration testing tools becomes essential in safeguarding IT systems and networks against potential vulnerabilities and cyberattacks. To provide an idea figure 1 illustrates the perspective on the matter.

Figure 1. The Kali Linux Distribution Package

B. Question Query: Is Kali Linux Suitable for You?

Kali Linux is a specialized operating system designed for professional penetration testers and security specialists. It offers unique features tailored to meet the demands of ethical hacking activities, but its suitability for individual users varies depending on their level of expertise and intended use. Key Features of Kali Linux: Kali Linux distinguishes itself with features such as default network service disabling, a custom Linux kernel patched for wireless injection, and a minimal set of trusted repositories.

These features ensure a secure environment for conducting penetration testing and security auditing tasks.

Considerations for Users: While Kali Linux offers powerful tools for cybersecurity professionals, it may not be suitable for individuals unfamiliar with Linux or seeking a general-purpose desktop distribution. Its small development team, limited upstream repositories, and strict security measures may pose challenges for inexperienced users.

Use Case Recommendations: Kali Linux is recommended for professional penetration testers and individuals studying penetration testing with the goal of certification. However, for users new to Linux or seeking a user-friendly desktop environment, alternative distributions like Ubuntu, Mint, or Debian are more suitable.

Cautions and Warnings: Users should exercise caution when adding repositories or packages outside of the officially supported sources, as this could compromise system integrity. Additionally, misuse of security tools without proper authorization may have legal and personal consequences.

While Kali Linux offers unparalleled tools for cybersecurity professionals, its suitability for individual users depends on their level of expertise and intended use. Users should carefully consider their needs and familiarity with Linux before choosing Kali Linux as their operating system. Figure 2 provides the various distros available on Kali Linux.

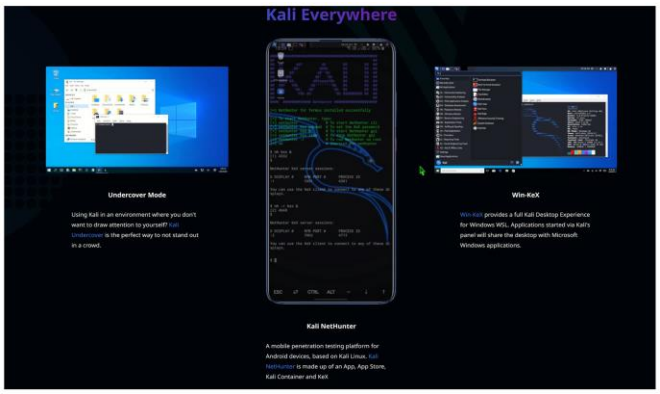


Figure 2. The Various types of Kali Distributions

V. EXPERIMENTATIONS AND TESTINGS

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and

import your prepared text file. You are now ready to style your paper.

A. Wi-Fi Penetration Testing with Vulnerability Analysis of Wi-Fi Network Encryption using Kali Linux

Kali Linux stands out as a robust open-source platform designed specifically for comprehensive penetration testing. With a diverse toolset tailored for penetration testing purposes, Kali Linux offers a comprehensive solution for assessing network security, including Wi-Fi networks.

Wi-Fi Penetration Test Process: The Wi-Fi penetration test process under Kali Linux is structured into four distinct stages: preparation, information gathering, simulated attack, and reporting. Each stage plays a crucial role in conducting a thorough assessment of Wi-Fi network vulnerabilities and ensuring the overall security of the network.

Preparation Stage: During the preparation stage, key tasks involve defining the scope of the penetration test, establishing boundaries, obtaining necessary authorization from the client, ensuring the legality of the test, planning the test execution, and evaluating the workload involved. This phase lays the foundation for the subsequent stages of the penetration test.

Information Gathering Stage: In the information gathering stage, the focus shifts towards collecting relevant data on wireless networks and associated devices within the test scope. This includes compiling information on network topology, identifying connected devices, determining network coverage, and pinpointing potential attack surfaces within the network range. Thorough information gathering is essential for identifying potential vulnerabilities and planning targeted attacks.

Simulated Attack Phase: The simulated attack phase involves verifying potential vulnerabilities identified during the information gathering stage through simulated attacks. This includes targeting Wi-Fi encryption methods, infrastructure components, and client devices. Attacks on Wi-Fi encryption methods entail analyzing encryption modes, selecting appropriate password cracking methods, and testing the strength of Wi-Fi passwords. Targeting infrastructure involves conducting penetration tests on licensed components, such as port scanning, service enumeration, and vulnerability exploitation. Client attacks are performed by establishing pseudo-APs and conducting penetration tests on client devices connected to these APs.

Reporting Phase: The final phase of the Wi-Fi penetration test is the reporting phase, where findings and recommendations are documented in a comprehensive test report. This report outlines detailed penetration testing procedures, technical methodologies employed, results of findings, and recommendations for improving security. The objective is to provide the client with actionable insights to enhance security awareness, address identified vulnerabilities, and elevate overall security posture.

Kali Linux serves as a powerful platform for conducting Wi-Fi penetration testing, offering a structured approach encompassing preparation, information gathering, simulated attacks, and reporting. By leveraging the tools and capabilities of Kali Linux, security professionals can effectively assess Wi-Fi network vulnerabilities and mitigate potential risks, thereby enhancing the overall security of the network. The vulnerability analysis of Wi-Fi network encryption is essential for assessing the security of wireless networks. It involves understanding the weaknesses inherent in different encryption methods and protocols used to secure Wi-Fi networks, such as Wi-Fi Protected Setup (WPS), Wired Equivalent Privacy (WEP), and Wi-Fi Protected Access (WPA).

WPS Encryption Vulnerability Analysis: WPS encryption simplifies the process of connecting devices to a wireless network by using a PIN code or Push Button Configuration (PBC). However, it suffers from significant security vulnerabilities, primarily due to the vulnerability of the PIN code authentication mechanism. With only 100 million possible combinations, the PIN code is susceptible to brute force attacks, making it relatively easy to crack. Many wireless network cards no longer support WPS due to its security flaws, but older access points may still have it enabled by default, making it a common target for penetration testing.

WEP Encryption Vulnerability Analysis: WEP encryption, based on the RC4 stream encryption technology, is known for its vulnerabilities. It uses XOR encryption, which becomes susceptible to attacks when the plaintext and ciphertext are known. WEP encryption relies on short initialization vectors that are transmitted in plaintext, making them easily accessible and reusable. By capturing enough data packets and performing XOR operations, attackers can analyze and calculate the WEP cipher, ultimately leading to its compromise.

WPA Encryption Vulnerability Analysis: WPA encryption, an improvement over WEP, is widely used in wireless networks. However, it is not without vulnerabilities. WPA employs the TKIP algorithm, while WPA2 uses the stronger AES-CCMP algorithm. Despite these improvements, WPA/WPA2 encryption is vulnerable to attacks involving the capture of four handshake packets and subsequent dictionary attacks. By capturing and analyzing these handshake packets, attackers can attempt to crack the WPA/WPA2 encryption and obtain the network password.

The vulnerability analysis of Wi-Fi network encryption involves identifying weaknesses in various encryption methods and protocols used to secure wireless networks.

From vulnerabilities in WPS encryption's PIN code authentication mechanism to the susceptibility of WEP encryption to XOR-based attacks and the dictionary attacks targeting WPA/WPA2 encryption, understanding these vulnerabilities is crucial for effectively securing Wi-Fi networks against potential threats.

To better understand figure 3 provides the diagram on the experimentations testing performed. Along with that, figure 4 provides the graphical representations on the experimental processing performed and initiated for the testing.

After the conduction of the experimental testing all the results and findings are represented within figure 5 for a better understanding on the perspective of the matter.

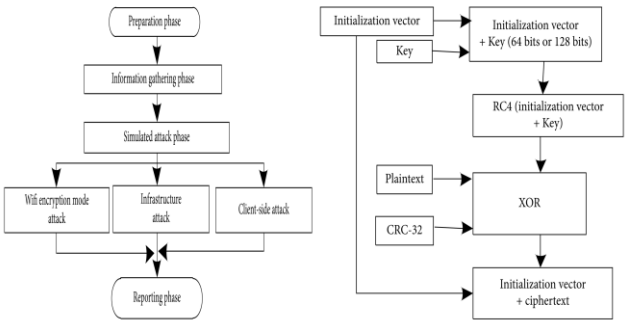


Figure 3. The Wi-Fi penetration testing process with its associated WEP encryption principle

B. Experimentations within the Wi-Fi Penetration Testing with Kali Linux deployments

Implementing a comprehensive security defense strategy involves a combination of proactive measures like vulnerability assessments and intrusion detection, reactive measures like virus protection, and continuous monitoring and analysis through auditing, accounting, and logging mechanisms. In the realm of Wi-Fi penetration testing, Kali Linux serves as a powerful toolset for assessing network vulnerabilities and ensuring robust security measures.

Through a series of experiments conducted in a controlled environment, the effectiveness of Wi-Fi penetration testing with Kali Linux was evaluated, yielding insightful results across various stages of the testing process.

Experimental Environment: The experimental setup comprises a wireless router, physical host, virtual attack machine (Kali Linux VM), USB wireless network card, pseudo-AP created by a wireless card, and two mobile devices. Leveraging VMware virtualization technology, the Kali Linux attacker operates on a host with high-performance specifications, ensuring optimal testing conditions.

Information Gathering: During the information gathering stage, tools like airudump-ng and Kismet in Kali Linux were employed to collect data on nearby wireless APs and connected clients. This information, including physical addresses, encryption modes, and MAC addresses, facilitated targeted attacks and offline analysis. Kismet scans provided crucial insights into

network topology, aiding in the identification of potential attack vectors within the network range.

Password Cracking: In password cracking experiments, the creation of a robust dictionary using Crunch facilitated efficient brute force attacks. The vulnerability of WPS encryption was exploited, showcasing rapid password cracking with known PIN codes using tools like Reaver. Additionally, experiments demonstrated successful cracking of passwords under various encryption modes, including WEP and WPA-PSK/WPA2-PSK, using Aircrack-ng with captured handshake packets.

Pseudo-AP Phishing Client Penetration Test: Utilizing a pseudo-AP hotspot created with Easy-Creds, phishing attacks were simulated to assess client vulnerabilities. Through tools like Airbase-NG, DMESG, SSLStrip, and URL Snarf, critical client information such as MAC addresses, IP addresses, system details, and open connection URLs were captured and analyzed. Wireshark facilitated comprehensive packet capture and analysis to extract desired information from the target AP.

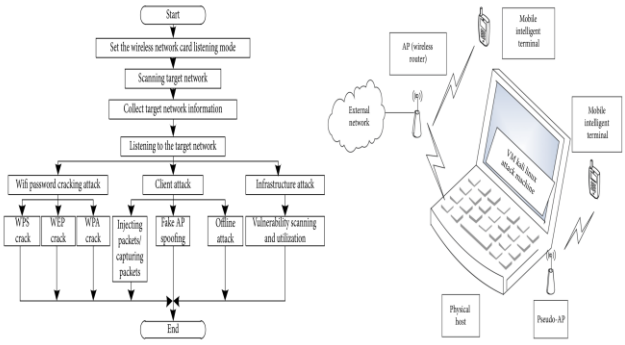


Figure 4. Kali Linux Wi-Fi penetration technical methods with proposed Experimental network topology

VI. RESULTS, FINDINGS, DISCUSSIONS

The experiments conducted in Wi-Fi penetration testing with Kali Linux demonstrated the platform's efficacy in assessing network vulnerabilities across various stages of the testing process. From information gathering to password cracking and client penetration tests, Kali Linux provided comprehensive toolsets and functionalities, enabling security professionals to conduct thorough assessments and mitigate potential risks effectively. Overall, the experimentations underscored Kali Linux's significance as a valuable tool for Wi-Fi penetration testing and enhancing network security measures.

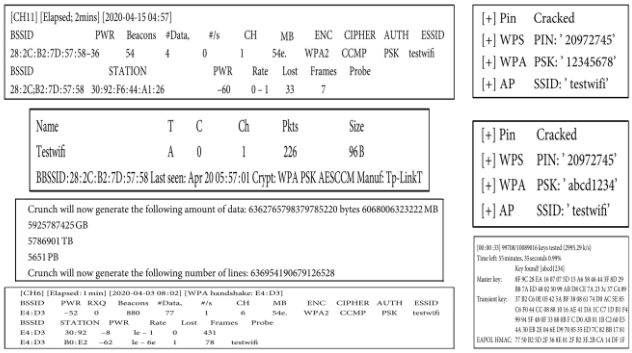


Figure 5. An overview of the results and findings of the proposed experimentations

Kali Linux stands out as a comprehensive platform for conducting web penetration testing, with its array of specialized tools, streamlined features, and emphasis on real-world security challenges. As organizations continue to grapple with evolving cybersecurity threats, the exhaustive capabilities of Kali Linux play a pivotal role in fortifying network defenses and ensuring the resilience of digital systems against malicious intrusions.

Wi-Fi networks are susceptible to vulnerabilities due to the wireless transmission of signals and inherent flaws in protocols. Based on the findings from the Wi-Fi penetration testing experiments conducted using Kali Linux, several suggestions are gathered to enhance the Wi-Fi security. These suggestions include modifying default router administrator passwords to complex ones, turning off vulnerable features like QSS, adopting more secure authentication encryption methods such as WPA2 + AES, implementing MAC address filtering, disabling SSID broadcasting, and turning off automatic WLAN connection.

Additionally, it's advised to restrict open network connections, set lengthy and complex Wi-Fi passwords, and enhance overall awareness of security measures to strengthen network supervision and intrusion detection capabilities.

This experimentation deployment provides a comprehensive analysis of Wi-Fi network vulnerabilities and common encryption methods, presenting a detailed penetration test flow and technical methods using Kali Linux for wireless networks.

The various stages and techniques involved in Wi-Fi penetration testing with Kali Linux, including listening, scanning, grabbing, password cracking, offline attacks, and pseudo-AP spoofing, are elaborated upon. Through simulated experiments, the efficacy of these Wi-Fi penetration testing methods using Kali Linux is demonstrated, showcasing their effectiveness in evaluating and improving Wi-Fi network security. The results underscore the importance of proactive security measures in identifying and addressing hidden security risks within Wi-Fi networks. Wi-Fi penetration testing with Kali Linux offers a proactive approach to security evaluation, transforming passive defense strategies into active defense mechanisms, ultimately contributing to the overall enhancement of Wi-Fi network security.

VII. CONCLUSIONS

Kali Linux, a renowned Linux distribution, is equipped with built-in tools specifically designed to facilitate web penetration testing. The development of Kali Linux has been centered around addressing security issues comprehensively, with a particular emphasis on enhancing capabilities for conducting penetration testing. This distribution has undergone continuous development and enhancements to bolster its ability to gather information about the target environment, a critical aspect of penetration testing.

Understanding the intricacies of the target system is paramount in penetration testing, and Kali Linux has integrated this fundamental process into its development framework.

Numerous security features have been incorporated into Kali Linux, underscoring its robustness as a platform for conducting comprehensive penetration tests. Given the exhaustive nature of penetration testing and the emphasis on understanding the target environment, it can be inferred that the research question regarding the effectiveness of Kali Linux in web penetration testing has been adequately addressed.

Kali Linux emerges as a powerful and reliable tool for conducting web penetration testing, offering a suite of built-in tools and security features tailored to the needs of security professionals and organizations. Its continuous development efforts and focus on addressing security issues underscore its commitment to providing a comprehensive platform for conducting rigorous penetration tests. Through its integration of essential processes such as information gathering and understanding the target environment, Kali Linux demonstrates its effectiveness in enabling security practitioners to identify and address vulnerabilities effectively.

REFERENCES

[1] "Official Kali Linux Releases". Retrieved 2020-08-29.

[2] "Kali Linux 2023.4 Release (Cloud ARM64, Vagrant Hyper-V & Raspberry Pi 5)". 5 December 2023. Retrieved 5 December 2023.

[3] Nestor, Marius (26 November 2019). "Kali Linux Ethical Hacking OS Switches to Xfce Desktop, Gets New Look and Feel". softpedia. Retrieved 2019-11-29.

[4] "Kali Linux 1.0 review". LinuxBSDos.com. 2013-03-14. Retrieved 2019-11-26.

[5] Simionato, Lorenzo (2007-04-24). "Review: BackTrack 2 security live CD". Linux.com. Retrieved 2019-04-10.

[6] Barr, Joe (13 June 2008). "Test your environment's security with BackTrack". Linux.com. Retrieved 2019-04-10.

[7] "BackTrack 4 - Hacking galore". Dedoimedo.com. 2009-05-15. Retrieved 2019-04-10.

[8] "BackTrack 5 R3 review". LinuxBSDos.com. 2012-08-17. Retrieved 2019-04-10.

[9] Watson, J.A. (2014-05-28). "Hands-on with Kali Linux 1.0.7". ZDNet.com. Retrieved 2019-04-10.

[10] "Kali Linux 1.0.7 review". LinuxBSDos.com. 2014-05-30. Retrieved 2019-04-10.

[11] "Kali Linux review". Dedoimedo.com. 2014-12-15. Retrieved 2019-04-10.

[12] Watson, J.A. (2016-01-22). "Hands-on with Kali Linux Rolling". ZDNet.com. Retrieved 2019-04-10.

[13] Smith, Jesse (2016-04-25). "Kali Linux 2016.1". DistroWatch Weekly. No. 658. Retrieved 2019-04-10.

[14] "Kali's Relationship with Debian". Kali Linux. 2013-03-11. Retrieved 2019-04-10.

[15] "Kali Linux Penetration Testing Tools". tools. kali.org. Retrieved 2019-04-10.

[16] "Kali Linux Metapackages". www.kali.org. 26 February 2014. Retrieved 2019-12-22.

[17] "Kali Linux arrives as enterprise-ready version of BackTrack - The H Open: News and Features". www.h-online.com. Retrieved 2019-12-22.

[18] "Mr. Robot and Kali Linux". 29 December 2020./

[19] Leroux, Sylvain (3 May 2017). "The Kali Linux Review You Must Read Before You Start Using it". itsfoss.com. Retrieved 2020-04-15.

[20] Grauer, Yael (2015-08-26). "A Peek Inside Mr. Robot's Toolbox". Wired. ISSN:1059-1028. Retrieved 2020-04-15.

[21] "Exploring the Hacker Tools of Mr Robot". HackerTarget.com. 2015-08-21. Retrieved 2020-04-15.

[22] "Kali Linux 2020.4 Release". www.kali.org. 18 November 2020. Retrieved 2021-01-12.

[23] "Kali Linux Hard Disk Install". Kali Linux Official Documentation. Archived from the original on 2020-05-19. Retrieved 2020-05-28.

[24] Pauli, Darren (2013-03-13). "BackTrack successor Kali Linux launched". SC Magazine. Retrieved 2019-04-10.

[25] Orin, Andy (2014-12-03). "Behind the App: The Story of Kali Linux". Lifehacker. Retrieved 2019-04-10. Mati Aharoni: One of our goals with Kali is to provide images of the operating system for all sorts of exotic hardware—mainly ARM based. This includes everything from Raspberry Pi's to tablets, to Android TV devices, with each piece of hardware having some unique property.

[26] "04. Kali Linux on ARM". Retrieved 2019-09-04.

[27] muts (2018-03-05). "Kali Linux in the Windows App Store". Kali Linux. Retrieved 2019-04-10.

[28] "Kali Linux NetHunter for Nexus and OnePlus". Retrieved 2019-04-10.

[29] "Kali Linux Forensics Mode". Retrieved 2019-04-10.

[30] Gray, Lerma (12 February 2021). "11 Best Linux Distros for Hacking And Penetration Testing in 2021 – dev.Count". Retrieved 2022-05-02.

[31] "Kali's Default Credentials | Kali Linux Documentation". Kali Linux. Retrieved 2022-05-02.

[32] "Burp Suite - Application Security Testing Software". portswigger.net. Retrieved 2023-09-29.

[33] "BeEF - The Browser Exploitation Framework Project". beefproject.com. Retrieved 2023-09-29.

[34] "cisco-global-exploiter | Kali Linux Tools". Kali Linux. Retrieved 2023-09-29.

[35] "sqlmap: automatic SQL injection and database takeover tool". sqlmap.org. Retrieved 2023-09-29.

[36] "WPScan: WordPress Security Scanner". wpscan.com. Retrieved 2023-09-29.

[37] Reverse Engineer's Toolkit, Mente Binária, 2023-09-28, retrieved 2023-09-29

[38] dev-gsniper (2023-09-27), Reverse-Engineering-toolkit, retrieved 2023-09-29.

[39] "Vulnerable By Design ~ VulnHub". www.vulnhub.com. Retrieved 2023-09-29.

[40] Hertzog, Raphael; O'Gorman, Jim; Aharoni, Mati (2017-06-05). Kali Linux Revealed: Mastering the Penetration Testing Distribution. Offsec Press. ISBN: 978-0-9976156-0-9.

[41] Kali Linux Revealed (PDF). Archived from the original (PDF) on 2021-01-02. Retrieved 2020-03-17.

[42] C. Balaji, B. Ramadoss, and N. Yasuyuki, "Secure information transmission framework in wireless body area networks," Journal of Applied Security Research, vol. 15, no. 2, pp. 279–287, 2020.

[43] Akhtar, Z.(2024).Securing Operating Systems (OS): A Comprehensive Approach to Security with Best Practices and Techniques. International Journal of Advanced Network, Monitoring and Controls, 9(1) 100-111. https://doi.org/10.2478/ijanmc-2024-0010.

[44] Bin Akhtar, Z. (2022). A Revolutionary Gaming Style in Motion. IntechOpen. doi: 10.5772/intechopen.100551.

Research on Machine Learning Program Generation Algorithm Based on AORBCO

Shiqian Wang
School of Computer Science and Engineering
Xi'an Technology University
Xi'an, China
E-mail:1178208937@qq.com

Songhan Wang
School of Computer Science and Engineering
Xi'an Technology University
Xi'an, China
E-mail:598226253@qq.com

Wuqi Gao
School of Computer Science and Engineering
Xi'an Technology University
Xi'an, China
E-mail:gaowuqi@126.com

Abstract*—The design and development of machine learning programs require selecting appropriate data and algorithms, and coding and debugging based on specific task requirements and the programming experience of developers. However, the current knowledge structure in the field of machine learning is relatively complex, lacking systematic organization, and developers often face the problem of lack of experience when choosing algorithms and designing programs, resulting in a long development cycle and easy errors in machine learning programs. In response to the above issues, this article proposes and designs a machine learning program generation algorithm based on the AORBCO model. The program generation ability includes two sub abilities: algorithm decision-making ability and code generation ability. AD-EKG has been designed for algorithmic decision-making ability, allowing Ego to select appropriate machine learning algorithms based on datasets in massive data. This algorithm combines the characteristics of the AORBCO model's domain knowledge base, knowledge graph based recommendation algorithm, and collaborative filtering algorithm. By calculating the descriptive and structural information between the dataset and algorithm, the interaction probability between the dataset and algorithm is obtained, allowing Ego to make algorithmic decisions interaction probability based. Results of the experiment have shown that the AD-EKG algorithm can fully utilize structural and descriptive information to improve the accuracy of Ego algorithm decision-making. CodeT5-EKG has been designed for code generation capability, allowing Ego to automatically generate machine learning program code. This algorithm combines the CodeT5 generative model with the domain knowledge base of the AORBCO model, by adding auxiliary information extracted using DPR technology to the code generation task, and performing diversified fusion operations to improve code generation quality. The CodeT5-EKG algorithm combines the creativity and efficiency of generative models and DPR technology, and is an algorithm that can improve the quality of generated code while also having the advantages of generative models. The experiments have proved that the code generated by this algorithm has better quality compared to other generative models with the same number of parameters.

Keywords-Program Generation; Recommendation Algorithms; AORBCO Model; Machine Learning

I. INTRODUCTION

The so-called automatic generation of machine learning programs is the process of providing a dataset by the user and letting the computer automatically generate machine learning algorithm code that meets the user's requirements based on the dataset. The challenges of time pressure and complexity in machine learning program development are addressed through automation and intelligence. However, due to the rapid growth of its associated data, machine learning technology, one of the most central techniques of artificial intelligence, is now facing a serious information overload problem [1].

DOI: 10.2478/ijanmc-2024-0013

For the past few years, researchers have attempted to use machine learning, deep learning, and other techniques to allow machines to automatically generate code. Beltramelli proposed a neural network model pix2code [2] which automatically reverse engineers the user interface and generates code based on GUI screenshots. Ahmad et al. proposed PLBART [3] pre-training model for program generation task. Wang et al. [4] raised a CodeT5 pre training model for code generation by incorporating tasks related to program identifiers during the pre-training process based on the principles of the T5 model [5]. OpenAI has released the ChatGPT model, which performs well in handling basic programming questions, answering technical questions, and generating basic code snippets. However, when dealing with complex and specific programming tasks, especially those that require a lot of details, ChatGPT's performance may have certain limitations.

In this context, the AORBCO model (Agent-Object-Relationship Model Based on Consciousness-Only) [6], as an intelligence model derived from the results of research on human intelligence consisting of the relevant theories of Consciousness-Only [7], and its idea of modeling knowledge can effectively alleviate the information overload problem. This model adopts the concept of "one person, one world" and proposes a reasonable abstraction of the objective world centered on Ego (self). From the perspective of human intelligence, thinking, and application, combining recommendation algorithms with code generation technology can leverage machine learning algorithms to improve user efficiency, thereby promoting the popularization and widespread application of machine learning algorithms.

Based on analyses and summaries of existing research, this paper proposes a machine learning program generation algorithm based on the AORBCO model. The program generation ability includes two sub abilities: algorithm decision-making ability and code generation ability. AD-EKG has been designed for algorithmic decision-making ability, allowing Ego to select appropriate machine learning algorithms based on datasets in massive amounts of data. Experimental results have shown that the AD-EKG algorithm can fully utilize structural and descriptive information to improve the accuracy of Ego algorithm decision-making. CodeT5-EKG has been designed for code generation capability, allowing Ego to automatically generate machine learning program code. The results of the experiment show that this algorithm generates higher quality code compared to other generative models with the same number of parameters.

II. Overall Design of Ego Program Generation Capability

The program generation capability of Ego refers to its ability to understand task requirements from user provided natural language descriptions and automatically complete the machine learning program generation process. The program generation ability includes two sub abilities: algorithm decision-making ability and code generation ability. The overall framework diagram of program generation capability is shown in Fig.1.

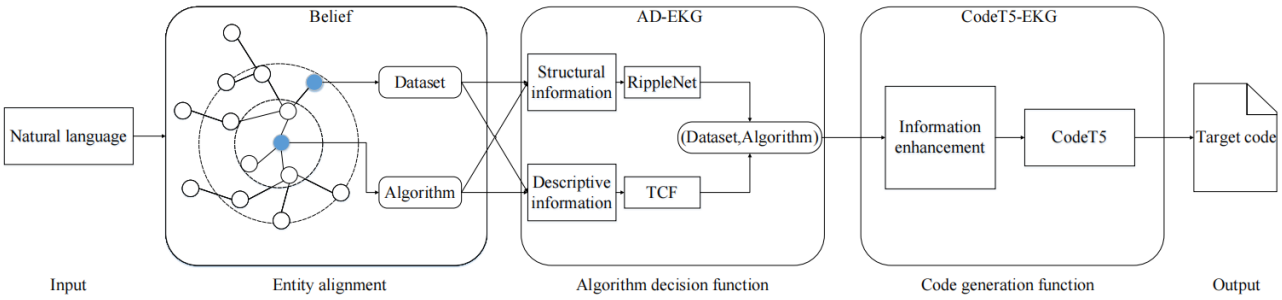


Figure 1. Overall framework diagram of program generation capability

Firstly, Ego receives natural language text sent by user Ego, which describes the characteristics of the dataset and the type of task. Subsequently, Ego utilized the knowledge alignment feature in the AORBCO model to align the knowledge of dataset objects in the domain knowledge base. At this stage, Ego will consider the dataset object described by user Ego to be the most similar object in the domain knowledge base after knowledge alignment. Then Ego will make capability calls for the object, including algorithm decision-making ability and code generation ability. The algorithmic decision-making ability will make decisions on the algorithmic object of the dataset, ultimately enabling Ego to find the algorithmic object that best matches the dataset object, forming a dataset algorithm binary. The subsequent code generation capability will be applied to the dataset algorithm binary for information augmentation and code generation, ultimately generating executable code. This end-to-end process enables Ego to understand task requirements from user provided descriptions and automatically complete the generation process of machine learning programs. Below, specific designs will be made for Ego's algorithm decision-making ability and code generation ability.

III. Design of AORBCO-ML Program Generation Algorithm

A. Design of Decision Ability for AD-EKG Algorithm

The algorithmic decision-making ability of Ego is its ability to select appropriate algorithms based on datasets in the knowledge graph of machine learning. This article designs an Algorithm Decision Based on Enhanced Knowledge Graph (AD-EKG) based on the characteristics of knowledge types in the domain knowledge graph of the AORBCO model. AD-EKG will combine structural information and descriptive information between objects to complete algorithm decision-making tasks, as shown in Fig.2.

AD-EKG includes a structural message calculation module and a descriptive information calculation module. The structural message calculation module is used to aggregate information from multi-level neighbors and extract structural information between objects. The descriptive information calculation module is used to extract linear and nonlinear relationships between object descriptive information. Below are two key components of AD-EKG.

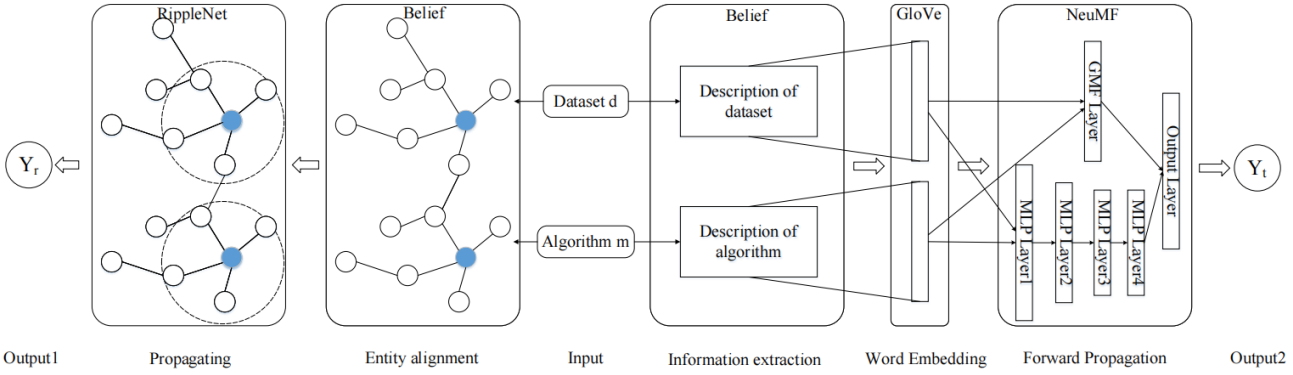


Figure 2. AD-EKG Overall Framework

1) Structural information calculation module

This article uses the RippleNet algorithm to implement the structural information calculation module of Ego. The input to the algorithm is a user item pair, the output is the probability of a user clicking on an item. This article views the dataset object as a user in a machine learning knowledge graph, an algorithm object as an item, and the relationship between the dataset object and the algorithm object as a historical interaction record. On this basis, use RippleNet to implement the algorithmic decision-making ability of Ego. The calculation process of RippleNet is shown in Fig.3.

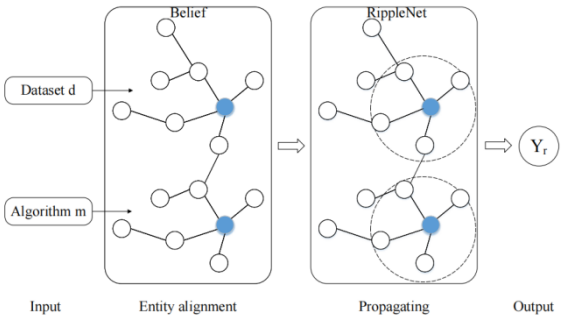


Figure 3. RippleNet calculation process

RippleNet calculates the interaction probability between two entities by searching for potential path information between user history click records and recommended items. The algorithm is executed as follows:

a) Model input: The model accepts users and items as inputs. User is represented by d, used to obtain the user's historical click history; The item is represented by m, representing the item to be predicted and clicked on.

b) Building a user seed set: The user seed set contains knowledge information that the user has operated on in the past. If it is a member of the seed set, the probability of clicking on the item during training is recorded as 1 (positive example), otherwise it is recorded as 0 (negative example).

c) First knowledge dissemination: obtain the set S_d^1 of first-order (hop-1) ripples of d, denoted by (h,r,t). To obtain valid recommendation information, there are certain invalid relationships between objects that need to be filtered.

d) Object embedding and similarity computation: normalized similarity is computed from the inner product of embedding vectors. The combination (h_i,r_i) of the head node h_i and the relation r_i in the first-level corrugated set S_d^1 given d is matrix-multiplied with the model input term m, and then the probability p_i of association of m with each (h_i,r_i) is obtained separately through the Softmax layer. Next, the model input term m is mapped into the embedding space, i.e., $m \in R^c$, and the dimension of the object embedding is denoted by c. The concrete representation of the association p_i probability is shown in equation(1).

$$p_i = \text{Softmax}(m^T R_i h_i) = \frac{\exp(m^T R_i h_i)}{\sum_{(h,r,t)\in S_d^1} \exp(m^T R_i h_i)} \quad (1)$$

At this point, the correlation probability p_i can be seen as the similarity between m and object h_i in the relationship space $R_i \in R^{c\times c}$, i.e., the degree to which the user's interests are preferred in the direction of the relationship in that r_i. Determine the correlation between m and each (h_i,r_i) based on the (h_i,r_i) in the ripple set S_d^k in the user's knowledge graph.

e) Calculate weighted average: In the previous step, the correlation probability p_i of m with respect to each (h_i,r_i) in the first level ripple set S_d^1 was obtained. p_i was multiplied by t_i in the first level ripple set S_d^1 and then summed to obtain the first order response o_d^1 of the current user d to m, as shown in equation(2).

$$o_d^1 = \sum_{(h,r,t)\in S_d^1} p_i t_i \quad (2)$$

Through the above process, the response of user hop-1 ripple set S_d^1 to m can be obtained. The process from step b to step e can be referred to as preference propagation.

f) Multiple knowledge propagation: The above steps are the first propagation of user's historical click records in the knowledge graph. To better mine knowledge, the first-order response o_d^1 obtained in step e is replaced by the embedded representation of m, and preference propagation continues. When the number of propagation is set to 3, the values of o_d^2 and o_d^3 can be calculated sequentially. Finally, the user's embedding representation is obtained by summing up the responses of each stage of d, as shown in equation(3).

$$d = o_d^1 + o_d^2 + ... + o_d^H \quad (3)$$

g) Predictive value calculation: The user embed representation and the item embed representation are inner products, and the predicted value Y_r is obtained through the Sigmoid function, which means the probability of user d clicking on item m, as shown in equation(4) Where σ represents the Sigmoid activation function.

$$Y_r = \sigma(d^T m) \quad (4)$$

2) Descriptive information calculation module

For the implementation of the descriptive information computing module, this paper proposes the TCF (Text-based Collaborative Filtering) algorithm that makes improvements to the NCF (Neural Collaborative Filtering) algorithm [8], which uses the NeuMF to jointly train the GMF and MLP, and utilizes text data to achieve recommendations. For joint training and utilizes text data to achieve recommendations. The algorithm takes descriptive information about the user-item as input and the output is the probability of the user clicking on the item.

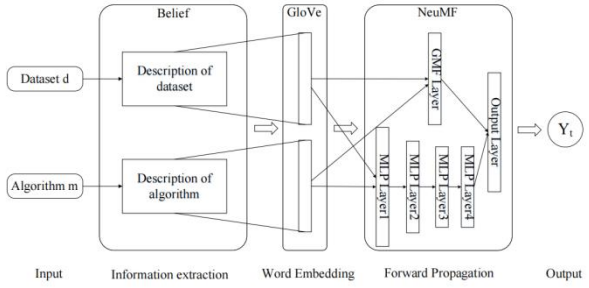


Figure 4. TCF Calculation Process

The overall structure of TCF is shown in Fig.4 above. Specifically, using t to denote the initial input text and n to denote that there are n words in the initial input text, then the descriptive text corresponding to the object can be denoted by $t = w_{1:n} = [w_1, w_2, ..., w_n]$. In this paper, we use GloVe [9] to initialize the embedding representation of each word w_i and obtain the sentence representation s by accumulating the representations of each word.

After text embedding, this paper obtains the vector forms s_d and s_m of the dataset and the descriptive text of the algorithm. In order to dig deeper into the interaction information between the two feature vectors, this paper uses the GMF and MLP layers to analyze the linear and nonlinear correlations of the dataset and the algorithm descriptive features, and then fuses these two types of information using the NeuMF layer to calculate the interaction probability of the dataset and the algorithm.

The linear interaction between the dataset and the algorithm description features can be obtained through the GMF layer as shown in equation (5) and equation (6) below:

$$\varphi_1 = s_d \odot s_m \quad (5)$$

$$y_1 = \sigma(G^T \varphi_1) \quad (6)$$

Here, $\odot$ denotes the product of elements. G^T is the weight matrix that can be obtained through learning. σ Denotes the Sigmoid activation function. This step can be interpreted as a special kind of matrix decomposition and has a higher expressive power than its original form.

While obtaining the linear interaction describing the features, this paper obtains the nonlinear interaction relationship between the two through the MLP layer. Specifically, this article connects two feature vectors and captures nonlinear interactions between features through a multi-layer fully connected network. The equation (7, 8 and 9) are as follows:

$$h_0 = s_d \,||\, s_m \quad (7)$$

$$\varphi_{nl} = h_n = \sigma(W_n^T h_{n\text{-}1} + b_n) \quad (8)$$

$$y_{nl} = \sigma(W^T \varphi_{nl}) \quad (9)$$

The symbol || represents the concatenation operation between feature vectors. In order to integrate linear and nonlinear interactions of text

features, this paper connects the last layer of GMF and MLP, and fuses linear φ_l and nonlinear feature φ_{nl} through NeuMF layer to better learn implicit interactions between descriptive texts and predict the final interaction probability of d and m. Its equation (10) is as follows:

$$Y_t = \sigma(W_t^T(\varphi_l \| \varphi_{nl})) \quad (10)$$

Finally, combining the two probabilities, equation (11) is as follows:

$$Y = W^T(\sigma(d^T m) + \sigma(W_t^T(\varphi_l \| \varphi_{nl}))) \quad (11)$$

Among them, $\sigma(d^T m)$ calculates the interaction probability between objects based on their structural information, $\sigma(W_t^T(\varphi_l \| \varphi_{nl}))$ calculates the interaction probability between objects based on their descriptive information, and W^T is the weight matrix.

B. Design of CodeT5-EKG code generation capability

The code generation capability of Ego is the ability to generate corresponding code based on datasets and algorithms. In order to build the code generation capability of Ego, this article uses the CodeT5+model as the basic model and integrates the domain knowledge base of the AORBCO model as auxiliary information for the generative model during code generation. A knowledge enhanced code generation algorithm (CodeT5-EKG) is constructed, which can improve the quality of generated code and has the advantages of generative modelling, as shown in Fig.5.

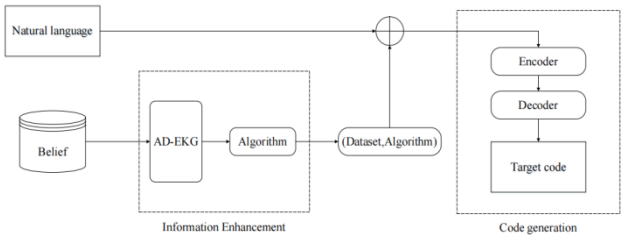


Figure 5. A Code Generation Algorithm Framework Based on Knowledge Enhancement

CodeT5-EKG consists of a code generation module and an information augmentation module. The code generation module is used to convert machine learning code templates into corresponding machine learning program code. The information enhancement module is used to extract relevant code from the domain knowledge base as auxiliary information for the code generation module, thereby improving the performance of the code generation module. Below are two key components of CodeT5-EKG.

1) Code generation module

In this paper, the CodeT5 family of models is used as the basic model for code generation, on the basis of which further innovations are made to obtain the CodeT5+ model. First, the model introduces a flexible mode selection mechanism, which enables it to run flexibly in encoder-only, decoder-only, or encoder-decoder modes according to the needs of different tasks. This design makes CodeT5+ more adaptable to different types of downstream tasks and improves the generality of the model. Second, CodeT5+ employs a multi-task pre-training strategy, including diverse tasks such as span denoising, causal language modeling (CLM), and text-code comparison learning. Such a set of pre-training tasks helps the model learn richer representations from both code and text data, allowing for better migration and adaptation in various applications.

In terms of model architecture, CodeT5+ adopts a "shallow encoder and deep decoder" architecture. The encoder and decoder get initialized by pre-training checkpoints and connected to the cross-attention layer. By freezing the deep decoder and training only the shallow encoder and the cross-attention layer, the computational efficiency is improved while the performance of the model is maintained. In addition, CodeT5+ introduces mechanisms for adjusting instructions to better align with natural language instructions. This mechanism makes the model more flexible in understanding and following natural language instructions, thus better meeting user expectations when generating code.

The CodeT5+ model was trained using the expanded CodeSearchNet pre-training dataset, which contains nine programming languages, as shown in Table 1.The model was divided into two groups for pre-training, the first group being CodeT5P-220M, CodeT5P-770M, and the second group is CodeT5P-2B, CodeT5P-6B, and CodeT5P-16B.The first group is trained from scratch according to the architecture of T5; while in the second group, the decoders of the models are initialized from the CodeGen-mono-2B, CodeGen-mono-6B, CodeGen-mono- 16B models were initialized, and the encoder was initialized from the CodeGen-mono-350M model.

TABLE I. PRE-TRAINING DATASET

Language	Sample quantity
Ruby	2,119,741
JavaScript	5,856,984
Go	1,501,673
Python	3,418,376
Java	10,851,759
PHP	4,386,876
C	4,187,467
C++	2,951,945
C#	4,119,796

In terms of model pre-training, CodeT5+ adopts two stages for pre-training: in the first stage of pre training, the model undergoes pre-training for span denoising tasks and joint training for two CLM tasks, and uses a linear decay learning rate (LR) scheduler with a maximum learning rate of 2e-4. The batch size of the denoising task is set to 2048, while the batch size of the CLM task is 512. In the second stage of pre-training, the model adopted a strategy of equal weight contrastive learning, matching, and joint optimization of two CLM losses, and underwent 10 cycles of training. Set the batch size to 256 and the learning rate to 1e-4. The maximum length of the code and text sequence is set to 420 and 128, respectively. The model uses the AdamW optimizer weights decay to 0.1. At the same time, the mixed precision training technique of ZeRO Stage 2 and FP16 using DeepSpeed [10] is utilized to accelerate the training process.

2) Information Enhancement Module

When facing problems, Ego usually consults and organizes relevant information in the knowledge base to enhance the specificity and accuracy of the answers. In recent years, some researchers have attempted to incorporate knowledge bases into generative tasks and perform diverse fusion operations to improve the efficiency of algorithms. They proposed a hybrid neural dialogue model with both response retrieval and generation capabilities. Lewis et al. [11] proposed a RAG framework for knowledge intensive NLP tasks, which utilizes the DPR(Dense Passage Retrieval) algorithm to extract information from search results, concatenates the extracted information with the original input, and finally inputs the concatenated results into a generator for processing [12]. Experimental results have shown that this method can produce more specific and accurate results.

In order to fully utilize DPR technology and its advantages in natural language processing and information retrieval, this paper adopts DPR technology to achieve code extraction of Ego in the domain knowledge base. DPR uses a text encoder to encode the questions and answers in question and answer data separately to convert the input text into a dense vector representation. By calculating the similarity between the two vectors to evaluate their correlation, it achieves fast retrieval in large-scale text datasets.

This paper constructs an information enhancer specifically designed for machine learning code generation tasks based on DPR technology. In DPR, by using question answer pairs as training data, the model can learn how to accurately match the correlation between questions and answers, thereby improving the accuracy of retrieval. Two of the encoders used pre trained CodeBERT to obtain better vector representations. Its structure is shown in Fig.6.

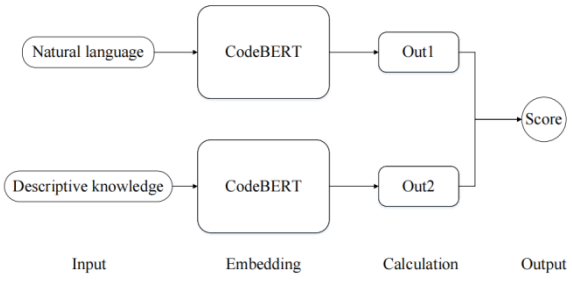


Figure 6. Diagram of DPR-based enhancer architecture

The objective likelihood function of the enhancer can be expressed as equation (12):

$$L(q_i, c_i^+, c_{i,l}^-, ..., c_{i,n-l}^-) = -\log \frac{e^{\text{sim}(q_i, c_i^+)}}{e^{\text{sim}(q_i, c_i^+)} + \sum_{j=1}^{n-1} e^{\text{sim}(q_i, c_{i,j}^-)}} \quad (12)$$

Among them, q_i represents the i th natural language input, c_i^+ refers to the correct descriptive information fragment related to natural language q_i, $c_{i,j}^-$ represents the j th descriptive information block except for c_i^+, n means the total number of samples, and sim stands for the calculation of dot product similarity. After being processed by an information enhancer, the processing method of Izacard et al [14] is referenced to splice and replace natural language inputs with descriptive chunks of information The process is shown in equation (13):

$$x' = x \oplus y_1 \oplus y_2 \oplus ... \oplus y_n \quad (13)$$

Where x denotes the original input text, y_k denotes the k th spliced and replaced descriptive information block, $\oplus$ denotes the splicing and replacing operation, and x' denotes the spliced and processed input text. The original input text for the CodeT5+ model is shown in Fig.7 below.

```
code_framework = f"""Write code for {task_area} to load and partition the {dataset_name} dataset,
define, train and evaluat {algorithm_name} models:
# Import necessary libraries
[EKG1]

# Load dataset
[EKG2]

# Split dataset
[EKG3]

# Initialize model
[EKG4]

# Train model
[EKG5]

# Evaluate model
[EKG6]

# Use the model for predictions on new data
[EKG7]
"""
```

Figure 7. original input

The above figure shows the original input text of the CodeT5+ model. Among them, task area represents the domain of the machine learning problem, dataset name represents dataset object's name, and algorithm name represents the algorithm object's name. In addition, the original input also includes module annotations related to machine learning programs, such as importing third-party libraries, loading and splitting datasets, model definitions, etc. The annotation texts of each module are connected with placeholders "[EKG]".

When performing information augmentation, the relevant fields such as domain, dataset name, algorithm name, and placeholder "[EKG]" will be replaced by the algorithm selected by the Ego algorithm's decision-making ability and the relevant information retrieved by DPR, forming the input source data after replacement processing. Partial retrieval information examples and text replacement examples are shown in Fig.8 and Fig. 9.

```
<attr name="datasetImportLib" weight="0.1">
    <![CDATA[
import os
import numpy as np
from tensorflow.keras import layers
from tensorflow.keras import models
from tensorflow.keras.preprocessing.image import ImageDataGenerator
    ]]>
</attr>
```

Figure 8. Retrieving information Example

```
Write code for image classification to load and partition the Intel Image Classification dataset,
define, train and evaluat ResNet models:
# Import necessary libraries
import os
import numpy as np
from tensorflow.keras import layers
from tensorflow.keras import models
from tensorflow.keras.preprocessing.image import ImageDataGenerator

# Load dataset
[EKG2]

# Split dataset
[EKG3]

# Initialize model
[EKG4]

# Train model
[EKG5]

# Evaluate model
[EKG6]

# Use the model for predictions on new data
[EKG7]
```

Figure 9. Text Replacement Example

When DPR fails to retrieve the corresponding text, the CodeT5+ model will directly generate code and replace the corresponding part of the placeholder "[EKG]". After DPR retrieval replacement and CodeT5+ model generation replacement, the original input will become a

complete code sequence. The final example of code generation is shown in Fig.10.

The above methods combine the advantages of information retrieval and generative models. DPR can quickly and accurately retrieve relevant code fragments, providing rich contextual information and prior knowledge. The retrieved code snippets help CodeT5+better grasp the context and generate code that matches the task requirements.

```
# Import necessary libraries
import os
import numpy as np
from tensorflow.keras import layers
from tensorflow.keras import models
from tensorflow.keras.preprocessing.image import ImageDataGenerator
import tensorflow as tf
from tensorflow.keras.applications import ResNet50

# Load dataset
# Set up data directories
train_dir = '/path/to/train'
validation_dir = '/path/to/validation'
test_dir = '/path/to/test'

# Set up data generators
train_datagen = ImageDataGenerator(rescale=1./255)
validation_datagen = ImageDataGenerator(rescale=1./255)

train_generator = train_datagen.flow_from_directory(
    train_dir,
    target_size=(img_height, img_width),
    batch_size=batch_size,
    class_mode='categorical')
```

Figure 10. Code Generation Example

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Verification of Decision Ability of AD-EKG Algorithm

This section mainly conducts validation experiments on the decision-making ability of the Ego algorithm, and the environmental information studied in the experiments is shown in Table 2.

TABLE II. EXPERIMENTAL ENVIRONMENT INFORMATION

Name	Configuration information
operating system	Windows 11
RAM	16G
Graphics card	NVIDIA GeForce RTX 3070 8G
development language	Python 3.7.8
Deep learning platform	TensorFlow 2.2.0

After studying the characteristics of data in the field of machine learning and the classification strategies of machine learning related information resources and network platforms, this article uses web scraping technology to collect data from websites such as Paperswithcode and Github. These data mainly include datasets, algorithms, and other related objects related to the field of machine learning, forming a knowledge graph based recommendation algorithm dataset. The dataset constructed in this article covers four fields (computer vision, semantic segmentation, image generation, and object detection). Including 256 machine learning datasets, 1482 machine learning algorithms, 4 machine learning tasks, 1366 academic papers, etc., a total of 5314 objects.

1) Building a dataset

After cleaning and preprocessing the crawled data, this article successfully screened 233 machine learning datasets and 1448 machine learning algorithms, which will be used for training models and analyzing user item interactions. As shown in Table 3.

TABLE III. DATASET STATISTICS

Domain knowledge graph		Dataset	
Number of objects	5262	Number of dataset objects	233
Relationship types	48	Number of algorithm objects	1448
Number of triples	14774	Number of interactions	1485
Average number of descriptive words	50.5	Sparsity	0.00440

2) Experimental plan

a) Evaluation indicators. This article models the decision-making ability of Ego algorithm as a recommendation algorithm, and in recommendation algorithms, the recommended results are usually viewed as a classification problem, that is, whether users like the items recommended by the recommendation system. Therefore, this article adopts commonly used indicators, including AUC, Precision, Recall, F1 score, and NDCG.

b) Parameter settings.

For RippleNet, The object embedding and relation embedding dimensions are configured to 16, with a maximum of 3 hops, 10 epochs, and a batch size of 32, optimized using the Adam optimizer. The learning rate and regularization coefficients are determined via grid search, and the search spaces are {10-4, 5×10-4, 10-3, 5×10-3}和{10-5, 10-4, 10-3, 10-2}；

For TCF, the dimension of the text embedding was set to 300, Multiply was used in GMF for linear computation, and 4 fully connected layers were used in MLP for nonlinear computation, and the outputs of GMF and MLP were connected by Concatenate of NeuMF.

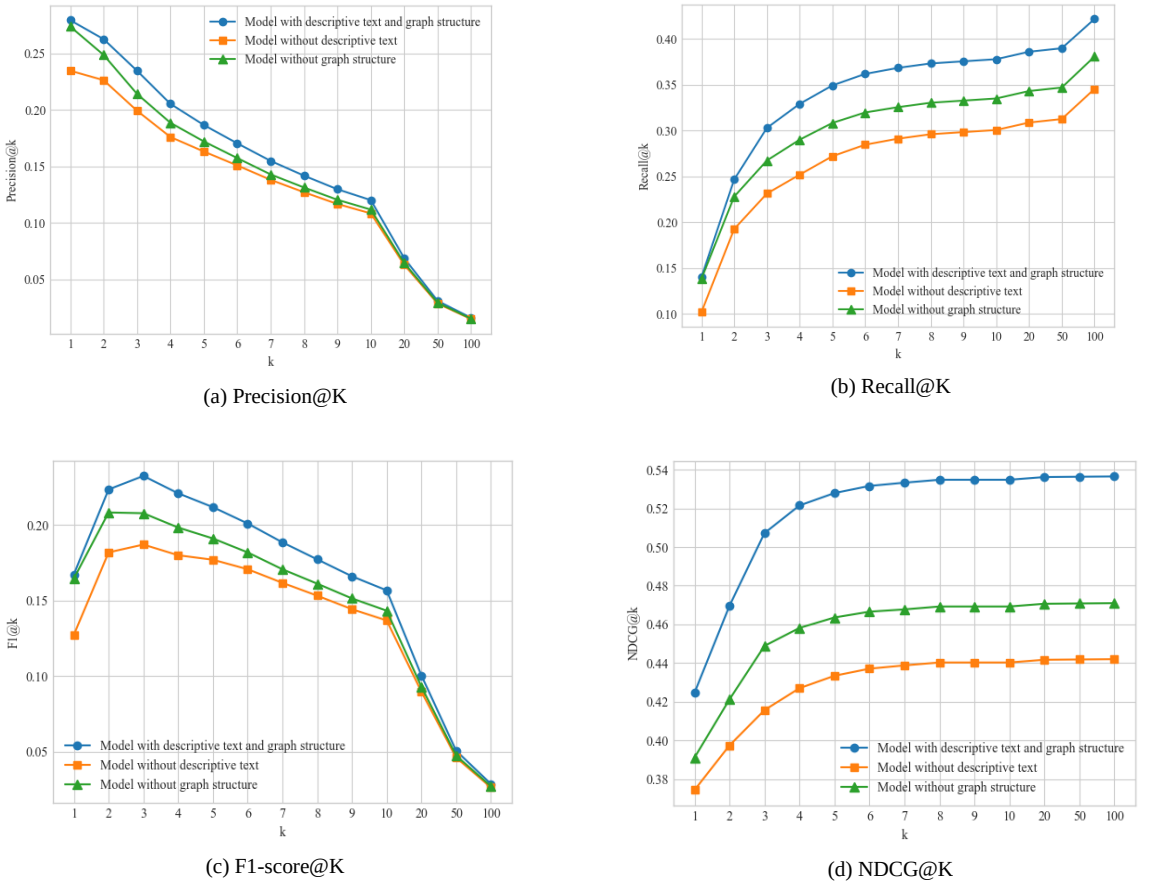
c) Comparison experiment. We compare AD-EKG with KGNN-L[14] and KGCN [15]recommendation models

d) Ablation experiment. To investigate the validity of the algorithm, i.e., whether both graph structural information and textual descriptive information are helpful for recommendation, this paper sets up the following scenarios for Top-K evaluation:

- Using only structural information (RippleNet);
- Using only descriptive information (TCF);
- Using both structural and descriptive information (AD-EKG).

3) Experimental analysis

The results of AD-EKG on the CTR prediction task and Top-K's recommendation are shown in Fig.11 and Table 4, respectively.



(a) Precision@K

(b) Recall@K

(c) F1-score@K

(d) NDCG@K

Figure 11. Top-K ablation experiments of AD-EKG under different variants

TABLE IV. CTR PREDICTION COMPARISON EXPERIMENT (%)

Model	AUC	Precision	Recall	F1-score
KGNN-LS	80.01	71.63	76.10	73.80
KGCN	71.62	62.78	64.38	63.57
RippleNet	82.55	69.43	86.91	77.19
TCF	82.16	78.24	82.81	80.46
AD-EKG	88.20	83.80	86.82	85.28

The experimental results from the experiments are presented in Table 4 and Fig.11. From the metrics Precision, Recall, F1-score and NDCG, AD-EKG outperforms the model that does not use both information in the recommendation task.The Recall value of AD-EKG increases with the length of the recommendation list, which indicates that the model is better able to capture the user's interests and needs.The NDCG metric is a ranking quality and relevance to measure the performance of ranking models in recommendation algorithms, AD-EKG is also higher than traditional models in NDCG metrics, indicating that AD-EKG can provide more relevant and higher quality recommendation results.

In summary, the AD-EKG model outperforms the single method traditional model on the CTR prediction task and Top-K recommendation. This suggests that the simultaneous use of structural and descriptive message from the knowledge graph can significantly improve the effectiveness of recommendation models.

B. Validation of CodeT5-EKG Code Generation Capabilities

This section focuses on the validation experiments of Ego code generation capability. Considering the performance requirements of the large language model, the experiments in this section are chosen to be conducted on the cloud platform. The specific environment information of the cloud platform is shown in Table 5 below.

TABLE V. CLOUD PLATFORM EXPERIMENTAL ENVIRONMENT INFORMATION

Name	Configuration information
operating system	Ubuntu 20.04.5 LTS
memory	64G
graphics card	NVIDIA A100 40GB
development language	Python 3.8
Deep learning platform	Pytorch 2.0.0

1) Dataset

In order to verify the performance of the DPR technique on the code generation task, this paper constructs a dataset of questions related to machine learning program generation.

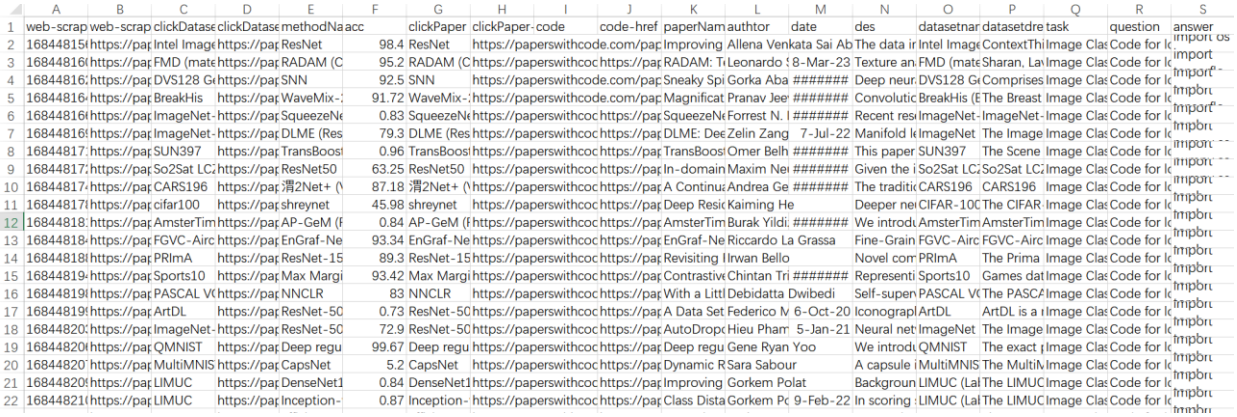


Figure 12. Example plot of a sample dataset

The dataset mainly consists of 122 question and answer data on machine learning image classification questions, as shown in Fig.12 below. The dataset of the machine learning program constructed in this paper to generate relevant questions is shown in Fig.12 above, where columns 3, 5, 14, 16, 17, and 18 of the file correspond to the dataset, algorithm, description

of the algorithm, description of the dataset, type of the task, relevant questions, and the answers of the machine learning domain, respectively, and the detailed data about the question and answer section is shown in Table 6.

TABLE VI. STATISTICAL DATA ON Q&A DATASET

Dataset	Attribute
source language	English
target language	Python
quantity	121
Average number of words in the source language	52
Maximum number of words in the source language	69
Average number of words in the target language	1365
Maximum number of words in the target language	1593

2) Evaluation Metrics

In this paper, the CodeBLEU metric [16] and ROUGE [17] metrics are used for assessing the quality of the code generated by the model. The CodeBLEU metric is a variant of the BLEU (Bilingual Evaluation Understudy) metric [18], and the BLEU metric is calculated as follows:

$$\mathrm{BLEU} = \mathrm{BP} \cdot \exp(\sum_{n=1}^{N} w_n \log P_n) \tag{14}$$

$$\mathrm{BP} = \begin{cases} 1 & if \quad c > r \\ e^{1-r/c} & if \quad c \le r \end{cases} \tag{15}$$

w_n denotes the weight of the n-tuple and p_n is the precision of the co-occurring n-tuple.BP is a penalty factor used to ensure that the scoring takes into account the length of the generated sequence and does not just focus on how accurate the generation is.

CodeBLEU is based on BLEU, additional syntactic matching as well as semantic matching score items are introduced, and the final score is weighed by a certain proportion, and its calculation formula16 is as follows:

$$\mathrm{CodeBLEU} = \alpha \cdot \mathrm{BLEU} + \beta \cdot \mathrm{BLEU}_{\mathrm{weight}} + \gamma \cdot \mathrm{Match}_{\mathrm{ast}} + \delta \cdot \mathrm{Match}_{\mathrm{df}} \tag{16}$$

In BLEU calculation, different tokens have the same weight, and different tokens have different weights in CodeBLEU calculation. In equation (16), $\mathrm{BLEU}_{\mathrm{weight}}$ is a weighted n-gram matching metric, similar to the BLEU computation; $\mathrm{Match}_{\mathrm{ast}}$ is the similarity of the abstract syntax tree, which is used to measure the syntactic information of the code; and $\mathrm{Match}_{\mathrm{df}}$ is the similarity of the semantic data flow, which takes into account the semantic similarity between the generated code and the reference code.

ROUGE metrics are mainly used to measure the degree of overlap between computer-generated code and reference code to evaluate assess the quality of automatically generated code. Commonly used evaluation metrics include ROUGE-N and ROUGE-L.

ROUGE-N mainly evaluates the code quality by calculating the number of n-grams that are the same in all the sentences in the automatically generated code and the reference code, and the proportion of them in the reference code. The detailed calculation formula17 is given below:

$$\mathrm{ROUGE\text{-}N} = \frac{\sum_{S \in \{Reference\}} \sum_{gram_N \in S} Count_{match}(gram_N)}{\sum_{S \in \{Reference\}} \sum_{gram_N \in S} Count(gram_N)} \tag{17}$$

$gram_N$ means that the length of the word is n, $Count_{match}(gram_N)$ represents the frequency with which words of length n exist both within the automatically generated code and within the reference code, as opposed to $Count(gram_N)$ which represents the frequency with which words of length n exist only within the reference code.

ROUGE-L counts the longest common substring that exists between the automatically generated code and the reference code to evaluate the overall coherence of the code, with Eqs. (18, 19, and 20) as follows:

$$R_{lcs} = \frac{LCS(X,Y)}{m} \tag{18}$$

$$P_{lcs} = \frac{LCS(X,Y)}{n} \tag{19}$$

$$F_{lcs} = \frac{(1+\beta^2)LCS(X,Y)}{R_{lcs} + \beta^2 P_{lcs}} \qquad (20)$$

Equation (19) and equation (20) denote the calculation of recall R_{lcs} and accuracy P_{lcs}, respectively. The F_{lcs} in equation (21) denotes the final calculated ROUGE-L value. Where X denotes the text of the reference code, and its length is identified by m. Y denotes the text of the model-generated code, and its length is identified by n. $LCS(X,Y)$ denotes the length of the longest common subsequence of X and Y. The parameter β is generally set to a larger number, which is used to indicate that the calculation of P_{lcs} recall holds a larger weight in the calculation of F_{lcs}.

3) Experimental analysis

Comparison of the experimental results is shown in Table7. It indicates that combining DPR technology with generative models is more effective in handling code generation problems than pure generative models when using the same parameter quantity model.

TABLE VII. COMPARATIVE EXPERIMENT (%)

label	model	Parameter quantity	CodeBLEU	ROUGE-1	ROUGE-2	ROUGE-L
1	CodeT5	770M	12.62	7.62	3.02	5.29
2	CodeT5-EKG	770M	23.93	13.52	4.62	10.02
3	CodeT5	2B	32.83	20.04	6.43	14.32
4	CodeT5-EKG	2B	47.94	24.30	9.22	17.60
5	CodeT5	6B	46.27	32.96	14.21	25.68
6	CodeT5-EKG	6B	51.12	35.58	16.11	27.54

TABLE VIII. COMPARISON WITH OTHER MODELS (%)

label	model	Parameter quantity	CodeBLEU	ROUGE-1	ROUGE-2	ROUGE-L
1	CodeT5-EKG	770M	23.93	13.52	4.62	10.02
2	CodeT5-EKG	2B	47.94	24.30	9.22	17.60
3	CodeT5-EKG	6B	51.12	35.58	16.11	27.54
4	CodeGen-Mono	2B	34.08	20.23	6.52	14.94
5	GPT-Neo	2.7B	19.82	12.57	2.79	11.28
6	InstructCodeT5	16B	43.71	25.00	9.63	21.06

Analyze the results in Table 8. As the number of model parameters increases, CodeT5-EKG shows significant improvements in both CodeBLEU and ROUGE metrics. Compared to purely generative models such as CodeGen Mono and GPT Neo, CodeT5-EKG exhibits higher code generation accuracy and consistency at smaller parameter sizes. In conclusion, the comparative results in Table 8 show that combining DPR techniques with generative models has significant advantages in English code tasks.

The retrieval + generative model constructed in this article combines the efficiency of the retrieval model with the creativity of the generative model. This combination enables the model to better control the generated content and make the generated content more reasonable. Compared to pure generative models, retrieval + generative models require fewer parameters and computational resources, making them easier to train and deploy. However, this model also has some limitations. It relies on information from prior data for retrieval, so different prior knowledge needs to be stored for different fields or tasks. Secondly, this behavior of retrieval may lead to a lack of diversity in the generated content, which may not be as flexible as pure generative models in some cases.

V. CONCLUSIONS

This article details the design and validation of a programme generation method based on the AORBCO model, including the design of algorithm decision-making ability and code generation ability. In the design of algorithm decision-making ability, this article proposes the AD-EKG algorithm. This algorithm combines the

characteristics of AORBCO model's domain knowledge graph, RippleNet, and TCF algorithm to enable Ego to intelligently select machine learning algorithms suitable for different tasks and datasets. The experimental results show that the AD-EKG algorithm can intelligently select suitable machine learning algorithms on different tasks and datasets, providing reliable decision-making basis for automatic program generation. In the design of code generation capability, this article adopts CodeT5+as the basic model for program generation. CodeT5+ is a pre-trained converter architecture that combines the information enhancer DPR to transform abstract algorithm descriptions into executable code. The experimental results show that the code generated by the CodeT5-EKG model has good accuracy and readability, providing support for the practicality of automatic generation of machine learning programs.

This article proposes a novel machine learning program automatic generation algorithm in the context of the AORBCO model, which has made important contributions to promoting research and application in the field of automated machine learning program design. In future research, the method proposed in this article can be further optimized and expanded to better adapt to the needs of different fields and tasks, providing more possibilities for the development of artificial intelligence.

REFERENCES

[1] Huang Liwei, Jiang Bitao, Lu Shouye et al. Review of recommendation systems based on Deep Learning [J]. Journal of Computers, 2018, 41(07):1619-1647.

[2] Beltramelli T. pix2code: Generating Code from a Graphical User Interface Screenshot [C]//Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems. 2018: 1-6.

[3] Ahmad W U, Chakraborty S, Ray B, et al. Unified Pre-training for Program Understanding and Generation [J]. 2021. DOI: 10.18653/v1/2021.naacl-main.211.

[4] Wang Y, Wang W, Joty S, et al. CodeT5: Identifier-aware Unified Pre-trained Encoder-Decoder Models for Code Understanding and Generation [J]. 2021.

[5] Raffel C, Shazeer N, Roberts A, et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer [J]. 2019. DOI:10.48550/arXiv.1910.10683.

[6] Feng Yanxing, Research on Program Generation in AORBCO Model [D]. Xi'an University of Technology, 2021. DOI:10.27391/dcnki.gxagu.2021.000121

[7] Xiao Liangshun, Research on Knowledge Fusion in AORBCO Modeling [D]. Xi'an University of Technology, 2023.

[8] He X, Liao L, Zhang H, et al. Neural Collaborative Filtering [J]. International World Wide Web Conferences Steering Committee, 2017. DOI:10.1145/3038912.3052569.

[9] Pennington J, Socher R, Manning C. Glove: Global Vectors for Word Representation [J]. 2014. DOI:10.3115/v1/D14-1162.

[10] Rasley J, Rajbhandari S, Ruwase O, et al. DeepSpeed: System Optimizations Enable Training Deep Learning Models with Over 100 Billion Parameters [C]//KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. ACM, 2020. DOI:10.1145/3394486.3406703.

[11] Lewis P, Perez E, Piktus A, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks [J]. 2020. DOI:10.48550/arXiv.2005.11401.

[12] Karpukhin V, Ouz B, Min S, et al. Dense Passage Retrieval for Open-Domain Question Answering [J]. 2020. DOI:10.18653/v1/2020.emnlp-main.550

[13] Izacard G, Grave E. Leveraging Passage Retrieval with Generative Models for Open Domain Question Answering [J]. 2020. DOI:10.48550/arXiv.2007.01282.

[14] Wang H, Zhang F, Zhang M, et al. Knowledge-aware Graph Neural Networks with Label Smoothness Regularization for Recommender Systems [J]. SIGKDD explorations, 2019.

[15] Li Xiang, Yang Xingyao, Yu Jiong et al. A bipartite recommendation algorithm based on knowledge graph convolutional networks [J]. Computer Science and Exploration, 2022, 16(01):176-184.

[16] Ren S, Guo D, Lu S, et al. CodeBLEU: a Method for Automatic Evaluation of Code Synthesis [J]. 2020. DOI:10.48550/arXiv.2009.10297.

[17] Barbella, Marcello and Tortora, Genoveffa, Rouge Metric Evaluation for Text Summarization Techniques. Available at SSRN: https://ssrn.com/abstract=4120317

[18] Ehud Reiter; A Structured Review of the Validity of BLEU.Computational Linguistics 2018; 44 (3): 393–401. doi:https://doi.org/10.1162/coli_a_00322

Artificial Intelligence and MCS Innovation

Thoughts Inspired by WG 7 N 387 on AIEN

Qingsong Zhang
Hong Kong Future Network Standardization Institute
Hongkong, China
E-mail: kingston_zhang1999@126.com

Bing Li
Jiangsu GreatFree Networking Technologies ltd.
Jiangsu, China
E-mail: bing.li@asu.edu

Peng Wang
Xi'an Technological University
Xi'an, China
E-mail: wangpeng@xatu.edu.cn

Hongfei Yu
Sichuan Beidou Hongpeng Technologies, ltd.
Sichuan, China
E-mail: techsup@macsec.cn

***Abstract*—Upon the request and guidance of SC 6/AG 4 Convenor Kingston ZHANG, a research group has been formed for study the potential impact of Artificial Intelligence on MCS Innovation and to evaluate how can AG 4's work contribute to the next stage of SC 6/WG 7's Future Network standardization program. This is a preliminary report.**

Keywords- AIEN; Future Network;MCS Efficiency; Terminology

I. INTRODUCTION

On 23 Feb. 2024, AG 4 posted a document N 166 which was a document from ISO/IEC JTC 1/SC 6/WG 7 N387 entitled "Harmonious collaboration between base station modulation and user applications." In the cover page of N 166, AG 4 Convenor Kingston ZHANG attached the following note:

The impact of AI on Modulation and Coding Scheme innovation has not been touched in AG 4. WG 7 N387 presents an initial study on the use of AI to achieve harmonious collaboration between base station modulation and user applications. Considering the relevance, this document is posted as an AG 4 document for review and comments. It will be listed on AG 4 10th meeting agenda.

A brief survey on the history of AG 4 indicate that despite its short history (about 17 months since its first GD document released in October 2022), a remarkable number (167) of documents have been generated and many technical issues have been discussed. However, most of the technical discussions have been on MCS gap analysis of different application fields. With this background, AG 4 N 166/WG 7 N 387 brings a new perspective to AG 4 research.

- It provides information that leads to discovery that there are interests and resources in SC 6 about AI technology which has not been considered by AG 4.
- It indicates that AI can be applied to the improvement of MCS efficiencies and is thus a relevant field for AG 4.
- The idea of AI enabled networking (AIEN) indicates AI is an enabler and not "gap", thus the current Gap Analysis focus does not include AI. It should be covered in the second phase of MCS Innovation research which is expected to start in 2024.
- AG 4 needs to include AI into its research scope.

This document hopes to start the discussion in AG 4 on the impact and applicability of AI in the development of MCS innovation.

DOI: 10.2478/ijanmc-2024-0014

II. SCOPE

This document provides a preliminary analysis on the relevance of AI with MCS Innovation focusing on the follow issues:

- An survey on AI research in SC 6.
- Analyzing relevance of AI with MCS Innovation.
- Thoughts and comments on WG 7 N 387
- Future directions for AG 4 on AI research

Exclusion: This document is a research work, which could lead to preparation for AI related MCS standard proposals. But it does not present an PWI or NP project proposal.

III. ACRONYMS

AI	Artificial Intelligence
AIEC	Artificial Intelligence Enabled Communication
AIEN	Artificial Intelligence Enabled Networking
DL	Deep Learning
EE	Energy Efficiency
FN	Future Network
IS	International Standard
MAC	Media Access Control
MCS	Modulation and Coding Scheme
MF-RIS	Multi-Functional Re-configurable Intelligent Surface
MFN	Modulation for Networking
NP	New Project
PHY	Physical
PWI	Preliminary Work Item
QoS	Quality of Service
RIS	Re-configurable Intelligent Surface
SE	Spectral Efficiency
SUE	Spectral Utilization Efficiency
TR	Technical Report
WD	Working Draft

IV. THE DEVELOPMENT OF AIEN WORK IN SC 6

A. Relevance

This document intends to analyze AI impact on MCS Innovation and assess the possibility of opening up a research field in AG 4 to study AI enabled MCS Innovation. As advisory group for SC 6, AG 4 should first understand what has been done in SC 6 in AI related research to avoid duplication/ contradictions and to make better use of existing resources.

B. AI Research in SC 6

The research of AI in SC 6 started with the concept of AIEN "Artificial Intelligence Enabled Networking" in 2019. It is a research project in SC 6 WG 7. It is also the only such project so far in SC 6. It has a long history of development and had gone through several major changes in concept.

C. Evolution of AIEN

The following table provides a highlight of the timeline of AIEN history:

TABLE I. AIEN HISTORY

Time	Event	Reference
2007	SC 6 Xi'an Resolution started Future Network project	6N 13307
2009-2010	SC 6 started TR 29181 series	6N 13602
2013-2014	ISO/IEC TR 29181 part 1-7 published	
2014	1st Draft of 29181-8 (FNQoS) presented	6N 15910
2016	ISO/IEC 21558 and 21559 (Future Network architecture and protocols) NP approved	6N 14601-14602

Time	Event	Reference
2017	ISO/IEC TR 29181-8 QoS published (problem statement and Requirement	6N 16519
2017	ISO/IEC 21558 and 21559 divided, 21558-8 and 21559-8 are assigned to FNQoS architecture and protocols	6N 16556
2018	WD for 21558-8 and 21559-8 submitted	6N 16573-16574
2018	ISO/IEC 21558 and 21559 reorganized, FNQoS become 21558-2 and 21559-2	6N 16850
2019	CD ballot for ISO/IEC 21558-2 and 21559-2	6N16958
2019	AIEN concept first proposed in SC 6	WG7 N 205-206
2020-10-10	PWI-AIEN proposed in SC 6	6N 17371
2021-09-10	SC 6 approved PWI-AIEN (PWI 5096)	6N 17599
2022	Final Text for publication of FNQoS (ISO/IEC 21558-2 and 21559-2) submitted	6N 17802 6N 17803
2023	FNQoS (ISO/IEC 21558-2 and 21559-2) published	
2023-08	In Chongqing Interim Meeting, five IS directions were proposed for AIEN	WG 6 N384-389
2023-08-07	Harmonious collaboration between base station modulation and user applications	WG 7 N 387
2024-02-05	AG 4 pays attention to WG 7's AIEN modulation discussion	AG 4 N 166

Above timeline can be categorized into four major stages

TABLE II. AIEN BACKGROUND

Stage	*Time*	*Events*
Stage 1	2007 ~2014	Future Network TR
Stage 2	2014 ~2017	FNQoS TR
Stage 3	2018 ~2023	FNQoS 21558-2 and 21559-2
Stage 4	2019 ~2024	AIEN

The first stage is from 2007 to 2014. During this seven year period, SC 6 founded Future Network project and completed the first phase of FN standardization by publish seven Technical Reports focusing on problem statement and requirements.

The second stage is from 2014 to 2017 in which ISO/IEC TR 29181-8 (FNQoS) was added and completed.

The third stage is from 2018 to 2023 in which FNQoS architecture and protocols (ISO/IEC 21558-2 and 21559-2) were developed and published.

The fourth stage is from 2019 to current which is almost parallel with the third stage and the scope has been extended.

D. Conceptual Changes of AIEN

1) FNQoS

In SC 6, AI work started with ISO/IEC TR 29181-8 focusing on QoS for Future Network. The technical concept was the use of proxy to make a more intelligent QoS for Future Network (FNProxy). The following paragraph from TR 29181-8 explains:

FNProxy is actually a software unit with a high degree of intelligence and autonomous learning ability. Adopting intelligent FNProxy technology to build the FNQoS architecture, the advantages are it can configure services for user through intelligent FNProxy server according to user QoS requirements, then provide expected services for user through proxy communication protocols, and speculate user's intention,customize and adjust services autonomously. FNProxy technology is an intermediate technology. Because FN is a converged network, it is difficult to design universal technology for each network. However, the concreteness can be shielded and the unified strategies can be implemented through the proxy technology.

From ISO/IEC TR 29181-8 to ISO/IEC 21558-2 and ISO/IEC 21559-2, Proxy based QoS is the main focus of adding intelligent functions to Future Network.

2) The Concept of AIEN

Starting from 2019, while ISO/IEC 21558-2 and ISO/IEC 21559-2 are still under development, the concept of AIEN has been proposed. It is an effort to expand the element of intelligence beyond FNProxy based FNQos and to explore more intelligent functions to support more networking protocols.

The following table offers a comparison:

3) Definition of AIEN

According to 6 N 17371, AIEN is defined as:

AIEN is the abbreviation of the process of machine networking which is given by certain intelligence by using AI learning reasoning method in the process of machine networking.

4) AIEN Current Status

The current status of AIEN study in SC 6 is reflected in SC 6 N 18186 circulated by SC 6 on 2023-12-14.

- AIEN is given PWI status and the official number is PWI 5096.
- The project is intended to produce a WD.
- There have been four updates on the project report. The last updated text of the report is given in WG 7 N 389 on 2023-08-07.
- The fourth report was accompanied by five new sub-field reports, including:
  - AIEN Network Object OID Arc Identification Registration
  - Harmonious collaboration between base station modulation and user application
  - Harmony Degree Calculation Calibration Method of AIEN
  - AIEN Based QoS for Vehicular Communications and Applications
  - Study of AIEN application in LEO satellite Mega-constellations
- These five sub-fields are treated as topics for future IS.
- Modulation was included in the five sub-fields which needs special attention from AG 4.

V. AIEN AND MODULATION

A. A New Direction

Among the five new AIEN proposals, WG 7 N 387 is related to AG 4 and needs special attention. The title of WG 7 N 387 is "Harmonious collaboration between base station modulation and user applications". here is a summary of the proposal:

- In wireless communications, spectral utility and economic considerations require new technologies.
- There are over service issue.
- A better coordination between end users and service providers can reduce over service.
- A better management of modulation can increase spectral utility efficiency.
- Solution is described as "Dynamic modulation and demodulation based on AIEN".
- The FNProxy protocol in ISO/IEC 21558-2 and 21559-2 provides some technical support.

In short, WG 7 N 387 is offering a new standard which is AIEN based modulation technology with artificial intelligent enabled functions to reduce over service and other benefits.

B. AG 4's Relevance

In many ways, WG 7 N 387 is closed related to AG 4 and requires attention.

WG 7 N 387 is the first document connecting AIEN with modulation. Its usage in the SC6 scope is universal.. Therefore, it needs help and support from AG 4 which is a dedicated group for Modulation and Coding Scheme innovation standardization research.

AG 4 has another useful resource. WG 7 N 387 is proposing modulation technology based on AIEN which is also based on published standard on Future Network. AG 4 convener Kingston ZHANG is one of the veteran expert of ISO/IEC Future Network standardization. He was an active

participant and made numerous contributions to SC 6 Future Network work and is one of the editors of ISO/IEC TR 29181-2 (Future Network Naming and Addressing).

A third advantage for AG 4 is its position. In SC 6, WG 1 is responsible for MAC and PHY layer standardization (communication) and WG 7 is specializing in Network layer standardization. Usually, modulation is considered a PHY layer technology. AG 4 is an SC 6 advisory group and therefore it is not restricted by the layer restrictions. AG 4 identifies and develops MCS technologies and can recommend standardization proposals to any SC 6 working groups.

C. Modulation for Networking

WG 7 N 387 is interesting to AG 4 also because it opens up a direction to work on "modulation for Networking, MFN".

Modulation is a PHY technology and its primary function is to facilitate communication. But when the communication equipment such as routers and switches is interconnected to form various networks, modulation also becomes the base technology for networking. The capabilities of modulation schemes will have direct impact on the performances of networks. Therefore, after a new MCS technology is identified and adopted as a SC 6 standard, apply it into networking standards can be done in parallel with works in communication.

In short, modulation technology is a work in PHY layer and primary application is to enhance communication efficiencies, but the work can be extended to cover networking. On the other hand, networking standard development may also find new requirement for enhanced Modulation services which is the case of WG 7 N 387.

D. Coordination Requirement

Since N387 is relevant to AG 4's work, some procedural rules need to be established for coordination and mutual support. AG 4 needs to establish positions on a few issues which will not only affect WG 7 N 387 but also AG 4's work.

The issues include:

- Should Artificial Intelligence be included in MCS Innovation research?
- Will Artificial Intelligence be able to help MCS Innovation to reach its objectives?
- Is the modulation technology presented in WG 7 N 387 (collaboration between network providers and end users in modulation service requirement) part of MCS Innovation?
- What will be the impact of MCS Innovation on networking?

VI. AI AND MCS INNOVATION

A. AI as an Industry Trend

AIEN is the first basic standard in SC 6 to apply AI technology to modulation. It can guide the usage of AIEN. SC 6 needs to take a broader survey of the academics and technologies to make the proposals more inclusive.

Before the emergence of AI as a popular technology, ICT sector has long history of applying intelligent operation into communication and networking operations. Typical applications include channel and frequency selections, automatic adjustment of transmitting power levels, and automatic hibernation in no use situations.

However, traditional concept of automation is not complicated. It is intelligent but not at the advanced level. The trend is moving from being intelligent to highly advanced Artificial Intelligence which can utilize deep learning and logical reasoning to perform more complicated duties. The conceptual evolution from FNProxy in Future Network to AIEN in PWI-5096 reflect this trend.

It is clear that AI can be applied to make MCS operations more efficiently. As to whether AI can help AG 4 find innovative MCS solutions to overcome the theoretic and physical berreier, it needs more assessment.

B. AI/DL Semantic Communication

Another direction is the emergence of Semantic Communication concept in the past decade which is hoping to rely on AI/DL technology to allow communication systems avoid the constraints of

classic information theories. This direction has not been analyzed in AG 4 but should be given attention because AG 4 does have the resources to evaluate the technical feasibility and future impact of this technology.

The semantics of the three elements (syntax, semantics and timing) of communication protocol can be represented by proper semantic coding, which can improve the communication effect and avoid the restriction of Shannon's law.

But one problem with semantic communication approach is that sending less information could lead to misunderstanding or misinterpretation on the receiving end. Proponent of this direction points to the growth of AI/DL (Artificial intelligence/Deep Learning) as the future tool to deal with this problem.

C. AIEC

Discussion on Semantic Communication and AIEN leads to a concept of Artificial Intelligence Enabled Communication (AIEC). The proxy mechanism in Future Network leads to the awareness that more intelligent solutions are needed for networking, the semantic communication technology also indicates the need for more intelligent solutions to improve communication services. This technical direction will continue to attract interests from the industry. The only uncertainty is about to what degree can AI impact communication performances.

There will be a race between AI and human wisdom to decide who can break the theoretical and physical limits in communication channel capacity.

D. AI Assisted RIS Modulation

RIS is the abbreviated term for Re-configurable intelligent surfaces, a new material based technology that is seen as have many benefits for enhanced communication and networking performances. The Surface material is capable of receiving, re-configuring, enhancing, and redirecting received radio frequency signals. Due to its potential to enhance the capacity and coverage of wireless networks by smartly re-configuring the wireless propagation environment, RIS is considered a promising technology for the sixth-generation (6G) of mobile communication, NFC and other wireless communication systems.

On 14 Dec. 2023, SC 6 circulated a Chinese contribution providing the first explanation on how RIS can enhance wireless communication and networking. The document introduced a new generation of RIS technology with abbreviated term of "MF-RIS". Each element of the MF-RIS is composed of an amplifier and a phase-shifter, which can be used to achieve signal amplification, reflection and refraction. The MF-RIS first amplifies the incident signals, and then divides them into reflection and refraction parts by power split circuits. Thus, all users around the MF-RIS can be served simultaneously.

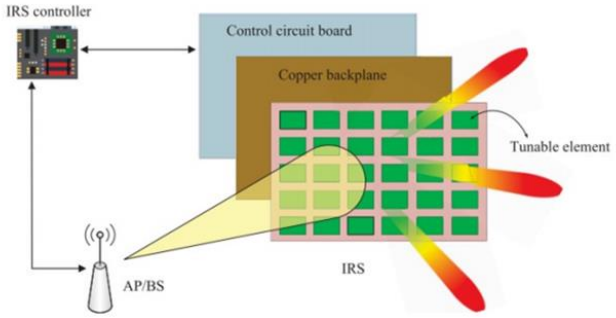


Figure 1. RIS Architecture

Clearly, RIS technology is intelligent and will most likely needs the support of Artificial Intelligence. On the other hand, from AG 4's perspective, RIS is capable of providing better and enhanced spectral utility rates and enhance user end efficiencies. Whether and how the technology can enhance MCS efficiencies or even provide an alternative direction different from AG 4's directions is unclear and is worthy more attention.

E. AI for MCS Innovation

How AI can be used to promote MCS Innovation has not been studied in AG 4, but some positions can be stated:

- Traditional automation control technologies can help communication and networking become intelligent and is useful in modulation and coding scheme services. But usually the technology is used to make micro-management of available resources.

- The MCS technology presented in WG 7 N 387 is not in the main direction of AG 4's research. But AG 4 can help AIEN program find more application scenarios for AI enabled management of MCS operations.
- The application of AI/DL in semantic communication is mainly for the purpose of restore original information but requires advance knowledge and availability of material support.
- Semantic Communication is pursuing a direction which is totally opposite of SC 6's MCS Innovation program because it put emphasis on transmitting less data, which SC 6's effort is to break through the limit and continue long term expansion of spectral and energy efficiencies.
- Due to the fact that MCS innovation shall try to break not only theoretic limits but also physical constraints such as in-channel and inter-channel interference, learning and reasoning capabilities of AI does not offer much help. Human mind will still be the main driving force for MCS Innovation.
- AG 4 should pay more attention to the development of Semantic Communication technology. And a special report on the topic should be developed. Since there is no committees in ISO/IEC involved in Semantic Communication standardization research, this work has to be carried out by AG 4.

VII. MCS Innovation for Networking

A. *Networking into Scope*

The discussion on AIEN is beneficial because it raised an overlooked perspective on MCS Innovation. In AG 4 research so far, MCS Innovation are exclusively treated as a communication technology. WG 7 N 387 revealed a situation that networking also needs the support of advanced modulation technologies. MCS Innovation thus needs to consider the gaps and requirements of MCS operations in the perspective of networking.

Before discussion the relationship between MCS innovation and Networking, the relationship between MCS Innovation and communication needs to be first explained.

B. *Defining Communication*

In SC 6, there is no definition for "communication." A simple definition of communication is that it is a process of exchanging information between two entities. In ICT sector, communication is also known as telecommunication as shown in JTC 1/SC 6 title. Communication can take many forms and mediums such as gesture, voice, music, writing, drawing, etc. Adding prefix "tele" to communication indicates that the exchange of information can take place in long distances.

Communication requires at least two entities, one to send and the other to receive information. MCS is a key technology involving both end offering modulation and demodulation, coding and decoding of messages.

Because the MCS modulation-demodulation and coding-decoding procedures take place below the MAC layer it is designed as PHY layer technology and belongs to SC 6/WG scope.

According to SC 6 Business Plan, the work of WG1 is closely related to all sorts of physical communication technologies.

SC 6/WG 1 is to deliver the standards on services and protocols in the physical and data link layers. As new types of communication technologies emerge, it extends the working boundary and delivers the standards the market requires.

C. *MCS Innovation for Communication*

MCS Innovation technical objectives have not been clearly defined, but based on previous SC 6 and AG 4 documents, the following principles should be established:

Rule 1: MCS Innovation's objective is to find a way for long term advancement and improvement for communication efficiency.

Rule 2: MCS Innovation's primary focus is on the design of wave forms that

Rule 3: MCS Innovation shall not be bound by classic information theory but shall strive to find ways to surpass the channel capacity limits and boundaries.

Rule 4: MCS Innovation shall develop a plan for continued growth of spectral efficiency for communication systems for the next two-three decades.

Rule 5: MCS Innovation shall overcome the conflict between spectral efficiency and energy efficiency as reflected in classic information theory.

Rule 6: MCS Innovation shall develop an evaluation scheme to evaluate capabilities and impact of new MCS technologies on communication systems.

D. Defining Networking

The definition of "network" can be found in ISO/IEC standards including:

- A collection of interconnected elements that provides connection services to users (ISO 1745:1975)
- The ensemble of equipment through which connections are made between terminal installations. These equipment operate in real time and do not introduce store and forward delays. (ISO/IEC 12139-1:2009)

If compared with "communication", networking can be viewed as an advanced form of communication. Networks can not only provide service to one to one communication, but also one to many communication (broadcasting) and many to many communication (meetings). Because of the more diversified application scenarios, networking is more complicated and sophisticated information system.

E. Network Standardization

In ISO/IEC, network standardization is the duty of JTC 1/SC 6 whose title includes "information exchange between systems" which is another way to refer to "networking". In SC 6, the work of networking standardization is the responsibility of WG 7 whose title is described as "Network, Transport, and Future Network".

WG 7 had developed lots of protocols on OSI network and transport by the early 1990s. Currently, WG 7 has been working on Future Network including, other emerging networking issues, to prepare the future of the networks. Considering recent market requirements and activities of other SDOs on 5G/6G network technologies, WG 7 plans to challenge this important technical area as well as future networking technologies including multi-access edge computing, communication protocols in low earth orbit inter-satellite networks and application of AI to networks to support emerging new applications and services such as Blockchain, and metaverse.

The challenge is that almost all new emerging application and services have dedicated committees to handle International Standardization. WG 7 needs to possess unique resource and advantage to attract interest in seeking support for networking standardization. In this perspective, MCS Innovation could possibly be the new driving force.

F. MCS Innovation For Networking

On this subject, there are a few issues needing clarification and understanding:

- Modulation and Coding Schemes are considered PHY layer protocols.
- In SC 6, WG 1 is responsible for MAC and PHY layer protocol standardization.
- Therefore, if it is strictly about MCS technological specifications and used in communication systems such as WLAN, Radio and PLC etc., the standardization projects should be processed in SC 6/WG 1.
- On the other hand, networks are composed of communication systems such as routers, access points, switches, servers etc., all with MCS systems embedded. Therefore, innovative MCS technologies if successfully identified, standardized and implemented, will not only enhance communication efficiencies, but also have an impact on networking capabilities and performances. Network designers and operators should be aware of the advances of MCS technology so that they can

integrate innovative MCS technology into their planning.

- AG 4 is a SC 6 advisory group and can recommend proposals to all SC 6's WG's according to their scopes.

G. MCS Innovation For Future Network

1) Natural Connection between AG 4 and WG 7

MCS Innovation has some natural linkage with the Future Network standardization activity in WG 7 and can also make significant contribution for its future development.

The MCS Innovation initiative was started by Kingston ZHANG in August 2021 when he submitted a contribution to WG 1 on the MCS challenges faced by WLAN. He is also the convenor of AG 4 on MCS Innovation since its founding on 2022-07-01.

On the other hand, Kingston ZHANG is also a senior expert member of SC 6 since 2004. Ever since the first SC 6 meeting creating a study project on Future Network, Mr. Zhang had been a key member of the Future Network standardization expert group. He is one of the editors of ISO/IEC TR 29181-2 on FN naming and addressing. He also contributed many WG 7 reports on Future Network planning and outlooks.

2) MCS Innovation's Potential Impact on Future Network

According to SC 6 Business Plan:

The current network has become an essential communication infrastructure for data transfer and social applications. Even though the current network is such an essential infrastructure, SC 6 notices many concerns, such as scalability, security, robustness, mobility, Quality of Service (QoS), reconfigurability, context awareness, etc. Since 2010, SC 6 has studied and published nine parts of ISO/IEC TR 29181 Future Networks-Problem statement and requirements: General aspects, Naming and addressing, Switching and Routing, Mobility, Security, Media transport, Service composition, Quality of service and Networking of everything. Driven by emerging application requirements, SC 6 has developed International Standards on Architecture for future networks (ISO/IEC 21558 series) and the corresponding protocols and mechanisms (ISO/IEC 21559 series).

ISO/IEC's Future Network has had significant impact on global ICT standardization and industry. Examples include the Decimal naming and addressing impact on digital identifiers standards in other SDO's and ISO/IEC TR 29181-5's impact on "Zero-Trust Security Architecture" which is becoming a hot field in information security. Furthermore, since ISO/IEC Future Network in the very beginning had set the year of 2020 as the time for commercial deployment, China has designated Future Network as one of the several "future industries" to devote more resources for its implementation and deployment.

Based on this observation, it is safe to say that ISO/IEC Future Network has reached its objective set for the past 15 years. Looking to the future, what will Future Network be after 15 years later? Is Future Network looking for increase its technological strengths and expand its capabilities? What will be the directions?

3) Future Directions for Future Network

From WG 7 N 387 and related proposals on AIEN, it gives a sign that WG 7 experts are planning to expand Future Network technological outreach capabilities, and Artificial Intelligence has been identified as one direction. In SC 6 Business Plan, WG 7 also shows intention to further build on Future Network's success to look for standardization opportunities:

Due to the new idea and expertise are required for future network project, WG 7 plans to invite experts from liaison organizations to participate in the relevant WG 7 activities, as well as contributions to be submitted from 'in-active' P-member national bodies. For the successful achievement of WG 7 work on Future Network related subjects, close collaboration with other SDOs, especially ITU-T SG11 (protocols) and SG13 (requirements and architectures) on Future Networks, IMT-2020 networks and beyond, should be strengthen. Also, considering recent market requirements and activities of other SDOs on 5G/6G network technologies, WG 7 needs to challenge this important technical area as well as

future networking technologies including multi-access edge computing, communication protocols in low earth orbit inter-satellite networks and application of AI to networks to support emerging new applications and services such as Blockchain, and metaverse.

4) Future Network 2.0

Besides the new directions such as AI, 6G, Blockchain and Metaverse, this document offers another perspective based on AG 4's works on MCS Innovation. Over 15 years ago, SC 6 experts set a few performance characteristics for Future Network including better performances in scalability, ubiquity, security, robustness, mobility and heterogeneity. Looking at the current situations and long-term objectives, a new and quite significant capability is network capacity.

Over the years, the quantity of information transmitted through networks has gone through explosive growth. This trend will continue as more application scenarios are emerging such as Big Data, AI, Cloud Computing and Metaverse which transmit huge amount of data through networks. In the meantime, network capacities have also expanded with the fast expansion of Fiber-optical transmission technology. However, due to the hard to overcome barriers of information capacity limit and physical constraints, communication systems are facing slowing down and even stagnation in MCS efficiency which forms bottle necks in networking systems.

Some technical development strategies try to bypass the barriers with ideas to reduce the volume of information as in Semantic Communications or to put the computing work load on the perimeters to reduce reliance on network (Edge Computing).

In SC 6, there are two promising technical directions. One is Future Network with its potential in clean slate network structural design and new ideas such as AIEN. The other is AG 4's MCS Innovation program which aims at breaking information channel capacity limit and continuous increase of spectral and energy efficiencies of future communication and network systems.

Therefore, based on AIEN, MCS Innovation and some other innovative technologies such as Stealthy Defense Information Security technologies, SC 6 may consider starting "Future Network 2.0" standardization initiative.

VIII. MCS Innovation Terminology

A. General

WG 7 N 387 has several definitions related to Modulation and Coding Scheme efficiencies. Some of the terms are also used in AG 4 technical research documents, but the meanings are different. AG 4 has not started offering terms and definitions in MCS Innovation standardization. The analysis below indicates that AG 4 needs to start working on definitions in order to avoid confusions.

B. Defining Spectral Efficiency

WG 7 N 387 offers this definition: "Spectral Efficiency (bps/Hz/cell, bps/Hz/Km2): Throughput provided by unit spectrum resources per cell or unit area."

The following offers some observations and brief comments regarding this definition:

- N 387 definition of Spectral Efficiency is different from the definition used in MCS innovation programs.
- MCS Innovation defines Spectral Efficiency as bps/Hz without concern for "cell" or "Km2".
- The reason of leaving out "cell" or "Km2" is that MCS Innovation focuses on MCS efficiency which is deployed in all communication systems.
- In many communication and networking scenarios, there are no performance measurement requiring factors of "cell" or "Km2". For example, for WLAN, NFC, Radio, Scatter communications etc., the communication is basically point to point and does not involve many cells or square kilometer calculations.
- The "bps/Hz/cell, bps/Hz/Km2" definitions for Spectral Efficiency is a terminology used in Mobile or cellular communications.

- According to ITU-R, "bps/Hz/cell, bps/Hz/Km2" is the format to describe "Spectral Utilization Efficiency" (SUE) which is quite different from MCS spectral efficiency (ITU-R SM1046-3).
- Furthermore, a SC 6 document (6N 17676) pointed out in 2022 that it should be further noted that Spectral Efficiency and Throughput are two different terms. They are not interchangeable terms. They represent different technical aspects even though having close relationship. Higher Spectral Efficiency can lead to higher throughput. On the other hand, use of higher frequencies can also lead to higher throughput.

More analysis on Spectral Efficiency terminology is needed.

C. Defining Energy Efficiency

Energy Efficiency is also a key vocabulary frequently used in MCS Innovation. WG 7 N 387 offers this definition: "Energy efficiency (bit/J): The number of bits that can be transmitted per joule of energy."

The following offers some observations and brief comments regarding this definition:

- N 387 definition of Energy Efficiency is different from the definition used in MCS innovation programs.
- In MCS Innovation, energy efficiency is equivalent to power efficiency.
- For calculation of MCS energy efficiency, the format is SNR/SE or SRN/bps/Hz. It shows how much energy is used to raise every bit of Spectral Efficiency.
- This format is useful to evaluate whether an innovative MCS system can offer better energy efficiency than legacy MCS systems and whether the new technology can break classical information capacity limit. The format can also be used to evaluate different innovative MCS proposals and to make long term technological road maps.

More analysis on Energy Efficiency terminology is needed.

D. Cost Efficiency

WG 7 N 387 offers a definition for Cost efficiency：

"Cost-efficiency (bit/dollar): Number of bits transmitted per Unit cost. Compared with 4G system, 5G system needs to be significantly improved in Spectral efficiency, energy efficiency and cost efficiency. Spectral efficiency needs to be increased by 5~15 times; Energy efficiency has increased by over a hundred times. Cost efficiency has increased by over a hundred times."

- This definition for cost efficiency applies to Mobile communications.
- In MCS Innovation studies, a format to calculate the cost efficiency of innovative MCS proposals has not been developed.
- It is unclear whether the bit/dollar format in WG 7 N 387 can be used in MCS Innovation.
- Probably, in MCS innovation, there is no need to set cost efficiency criteria. It can be treated as collateral benefits associated with a new generation of MCS technologies that possess extraordinary high Spectral and Energy Efficiencies.

E. AG 4 Terminology

Above discussions have revealed discrepancies about some key terms such as Spectral Efficiencies and Energy Efficiencies. These discrepancies will cause confusions and misunderstanding if they are not effectively resolved.

Since WG 7 N 387 definitions are based on well established Mobile communication standards and in special application scenarios, AG 4 does not need to request changes from other organizations. What AG 4 can and should do is to establish its own system of terminologies with clearly defined definitions for terms used in MCS Innovation standard systems. For start, AG 4 can work out definitions for:

- MCS
- Innovative MCS
- Spectral Efficiency
- Energy Efficiency

- Over-capacity Communications
- Over-limit Communications
- Spectral Efficiency Rating System
- Energy Efficiency Rating System

...

IX. Recommendation for More AG 4 Actions

- AG 4 should pay attention to the AIEN concept in WG 7 and contribute to the study of AI enabled modulation services.
- AG 4 should conduct a comparative research on Semantic Communications and develop a special report to compare different approaches on MCS technology development.
- AG 4 should continue watch on the development of AI/DL to evaluate its applicability in MCS technology development.
- AG 4 should consider the application of MCS innovative technologies to networking, with special attention to how the new technologies can make Future Network perform better.
- AG 4 should start a research project to provide terms and definitions for key MCS terminologies including the vital important Spectral Efficiency and Energy Efficiency.
- After the review and discussion at the 11th meeting on 2024-05-15, AG 4 needs to send this document (with revisions if suggested) to SC 6/WG 7 for their review and consideration at its Interim meeting 28~31 May 2024 in Hong Kong, China.

Acknowledgment

Upon request and guidance from AG 4 Convenor Kingston ZHANG, a special group of Chinese experts are organized to compile this report.

- WANG Zhongsheng, (SC 6 WG 1 and WG 7 Experts, CN)
  Xi'an Technological University
  wzhsh1681@163.com
- LI Bing
  Jiangsu GreatFree Networking Technologies ltd.
  bing.li@asu.edu
- YU Hongfei
  Sichuan Beidou Hongpeng Technologies, ltd.
  techsup@macsec.cn
- YANG Qin
  Computer Center of Kunming University of Science and Technology
  liruiking@kust.edu.cn
- WANG Yubian
  Belarusian State University of Transport
  wangirlux@mail.ru
- Shutong SHEN (DE)
  Electrical Engineering and Information Technology
  Chemnitz University of Technology, Germany
  shutong.shen@s2021.tu-chemnitz.de
- Gaoxiang LIU (CN)
  Jiangsu GreatFree Networking Technology ltd.
  Axe1991@163.com
- YANG Dong
  Institute of Psychology Bit Data (IPBD), Nanjing, China
- Ge Xuequan
  Jiangsu Ruanyi Science And Technology Stock Co.,Ltd.
- WU Zhengli
  Chongqing Yuliuming Networking Technology Ltd.
- YE Jing
  Nanjing Bofeng Communication Technology ltd.
- YIM Wai Ning
  Institute of Future Network Standardization, H.K.S.A.R
- Ouyang Quan
  Chalmers University of Technology, Sweden
- LI Yuanjie
  Sichuan Beidou Hongpeng Technology ltd.

- WANG Jun
  Institute of Psychology Bit Data (IPBD), Nanjing, China
- ZHANG Hanqing
  Nanjing University of Information Science and Technology

The team also wishes to acknowledge the WG 7 experts for their work on WG 7 N 387 that inspired this call for contributions:

- Jim Tian, CN Expert (jim.tian@jit.edu.cn)
- Zhenghua LI, CN Expert (lizh@cigit.ac.cn)
- Zhaohua LONG, CN Expert (zhaohua-long@qq.com)
- Liangxin SI, Expert (liangxin.si@gmail.com)
- Jie YANG, Expert (305994544@qq.com)

REFERENCES

[1] ISO/IEC JTC 1/SC 6/WG 7 N 387: Harmonious collaboration between base station modulation and user applications, 2023-08-07

[2] ISO/IEC TR 29181-8:2017 Information technology — Future Network — Problem statement and requirements — Part 8: Quality of Service

[3] ISO/IEC 21558-2:2023 Telecommunications and information exchange between systems — Future network architecture — Part 2: Proxy model-based quality of service

[4] ISO/IEC 21559-2:2023 Telecommunications and information exchange between systems - Future network protocols and mechanisms - Part 2: Proxy model-based quality of service

[5] ISO/IEC JTC 1/SC 6/WG 7 N 205:Artificial Intelligence Enabled Networking (AIEN) 2019-04-22

[6] ISO/IEC JTC 1/SC 6/WG 7 N 206: Introduction of Artificial Intelligent Enabled Networking (AIEN) 2019-04-22

[7] ISO/IEC JTC 1/SC 6 N 17371 : PWI on Artificial Intelligence Enabled Networking, 2020-10-30.

[8] ISO/IEC JTC 1/SC 6 N 18186 : Report on the Status of PWI 5096 (AIEN) study, 2023-12-14.

[9] ISO/IEC JTC 1/SC 6/WG 7 N 389 : The Fourth Study report on PWI-AIEN, 2023-08-07

[10] ISO/IEC JTC 1/SC 6/WG 7 N386: AIEN Network Object OID Arc Identification Registration, 2023-08-07

[11] ISO/IEC JTC 1/SC 6/WG 7 N388: Harmony Degree Calculation Calibration Method of AIEN, 2023-08-07

[12] ISO/IEC JTC 1/SC 6/WG 7 N384: AIEN Based QoS for Vehicular Communications and Applications, 2023-08-07

[13] ISO/IEC JTC 1/SC 6/WG 7 N385: Study of AIEN application in LEO satellite Mega-constellations, 2023-08-07

[14] Huiqiang Xie, Zhijin Qin, Geoffrey Ye Li, and Biing-Hwang Juang Deep Learning Enabled Semantic Communication Systems, arXiv:2006.10685v3 [eess.SP] 19 Apr 2021.

[15] Zhijin Qin, Xiaoming Tao, Jianhua Lu,and Geoffrey Ye Li, Semantic Communications: Principles and Challenges arXiv:2201.01389v1 [cs.IT] 30 Dec 2021

[16] Christina Chaccour, Walid Saad, M ´ erouane Debbah,, Zhu Han, and H. Vincent Poor : Less Data, More Knowledge: Building Next Generation Semantic Communication Networks , arXiv:2211.14343v1 [cs.AI] 25 Nov 2022

[17] Jincheng Dai, Ping Zhang, Kai Niu, Sixian Wang, Zhongwei Si, and Xiaoqi Qin : Communication Beyond Transmitting Bits: Semantics-Guided Source and Channel Coding, arXiv:2208.02481v1 [cs.IT] 4 Aug 2022

[18] ISO/IEC JTC 1/SC 6 N 18204 : SC 6 Business Plan (PERIOD COVERED: January 2024 – October 2024), 2024-01-05

[19] ISO 1745:1975 Information processing — Basic mode control procedures for data communication systems.

[20] ISO/IEC 12139-1:2009 - Information technology - Telecommunications and information exchange between systems - Powerline communication (PLC) - High speed PLC medium access control (MAC) and physical layer (PHY) - Part 1: General requirements.

[21] ISO/IEC JTC 1/SC 6 N18178: Presentation for PWI Proposal on Efficient Eavesdropping Combat Using UAV Mounted MF-RIS, 2023-12-14

[22] ISO/IEC JTC 1/SC 6 N 18179 : Presentation for PWI Proposal on MF-RIS for Wireless Communications, 2023-12-14

[23] Zheng A, Ni W, Wang W, et al. Enhancing NOMA Networks via Reconfigurable Multi-Functional Surface[J]. IEEE Communications Letters, 2023, 27(4): 1195-1199.

[24] Zheng A, Ni W, Wang W, et al. Next-Generation RIS: From Single to Multiple Functions[J]. IEEE Wireless Communications Letters, 2023.

[25] Wang W, Ni W, Tian H. Multi-Functional RIS-Aided Wireless Communications[J]. IEEE Internet of Things Journal, 2023.

[26] Jun Zhao , Yang Liu : A Survey of Intelligent Reflecting Surfaces (IRSs): Towards 6G Wireless Communication Networks , arXiv:1907.04789v3 [eess.SP] 2 Nov 2019.

[27] Mohamad H. Dinan and Arman Farhang : RIS-Assisted OTFS Communications: Phase Configuration via Received Energy Maximization Mohamad H. Dinan and Arman Farhang, arXiv:2404.07759v2 [cs.IT] 12 Apr 2024.

[28] ISO/IEC TR 29181-2:2014 Information Technology Future Network Problem statement and requirements Part 2: Naming and addressing.

[29] ISO/IEC TR 29181-1:2012 Information technology Future Network Problem statement and requirements Part 1: Overall aspects.

[30] ISO/IEC JTC 1/SC 6 N 18204 : SC 6 Business Plan (PERIOD COVERED: January 2024 – October 2024), 2024-01-05.

[31] ISO/IEC TR 29181-5:2014 Information technology Future Network Problem statement and requirements Part 5: Security.

[32] ISO/IEC JTC 1/SC 6 N 17676: Chinese experts' contribution on High-efficiency (HE) modulation and coding scheme (HE-MCS) , 2022-03-02

[33] ITU-R Recommendations SM.1046-3: Definition of spectrum use and efficiency of a radio system, 2017.

[34] Shannon, C.E. Communication in presence of noise. IRE.1949,37(1), 10–21.

[35] Saso Tomazic : Spectral Efficiency, January 2008DOI: 10.1081/E-EWMC-120043448

[36] Iyer, S., Patil, A., Bhairanatti, S. et al. A Survey on Technological Trends to Enhance Spectrum-Efficiency in 6G Communications. Trans Indian Natl. Acad. Eng. 7, 1093–1120 (2022). https://doi.org/10.1007/s41403-022-00372-w.

[37] Wu, L. Energy-efficient wireless communications. Nat Rev Electr Eng 1, 77 (2024). https://doi.org/10.1038/s44287-024-00027-8.

[38] F. Mahmood, E. Perrins and L. Liu, "Energy-Efficient Wireless Communications: From Energy Modeling to Performance Evaluation," in IEEE Transactions on Vehicular Technology, vol. 68, no. 8, pp. 7643-7654, Aug. 2019, doi: 10.1109/TVT.2019.2921304.

Deep Learning Based Defect Detection Research on Printed Circuit Boards

Qihang Yang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: gnay_qh@163.com

Fan Yu
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: yffshun@163.com

***Abstract*—As computer vision and deep learning detection techniques advance rapidly, their use in identifying defects has become more common across various industries. The significance of Printed Circuit Boards (PCBs) in contemporary electronic devices is undeniable, as they substantially influence the functionality and durability of these products. Thus, utilizing deep learning models for identifying flaws in Printed Circuit Boards (PCBs) is of particular importance. The focus of this study is primarily on examining PCB defect identification utilizing deep learning techniques. Firstly, it introduces the importance and development history of PCBs in the electronics and information industry. It then offers a comprehensive review of the existing research on conventional PCB defect detection approaches alongside methodologies grounded in deep learning. Following that, the structure of the YOLOv8 object detection model and its key technologies are elaborated. Lastly, the superior performance of YOLOv8 in PCB defect detection tasks is verified through comparative experiments. According to the evaluation metrics of the algorithm, the average detection accuracy reaches 92.3%, and the Frames Per Second (FPS) value reaches 157.2, meeting the accuracy requirements for PCB defect detection in the industrial domain.**

Keywords-Deep Learning; Defect Detection; YOLOv8

I. INTRODUCTION

The Printed Circuit Board (PCB), an essential element in electronic devices, is produced through processes that involve electronic chemistry, the industry is honored as the "mother of electronics". PCB plays an indispensable role in the electronic information industry. Widely used in a variety of fields, including but not limited to integrated circuits, artificial intelligence, medical equipment, aerospace and industrial equipment, PCB's main function is to connect the circuit components, so as to facilitate the electronic equipment to achieve higher performance and efficiency. By connecting various electronic components and devices together in an orderly manner, PCB realizes the effective assembly of circuits, which is crucial for the proper functioning of electronic devices. In modern technology, the role of the PCB is not only to provide circuit connectivity, it also helps to optimize the stability, reliability and performance of electronic equipment. With the help of PCBs, device manufacturers are able to achieve higher performance standards and more sophisticated functionality, thus providing a better user experience. Therefore, PCB is considered one of the cornerstones in the field of electronics and is important for the development of modern society and technological advancement.

Modern electronic and electrical devices must rely on PCBs for electrical interconnections. Therefore, the quality of PCB boards is crucial for electronic devices and directly affects the success of the product. With the rapid development of emerging fields such as Internet of Things (IoT) technology, automotive electronics and 5G communications, the quality of PCB board design and production is vital for the efficiency and dependability of electronic products. Following the reform and opening up period, China's electronics sector has seen swift advancement, particularly in recent times, coinciding with the boom in the electronics industry, China's PCB manufacturing industry has continued to develop rapidly, with output value and production steadily ranking among the global leaders, making important contributions to the national economy and employment. With the continuous renewal of

electronic products, price competition continues to reshape the supply chain structure. With its industrial distribution, cost and market advantages, China has rapidly emerged as one of the most important PCB production bases in the world.

II. Related Works

The PCB contains an extensive array of components, each with their intricate and varied characteristic details. During the production process, various uncertainties such as raw materials, production environment and manual operation often lead to various surface defects. Regular and prompt inspections are essential to maintaining production line functionality. As a result, there has been an ongoing investigation into PCB surface defect detection both domestically and internationally.

Currently, defect detection technologies are primarily divided into two approaches: conventional techniques and those based on deep learning. Next, the research status of these two directions will be described in detail.

Conventional methods for detecting defects in PCBs encompass manual examination, electrical measurement, and automated optical evaluation techniques. Manual inspection is one of the earliest approaches, which requires operators to use microscopes or magnifying glasses to discover various complex defects on the circuit board [1,2]. However, this approach is susceptible to subjective factors such as visual fatigue, which may lead to problems such as misdetection and missed detection. Simultaneously, the challenges associated with manual inspection increase because of the high cost of labor and reduced efficiency, compounded by the growing scale of integrated circuits and the intricacy of their wiring. Another method is electrical testing, which uses a probe instead of a needle bed, and a fast-moving electrical probe is placed in contact with the pins of the PCB for electrical measurement [3]. However, this approach requires the probes to be in contact with the board, which may damage the PCB surface and lead to unnecessary losses. Automated optical inspection, on the other hand, is limited by multi-sensor imaging, light source, field of view, and resolution. In real production environments, debugging is complicated and lacks good portability. All of these traditional methods have limitations, the process is complex, and any one of the errors may lead to misdetection. A single inspection method can no longer meet the demand for efficient and rapid inspection of production lines.

The swift advancement of deep learning algorithms has sparked a surge in research focusing on the detection of PCB surface flaws using Convolutional Neural Networks (CNNs) [4]. Various researchers have introduced multiple deep learning-oriented techniques to address the issue of PCB defect detection. For instance, Ding and colleagues [5] have presented the Tiny Defect Detection Network (TDD-Net), which integrates the fundamental network of Faster R-CNN with Feature Pyramid Networks (FPN) [6] to enhance the precision in detecting PCB defects. In addition, Li et al [7] trained a mixture of Faster R-CNN and YOLOv2 models and integrated the detection results of the two models to achieve a high degree of precision. Hu and colleagues [8] have combined FPN with ResNet50 as the foundational network for Faster R-CNN, incorporating the ShuffleNetV2 framework to enhance the model's detection precision. Meanwhile, Tang and associates [9] employed a dual network alongside a pyramid pooling module (PPM) for defect identification, attaining high performance in detection. However, although the above algorithms have achieved better results in reducing false and missed detections, the intricate nature of the comprehensive model and its elevated time complexity pose challenges in fulfilling the real-time detection necessities within PCB manufacturing settings. This complexity may affect the application of the algorithms, especially in production sites where fast detection is required.

Yuan Li et al [10] integrated the Multi-Residual Attention Mechanism (MRHAM) into the YOLOv4 algorithmic model to enhance the ability of sensory field attention for defective target features. Concurrently, they employed the K-means++ clustering method to conduct an in-depth analysis of the PCB dataset, enhancing model robustness and ensuring rapid processing. Liao and colleagues [11] introduced a PCB defect

detection framework using YOLOv4-MN3, which streamlined and optimized the backbone network, feature fusion unit, and prediction module, diminishing the parameter count and resulting in a detection rate of 56.98 frames per second (FPS). In contrast, Wang and colleagues [12] presented a lightweight network for defect detection, YOLOX-MC-CA, which incorporated Coordinate Attention (CA) [13] and enhanced the CSPDarkNet backbone to accelerate detection speed, achieving satisfactory performance on the PCB open dataset. These research efforts meet the real-time criteria by streamlining the network architecture, thereby accelerating detection. Nevertheless, such simplification can potentially compromise the model's capacity to extract features from the input images. Therefore, further improving the detection efficiency while maintaining the detection accuracy is an important issue of concern for current scholars at home and abroad.

III. Algorithms Model

YOLOv8 was proposed by Ultralytics in January 2023 as an improved version of the YOLOv5 algorithm model. Similar to YOLOv5, it does not have an associated paper at the moment, but its code has been open-sourced on the GitHub repository. YOLOv8 continues the overall architecture of the algorithm model since YOLOv4, as illustrated in Figure 1. The framework principally comprises three elements: the Backbone, which is the feature extraction network, the Neck, responsible for feature fusion, and the Head, the detection head component. Within this configuration, the Backbone commences the process by extracting attributes from the samples, resulting in the creation of feature maps at three distinct scales. The Neck integrates these three feature maps with surface and depth information, producing three new feature maps. The Head performs classification and regression on each sample point of the three new feature maps. The YOLOv8 algorithm model is categorized into N, S, M, L, and X versions based on the model's width and depth. In the following sections, a detailed analysis of YOLOv8 will be conducted, covering Backbone, Neck, Head, and the loss function.

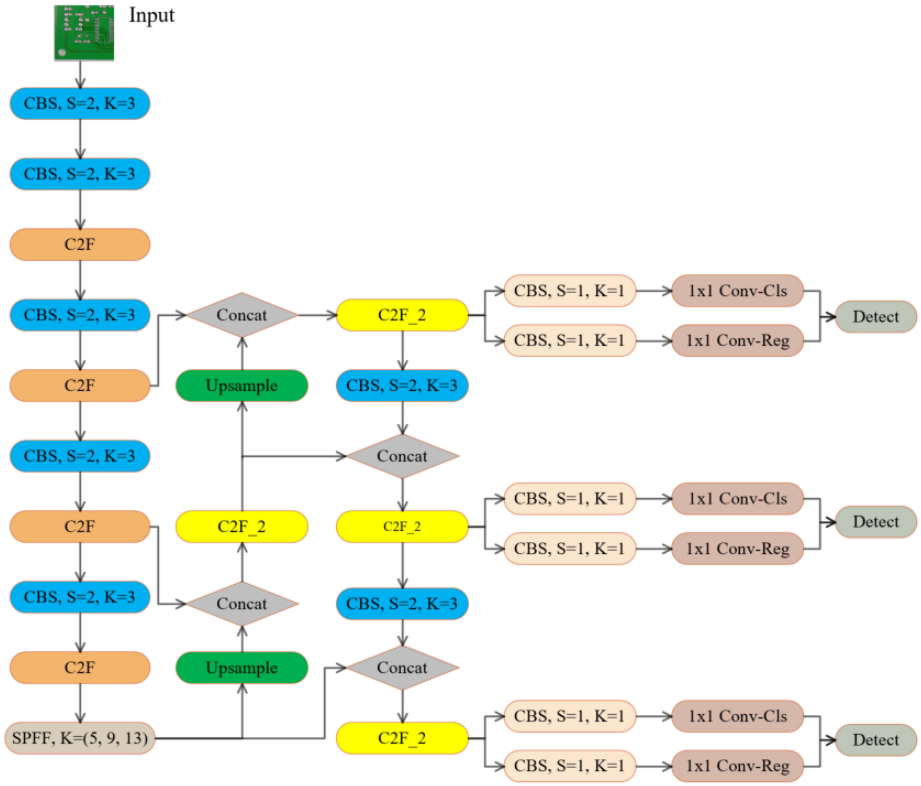


Figure 1. Diagram of YOLOv8 structure

A. Feature Extraction Network

The attribute extraction network of YOLOv8 represents an enhancement upon YOLOv5. Rather than employing the Focus component for parameter reduction, YOLOv8 utilizes a 2D convolution with a stride of 2 and a kernel size of 3 for expanding channels from the initial input. It also introduces a novel convolutional unit, C2f, which takes the place of the C3 unit in YOLOv5, while maintaining the SPPF pooling element.

The C2f module is still constructed based on the CSPNet (Cross Stage Partial Network) idea, and the architecture is depicted in Figure 2. Upon inputing the attributes into the C2f unit, the 1x1 convolution will be used for channel integration, and then the feature tensor will be sliced into two parts according to the channel Split, one of which will enter the Bottleneck block to further extract the features, and the other part will be spliced with the features processed by the Bottleneck block according to the channel. This configuration augments the ability of the convolutional neural network to extract features and minimizes the time spent on memory access. The Bottleneck is illustrated in Figure 3, which utilizes the residual idea, where the original input is convolved twice to extract features and then pointwise added to the original input. One part of the output of the Bottleneck continues to be used for the Bottleneck operation, and the other part is spliced with the half after Split for per-channel splicing. The advantage of the residual idea is that it not only preserves the basic features of the original input, but also avoids the problem of vanishing gradients.

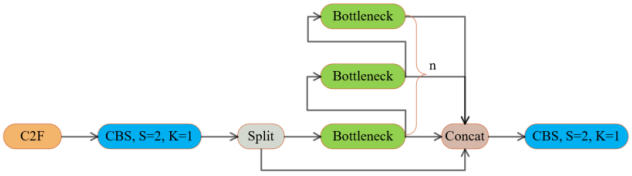


Figure 2. Schematic diagram of C2F module

Figure 3. Bottleneck Schematic Diagram

SPPF (Spatial Pyramid Pooling Fast) is an improved version of SPP, the structure is able to extract features at different scales of the object as SPP, enrich the feature information of the output layer of this feature, and present the same effect, while the time consumed is half of SPP, the structure is shown in Figure 4. The main process is: after the input features pass through the 1x1 convolutional integration channel, the output is copied in two copies, one for pooling kernel of 5 for maximum pooling, and the other is involved in the splicing with the output of the pooling layer. The pooled feature output is also copied into two copies, one to participate in the splicing of other pooled outputs, and the other to continue the maximum pooling, repeat this step three times, a total of four copies of the output features, four copies of the output features are spliced according to the channel, and the channel information is integrated using the 1x1 convolution.

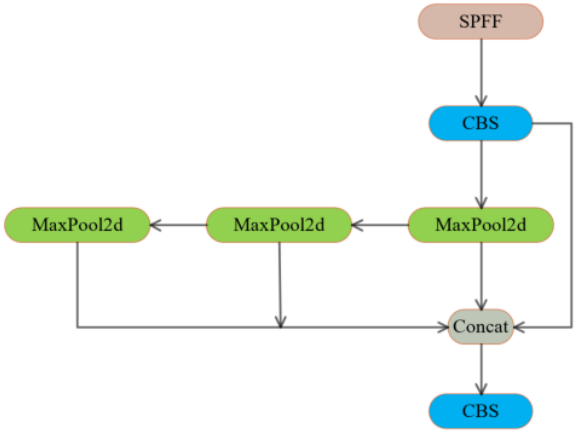


Figure 4. Schematic diagram of SPPF structure

B. Feature Fusion Networks

The Neck segment of YOLOv8 predominantly utilizes an enhanced version of the PANet (Path Aggregation Network) design, which builds upon the FPN (Feature Pyramid Networks). This structure involves upsampling the feature layers of smaller dimensions and fusing them with feature layers of larger dimensions, directly outputting the FPN to enrich the feature information contained in the large-dimensional feature maps. On the other hand, PANet further downsamples the fused large-dimensional features, merges them again with the

feature maps of smaller dimensions, enriching contextual information, and enhancing the expression capability of the smaller-dimensional features. In contrast to YOLOv5, YOLOv8 employs the C2f unit in place of the C3 unit, and integrates channel-wise before upsampling the feature maps of smaller dimensions by removing the 1x1 convolution.

C. Detection Head

The Head component of YOLOv8 has transitioned from the initial coupled head design to the present prevalent decoupled head architecture. This decoupled head represents a standard design in object detection, tasked with deriving target location and class information from the detection network's feature map.

Specifically, decoupling the head involves separating the main part of the neural network model from the classifier part for training. The advantage of this design is the flexibility to modify and replace the classifier without altering the backbone network. Moreover, the decoupled head efficiently diminishes the quantity of parameters and the computational burden, which enhances the model's capacity for generalization and robustness, all while maintaining the backbone network architecture.

Within the YOLOv8 decoupled head structure, the prior Obj branch has been eliminated. Now, only separate branches for classification and regression persist. The regression branch employs an integral form derived from the Distribution Focal Loss concept. It is important to highlight that the channel counts in the classification and regression branches of the decoupled head may differ.

D. Loss Function

The loss function is a critical element in the training of algorithmic models; an appropriately designed loss function can lead to quicker model convergence and enhanced robustness. The loss function utilized by YOLOv8 comprises a combination of classification loss VFL (Varifocal Loss) and regression loss CIOU + DFL (Distribution Focal Loss).

The classification loss VFL is shown in Equation 1, where q represents the overlap and concurrent alignment between the projected coordinate system and the actual coordinate system, p is the softmax output value of the category, α considers the spectrum of values [0,1], and γ considers the spectrum of values [0,5]. Since there are too few positive samples during training, reducing the Loss contribution of negative samples makes the model more inclined to the training of high quality positive samples.

$$VFL(p,q)=\begin{cases}-q(q\log(p)+(1-q)\log(1-p)) & q>0\\ -\alpha p^{\gamma}\log(1-p) & q=0\end{cases} \quad (1)$$

The regression loss L_{reg} mainly calculated using the summation of the *CIOU* loss L_{CIOU} and the *DFL* loss is depicted in Equation 2, the L_{CIOU} mathematical expression is illustrated in Equation 3, and the *DFL* mathematical expression is illustrated in Equation 6.

$$L_{reg}=L_{CIOU}+DFL \quad (2)$$

$$L_{CIOU}=1-IOU(A,B)+\frac{\rho^2(b,b^{gt})}{c^2}+\alpha v \quad (3)$$

$$\alpha=\frac{v}{(1-IOU)+v} \quad (4)$$

$$v=\frac{4}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}}-arctan\frac{w}{h})^2 \quad (5)$$

$$DFL(S_i,S_{i+1})=-((y_{i+1}-y)\log(S_i)+(y-y_i)\log(S_{i+1})) \quad (6)$$

In Equation 3, $IOU(A,B)$ signifies the degree of overlap and concurrent alignment between the actual and estimated coordinate systems, $\rho^2(b,b^{gt})$ notes the geometric distance measured in Euclidean space between the central points of the forecasted and actual coordinate systems, c is

the diagonal distance of the outer rectangle containing the real and predicted target frames, and α is the coefficient used for balancing the ratio as shown in Equation 4. v is the distance between the height-width ratio $\frac{w^{gt}}{h^{gt}}$ of the real coordinate frame and the height-width ratio $\frac{w}{h}$ of the predicted coordinate frame, used to measure the height-width scale consistency as shown in Equation 5.In Equation 6, y denotes the true label value, y_i and y_{i+1} denote the two closest values of y, respectively, and S_i and S_{i+1} correspond to the probabilities of the two values.

IV. Experiments

A. Experimental Environment

The experimental environment is shown in Table I

TABLE I. Experimental Environment

Experimental environment	Version
CPU	Intel Core i7-11800H
GPU	NVIDIA GeForce RTX307
Language	Python3.7
Deep Learning Framework	Pytorch1.11.0
CUDA	11.3.0
Compiler	Pycharm2021

B. Dataset

The dataset in question is a PCB (printed circuit board) defect collection made available by the Open Lab of Peking University. The types of defects are missing hole, mouse bite, open circuit, short circuit, spur, and spurious copper, and it contains a total of 11,361 images.

C. Evaluation Metrics

When assessing single-target detection models, it's customary to utilize metrics such as accuracy and recall to gauge the model's detection efficacy. The accuracy rate is defined as the ratio of correctly identified actual samples to the total number of samples, while the recall rate represents the ratio of correctly identified actual positive samples to the total number of actual positive samples. The precise formulations for the accuracy rate P and the recall rate R are depicted in Equations 7 and 8, respectively.

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

Within this evaluative measure, TP signifies the count of positive samples rightly identified, FP denotes the quantity of negative samples erroneously labeled as positive, and FN indicates the number of positive samples mistakenly classified as negative. For a detection model, it is often desirable to have higher precision and recall, but it is often the case that a rise in one metric causes a fall in the other. In order to comprehensively assess the performance metric of a detection model, researchers introduce the P-R curve. The P-R curve is a curve that describes the change in the relationship between the model's accuracy and recall, and by determining the area beneath the curve, it can intuitively reflect the model's goodness or badness.

In the assessment of multi-target detection models, the metric often employed to gauge the comprehensive efficacy is the Mean Average Precision (mAP). This metric encapsulates the model's detection acuity across all categories. The mAP score is derived from the average of the Average Precision (AP) for each category. It mirrors the precision in detecting individual targets and is a function of the model's precision and recall rates. The underlying computational expressions are illustrated in Equations 9 and 10.

$$\mathrm{AP} = \int_0^1 PRdR \tag{9}$$

$$\mathrm{mAP} = \frac{1}{c}\Sigma_{c_i \in \mathcal{C}} AP_{c_i} \tag{10}$$

D. Results

To validate the detection capabilities of the models presented in this study, a comparative analysis with contemporary mainstream object detection methodologies was conducted. The training of all models was carried out utilizing transfer learning techniques, with the COCO2017 dataset serving as the pre-training dataset. The outcomes of these experiments are documented in Table Ⅱ.

TABLE II. EXPERIMENTAL RESULTS

Model	mAP(%)	FPS	Quantity of participants /M
Faster R-CNN	89.5	20.5	125.3
YOLOv5	85.3	79.4	42.7
YOLOv7	87.1	102.5	37.2
YOLOv8	92.3	157.2	28.5

The model introduced in this paper achieves superior overall performance regarding detection accuracy and speed, with a mAP of 92.3 and a detection rate of 157.2 frames per second (FPS), while only having 28.5 million parameters. Although the two-stage approach, Faster R-CNN, offers higher detection accuracy due to its maximum input image resolution, its large parameter count and slower computation make it less practical for real-world applications. In contrast, single-stage models like YOLOv5 and YOLOv7 still trail behind YOLOv8 in both accuracy and speed. These findings are presented visually in Figures 5 and 6.

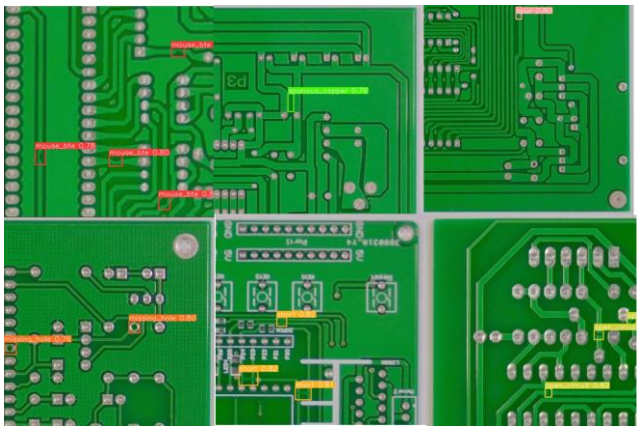


Figure 5. Pcb defect detection effect

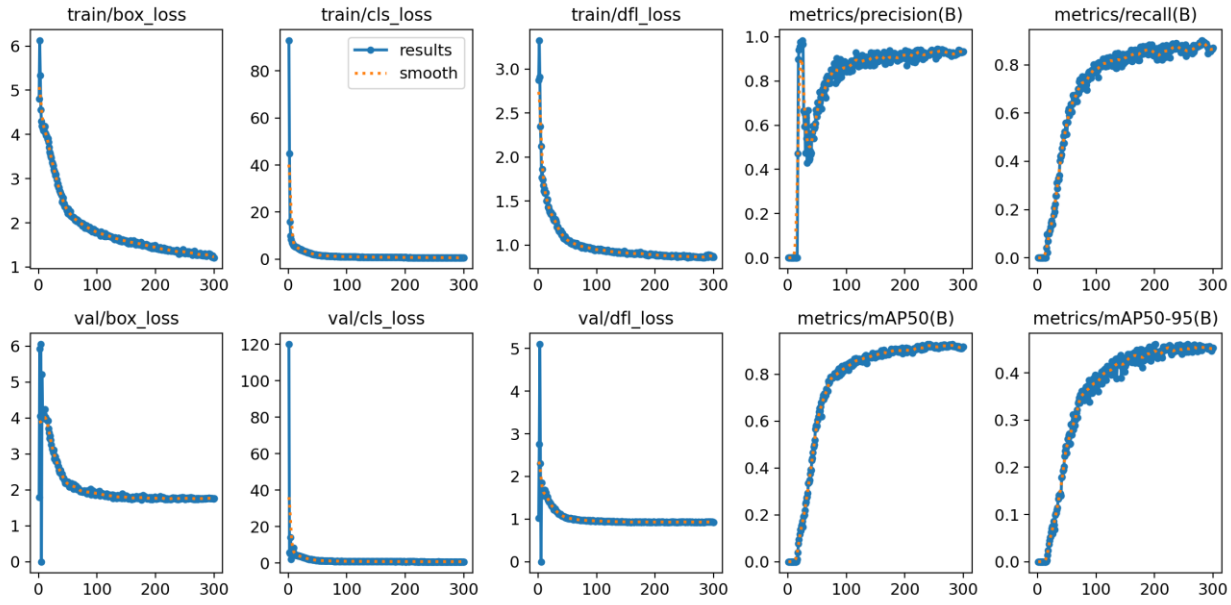


Figure 6. Loss, Precision, Recall, mPA0.5 and mAP0.5-0.95 curves

V. CONCLUSIONS

In this paper, we have explored the problem of PCB surface defect detection based on printed circuit boards. Traditional PCB defect detection methods such as expert visualization or machine vision have more or less limitations, and with the evolution of deep learning target detection

algorithms in full swing, the application of deep learning methods to the PCB defect detection problem is also a general trend. Due to the industrial production of PCB surface defects there are difficult to detect the problem, this paper carries out a series of analyses of the existing detection model, and finally in the experimental environment under the same conditions, for the detection model of YOLOv8 and the other mainstream detection models in the industry in the detection of the average accuracy of the average value, detection speed, model complexity and other indicators of the assessment of a comprehensive comparison of the experiments. Furthermore, the experimental results suggests that the YOLOv8 detection algorithm exhibits a significant enhancement in the mentioned metrics, demonstrating its relevance for identifying defects in electronic products. This underscores its valuable research potential within the realm of object detection.

REFERENCES

[1] Li Y, Li S. Defect detection of bare printed circuit boards based on gradient direction information entropy and uniform local binary patterns [J]. Circuit World, 2017, 43(4): 145-151.

[2] Chomsuwan K, Yamada S, Iwahara M. Improvement on defect detection performance of PCB inspection based on ECT technique with multi-SV-GMR sensor [J]. IEEE Transactions on magnetics, 2007, 43(6): 2394-2396.

[3] Verdingovas V, Jellesen M S, Ambat R. Colorimetric visualization of tin corrosion: A method for early stage corrosion detection on printed circuit boards [J]. Microelectronics Reliability, 2017, 73: 158-166.

[4] Yang J, Li S, Wang Z, et al. Using deep learning to detect defects in manufacturing: a comprehensive survey and current challenges [J]. Materials, 2020, 13(24): 5755.

[5] Ding R, Dai L, Li G, et al. TDD‐net: a tiny defect detection network for printed circuit boards [J]. CAAI Transactions on Intelligence Technology, 2019, 4(2): 110-116.

[6] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[7] Li Y T, Kuo P, Guo J I. Automatic industry PCB board DIP process defect detection with deep ensemble method [C]//2020 IEEE 29th International Symposium on Industrial Electronics (ISIE). IEEE, 2020: 453-459.

[8] Hu B, Wang J. Detection of PCB surface defects with improved faster-RCNN and feature pyramid network [J]. Ieee Access, 2020, 8: 108335-108345.

[9] Tang S, He F, Huang X, et al. Online PCB defect detector on a new PCB defect dataset[J]. arXiv preprint arXiv:1902.06197, 2019.

[10] Li Yuan. Research and application of target detection method based on improved YOLOv4 [D]. Chongqing: Chongqing University of Posts and Telecommunications, 2021.

[11] Liao X, Lv S, Li D, et al. Yolov4-mn3 for pcb surface defect detection [J]. Applied Sciences, 2021, 11(24): 11701.

[12] Xuan W, Jian-She G, Bo-Jie H, et al. A lightweight modified YOLOX network using coordinate attention mechanism for PCB surface defect detection [J]. IEEE Sensors Journal, 2022, 22(21): 20910-20920.

[13] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.

Research on the Expanded Night Road Condition Dataset Based on the Improved CycleGAN

Lei Cao
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: 2443816614@qq.com

Li Zhao
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: 332099732@qq.com

***Abstract*—Image style transfer is a major area of study in image processing and has applications in creative production, special effects for film and television, and other areas. Image style transfer is the process of using style transfer technology to change a common image into one with a different style without changing the content. Image style transfer methods are mainly divided into traditional image style transfer methods and deep learning image style transfer methods. The two primary classifications of picture style transfer techniques are deep learning technologies and conventional methods. Traditional image style transfer methods have poor results and are difficult to apply in people's lives. With the quick advancements in machine learning, digital image processing, and computer vision, deep learning image style transfer methods have received widespread attention from researchers. Most of these methods use convolutional neural networks to achieve image style transfer on the premise of paired data sets, but obtaining paired data sets is difficult and costly. Accordingly, it is of great significance to study unpaired images to implement style transfer algorithms. The primary focus of this study is the CycleGAN network-based picture style transfer technique, and improves this algorithm in content compiler, style compiler. It is applied to the generation of night road conditions during autonomous driving training.**

Keywords-Generative; Adversarial Networks; Style Migration; Deep Learning

I INTRODUCTION

In recent years, generative adversarial network-based image style migration methods have advanced significantly, but many image style migration algorithms can only complete data that match each other, but it's not possible that all the training sets are complete pairs, which increases the difficulty of obtaining them, and to a certain extent limits the application of the model. Naturally there are inconveniences and struggling in some topics. CycleGAN model is a general framework to solve different types of image migration without matching data, and its task is to firstly transform different image domains by learning the correspondence between the real image domain and the art style domain. [1] The implication of this correspondence is that the generator can transform an image into an image in the artistic style domain. In a similar vein, CycleGAN [2] must translate images from the Y domain to the X domain during training, and the process is similar to the process described above, only exchanging X and Y letters. Such a cycle forms a cyclic network. It is on the basis of the Cycle GAN model that the research content of the text is built, and through the understanding and study of the CycleGAN model, it is improved upon and extended in the application of migration of artistic image styles.

In generative adversarial networks, most of the generative networks are encoded and then decoded, whereas encoding uses encoders to convert the low dimensional spatial properties of the input into high dimensional spatial properties and decoders are used to decode and output the high dimensional spatial properties. [3] While the picture change is carried out in the decoding phase, the content features of the initial picture are preserved to the greatest extent possible throughout the migration process.

The proposal of Cyc1eGAN introduces a new loss, which is similar to what we call content loss in the style transfer process. To the extent that it limits the generator, it's the loss of cycle consistency. When the cycle consistency loss does

not match the actual image of the target area, the algorithm will restrict the content of the model and find the corresponding relationship in the two unmatched data sets.

II NETWORK ARCHITECTURE MODEL

The image style transfer algorithm based on Cyc1eGAN proposed in this article. In the generative adversarial network, most of the generating networks are encoded first and then decoded. The encoding uses the encoder to convert the input low-dimensional space characteristics into high-dimensional ones. Spatial characteristics, and the decoder uses the decoder to decode and output high-dimensional spatial characteristics. During the migration process, try to keep up the writing characteristics of the initial picture, and perform image conversion in the decoding part. To make things easier to understand, we call the network that is used for encoding the encoder and the network that is used for decoding the generator. It uses residual networks to fuse images and achieve image style conversion. On this basis, a multi-scale discrimination method is adopted to increase the image style transfer's correctness.

The training of GAN is optimized alternately, and the discriminative and generative models are interleaved. The first is to keep the generated model unchanged and improve the accuracy of the discriminant model. After completion, we leave the discriminative model unchanged to increase the probability that the generative model generates real data. After training to a certain extent, the generative model produces samples that are quite similar to the original samples.

At this time, relying on the discriminant model to determine will cause difficulties and it will not be so accurate. It will enter an equilibrium state and the training will end.

The CycleGAN network can transform the style between two data sets, and its network training is bidirectional. At different network levels, the retrieved photos' content and style also differ. Content characteristics are extracted using the autoencoder, while style features are extracted using the variational autoencoder.

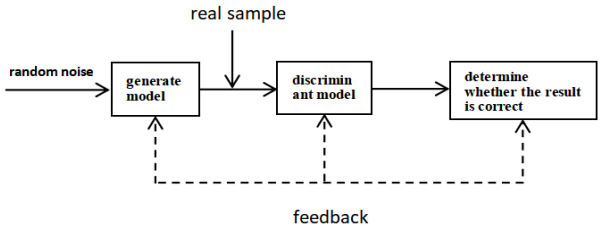


Figure 1. Example of a ONE-COLUMN figure caption.

Next, the image's content and style components are retrieved, and lastly, a style-transferred image is produced.

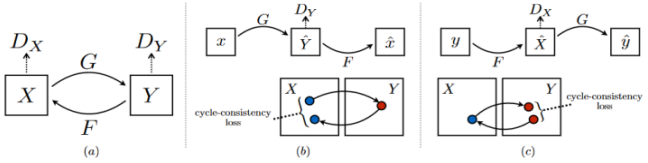


Figure 2. CycleGAN Network structure diagram

A. *Encoder structure*

1) *Content coding:*

The content encoder uses the autoencoder network model. Autoencoder technique [4] is a data compression-based dimensionality reduction algorithm which is a new learning method for data compression and decompression based on data. Its function is to compress the data. In this paper, only the coding component of the autoencoder is used to extract the image. The generative adversarial network's generator will perform the function of generating the image.

The content encoders in the network structure of this document consist mainly of the residual network. Deep residual network is a very effective method in picture categorization, target location and detection.

The content encoder consists of 3 convolutional layers and 2 residual modules.

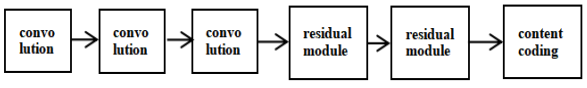


Figure 3. Content encoder structure diagram

2) *How the content encoder works:*

a) The pixels of the input image are 256 x 256, and a down sampling operation is performed on it.

The first layer: 64 7 x 7 convolution kernels, sliding step size: 1, padding size: 3, and instance normalization processing, and then perform Relu activation processing.

b) The second and third layers: 128 convolution kernels of 4 x 4 size, sliding step size: 2, padding size: 1, and the subsequent steps are the same as the first step.

c) There are two convolutional layers with 256 3 x 3 convolution kernels in the residual module. Both the cushioning and the sliding step sizes are one. The subsequent steps are the same as the first step.

3) Style coding

The variational self-encoder network model is used in this paper's network structure. The variational auto-encoder needs to generate implicit vectors conforming to a Gaussian distribution during the encoding process, a constraint that enables the training data's ground rules to be understood by the encoder so that it can learn the implicit variables of the input data. With the training data variational autoencoders figure out the parameters' probability distribution. Variational autocoding techniques are used to encode image formats to give them greater randomness.

Variational autoencoder is implemented by converting the output of the encoder into two vectors corresponding to the vectors representing the standard deviation and mean. Using the mean and standard deviation vectors, an implicit variational model can be constructed to derive the coding vectors.

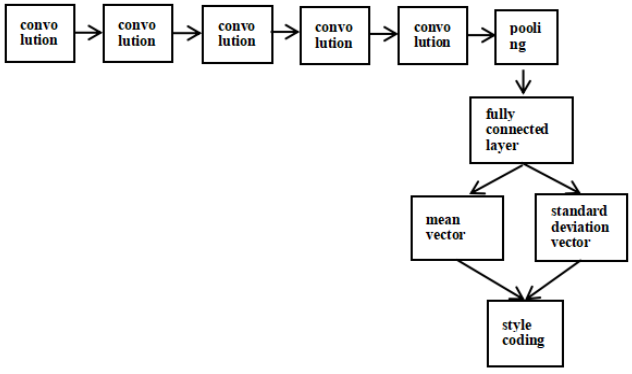


Figure 4. Style encoder structure diagram

Five convolutional layers, one pooling layer, and one fully linked layer make up the style encoder.

B. Generative network structure

In generative networks, the residual structural unit normalization [5] used is an algorithm based on adaptive normalization that speeds up the stylization of an image while maintaining its style. Here, we first input the content and style data at the same time, then go through the four residual modules, and finally the style transfer is completed by the up-sampling and convolution operation.

Neural networks for different tasks often choose different normalization functions, and different normalization functions will have different effects on the final results. Compared with the original GAN model, the CycleGAN model replaces batch normalization with instance normalization in order to make the model more stable.

CycleGAN uses instance normalization to perform separate normalization processing on each image, making the content information between each image independent of each other, avoiding the mutual influence between images and resulting in blurred generated images. Some progress has been made using instance normalization, but this method only works better in image style conversion when there are small differences in the shape and texture of the image domain. In aspects such as face style conversion, which have large differences, the effect is not very ideal. Instance normalization operates on each image and normalizes the H and W dimensions on a single channel, so in IN, different channels of different features are irrelevant, which allows the content feature information to be transferred to the style feature When applied to information, good results will always be achieved, but if it is only applied to a single channel, it will interfere with the original semantic information. Layer normalization acts on all channels. Layer normalization normalizes the C, H, and W dimensions of each image. It considers more global information, so some detailed information is often ignored. During training, the parameters of adaptive layer instance normalization (AdaLIN) [11], can be learned from the data set by adaptively selecting the proportion between instance and layer

normalization. Adaptive layer instance normalization combines the two, offsets their shortcomings, and combines the advantages of both. The specific formula is as follows:

$$\hat{\partial}_I = \frac{\partial - \mu_I}{\sqrt{\sigma_I^2 + \varepsilon}}, \hat{\partial}_L = \frac{\partial - \mu_L}{\sqrt{\sigma_L^2 + \varepsilon}} \quad (1)$$

$$AdaLIN(\partial, \gamma, \beta) = \gamma.(\rho.\hat{\partial}_I + (1-\rho).\hat{\partial}_L) + \beta \quad (2)$$

The adaptive layer instance normalization method is introduced into the decoder of the generator network, and combined with the self-attention mechanism, important local feature information and global feature information are automatically learned during network training.

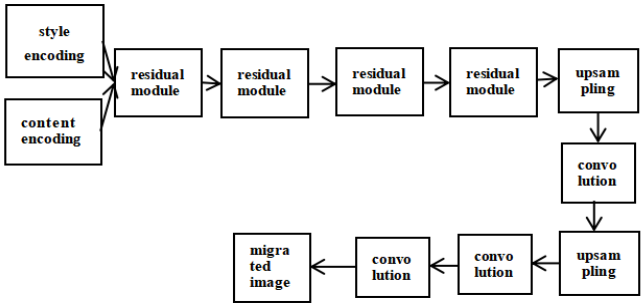


Figure 5. Structure of the generative network

The migrating image is the result of the generative network, which takes style and content elements as inputs.

C. *Discriminative network design*

Convolutional neural network [6] is still used in feature extraction of the whole image, and then recognition is performed based on the extracted features. Nevertheless, this approach frequently overlooks the image features, leading to issues like detail loss and image blurring. Therefore, we use a multi-scale discriminative network, which is an improved chunked discriminative network.

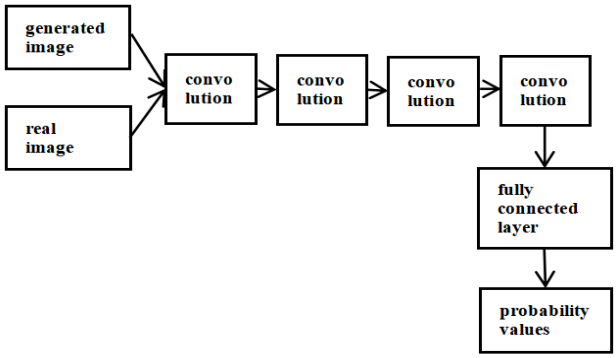


Figure 6. Diagram of the structure of the discriminating network

The PatchGAN discriminator is essentially a multi-layer convolutional network. The input image of the network undergoes multi-layer feature extraction and finally outputs a 30*30 feature map. The network also divides the input image into 30*30 patches. The pixels in the feature map match the input image patch, and the value in the feature map is between 0 and 1, which is employed to convey how close the patch component image is to the original image.

Finally, to show how similar the complete image is to the original image, the scores of each patch are put together and averaged. The PatchGAN discriminator can identify patches that are significantly different from other patches in image information and give the corresponding patch a lower discriminant score. This way, the local information of the generated image can be consistent with the overall information, which significantly lowers the expense of producing the picture mistake. In addition, because the local image and the overall image are closely related, the patchGAN network based on local image area discrimination plays a positive role in the fusion of the overall information and local information of the image.

The discriminative network has one fully connected layer and four convolutional layers making up its network structure.

III LOSS FUNCTION DESIGN

A. *Adversarial Loss*

Adversarial loss [7] is used to bring the generated image closer to the actual target style

image. Equation displays the adversarial loss computation formula.

$$L^{y_1}_{GAN} = E_{c_1-p(c_1),s_2-q(s_2)}[\log(1-D_1(G_1(c_1,s_2)))] + E_{y_1-p(Y),s_2-q(s_2)}[\log D_1(y_1)] \quad (3)$$

$$L^{x_1}_{GAN} = E_{c_2-p(c_2),s_1-q(s_1)}[\log(1-D_2(G_2(c_2,s_1)))] + E_{x_1-p(X),s_2-q(s_2)}[\log D_2(x_1)] \quad (4)$$

$G_1(c_1,s_2)$ denotes the output migration image, and $D_1(G_1(c_1,s_2))$, $D_1(y_1)$ are the discrimination results, respectively.

B. Image Reconstruction Loss

In the article, the generative network is fed both the style and content elements that were taken out of the style and content encoders in order to improve the capability of generative network in generative adversarial network.

The model is characterized by clearer and richer images clearly state the units for each quantity that you use in an equation.

C. Content encoding loss

Because a content encoder is included, the content coding loss is added to the loss function. The article proposes a method for style transfer using mismatched data and content feature extraction of the image using the content encoder, which ensures the invariance of the image during the style transformation. The style-shifted image [8] is re-inputted into the content encoder Ec, and as much as feasible, the content information of the picture is passed through the content information of the image that is output by the content encoder c so that the content information obtained has the same degree as the actual image described in order to train the content encoder.

D. Style coding loss

The style encoding loss is added to the loss function since the style encoder is now included. In order to ensure that the style converted image maintains the same style as the target image, the modified image is once again entered into the style encoder Es, and the extracted image style information must be as similar as possible to the style information outputted via the style encoder, so as to achieve the training of the style encoder.

E. Loss of cyclic consistency

CycleGAN network model, both to stylize the image and to recover the transformed image again, so the cyclic consistency loss of the image needs to be considered [9].

$$L^{y_1}_{cyc} = E_{y_1-p(Y)}[\| G_1(G_2(y_1)) - y_1 \|_1] \quad (5)$$

$$L^{x_1}_{cyc} = E_{x_1-p(X)}[\| G_2(G_1(x_1)) - x_1 \|_1] \quad (6)$$

IV ANALYSIS OF IMAGE STYLE MIGRATION RESULTS

A. Training and Testing

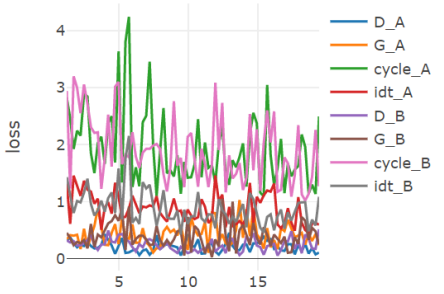
The training process of the network optimizes the parameters using Adam's algorithm [10], this uses adaptive low-order moment estimation as the basis for a one-time gradient optimization of any objective function.

It mainly contains the following features:

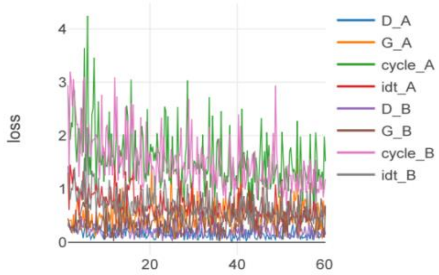
1) *Straightforward realization*
2) *Efficient Computing*
3) *Less memory usage*
4) *Gradient diagonal scaling's invariance*
5) *Appropriate for addressing optimization issues with a lot of data and parameters*
6) *Suitable for non-stationary objectives*
7) *Ideal for resolving issues with sparse gradients or extremely high noise*
8) *Hyper-parameters essentially just require a very tiny amount of intuition to interpret.*

We will train the content encoder, style encoder, generator, and input the content and style images to the appropriate content and style encoders for feature extraction during the project's testing phase, and then input the extracted content from the encoder to the generator to finally produce the style migration image on the generator.

Below are some loss function images from the training:



(a)



(b)

Figure 7. Image of the loss function

Specific values for some of the losses are listed in Table I:

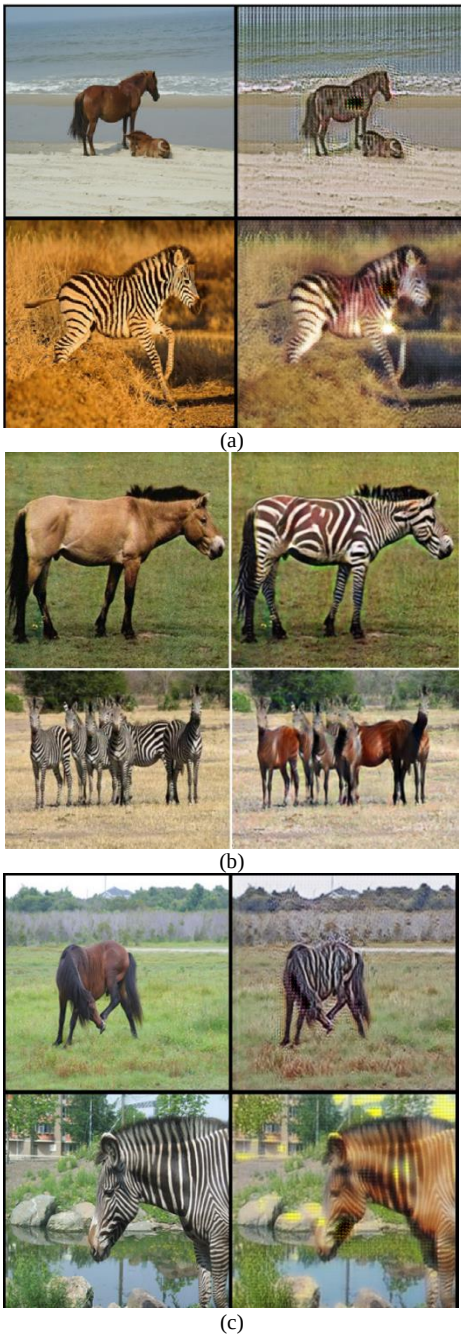
TABLE I. TABLE OF INDIVIDUAL LOSS VALUES

epoch	D_A	G_A	cycle_A	idt_A	D_B	G_B	cycle_B	idt_B
epoch:1	0.31	0.34	2.83	1.31	0.32	0.45	2.94	1.43
epoch:1	0.37	0.41	2.46	0.61	0.28	0.36	1.23	1.09
epoch:1	0.28	0.40	1.93	1.46	0.31	0.23	3.20	0.80
epoch:1	0.20	0.42	2.24	1.30	0.23	0.29	3.00	1.01
epoch:2	0.24	0.17	2.16	1.10	0.23	0.27	2.55	0.99
epoch:2	0.22	0.52	2.89	1.37	0.27	0.32	3.06	1.16
epoch:2	0.29	0.20	2.85	1.21	0.32	0.18	2.72	1.40
epoch:2	0.22	0.41	1.85	1.18	0.42	0.61	2.30	0.96
epoch:3	0.18	0.57	1.50	0.96	0.23	0.20	2.21	0.76
epoch:3	0.19	0.88	2.03	1.05	0.15	0.42	2.22	0.85
epoch:3	0.24	0.52	2.14	0.50	0.25	0.37	1.22	1.02
epoch:3	0.43	0.24	1.66	0.90	0.25	0.49	1.76	0.82
epoch:4	0.33	0.23	2.37	0.99	0.49	0.62	2.53	1.06
epoch:4	0.23	0.45	2.50	0.79	0.30	0.66	1.61	1.16
epoch:4	0.08	0.60	1.69	1.33	0.28	0.76	3.02	0.80
epoch:4	0.24	0.25	3.64	1.09	0.42	0.66	3.10	1.58
epoch:5	0.36	0.72	1.34	0.72	0.43	0.87	1.66	0.68
epoch:5	0.10	0.35	3.81	0.73	0.34	0.15	1.82	1.75
epoch:5	0.11	0.35	4.24	0.88	0.30	0.91	1.78	2.15
epoch:5	0.14	0.60	1.26	1.10	0.20	0.24	2.22	0.63
epoch:6	0.16	0.28	1.60	0.93	0.12	0.34	1.79	0.71
epoch:6	0.06	0.28	1.27	0.80	0.25	0.40	1.59	0.61
epoch:6	0.14	0.53	2.40	0.91	0.20	0.40	1.86	1.34
epoch:6	0.16	0.28	2.51	0.90	0.32	0.34	1.93	1.23
epoch:7	0.07	0.22	3.46	0.93	0.34	0.24	1.93	1.30
epoch:7	0.14	0.11	1.99	0.91	0.31	0.25	1.98	0.94
epoch:7	0.41	0.18	1.10	1.10	0.23	0.57	2.01	0.58
epoch:7	0.16	0.18	1.84	0.80	0.16	0.32	1.93	0.83
epoch:8	0.23	0.47	2.64	0.72	0.21	0.23	1.48	1.20
epoch:8	0.25	0.56	1.50	0.65	0.28	0.32	1.18	0.73
epoch:8	0.24	0.46	1.55	0.83	0.18	0.24	1.61	0.70
epoch:8	0.20	0.51	1.44	1.05	0.13	0.34	2.76	0.70
epoch:9	0.20	0.40	1.81	0.83	0.10	0.48	1.63	0.84
epoch:9	0.05	0.59	1.14	0.84	0.20	0.49	1.77	0.42
epoch:9	0.20	0.29	1.70	0.64	0.14	0.05	1.25	0.86
epoch:9	0.43	0.15	1.42	0.86	0.10	0.73	2.14	0.61
epoch:10	0.23	0.73	1.43	0.75	0.12	0.11	2.20	0.74
epoch:10	0.20	0.50	1.65	0.69	0.16	0.57	1.71	0.74
epoch:10	0.34	0.41	2.45	0.73	0.10	0.51	1.26	1.17
epoch:10	0.09	0.43	1.43	1.04	0.16	0.32	1.93	0.71

During the actual training process, the training results can be judged according to the loss value.

The smaller the value, the more successful the training is. Generally, the training results improve with a lesser decrease of D. As can be seen from Figure 7, as training continues, each loss value has a decreasing trend. It can also be observed during the training process that the effect of style transfer gradually becomes stronger as the epoch increases.

There are also some demonstrations of images during training, as in Figure 8:



(a) (b) (c)

Figure 8. Training presentation diagram

It can be seen that at the beginning of the training, the effect is not very obvious. As the number of iterations and training increases, the style migration can slowly be seen, and eventually the transformation is completely clear.

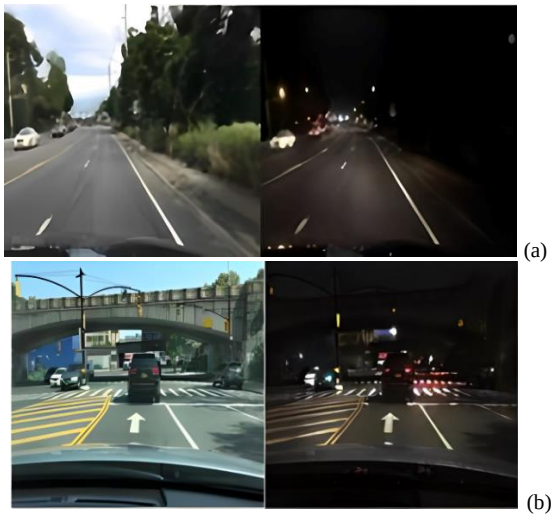
B. Experimental application and analysis

The topic of autonomous driving has advanced quickly in recent years, and the quality and quantity of datasets are crucial for the research in the domain of self-driving cars, and excellent datasets can help the autonomous.

Diving system better adapt to different lighting and environmental conditions, to enhance the self-driving system's functionality and safety at night and in low light conditions. At night or in low-light conditions, objects in images are often more difficult to identify and locate than during the day, which can lead to degraded performance of autonomous driving systems, increasing the risk of traffic accidents. Therefore, the dataset can be enriched and enhanced by using the method of daytime and nighttime road map style transfer, to enhance the capacity for generalization and the training effect of the autonomous driving system.

In this experiment, 1000 actual traffic maps were selected as the training set using the BDD100K dataset, including 500 road maps during the day and 500 road maps during the night period. 400 actual traffic maps were used as training sets, including 200 during the day and 200 at night.

The conversion effect is shown in Figure 9:



(a) (b)

Figure 9. Day and night conversion diagram

V CONCLUSIONS

With deep learning technology developing continuously in recent years, there is a wave of new development trend within the domain of image processing, and image style migration is one of the research hotspots. The focus of this thesis is to implement this on the basis of generative adversarial networks. The article covers relevant techniques and effects as well as the current status of picture style migration development. The picture style migration method based on deep learning is explained.

Then it explains the corresponding theory and knowledge, convolutional neural networks, generative adversarial networks, and artificial neural networks. Emphasis is placed on the fundamental idea behind the generative adversarial network model, its training procedure, its style of derivation, and its application to the field of graphics transformation.

The generative adversarial network-based technique for image style transfer still has many issues that need further research. In network training, in order to lessen the difficulty of network training and increase the network's learning efficiency the model must be optimized and trained. The model-optimized offline picture style transfer approach has essentially achieved real-time performance in tests, but model learning still requires a significant amount of time.

Through image detail processing, the resultant style transfer images can be further enhanced. To provide a more realistic picture style transfer effect, the network's activities are refined further during the network design phase.

REFERENCES

[1] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, WardeFarley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In: Proceedings of the 2014 Conference on Advances in Neural Information Processing Systems 27. Montreal, Canada: Curran Associates, Inc., 2014. 2672−2680.

[2] Dong H, Neekhara P, Wu C, et al. Unsupervised image-to-image translation with generative adversarial networks [J].2017.

[3] Huang X, Liu M Y, Belongie S, et al. Multimodal unsupervised image-to-image translation[A].// Proceedings of the European Conference on Computer Vision [C]. 2018:172-189.

[4] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation [J]. Springer, Cham, 2015.

[5] YU X, FATIH P. Imagining the Unimaginable Faces by Deconvolutional Networks [J]. IEEE Transactions on Image Processing, 2018:2747-2761.

[6] Odena A, Dumoulin M et al. "Deconvolution and Checkerboard Artifacts", Distill, 2016.http://doi. org/10.23915/distill.

[7] Gupta A, Zou J. Feedback GAN for DNA Optimizes Protein Functions [J]. Nature Machine Intelligence, 2019, 1(2): 105-111.

[8] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generat ive adversarial networks [J]. Communications of the AC M, 2020, 63(11): 139-144.

[9] Liang X, Chen L, Nguyen D, et al. Generating synthesized computed tomography (CT) from conebeam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy [J]. Physics in Medicine & Biology, 2019, 64(12): 125002.

[10] Zhu J-Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle consistent adversarial networks [C]//IEEE international conference on computer vision, 2017: 2223-2232.

[11] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved Training of Wasserstein GANs [C]// Advances in Neural Information Processing Systems, Long beach, USA,2017: 5769-5779.

An Improved Hybrid Path Planning Algorithm in Indoor Environment

Jiaxiang Fang
School of Computer Science & Engineering
Xi'an Technological University
Xi'an, China
E-mail: 2293950221@qq.com

Shuping Xu
School of Computer Science & Engineering
Xi'an Technological University
Xi'an, China
E-mail: 563937848@qq.com

***Abstract*—During the path planning of robots in the indoor unstructured complex environment, there are often problems such as unreachable target points, deflection in the planning process, and failure to avoid dynamic obstacles in time. To solve these problems, an improved hybrid indoor path planning algorithm was proposed, wherein the improved global path planning algorithm was effectually integrated with improved local path planning algorithm. Firstly, the heuristic factor of traditional A-Star algorithm was optimized, search range and nodes were reduced, and then the path generated by traditional A-Star algorithm for path planning was smoothed using the angle bisector tangent point method. Secondly, combining path and environment information, local path planning was undertaken by utilizing the improved artificial potential field algorithm, and the unreachable target points problem was addressed by adjusting the repulsive field parameters. Additionally, dynamic potential field function was constructed to make it have the ability to resolve dynamic obstacles. Finally, in the part of actual environment verification, a comparison was made in this paper to assess the performance of the traditional hybrid algorithm against the improved algorithm in terms of path planning. The consequences showed that, by the hybrid algorithm proposed in this paper, the path planning length was reduced by 10.3%, the running time was decreased by 12.5%, and 34 redundant nodes were eliminated. The consequences indicated that the hybrid algorithm can effectively address the indoor unstructured and complex path planning problems.**

Keywords-Mobile Robot; Path Planning Technology; A-Star Algorithm; Artificial Potential Field Algorithm; Autonomous Obstacle Avoidance

Driven by the swift evolution of technology, indoor mobile robot path planning has become a topic of keen interest in the research community. In essence, mobile robot path planning is in the case of excluding human manipulation, the mobile robot identifies and processes the data obtained by the sensor, and calculates an optimal path which is safe and collision-free at the same time [1]. At present, the widely used global path planning algorithms currently include A-Star algorithm [2], D-Star algorithm [3], etc. Frequently applied algorithms for local path planning are the artificial potential field method [4], dynamic sliding window method [5], fuzzy logic method [6], etc. A-Star algorithm is a well-known global path planning algorithm, which is a heuristic algorithm [7] that mainly uses heuristic information to find the optimal path [8]. The optimal path is selected by the artificial potential field algorithm in a manner that the potential function decreases within the obstacles force field.

The A-star algorithm excels in its direct search methodology, effectively delivering satisfactory planning solutions, but it is plagued by poor real-time performance and issues such as deviation caused by turning angles. The artificial potential field method stands out as a prominent local path planning technique. It boasts simplicity in calculation and analysis, easy control, and superior real-time performance. However, it is prone to issues such as local minima and unreachable targets. Moreover, both of them cannot get satisfactory results in the treatment of dynamic obstacles. To address these problems, relevant scholars have proposed various improvements and optimizations to the algorithm. For instance, literature [9] described a bidirectional time-efficient A-Star algorithm for path finding, employing a multi-neighbor grid distance calculation scheme to achieve improved efficiency and smoother paths. However, this approach tends to deviate during navigation when dealing with

large, complex maps. Literature [10] avoided generating paths through obstacle grid vertices by adding a priority-based child node generation strategy into the A-Star algorithm, but the path smoothness during navigation is not enough, which has certain security risks. Literature [11] improved the key node selection strategy of A-star algorithm, thus optimizing path planning in static environments to some extent, but still unable to resolve errors in dynamic environments. Literature [12] put forward an iterative searching strategy capable of skipping intermediate nodes, leading to a decrease in the number of accessed nodes and an enhancement in the overall speed of the algorithm. Nevertheless, the path still contains numerous turning points, prone to issues such as path drift. Literature [13] proposed an improved ant colony algorithm in global path search. The smoothness and calculation effect of path planning is more obvious than A-star algorithm, but the amount of data in the search process is too large to be suitable for equipment with poor performance. In terms of artificial potential field algorithms, literature [14] used the improved Pseudo-Dubins curve to smooth the path, which can obtain better local path calculation effect, but the calculation is cumbersome and the operation efficiency is poor. The study in literature [15] introduced a hierarchical modification technique to tackle the dynamic obstacle avoidance challenge in artificial potential fields, resulting in improved obstacle avoidance performance during movement. Nevertheless, it still faces issues of unreachable targets and local optima. Literature [16] proposed a method to introduce the motion direction of the robot as the control variable in the operation of the potential field function, which can effectively reduce the repulsion of obstacles other than the motion direction to the robot, eliminate the inaccessibility of unreachable target points, but increase the risk of robot collision. Literature [17] used a method of adding virtual sub target points to address the local minimum issue in potential field algorithms, but it has the shortcomings of comprehensive path planning error and excessive fold angle.

Through the above research, it can be found that in indoor unstructured complex environments, the traditional A-Star algorithm tends to encounter path slippage and corner problems when dealing with global paths [18]; The traditional artificial potential field algorithm often fails to achieve optimal path planning results during local planning [19]. Therefore, an integrated path planning algorithm was presented in this paper, which combines an angle bisector tangent optimization for the A-Star algorithm with an enhanced artificial potential field method that constructs a dynamic force field. The improved fusion algorithm introduced ideas of angle bisector tangent points to perform global path planning based on the original A-Star algorithm. Moreover, for local planning, the utilization of an improved artificial potential field algorithm optimized the effect of obstacle avoidance when obstacles are detected during path traversal, achieving autonomous obstacle avoidance and overall path optimization. This results in a smoother path trajectory, a larger detection range, and a reduction in possible drift or errors in the path.

I. IMPROVEMENT OF A-STAR GLOBAL PATH PLANNING ALGORITHM

A. Grid map environment modeling

The construction of a grid map is the premise of mobile robot to perform path planning. Grid map is popular among scholars because of its simplicity and easy implementation. In the grid map, each grid contains information about obstacle occupancy, and each occupancy information can be represented by a specific letter, where 0 represents an occupied grid and 1 represents a free grid [20]. The grid size will affect the speed of path planning, so it is important to construct grid map reasonably.

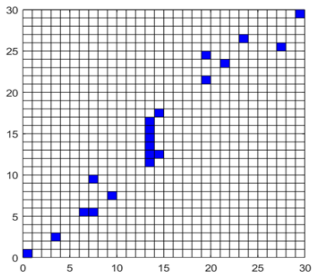


Figure 1. Raster map model

B. Traditional A-star algorithm

A-Star algorithm, as a heuristic search algorithm, incorporates a heuristic factor function into the Dijkstra algorithm to steer its search direction, thus allowing it to compute globally optimal paths in static settings [21]. The evaluation function of the A-Star algorithm is:

$$f(n) = g(n) + h(n) \tag{1}$$

The above formula is the evaluation function of A-Star algorithm. $g(n)$ is the estimated surrogate value from the starting node to the current node. It can usually be expressed by the Euclidean distance between two points.

$$h(n) = \sqrt{(x_{start} - x_{end})^2 + (y_{start} - y_{end})^2} \tag{2}$$

In the above formula, $(x_{start} - x_{end})$ represents the abscissa distance between the current node and the target node, and $(y_{start} - y_{end})$ represents the ordinate distance.

C. Improvement of A-star algorithm

The traditional A-star algorithm can plan an effective optimal path, but there are problems with too many redundant nodes and unsmooth paths. However, heuristic factor can effectively guide the search direction of the A-star algorithm. Therefore, this paper first optimized the heuristic factor function, and then smoothed the A-Star algorithm path.

1) Optimization of heuristic factors

Heuristic factor plays a key role in optimal path planning, which can guide the search direction of A-star algorithm. When $h(n) = 0$, A-star algorithm is equivalent to Dijkstra algorithm; When the estimated generation value of the heuristic factor $h(n)$ is less than the real cost value, the search range of A-star algorithm becomes larger and the number of search nodes becomes more, which can ensure the generation of the optimal path; When the estimated generation value of the heuristic factor $h(n)$ is greater than the actual cost value, the search range of A-star algorithm becomes smaller, the number of search nodes becomes less, which cannot ensure the generation of the optimal path. Seeking more realistic path planning results, the optimized heuristic factor $h(n)$ was introduced as follows:

$$h(n) = \begin{cases} \sqrt{2}|y_s - y_{end}| + |x_s - x_{end}| - |y_s - y_{end}|, |y_s - y_{end}| \geq |x_s - x_{end}| \\ \sqrt{2}|x_s - x_{end}| + |y_s - y_{end}| - |x_s - x_{end}|, |y_s - y_{end}| < |x_s - x_{end}| \end{cases} \tag{3}$$

2) Smooth path processing

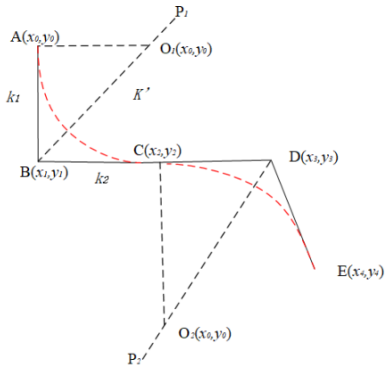


Figure 2. A-Star algorithm smoothing processing diagram

Since the traditional A-Star algorithm has an unsmooth path problem in the path planning process, the optimization method of tangent point of angle bisector was introduced to optimize the A-star algorithm, which helps generate smoother paths. The optimization model is shown in Fig. 2.

Assuming that the initial node position of the mobile robot is $A(x_0, y_0)$, the turning point is smoothed in turn, and the turning points B and D are smoothed in turn until reach the end point $A(x_i, y_i)$. The following are the steps for path optimization:

Step 1: determine if the three points are collinear, that is, if the node is a turning point. Make a vertical line crossing the angle bisector at the previous node of the turning point. In the figure above, the slope of edge AB is k_1, the slope of edge BC is k_2, and the slope of the angle bisector of the turning angle is denoted as $k^{'}$. Where:

$$k_1 = \frac{y_2 - y_1}{x_2 - x_1}, k_2 = \frac{y_3 - y_2}{x_3 - x_2}, k_n = \frac{y_{n+1} - y_n}{x_{n+1} - x_n} \quad (4)$$

According to the slope relation formula $\frac{k^{'} - k_1}{1 + k_1 \cdot k^{'}} = \frac{k_2 - k^{'}}{1 + k_2 \cdot k^{'}}$, the expression of the slope $k_1^{'}$ of BO_1 is as follows:

$$k^{'2} + \frac{2(1 - k_1 k_2)}{k_1 + k_2} \cdot k^{'} - 1 = 0 \quad (5)$$

Then, the coordinates of the intersection point O_1 is:

$$\begin{cases} x_0 = \dfrac{x_1 + k_1 y_1 + k_1 (k^{'} x_2 - y_2)}{1 + k^{'} k_1} \\ y_0 = k^{'} (x - x_0) + y \end{cases} \quad (6)$$

Step 2: make a circle with the vertical length as the radius through the intersection O_1, and judge whether there is an intersection with the tangent circle. The tangent circle equation is:

$$(x - x_0)^2 + (y - y_0)^2 = r_0^2 \quad (7)$$

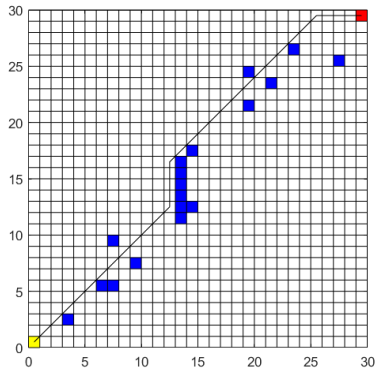
The radius r_0 of the tangent circle is represented as:

$$r_0 = \sqrt{x_0^2 + x_1^2 + y_1^2 + y_0^2 - 2x_0 x_1 - 2y_0 y_1} \quad (8)$$

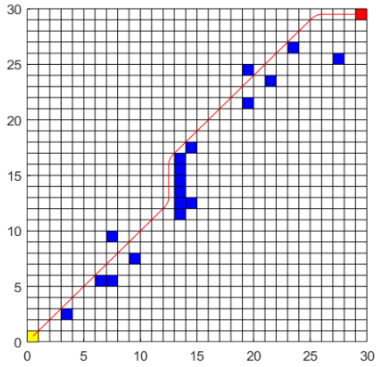
Step 3: determine whether there is a next turning point. If it exists, return to step 1; Otherwise, replace the distance between nodes with an arc.

Step 4: determine whether the optimized path contains every node from the path planned by A-star algorithm. If so, optimization process completes; Otherwise, return to step 1.

The following is a simulation experiment of the improved A-star algorithm on the grid model in Fig. 1 in MATLAB, where each grid in the abscissa and ordinate represents a distance of 1m. The verification results are as follows:



(a)Before path smoothing



(b)After path smoothing

Figure 3. A-Star algorithm path smoothing results in 30×30 environment

As evident in Fig. 3(b), compared to Fig. 3(a), the red trajectory was obviously smoothed and optimized in the trajectories of (12,12), (26,30) two meshes. By comparing various indicators in Table 1, it is evident that the improved A-Star algorithm effectively eliminated turning points, shortened the path length, improved path smoothness, reduced the time required to search for paths, and greatly reduced the number of nodes that need to be expanded during the path planning

process. This shows that the A-Star algorithm modified with the heuristic factor surpasses the traditional A-Star in smoothness of path planning, reliability, and in judging actual trajectories.

TABLE I. COMPARISON OF THE EFFECTS OF A-STAR ALGORITHM IMPROVEMENT

Algorithm	Path length/m	Time for path finding/s	Number of expansion nodes	Is there a turning point
A-Star	22.42	5.9	166	Yes
Improved A-Star	21.56	5.5	59	No

II. IMPROVEMENT OF LOCAL PATH PLANNING ALGORITHM FOR ARTIFICIAL POTENTIAL FIELD

A. Traditional artificial potential field algorithm

Artificial potential field algorithm is a popular choice among local path planning algorithms. Fig. 4 illustrates the force analysis conducted on the robot within this artificial potential field.

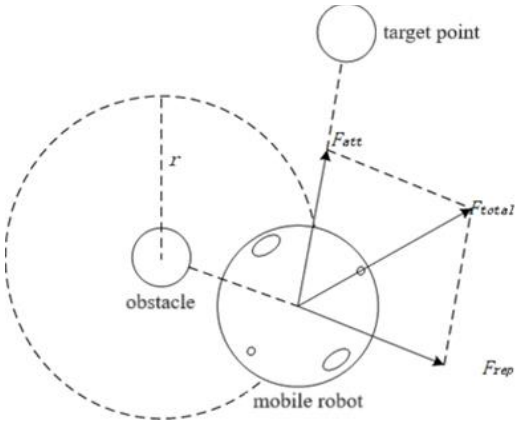


Figure 4. Force analysis diagram of mobile robot

Artificial potential field algorithm's core concept is implemented through the simulation of an imaginary force field, and its theoretical idea can be summarized as follows: the robot is abstracted into a particle with point charge in a virtual force field. The target node generates an attractive potential field that pulls the mobile robot towards it, while obstacles create a repulsive potential field that pushes it away. As the mobile robot draws closer to the target node, the gravitational field intensifies. Similarly, the repulsive force field will also increase when approaching an obstacle. Under the influence of both gravitational and repulsive potential fields, the mobile robot is propelled towards the target node and finally generates an optimal path that can avoid obstacles autonomously [22].

As the mobile robot enters the obstacle's range of influence, it is simultaneously influenced by both the repulsive force F_{rep} of the obstacle as well as the attractive force F_{att} of the target point, resulting in the total force F_{total} acting on the mobile robot, which determines its actual direction of movement [23]. For the artificial potential field represented in Fig. 4, the virtual force field size can be expressed in terms of the negative gradient of the potential field. Specifically, assuming that mobile robot is located in space m, the repulsive force received by mobile robot continues to increase when it approaches the obstacle, and the gravitational force it receives gradually decreases when it approaches the target point. Thus, the potential field function of mobile robot can be expressed as:

$$\vec{E}_{total}(m) = \vec{E}_{att}(m) + \vec{E}_{rep}(m) \tag{9}$$

In the above formula, $\vec{E}_{total}(m)$ is the total potential field force function of the robot located at m, $\vec{E}_{att}(m)$ is the gravitational potential field function, and $\vec{E}_{rep}(m)$ is the repulsive potential field function. Similarly, below is the definition of the repulsive force field function utilized by mobile robots:

$$\vec{F}_{rep}(m) = -\nabla \vec{E}_{rep}(m) = \begin{cases} \frac{p}{2} k_{rep} (\frac{1}{m - m_{pobs}} - \frac{1}{r})^{p-1} \cdot \frac{1}{m - m_{pobs}^{\ 2}} & , m - m_{pobs} \le r \\ 0 & , m - m_{pobs} > r \end{cases} \quad (10)$$

In the above equation, k_{rep} is the gain coefficient of the repulsive field, $m - m_{pobs}$ is the Euclidean distance between the current robot position and the obstacle, r is the repulsive radius, and p is an adjustable parameter.

Due to the presence of complex obstacle environments in practical environments, the total force field function can be modified to the sum of the gravitational function and the combined repulsive function, namely:

$$\vec{E}_{total}(m) = \vec{E}_{att}(m) + \sum_{i=1}^{n} \vec{E}_{rep}(m) \quad (11)$$

The net force experienced by the mobile robot at point m in space can be expressed as:

$$\vec{F}_{total(m)} = -\nabla \vec{E}_{total}(m) = \vec{F}_{att}(m) + \sum_{i=1}^{n} \vec{F}_{rep}(m) \quad (12)$$

B. Improvement of artificial potential field algorithm

Through the above analysis, it can be seen that the artificial potential field algorithm has the advantages of easy implementation and high performance, but it is also prone to the problem of unreachable target points, and it cannot play a good role in identifying and avoiding local dynamic obstacles [24]. To address these limitations of the traditional artificial potential field algorithm, this paper proposed solutions. Firstly, the repulsion field parameters were modified to resolve the problem of unreachable target points. Secondly, to enable the algorithm to handle dynamic obstacles, a dynamic potential field function was constructed. This approach solved the problem of dynamic obstacle avoidance.

1) Correction of repulsion field parameters

Although the traditional artificial potential field algorithm is relatively easy to implement, when there are obstacles around target point and target point is within the range of obstacles influence, there may be a phenomenon where the target point is unreachable. To make the magnitude of the repulsive force change as the gravitational force changes with the distance, the repulsive force function was modified by referring to the gravitational potential field function. By introducing the relative position $(m - m_{end})^p$, the repulsive force decreases continuously when it approaches the target point. The modified repulsion field function can be expressed as:

$$\vec{E}_{rep}(m) = \begin{cases} \frac{1}{2} k_{rep} (\frac{1}{m - m_{pobs}} - \frac{1}{r})^2 (m - m_{end})^p & , m - m_{pobs} \le r \\ 0 & , m - m_{pobs} > r \end{cases} \quad (13)$$

In the above formula, k_{rep} is the repulsion field gain coefficient, $m - m_{pobs}$ is the Euclidean distance between robot position and the obstacle, r is the repulsion radius, and p is an adjustable parameter. Based on the traditional gravitational potential field function, the relative position $(m - m_{end})^p$ was introduced and the repulsive field parameters are adjusted. This made the repulsive force of the mobile robot decreased as it approached the target point.

With the negative gradient of the repulsive force field function signifying its intensity, the repulsive force can be expressed as:

$$\vec{F}_{rep}\left(m\right)=-\nabla\vec{E}_{rep}\left(m\right)=\begin{cases}\vec{F}_{rep1}+\vec{F}_{rep2} & ,m-m_{pobs}\le r\\ 0 & ,m-m_{pobs}>r\end{cases} \tag{14}$$

Where:

$$\vec{F}_{rep1}=k_{rep}\left(\frac{1}{m-m_{pobs}}-\frac{1}{r}\right)\frac{\left(m-m_{end}\right)^{p}}{m-m_{pobs}{}^{2}} \tag{15}$$

$$\vec{F}_{rep2}=-\frac{p}{2}k_{rep}\left(\frac{1}{m-m_{pobs}}-\frac{1}{r}\right)^{2}\left(m-m_{end}\right)^{p-1} \tag{16}$$

The repulsive force $\vec{F}_{rep1}$ points to the direction of the robot from the obstacle. The repulsive force $\vec{F}_{rep2}$ points to the target point from the robot.

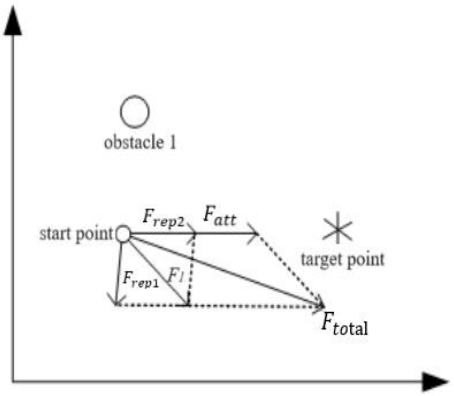


Figure 5. Modified repulsion field parameters force analysis

The force analysis after adjusting the repulsion field parameters is shown in Fig. 5, where the x-axis and y-axis represent the distance traveled, measured in meters.

2) Construction of dynamic force field

The artificial potential field being a blend of gravitational and repulsive fields, therefore, it is logical to divide the dynamic potential field into two sections: a gravity field based on relative velocity and a repulsion field based on relative velocity.

Gravity field based on relative velocity: in an indoor unstructured environment, a mobile robot faces the challenge of path planning in the presence of dynamic obstacles. By adding a relative velocity term to the traditional gravitational potential field function, a dynamic gravitational potential field function is formed. The function for the relative velocity gravitational field can be formulated as:

$$\vec{E}_{att}\left(m,v\right)=\frac{1}{2}k_{att}\left(m-m_{end}\right)^{p}+\frac{1}{2}k_{att}\left(v-v_{end}\right)^{p} \tag{17}$$

In the above formula, $\left(v-v_{end}\right)$ is the relative speed of the mobile robot to the target point in space m. k_{att} is the gravitational gain coefficient, p is the adjustable parameter, m_{end} is the endpoint coordinates $\left(x_{end},y_{end}\right)$, m is the robot current point coordinates $\left(x,y\right)$, and $\left(m-m_{end}\right)$ represents the Euclidean distance between point m and the end point.

Then the gravity function is expressed as:

$$\vec{F}_{att}(m,v) = -\nabla \vec{E}_{att}(m) = \frac{p}{2} k_{att} \left(m - m_{end}\right)^{p-1} + \frac{p}{2} k_{att} \left(v - v_{end}\right)^{p-1} \quad (18)$$

The gravity $\vec{F}_{att}(m)$ points to the target point, and the size of the velocity gravity function $\vec{F}_{att}(v)$ is related to the relative velocity between mobile robot and target point.

Repulsion field based on relative velocity: similar to the dynamic gravitational force field, the relative velocity term is also added in the construction of the dynamic repulsive field. Additionally, considering that there may be dynamic obstacles in the environment, the relative velocity based on obstacles is added. Thus, the repulsion field function based on relative velocity is as follows:

$$\vec{E}_{rep}(m,v) = \begin{cases} \vec{E}_{rep}(m) + \vec{E}_{rep}(v), m - m_{pobs} \le r \ and \ 0 \le v \\ \vec{E}_{rep}(m) \quad , m - m_{pobs} \le r \ and \ 0 \le v \\ 0 \quad , m - m_{pobs} > r \end{cases} \quad (19)$$

In which,

$$\vec{E}_{rep}(m) = \frac{1}{2} k_{rep} \left(\frac{1}{m - m_{pobs}} - \frac{1}{r} \right)^2 \left(m - m_{end}\right)^p \quad (20)$$

$\vec{E}_{rep}(m)$ is the repulsion field function of the mobile robot in space m, $\vec{E}_{rep}(v)$ is the velocity field function. $\left(v - v_{pobs}\right)$ represents the moving speed of the mobile robot relative to the obstacle. e represents a unit vector. Then the repulsion function is:

$$\vec{E}_{rep}(v) = k_{att} \left(v - v_{pobs}\right)^e \quad (21)$$

$$\vec{F}_{rep}(m,v) = -\nabla \vec{E}_{rep}(m,v) = \begin{cases} \vec{F}_{rep}(m) + \vec{F}_{rep(v)}, m - m_{pobs} \le r \text{ and } 0 \le v \\ \vec{F}_{rep}(m) \quad , m - m_{pobs} \le r \text{ and } v \le 0 \\ 0 \quad , m - m_{pobs} > r \end{cases} \quad (22)$$

In which, $\vec{F}_{rep}(m) = \vec{F}_{rep1} + \vec{F}_{rep2}$:

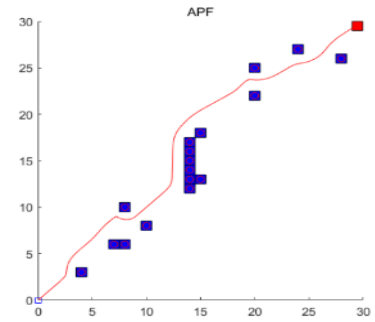
$$\vec{F}_{rep1} = k_{rep} \left(\frac{1}{m - m_{pobs}} - \frac{1}{r} \right) \frac{\left(m - m_{end}\right)^p}{m - m_{pobs}^{\ 2}} \quad (23)$$

$$\vec{F}_{rep2} = -\frac{p}{2} k_{rep} \left(\frac{1}{m - m_{pobs}} - \frac{1}{r} \right)^2 \left(m - m_{end}\right)^{p-1} \quad (24)$$

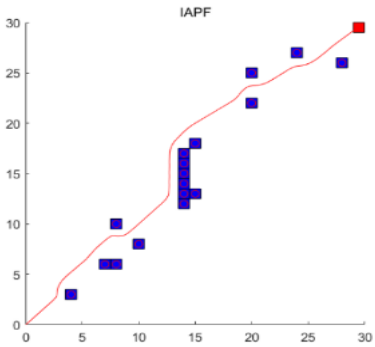
$$\vec{F}_{rep}(v) = -k_{rep}^{\ e} \quad (25)$$

For the purpose of confirming the validity of the improved algorithm, a 30m×30m grid was constructed in the MATLAB environment, and two sets of experimental comparisons were carried out to analyze the improved artificial potential field algorithm, comparing it with the traditional version. The first set of experiments selected the grid map environment in Fig. 1 for simulation comparison. The unit of horizontal and vertical

axes is m. As depicted in Fig. 6, these are the experimental results.



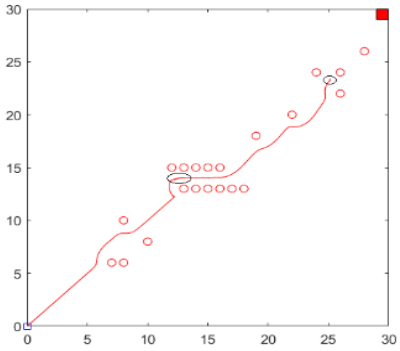
(a) Artificial potential field algorithm



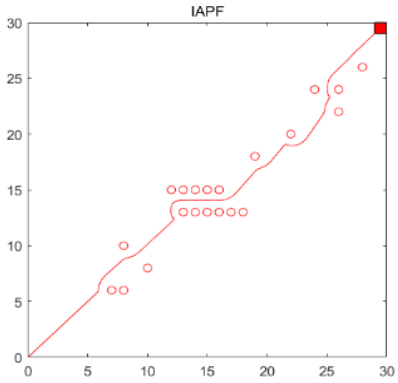
(b) Improved artificial potential field algorithm

Figure 6. Path planning comparison

The second set of experiments was simulated, using red circles to represent obstacles set up, to test the path planning effectiveness of the two algorithms when encountering local obstacles.



(a) Artificial potential field algorithm



(b) Improved artificial potential field algorithm

Figure 7. Complex obstacle test comparison

Fig. 6 reveals that in a static setting, the improved artificial potential field algorithm achieves smoother path planning and more precise obstacle detection, outperforming the traditional approach. However, in the dynamic environment portrayed in Fig. 7, the traditional algorithm fails to navigate to the target due to challenges in obstacle recognition. The lengthy path and extended runtime depicted in Fig. 7(a) are symptomatic of the complex obstacle environment, which poses a challenge for the traditional algorithm. Fortunately, the enhanced algorithm overcomes this limitation.

As Table 2 indicates, by modifying the repulsion field parameters, the challenge of inaccessible target points was conquered, the oscillation phenomenon was effectively eliminated, compared with the traditional artificial potential field algorithm. Its running time was reduced by 13.92%, the path length was reduced by 4%. By way of comparative experiments, the efficacy of the improved artificial potential field algorithm was thoroughly validated.

TABLE II. COMPARISON RESULTS OF IMPROVED ALGORITHM

Experiment Name	Algorithm	Path length/m	Run time/s	Number of cycles
Path planning testing	APF	49.970710	6.186677	447
	IAPF	48.003037	5.430491	440
Complex obstacle testing	APF	∞	∞	∞
	IAPF	51.519690	6.801836	451

III. Hybrid Algorithm Simulation Experiment and Result Analysis

A. Principle of hybrid algorithm

In indoor unstructured complex environments, neither the artificial potential field algorithm nor A-star algorithm is capable of accomplishing optimal path planning on its own. Consequently, this paper integrated these two path planning algorithms into a hybrid solution that combines the advantages of both and can complete the optimal path planning task.

During the implementation process of the hybrid path planning algorithm, the following two problems should be considered: Firstly, when the mobile robot does not enter the obstacle's radius of influence, the optimized A-star algorithm is utilized to obtain global initial paths for global path planning; Secondly, utilizing this comprehensive global path as a foundation, it is necessary to determine whether there are any reserved nodes generated by A-star algorithm within the obstacle's range of influence. If there are, further reductions should be made. Fig. 7 depicts the model for the hybrid path planning algorithm:

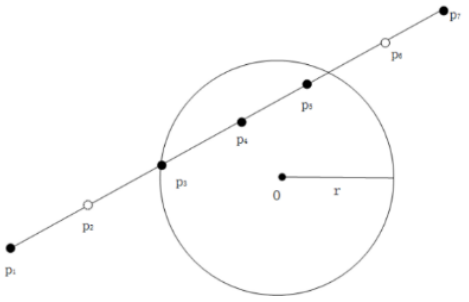


Figure 8. Hybrid algorithm model diagram

In Fig. 8, $\left(p_{1,} \ldots \ldots, p_i\right)$ are the path intermediate node of the A-Star algorithm, and r is the radius of influence of the obstacles. Among them, $p_1, p_2, p_3, p_4, p_5, p_6, p_7$ are the reserved nodes of the optimized A-star algorithm.

Assuming that the robot at time t is located at point p_1, upon advancing to point p_3, the mobile robot comes under the influence of the artificial potential field algorithm. Since points p_4 and p_5 fall within this influence, they are disregarded, and the next node p_7 is taken as the target point.

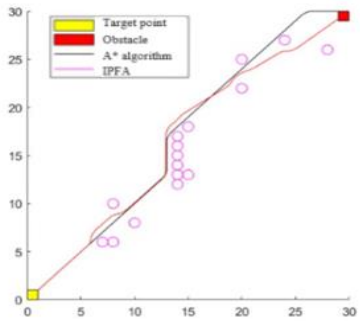
The hybrid path planning algorithm proceeds in the following execution stages:

1) Initialize map parameters;
2) Carry out global path planning utilizing the A-Star algorithm, while documenting and storing the path nodes for future reference;
3) Use node pruning strategy to retain key path nodes;
4) Treat the key path nodes as local target points in turn. If the key target node is within the radius of the obstacle, the node is deleted and the next key node is selected as the target node;
5) With the utilization of the improved artificial potential field algorithm, the robot systematically traverses from its current location to the next sub-target, ensuring accurate path planning;
6) Check if the robot has completed its journey to the final target node. Stop if it has; otherwise, repeat step 5 for further route calculation.

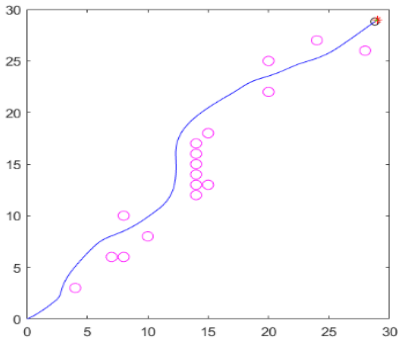
B. Simulation of experimental results of hybrid algorithm

1) Experimental results in static environment

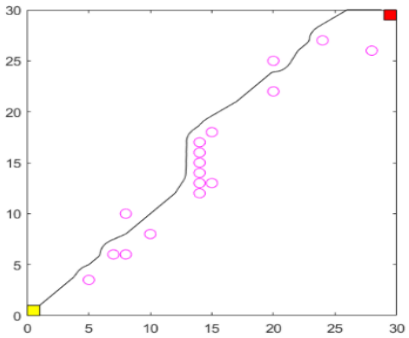
The experiment used MATLAB to conduct simulation tests on the hybrid path algorithm. The experimental environment was also set as the grid map model depicted in Fig. 1. The experiment was carried out in both dynamic and static environments, grouped according to whether the obstacles are movable or not. The figure below demonstrates the experimental results.



(a) the improved A-Star and the improved artificial potential field



(b) DWA



(c) hybrid path planning algorithm

Figure 9. Static path comparison diagram

The above three sets of experiments were conducted in a static environment to compare the path planning information generated by the improved A-star algorithm, the improved artificial potential field algorithm, DWA algorithm, and the hybrid algorithm. The comparison was based on their respective planning path lengths, search path durations and ability to handle dynamic obstacles.

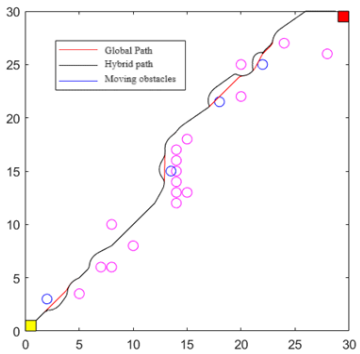
As evident from Table 3 and Fig. 9, in the context of identical obstacles, the A-Star algorithm demonstrates an ability to plan a short path with rapid search speed during the path planning process, but it does not have the ability to dynamically avoid obstacles. Despite its longer planned path and increased search duration, the artificial potential field algorithm boasts the crucial capability of dynamically steering clear of obstacles, making it a practical choice for complex indoor navigation scenarios. By fusing the benefits of the two algorithms, the hybrid approach achieves not only a shorter search path and time but also the flexibility to handle dynamic obstacles effectively.

TABLE III. ALGORITHM COMPARISON IN STATIC ENVIRONMENT

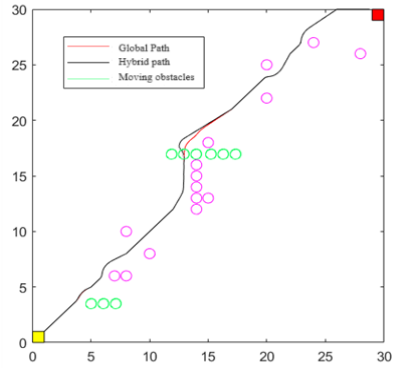
Algorithm	Path length/m	Search time/s	Does the algorithm have the ability to handle dynamic obstacles
A-Star	45.36	6.72	No
IAPF	48.00	10.43	Yes
DWA	48.86	28.21	Yes
Hybrid algorithm	46.54	8.14	Yes

2) Experimental results in dynamic environment

The path planning experiment of the hybrid algorithm was conducted in a 30m×30m environment. In the dynamic environment, the parameters for obstacle avoidance were set to include a step increment of 0.1, a gravitational and velocity gain of 3, an obstacle detection radius of 3 units, and a repulsive gain of 5. The predefined obstacle was programmed to traverse a path from the point (12, 17) to (16, 17). The figure below demonstrates the experimental results.



(a) Dynamic appearance of obstacles



(b)Moving obstacles

Figure 10. Dynamic path planning diagram

The purple circles in Fig. 10(a) represent static obstacles that already exist, while the blue circles represent dynamic obstacles that appear randomly. In Fig. 10(b), green circles represent persistent dynamic obstacles, which move left and right between positions with ordinates of 3 and 17. These images compared the path under the hybrid algorithm with the static simulated global path to show the impact of the introduction of dynamic and moving obstacles on the return of the global path. In Fig. 10(a), the hybrid algorithm can effectively avoid dynamically appearing obstacles in the simulated indoor environment and return to the global path, ensuring the real-time nature of the path planning task. Mobile obstacles were introduced in Fig. 10(b). Mobile robot effectively avoided dynamic obstacles and returned to the global path, which effectively verified the effectiveness of the improved hybrid algorithm. Additionally, this enhancement significantly improved the real-time performance of path planning.

C. Verification of hybrid algorithm experiment in real environment

The robot platform in this paper was built on the Hands-free mobile robot platform. RPLIDAR A1 lidar was selected as the main ranging sensor. To steer the mobile robot's motions precisely, the OpenRE Board controller was selected as the primary control mechanism. According to the composition structure of the Handsfree_Stone_V3 mobile robot, the various hardware parts were assembled. The finished assembly of Handsfree_Stone_V3 is shown in Fig. 11.

The physical environment verification was conducted in a room with dimensions of 4.5 meters in length and width. The laboratory was equipped with obstacles, and the test commenced from the doorway position. As shown in Fig. 12.

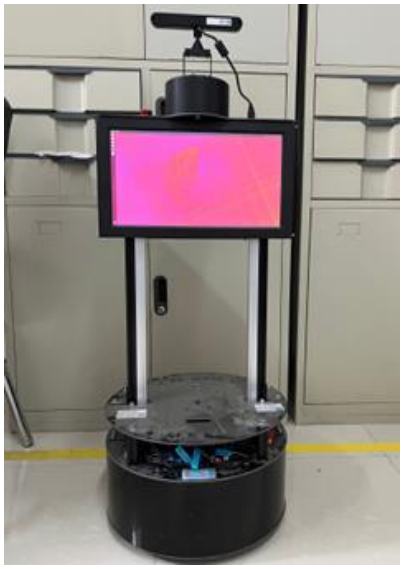
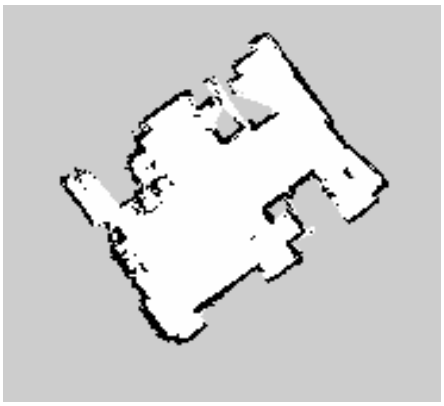


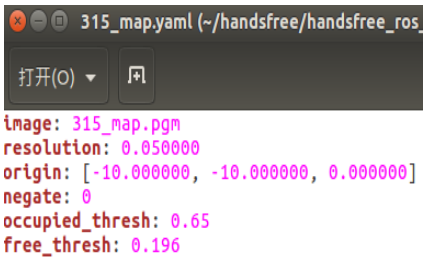
Figure 11. Hands-free robot platform

Figure 12. Actual test scenario

This paper used Gmapping to complete the mapping of the actual environment of the Hands-free robot. The mapping results and mapping parameters are shown in Fig. 13:



(a)Actual environmental mapping



(b)Actual environmental mapping parameters

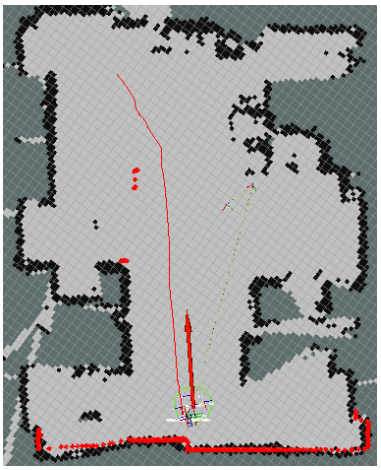
Figure 13. Actual scene construction effect

After completing the mapping of the actual environment as mentioned above, the next step is a comparative analysis of the path planning prowess of the traditional A-Star hybrid DWA algorithm and the improved A-Star hybrid improved artificial potential field algorithm presented herein. The comparison was executed in a real-life setting, mirroring actual environmental conditions. The figure below demonstrates the experimental results.

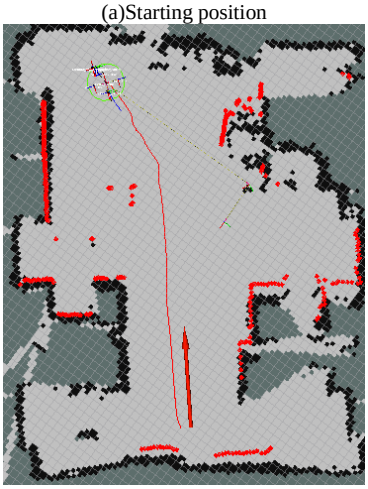
(a)Starting position

(b)Ending position

Figure 14. A-Star Hybrid DWA algorithm path planning



(a)Starting position



(b)Ending position

Figure 15. Improved A-Star hybrid improved artificial potential field algorithm

The navigation effects in Fig. 14 and Fig. 15 and the data in Table 4 show that, under the

condition of setting the same starting and ending points, by comparing the path length, search time and passing nodes number of the two path planning algorithms, it can be found that the hybrid algorithm proposed in this paper reduced path planning length by 10.3%, reduced the running time by 12.5%, and passed through 34 less redundant nodes. This proved that in real environments, the hybrid algorithm introduced in this paper offers distinct advantages in path planning over the traditional hybrid algorithm.

TABLE IV. Results of algorithm comparison in real environment

Path planning algorithm	Path length/m	Number of nodes passed through	Search time/s
A-Star Hybrid DWA	3.66	126	54.42
Hybrid algorithm in this paper	3.24	92	48.36

IV. CONCLUSIONS

The objective of this paper is to primarily identify and enhance the weaknesses in the conventional A-Star algorithm and the artificial potential field algorithm, thereby optimizing their performance. Firstly, the heuristic factor of A-star algorithm was adjusted, and its redundant nodes were eliminated using the node deletion strategy. Simultaneously, the path turning points of A-star algorithm were smoothed. The search time was shortened effectively. Secondly, the deficiencies of the artificial potential field algorithm were addressed in static and dynamic environments separately. In the static environment, the target unreachable problem is solved by modifying the repulsion field parameters. In the dynamic environment, the relative speed term was introduced so that the speed of the mobile robot becomes smaller when approaching the obstacle and becomes larger when it is further away. The simulation results indicated that the hybrid algorithm efficiently addressed the path planning dilemma for mobile robots within a complex and unstructured indoor space. The improved algorithm's running time was reduced by 13.92%, and the planned path length was reduced by 4%. In the part of actual environment verification, the path planning results between the traditional hybrid algorithm and the improved algorithm was compared. The results showed that the path planning length was reduced by 10.3%, the running time was decreased by 12.5%, and 34 redundant nodes were eliminated by the hybrid algorithm presented in this paper, demonstrating greater efficiency than the traditional hybrid algorithm.

ACKNOWLEDGMENT

The authors wish to thank the cooperators. This research is partially funded by the Project funds in Shaanxi province University Student Innovation and Entrepreneurship Fund Project (S202310702059).

REFERENCES

[1] HUO Jiajia, LIU Yu, ZHANG Xuanyi, et al. Research on Obstacle Avoidance and Path Optimization of Intelligent Vehicle [J]. Shandong Industrial Technology, 2017(11):163.

[2] XIN Peng. Research on Mobile Robot Path Planning Based on A* Algorithm [D]. Hebei University of Technology, 2022.

[3] REN Hongge, HU Hongchang, Shi Tao. Dynamic Path planning of Mobile Robot based on improved ant Colony Algorithm [J]. Modern Electronics Technique, 2021, 44(20):182-186. (in Chinese)

[4] KANG Chengbo. Research on Autonomous localization and Navigation System of logistics robot based on Laser SLAM [D]. Shanghai University of Engineering Science, 2020.

[5] Fox D, Burgard W, Thrun S. The dynamic window approach to collision avoidance[J]. IEEE Robotics & Automation Magazine, 2002, 4(1): 23-33.

[6] ZHU Manman, DU Yu, ZHANG Yonghua, et al. Local Path planning for Intelligent Vehicle based on Fuzzy Logic [J]. Journal of Beijing Union University, 2016, 30(04):29-32.

[7] Peter E H, Nils J N, Betram R. A Formal Basis for the Heuristic Determination of Minimum Cost Paths[J]. IEEE Transactions on Systems Science and Cybernetics, 1968, 4(2): 100-107.

[8] HUO Fengcai, CHI Jin, HUANG Zijian, et al. Mobile Robot path planning Algorithm Review (Information Science Edition) [J]. Journal of Jilin University, 2019, (10):19-21.

[9] GAO Mindong, ZHANG Yani, ZHU Lingyun. Bidirectional aging A* Algorithm for Robot Path

Planning [J]. Application Research of Computers, 2019, 36(3):792-795.

[10] GU Chen. Application of Improved A* Algorithm in Robot Path Planning [J]. Electronic Design Engineering, 2014, 22(19):96-98.

[11] WANG Shuaijun, HU Likun, WANG Yifei. Path planning of Indoor Mobile Robot based on improved D ~ * Algorithm [J]. Computer Engineering and Design, 2019, 41(4):1118-1124. (in Chinese)

[12] Duchon F, Babinec A, Kajan M, et al. Path Planning with Modified a Star Algorithm for a Mobile Robot[J]. Procardia Engineering, 2014, 96(96):59-69.

[13] Liu J, Yang J, Liu H, et al. An Improved Ant Colony Algorithm for Robot Path Planning[J]. Soft Computing, 2017, 21(19):5829-5839.

[14] Kaplan A, Kingry N, Uhing P, et al. Time-Optimal Path Planning with Power Schedules for a Solar-Powered Ground Robot [J]. IEEE Transactions on Automation Science and Engineering, 2017, 14(2):1235-1244.

[15] Xu X L, Pan W, Huang Y B. Dynamic Collision Avoidance Algorithm for Unmanned Surface Vehicles via Layered Artificial Potential Field with Collision Cone[J]. Journal of Navigation, 2020, 73(6):1306-1325.

[16] Zhang C. Path Planning for Robot based on Chaotic Artificial Potential Field Method[C]. //International Conference on Advanced Engineering and Technology, 2018, 317.

[17] XU Fei. Research on Obstacle Avoidance and Path Planning of Robots Based on Improved Artificial Potential Field Method [J]. Computer Science, 2016,43(12):293-296.

[18] ZHANG Qiang, LU Shouyin, ZHANG Jiarui, et al. Path planning of inspection robot based on Safety A* algorithm and improved artificial potential field method [J]. Computer Age, 2022, (11):29-33+37.

[19] ZHANG Shengnan. Research on Path Planning and Tracking Algorithm of Mobile Robot [D]. Heilongjiang University, 2021.

[20] ZHOU Shibo. Path planning of wheeled robot based on improved A* algorithm and artificial potential field method [D]. Taiyuan University of Technology, 2021.

[21] LV Jindou. Research on Robot Dynamic Path Planning Based on Improved A~* Potential Field Method [D]. Xi'an University of Architecture and Technology, 2020.

[22] ZHI Jin Zhu. Research on Path Planning of Indoor Mobile Robot in Dynamic Environment [D]. Xi'an Technological University, 2021.

[23] LU Jian. Path Planning and Navigation of Indoor Mobile Robot [D]. Fujian University of Technology, 2022.

[24] Qi B, Li M, Yang Y, et al. Research on UAV path planning obstacle avoidance algorithm based on improved artificial potential field method[J]. Journal of Physics: Conference Series, 2021, 1948(1): 14-16

Fabrication and Development of the Embedded Linux Based on the ARM System

Ruoyu Wang
School of Electronic Engineering
Xi'an University of Posts and Telecommunications
Xi'an, China
E-mail: 2544560286@qq.com

Jiyao Fan
School of Electronic Engineering, Xi'an Aeronautical Institute, Xi'an, 710077, China
E-mail: fjy_stu@126.com

Lulu Chen
School of Mechanical Engineering,
Xihua University,
Chengdu, China
E-mail: chenll1232022@qq.com

Xiaoheng Sun
School of Electronic Engineering
Xi'an University of Posts and Telecommunications
Xi'an, China
E-mail: xhsun_xiyou@126.com

Chenyu Zhang
School of Electronic Engineering
Xi'an University of Posts and Telecommunications
Xi'an, China
Email: zcy_xupt@163.com

Lei Tian
School of Electronic Engineering
Xi'an University of Posts and Telecommunications
Xi'an, China
E-mail: tla02@126.com

***Abstract*—With the development of the embedded system and improvement of the requirements from customers, embedded GUI which was used to communicate between users and embedded system became to be a key in researching the embedded system. This article introduced what is embedded system and what is embedded GUI. It described the porting of an embedded GUI on ARM Linux platform, including the establishment of compile environment, configuring and modifying of compile options and setting of runtime parameters. Above all, this article introduces the development of Embedded Linux GUI, describes the GUI development environment of Qt /Embedded and Qt, one procedure of the app location development based on Qt/Embedded is described in detail. This system adopts the B/S mode and uses the ARM11 development board with S3C6410 as the core. Sensors are used to obtain information, and the obtained data is judged and analyzed. When the preset conditions are not met, an alarm function is triggered. At the same time, data transmission is carried out through the Zigbee wireless module, and the receiving end analyzes and displays the received data. Each device is connected to the Zigbee module, which distinguishes it by assigning different network addresses and constantly transmits the status of the monitoring area to the receiving end. The receiving end distinguishes the monitoring status of different locations based on the different network addresses, increasing mobility.**

Keywords-Virtual System; Embedded System; GUI; Porting; Qt; Qt/Embedded

I ESTABLISHMENT OF VIRTUAL MACHINE AND INSTALLATION OF LINUX

A. The concept of virtual machines and virtual machine software

The so-called virtual computer (abbreviated as virtual machine) is actually an application software [1]. VMware, the narrow virtual machine software that will be introduced to you today, is actually just an application software. The virtual machine created by VMware is almost identical to the real computer [2-3]. The currently popular virtual machine software includes VMware and Virtual PC. This article is based on VMware Workstation to conduct various practical exercises [4-5].

1) Establishment of virtual machines

After clicking "New Virtual Machine". Click the "Next" button to enter the virtual machine configuration interface, where there are two options: one is the "Typical" method; The second is the "custom" method. We choose a typical approach here.

Click the "Next" button to enter the virtual machine operating system selection interface, where you can click "Linux". Select "Red Hat Linux". and so on, the user will be asked to set the size of the hard drive, which is 10GB.

2) Installing Linux on a Virtual Machine

The method to install Linux on a virtual machine is actually very simple, as follows [6-7]:

Before installation, we need to set up the optical drive of the virtual machine. At the beginning of the installation process, a welcome dialog box appears and the user presses “Enter” to continue. Red Hat will sequentially ask users what language they are using, the type of keyboard they are using, and the location where the software is installed. Select “Install” in the subsequent upgrade or system installation inquiry, and choose “Custom Installation” for the type of installation to use.

Now users need to create two partitions in the reserved hard drive space. The first partition serves as the root partition of Linux and is used to install Linux files. The second partition serves as a swap partition to supplement the user's physical memory, and it is recommended to set it to one to two times the amount of memory. Here, users will be prompted to set up some hardware, such as mouse, network, Linux Loader and Configurator.

B. Communication between virtual and board

The VMware Workstation software comes with the function of sharing host files. The specific method is to select Virtual Machine→Settings→Options in the menu, where one option is to share a folder. Then, you select the folder you want to share, click OK, and the file will appear in a directory of the virtual machine.

The video monitoring and alarm system based on wired and wireless data transmission is an embedded system developed based on the S3C6410 development board. This system is based on pure video monitoring, combined with data collected from human infrared sensors, sound sensors, light sensors, temperature and humidity sensors, to determine whether an alarm is needed. The data is transmitted through the Zigbee. wireless module in the wireless system to complete the data transmission of the monitoring system. The embedded board is in Figure 1.

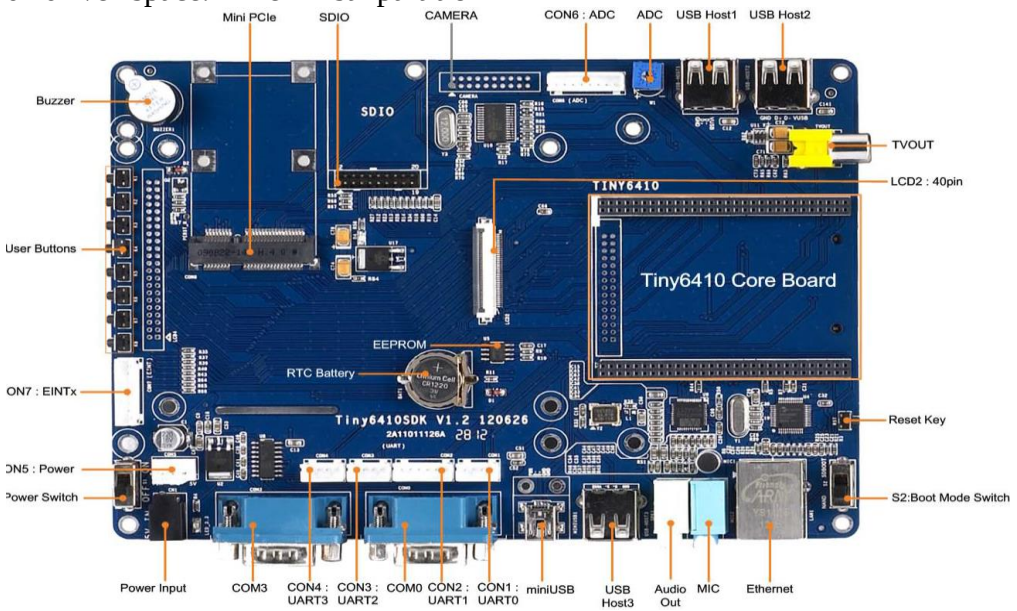


Figure 1. The Embedded Development Board

In Figure 1, the system is the core board which based on the S3C6410 ARM11 chip as the main processor. It integrates 256M of DDR RAM, 2GB of NAND Flash memory, and is powered by 5V. It can be connected to a PC through the COM1 port to access the development board. It can be connected to the Zigbee wireless module through the COM3 port for information transmission and reception, and has three USB interfaces that can be used to connect USB cameras for real-time video monitoring. The core board also comes with two rows of SDIO interface arrays, which can self-connect the required modules. This system requires the application of SDIO ports, human infrared sensor modules, sound sensor modules, temperature and humidity sensor modules, and light sensor modules. The core board comes with LED lights and PWM modules. This article uses an external USB camera and various sensors to obtain data from the monitoring site. It is wirelessly transmitted through the external Zigbee, and the alarm function is achieved through the system's integrated LED lights and PWM.

II EMBEDDED LINUX KERNEL AND FILE SYSTEM

A. Establish the Cross Compilation in S3C6410

The S3C6410 microprocessor is a low-power, highly integrated ARM920T core-based microprocessor designed by SAMSUNG for handheld devices. The development board selected in this article is YF6410II, which uses the S3C6410 microprocessor, making it suitable for developing high-performance handheld and portable smart devices or terminals.

Copy YUANFENG.tar.gz from the Linux directory on the development board's CD to the root directory. The execute script is:

[root@localhost YF6410] #./YFINSTALL.sh

After the script file is executed, the compiled development environment (arm Linux gcc-2.95.3) is successfully installed, and a tftp server is also installed. The default working directory for tftp server is/tfpboot

Select load configuration file and enter the path to the configuration file in the dialogbox. Next, select the LCD option is 3.9TFT. The touch screen is used in the design, so it is necessary to modify kernel/drivers/char/s3c6410 ts. c. *Download file system: YFLoader#load flash 0x40 0x1b00000 root.*

Turn off the power to the development board and connect the LCD. Restart the development board, and the calibration program for QT will appear on the LCD. After calibration, enter the QT interface.

B. Signal and slot principle

The signal and slot provide a mechanism for communication between objects, and the QT designer can complete it. GUI programs can respond to user actions. When a user clicks on a menu item or toolbar button, the program will execute some code. More generally, various objects need to be able to communicate with each other. Programmers must associate events with relevant code. Trolltech has invented a solution called "Signal and Slot". The signal and slot mechanism are a powerful object to object communication mechanism that can completely replace the rough callback and message mapping of traditional toolkits.

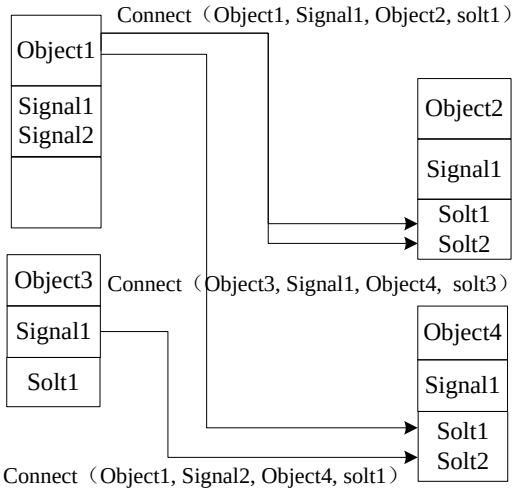


Figure 2. Flowchart of Signal to Slot Connection

In Figure 2, when using the old callback mechanism to connect some code with a button, a function pointer needs to be passed to the button. When the button is clicked, the function is called. Old toolkits cannot guarantee that the parameters given when a function is called have the correct type, making it easy to crash. Another issue with callback methods is that they tightly bind GUI

elements with their functionality, making it difficult to independently develop classes.

The signal and slot mechanism of QT are different. When an event occurs, the components emit signals. For example, when a button is clicked, a "clicked" signal is sent. By creating a function (slot) and calling the connect () function to connect signals, programmers can connect signals to slots. The signal and slot mechanisms do not require classes to know each other, making it easier to develop highly reusable classes.

III PROGRAMMING UNDER LINUX SYSTEM

A. Common System Instruction Set for Linux

For beginners of Linux, the biggest headache at first may be not being able to find the software they need to use. In fact, there are also some commonly used software under Linux that have the same functions as similar software under Windows. Linux's vi and Windows's Edit are very similar, and Open Office has almost all the functions of MS Office. The comparison of command lines in Linux and DOS systems is shown in Table 1.

TABLE I. COMPARISON TABLE OF COMMONLY USED COMMANDS IN LINUX AND DOS

Command	***MS-DOS***	***Linux***	***Example of Linux***
Using an editor to edit files	edit	vi	vi thisfile.txt
Copying files	copy	cp	cp thisfile.txt
Transfer files	move	mv	mv thisfile.txt
List files	dir	ls	ls
Delete files	del	rm	rm thisfile.txt
Create directory	mkdir	mkdir	Mkdir directory
Using the specified path	Cd path	Cd path	cd directory

From the Table 1, many Linux commands typed at Shell prompts are similar to commands typed in DOS. In fact, some commands are exactly the same. Table 1 provides commonly used commands under Windows DOS prompts and equivalent commands in Linux. The appendix also provides a simple example of how to use these commands at Linux Shell prompts. Please note that these commands usually have many options, and to further learn each command, please read the relevant manual (man) page (for example, typing man ls at the shell prompt can read information about the ls command).

If the user inserts a USB drive or CD under Windows, Windows may recognize them directly, but in Linux, they must be mounted. Under Red Hat Linux, the command mount is generally used to mount devices. Mount lists all partitions of the system.

B. Common software under Linux

VI is a widely used full screen document editor in the Unix world, and it can be said that almost any Unix machine will provide this software. Qt is a cross platform C++graphical user interface library produced by TrollTech in Norway. It currently includes Qt, Framebuffer and the Qt designer. Qt supports all Unix systems, including Linux. It also supports WinNT/Win2000 and Win95/98 platforms. Qt Designer is a tool used to design and implement user interfaces that can be used on multiple platforms. It can simplify user interface design experiments.

KDevelop is a fast window development tool in the X system. KDevelop itself does not include a compiler, and it uses the GNU compiler suite to generate executable code.

C. Design and Compilation of Qt Window

In order to have a more direct understanding of Qt programming, here is an example for everyone. Run Qt Designer, click on the menu File→new, and create a new project. Select the save path and file name hello.pro, then click on the menu File→New, select C++Source File, confirm, and enter the following content. Using the Qmake for program compilation.

Name the source file hello.cpp. Then open the terminal, enter the directory where the source file is located. At this point, a hello executable file will be generated. Then run it: #/ Hello. The results are as follows in Figure 3.

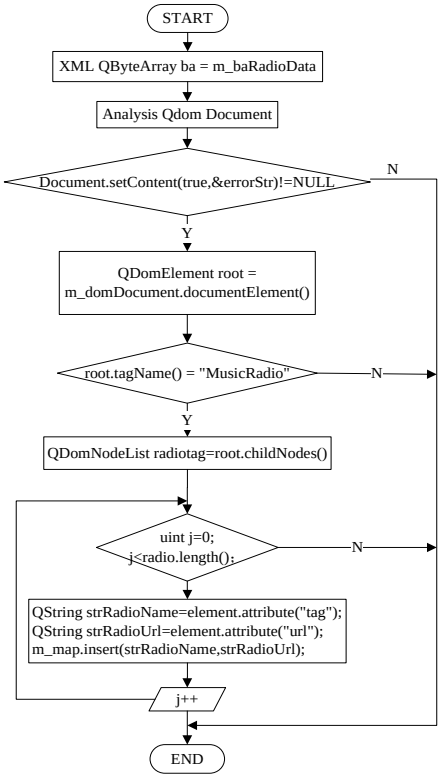


Figure 3. The flowchart of the XML file parsing

As shown in the Figure 3, in order to provide more convenient engineering management for user applications, the system parses XML files. A project includes pro, ui, ui.h and the main function. cpp. Among them, the role of engineering files is to manage all files and their relationships in the current project, while the role of form files is to manage various interface elements and their properties in the form. In addition to reading and writing to. ui files, Qt Designer also adds a form ui.h file. It is a regular C++ file, mainly used for customizing slot methods the ui.h file is the responsibility of Qt Designer to maintain consistency with the slots related to the form. Whenever a user adds, removes, or changes the connection of a slot to the form, Qt Designer will automatically update it the content of the ui.h file.

IV PORTING AND DEBUGGING OF QT GRAPHICS PROGRAM

A. Establishment of Qt/Embedded Development Environment

Due to Qt/Embedded using Framebuffer as the underlying graphics interface, in order to use Qt/Embedded [8]-[10]. It is necessary to start the Framebuffer driver program. Because by default. Linux does not configure frame buffer drivers [11]-[13]. Therefore, it is necessary to enable frame buffering under Linux. The specific methods are as follows:

Log in with root privileges to the/boot/grub directory, type the vi gruB.conf command to modify the gruB.conf file. If you restart and enter, you will see the cute little penguin in the upper left corner of the screen, which indicates that Linux has enabled the frame buffer driver [14-15].

B. The process of establishing a Qt/Embedded development environment

Qt/Embedded can create different development environments for different platforms. In this article, as the ARM system is used, it is necessary to establish an X86 development environment and ARM development platform. The specific methods are as follows. Here, the environment to run on the X86 platform. Run the following command in Linux command mode, it is show in Figure 4.

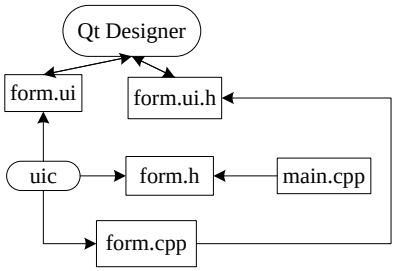


Figure 4. Working principle of all engineering files

After working the Qt files, then run the following command in Linux command mode. The specific usage of compilation options is feasible/ View the configure help and make help commands. According to the developer's own development environment, other parameters can also be added to the configure parameters, such as - *no opengl or - no xft.* The installation of QT/X11 is mainly to provide designer tools and qvfb tools to QT/E.

In fact, the method for establishing the QT/E environment of ARM is basically the same as that for establishing the environment of X86. Just because the libraries used on the ARM platform and X86 are different, it is necessary to add

corresponding parameters during compilation. Here, it is necessary to:

If the program is written using an interface designed by Qt Designer, the *. ui file needs to be converted to*. h file and *. cpp file. Then write a *. pro file (used to generate Makefile files), which has a relatively fixed format.The basic format of the hui. pro file is as follows. Here, runs the application on the host. Before compiling and executing QT/E, it is necessary to first set up the TMAKE and QT/E LIB environments.

After entering the directory where QT/X11 is installed and execute the program qvfb in the bin directory. Then execute the QT/E program, the corresponding program display will appear in the window opened by the qvfb program, it is showed in Figure 5.

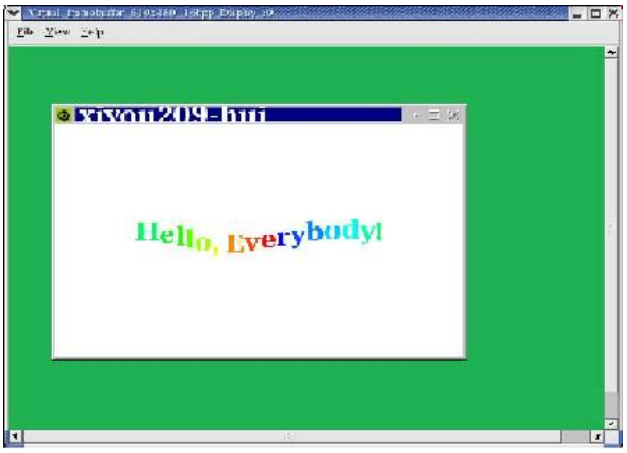


Figure 5. The execution result of hello

Compile programs for the development board. Similarly, the first step is to set up the TMAKE and QT/E LIB environments, with basic parameters identical to those on the host. There is only a difference when setting the TMAKEPATH path for tmake and the LD-LIBRARY-PATH path for QT/E. Need to convert:

C. The Embedded Development Board Control Platform

Due to the design of the remote monitoring and alarm system in this article, it is necessary to display the data obtained from modules such as USB cameras, temperature and humidity sensors, human infrared sensors, light sensors, sound sensors, etc. connected to the development board on a PC through a web page. Therefore, a web server should be established on the development board to enable it to run web programs, and the PC can access the data on the development board. The specific physical image is shown in Figure 6.

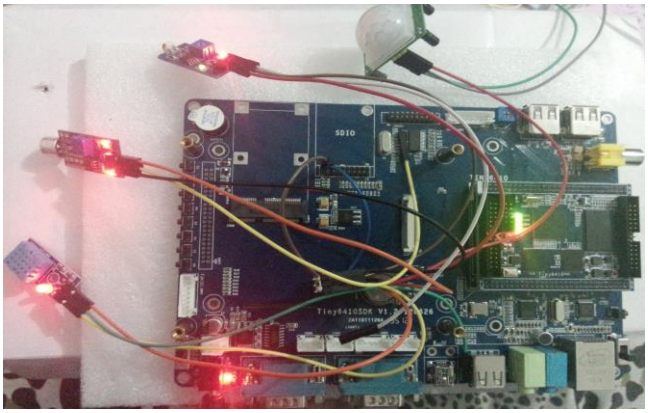


Figure 6. The specific physical chart

After the system is powered on, when the network addresses of multiple modules are in the same network segment, the modules automatically form a network, and the coordinator automatically assigns addresses to the routing. It can also be configured through module configuration software. If the module loses power, the network can be automatically repaired after power on. Firstly, the configured coordinator receives information from the serial port and automatically sends it to all routers in the network; After obtaining data from the serial port, the router will forward it to the coordinator; At this point, users can transmit to each other through specified addresses between any two routers or coordinators.

The Zigbee wireless transmission module mainly achieves transparent data transmission from serial port to Zigbee wireless through the Zigbee wireless transmission module. It is based on TI's CC2530 chip, which adds RF transceiver and runs the ZIGBEE 2007/PRO protocol stack internally, with all the characteristics of the ZIGBEE protocol. For the complex Zigbee protocol, the transparent transmission Zigbee module embeds the protocol stack into the module and, in conjunction with the UART to RS232 motherboard, converts the UART of the Zigbee module into a standard RS232 interface. It can be directly connected to the serial port of the

development board to transmit and receive data through the read/write serial port.

V Conclusions

So far, our porting work has been successfully completed. Qt has successfully run on our S3C6410 platform. Due to the fact that embedded systems belong to resource limited systems, embedded GUI development must also consider the issue of system overhead, which requires careful analysis of specific applications to meet the specific needs of users. Qt/Embedded continues all the functions of Qt in desktop systems, with rich API interfaces and component.

This paper designs a monitoring system based on wired transmission, a monitoring system based on wireless transmission, and a wired wireless joint debugging system. This system is a low-cost network monitoring embedded system based on ARM11's S3C6410 as the core. This system utilizes embedded devices, such as the S3C6410 development board. The early warning system studied through this ARM development board is based on a pure alarm system. In addition, data collected from human infrared sensors, sound sensors, light sensors, temperature and humidity sensors, and images captured by USB cameras are sent through wired or Zigbee wireless modules to determine whether an alarm is needed. And porting the client to Windows by accessing the development board to access the server meets the needs of more users.

Acknowledgment

This work was partly supported by the research project of Xi'an University of posts and telecommunications teaching reform JGA202304 and the undergraduate innovation and entrepreneurship plan 202311664058.

References

[1] N. Jagadeeswari and V. M. Rai, "Homogeneous Batch Memory Deduplication Using Clustering of Virtual Machines," Computer Systems Science and Engineering, vol. 44, pp. 929-943, June 2022.

[2] L. Huibo, W. Haiqiao, L. Ping and G. Peng, "VMware-based Hardware-in-the-loop Simulation Approach for Link11 Data Link," Binggong Xuebao/Acta Armamentarii, vol. 41, pp. 224-233, June 2020.

[3] L. Ying. "Lightweight information integration model of IOT based on VMware cluster technology," International Journal of Information and Communication Technology, vol. 18, pp. 41-56, December 2021.

[4] M. A. A. Mohamed, "VFST: Virtual and fully software-based toolchain for PC interfacing education and research," Computer Applications in Engineering Education, vol. 31, pp. 389-407, March 2023.

[5] A. Bhonde and S. R. Devane, "Priority-based security-aware virtual machine allocation policy," International Journal of Information and Computer Security, vol. 23, pp. 40-56, February 2024.

[6] N. Karapetyants and D. Efanov, "A practical approach to learning Linux vulnerabilities," Journal of Computer Virology and Hacking Techniques, vol. 19, pp. 409-418, September 2023.

[7] L. Kunli, L. Wenqing, Z. Kun and T. Bibo, "HyperPS: A Virtual-Machine Memory Protection Approach Through Hypervisor's Privilege Separation," IEEE Transactions on Dependable and Secure Computing, vol. 20, pp. 2925-2938, July 2023.

[8] A. K. George. "Real-time performance analysis of distributed multithreaded applications in a cluster of ARM-based embedded devices," International Journal of High-Performance Systems Architecture, vol. 11, pp. 105-116, December 2022.

[9] L. Vignati, S. Zambon and L. Turchet, "A Comparison of Real-Time Linux-Based Architectures for Embedded Musical Applications," AES: Journal of the Audio Engineering Society, vol. 70, pp. 83-93, January 2022.

[10] W. Tao, W. Li, Y. Pengfei, R. Popli, P. Rani and R. Kumar, "Application of embedded Linux in the design of Internet of Things gateway," Journal of Intelligent Systems, vol. 31, pp. 1014-1023, January, 2022.

[11] J. Qiaowen, M. Haoyu, L. Yan, W. Zheyu and S. Wenchang, "A Novel Integrity Measurement Architecture for Embedded Linux Systems," Jisuanji Yanjiu yu Fazhan/Computer Research and Development, vol. 59, pp. 2362-2375, October 2022.

[12] F. Bowen, L. Su, W. Jiangdong, L. Qiran, W. Qingnan and T. Jihui, "A Novel Intelligent Garbage Classification System Based on Deep Learning and an Embedded Linux System," IEEE Access, vol. 9, pp. 131134-131146, September 2021.

[13] Q. Jiaqing, W. Huachen and G. Fei, "A multiprocessor real-time scheduling embedded testbed based on Linux," International Journal of Embedded Systems, vol. 14, pp. 451-464, January 2021.

[14] L. Kaizheng, Y. Ming, L. Zhen, Y. Huaiyu, Z. Yue, F. Xinwen and Z. Wei, "On Manually Reverse Engineering Communication Protocols of Linux-Based IOT Systems," IEEE Internet of Things Journal, vol. 8, pp. 6815-6827, April 2021.

[15] D. E. Mera, R. A. R. Solis, L. Reyes, R. Armstrong, W. J. Hernandez and A. L. Guzman-Morales, "A Power and Performance Study of Compact L-Band Total Power Radiometers for UAV Remote Sensing Based in the Processing on ZYNQ and ARM Architectures," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 1103-1113, December 2022.

Digital Camouflage Generation Based on an Improved CycleGAN Network Model

Leixiang Xia
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, Shaanxi, China
E-mail: 2576341766@qq.com

Zhiyi Hu
Engineering Design Institute
Army Research Loboratory
Beijing, 100042, China
E-mail: 763757335 @qq.com

Jun Yu
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, Shaanxi, China
E-mail: yujun@xatu.edu.cn

Yunshan Xie
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, Shaanxi, China
E-mail: 2398576240@qq.com

Kuncai Jiang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, Shaanxi, China
E-mail: 2992165628@qq.com

***Abstract*—This paper proposes a digital camouflage generation method based on an improved CycleGAN to produce camouflage patterns with a high degree of fusion with the background and realistic texture details. Firstly, a SE-ResNet network structure is constructed by combining the residual network ResNet with the channel attention mechanism SENet, enabling flexible adjustment of channel weights to effectively extract crucial channel features and enhance the network's perception capability of important information in images. Secondly, a color preservation loss is introduced to improve the adversarial loss function, thereby avoiding training instability and fluctuation in pattern quality. Experimental results demonstrate that the camouflage patterns generated using the proposed method achieve a Structural Similarity Index (SSIM) of 0.77 and a Peak Signal-to-Noise Ratio (PSNR) of 18.9, representing improvements of 0.27 and 3.3, respectively, compared to the original CycleGAN. This method can generate digital camouflage patterns with richer details, textures, and high fusion with the background.**

Keywords-Digital Camouflage; CycleGAN; Channel Attention Mechanism; Residual Network

I. INTRODUCTION

With the rapid development of military science and technology, military reconnaissance methods are exhibiting trends of diversification, intelligence, and high-precision resolution capabilities, which have significantly improved battlefield information acquisition capabilities, posing greater challenges to military camouflage technology [1]. Against this backdrop, digital camouflage technology, as a part of military camouflage techniques, plays an important role in military reconnaissance countermeasures.

To cope with diversified military reconnaissance techniques, domestic scholars have conducted in-depth research on digital camouflage technology. Cai Yunxiang [2] and colleagues utilized fractal dimension estimation based on fractal Brownian motion and a layer-by-layer fuzzy C-means clustering algorithm with a pyramid structure to extract texture and primary color features from background images, achieving the generation of digital camouflage patterns.

DOI: 10.2478/ijanmc-2024-0019

However, this digital camouflage the generation method is relatively cumbersome and requires sufficient practical experience to produce high-quality camouflage. To automatically generate digital camouflage patterns, Yang Wuxia [2] et al. extract the main colors of the background using the K-means algorithm on the color grayscale histogram based on the target background image to generate digital camouflage. Jia Qi [3] et al. used Markov random fields and pyramid models to build a digital camouflage design system, initially achieving automation in digital camouflage design. With the development of deep learning technology, Teng Xu [3] et al. combined cyclic adversarial networks with densely connected convolutional networks to quickly and automatically generate digital camouflage patterns, but the generated patterns still lacked richness in details and textures.

To generate camouflage patterns that blend seamlessly with the background and exhibit realistic texture details, this paper enhances a generative adversarial network model based on CycleGAN [4] (Cycle-Consistent Generative Adversarial Network). Building upon the traditional framework of Cycle-Consistent Generative Adversarial Networks, this study introduces a channel attention mechanism into the existing residual network to extract image features. Moreover, it incorporates a color loss function and enhances the adversarial loss function. These modifications effectively resolve issues related to the fine details and textures of the generated patterns.

II. Cycle-Consistent Generative Adversarial Network

A. Network Architecture

CycleGAN is an unsupervised generative adversarial network designed for translating images from one domain to another, such as from class X to class Y, without the need for paired training data. The model's key innovation lies in enforcing bidirectional image translation through a cycle consistency loss. This loss ensures that a translated image can be reconstructed into its original form within the same domain, maintaining fidelity to the original image throughout the translation process.

A complete CycleGAN model consists of two sets of generators and discriminators, each targeting a specific translation direction. Specifically:

The first set includes generator G and discriminator D_Y. Generator G is responsible for translating images from class X to class Y, while discriminator D_Y distinguishes between generated Y-class images and real Y-class images.

The second set includes generator F and discriminator D_X. Generator F translates images from class Y back to class X, while discriminator D_X distinguishes between generated X-class images and real X-class images [5].

During the training process, two sets of generators and discriminators are trained alternately. By optimizing the adversarial loss and cycle consistency loss, the model can gradually learn the mapping between the two image types and generate images that conform to the characteristics and distribution of the target image type.As shown in Figure 1.

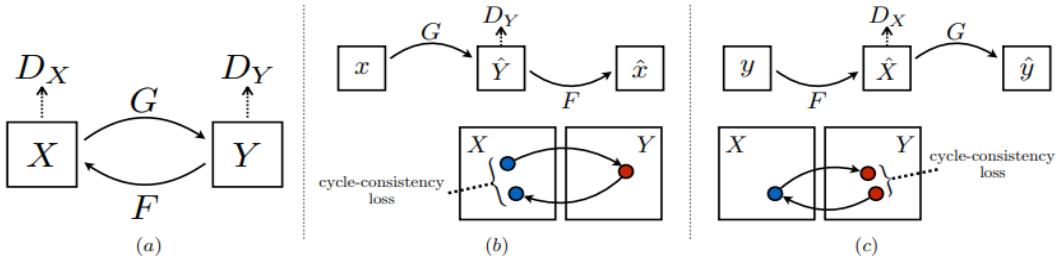


Figure 1. Structure of CycleGAN

ResNet (Residual Networks) is a deep learning network architecture that addresses the gradient issues in training deep networks by introducing residual blocks and skip connections, thereby enhancing network performance. CycleGAN utilizes multiple basic residual blocks in its generators to construct the network [6]. While ResNet solves the gradient vanishing problem through residual blocks, which enhances the efficiency of data transmission and mitigates the issues of gradient diffusion and network degradation, it does not significantly improve the quality of image generation.

B. Loss function

The loss functions of the CycleGAN network comprise three types: the adversarial loss (referred to as GAN loss in this context), the cycle consistency loss (referred to as cycle loss), and the identity mapping loss (referred to as identity loss). The adversarial loss aims to promote the adversarial learning between the generator and the discriminator, encouraging the generator to produce more realistic samples. CycleGAN employs two adversarial losses, taking the generator G and the discriminator D_Y as an example, which can be expressed as in formula (1); the cycle consistency loss enables translation between two different domains, and its formula is given in (2); the identity mapping loss is primarily designed to train the network's recognition capabilities, and its formula is presented in (3).

$$L_{GAN}\left(G, D_Y, X, Y\right) = E_{y\sim p_{\text{data}}(y)}\left[\lg D_Y(y)\right] + E_{x\sim p_{\text{data}}(x)}\left[\lg\left(1 - D_Y(G(x))\right)\right] \tag{1}$$

$$L_{cyc}(G, F) = E_{x\sim p_{\text{data}}(x)}\left[\| F(G(x)) - x\|_1\right] + E_{y\sim p_{\text{data}}(y)}\left[\| G(F(y)) - y\|_1\right] \tag{2}$$

$$L_{\text{identity}}(G, F) = E_{y\sim P_{\text{data}}(y)}\left[\| G(y) - y\|_1\right] + E_{x\sim P_{\text{data}}(x)}\left[\| F(x) - x\|_1\right] \tag{3}$$

Wherein, $p_{\text{data}}(x)$ represents the sample distribution of type X, and $p_{\text{data}}(y)$ represents the sample distribution of type Y. x represents images of type X, and y represents images of type Y[7]. JS (Jensen-Shannon) divergence is a parameter used in traditional GAN loss to measure the difference between the distributions of real images and generated images. When the two distributions do not overlap, the gradient of the JS divergence may become very small or zero, leading to issues such as gradient vanishing and, in extreme cases, gradient explosion

III. IMPROVED CYCLEGAN NETWORK MODEL

The image generation quality of traditional CycleGAN network models is not high, as the simple stacking of residual blocks in the generator introduces excessive redundant channels. These redundant channels are not conducive to extracting finer-grained information features, thus affecting the quality of image generation. To address this issue, this paper introduces the channel attention mechanism of SENet to optimize the residual blocks and improves the loss function, aiming to enhance the quality of image generation.

A. Improvement of network structure

In the original CycleGAN architecture, convolutional layers and pooling layers are typically employed to extract image features. However, the feature information extracted using this method is often overly redundant. To address this, a channel attention mechanism is introduced. The channel attention mechanism calculates the importance of each channel, enabling the network to focus on more significant channels and filter out redundant information. This paper utilizes the classic channel attention method, SENet [8] (Squeeze-and-Excitation Networks), and combines the SENet model with ResNet to generate SE-ResNet modules, thereby enhancing the network's feature extraction capabilities.

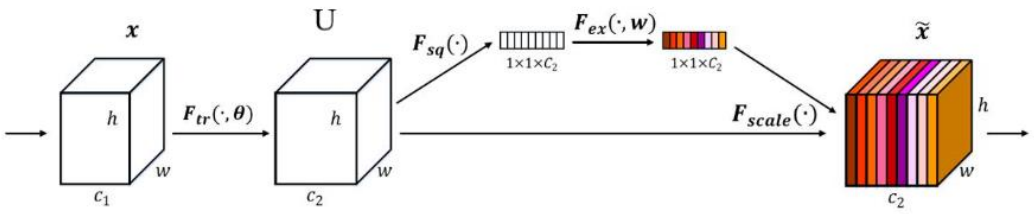


Figure 2. Schematic diagram of SENet

1) SENet Channel Attention Mechanism

The SENet module consists of two core operations: Squeeze and Excitation. The Squeeze operation compresses the feature channels through global average pooling. The Excitation operation learns the dependencies between channels through fully connected layers, generating weights for each channel, thus achieving the recalibration of feature channels. As shown in Figure 2, the input data X undergoes a convolutional operation $F_{tr}(\cdot, \theta)$ to obtain data U. Subsequently, it passes through the Squeeze operation $F_{sq}(\cdot)$, the Excitation operation $F_{ex}(\cdot, \mathbb{W})$, and the Scale operation $F_{scale}(\cdot)$, where C represents the number of channels, and H and W represent the height and width of a single channel. The specific implementation steps are as follows:

Step One: Squeeze（F_{sq}）: Global average pooling is used to compress the 2D features of each channel, with dimensions H×W, into a one-dimensional numerical value, generating a feature description along the channel dimension.

Step Two: Excitation（F_{ex}）: In SENet, there is a fully connected layer that takes the feature vector obtained from the previous step as input. This fully connected layer has an intermediate hidden layer to learn the dependencies between channels. Through an activation function (such as ReLU) and a parameterized scaling operation, the fully connected layer can learn the weight of each channel. These weights represent the importance of information in different channels [9].

Step Three: Scale（F_{scale}）: The learned channel weights are applied to the input feature map. Each channel's weight is multiplied by the corresponding channel's feature map, reweighting the feature map to enhance the representation of important channels and suppress the representation of less important channels.

During the above process, the Squeeze operation reduces the dimensions of the channel from C×H×W to C×1×1 through global pooling. The Excitation operation utilizes a fully connected layer to generate a weight vector for each channel's feature. Finally, the Scale operation multiplies the output of the Excitation operation with the input feature map. Through these operations, the weight vectors are assigned to each channel of the feature map, weighting the features of different channels accordingly.

2) SE-Resnet Module

SE blocks are plug-and-play and can be integrated into ResNet, resulting in SE-ResNet. The structure of SE-ResNet is illustrated in Figure 3.

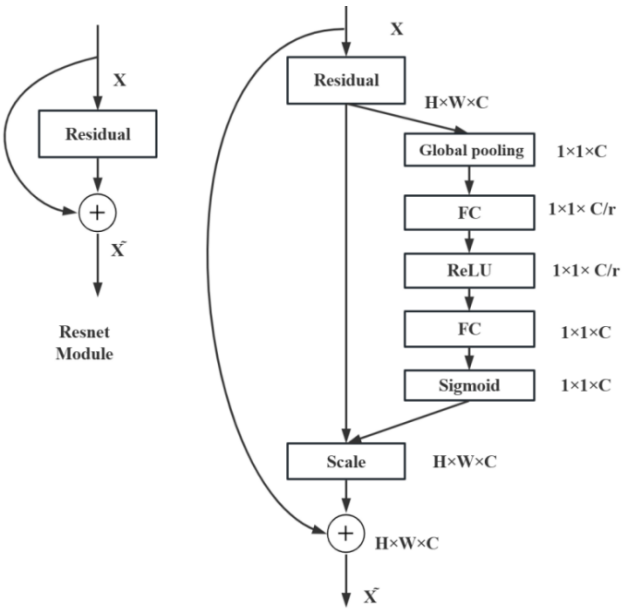


Figure 3. Diagram of the SE-Resnet module

SE-ResNet's basic structure is similar to that of traditional Residual Networks, composed of multiple residual blocks. The key difference lies in the addition of SE modules within each residual block to extract attention information. As shown in Figure 4, by introducing SE modules, SE-ResNet can more accurately mine feature information from images and enhance the network's utilization of features from different channels.

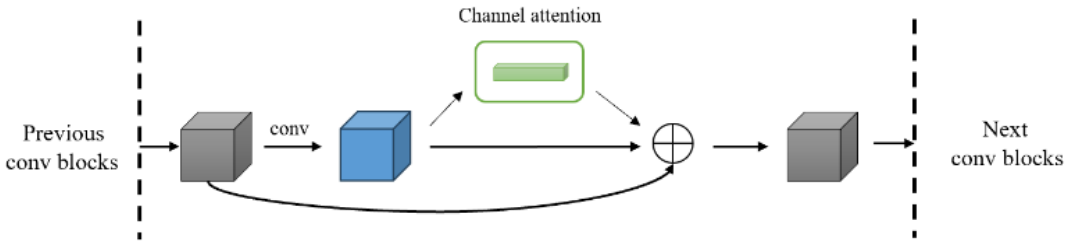


Figure 4. Schematic diagram of the attention mechanism for joining channels

The workflow of SE-ResNet is as follows:

a). Adjust the input image to a uniform size and pass the 3-channel image through a 7×7 convolution kernel to transform it into a 64-channel image. Subsequently, use 3×3 convolutions to extract features and generate feature maps.

b). Through a 9-layer residual structure, the image features are transformed from the source type to the target type. During this process, the input feature information is propagated and output.

c). Utilize deconvolution operations to restore the high-dimensional feature maps from the input, in order to reconstruct the surface features of the image.

d). The final layer performs a convolution operation to modify the dimensionality of the output from the previous step. The network structure of this generator is illustrated in Figure 5.

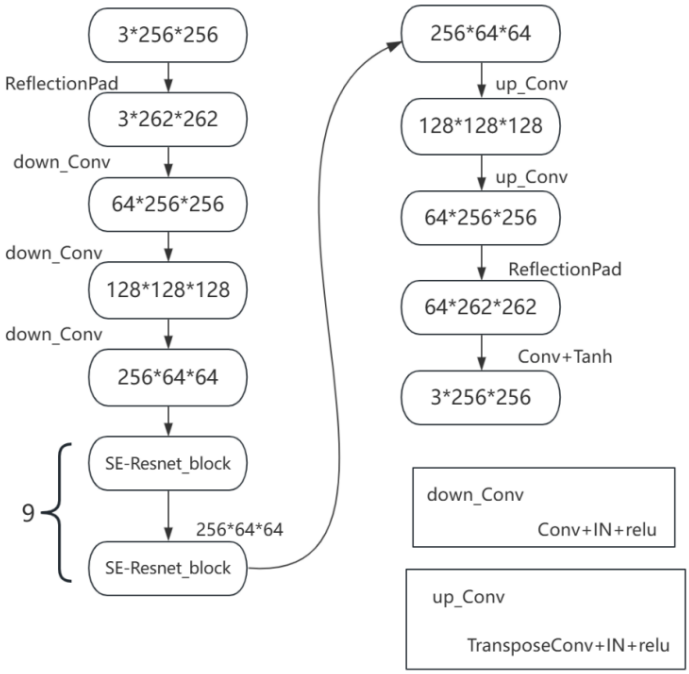


Figure 5. Improving the Generator Network Structure

B. Improvements to the loss function

In this paper, the loss function from WGAN-GP (Wasserstein Generative Adversarial Networks-Gradient Penalty) is adopted instead of the original GAN loss. WGAN-GP focuses more on matching the overall distribution, better preserving global consistency, and making the generated samples more natural. Additionally, a color preservation loss function is also included to ensure consistency in color features before and after image transformation.

1) Improving the Adversarial Loss Function

To avoid gradient issues caused by the JS divergence, this paper uses the WGAN-GP loss instead of the original GAN loss. The Wasserstein distance (also known as the EM distance) in WGAN-GP loss exhibits superior smoothness and smoother curve changes. Even when the generated image distribution does not overlap with the real image distribution, the discriminator D can still accurately calculate the actual distance between them, effectively evaluating the quality of the images produced by the generator G. This allows the use of the backpropagation algorithm to optimize the discriminator D, theoretically overcoming the difficulties of gradient vanishing and gradient exploding.

The meaning of EM distance is explained in equation (4).

$$W\left(p_{\text{data}}, p_{\text{model}}\right) = \sup_{\|f\|_L \le 1} E_{x\sim p_{\text{data}}}[f(x)] - E_{x\sim p_{\text{model}}}[f(x)] \tag{4}$$

In this context, p_{data} represents the distribution of real data, while p_{model} represents the model distribution output by the generator. $f(x)$ is a Lipschitz continuous function.

To achieve gradient penalty, a gradient penalty term is introduced on the basis of the EM distance. By penalizing the gradient of the discriminator's output for samples, the calculation is performed as shown in formula (5). During each parameter update, samples are taken from the linear interpolation points between real samples and generated samples. The gradients of the discriminator corresponding to these points are then calculated. By applying a penalty term to these gradient norms, the gradient norms are restricted to a reasonable range. Where $\hat{x}$ is the linear interpolation between a real sample x and a generated sample $G(z)$,and λ is the weight coefficient of the gradient penalty.

$$L_{\text{GP}} = \lambda E_{\hat{x}\sim P_{\hat{x}}}\left[\left(\left\|\nabla_{\hat{x}} D(\hat{x})\right\|_2 - 1\right)^2\right] \tag{5}$$

The objective function in the WGAN-GP model is shown in equation (6).

$$L = \underbrace{\underset{x\sim P_{\text{g}}}{E}[D(\tilde{x})] - \underset{x\sim P_{\text{r}}}{E}[D(x)]}_{\text{Original\_critic \_loss}} + \underbrace{\lambda \underset{x\sim P_x}{E}\left[\left(\left\|\nabla_{\hat{x}} D(\hat{x})\right\|_2 - 1\right)^2\right]}_{\text{gradient\_penalty}} \tag{6}$$

In this paper, by reassessing the weight parameters, the images generated from random noise are taken as penalty terms and input into the network, and gradient clipping is applied to address the issue of training instability when using the EM distance. The improved objective function is shown in formula (7), where $P_{\text{penalty}}(x)$ serves as a penalty term positioned between $P_{data}(x)$ and $P_{\text{z}}(z)$, and $P_{\text{z}}(z)$ aims to continuously approach $P_{data}(x)$.

$$V(D,G) = \max_D \left\{E_{x\sim P_{\text{data}}(x)}[(D(x))] - E_{z\sim P_{\text{z}}(z)}[D(G(z))] - \lambda_1 E_{x\sim \text{ penalty}}\left[\left(\left\|\nabla_x D(x)\right\|_2 - 1\right]\right)^2\right]\right\} \tag{7}$$

The GAN loss typically encourages the generated samples to be as similar to the real samples as possible on a pixel-level basis. However, this pixel-level similarity can sometimes lead to inconsistencies in the details of the generated images. WGAN-GP (Wasserstein GAN with Gradient Penalty) incorporates a gradient penalty term that constrains the gradient norm of the discriminator, preventing gradient explosion and vanishing issues. This gradient penalty term also enhances the stability of the network and allows it to generate more realistic and clearer samples.

2) Add color retention loss function

To enhance the similarity between the generated images and the original images, a color preservation loss is integrated into the cycle consistency loss. This color preservation loss function is used to calculate the distance between the distributions of real samples and generated samples. Compared to other loss functions, the color preservation loss is relatively less affected by outliers, making it more robust to noise or abnormal situations. Through regularization using the color preservation loss function, the model can be encouraged to select features with fewer non-zero weights, which improves the model's generalization ability and interpretability. The formula for the color preservation loss is shown in equation (8).The formula for the total loss function is shown in equation (9).

$$L_{\text{color}}(G,F)=E_{x-P_{\text{data}}(x)}\left[\| x-G(x)\|_1\right]+E_{x-P_{\text{data}}(x)}\left[\| x-F(G(x))\|_1\right] \tag{8}$$

$$L_{\text{total}}\left(G,F,D_X,D_Y\right)=V(D,G)+\lambda L_{\text{cyc}}(G,F)+\mu L_{\text{identity}}(G,F)+L_{\text{color}}(G,F) \tag{9}$$

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental platform used Ubuntu 18.04 system and employed a graphics processing unit (GPU) to accelerate the training speed. The deep learning framework adopted in this experiment is Pytorch 1.4.0. The datasets used in this paper are from UCMerced_LandUse [10], Places365 [11], and others, with pre-training conducted on the public dataset monet2photo.

In this experiment, the images in the dataset were preprocessed by eliminating unqualified images, removing poor learning samples, excluding overly similar learning samples, and fixing the image sizes. Subsequently, data augmentation techniques such as scaling up, random rotation, random flipping, and adding noise were applied to expand the dataset to 4,200 images.

To evaluate the experimental results, Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) were used in this paper.

Given two image sets I and K of dimensions m×n, for any i and j, PSNR can be defined as equation (10).

$$PSNR=10\lg\left[\frac{\left(2^n-1\right)^2}{MSE}\right] \tag{10}$$

In equation (10), MSE represents the Mean Squared Error, and its expression is given in equation (11).

$$MSE=\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I(i,j)-K(i,j)\right]^2 \tag{11}$$

The higher PSNR value is, the lower the degree of image distortion and the better the quality of the image.

Luminance, Contrast, and Structure are the three essential elements that constitute an image. Their respective definitions are given in equations (12), (13), and (14):

$$l(x,y)=\frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1} \tag{12}$$

$$c(x,y)=\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2} \quad (13)$$

$$s(x,y)=\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3} \quad (14)$$

The SSIM algorithm consolidates the calculation formulas for luminance, contrast, and structural similarity to obtain the overall calculation formula for image similarity (12).

$$\mathrm{SSIM(x,y)}=\left[l(x,y)^{\mathrm{a}},c(x,y)^{\beta},s(x,y)^{\gamma}\right] \quad (15)$$

SSIM values range from -1 to 1, with SSIM = -1 indicating complete dissimilarity between two images, and SSIM = 1 indicating their absolute identity.

This paper conducted ablation experiments and comparative experiments. In the ablation experiments, three improvements were made to the model with the addition of the color preservation loss function: CycleGAN + improved loss function, CycleGAN + SENet, and CycleGAN + SENet + improved loss function. These experiments aimed to validate the efficacy of the proposed approach. During comparative analysis, the proposed model was assessed alongside two image generation methods, GAN and DRIT, to gauge the image generation quality across different models.

A. Ablation experiments

In the ablation experiments, initially, the color preservation loss function was incorporated, with the outcomes presented in Figure 6. It is evident that upon integrating the color preservation loss function, the colors of the generated images more closely resemble those of the input images.

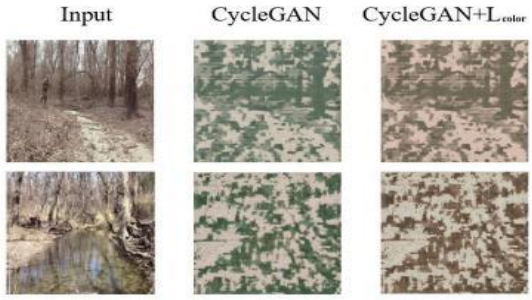


Figure 6. Comparison of Results before and after Adding Color Preservation Loss

1) CycleGAN + Improved Loss Function

The CycleGAN was combined with the improved loss function. Three models were trained using the original GAN loss, the WGAN objective function, and the WGAN-GP objective function. Figure 7 shows the changes in the loss curves during the training process for the three different loss functions: the original GAN loss, WGAN, and WGAN-GP.

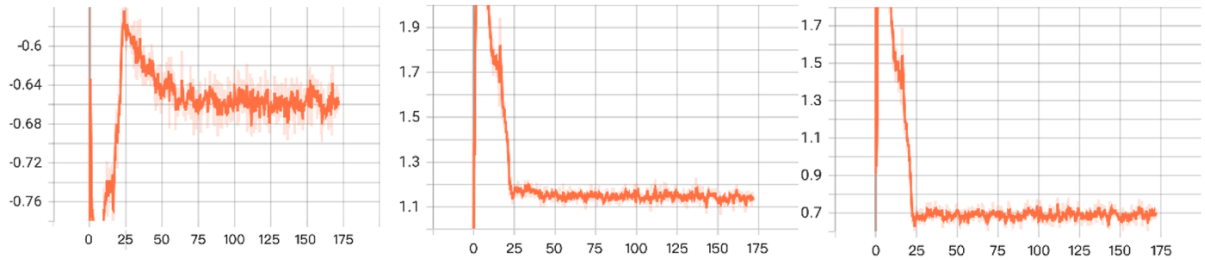


Figure 7. Loss Curve Variation Chart

As can be seen from Figure 7, compared to the loss function graphs of the original GAN and WGAN, the WGAN-GP loss adopted in this paper achieves a lower loss value when reaching the Nash equilibrium. This phenomenon indicates that the images generated by the WGAN-GP network have a higher quality.

2) CycleGAN + SENet

The CycleGAN was combined with SENet. The curve comparisons of the various loss functions between the original CycleGAN model and the improved CycleGAN model in this paper are shown in Figure 8 and Figure 9, where the horizontal axis represents the number

of epochs (iterations), and the vertical axis represents the loss value.

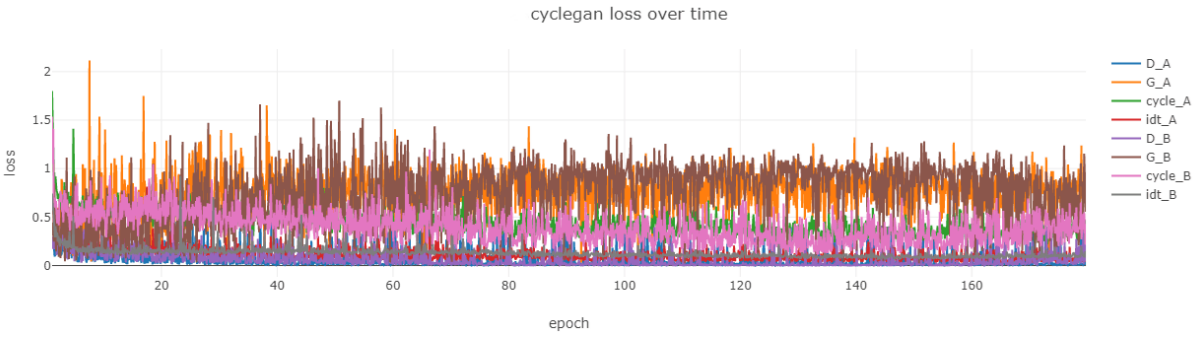


Figure 8. Loss Function Change Trend for the Original CycleGAN Model

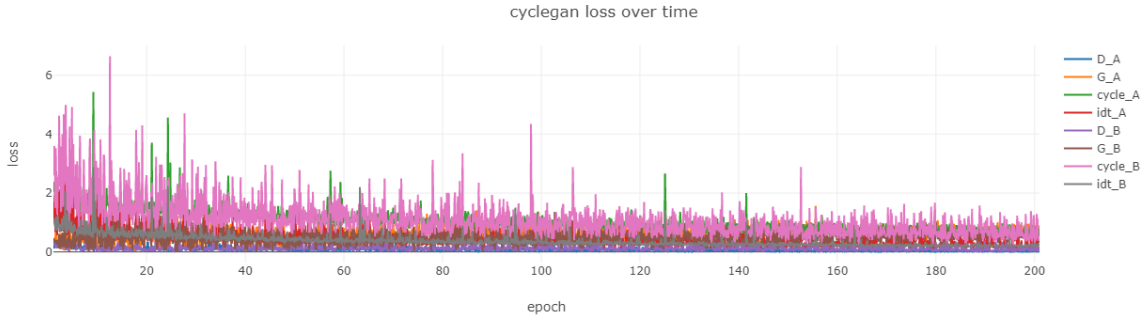


Figure 9. Loss Function Change Trend for the Improved CycleGAN Model

As can be seen from Figures 8 and 9, the loss function of the original CycleGAN model has not yet shown a converging downward trend when reaching 160 epochs, while the loss function of the improved model begins to show a downward trend starting from 40 epochs and converges at 200 epochs.

The ablation experiments were conducted to verify the effects of introducing the SENet network structure and the improved loss function. Using the method of controlled variables, four sets of experiments were performed by adding different improved modules. The final evaluation index data are shown in Table Ⅰ.

TABLE I. EVALUATION INDEX TABLE OF ABLATION EXPERIMENTS

Model	SSIM	PSNR
CycleGAN	0.50	15.6
CycleGAN+SENet	0.61	16.7
CycleGAN+improved loss function	0.68	15.9
CycleGAN+SENet+improved loss function.	0.77	18.9

According to Table I, the baseline CycleGAN model exhibited the poorest performance. Incorporating the SENet channel attention mechanism led to an SSIM increase of 0.11 and a PSNR increase of 1.1. Compared to the original CycleGAN, the version with enhanced loss functions showed SSIM and PSNR improvements of 0.18 and 0.3, respectively. By combining these approaches, SSIM reached 0.77, while PSNR achieved 18.9. The experimental results indicate that both methods, when used independently, can enhance the pattern details and texture quality. By combining these two methods in this paper, a more ideal pattern generation effect can be achieved.

B. Comparative experiments

In this paper, two image generation methods, GAN (Generative Adversarial Network) and DRIT (Diversity Regularized Image-to-Image Translation), were selected to conduct comparative experiments and evaluate the effectiveness of image generation. GAN is a form of generative adversarial network that generates images by allowing two neural networks (generator and discriminator) to learn from each other in an adversarial manner. DRIT is a model that can achieve image translation with different styles.

(a)Input image (b) GAN (c) DRIT (d)Improved CycleGAN

Figure 10. Comparison of images generated by different models

As shown in Figure 10, (a) represents the target-type image and the original image; (b) is the image generated by the GAN model, which performs poorly in terms of color and texture, significantly differing from the original image; (c) depicts the image produced by the DRIT model, which although retains some texture features, differs greatly in color features and appears blurry. (d) is the image generated by the CycleGAN model, which, compared to DRIT, not only preserves texture and color features but also exhibits richer details.

To evaluate the quality of the generated camouflage patterns, two metrics, SSIM and PSNR, were used, and the final evaluation results are presented in Table Ⅱ.

TABLE II. COMPARATIVE EXPERIMENTAL EVALUATION SCORE TABLE

Generative model	SSIM	PSNR
GAN	0.19	13.5
DRIT	0.48	14.8
CycleGAN+SENet+improved loss function	0.77	18.9

As seen from Table Ⅱ, the SSIM value of our proposed method is 0.58 higher than that of the GAN model and 0.29 higher than the DRIT model. Similarly, the PSNR value of our method is 5.4 higher than the GAN model and 4.1 higher than the DRIT model. Combining the results from Table Ⅱ and Figure 10, it can be concluded that our proposed method achieves superior performance, both in terms of visual perception and the SSIM and PSNR scores.

V. CONCLUSIONS

This paper investigates a digital camouflage generation method based on an improved CycleGAN. Firstly, the combination of residual networks with channel attention mechanisms is explored, which enables the network to focus more on important channel features. Secondly, the use of WGAN-GP loss instead of the original GAN Loss improves the adversarial loss function, avoiding the common problem of unstable outputs in traditional generative adversarial networks, thus generating more realistic and clearer samples. Lastly, a color preservation loss function is added

to prevent the generator from autonomously modifying the image's hue before and after image translation, avoiding color changes. Experimental results demonstrate that compared to other methods, the proposed approach produces camouflage patterns with more realistic texture details and higher fusion with the background, confirming the effectiveness of this method.

REFERENCES

[1] Yang Di. Evaluation Method of Camouflage Effect in Dynamic 3D Scenes [Dissertation]. Jiangnan University, 2024. DOI: 10.27169/d.cnki.gwqgu.2023.001844.

[2] Teng Xu. Research on Digital Camouflage Generation Based on Generative Adversarial Networks [Dissertation]. Southwest University of Science and Technology, 2021. DOI: 10.27415/d.cnki.gxngc.2021.000342.

[3] Teng Xu, Zhang Hui, Yang Chunming, Zhao Xujian, Li Bo. Digital Camouflage Disguise Generation Method Based on Cycle-Consistent Adversarial Networks [J]. Computer Applications, 2020, 40(02): 566-570.

[4] Luo Li. Research and Application of Adversarial Sample Generation Method Based on CycleGAN [Dissertation]. Guilin University of Electronic Technology, 2023. DOI: 10.27049/d.cnki.ggldc.2023.001423.

[5] Liu, Zunyang et al. "Background dominant colors extraction method based on color image quick fuzzy c-means clustering algorithm." Defence Technology (2020): n. pag.

[6] Lü Lin. Research and Implementation of Image Translation Algorithm Based on Cycle Generative Adversarial Networks [D]. Shandong University, 2024. DOI: 10.27272/d.cnki.gshdu.2023.002296.

[7] Fu JinHuang Shan. Research on Image Denoising Based on Improved Cycle Generative Adversarial Networks [J]. Computer Engineering and Applications, 2023, 59(17): 178-186.

[8] JHU J, SHEN L, ALBANE S, et al. Squeeze-and-excitation networksL [J]. EEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8) 2011-2023.

[9] Li Yongjiang, Yang Zhigang. Image Translation Method Based on Improved CycleGAN [J]. Journal of Harbin University of Commerce (Natural Sciences Edition), 2023, 39(03): 281-286. DOI: 10.19492/j.cnki.1672-0946.2023.03.005.

[10] Li Yongjiang, Yang Zhigang. Image Translation Method Based on Improved CycleGAN [J]. Journal of Harbin University of Commerce (Natural Science Edition), 2023, 39(03): 281-286. DOI: 10.19492/j.cnki.1672-0946.2023.03.005.

[11] Li .Research on Artistic Image Style Transfer Method Based on Generative Adversarial Networks [D]. Guangdong University of Technology, 2022. DOI: 10.27029/d.cnki.ggdgu.2021.000308.

A Review of Manned/Unmanned Aerial Vehicle Cooperative Technology and Application in U.S. Military

Wenguang Li
Shijiazhuang campus
Army engineering university of PLA
Shijiazhuang, China
E-mail: oec_lwg@163.com

Yong Liu
Unit 75737
People's Liberation Army
Guangzhou, China
E-mail: 1292243594@qq.com

Fengming Shi *
Shijiazhuang campus
Army engineering university of PLA
Shijiazhuang, China
E-mail: shi_fengming@163.com

Yuefei Zhao
Shijiazhuang campus
Army engineering university of PLA
Shijiazhuang, China
E-mail: 919962023@qq.com

Weizhao Zhang
Department of teaching and research
Army engineering university of PLA
Shijiazhuang, China
E-mail: 1141343749@qq.com

Zhili Wang
Shijiazhuang campus
Army engineering university of PLA
Shijiazhuang, China
E-mail: yangbo2000_8138@163.com

***Abstract*—The paper is to learn the experience in the field of Manned/Unmanned Aerial Vehicle (MUAV) cooperative technology and enlighten the technology development and progress of our country. First of all, it is to collect the typical research projects and tracke experiment dynamic, excavate the development in the field of MUAV cooperative technology route; And then, attention is paid to the related typical application cases to analyze the current technical maturity; Second, from the perspective of technology development, application, analysis is executed from the human-computer interaction control technology, intelligent decision-making technology, communication anti-jamming technology, open hardware and software technology to MUAV; Finally, the experience of the U.S. military in the field of MUAV collaborative technology are summarized, and the enlightenment for the development of China in the field is given.**

Keywords-Manned/Unmanned Aerial Vehicle; Collaborative technology; Test platform; Application case

I. INTRODUCTION

In 2010, the US Army released "the Army UAV System Roadmap (2010-2035)". Thereafter, the Army reevaluates the roadmap every two years to keep it up with the rapid pace of technological developments. The roadmap divides unmanned aerial vehicle (UAV) development into three phases. The goal of the near-term development phase (2010-2015) is to rapidly integrate UAV systems into tactic-level units to meet the force's current operational requirements. The medium-term development phase (2016 to 2025) aims to increase the degree of automation of UAV systems that can quickly and fluently support combat operations, and establish the link between manned and unmanned systems, opening the way for collaboration between manned and unmanned systems. The long-term development phase (2026 to 2035) aims to greatly improve the ability of

manned and unmanned systems to cooperate in combat operations.

In the "2011-2036 Comprehensive Roadmap of Unmanned Aerial Vehicle Systems" issued by the U.S. Department of Defense, the definition of Manned/Unmanned Aerial Vehicle (MUAV) cooperative operations is that a unified formation is established between manned and unmanned aerial vehicles (UAVs) to complete a certain mission, and the process of finally completing the established mission objectives through cooperative control, platform interoperability and resource information sharing [1]. This is for the first time, the American MUAV cooperative engagement has made the definition.

In 2018, the US Navy released a Roadmap for Naval Unmanned Systems. The Navy's vision for the development of unmanned systems is to seek to establish a future force that can adapt to all-domain operations with efficient manned/unmanned cooperation. It aims to improve the Navy's combat capabilities by reducing the operating costs and risks of unmanned systems in the air, surface, underwater and ashore, enhancing the endurance of unmanned systems, and providing better situational awareness.

It is easy to see that the US military has always attached great importance to the construction of its unmanned system combat capability, and has taken the realization of manned/unmanned cooperative operations as its long-term development goal and vision. This article is embarked from the American MUAV cooperative development, tracking its research projects and typical test platform, excavating its technical maturity, analyzing the American MUAV cooperative engagement case and its capabilities; After that, combined with the next development plan of the US military, the enlightenment of our MUAV collaborative combat capability construction is given.

II. TYPICAL RESEARCH PROJECTS

A. Skyborg

Skyborg is a large-scale artificial intelligence program of the U.S. Air Force Research Laboratory. It aims to develop an artificial intelligence software system that can control UAVs to fly, take off and land autonomously, and cooperate with human pilots in combat. The ultimate goal is to improve the efficiency of UAV mission planning and the performance of human-machine collaboration to deal with various threats in high-confrontation combat environments [2].

In March 2019, the Air Force first unveiled the Aerial Borg program, and officially released the project proposal in May 2020. In December 2020, the Air Force awarded contracts to General Atomics, Boeing, and Kratos Defense and Security to build the Loyal Wingman prototype to carry the artificial intelligence software systems of the Aerial Borg program.

At the end of April 2021, the UAP-22 Mackerel Shark UAV developed by Kratos Defense and Security became the first test flight platform of the Aerial Borg program. On May 5, 2021, the second flight test was carried out, and it was flown in coordination with an F-16C fighter jet. The U.S. military claimed that this flight test was the closest coordinated flight between an autonomous UAV and a manned fighter jet in the history of the U.S. Department of Defense. The Air Force has also conducted a series of flight tests to test the capability of manned aircraft and multiple unmanned aerial vehicles to cooperate with each other. Figure 1 is the air's projects designed by MUAV layered fighting style.

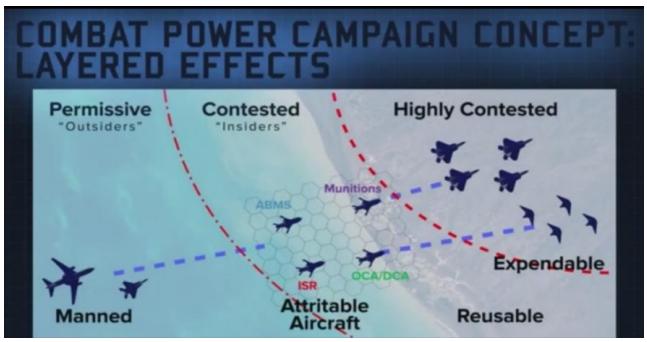


Figure 1. Hierarchical combat style of the Aerial Borg project

B. Project Carrera

John Clark, Lockheed Martin Vice President and general manager of "Skunk Works," detailed the company's newly launched Flexible Autonomy program, internally known as "Project Carrera,"

during a press conference call with the release of the official concept video on Sept 14, 2022 [3-4]. The project's first experiments was planned to take place at Skunk Works and involve Racing drones and F-35 strike fighters, which fly four Racing drones under the wings of the F-35 and launch the drones in the air. Figure 2 shows the manned/unmanned combat model designed by Project Carrera.

Figure 2. Lockheed Martin manned-unmanned combat scenario

Skunkworks' vision of Project Carrera, especially in the first phase, is mainly to explore and refine various future manned/unmanned cooperative combat concepts, and is very optimistic about the vision of manned/unmanned cooperative combat. More details are expected to emerge as Project Carrera nears the start of flight tests.

C. Air Combat Evolution Project (ACE)

The Defense Advanced Research Projects Agency (DARPA)'s Office of Strategic Technology launched the Air Combat Evolution project in May 2019. The purpose is to overcome the problem of autonomous performance in human-machine cooperative close air combat, enhance the confidence of warfighters in autonomous combat systems, and develop a scalable, trusted, human-level artificial intelligence autonomous system [5].

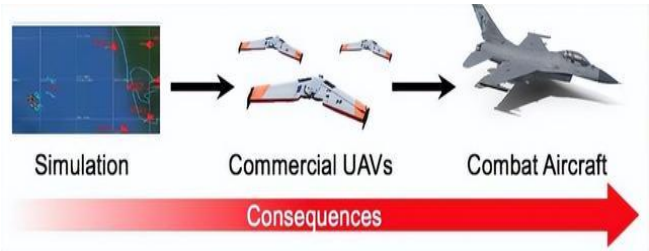


Figure 3. The three phases of the Air Combat Evolution program

The ACE program can be specifically divided into three phases, which are expected to last five years: The first phase is a simulated environment study lasting 18 months, focusing on the validation and development of its key capabilities in a simulated environment and simulation. The second phase, UAV flight trials, will last 16 months. The third phase, MUAV flight trials, will use full-size aircraft and small aircraft to conduct tests. Figure 3 shows the three phases of the ACE program. The DARPA announced on March 18, 2021 that more than half of the first phase of the Air Combat Evolution program has been completed, and several key phase results have been achieved.

Information on other manned/unmanned collaborative technology projects in the United States is shown in Table Ⅰ.

TABLE I. OTHER PROJECT INFORMATION

Time	Project name	Verification platform	Main content
1993	Bird Dog	/	Develop the concept of manned/unmanned collaboration
1997	TCS	P-3C, F/A-18, AV-8B/ RQ-8A, MQ-4C, X-47B, UTAP-22	The focus is on the test and verification of cooperative combat projects to achieve the 5-level control of UAVs
2006	HSKT	AH-64D /RQ-5B	Verify manned helicopter /UAV/ fighter aircraft coordination capabilities
2021	UxS IBP21	E-2C, EA-18G, MH-60/ MQ-9	The focus is on assessing the capabilities of manned/ unmanned systems in areas such as intelligence, surveillance and reconnaissance
2022	The Marines train in tandem with the Navy	UH-1Y, AH-1Z/ MQ-8C	Verify the ability of manned/ unmanned formations to work together in future coastal environments

III. TYPICAL TEST PLATFORMS

A. XQ-58A Valkyries UAV

Figure 4. XQ-58A Valkyria UAV

American "loyal wingman" project, designed to mix fifth-generation fighter with UAV formation, coordinated various combat missions. In the composite wing, human-computer play the role of "commander", is responsible for his orders, UAV is responsible for "charge" and strike enemy [6]. XQ - 58 Valkyries UAV is a typical representative of "loyal wingman" project, by "carat," defense security company research and development, in collaboration with the air force research laboratory is a ranged attack, high subsonic UAV, mainly for the combined with man-machine, surveillance and reconnaissance and long-range strike missions. Figure 4 shows the XQ-58A Valkyrie UAV, and its performance parameters are shown in Table Ⅱ.

Kratos initially developed three demonstrators, one of which was delivered to the US Air Force Research Laboratory for flight testing, and the remaining two were used for internal testing. In March 2019, XQ - 58 Valkyries demonstrate a prototype successfully in Yuma proving ground for the first time for the flight. Subsequently, the second and third flight tests were conducted in June and October 2019, respectively, and the fourth flight test was conducted in January 2020. At this point, XQ - 58 a Valkyries the accumulative total of UAV flight time has been more than five hours, is beyond the goal of the programmed of flight test. In August 2023, the U.S. Air Force conducted its first flight test with an AI-autonomously piloted XQ-58A Valkyria. At present, XQ - 58 Valkyries drones can have a and F - 22 or F - 35 fighter fleet, realize the cooperative engagement, to carry out reconnaissance and combat tasks.

TABLE II. XQ-58A PARAMETERS

Serial number	Parameters
1	Length 8.8m
2	Wingspan 6.7m
3	Maximum take-off weight 2.7t
4	Maximum speed Mach 0.72
5	Maximum range 4800km
6	The maximum combat radius is 2500km

B. Unmanned modification of the F-16

In recent years, the US Air Force has begun to refit the F-16 fighter jet in depth, mainly planning to put the transformed F-16 fighter jet completely under the command of artificial intelligence system in actual combat. At present, the first flight test of the program has been completed, which also lays the foundation for the US Air Force to explore MUAV cooperative operations, and provides the possibility for the US military to form the world's first unmanned fighter formation.

Compared with the XQ-58A Valkyries, the unmanned F-16 has unparalleled advantages in range, bomb load and maneuverability. The advantages of the XQ-58A Valkyries are light and small fuselage, stealth ability and low manufacturing cost, but its independent combat ability, especially air combat ability, is far from the F-16 combat UAV, which plays the role of fire multiplier and is a real unmanned wingman.

It is said that in the future, the US Air Force plans to develop a fleet of at least 1,000 UAVs to serve as wingmen for the F-35 fighter and the next generation of advanced fighters. The coordinated drones could carry missiles or other weapons, carry out electronic warfare missions, or perform intelligence, surveillance and reconnaissance missions ahead of other manned aircraft [7].

C. Elf UAVs

The Elf is a U.S. DARPA program of UAVs, code-named X-61A, launched in 2015. It is capable of being launched in the air by a manned

transport aircraft, retrieved, and maintained before being used again [8].

Figure 5. A C-130 transport aircraft carrying a Gremming drone

During the flight demonstration, the X-61A was launched from a C-130 wing hantry, and the drone was controlled by two control platforms, one mounted on the C-130 and the other on the ground. During the recovery, the cap sting-recovery system on the tailgate of the C-130 transport aircraft became the key technology, and the X-61A UAV completed the automatic docking with the C-130 through the docking system at the end of the cap sting-towed cable. When the X-61A is successfully docked, the UAV will shut down its engines, redraw its wings, and then be pulled back to the aircraft cargo bay by the towing cable.

In November 2019, the X-61A was launched for the first time on a C-130A transport aircraft. In August 2020, the X-61A made its second free flight and finally landed successfully with the help of a parachute. In October 2021, the recovery of the X-61A by a C-130 transport aircraft was successfully tested. Figure 5 shows the C-130 transport aircraft carrying X-61A UAVs.

IV. TYPICAL APPLICATION CASES

A. Manned/ unmanned cooperative reconnaissance

Since 2020, the US MQ-4C Poseur unmanned aerial vehicle has begun to deploy around the South China Sea and conduct joint reconnaissance with P-8A Poseur anti-submarine patrol aircraft. As an upgraded version of the RQ-4N Global Hawk UAV, the MQ-4C Mermaid Posei UAV not only improves the strength of the fuselage to better adapt to the wind environment at sea, but also adds anti-ice measures, which can repeatedly cross the clouds to the lower airspace for close reconnaissance, which is very suitable for carrying out a wide range of reconnaissance and surveillance missions at sea. Figure 6 show the MQ-4C and P-8A are working together on reconnaissance

Figure 6. MQ-4C and P-8A are working together on reconnaissance

According to the plan, the US Navy will use 68 MQ-4C Merman Posei UAVs and 117 P-8A Posei anti-submarine patrol aircraft to jointly carry out maritime patrol and surveillance and reconnaissance missions to replace the old P-3C patrol aircraft. MQ - 4 c "mermaid" a sea drones in the anti-submarine mission for P - 8 a maritime intelligence support, and P - 8 a is MQ - 4 c "mermaid" a sea of UAV control and communications relay platform.

In December 2020, two US Navy B-1B strategic bombers opened the way for the MQ-4C Poseur to conduct reconnaissance close to the South China Sea. Some scholars believe that if the MQ-4C UAV and B-1B bombers carry out missions in coordination, the MQ-4C UAV can be used for search, detection and targeting missions, and the B-1B will drop AGM-158C long-range anti-ship missiles on the target.

B. Cooperative manned/unmanned anti-submarine

In May 2023, the US Navy held the Unmanned Systems Integration Operations-23 (USIBP-23) exercise, which focused on surveillance,

reconnaissance, maritime and submarine long-range firing, command and control, and reorganizing intelligence, and focused on cooperative manned/unmanned anti-submarine operations [9]. During the exercise, the participants used the MQ-9B Ocean Guardian UAV and the MH-60R Seahawk helicopter to quickly complete the association and location of the target, and then transmitted the tactical report to the commander of the anti-submarine Warfare Center at the Pearl Harbor Naval Base through the MQ-9B Ocean Guardian UAV [10]. The ASW commander then used the MH-60R Seahawk helicopter to drop torpedo training rounds against the simulated submarine.

C. Coordinated manned/unmanned strike

"In 2015, the sea - air l" exercise, F - 16 VISTA (flight simulator) successfully verified the route to follow, autonomous formation flight, air collision avoidance ability and the ability to rejoin the formation, promote the autonomy of the craft level, and command the autonomy of the command and control of UAV [11]. In addition, "sea - air l" also verified the F - 16 VISTA have the capacity of planning and ground attack missions. While performing a mission, the F-16 VISTA determines mission priorities based on the priorities given by the command aircraft to achieve all mission objectives.

The US military successfully verified the autonomous strike capability of the unmanned combat aircraft during the "Hepher-Air Raider II" exercise held in 2017. The demonstration set and achieved three key objectives: first, it successfully planned and completed an autonomous strike mission to the ground; Second, the ability to adapt to changes in the environment during the execution of a mission; And the successful use of a software integration environment that is fully compliant with the Air Force's Open Mission System (OMS) standard[12-13].

V. SYNERGISTICAL KEY TECHNOLOGIES

A. Human-computer interaction control technology

In the future air combat battlefield, human-computer cooperative operation is a dialectical unity. The pilot plays the role of driving the fighter to carry out air combat tasks, and also acts as the commander of the MUAV mixed formation, conveying the decision-making instructions of the rear command center, integrating battlefield information, and issuing instructions such as interim guidance. Pilots need to be responsible for the "fighting" in the past, but in the future pilots would need to be "battlefield commander", to be able to communicate, to other UAV in the composite wing give orders. The clear, fast and accurate transmission of commands by human-computer interaction control technology has a direct impact on the battlefield situation. The development of deep learning, data mining and other technologies has provided infinite possibilities for human-computer interaction control technology based on natural language understanding. Even in the noisy and thundering environment of the battlefield, the pilot can easily express the intention and clearly indicate "how to do", "what to do" and "when to do", and the UAV can quickly understand and execute these instructions to realize the command and control of the pilot over the UAV.

B. Intelligent assistant decision-making technology

In MUAV hybrid formation, pilots can clearly and quickly convey strategic and tactical intentions to UAVs through natural language, thus reducing the burden of pilots. The hybrid formation usually consists of one manned aircraft and multiple UAVs. Even if the interactive control mode between the pilot and the UAV is simplified, the interaction pressure is still large. Therefore, how to further reduce the pressure of pilots has become one of the important issues in the maturation of MUAV cooperative operations. Intelligent assistant decision-making technology can assist pilots in decision-making through intelligent algorithms. A small number of decision-making tasks with higher priority are

assigned to the pilot for decision-making, and a large number of decision-making tasks with lower priority are completed by the machine independently. This can effectively reduce the decision-making pressure faced by pilots, so that they can have more energy and time to deal with more important affairs.

C. Communication Anti-jamming technology

In MUAV cooperative operations, maintaining good communication is the key to maintaining cooperation. MUAVs can usually achieve effective communication before entering enemy airspace. However, once inside hostile airspace, they will face various means of communication jamming. Man-machine and UAV at the individual level can keep certain single combat ability. However, on the whole, MUAVs still need to rely on high-speed and secure communication to exchange instructions and important intelligence information. At the same time, using the frequency hopping communication method with strong anti-jamming ability can minimize the possibility of being interfered by the other side, which is also one of the technologies worthies of continuous research and application.

D. open hardware and software technology

As a hybrid combat formation, the MUAV cooperative combat system needs to have multiple functions such as surveillance, reconnaissance, electronic jamming, damage assessment, and strike implementation. The UAVs in the hybrid formation need to adopt various types or carry different payloads and sensors to perform diverse tasks. Therefore, integrating heterogeneous UAVs from different manufacturers and efficiently integrating sensors and payloads from different manufacturers into the overall system are realistic challenges faced by the MUAV cooperative combat system in terms of operation, maintenance and upgrading. Open hardware and software technology has become an effective means to solve this problem, which can realize the seamless integration of software systems and hardware products produced by different manufacturers, thus simplifying the subsequent maintenance and upgrading work.

VI. REFLECTIONS AND IMPLICATIONS

A. Continually promote the collaborative combat between the fifth-generation fighter and UAV

In the future, the US military will still attach importance to the collaborative operations between the fifth-generation fighter jets and advanced UAVs. The US military is not only studying and testing the networking and intelligent upgrade of the fifth-generation fighter, but also actively exploring the formation combat of the fifth-generation fighter with stealth unmanned reconnaissance aircraft and unmanned bombers. The goal is to build and improve the overall collaborative combat ability of the fifth-generation fighter and UAV/UAV group based on information sharing. So that the fifth-generation fighter can effectively command and control the UAV or even the UAV group for combat.

B. Vigorously try to make unmanned transformation of manned fighter aircraft

The unmanned transformation of manned aircraft has huge potential advantages, which can not only save research and development costs, but also greatly improve the combat capability after transformation by taking advantage of the load advantage of manned aircraft. The DARPA announced that the artificial intelligence system developed by the agency has been installed on the F-16 fighter jet after a special unmanned modification, making it an artificial intelligence unmanned aircraft, which can be used as a "loyal wingman" to cooperate with the F-35 fighter jet, improving the combat capability of the US military to a certain extent.

C. Focus on the development and actual use of multifunctional large unmanned wingmen

Large unmanned wingmen usually have the advantages of high ceiling, long flight range, long flight duration and heavy load, which are suitable for long-distance combat. The United States is actively developing large unmanned loyal wingmen, such as the XQ-58A Val force, which can cruise at Mach 0.72, reach a maximum altitude of 13,000 meters, have a maximum range of 5,556 kilometers, a maximum takeoff weight of 2,700 kg,

and a payload of 540 kg. It can form a formation with F-35 fighters or F-22 fighters. To carry out long-range strike and reconnaissance missions, which can be commanded by manned aircraft?

China should also speed up the research and development and practical use of large multi-functional unmanned wingmen with high ceiling, long range, long endurance and heavy payload, which requires continuous upgrading of existing UAV platforms to improve their tactical and technical performance and fully tap the combat potential of large multi-functional unmanned wingmen.

VII. CONCLUSIONS

This paper focuses on the research status, application cases and key technologies of the US military in the field of MUAV cooperative operations. The results show that the US military has long been committed to the research and development of MUAV cooperative operations in the future strong confrontation combat environment, and has made great progress in the exploration of cooperative operations theory, the research and development of key technology platforms, and the innovation of operation modes. By dynamically tracking the current situation of the US military in the field of MUAV cooperative operations, the direction and suggestions are presented for China to strengthen the construction of MUAV cooperative operations.

REFERENCES

[1] Dong Kangsheng, Hu Weibo, Shen Yanming et al. Development trend and enlightenment of intelligent unmanned air combat equipment of U.S. military [J]. Modern Defense Technology, 2022, 50(04): 28-37.

[2] LI Lei. Development analysis of typical foreign manned/unmanned Aerial vehicle cooperative combat projects [J]. Unmanned Systems Technology, 2020, 3(04): 83-90.

[3] Jiang Peng, Wang Rui, Zheng Lihui et al. Research status and development trend of manned/Unmanned Aerial Vehicle cooperative combat abroad [J]. Ordnance Automation, 2023, 42(03): 84-89.

[4] WANG Ruijie, WANG Dechao, FENG Lu et al. [4] Research on the development of foreign UAV swarm warfare style and anti-swarm strategy [J]. Modern Defense Technology, 2023, 51(04): 1-9.

[5] FOUSE S, CROSS S, LAPIN Z J. DARPA's Impact on Artificial Intelligence [J]. AI Magazine, 2014, 20 (4) : 135-136. Association for the Advancement of Artificial Intelligence, 2020, 41(2): 3–8.

[6] Congressional Research Service. Unmanned aircraft sys - tems: Roles, missions, and future concepts [EB/OL]. 2022-07-18. https://sgp. fas.org./CRs/weaps/r47188.pdf.

[7] GUNZINGER M, REHBERG C, COHN J, et al. An Air Force for an Era of Great Power Competition[M]. Washington DC: Center for Strategic and Budgetary As- sessments, 2019.

[8] Sun Linhui, Li Xun. Research status, hotspots and development trends of international manned/Unmanned Aerial Vehicle cooperative Operations: BiblioShiny Visual Analysis based on WoS data from 2000 to 2022 [J]. China-arab States Science and Technology Forum (Chinese and English), 2023(05): 48-58.

[9] Xu Liang, Pan Xuanhong, Wu Ming. Research on Manned/Unmanned Aerial Vehicle Cooperative anti-submarine warfare mode [J]. Chinese Ship Research, 2018, 13(06): 154-159.

[10] LUO Xuefeng, LEI Yongchun, FAN Jun. Review of foreign research on cooperation between manned helicopter and UAV [J]. Helicopter Technology, 2018(03): 61-67.

[11] LUO W, WEI R X. Intelligent Path Planning for Manned/Unmanned Aerial Vehicle Cooperative Strike [J]. Control Theory and Applications, 2019, 36(07): 1090-1095.

[12] WANG Ronghao, GAO Xingyu, XIANG Zhengrong. Review of manned/Unmanned Aerial Vehicle cooperative system and key technologies [J]. Journal of Ordnance Equipment Engineering, 2023, 44(08): 72-80.

[13] DU Zhuang, LIU Gang. Research on Key Technologies of Manned and Unmanned Aerial Vehicle Cooperative Combat System [J]. Science and Technology Innovation and Application, 2018(24): 139-140.