AI Chinese sign language recognition interactive system based on audio-visual integration

Xiaoran He
Tianjin University of Science and Technology
Tianjin, China
2489615190@qq.com

Yangde Lin
Tianjin University of Science and Technology
Tianjin, China
linyangde@mail.tust.edu.cn

Zhiyuan Hu
Tianjin University of Science and Technology
Tianjin, China
1476754996@qq.com

Xintong Xu
Tianjin University of Science and Technology
Tianjin, China
1607784937@qq.com

Rui Xu
Tianjin University of Science and Technology
Tianjin, China
1258968133@qq.com

Weiming Xiang
Tianjin University of Science and Technology
Tianjin, China
3113126511@qq.com

2023 IEEE International Conference on Electrical, Automation and Computer Engineering (ICEACE) | 979-8-3503-0961-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICEACE60673.2023.10442295

***Abstract*—Sign Language recognition is an important research direction in the field of computer vision. It is very important to recognize sign language efficiently and accurately to promote the two-way communication between the healthy people and the deaf-mutes so as to help the disabled and other people to integrate into the society harmoniously. There are two types of sign language recognition: Static Sign Language and dynamic sign language. The former can be recognized by image classification network and other technologies, and it is relatively mature for now. At present, the research in this field is not perfect at home and abroad. In this paper, an integrated sign language recognition system based on Yolov5 target detection algorithm combined with LSTM network and OpenPose technology is proposed, eventually the system will be deployed on the raspberry pie to increase portability. In order to improve the accuracy and efficiency of model recognition, this paper adopts data enhancement algorithm and changes the skeleton, detection head and loss function of Yolov5 model, finally, the model recognition accuracy of 98.87% was obtained on the open source Chinese sign language data set recognition test set of China University of Science and Technology. Our experiments show that our improved model has strong dynamic sign language recognition ability, and we hope that this can provide an improved idea for future sign language recognition, especially dynamic sign language recognition.**

Keywords*—Sign Language recognition, target detection, data enhancement, Sign Language Recognition System, Chinese sign language

I. INTRODUCTION

According to the World Hearing Report released by the World Health Organisation, one fifth of the world's population is currently hearing impaired, with hearing loss affecting more than 1.5 billion people globally, and by 2050, it is expected that one quarter of the world's population will have hearing problems and nearly 2.5 billion people will suffer from some degree of hearing loss. As of today, there are about 27.8 million hearing impaired people in the country and about 200,000 new people are added every year. Hearing impairment has already caused great disturbance to the communication of information for the hearing impaired and continues to affect all aspects of their life, education and employment. It is urgent to solve the communication barriers of the hearing impaired people so that they can lead a normal life.

It is well known that the most common sign language communication for hearing impaired people has a certain threshold, and they cannot communicate with normal people who have not learnt sign language. For the analysis of existing sign language information communication solutions[1], compared with the immaturity of electromechanical acquisition sign language translation technology, mechanical vision solutions facing the camera sign language scene limitations, "sign language translator" in the daily use of hearing-impaired people stand out. Therefore, a sign language recognition interactive system based on the Internet of Things and artificial intelligence technology, which consists of user APP, Web management terminal, information processing terminal and intelligent sign language interpreting glove hardware, was born[2].

II. PROJECT SUMMARY

This project is dedicated to creating an audio-visual integrated AI Chinese sign language recognition interactive system, with intelligent sign language interpreter as the main hardware product, based on the Internet of Things and artificial intelligence technology, consisting of four parts: the interpreter and APP, as well as the WEB management terminal[3] and information processing terminal[4].

1)Bluetooth connection will be used between the interpreter and the mobile phone APP, the Web management terminal will be used for the administrator to promote and manage the product, and the information processing terminal will be used to recognise the sign language.

2)The hearing impaired person holds the interpreter in his hand and expresses himself through sign language, the APP recognises the sign language through the information processing terminal and translates it into text to be displayed on the mobile phone, and at the same time plays sound to facilitate the hearing impaired person to receive the information; on the other hand, the APP recognises the voice of the hearing impaired person and translates it into text to be used by the hearing impaired person to receive the information.

3)The transformed information will be transmitted to the

management terminal for iterative optimisation of the product.

This project combines the functions of the four modules of the sign language recognition system with each other, and tries its best to solve the two-way communication between the hearing-impaired person and the listener.

III. SOFTWARE DESIGN

Product technology design on the combination of hardware and software, relying on intelligent sign language recognition technology and voice recognition technology to achieve the translation of sign language, software design on the development of websites[5] and apps for multi-application[6], the front-end using the mainstream Vue framework, the back-end using the Django framework, the service side, i.e., the database is used MySQL for data storage, the overall software design follows the principle of multi-application and easy operation to ensure that the client is convenient to use. The overall software design follows the principle of multi-application and easy operation to ensure that the client side is easy to use, and the hardware device adopts a card-type computer - Raspberry Pi[7] for functional deployment to complete the management and scheduling of the complex task of sign language recognition, which ensures that the functionality is complete and at the same time is more conducive to the user to carry and use. The company also seeks to further fill the gap of communication aids for the deaf and able-bodied.

A. Intelligent sign language recognition technology

Intelligent sign language recognition technology is the core of AI sign language interaction platform, and its flow chart is shown in Fig1.

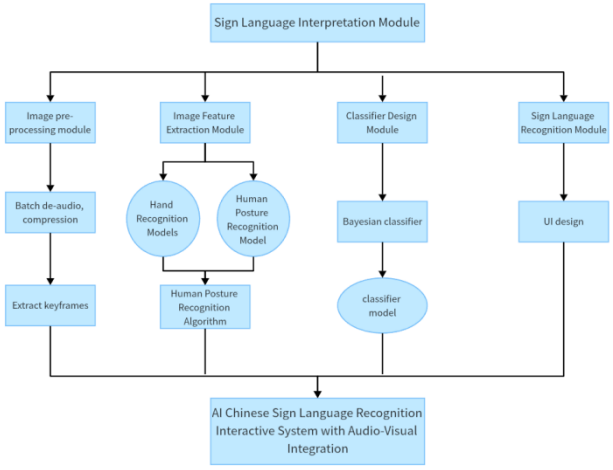

Fig. 1. Flow chart of intelligent sign language recognition technology

1) LSTM algorithm

RNN has been successful in speech recognition, machine translation, computer vision, etc. One of its significant advantages is that it can handle inputs of different lengths and effectively extract inter-frame temporal features.LSTM, as an improvement of RNN, adds a processor for judging whether the information is useful or not, and thus LSTM is commonly used in temporal classification.LSTM is not only able to sense temporal changes in sign language LSTM is not only able to sense temporal changes in sign language, but also able to learn the correspondence of gesture changes, thus further improving sign language classification. We constructed an LSTM-based sign language recognition model with the motion trajectories of four skeleton joints as inputs, and the experiments show that the results are satisfactory. The following is a graph of the effect of skeleton joint point detection，and its flow chart is shown in Fig2.

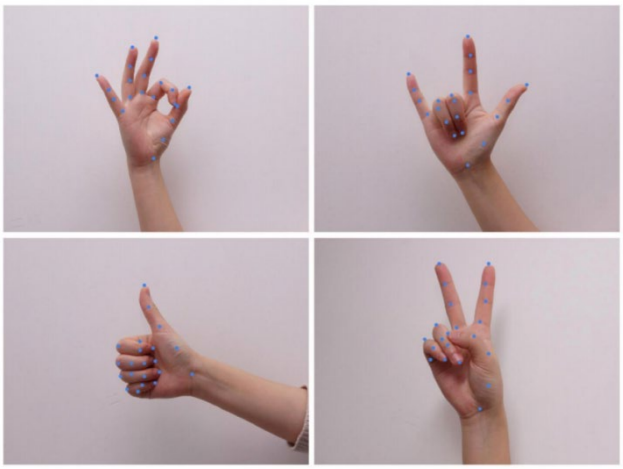

Fig. 2. Effect diagram of skeleton joint detection

When the hand is recognised first, the extended finger parts will be recognised and analysed individually, the coordinate points of these recognised parts will be used as inputs to the LSTM, and the type of sign language will be used as labels for the supervised learning, and the final network will be obtained after a number of iterations of the network in several layers. The following is the theoretical diagram of the LSTM algorithm where x represents the input and h represents the output of the layer,A represents an LSTM unit of the neural network，and its flow chart is shown in Fig3.

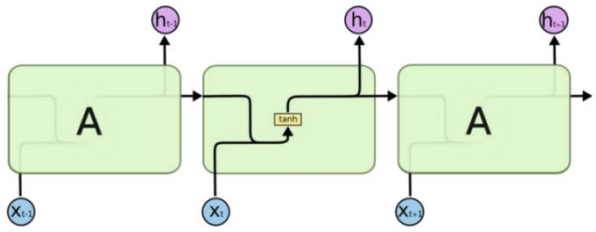

Fig. 3. LSTM algorithm diagram

2) Target Detection Assisted Algorithm (SSD) and Model Fine-tuning

The SLTM model alone is sometimes not effective in detecting sign language, especially when the movement is not large, and the translation results are not satisfactory. We have fine-tuned the network by borrowing ideas from the target detection algorithm (SSD).

SSD algorithms are a very famous class of algorithms in target detection, which are characterised by their speed while ensuring recognition accuracy, especially for small targets, and this is the core of our model optimisation.

Based on the SLTM network architecture, we change the activation function before layer to layer to the activation function of SSD, and the experiment shows that the model optimisation is feasible. The following is the SSD activation function.

The meaning of this function is that when the absolute value of the input x is less than 1, it outputs 0.5x squared, and in all other cases it outputs the absolute value of x -0.5. The

platform uses Tensorflow 2.6 to build the neural network, and after data fitting, our sign language recognition accuracy reaches 9%, and the recall rate reaches 92%, which is ideal.

B. Speech Processing Technology

In order to reduce the cost of voice processing, we use a third-party API[10], "convert sound into text, let your application grow ears", Baidu speech recognition technology through the Baidu voice open platform to provide developers with accurate, free, secure and stable services. Baidu's speech recognition technology adopts a simpler and more effective method than the mainstream speech recognition system, they use deep learning algorithms similar to neural networks to replace the previous recognition module, which significantly improves the recognition efficiency,and its flow chart is shown in Fig4.

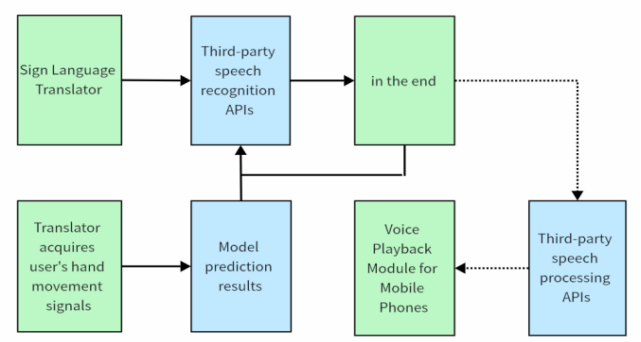

Fig. 4. Speech processing part flow chart

C. OSS Object Storage Service Technology

OSS has a platform-independent RESTful API interface that can store and access any type of data in any application, at any time, in any location, using the AliCloud management console to complete the OSS operation process is as follows.

Our product stores the model in the OSS as a separate API call, on the one hand, to solve the problem of server pressure, on the other hand, let the model as a pluggable application can be easily called and integrated, and its flow chart is shown in Fig5.

Fig. 5. OSS operation flow chart

D. Multi-terminal application development technology

The front-end uses Vue framework, the back-end uses Django framework, the database uses MySQL, using uni-app hybrid development platform for development, one-stop generation of multi-application: website, APP, applet. The following is the system design architecture diagram.

IV. Hardware Design

Product Structure Design Concept: For hearing impaired people, their needs are focused on the experience of information communication, and the problem of information communication affects their life, education, employment and personal safety. The common solution at this stage is for the hearing impaired to buy and wear hearing aids. The bad thing is that the price of hearing aids in the market is extremely high and they are not suitable for every hearing impaired person. In contrast, the use of sign language greatly facilitates the daily life of the hearing impaired. In view of the above, we conceived a "Magic Box" which is easy to carry and easy to communicate with the hearing impaired people.

Firstly, considering the problem of carrying, we designed the product to be similar to the size of a mobile phone, and there is a lanyard in the upper right corner of the product, which can be used with keychain, neck rope, etc., which is very convenient for people to carry; secondly, we chose the Raspberry Pi as the main design base, with the Raspberry Pi camera, Raspberry Pi display, which greatly reduces the cost of the product under the guarantee of the quality, and it is also more closely related to our design concept. Our design concept. On the one hand, we hope that the hearing impaired can use this product to strengthen their communication with people and actively integrate into the society; on the other hand, we hope to expand the language environment of sign language, so that sign language can enter into people's lives and integrate into people's lives. In this regard, the original intention and concept of our design is to enhance the relationship between the hearing impaired and people and society, to create a large environment for the sign language culture, and to make the sign language culture closer to people's lives,and its flow chart is shown in Fig6.

Fig. 6. Three-dimensional product visualization

Based on the Raspberry Pi and the camera to complete the hardware design, after deploying the project to the Raspberry Pi, the camera will be used to identify the sign language and convey the message that the deaf person wants to express.

V. Experimental Tests and Results

A. Experimental contents

The audio-visual AI Chinese sign language recognition interactive system adopts a combination of hardware and software. The hardware part is mainly the Raspberry Pi used to capture sign language images and the camera. The software part is mainly used to process the video images through FFmpeg, and then configure the development environment of Python3.9 under Anaconda, then combine with Cmake to compile the OpenPose model, and finally combine with the image algorithms of OpenCV in the PyCharm compiler to achieve all the programs of the sign language image recognition system compilation.

1) Hardware Environment

The main hardware devices used for sign language image acquisition are laptop camera and Raspberry Pi camera. The detailed parameters of the hardware environment in which the programme runs are as follows.

OS: Windows 11, 64bit

CPU: Intel(R) Core(TM) i7-10750H CPU@2.60GHz

GPU: NVIDIA GeForce GTX 1660 Ti

Memory: 16G

2) Software environment

Video processing tools: FFmpeg

Integrated development environment: PyCharm Professional, Anaconda3

Programming language environment: Python3.9

B. Experimental process

The audio-visual AI Chinese sign language recognition interactive system, based on OpenPose human posture open source model[11] and YOLOv5[12], LSTM and other neural networks for migration learning, training, optimisation of the network structure to train the hand model to detect the video and image, and then the digital features for the classifier model prediction, the prediction results will be presented in the form of text.

This experiment is based on OpenCV, YOLO, by capturing the image captured by the camera to identify the three-dimensional coordinates of the key points of the hand, the following figure shows the schematic diagram of the key points of the hand as well as the hand contour recognition,and its flow chart is shown in Fig7,8,9.

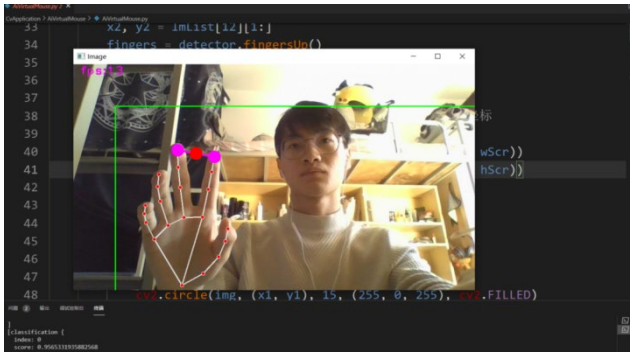

Fig. 7. Key Point detection of hand

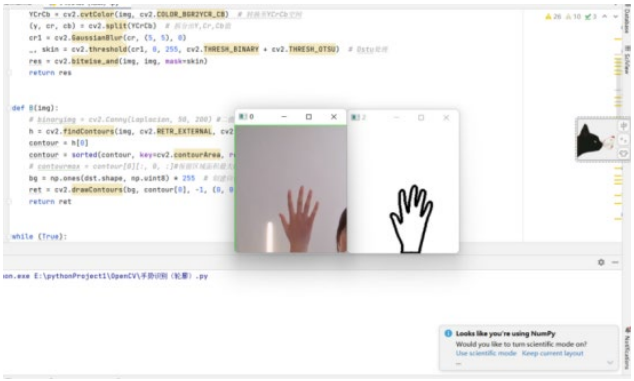

Fig. 8. Hand contour detection diagram 1

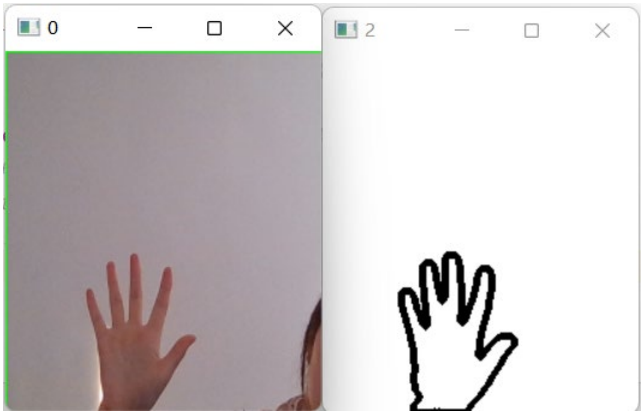

Fig. 9. Hand contour detection diagram 2

The design and development of the voice module, through the voice module to identify the common words of Chinese sign language based on Mandarin, and support the implementation of voice input to text and other functions, the following is the core code of the voice module,and its flow chart is shown in Fig10.

```
import pyttsx3


engine = pyttsx3.init()

rate = engine.getProperty('rate')
engine.setProperty('rate', rate)

engine.say(f"{sentence}")
engine.runAndWait()
```

Fig. 10. Voice to text function core code

Based on OpenPose neural network small dataset sign language detection, in the OpenPose design to solve the distance and angle of the formula and method, ultimately because of individual differences each person's skeleton may be different, currently optimised for the distance ratio (i.e., the distance between the 3-4 key points of the small arm and the length of the neck 0-1 key points of the distance ratio). The following is a simple demonstration of sign language as well as digital recognition results,and its flow chart is shown in Fig11,12.

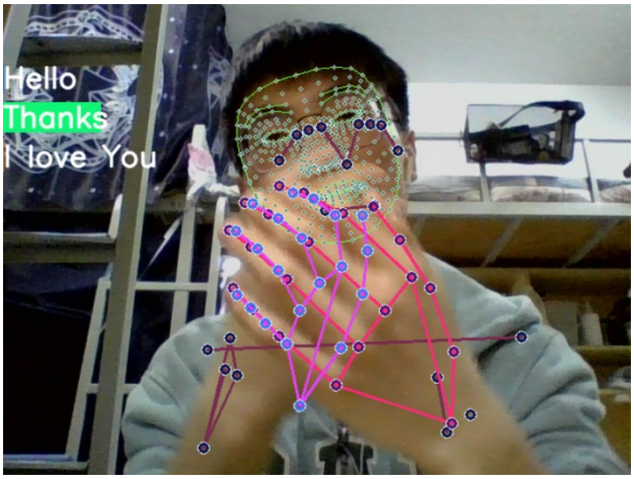

Fig. 11. Simple Sign Language recognition effect chart

Fig. 12. Digital recognition effect chart

C. Experimental results

The sign language recognition human-computer interaction platform based on OpenCV and YOLO has been fully implemented, and is able to recognise 25 commonly used sign language gestures, with an accuracy of about 96%. Considering that with the increase in the number of sign language recognition, the overlap between different sign language gestures is also getting higher and higher, and the effect of simply adopting YOLO may not be good, so the team will consider joining OpenPose, LSTM neural network to assist in evaluating human posture, and has already achieved sign language detection on a small dataset with an accuracy rate of about 97%.

Experimental validation on a self-built dataset, the number of iterations is 100, our model for the detection of small targets of sign language location loss is less than 0.03, confidence loss is less than 0.01, classification loss is less than 0.01, the final detection accuracy reaches 98.87%, close to 100%, and the recall rate reaches 99.98%. This shows that our model works well for static sign language recognition, and its flow chart is shown in Fig13.

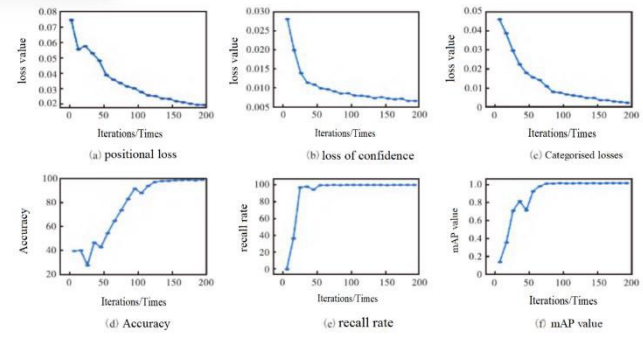

Fig. 13. Training set experimental results

D. Feasibility Analysis

The AI Chinese sign language interaction system studied in this project is to build a hardware and software platform using deep migration learning, Web development and Raspberry Pi technologies, based on the open source Chinese sign language dataset, combined with data enhancement methods. At the same time, the platform is capable of recognising commonly used words in Mandarin-based Chinese sign language and supports functions such as voice input to text, which can effectively promote two-way communication between deaf and normal people. The platform is different from the traditional static sign language recognition method. Since sign language is a series of movements consisting of rapid movements with similar characteristics, the traditional static sign language recognition method is difficult to deal with the complex vocabulary

expressions and large-scale changes in dynamic sign language hand movements. This system is to use the method of dynamic sign language recognition, based on neural networks such as LSTM, YoLoV5, OpenPose, etc., using Pytorch, OpenCV and other open-source frameworks for migration learning, initially locating and identifying 50 commonly used sign language vocabularies to achieve the effect of sign language recognition, and using speech processing libraries to achieve the function of recording to text, and the recognition results to speech. Compared with static sign language, dynamic sign language has a large vocabulary, more varieties, rich expressions, more practicality, and is the current hot spot of sign language recognition research. After summarising and sorting out the sign language recognition methods and techniques in recent years, considering the accuracy as well as the cost of implementation, we adopt Long Short-Term Memory Network (LSTM) and Target Detection Algorithm (SSD) as the core algorithm of sign language recognition and make algorithmic improvement based on them.

The part of our product used to voice processing is only the user side, so in order to reduce costs, using access to third-party API services. "Baidu's speech recognition technology provides developers with accurate, free, secure and stable services through the Baidu Speech Open Platform, which converts sound into text and lets your apps grow ears. Baidu's speech recognition technology adopts a simpler and more effective approach than mainstream speech recognition systems, and they use deep learning algorithms similar to neural networks to replace the previous recognition module, which significantly improves recognition efficiency.

The platform uses Tensorflow 2.6 to build an improved LSTM model based on SSD, collects data based on a miniature camera, hosts the model with the help of OSS service for sign language prediction, and translates the sign language into speech with the help of a third-party speech processing service, and uses APPs and mini-programmes as the carriers for the users to use, and at the same time, it develops a backend management system for the administrators to use. The technologies used in the platform are all open source, free of charge, with high security and development.

VI. Conclusion

In our work, we constructed a complete sign language recognition system for the dynamic sign language recognition problem. Our model combines the ability of LSTM networks to store transient temporal information and the ability of YoLoV5 to quickly and accurately capture sign language actions, while adding OpenPose to assist in evaluating human poses. Based on this, we performed data augmentation on the open-source Chinese sign language dataset from CUHK, which ensures that the model is able to learn adequately and improves the robustness of the model to a certain extent. We conducted experiments on the dataset using migration learning techniques and obtained a model recognition accuracy of up to 98.87% on the test set. Finally, we have built a complete sign language recognition system by combining migration learning along with web development and Raspberry Pi deployment techniques. Possible future research can be carried out in the direction of adding attentional mechanisms to neural networks, modelling sign language action timing information, and sign language data enhancement.

Acknowledgments

The writing is set aside at the end of the article, and my thoughts are mixed. I once read a sentence: all experience is learning. Wreaking summer and winter, whether it is joy or pain, all experience, in me is a gift. All encounters are treasures for me. All the bonds in the past four years may not be remembered for a lifetime, but I am definitely grateful for a lifetime. Heron Island Ming Ming, thank you for Huayuan four years together.

I wish to go through the years, in all things in all beings to do people. I will tell the world what it means to be brave. The world is a long way away, and the road is long and the horses are in a hurry. I would like to take this opportunity to wish you all the best in your encounters:

The sky is high, the sea is wide, and all things are favourable.

The landscape has a way, sooner or later we will meet again.

References

[1] Mary B,M. C A,Robert C, et al. Changes in discourse informativeness and efficiency following communication-based group treatment for chronic aphasia[J]. Aphasiology,2023,37(3).

[2] Julia R ,Lorraine J O ,Adelheid K , et al. Deaf and hard-of-hearing patients are unsatisfied with and avoid German health care: Results from an online survey in German Sign Language[J]. BMC Public Health,2023,23(1).

[3] Huang W ,Yang P . Application Analysis and Research of Artificial Intelligence Technology in the Creative Stage of Web Design[C]//Wuhan Zhicheng Times Cultural Development Co., Ltd..Proceedings of 5th International Workshop on Education Reform and Social Sciences (ERSS 2022).BCP Social Sciences & Humanities,2022:9.DOI:10.26914/c.cnkihy.2022.067738.

[4] Brother Kogyo Kabushiki Kaisha; "Computer Readable Recording Medium, Information Processing Terminal Device, And Control Method Of Information Processing Terminal Device" in Patent Application Approval Process (USPTO 20200285801)[J]. Computer Technology Journal,2020.

[5] Aditra G P ,Mahendra G I D . Web-Based System for Bali Tourism Sentiment Analysis during The Covid-19 Pandemic using Django Web Framework and Naive Bayes Method[C]//..Atlantis Press,2022.

[6] Zhang Q ,Yang S ,Ren R . Research on Uni-app Based Cross-platform Digital Textbook System[C]//International Association of Applied Science and Engineering.Proceedings of the 2020 3rd International Conference on Computer Science and Software Engineering (CCSE 2020). ACM, 2020:6.DOI:10.26914/c.cnkihy.2020.060802.

[7] Yang J. Raspberry Pi-based IoT Printing System[C]//Institute of Management Science and Industrial Engineering.Proceedings of 2019 3rd International Conference on Mechanical and Electronics Engineering(ICMEE 2019). Clausius Scientific Press, 2019:6. DOI:10.26914/c.cnkihy.2019.039715.

[8] Y.Hu ,W.Y.Shen ,W.Zheng . Fourier Descriptor Based Features for Static Sign Language Recognition[C]//Advanced Science and Industry Research Center.Proceedings of 2015 International Conference on Automation, Mechanical and Electrical Engineering(AMEE 2015). DEStech Publications,2015:7.

[9] Xingyu C,Junzhi Y,Zhengxing W. Temporally Identity-Aware SSD With Attentional LSTM[J]. IEEE Transactions on Cybernetics, 2019, 50(6).

[10] Li J T ,Cao K ,Liu Z , et al. Linking Consumer's Needs with Product Performance on E-shopping Platforms with Fashion Big Data(FBD) API Plugins[C]//Technical University of Liberec Faculty of Textile Engineering,Textile Bioengineering and Informatics Society.Textile Bioengineering and Informatics Symposium Proceedings(TBIS 2022) .Textile Bioengineering and informatics Society, 2022:11. DOI:10.26914/c.cnkihy.2022.067958.

[11] Ma Y ,Wei C ,Long H . A Gait Recognition Method Based on the Combination of Human Body Posture and Human Body Contour[C]//Hubei Zhongke Institute of Geology and Environment

Technology.Proceedings of the 2nd International Conference on Artificial Intelligence and Computer Science (AICS 2020), 2020:8. DOI:10.26914/c.cnkihy.2020.028983.

[12] Tao W ,Qixin Z ,Jiacheng W , et al. An improved YOLOv5s model for effectively predict sugarcane seed replenishment positions verified by a field re-seeding robot[J]. Computers and Electronics in Agriculture, 2023,214.