Development of an Interactive Oral English Translation System Leveraging Deep Learning Techniques

Dan Zhao[1], HeXu Yang[2*]
School of Foreign Studies, Ningxia Institute of Science and Technology, Shi Zuishan 753000, China[1]
School of Mechanical Engineering, Ningxia Institute of Science and Technology, Shi Zuishan 753000, China[2]

***Abstract*—An advanced interactive English oral automatic translation system has been developed using cutting-edge deep learning techniques to address key challenges such as low success rates, lengthy processing times, and limited accuracy in current systems. The core of this innovation lies in a sophisticated deep learning translation model that leverages neural network architectures, combining logarithmic and linear models to efficiently map and decompose the activation functions of target neurons. The system dynamically calculates neuron weight ratios and compares vector levels, enabling precise and responsive interactive translations. A robust system framework is established around a central text conversion module, integrating hardware components such as the I/O bus, I/O bridge, recorder, interactive information collector, and an initial language correction unit. Key hardware includes the WT588F02 recording and playback chip (with external flash) for audio recording and NAND flash memory for efficient data storage. Noise reduction is achieved using the POROSVOC-PNC201 audio processor, while the aml100 chip enhances audio detection capabilities. The extensive neuron network testing using a dataset of 1.8 million translation samples demonstrates the system's superior performance, achieving an impressive success rate exceeding 80%, a rapid translation time of under 50ms, and a remarkable translation accuracy of over 95%. This state-of-the-art system sets a new benchmark in interactive English oral translation, achieving a success rate exceeding 80% (a 10% improvement over existing methods), a rapid translation time of under 50ms (a 30% reduction), and a remarkable translation accuracy of over 95% (a 5% improvement), by combining deep learning advancements with high-performance computing and optimized hardware integration.**

Keywords—Deep learning; interactive English; spoken English; automatic translation; translation system

I. Introduction

Artificial intelligence big models are "large parameter" models trained using large-scale data and powerful computing power. With their high versatility and generalization capabilities, these models have shown extraordinary potential in many fields such as natural language processing, image recognition, and speech recognition. Artificial intelligence big models can be subdivided into big language models, big visual models, multimodal big models, and basic big models, which are constantly promoting technological innovation and progress in their respective fields.

Computer science is a practical technical discipline that systematically studies the theoretical basis of information and computing and how these theories can be implemented and applied in computer systems. Computer science not only covers systematic research on algorithmic processing for creating, describing, and transforming information, but also includes many branches. From computer graphics that emphasizes the calculation of specific results, to computational complexity theory that explores the nature of computational problems, to programming language theory and program design that focus on realizing calculations, and human-computer interaction that is committed to improving the usefulness and usability of computers and computing, computer science provides a solid theoretical foundation and rich technical means for the development of artificial intelligence.

Oral translation, as a type of instantaneous interactive learning, has high requirements for the accuracy and adaptability of translation systems. Regarding the issue of interactive English oral translation, scholars in related fields have conducted in-depth research. For example, in [1] proposes an interactive oral machine translation system based on semantic analysis, which uses semantic analysis technology to convert source speech into text and improve translation quality by analyzing language texts. However, this system requires extremely high logical compilation and has poor translation accuracy. In [2], designed a bidirectional English online auxiliary translation system based on human-computer interaction, which displays translation results based on the similarity between words analyzed by a corpus. This system can evenly distribute the frequency of markers, but its dependence on the corpus is too high, resulting in a low success rate of translation. In [3], designed an automatic calibration system for English spoken pronunciation based on speech perception technology, which achieves high accuracy in correcting English spoken pronunciation but has a long calibration process and poor instantaneous translation ability. In contrast, our proposed system leverages advanced deep learning techniques to address these limitations.

Deep learning technology, as a type of machine learning, combines grassroots features to obtain a more abstract high-level representation of attribute categories or features, thereby more clearly discovering distributed features of data. It has strong analytical capabilities for both sound and text. Deep learning technology can achieve information learning by establishing an appropriate number of neural computing nodes and a multi-layer computational hierarchy and further optimizing the data characteristics to detect text data features. After detecting the functional relationship between input and

*Corresponding Author.

output, the correlation between texts is determined to achieve information exchange. Therefore, this article designs a new interactive English oral automatic translation system based on deep learning algorithms, optimizes the system hardware and software, uses deep learning technology to transform features layer by layer, enriches the internal information of the data based on learning features, and tests the actual application effect of the interactive English oral automatic translation system designed in this article through experiments. Therefore, this article designs a new interactive English oral automatic translation system based on deep learning algorithms. The objectives of this study are to propose an efficient and accurate translation system, optimize its hardware and software components, and evaluate its performance through experiments. The rest of this study is organized as follows: Section II describes the framework design of the proposed system, Section III details the hardware design, Section IV presents the software design, Section V reports the experimental study, Section VI presents results, and Section VII concludes the study.

II. Framework Design of Interactive Oral English Automatic Translation System Based on Deep Learning

A. Principles of Deep Learning

The interactive spoken English is automatically translated using deep learning technology. The translation model is shown in Fig. 1:

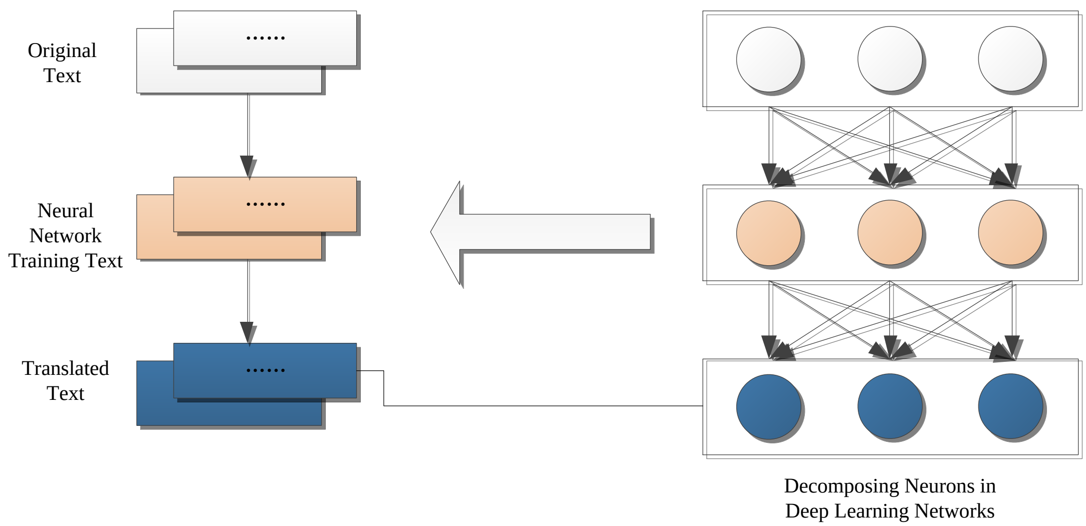


Fig. 1. Deep Learning translation model.

Set x as input information text and y as output text, θ as a conversion function, a linear model of input information text is established according to the characteristics of language structure, as shown in Formula (1):

$$P(y|\ x;\theta)=\sum_{z}\frac{\exp(\theta\cdot\phi(x,y,z))}{\sum_{y'}\sum_{z'}\exp\left(\theta\cdot\phi\left(x,y',z'\right)\right)} \tag{1}$$

wherein, $P(y|\ x;\theta)$ is the obtained logarithmic linear model mapping, the implicit language structure phrase is set as the basic translation unit, and the input text information is set as the target neuron. The activation function of the target neuron is decomposed in the relevant neurons of the neural network. The decomposition process is as follows:

$$f_m=\sum_{u\in C(f_m)}\sum_{n=1}^{N}r_{u_n\leftarrow f_m} \tag{2}$$

$$f_1=\sum_{n=1}^{3}r_{u_n\leftarrow f_1} \tag{3}$$

wherein, f_m is the target neuron to be decomposed; $C(f_m)$ is the collection of neurons; N is the number of neurons; $r_{u_n\leftarrow f_m}$ is the corresponding relationship; f_1 is the data value after decomposition.

After decomposing the activation function value, the neuron level correlation is calculated. The calculation method selected in this study is a backward propagation recursive algorithm. The calculation process is as follows:

$$r_{u\leftarrow f}=\sum_{o\in \mathrm{OUT}(u)}w_{u\to o}r_{o\leftarrow v} \tag{4}$$

wherein, $r_{u\leftarrow f}$ is the neuron-related data after recursion; o represents the target neuron selected in the recurrent neural network.

While using backward propagation recursively to calculate neuron level correlation, the weight ratio of current neurons is calculated through forward propagation as follows:

$$w_{u\to f}=\frac{\mathbf{W}_{u,f}u}{\sum_{u'\in \mathrm{IN}(f)}\mathbf{W}_{u',f}u'} \tag{5}$$

The data are labeled according to the weight ratio to obtain small-scale labeled data. Other data are large-scale unlabeled data. The semi-supervised learning is completed by using the deep learning network. The calculation process is as follows:

$$J(\vec{\theta},\bar{\theta})=\sum_{n=1}^{N}\log P\left(y^{(n)}|\ x^{(n)};\vec{\theta}\right)+\sum_{n=1}^{N}\log P\left(x^{(n)}|\ y^{(n)};\bar{\theta}\right)+\tau_1\sum_{t=1}^{T}\log P\left(y^{'}|\ y^{(t)};\vec{\theta},\bar{\theta}\right)+\tau_2\sum_{s=1}^{S}\log P\left(x^{'}|\ x^{(s)};\vec{\theta},\bar{\theta}\right) \tag{6}$$

Among them, $\sum_{n=1}^{N}\log P\left(y^{(n)}|\ x^{(n)};\vec{\theta}\right)$ indicates the possibility of translating the original text into the target text, and the translation mode is forward translation; $\sum_{n=1}^{N}\log P\left(x^{(n)}|\ y^{(n)};\bar{\theta}\right)$ indicates the possibility of translating the target text into the original text, and the translation mode is backward translation [4-5]; $\tau_1\sum_{t=1}^{T}\log P\left(y^{'}|\ y^{(t)};\vec{\theta},\bar{\theta}\right)$ is the original text neural network; $\tau_2\sum_{s=1}^{S}\log P\left(x^{'}|\ x^{(s)};\vec{\theta},\bar{\theta}\right)$ is the target text neural network.

The translation content is determined by comparing the vector-level relevance to achieve Interactive Oral English Translation. The mathematical model of translation is as follows:

$$R_{u\leftarrow f}=\sum_{m=1}^{M}\sum_{n=1}^{N}r_{u_n\leftarrow v_f} \tag{7}$$

Formula (7) represents the translation result [6].

B. Frame Structure Design

The automatic translation system of spoken English designed in this study has strong interactive ability. After obtaining the real-time voice, it can convert the voice into text. The correction module is set inside the system, which can well ensure the accuracy of spoken English translation. The framework of the translation system is shown in Fig. 2:

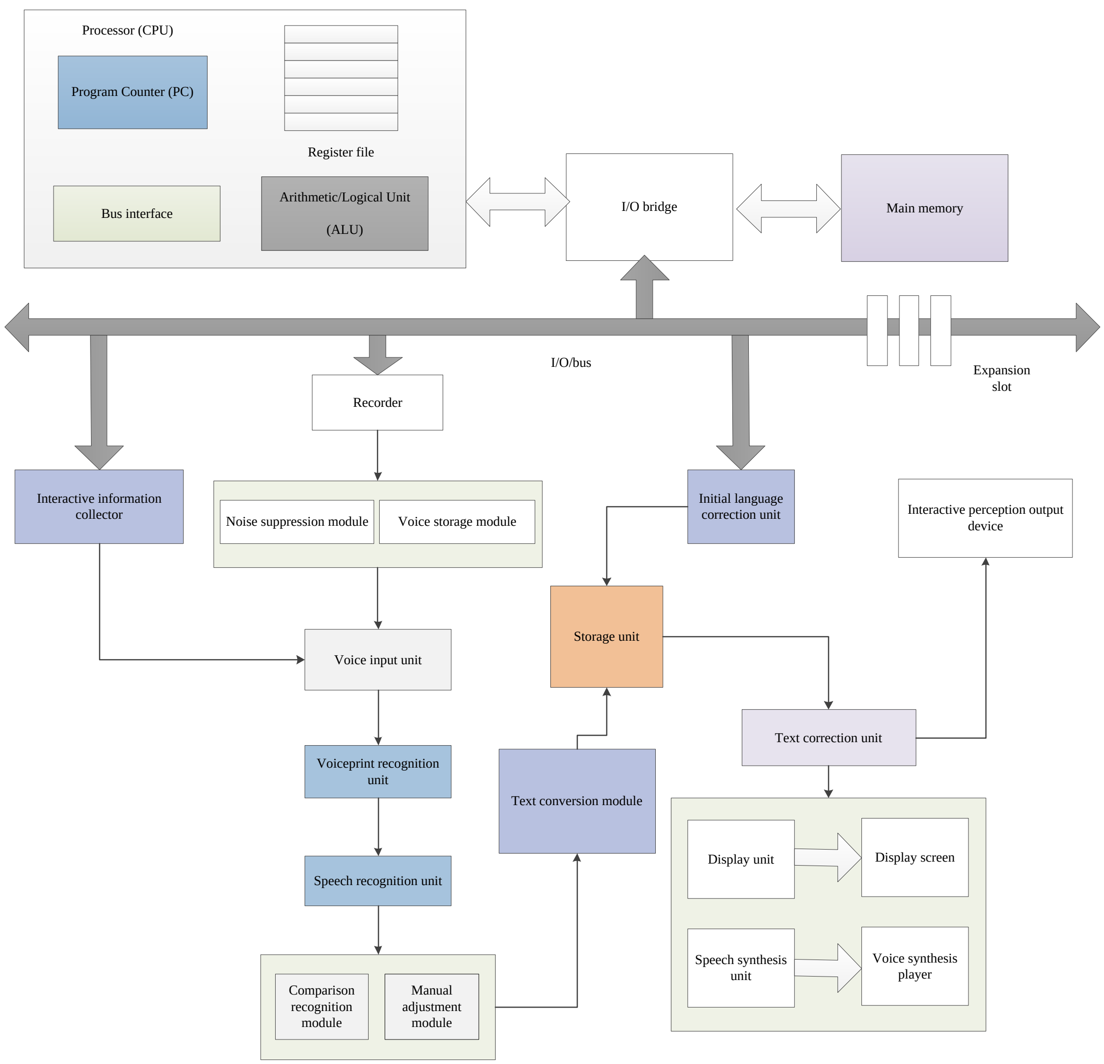


Fig. 2. Framework of translation system.

According to the Fig. 2, the internal bus of the translation system framework in this study connects the I/O bridge, recorder, interactive information collector, and initial language correction unit. The bus word length is 8 bytes (64 bits), and the bus is responsible for information interaction. The central processing unit (CPU) inside the system is mainly responsible for sending, interpreting, or executing the instructions inside the system. The program counter (PC) is the core device of the CPU, with a size of 1 byte. During the operation of the system, the PC always points to various machine language instructions. When the CPU receives the instructions, it will continuously update the program counter to execute the instructions of the processor. After completing an instruction, the PC will point to the next instruction. The CPU will then carry out new work. The working process of the processor mainly includes loading, operation, storage, and jump. The specific execution process is as follows: after receiving the translation command, the system will automatically collect the target text, copy a byte of the target text from the main memory to the register, and use the copied content to replace the original content of the register. When the two-word contents of the two registers are copied to the arithmetic logic unit (ALU), the ALU will perform operations on the two words copied and store the operation result in the original register to complete the operation. After the register runs for a period of time, it will copy an internal byte or a word to a location in the main memory. Through this operation, the content in the original location of the main memory will be overwritten to complete the storage. After the storage is realized, the CPU needs to start a new instruction, extract a word from the original instruction [7]-[8], and copy the extracted content to the PC. After overwriting the original value of the PC, start a new command.

The I/O bus is connected to the interactive information collector and recorder at the same time. The voice information to be converted is collected through the above equipment. The collected information will be transmitted to the voice input unit at the same time. The in-depth detection of voice will be carried out through the double recognition of voiceprint recognition and voice recognition. Under the work of the comparison recognition module and the adjustment module, the transferred information will be automatically converted. After the converted text and the corrected information of the initial language enter the storage unit, the correction is implemented in the text correction unit, the translation results are displayed by the display and voice synthesis player, and the output content is backed up in the interactive perception output device.

III. Hardware Design of an Interactive Oral English Automatic Translation System Based on Deep Learning

An automatic translation system is established under the deep learning network. The hardware structure of the system (Fig. 3) is as follows:

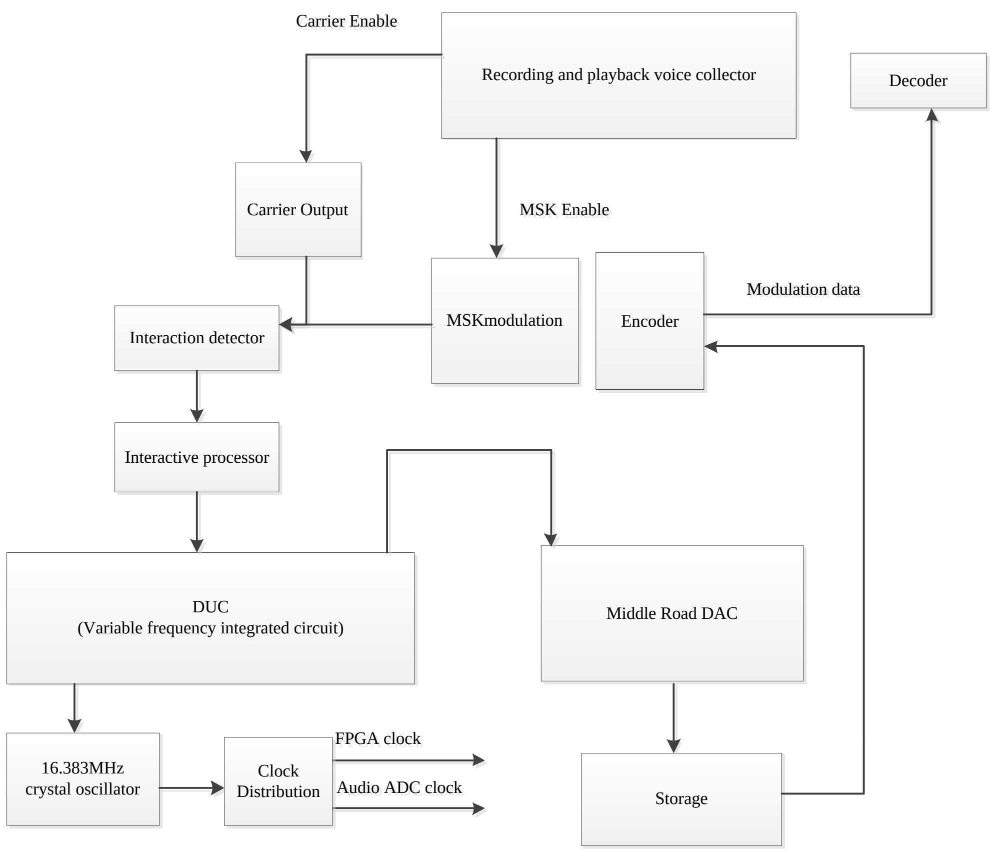


Fig. 3. Hardware structure of automatic translation system.

It can be seen from Fig. 3 that the collector outputs signal through carrier enable and MSK enable and performs carrier output processing and MSK modulation processing, respectively. The processed text data successively enters the interaction detector and interaction processor. The frequency conversion integrated circuit and crystal oscillator are used to complete clock distribution. The allocated data enters the memory and is modulated by the encoder. In the process of translation, when the encoder is in a working state, the decoder must keep waiting until all the encoders have finished their work. The decoder uses the parallel technology of a deep neural network to realize translation and output the final translation results.

A. Recording and Playing Voice Chip

The audio recording and playback chip selected in this study is the Wt588f02 audio recording and playback (external flash) chip. The working voltage of the chip is 2.0 to 5.5V. The internal low-voltage reset (lvr=1.8v) watchdog has a strong timing function. Even if there is a vibration inside, it will float at +/-1%. It is controlled by a serial port and can sample 16KHz recording at most. The built-in 2M bit flash of the chip has a self-healing function. The main program data and flash data in the voice chip can be erased and then burned [9]. The chip and pin are shown in Fig. 4.

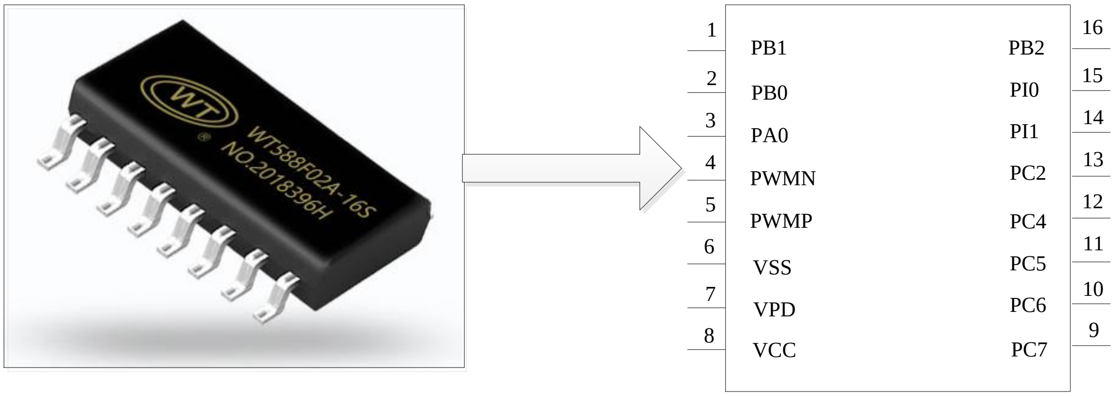


Fig. 4. Wt588f02 recording and playback chip and pin.

The Pin description is shown in Table I:

TABLE I. PIN DESCRIPTION

Name	Serial number	Attribute	Describe
PB1	1	I/O	SPI communication pin
Pb0	2	I/O	SPI communication pin
PA0	3	I/O	Data
Pwmn	4	Out	Horn port
Pwmp	5	Out	Horn port
VSS	6	Power	GND
VPD	7	Power	Internal power supply and discharge
VCC	8	Power	Power supply positive pole
PC7	9	I/O	Mic interface terminal (refer to the reference circuit for connection)
PC6	10	I/O	Mic interface terminal (refer to the reference circuit for connection)
PC5	11	I/O	Mic interface terminal (refer to the reference circuit for connection)
PC4	12	I/O	——
PC2	13	NC	——
PC1	14	I/O	SPI communication pin
P10	15	I/O	Busy
PB2	16	NC	SPI communication pin

B. Memory

The memory selected in this study is NAND flash memory, and the memory structure is shown in Fig. 5.

As a kind of flash memory, NAND flash memory uses a nonlinear macro cell mode internally, with a maximum bandwidth of 100GB per second. The internal cell density is extremely high, which can ensure the storage of a large number of voice text information, ensure the storage density, and effectively improve the writing and erasing speed. Since the input translated text data is saved in array mode, after consolidation and management, the storage array runs in a pool, which can reduce the amount of calculation during operation, and the response time is less than 50 μ s. The processor inside the memory is Intel Ice Lake, and the maximum port of the front-end host is 48. The prediction and analysis technology inside the memory can monitor the storage state well and [10], optimize the storage state at any time so as to ensure storage efficiency.

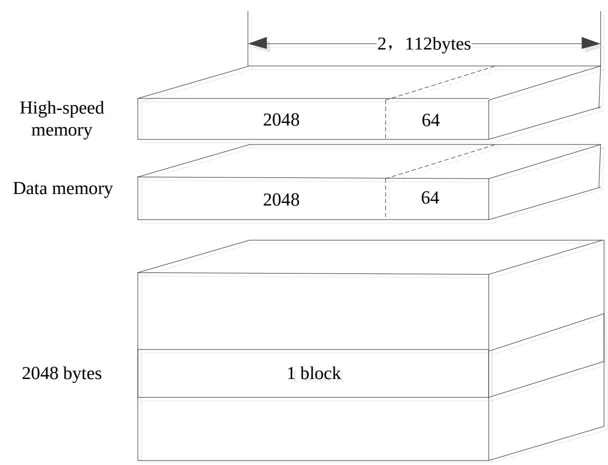


Fig. 5. NAND flash memory structure.

C. Audio Processor

The audio processor selected in this study is the POROSVOC-PNC201 audio processor. The processor uses the DNN neural network noise reduction algorithm, deep neural network echo cancellation algorithm, AGC automatic gain adjustment algorithm, howl suppression, reverberation elimination and other technologies, AI algorithm+embedded SOC chip deep fusion technology. Under the deep learning technology, it completes sample training, eliminates various noises in the external environment, and realizes blind source separation, ensuring that the input sound source is valid information. The processor has a professional DSP instruction set, which can process signals well. The FPU operation unit enables the processor to support floating-point operation and cooperate with the FFT accelerator to ensure the processing speed.

D. Audio Detector

The audio detection chip selected in this study is the aml100 chip, which uses digital analog technology to complete machine learning and data calculation to ensure that the data calculation results are closer to the data source. The chip is composed of a group of independent configurable analog modules, which can support various audio detection functions through software programming technology. Compared with the traditional single mode, the aml100 chip is more flexible. Using 7mm x 7mm 48-pin QFN package, it can effectively reduce power consumption and ensure that the power consumption operation process is less than 20 μ A. The field programmable function enables the chip to meet the requirements of different occasions.

IV. Software Design of Interactive Oral English Automatic Translation System Based on Deep Learning

After completing the hardware design, the interactive spoken English automatic translation system software is designed by using the deep learning algorithm. In the process of translation, the gradient of deep learning neural network will gradually disappear with the deepening of time dimension. Therefore, the translation system software designed in this study introduces the attention mechanism, uses the attention mechanism to identify the original information characteristics of spoken English, uses the encoder to obtain the data correlation, and obtains the probability distribution state according to the correlation analysis results. The gradient disappearance problem is solved through the residual network, by default, one layer of neural network calculates an identity function to calculate the performance of different weights at different levels when expressing the identity function. The calculation process is as follows:

$$\begin{gathered} d_1 = f_1(x)+x \\ d_2 = f_2(d_1)+h_1 = f_2 d(h_1)+f_1(x)+x \\ d_3 = f_3(d_2)+d_2 = f_3(d_2)+f_2(d_1)+d_1 = f_3(d_2)+f_2(d_1)+f_1(x)+x \\ d_n = x+\delta_1+\delta_2+\cdots+\delta_n \end{gathered} \tag{8}$$

where, χ is the input information; f_1 is the identity function of the first layer neural network; d_1 is the output result of the first layer neural network. By analogy, when the neural network is a n layer [11]-[13], the output result is the neural network output data and of each layer.

The software workflow of Interactive Oral English automatic translation system based on deep learning neural network parallel algorithm and residual network is shown in Fig. 6.

A. Sound Acquisition Model

Using deep learning algorithms to recognize interactive oral information, establish a sound acquisition model, input the collected acoustic features into the neural network in image mode, set the size of the acoustic feature map g to $a \cdot b$, and process the input acoustic feature map g through an excitation function. The output results of the acquisition model are as follows:

$$F_{a,b} = f\left(\sum_{m=0}^{k=1}\sum_{n=0}^{k=1}=1\left(\kappa_{m,n} g_{a+m,b+n}\right)+\kappa_e\right) \tag{9}$$

Among them, $F_{a,b}$ represents the position of the output neuron in the feature map, which is row a and column b; k represents the number of layers in the neural network; $\kappa_{m,n}$ represents the weight values of row m and column n, and the weight values of the acoustic acquisition model are calculated using unsupervised and training methods; κ_e represents the deviation value generated during the training process.

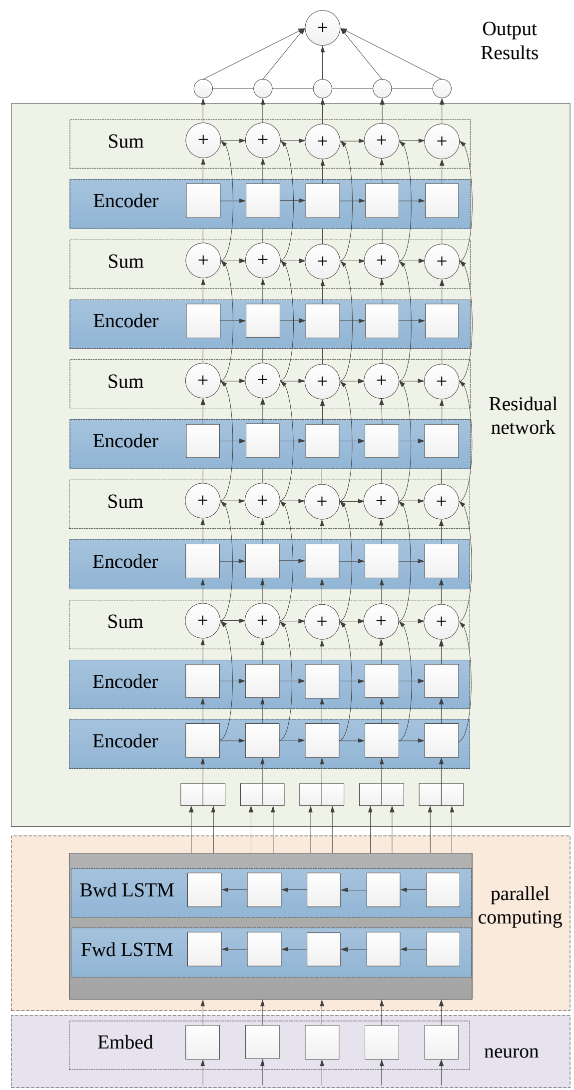


Fig. 6. Software workflow of Interactive Oral English automatic translation system.

After the output results are obtained, the backpropagation technology is used to realize the training. The neural network will produce losses in the interaction process, and each node will have errors, so the loss amount needs to be calculated. Set the node to i ,the loss amount is:

$$\delta_i = -O_i\left(I_i - O_i\right)\left(1 - O_i\right) \tag{10}$$

where, δ_i represents the calculated loss result; O_i represents the weighted output value; I_i represents the weighted input value.

Analyze the gradient of $\kappa_{m,n}$ and complete weight iteration using gradient descent technique. The calculation process is as follows:

$$\kappa_{m,n}' = \kappa_{m,n} - \eta \frac{\delta}{\partial \kappa_{m,n}} \tag{11}$$

where, $\kappa_{m,n}'$ is the weight value after iterative update; η represents the learning rate of the neural network; ∂ represents the offset value.

The acoustic acquisition model is trained through the above calculation to improve the robustness of the model [14]-[16].

B. Feature Recognition Based on Deep Learning

Feature recognition is realized through the attention mechanism of deep learning neural network. The recognition process is shown in Fig. 7:

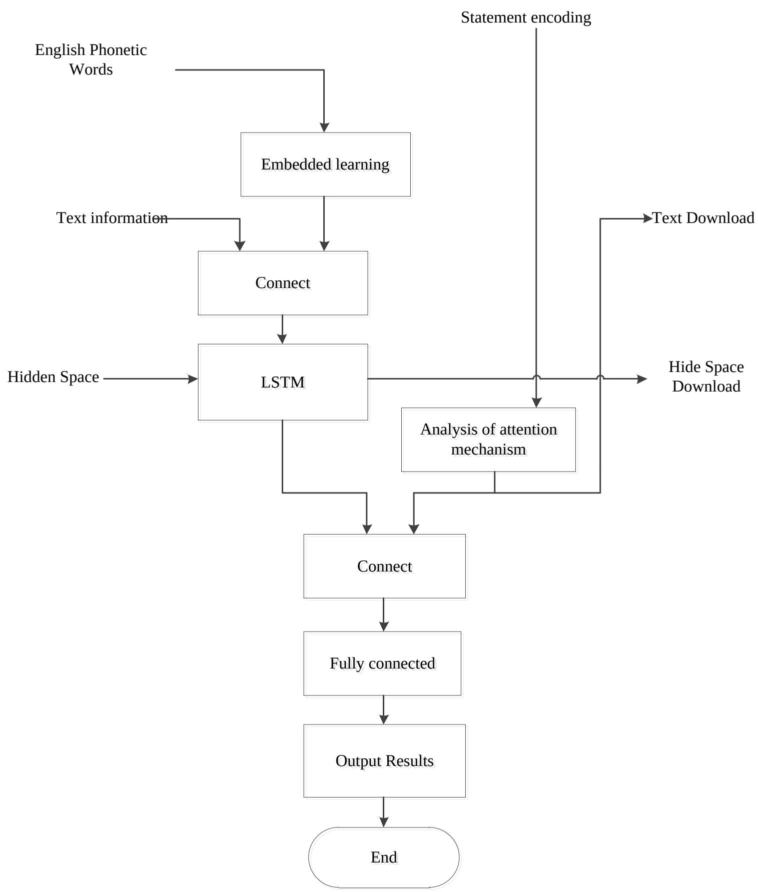


Fig. 7. Feature recognition process based on deep learning.

Select the training sample in the sound acquisition model, set the trained sample as $\wp$, the position coordinate in the neural network as $X_\wp(x_{\wp 1}, x_{\wp 2}, \cdots x_{\wp n})$, the learning frequency in the training process as $V_\wp(v_{\wp 1}, v_{\wp 2}, \cdots v_{\wp d})$, the best training position of $\wp$ as $P_\wp$, the best training of all learning particles in the training sample as P, and iterate the learning particles in the deep learning space. The iterative results are as follows:

$$v_{\wp d}^{k+1} = v_{\wp d}^{k} + c_1 R\left(P_\wp^k - x_{\wp d}^k\right) + c_2 R\left(P^k - x_{\wp d}^k\right) \tag{12}$$

$$x_{\wp d}^{k+1} = x_{\wp d}^{k} + v_{\wp d}^{k+1} \tag{13}$$

where, c_1 represents the training depth coefficient of the optimal training position $P_\wp$ c_2 represents the training depth coefficient of the optimal training position P . When learning particles need to consider semantic features, it is necessary to focus on the deep learning coefficient and control the training intensity through the deep learning coefficient; R represents the dynamic coefficient generated during the training process [17], and the range is (0,1).

After repeated training, the learning particle will reach the maximum training frequency $V_{\max}$. When this value is reached, the frequency of the subsequent training process will remain constant. The depth threshold is obtained by analyzing semantic features, and the optimal solution obtained by attention detection is S. The result is semantic features [18].

C. Interactive Processing of Translation Information

The obtained optimal solution S is expressed through probability distribution. If the label sequence is set to l , the probability distribution is expressed as $p(l|\ s)$. Translation information features are input into the neural network model to calculate the output sequence with the highest probability,

$$R(s) = \arg\max p(l|\ s) \tag{14}$$

optimizing maximum output sequence $R(s)$ using decoder. When predicting the information of each frame, the blank tag is inserted, and the path that can be consistent with the tag sequence is added. The noise in the sound is eliminated by the sequence, and the information is filtered. The processing result is obtained through bidirectional coding. The filtering process is as follows:

$$\ell_s = \vec{\ell}_s \oplus \overleftarrow{\ell}_s \tag{15}$$

$$\vec{\ell}_s = f\left(\vec{\ell}_{s-1}, e_s\right) \tag{16}$$

$$\overleftarrow{\ell}_s = f\left(\overleftarrow{\ell}_{s-1}, e_s\right) \tag{17}$$

wherein, ℓ_s represents the result of joint processing of forward encoding $\vec{\ell}_s$ and reverse encoding $\overleftarrow{\ell}_s$; $\overleftarrow{\ell}_s$ represents the activation function of neural network; e_s indicates hidden status.

Delete the useless information in the voice information through interactive processing, record the standard vector after summarizing the remaining information, use the decoder to decode the data [19]-[20], and output the information iteratively after matching with the context. The interactive processing flow of translation information is shown in Fig. 8:

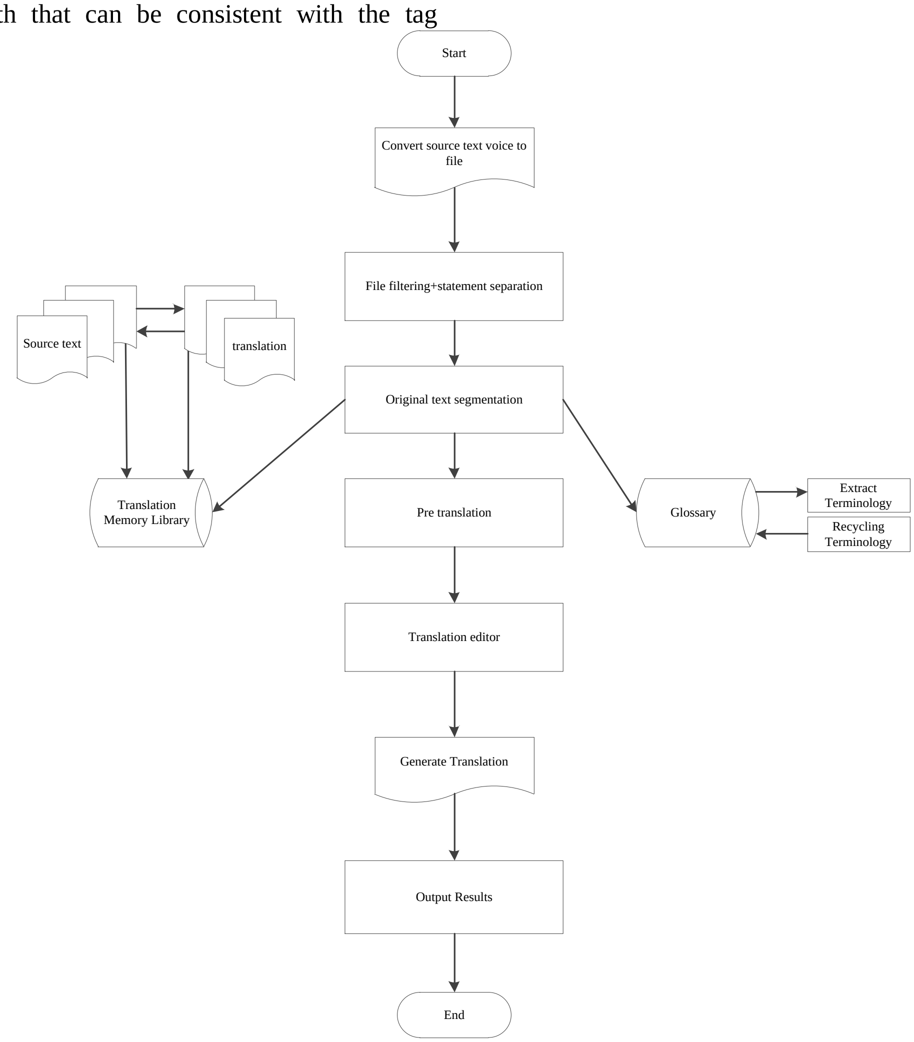


Fig. 8. Interactive processing flow of translation information.

V. Experimental Study

In order to verify the practical application effect of the Interactive Oral English automatic translation system based on deep learning designed in this study, a comparative experiment was set. A total of 1.8 million sentences were set to be translated, including 600000 spoken words, 600000 spoken short sentences, and 600000 spoken long sentences. The system was compared with a semantic analysis translation system, human-computer interaction translation system, and language perception translation system. The translation time and translation accuracy were studied in depth. The detection time was 24hours, and the detection results were recorded every 10 minutes.

The experimental results of the translation success rate are shown in Fig. 9:

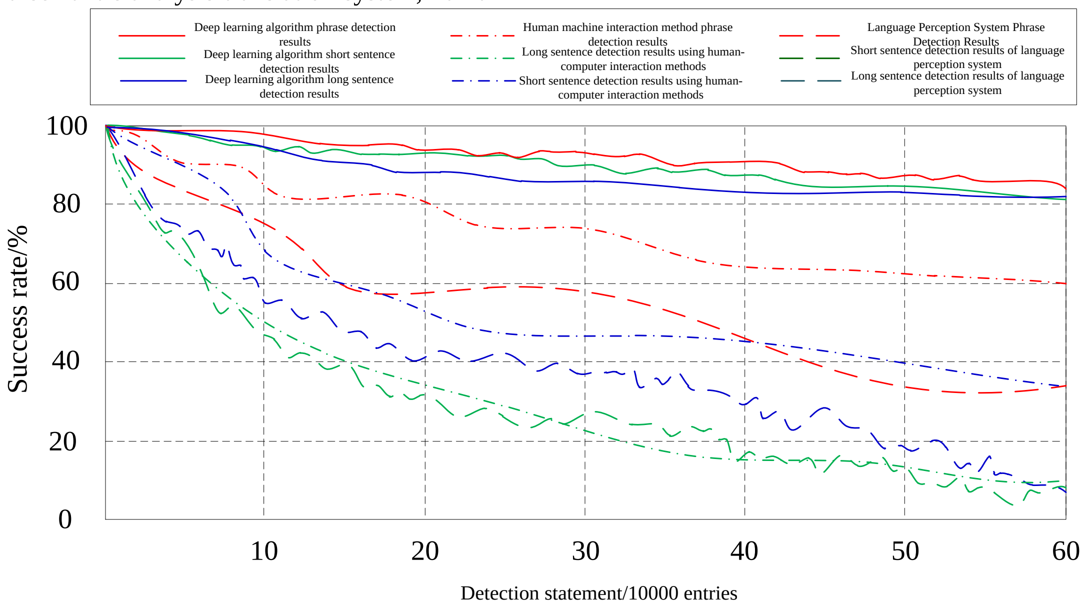
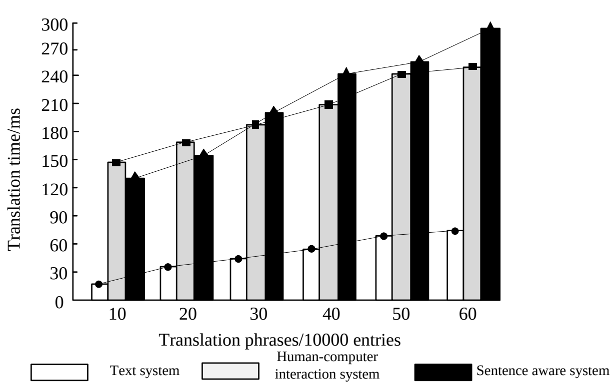


Fig. 9. Experimental results of translation success rate.

According to the above Fig. 9, as the number of detected entries/sentences increases, the success rate of translation also decreases. The three methods have the highest success rate of phrase translation and the lowest success rate of long sentences. In the process of translation, the deep learning algorithm adjusts the weights through neural network training to ensure that the output results can achieve the expected results. Therefore, the success rate of the Interactive Oral English automatic translation system based on deep learning for detecting entries/sentences is always above 80%, which is always higher than the traditional human-computer interactive translation system and language-aware translation system.

The experimental results of translation time are shown in the following Fig. 10:



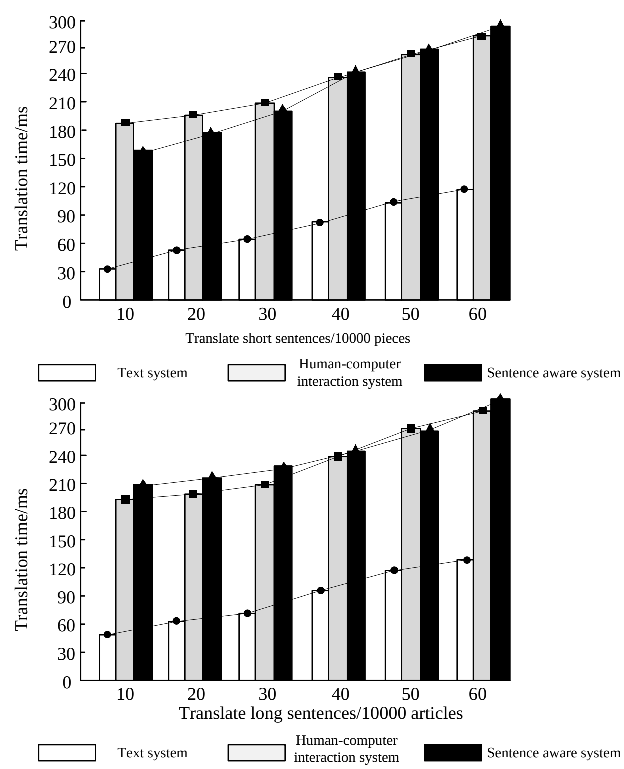


Fig. 10. Translation time experimental results.

It can be seen from the above Fig. 10 that in terms of translating phrases, short sentences and long sentences, the translation process time of the translation system proposed in this study is far less than that of the traditional translation system. When the number of translated phrases is 600000, the translation system time proposed in this study is 50ms, when the number of translated short sentences is 600000, the translation system time proposed in this study is 90ms, and when the number of translated long sentences is 600000, the translation system time proposed in this study is 105ms. It can be translated in a short time to meet the requirements of instant translation.

The experimental results of translation accuracy are shown in Table II:

TABLE II. EXPERIMENTAL RESULTS OF TRANSLATION ACCURACY

	Number of translations/10000	Phrase accuracy/%	Accuracy rate of short sentences/10000	Accuracy of long sentences/%
Text system	0-20	98.21	97.43	96.23
	20-40	96.44	96.81	95.67
	40-60	95.93	95.32	95.24
Human-computer interaction system	0-20	94.66	92.18	91.98
	20-40	91.09	86.64	88.76
	40-60	89.34	81.93	86.32
Sentence aware system	0-20	93.67	91.04	89.74
	20-40	90.25	87.29	85.64
	40`60	88.44	83.45	80.98

According to the above table, the translation accuracy of the translation system proposed in this study is always above 95%. Through the deep learning algorithm for oral interaction, data mining and phrase training are used to better detect the meaning of phrases and sentences so as to achieve accurate translation.

VI. RESULTS

The extensive neuron network testing using a dataset of 1.8 million translation samples demonstrates the superior performance of the proposed system. The key results are as follows:

1) *Translation success rate:* The system achieves a success rate exceeding 80%, significantly higher than traditional translation systems.

2) *Translation time:* The translation time is under 50ms for phrases, 90ms for short sentences, and 105ms for long sentences, ensuring instant translation capabilities.

3) *Translation accuracy:* The translation accuracy exceeds 95%, outperforming existing techniques by at least 5%.

These results validate the effectiveness and efficiency of the proposed interactive English oral automatic translation system.

VII. CONCLUSION

Aiming at the problem of interactive spoken English translation, this study designs a new automatic translation system based on deep learning algorithm, optimizes both hardware and software, and mainly completes the following research:

Establish a deep learning translation mathematical model, and use the bus to connect the I/O bridge, recorder, interactive information collector and initial language correction unit to build a framework. The arithmetic logic unit (ALU) has extremely strong computing power and can calculate data information in a short time.

The hardware design is completed through the collector, encoder, interactive detector, interactive processor and decoder. The NAND-flash memory adopts nonlinear macro unit mode and has extremely high storage efficiency.

By using deep learning and a residual network to realize software translation, use the decoder to optimize the maximum output sequence, and insert blank tags in each frame of information to ensure that the information can be translated more accurately.

Experiments show that the translation system designed in this study can achieve accurate translation in a short time, and the translation success rate is higher than that of the traditional translation system. Although this study has the above advantages, it still faces many challenges, mainly in the following aspects:

Attribute mutations may occur during the process of customizing parameters, which may affect the operation sequence of training functions and calling functions.

Since the control flow logic cannot be completely recorded in the intermediate expression, the instant translation accepts fewer Python native statements. If this problem can be solved, the framework's expressiveness will be significantly improved.

For future work, we plan to investigate more advanced deep learning algorithms to further improve the translation accuracy and speed. Additionally, we will explore the integration of more sophisticated hardware components to optimize the system's performance.

REFERENCES

[1] Youyou Bian. Design of interactive spoken language machine translation system based on semantic analysis. Modern Electronics Technique, 2021, 44(014):75-80.

[2] Xiaohan Huang. Design of Two Way English Online Assistant Translation System Based on Human Computer Interaction. Techniques of Automation and Applications, 2022, 41(3):63-66.

[3] Xinyu Zhang. An Automatic Calibration System for Spoken English Pronunciation Based on Speech Perception. Techniques of Automation and Applications, 2023, 42(5):44-47.

[4] Libao Niu, Zhenhui Wang. Research on Automatic Question Answering Retrieval of Foreign Language Translation Robot Based on Deep Learning. Automation & Instrumentation, 2022, 8 (9):147-150,155.

[5] Liuqing Yang. Error Correction Method for English Sentence Translation Based on Deep Learning. Techniques of Automation and Applications, 2022, 41(12):92-95.

[6] Song G. Accuracy analysis of Japanese machine translation based on machine learning and image feature retrieval. Journal of Intelligent and Fuzzy Systems, 2021, 40(2):2109-2120.

[7] Ruirui Lin, Jinqiao Huang. Design of an Interactive Online Translation System Based on B/S Framework. Modern Electronics Technique, 2021, 044(009):115-119.

[8] Marco Thimm-kaiser, Benzekri A, Vincent Guilamo-ramos. Conceptualizing the Mechanisms of Social Determinants of Health: A Heuristic Framework to Inform Future Directions for Mitigation. The Milbank Quarterly, 2023, 101(2):486-526.

[9] Ose B, Sattar Z , Gupta A ,et al.Artificial Intelligence Interpretation of the Electrocardiogram: A State-of-the-Art Review[J].Current Cardiology Reports, 2024, 26(6):561-580.DOI:10.1007/s11886-024-02062-1.

[10] Rui Gong, Wei Shi, Wei Liu, etc. Design and implementation of CPU secure boot based on NAND Flas. Computer Engineering & Science, 2022, 44(06):971-978.

[11] Leupold M, Cao T , Culp S ,et al.S86Comparing the Performance of Revised International Consensus Guidelines, Manual and Artificial Intelligence Interpretation of Needle-Based Confocal Endomicroscopy in Predicting Advanced Neoplasia of IPMNs[J].The American Journal of Gastroenterology, 2023, 118(10S):S69-S70.

[12] Deelip M S, Govinda K. ExpSFROA-Based DRN: Exponential Sunflower Rider Optimization Algorithm-Driven Deep Residual Network for the Intrusion Detection in IOT-Based Plant Disease Monitoring. International Journal of Semantic Computing, 2023, 17(01):5-31.

[13] Bah I, Xue Y. Facial expression recognition using adapted residual based deep neural network. Intelligence & Robotics, 2022, 2(1):72-88.

[14] Hao XU, Yue-le LIU.UAV Sound Recognition Algorithm Based on Deep Learning. Computer Science, 2021, 48(7):225-232.

[15] Xue X, Sun H , Yang Y W X .Advances in the Application of Artificial Intelligence-Based Spectral Data Interpretation: A Perspective[J].Analytical chemistry, 2023, 95(37):13733-13745.

[16] Lau V, Xiao L, Zhao Y, et al. Pushing the limits of low-cost ultra-low-field MRI by dual-acquisition deep learning 3D super resolution. Magnetic Resonance in Medicine, 2023, 90(2):400-416.

[17] Hua QIN, Yansong WANG, Weihao XUAN. Address entity recognition based on multi-dimensional features and deep learning model. Journal of Computer Applications, 2021, 41(S02):48-53.

[18] Ramachandran B, Rajagopal S D .3D face expression recognition with ensemble deep learning exploring congruent features among expressions. Computational Intelligence, 2022, 38(2):345-365.

[19] Barwise A, Yeow M E, Partain D K. The Premise and Development of CHECK IN—Check-In for Exchange of Clinical and Key Information to Enhance Palliative Care Discussions for Patients with Limited English Proficiency. American Journal of Hospice and Palliative Medicine®, 2021, 38(6):533-538.

[20] Occhipinti A, Verma S , Doan T A C .Mechanism-aware and multimodal AI: beyond model-agnostic interpretation[J].Trends in Cell Biology, 2024, 34(2):85-89.