A Hybrid Interpretation Model Leveraging Improved Grey Wolf Optimization and Graph Attention Networks for Intent Recognition and Dynamic Interaction

Haibo Zhang
School of Humanities, Lishui University, Lishui, 323000, China
E-mail: haibozhang688@outlook.com

Keywords: intent recognition, independent dynamic interaction, interpretation model, IGWO

Received: May 20, 2025

The study presents an integrated interpreting model that combines intent recognition via Improved Grey Wolf Optimization (IGWO) with independent dynamic interaction via Graph Attention Networks (GAT). It builds an intent-relation graph, uses multi-head attention to capture dynamic links among intents, and leverages IGWO to adapt intent thresholds and attention-head weights. Goal: more reliable multi-intent recognition and better adaptation to changing contexts. On WMT14 (EN–FR) it achieves 94.37% accuracy (DNN & HMM 79.31%, EEMD 75.09%, LLM 64.97%). For 60-s audio it reaches 47.35 dB SNR (EMD 29.74 dB, LLM 26.72 dB); at 160 s it remains highest (47.68 dB). IGWO boosts accuracy via chaotic initialization and Gaussian mutation; a heterogeneous GAT models ternary relations. WMT14 and LibriSpeech are used for translation/ASR, and MixSNIPS/MixATIS for multi-intent understanding. After 500 iterations IGWO hits 92.91% (deep bidirectional pre-trained language model 69.86%, HMM 63.79%); recall exceeds 90% across datasets. Results indicate more accurate, natural translations and stronger handling of technical terminology.

Povzetek: Raziskava predstavi hibridni interpretacijski model, ki združuje izboljšano optimizacijo sivih volkov in grafne pozornostne mreže, kar omogoča večnamensko prepoznavo namenov ter prilagodljivo, naravno interakcijo v dinamičnih prevajalskih okoljih.

1 Introduction

In today's world, where globalization and digitalization are deeply integrated, real-time cross-language interaction has become indispensable across various fields. The demand for multilingual real-time interaction in cross-cultural meetings, international business negotiations, and emergency medical situations has been growing exponentially as the process of globalization accelerates [1]. Although existing interpretation models perform stably in general scenarios, they still rely on static rules and predefined interaction patterns. When faced with dynamic and ever-changing real-world environments, these models often expose issues such as translation delays, intent misjudgment, and rigid interactions due to a lack of context awareness and adaptive ability [2-3]. For example, dialogues with ambiguous intentions, terminology-heavy discussions, or semantic ambiguities caused by cultural differences can easily lead to reduced accuracy in interpretation results [4]. In recent years, speech recognition and neural machine translation technologies based on deep learning have significantly improved the benchmark performance of interpretation systems. However, their core bottleneck has gradually shifted to scene-specific intent recognition and independent dynamic interaction [5]. Intent recognition, as a natural language processing technique that understands the intent behind users' input texts or speech, has been widely applied in fields such as intelligent customer service and virtual assistants [6, 7]. By accurately understanding the user's intent, systems can provide more personalized and efficient services. L. L. Li et al. proposed a systematic review method for human lower limb movement intention, verifying the effectiveness of the method through analysis of the intention perception signals [8]. Independent dynamic interaction allows each module in a system to independently adjust and interact during operation, without relying on the states or behaviors of other components [9, 10]. With the development of intelligent technology in recent years, independent dynamic interaction has shown significant development trends and application potential in multiple fields. Some scholars have studied its advantages of independent interaction during operation. For instance, P. Sun et al. proposed a hybrid system to improve the dynamic stability of converters, analyzing its dynamic interactions. Experimental results demonstrated that the method possesses high accuracy and effectiveness [11]. Current translation technology has entered a new stage of development centered around large language models. For example, Y. Xiao addressed the issue of low accuracy in non-autoregressive generation in machine translation tasks. By exploring its expansion in areas such as grammar correction and text summarization, it was found

that non-autoregressive generation could promote its industrial application [12]. Xiang et al. put forward an integrated improved fusion model tailored to the characteristics of the Internet of Things platform. The model used a computer-aided translation system to extract specific words from the text corpus, and the results showed that the translation accuracy and recall rate of this method were both higher [13]. The study summarized the performance comparison of deep bidirectional pre-training language model (Bidirectional Encoder Representations from Transformers, BERT), deep neural network combined with hidden Markov model (Deep Neural Network and Hidden Markov Model, DNN-HMM), etc. and the results are shown in Table 1.

Table 1: Comparison of performance indicators and Limitations of existing methods

Category	Method	Performance index	Limitations
Intention recognition	BERT	Intent classification accuracy rate: 85%-90%	Insufficient semantic understanding in complex contexts, weak ability to recognize implicit intentions, and poor multilingual adaptability.
	Grey Wolf Optimization Algorithm	Optimization efficiency: Medium, with approximately 50 to 100 iterations	It is prone to fall into local optimum and has difficulty adapting to the demands of dynamically changing interpretation scenarios.
Dynamic interaction	Graph neural network	The accuracy rate of interaction relationship capture: 80%-85%	It has limited processing capabilities for long sequence dependencies, lacks a dynamic update mechanism, and has a slow response to real-time interaction scenarios.
	Recurrent neural network	Latency: approximately 300-500ms, signal-to-noise ratio: 7-10dB	There exists the problem of vanishing gradient, the ability of multimodal information fusion is weak, and it is difficult to adapt to complex interpretation scenarios.
Comprehensive model	DNN-HMM	Interpretation accuracy rate: 75%-80%, delay: 400-600ms	The model structure is fixed and cannot flexibly cope with multi-language and multi-domain scenarios, with poor intent understanding and interaction coordination.

As shown in Table 1, currently, some algorithms have difficulty identifying weak signals in complex environments, often resulting in the omission of key information. Most existing interpretation models deal with intent recognition and interaction respectively. This innovative model integrates an improved Grey Wolf Optimization (IGWO) algorithm for intent recognition with an independent dynamic interaction mechanism based on Graph Attention Networks (GAT), aiming to overcome the limitations of traditional algorithms in capturing intentions in complex environments. Real-world interpretation scenarios require handling multiple intents and complex semantic interactions. This proposed model ensures accurate intention recognition to support semantic understanding and promotes the development of intelligent response technology, which is both innovative and necessary.

Current interpretation models struggle to capture users' implicit intentions, leading to translations that deviate from the intended dialogue goals. For example, non-autoregressive generation can significantly improve the reasoning speed in tasks such as machine translation, but there is a loss of accuracy. Traditional systems rely on linear input-output mechanisms and lack the ability to dynamically adjust translation strategies based on real-time dialogue states, such as speaker emotions, environmental noise, and multimodal context. This limitation causes a significant drop in reliability in noisy environments or during multi-speaker exchanges [13-14]. Therefore, this study innovatively proposed to build a fusion interpretation model, which uses IGWO and GAT for intent recognition and the design of independent dynamic interaction mechanisms. At the same time, by introducing the intent recognition module to analyze the speaker's semantic intent in real time, the core information of the source language can be captured more accurately. The independent dynamic interaction mechanism allows different modules to be independently optimized and updated during the interaction process. The model innovatively employs IGWO with a dynamic weight adjustment mechanism and a population diversity optimization strategy to enhance the global search

capability and convergence speed of intent recognition, and GAT dynamically models the relationship between nodes through the attention mechanism. The study aims to improve the accuracy of the interpretation model in capturing the user's implicit intentions in complex scenarios through this fusion model, and offers robust data to support the advancement of intelligent interpretation technologies.

2 Methods and materials

2.1 Design of intent recognition algorithm based on IGWO

Intent recognition is an important task in natural language processing. It primarily analyzes text, speech, and other user inputs to understand underlying intentions or needs [15-16]. Although intent recognition demonstrates strong adaptability and interactivity, its effectiveness heavily depends on the quality and quantity of training data [17-18]. Moreover, intent recognition technologies face challenges with ambiguous, complex, or vague statements, leading to increased uncertainty and risk in recognition outcomes [19]. GWO efficiently optimizes the parameters of intent recognition models by simulating the hunting behavior of grey wolves. This study improves GWO by introducing a dimensional reasoning hierarchy search strategy, resulting in IGWO, which is applied to optimize intent recognition. The optimization process of IGWO is shown in Figure 1.

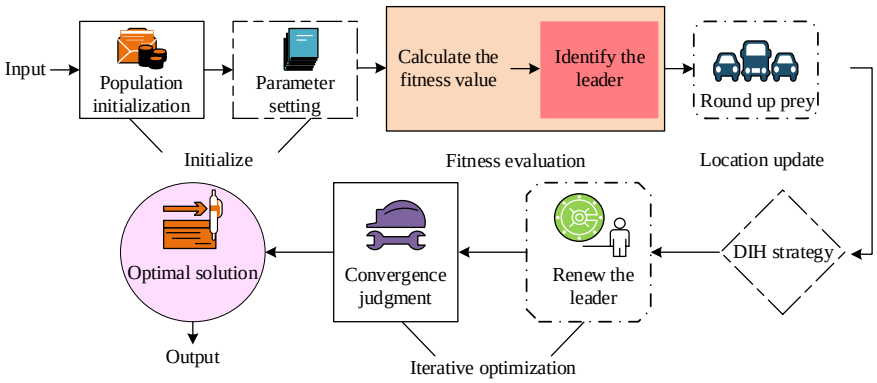


Figure 1: Schematic diagram of IGWO optimization process

In Figure 1, the position of each grey wolf represents a potential solution to the problem. The population initialization part randomly generates the grey wolf population and sets the population size based on the problem complexity. The study calculates the fitness value of each grey wolf to evaluate the quality of its solution. The complexity of the real-time input speech is used to dynamically adjust the convergence factor of GWO, and a random dropout mechanism is introduced in the GWO iteration to randomly mask the weight updates of some noise nodes. The cross-attention mechanism aligns speech, text, and temporal behavior features, generating a fused embedding vector as the fitness evaluation criterion for GWO. The cooperation of the grey wolf group searches for optimal feature weight distribution to enhance the consistency of cross-lingual intent expression. The calculation of the grey wolf group approaching and surrounding the prey is shown in Equation (1).

$$\begin{cases} \vec{D} = \left| \vec{C} \cdot \vec{X}_p(t) - \vec{X}(t) \right| \\ \vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \end{cases} \quad (1)$$

In Equation (1), $\vec{D}$ represents the position of the individual relative to the prey, t and $\vec{X}_p$ are the current iteration number and prey's position vector, $\vec{X}$ and $\vec{X}(t+1)$ represent the position vectors of the grey wolves and their position updates, $\vec{A}$ and $\vec{C}$ are coefficient vectors. In the intention recognition task, the dimension of each variable is consistent with the number of features. The calculation of the coefficient vectors is shown in Equation (2).

$$\begin{cases} \vec{A} = 2 \cdot \vec{\partial} \cdot \vec{r}_1 - \vec{\partial} \\ \vec{C} = 2 \cdot \vec{r}_2 \end{cases} \quad (2)$$

In Equation (2), $\vec{\partial}$ represents the convergence factor that linearly decreases from 2 to 0 as the iteration count increases, and $\vec{r}_1$ and $\vec{r}_2$ are random vectors within the range of $[0,1]$, the number of dimensions is consistent with the number of features. IGWO simulates the hunting behavior of grey wolves to efficiently find the optimal solution in the search space. Therefore, IGWO optimizes the intent recognition algorithm, and the optimized algorithm is shown in Figure 2.

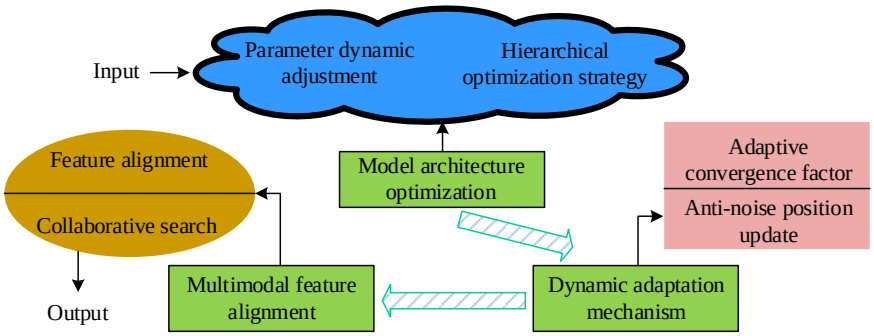


Figure 2: Optimization process of intent recognition algorithm based on IGWO

In Figure 2, the design process first treats the hyperparameters of intent recognition as optimization variables for GWO. The parameters are dynamically adjusted by the grey wolf group's cooperative search to find the optimal parameter combination. In the hierarchical optimization strategy, the decision wolves and auxiliary decision wolves represent the global and second-best solutions. The two wolves optimize the main network parameters of the model and adjust the weights of the local feature extraction module, respectively. The feature extraction calculation in intent recognition is shown in Equation (3).

$$TF-IDF(t,d)=TF(t,d)\times IDF(t) \quad (3)$$

In Equation (3), $TF(t,d)$ and $IDF(t)$ represent the term frequency of word t and the inverse document frequency of word d in document t. The dimension of $TF(t,d)$ is the statistical value of a single document and a single word, while $IDF(t)$ is the statistical value of a single word in the global document set. The calculation of the inverse document frequency is shown in Equation (4).

$$IDF(t)=\log\frac{N}{1+DF(t)} \quad (4)$$

In Equation (4), N and $DF(t)$ represent the total number of documents and the number of documents containing word t. N is the statistical value of the global document set, which is a fixed value, and the dimension of $DF(t)$ is the statistical value of a single word in the global document set. Then, the convergence factor of GWO is dynamically adjusted based on the complexity of real-time input speech, and the random dropout mechanism is introduced to mask the weight updates of noise nodes in GWO iterations. The cross-attention mechanism aligns temporal behavior features to generate a fused embedding vector used as the fitness evaluation criterion for GWO. The cooperation of the grey wolf group enhances the consistency of language intent expression. In the cross-attention module, the attention weights between different features are calculated to explore feature associations. During cross-attention fusion, features are weighted and summed according to the attention weights to achieve deep integration. Moreover, IGWO can optimize the attention parameters, enabling the model to better adapt to various intent recognition scenarios.

2.2 Design of independent dynamic interaction method based on GAT

The IGWO-based intent recognition algorithm leverages GWO's strong global search capability to prevent convergence to local optima. Additionally, IGWO has a straightforward structure and requires fewer adjustable parameters, enhancing its convenience and efficiency in intent recognition. Furthermore, independent dynamic interaction, as an interactive mode that enhances user experience through dynamic elements and user interaction functions, can independently perceive changes in user behavior and the environment, dynamically adjusting the interaction mode based on this information [20-21]. This study incorporates independent dynamic interaction into the construction of the interpretation model. The model dynamically adjusts translation results according to personalized needs, such as users' language habits and professional terminology preferences, providing translation services more tailored to users' needs. The modular architecture design and technical implementation path of the independent dynamic interaction unit are shown in Figure 3.

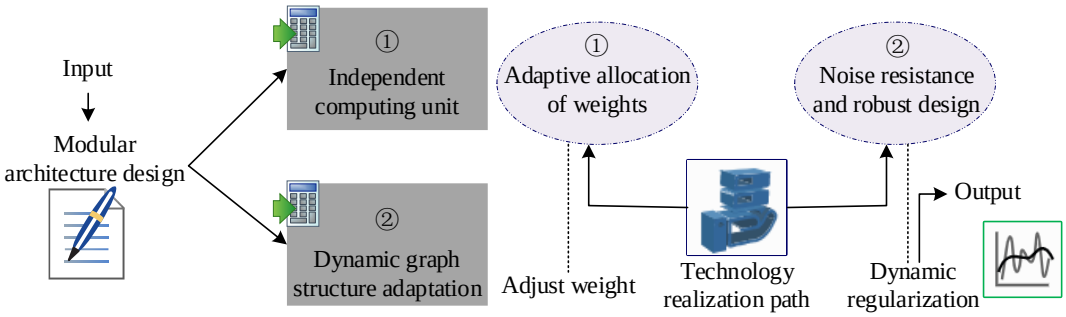


Figure 3: Independent dynamic interaction unit design and technical implementation path

Figure 3 illustrates the modular architecture and technical implementation path of the independent dynamic interaction unit. In the modular architecture design, the independent computation unit designs an independent GAT module for each interaction node, dynamically allocating the weight between nodes via multi-head attention mechanism and supporting real-time updates of local dynamic relationships. The adaptive weight allocation aligns multi-modal inputs via the cross-attention mechanism to generate dynamic feature fusion graphs, while IGWO adjusts the attention weights. This study uses Fitz's Law as the basis for modeling interactive behavior. The calculation equation of Fitz's Law describing the time law of human movement is shown in Equation (5) [22].

$$T_1 = a + b\log_2\left(\frac{d}{W} + 1\right) \quad (5)$$

In Equation (5), a and b represent empirical constants, W and d represent the width of the target area and the distance between the starting position and the target position. The equation describing the calculation of the time law of multiple-choice decision is shown in Equation (6) [23].

$$T_2 = a + b\log_2(n+1) \quad (6)$$

In Equation (6), n represents the number of choices. By using Formula (5) and Formula (6), the study extracted the core rules of distance and quantity affecting interaction efficiency. The human-computer interaction rules of Fitz Law were transformed into design constraints and mechanism inspiration for the model, which was used to guide the design of GAT dynamic interaction module and make GAT dynamic interaction module closer to the human behavior characteristics in real interpretation scenarios. Independent dynamic interaction adjusts translation strategies in real time based on context and environment. However, its dynamic design may hide some content, preventing users from directly accessing needed information and reducing interaction intuitiveness. GAT can automatically learn the relationships between nodes through the attention mechanism, dynamically adjusting the interaction strategy to improve flexibility and adaptability. Therefore, this study employs GAT to optimize dynamic interaction design, enabling efficient and flexible capture of interaction relationships. The independent dynamic interaction method based on GAT is shown in Figure 4.

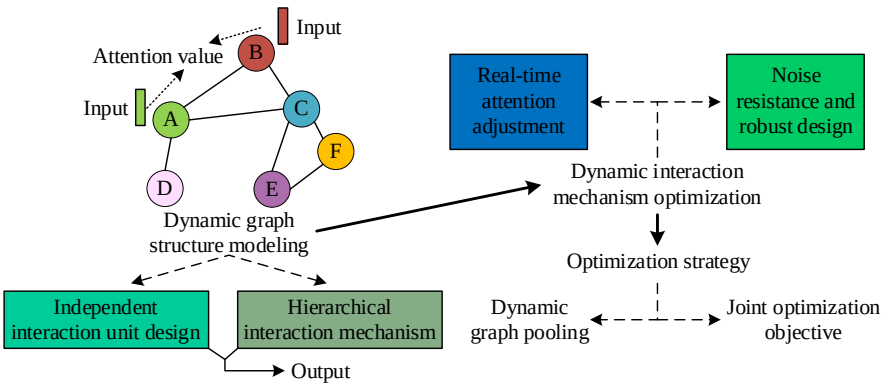


Figure 4: Schematic diagram of independent dynamic interaction method based on GAT

As shown in Figure 4, in the independent dynamic interaction process based on GAT, independent GAT modules are designed for each user and device interaction node to support real-time updates of local dynamic relationships. Then, the multi-head attention mechanism calculates the interaction weight between nodes by combining timestamp information to capture the temporal variation of interaction relationships. The hierarchical mechanism includes intra-session and inter-session interactions. Intra-session interaction mainly uses a position-aware attention network to capture local dynamic sequences, while inter-session interaction requires constructing a global hypergraph encoder to establish long-term dependencies through

higher-order connectivity. The linear transformation of node features is calculated as shown in Equation (7).

$$\vec{h}'_i = w\vec{h}_i \quad (7)$$

In Equation (7), w and $\vec{h}_i$ are the learnable weight matrix and the original features of node i. The attention coefficient between nodes is calculated as shown in Equation (8).

$$e_{ij} = Leaky\,\mathrm{Re}\,LU(\vec{a}^T\left[w\vec{h}_i // w\vec{h}_j\right] \quad (8)$$

In Equation (8), $\vec{a}$ and $//$ represent the learnable attention vectors and the feature concatenation operation. The study optimizes the strategy during dynamic interaction by designing dynamic graph pooling and multi-task loss functions, using learnable projection vectors to filter key nodes and generate compact subgraph representations, thus improving the discrimination ability of dynamic interaction mode and the modeling ability of interaction relationships.

2.3 Construction of interpretation model based on improved intent recognition and independent dynamic interaction

Optimizing intent recognition and independent dynamic interaction can improve the accuracy and efficiency of intent recognition and enhance the flexibility and adaptability of independent dynamic interaction in complex dialogues [24]. Therefore, this study builds the interpretation model on improved intent recognition and independent dynamic interaction, dynamically adjusting strategies to better handle multi-task scenarios. The cognitive process of interpretation is shown in Figure 5.

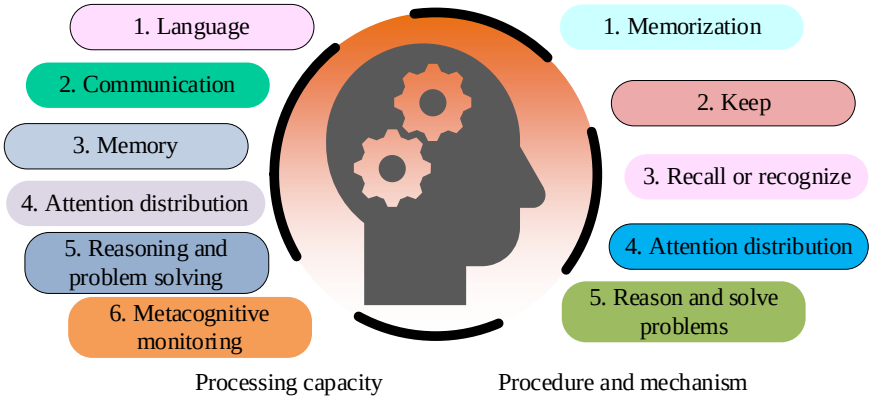


Figure 5: Schematic diagram of the cognitive processing of interpretation

In Figure 5, the cognitive processing of interpretation involves the integration of multiple cognitive mechanisms and skills, mainly including communication, attention, problem-solving, and reasoning ability. The classic formula of human interpretation cognitive process is introduced as shown in Equation (9).

$$OI = L + G + M + O \quad (9)$$

In Equation (9), L and G represent hearing analysis and notes, M and O represent short-term memory and coordination. The calculation for the speech output phase is shown in Equation (10).

$$OI_2 = R + GR + P \quad (10)$$

In Equation (10), R, GR, and P represent memory, note reading, and production. Equations (9) and (10) reveal cognitive patterns in human interpreting, such as listening analysis, memory coordination, and speech production. These provide valuable references for simulating real interpreting processes in models. This study designs the computational components based on these principles, constructing multi-task encoders and hierarchical decoders within the model to enable artificial intelligence systems to implicitly replicate the cognitive processes of human interpreting in their computational logic. During interpretation, the interpreter listens to, comprehends, and stores the original language information in short-term memory. When needed in the scene, the interpreter extracts information from short-term memory and expresses it in the target language. The interpreter also needs to quickly allocate attention, perform comprehensive analysis, reasoning, and problem-solving to ensure the accuracy and logic of the translation. The application foundation of intent recognition and independent dynamic interaction in interpretation is shown in Figure 6.

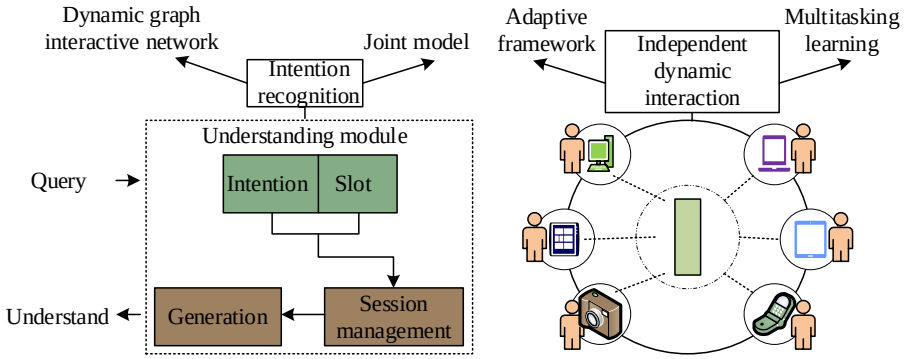


Figure 6: Application principles of intent recognition and independent dynamic interaction

Figure 6 shows that intent recognition in interpretation primarily relies on constructing a dynamic graph interaction network and a joint model. Constructing a dynamic graph model to capture the relationship between intent and context improves the accuracy of intent recognition, and combining intent recognition with slot filling tasks enables more comprehensive semantic understanding. In interpretation, independent dynamic interaction adjusts translation strategies using an adaptive framework. The combination of intent recognition, slot filling, and dynamic interaction further enhances the robustness and timeliness of interpretation. The graph structure of the time step is calculated as shown in Equation (11).

$$g_T = f(g_{T-1}, x_T) \quad (11)$$

In Equation (11), f and x_T represent the dynamic update function and the input features of time step T, and g_{T-1} is the graph structure of the previous time step. The interpretation model based on improved intent recognition and independent dynamic interaction is shown in Figure 7.

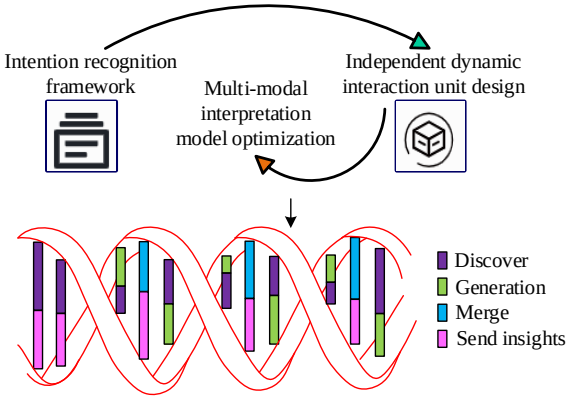


Figure 7: Framework diagram of the proposed interpretation model

In Figure 7, the fused interpretation model first uses GAT to construct an intent relationship graph, leveraging the multi-head attention mechanism to capture the dynamic associations between multiple intents. IGWO is then introduced to dynamically adjust the intent boundary thresholds and attention head weights. IGWO's global search capability improves intent confidence calculation, enabling more accurate and interpretable intent recognition. The local graph update strategy of GAT, combined with the encircling mechanism of IGWO, enables the interactive unit to respond to sudden context changes within milliseconds. The fitness value calculation for the grey wolf's position is shown in Equation (12).

$$f(\vec{X}) = Accuracy(Intent\,\mathrm{Re}cognition(\vec{X})) \quad (12)$$

In Equation (12), $Accuracy$ represents the accuracy of the intent recognition model. The calculation of attention weights is shown in Equation (13).

$$\alpha_{ij} = \frac{\exp(Leaky\,\mathrm{Re}\,LU(a'\left[wh_i // wh_j\right]))}{\sum_{k\in N_i}\exp(Leaky\,\mathrm{Re}\,LU(a^T\left[wh_i // wh_k\right]))} \quad (13)$$

In Equation (13), a represents the parameter vector of the attention mechanism. The calculation for node feature updates is shown in Equation (14).

$$h_i' = \sigma(\sum_{j\in N_i}\alpha_{ij} wh_j) \quad (14)$$

In Equation (14), σ represents the activation function. The combined calculation of IGWO and GAT is shown in Equation (15).

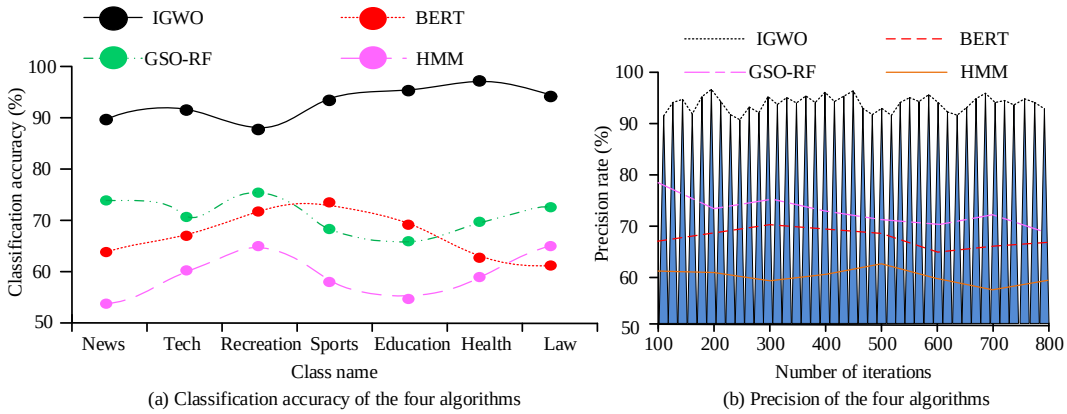
$$Interpretation = Decoder(AttentionFusion(IGWO_{feat}(Intent\,\mathrm{Re}c(S)), GAT_{feat}(DynamicInter(C))) \quad (15)$$

In equation (15), $IGWO_{feat}(Intent\,\mathrm{Re}c(S))$ represents the extraction of intent feature vectors from input S after the IGWO optimization intention recognition model. $GAT_{feat}(DynamicInter(C))$ denotes the GAT extracting semantic interaction feature vectors from dynamic interaction content C. $AttentionFusion(\cdot,\cdot)$ is the attention mechanism that integrates the intent and semantic features, enabling the model to focus on key interaction content driven by intent. $Decoder(\cdot)$ represents the final interpretation output generated by the decoder after the fusion of these features. In the interpretation model, a temporal graph attention mechanism is introduced. Through GAT and IGWO, long-range contextual dependencies are captured, and the temporal window length and attention decay coefficient are adjusted, balancing translation accuracy and real-time performance. The intent feature vector is output by the IGWO intent recognition module and, after being adapted to the dimensions by the projection layer, is fed into the GAT interaction module. The real-time update cycle is set to 50ms. Within each cycle, IGWO outputs new intent features after 10 iterations and triggers the GAT to synchronize updates. When IGWO outputs updates, GAT immediately responds and recalculates the interaction weights, ensuring real-time coordination between the intent and interaction modules. In terms of graph construction, nodes are represented by interpreting sentences and semantic units, with node features integrating speech text encoding and the intent vectors identified by IGWO. Next, edge weights are calculated based on semantic relevance and the degree of intent alignment, with stronger associations receiving higher weights. As the interpretation progresses and new speech texts are input, the graph dynamically updates, triggering changes in node additions and deletions and edge weight updates. The Graph Attention Network (GAT) is used to aggregate new features in real-time, accurately capturing changes in the flow of semantics and intentions. IGWO and GAT focus on intent recognition and dynamic interaction, respectively. IGWO is designed to uncover deep-level intentions and generate intent feature vectors, while GAT uses statements and semantic blocks as nodes. The node features are integrated into the intent vectors of IGWO, and edge weights are constructed based on semantic relevance and intent alignment. During operation, each IGWO recognition result triggers an update of the GAT graph structure, transforming the two processes from independent functions into a deeply integrated, collaborative workflow.

3 Results

3.1 Performance comparison of IGWO intent recognition algorithm

To evaluate the performance of the IGWO-based intent recognition algorithm, this study compares it with three other algorithms. These algorithms include the Glowworm Swarm Optimization combined with Random Forest (GSO-RF) intent recognition algorithm, the BERT-based intent recognition algorithm, and the Hidden Markov Model (HMM)-based intent recognition algorithm. The experimental datasets used were the MixSNIPS and MixATIS datasets. Both are public data sets. The MixSNIPS dataset serves as the training set and contains speech records from multiple domains, each labeled with multiple intents and associated slot data. This dataset was mainly used for intent recognition and slot filling tasks. The MixATIS dataset served as the test set, containing dialogue data from various domains, each with multiple intents and corresponding slot information, simulating real-world complex language expressions by users. The operating system used was Ubuntu 20.04.2 LTS, with PyTorch as the deep learning development framework. The GPU and CPU were NVIDIA GeForce RTX 2080Ti and Intel Core i7-9700F, respectively. The Adam optimizer was used, with beta 1 set to 0.5 and beta 2 set to 0.999. The IGWO population size was set to 50, iterations to 200, while other parameters were fixed using the control variable method and adjusted individually. The study first compared the classification accuracy and precision of the four algorithms on the MixSNIPS dataset, with results shown in Figure 8.



(a) Classification accuracy of the four algorithms (b) Precision of the four algorithms

Figure 8: Classification accuracy and precision results

Figure 8(a) shows that IGWO achieved significantly higher classification accuracy than the other three algorithms. The classification accuracy of the comparison algorithms was all below 80%. The classification accuracy of IGWO for news-related intents was 89.98%, while the classification accuracy for GSO-RF and BERT were 74.05% and 63.42%, respectively. HMM had the lowest accuracy, at only 53.61%. In Figure 8(b), IGWO's precision stabilized between 90% and 99% as the number of iterations increased. Among the comparison algorithms, GSO-RF achieved the highest precision of only 78.86%. When the number of iterations reached 500, the precision rates of BERT and HMM were 69.86% and 63.79%, respectively, while IGWO's precision remained higher than the comparison algorithms at 92.91%. These results demonstrate that IGWO achieved higher classification accuracy and precision. By improving the traditional grey wolf optimization algorithm, it searches more efficiently for optimal solutions, showing strong capability in feature extraction and key information selection. The study then analyzed the recall rates of the four algorithms on different datasets, with results shown in Table 2.

Table 2: Recall rate experimental results in different datasets

Dataset	Class name	Algorithm			
		IGWO	GSO-RF	BERT	HMM
MixSNIPS dataset	News	92.35%*	80.94%	80.09%	72.53%
	Tech	90.34%*	81.63%	76.54%	69.43%
	Entertainment	91.76%*	85.47%	74.33%	71.81%
	Sports	93.71%*	86.54%	72.91%	72.35%
	Education	92.33%*	88.14%	78.53%	68.18%
	Health	94.08%*	83.46%	73.16%	67.69%
	Law	91.35%*	83.06%	76.37%	73.41%
MixATIS dataset	News	95.03%*	84.11%	79.29%	69.34%
	Tech	91.36%*	80.31%	77.68%	66.29%
	Entertainment	94.14%*	83.21%	74.31%	68.79%
	Sports	93.54%*	79.61%	76.81%	64.32%
	Education	92.01%*	77.39%	80.34%	68.81%
	Health	93.33%*	81.07%	73.64%	67.94%
	Law	93.46%*	82.38%	72.19%	63.22%

Note: "*" indicates statistical significance, $p < 0.05$.

As shown in Table 2, the IGWO-based intent recognition algorithm achieved higher recall rates across the different datasets, all reaching over 90%. In the MixSNIPS dataset, when the category was health, IGWO achieved the highest recall rate of 94.08%. Among the comparison algorithms, GSO-RF had the highest recall rate for the education category at 88.14%. In the MixATIS dataset, IGWO achieved the highest recall rate of 95.03%, while the highest recall rate for the other three comparison algorithms was only 84.11%. The recall rates for BERT and HMM were the lowest, at 72.19% and 63.22%, respectively. These results show that IGWO offers better comprehensiveness in intent recognition and can more efficiently search for the optimal solution, ensuring that the algorithm covers all relevant results in intent recognition tasks. The F1 value, translation quality BLEU score, interpretation delay and task success rate of the four algorithms in different data sets were then tested. The results are shown in Table 3.

Table 3: compares the performance indexes of the four algorithms in different data sets

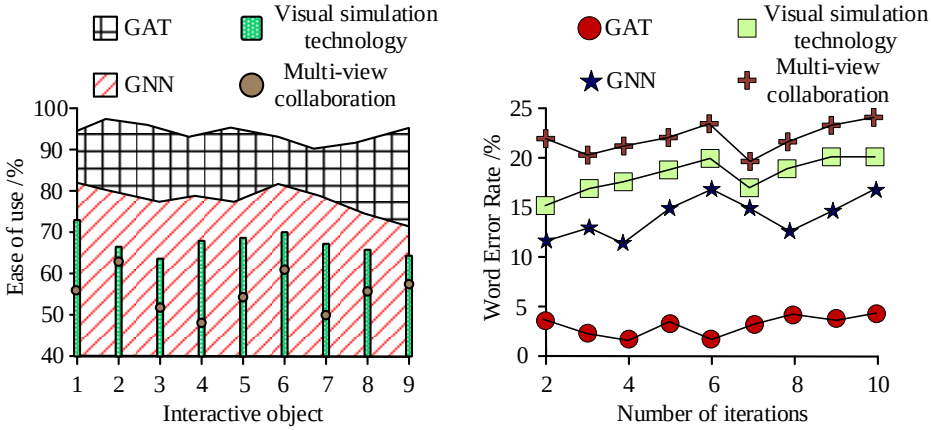
Data set	Index	Algorithm			
		IGWO	GSO-RF	BERT	HMM
MixSNIPS data set	F1/%	92.23	82.58*	85.75*	76.31*
	BLEU score of translation quality	78.51	61.35*	65.23*	54.70*
	Interpretation delayed /ms	113.52	156.71*	138.49*	189.23*
	Task success rate /%	95.34	86.15*	82.43*	79.45*
MixATIS data set	F1/%	91.51	84.26*	87.94*	78.62*
	BLEU score of translation quality	85.29	63.13*	67.51*	56.98*
	Interpretation delayed /ms	105.36	149.27*	129.64*	181.58*
	Task success rate /%	96.22	81.45*	78.21*	76.54*

Note: "*" indicates statistical significance, $p < 0.05$.

In Table 3, the F1 score of IGWO in the MixSNIPS dataset reached 92.23%, significantly outperforming the comparison algorithms. The F1 scores for GSO-RF, BERT, and HMM were 82.58%,85.75%, and 76.31%, respectively. In the MixATIS dataset, IGWO's interpretation delay was only 105.36ms, lower than the comparison algorithms, and its task success rate was 96.22%. In contrast, HMM's task success rate was only 76.54%, the lowest among the four algorithms. The data indicates that this algorithm captured intent more accurately and improved translation quality and efficiency. Moreover, the algorithm demonstrated superior performance in intent recognition F1, BLEU score, interpretation delay, and task success rate, highlighting its robustness and effectiveness across multiple datasets.

3.2 Performance analysis of independent dynamic interaction method based on GAT

This study tests the GAT-based independent dynamic interaction method by comparing it with multi-view collaborative, scene simulation technology, and Graph Neural Network (GNN)-based methods. The GAT consists of 4 layers with 8 attention heads and a dropout rate between 0.1 and 0.3 to prevent overfitting. Unity was used as the virtual reality simulation platform, and the GPS system positioning accuracy of the Kinect camera was set to 8.4m. The camera resolution and depth range were set to 640×480 and 0.5m–4.5m, respectively, while the image refresh rate and texture resolution of the virtual scene were 95Hz and 3840×2160, respectively. The user's field of view angle and the camera frame rate were set to 120° and 30fps. The study first compared the ease of use and word error rates of the four methods, as shown in Figure 9.



(a) Ease of use of the four methods (b) The word error rates of the four methods

Figure 9: Comparison of ease of use and number of dynamic interactions stops

As shown in Figure 9(a), the usability of the GAT used in the study reached more than 90% across different interactive objects, while the usability of multi-view collaboration showed large fluctuations, with the lowest being only 48.37%. The highest and lowest usability of GNN were 81.71% and 74.69%. When there were 9 interactive objects, the usability of GAT reached 96.98%, and the usability of visual simulation technology was 65.07%. As shown in Figure 9(b), the word error rate of GAT is only 4.89%, while the highest word error rates for GNN and scene simulation technology are 17.23% and 20.47%, respectively. The highest word error rate for multi-view collaboration is 23.49%. These data indicate that GAT effectively handles user input and prevents interaction interruptions caused by system delays or errors. Based on the above data results, GAT was more usable and effectively processed user input to avoid interaction interruptions caused by system response delays or errors. To verify the enhancement of multi-modal semantic understanding and language translation quality through the independent dynamic interaction method based on GAT, the study indirectly assesses the model's ability to capture language-action associations by observing the accuracy of thumb positioning. The millimeter-level error in thumb positioning indicates the GAT network's capability to capture local dynamic features. Accurate hand positioning is essentially the external manifestation of the model achieving semantic and action alignment, ultimately serving to optimize language translation quality. Based on this, the study analyzes the thumb positioning effects under semantic association scenarios for four dynamic interaction methods and further correlates these with language translation quality metrics to validate the impact of the interaction mechanism on core outputs. Therefore, a comparative analysis of the thumb positioning effects of four dynamic interaction methods was conducted, and the results are shown in Figure 10.

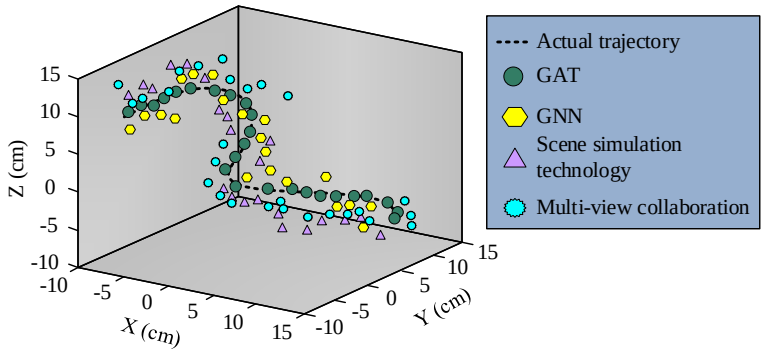
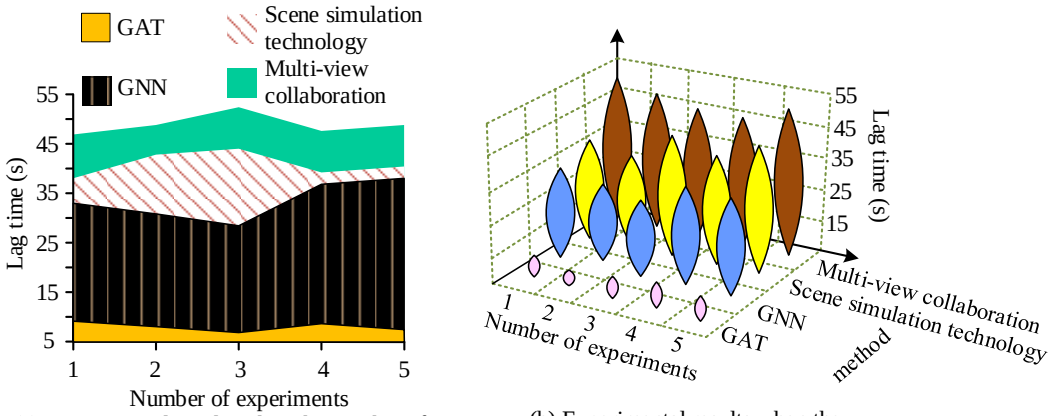


Figure 10: Thumb positioning results of the four methods

As shown in Figure 10, the improved intent recognition module in the interpretation model needs to synchronize with dynamic interactions, and the detection of sudden changes in the thumb trajectory can validate the semantic-action alignment of the two. The GAT-based independent dynamic interaction method outperformed the other three methods in thumb positioning. The three-dimensional coordinates of the user's thumb tracked by GAT consistently align with the actual trajectory. For the comparison methods, the GNN-based method was the furthest from the actual thumb trajectory and had no three-dimensional coordinates overlapping with it. Scene simulation technology, compared to GNN, showed even more noticeable deviations from the actual trajectory. While the multi-view collaborative method had a few three-dimensional coordinates overlapping with the user's actual thumb trajectory, its positioning effect was still inferior to the GAT-based method. Based on these results, it can be concluded that the GAT-based independent dynamic interaction method provides more accurate hand shape positioning, supporting more complex operational tasks. The study then tested the dynamic interaction latency time of the four methods with different interaction objects, with results shown in Figure 11.



(a) Experimental results when the number of interaction objects is 20

(b) Experimental results when the number of interaction objects is 50

Figure 11: Dynamic interaction lag time of different interactive objects

As shown in Figure 11, when the number of interaction objects was 20, GAT had the lowest dynamic interaction latency time, with a value of 7.93s at 3 experimental trials. The dynamic interaction latency times for GNN and scene simulation were 28.34s and 44.73s, respectively, while the multi-view collaborative method had a latency time of 53.69s. When the number of trials was increased to 5, GAT's latency time was 8.32s, while multi-view collaborative had a latency time of 49.71s. The highest latency time for GAT was 13.24s, and the highest for multi-view collaborative was 53.06s. These results show that the GAT-based independent dynamic interaction method achieves the lowest latency across various interaction object counts, demonstrating superior computational efficiency and reasoning ability in complex tasks.

3.3 Performance comparison and analysis of fusion interpretation models

To verify the performance of the fused interpretation model based on intent recognition and independent dynamic interaction, the study selected the following models for comparison: interpretation models based on DNN-HMM, Ensemble empirical mode decomposition (Ensemble Empirical Mode Decomposition, EEMD) and large language model (LLM). The experimental datasets used were the Wmt14 and LibriSpeech datasets. Both are public data sets. The Wmt14 dataset served as the training set, containing parallel corpora for multiple language pairs, such as English-French, English-German, etc. Each sentence pair included source and target language texts. The LibriSpeech dataset served as the test set, containing approximately 1000 hours of 16kHz

English speech corpus. During the preprocessing stage, multi-language texts must undergo unified cleaning to remove special characters and non-linguistic symbols. Cross-language symbols are standardized through predefined dictionaries, including their professional terms and colloquial expressions. In the post-processing of ASR, language models correct word order errors, providing each algorithm with standardized input data to ensure fair performance comparisons. The WMT14 dataset is split into training and validation sets at an 8:2 ratio, while the LibriSpeech dataset is reserved entirely for testing to maintain evaluation integrity. During training, the WMT14 training set is fed into both the fusion interpretation model and the comparison model, with cross-entropy loss function used to optimize parameters. After training, the LibriSpeech test set is input into each model, and multiple rounds of experiments are conducted to calculate the mean and standard deviation, ensuring the reliability of the results. The experimental environment setup is shown in Table 4.

Table 4: Experimental environment and configuration

Hardware configuration	CPU	Intel Xeon Gold 5218 @ 2.30GHz
	GPU	NVIDIA Tesla V100 32GB
	Internal memory	256GB DDR4
	Storage	2TB NVMe SSD
Software environment	Operating system	Ubuntu 20.04 LTS
	Deep learning framework	PyTorch 1.12.0
	Python version	3.8.10
	CUDA version	11.3
	CuDNN version	8.2.1

The study first compared the training and actual loss values of the four models, using the BERT pre-trained language model to encode the input text and generate high-dimensional vector representations. The text sequences were uniformly truncated to a fixed length, and the results are shown in Figure 12.

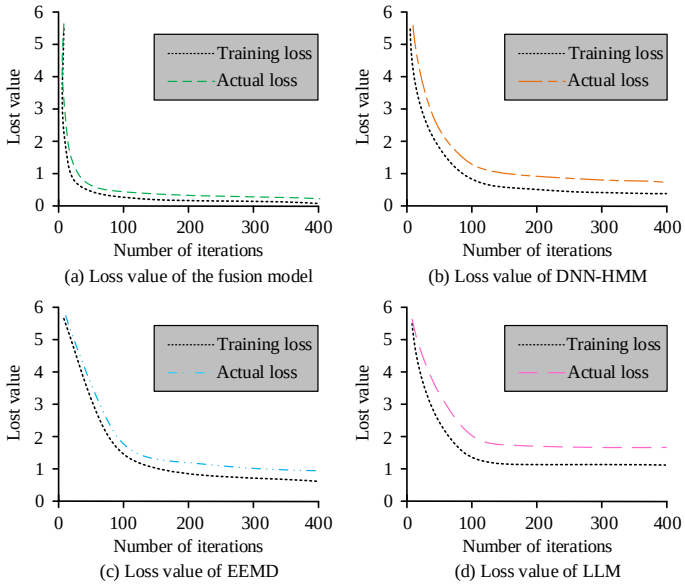


Figure 12: Comparison of loss value results of four models

From Figure 12, it can be seen that the loss values for the study's model were the lowest and most consistent between training and actual losses, with convergence occurring around 30 iterations and stabilizing at approximately 0.35. The DNN-HMM model's actual loss curve was higher than its training loss curve and only converged at 100 iterations. The EEMD model's loss curve converged around 110 iterations, with a stable loss value of around 0.78. The loss curve for the LLM model showed a large discrepancy between the actual and training losses. Based on these results, it can be concluded that the study's model had lower loss values and could better fit the training data, with results more consistent with the true labels. It also exhibited superior generalization and stability compared to the comparison models. The study then compared the interpretation accuracy of the four models on the Wmt14 and LibriSpeech datasets, with results shown in Table 5.

Table 5: Interpretation accuracy results in different datasets

Dataset	Interpretation type	Interpretation accuracy (%)			
		Fusion model	DNN-HMM	EEMD	LLM
Wmt14	Anglo-French	94.37	79.31*	75.09*	64.97*
	Franco-English	96.21	77.59*	71.32*	69.85*
	Anglo-German	90.24	74.69*	70.24*	67.19*
	German-English	91.54	79.18*	72.31*	68.32*
	UK-China	93.36	80.23*	69.58*	69.04*
	China-UK	94.14	80.06*	66.19*	65.42*
LibriSpeech	Anglo-French	96.88	76.43*	69.87*	57.41*
	Franco-English	94.75	79.83*	68.52*	58.39*
	Anglo-German	92.05	76.52*	65.31*	70.35*
	German-English	90.04	77.68*	63.93*	61.23*
	UK-China	93.34	79.53*	69.57*	65.34*
	China-UK	91.72	74.18*	70.25*	58.17*

Note: * indicates that the difference between the proposed model and the comparison model is statistically significant, $p<0.05$.

As shown in Table 5, the fusion model used in the study showed a high interpretation accuracy across different datasets. In the Wmt14 dataset, when the interpretation type was German-English, the interpretation accuracy of the proposed model was the highest at 91.54%, while the interpretation accuracy of the LLM model was 68.32%. In the LibriSpeech dataset, the English-German interpretation accuracy of the proposed model reached 92.05%, while the accuracy of the DNN-HMM, EMMD, and LLM models were 76.52%, 65.31%, and 70.35% respectively, and their interpretation effects were not as good as the proposed fusion model. It could be seen that the proposed model more accurately captured the meaning of the source language and converted it into the target language in the translation task. To ensure the reliability of the results, all models underwent 10 independent experiments for each dataset and interpretation type. The final accuracy was calculated as the average of these 10 experiments, with the standard deviation and 95% confidence interval also calculated. The improved GWO algorithm uses a population size of 30 and a maximum of 100 iterations. The attention heads and hidden layer dimensions were both set to 8. Increasing the GWO population size aims to enhance the fusion model's adaptability in complex contexts. The statistical analysis results for the four models' interpretation accuracy across different datasets are presented in Table 6.

Table 6: Statistical analysis results of the four models on interpretation accuracy in different data sets

Data Set	Interpretation type	Mean ± standard deviation, (confidence interval)			
		proposed model	DNN-HMM	EEMD	LLM
Wmt14 data set	Anglo-French	94.37±0.82(93.91-94.83)	79.31±1.54(78.23-80.39)	75.09±2.11(73.57-76.61)	64.97±2.85(62.73-67.21)
	Franco-English	96.21±0.75(95.78-96.64)	77.59±1.62(76.41-78.77)	71.32±2.25(69.53-73.11)	69.85±2.68(67.72-71.98)
	Anglo-German	90.24±0.91(89.65-90.83)	74.69±1.78(73.42-75.96)	70.24±2.33(68.39-72.09)	67.19±2.92(64.88-69.50)
	German-English	91.54±0.87(91.02-92.06)	79.18±1.59(78.05-80.31)	72.31±2.18(70.65-73.97)	68.32±2.75(66.28-70.36)
	UK-China	93.36±0.84(92.89-93.83)	80.23±1.51(79.17-81.29)	69.58±2.29(67.81-71.35)	69.04±2.81(66.95-71.13)
	China-UK	94.14±0.79(93.68-94.60)	80.06±1.65(78.89-81.23)	66.19±2.42(64.35-68.03)	65.42±2.98(63.07-67.77)
LibriSpeech data set	Anglo-French	96.88±0.72(96.47-97.29)	76.43±1.68(75.19-77.67)	69.87±2.37(68.05-71.69)	57.41±3.12(55.00-59.82)
	Franco-English	94.75±0.81(94.26-95.24)	79.83±1.55(78.70-80.96)	68.52±2.22(66.85-70.19)	58.39±3.05(56.03-60.75)
	Anglo-German	92.05±0.89(91.48-92.62)	76.52±1.72(75.29-77.75)	65.31±2.45(63.43-67.19)	70.35±2.88(68.13-72.57)
	German-English	90.04±0.93(89.45-90.63)	77.68±1.61(76.53-78.83)	63.93±2.38(62.17-65.69)	61.23±2.95(58.97-63.49)
	UK-China	93.34±0.86(92.85-93.83)	79.53±1.53(78.46-80.60)	69.57±2.31(67.82-71.32)	65.34±2.87(63.22-67.46)
	China-UK	91.72±0.83(91.22-92.22)	74.18±1.75(73.01-75.35)	70.25±2.27(68.47-72.03)	58.17±3.15(55.68-60.66)

In Table 6, the proposed model outperforms the other three comparison models in terms of mean and standard deviation across all language pairs and datasets. The mean accuracy for the English-French pair on the Wmt14 dataset is 94.37%, significantly higher than the other models. Meanwhile, the mean accuracies for DNN-HMM and EEMD are 79.31% and 75.09%, respectively, while the mean accuracy for LLM is 64.97%. On the LibriSpeech dataset, the mean accuracy for the German-English pair is 90.04%, which remains significantly higher than the comparison models. These data indicate that the proposed model demonstrates stronger robustness in interpretation accuracy, effectively enhancing translation precision. In order to further verify the performance of the interpretation model in actual application scenarios, the study chose to simulate a large international conference and used simultaneous interpretation equipment, selecting high-quality and low-latency translation headphones, transmitters, and receivers. The data Signal-to-Noise Ratio (SNR) of the four models at different audio lengths was compared by simulating multilingual scenarios, and the results are shown in Figure 13.

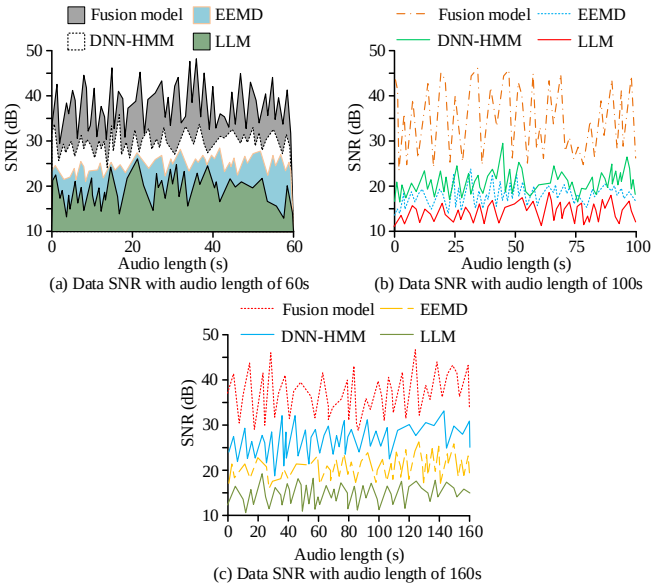


Figure 13: Data SNR results at different audio lengths

From Figure 13, it can be seen that when the audio length was 60s, the highest and lowest SNRs of the proposed model were 47.35dB and 29.74dB respectively. The highest SNR of the DNN-HMM model was 35.81dB, and the lowest data SNR was 24.06dB. The highest SNR of the proposed model was 46.05dB when the audio length was 100s, while the highest data SNR of the LLM model was 19.25dB. When the audio length was 160s, the SNR of the proposed model among the four models was still higher than that of the other three comparison models, and its highest data SNR reached 47.68dB. It could be seen that the training effect of the proposed model was better. The high SNR data enabled the model to converge faster, reduced its training time and resource consumption, and allowed the proposed model to achieve higher translation quality when dealing with complex sentences, professional terms, and cultural background differences. Finally, the study designed ablation experiments on isolated IGWO and GAT, and the results are shown in Table 7.

Table 7: Ablation results

Type of experiment	Compare models	F1 /%	Interpretation delayed /ms
IGWO vs GWO	IGWO	89.2±1.1	123.4±8.3
	GWO	82.5±1.3	156.7±9.2
GAT vs GNN	GAT	91.3±1.2	115.5±7.7
	GCN	85.6±1.4	148.9±8.6
There is cross attention vs no cross attention	Complete pattern	89.2±1.1	123.4±8.3
	Remove cross attention	85.1±1.2	145.6±9.0

In Table 7, in the IGWO vs GWO experiment, the F1 score and interpretation delay of IGWO were both higher than those of GWO, indicating that the improved IGWO algorithm offers better accuracy and efficiency in intent recognition. In the GAT vs GNN experiment, GAT achieved an F1 score of 91.3±1.2% with an interpretation delay of 115.5±7.7ms, demonstrating superior model performance compared to GNN. In the cross-attention vs non-cross-attention experiment, the full model showed a higher F1 score and lower interpretation delay, suggesting that the cross-attention mechanism enhances module collaboration and optimizes the model's

intent-semantic interaction.

4 Discussion

The IGWO intent recognition algorithm demonstrated higher accuracy and precision on the MixSNIPS dataset compared to other models. IGWO's classification accuracy significantly exceeded that of the other three algorithms, all of which scored below 80%. This was mainly because IGWO dynamically adjusted the search weight according to the fitness of the population by introducing a nonlinear weight allocation strategy, thus avoiding the problem that GWO was prone to fall into the local optimum in intent recognition. This mechanism significantly improved the global search ability of the model in complex intent classification tasks, thereby enhancing its classification accuracy. This result was similar to the research results of I. Kim's group [25]. The IGWO algorithm achieved higher recall rates across datasets by employing diversity maintenance strategies, such as random reverse learning, which prevent premature convergence and enable more thorough exploration of the intent feature space. This coincided with the research results of G. Leus et al. [26]. Usability tests show that the proposed GAT-based independent dynamic interaction method improves user experience. This is mainly due to GAT's introduction of an attention mechanism, which dynamically assigns weights to interaction nodes, allowing for a more precise capture of user intent and contextual information. This mechanism significantly reduces interaction interruptions caused by incomplete or misjudged information. Previous studies by Z. He et al. showed that combining operation fusion, multi-level pipelines, and graph partitioning optimization significantly enhances GAT performance [27]. In the comparison of thumb positioning effectiveness among four dynamic interaction methods, the GAT-based independent dynamic interaction method outperformed the other three methods. This is mainly due to GAT's real-time feedback mechanism, which dynamically adjusts interaction strategies based on user behavior. This result is consistent with the findings of A. Coscia et al. [28]. When examining the dynamic interaction latency across different interaction objects, the proposed method showed lower latency. By integrating multimodal data such as visual and tactile information, the GAT method was able to more comprehensively analyze the behavior features of the interaction objects, reducing the latency caused by the insufficiency of a single data source. Meanwhile, Y. Li et al. found in 2023 that the GAT structure design is interpretable, with simulations showing its near-optimal performance and real-time computation support [29]. Based on training and validation loss values, the proposed model demonstrates strong data fitting ability, primarily driven by IGWO's efficient intent recognition and effective resource optimization. This coincided with the results obtained by K. Jang et al. in 2025 [30]. Compared with the research conducted by J. Torres Gómez's team in 2023 [31], the interpretation accuracy of the four models in different data sets showed the highest interpretation accuracy. This improvement results from integrating multimodal data like speech and text, enabling more comprehensive analysis of user intent and context while reducing translation bias from limited data sources. Through the data signal-to-noise ratio results of the four models at different audio lengths, it was found that the training effect of the proposed model was better and had higher signal-to-noise ratio data. This was mainly because the model dynamically adjusted the interaction strategy according to user behavior combined with the real-time feedback mechanism, and further optimized the anti-noise effect. This result was similar to the research conducted by M. Sajid et al. in 2024 [32]. Traditional GWO tends to fall into local optima in complex multilingual environments, while standard GAT struggles with feature differentiation due to high node similarity when processing long multilingual sequences. However, the proposed model demonstrates a higher SNR across various audio lengths compared to the other three models, with its highest SNR reaching 47.68dB. Moreover, the proposed model achieves higher translation quality when handling multilingual differences, accurately capturing and converting the meaning of the source language into the target language in translation tasks. This research considers the potential of multimodal interpretation and plans to incorporate gesture and visual information integration in future work. In the future, it will incorporate gestures and visual information integration. It plans to collect video data of interpretation with gestures and design a multimodal feature extraction module to explore the integration mechanism of visual, audio, and text features. By optimizing dynamic interaction strategies, it aims to adapt to more realistic interpretation scenarios.

In summary, the interpretation model based on IGWO intention recognition and GAT independent dynamic interaction made significant contributions to improving the accuracy of intention recognition, enhancing the ability to capture contextual information, achieving real-time feedback and adaptive adjustment, and promoting the development of interpretation technology. These contributions provided a new theoretical framework and practical methods for interpretation technology, promoted the application and development of artificial intelligence in the field of interpretation, and offered an important reference for future interpretation research and practice.

5 Conclusion

The existing interpretation models suffer from low translation accuracy and weak sensitivity to data bias. To address these limitations, this study proposed a new interpretation model combining the IGWO-based intent recognition algorithm and the GAT-based independent dynamic interaction. By constructing a dynamic graph model to capture the relationship between intent and context, the model enhances intent recognition accuracy and enables a more comprehensive semantic understanding. The results showed that IGWO's classification accuracy was significantly higher than that

of the other three intent recognition algorithms, with HMM having the lowest classification accuracy. The proposed model demonstrated good performance in terms of data SNR, loss values, and dynamic interaction stop counts across different audio lengths. Overall, the proposed integrated model exhibited excellent classification accuracy and translation performance. Its capabilities in speech recognition, language conversion, and contextual understanding enable it to efficiently and accurately complete translation tasks. Although the current model shows excellent classification and translation accuracy, future work will focus on reducing model complexity and enhancing robustness in extreme scenarios to further improve performance.

Fundings

The research is supported by: "14th Five-Year Plan" Teaching Reform Project for Regular Undergraduate Universities in Zhejiang Province, Research on the Digitalized Training Model and it's Effectiveness for English Interpretation Talents, (NO. jg20220589).

References

[1] S. Ni, L. Zhao, A. Li, D. Wu and L. Zhou. "Cross-View Human Intention Recognition for Human-Robot Collaboration," *IEEE WIREL COMMUN*, vol. 30, no. 3, pp. 189-195, June, 2023, DOI: 10.1109/MWC.018.2200514.

[2] Y. Bai, X. Lu and B. Xu. "A Probabilistic Fuzzy Classifier for Motion Intent Recognition," *IEEE T FUZZY SYST*, vol. 32, no. 3, pp. 1098-1107, March, 2024, DOI: 10.1109/TFUZZ.2023.3317938.

[3] G. Jiang, K. Wang, Q. He and P. Xie. "E2FNet: An EEG- and EMG-Based Fusion Network for Hand Motion Intention Recognition," *IEEE SENS J*, vol. 24, no. 22, pp. 38417-38428, Nov, 2024, DOI: 10.1109/JSEN.2024.3471894.

[4] K. Cheng, D. Sun, J. Jian, D. Qin, C. Chen and G. Liao. "Deep Learning Approach for Driver Speed Intention Recognition Based on Naturalistic Driving Data," *IEEE T INTELL TRANSP*, vol. 25, no. 10, pp. 14546-14559, Oct, 2024, DOI: 10.1109/TITS.2024.3398083.

[5] Chauhan S, Saxena S and Daniel P. "Analysis of Neural Machine Translation KANGRI Language by Unsupervised and Semi Supervised Methods," *IETE J RES*, vol. 69, no. 10, pp. 6867-6877, Apr, 2023, DOI: 10.1080/03772063.2021.2016506.

[6] Z. Zuo, P. Wu, X. Sun, X. Tong, R. Guo and H. Huang. "Interactive Feature Extraction Network for Target Intention Recognition," *IEEE SENS J*, vol. 25, no. 3, pp. 4850-4868, Feb, 2025, DOI: 10.1109/JSEN.2024.3488695.

[7] B. Wang. "A Hybrid Fuzzy Logic and Deep Learning Model for Corpus-Based German Language Learning with NLP", *Inform.*, vol. 49, no.21, PP. 1–14, 2025, DOI: 10.31449/inf.v49i21.7423

[8] L. L. Li, G. Cao, H. Liang, Y. Zhang and F. Cui. "Human Lower Limb Motion Intention Recognition for Exoskeletons: A Review," IEEE SENS J, vol. 23, no. 24, pp. 30007-30036, Dec, 2023, DOI: 10.1109/JSEN.2023.3328615.

[9] C. Wong, L. Vergez and W. Suleiman. "Vision- and Tactile-Based Continuous Multimodal Intention and Attention Recognition for Safer Physical Human–Robot Interaction," *IEEE T AUTOM SCI ENG*, vol. 21, no. 3, pp. 3205-3215, July, 2024, DOI: 10.1109/TASE.2023.3276856.

[10] F. Dinarta, A. "Wicaksana. Enhanced Hate Speech Detection in Indonesian-English Code-Mixed Texts Using XLM-RoBERTa", *Inform.*, vol. 49, no.21, PP. 45–56, 2025, DOI: 10.31449/inf.v49i21.7713

[11] P. Sun, H. Xu, J. Yao, Y. Chi, S. Huang and J. Cao. "Dynamic Interaction Analysis and Damping Control Strategy of Hybrid System with Grid-Forming and Grid-Following Control Modes," *IEEE T ENERGY CONVER*, vol. 38, no. 3, pp. 1639-1649, Sept, 2023, DOI: 10.1109/TEC.2023.3249965.

[12] Y. Xiao, L. Wu, J. Guo, J. Li, M. Zhang, T Qin. "A Survey on Non-Autoregressive Generation for Neural Machine Translation and Beyond," *IEEE T PATTERN ANAL*, vol. 45, no. 10, pp. 11407-11427, Oct, 2023, DOI: 10.1109/TPAMI.2023.3277122.

[13] Xiang Y, Chen Y, Ye F H. "Enhancing computer-aided translation system with BiLSTM and convolutional neural network using a knowledge graph approach," *J SUPERCOMPUT*, vol. 80, no. 5, pp. 5847-5869, Oct, 2024, DOI: 10.1007/s11227-023-05686-2.

[14] E. Wang, X. Chen, Y. Li, Z. Fu and J. Huang. "Lower Limb Motion Intent Recognition Based on Sensor Fusion and Fuzzy Multitask Learning," *IEEE T FUZZY SYST*, vol. 32, no. 5, pp. 2903-2914, May, 2024, DOI: 10.1109/TFUZZ.2024.3364382.

[15] H. Zhang, D. Guo, Y. Guo, F. Wu and S. Gao. "A Novel Method for the Driver Lane-Changing Intention Recognition," *IEEE SENS J*, vol. 23, no. 17, pp. 20437-20451, Sept, 2023, DOI: 10.1109/JSEN.2023.3299253.

[16] Y. Zhang, W. Ma, F. Huang, X. Deng and W. Jiang. "A Novel Air Target Intention Recognition Method Based on Sample Reweighting and Attention-Bi-GRU," *IEEE SYST J*, vol. 18, no. 1, pp. 501-504, March, 2024, DOI: 10.1109/JSYST.2023.3319643.

[17] M. Yi, W. Lee and S. Hwang. "A Human Activity Recognition Method Based on Lightweight Feature Extraction Combined with Pruned and Quantized CNN for Wearable Device," *IEEE T CONSUM ELECTR*, vol. 69, no. 3, pp. 657-670, Aug, 2023, DOI: 10.1109/TCE.2023.3266506.

[18] D. Roldán-Álvarez and F. Mesa. "Intelligent Deep-Learning Tutoring System to Assist Instructors in Programming Courses," *IEEE T EDUC*, vol. 67, no. 1, pp. 153-161, Feb, 2024, DOI: 10.1109/TE.2023.3331055.

[19] X. Fu, M. Huang, C. K. Tse, J. Yang, Y. Ling and X.

Zha. "Synchronization Stability of Grid-Following VSC Considering Interactions of Inner Current Loop and Parallel-Connected Converters," *IEEE T SMART GRID*, vol. 14, no. 6, pp. 4230-4241, Nov, 2023, DOI: 10.1109/TSG.2023.3262756.

[20] J. Liang, Y. Lu, F. Wang, G. Yin, X. Zhu and Y. Li. "A Robust Dynamic Game-Based Control Framework for Integrated Torque Vectoring and Active Front-Wheel Steering System," *IEEE T INTELL TRANSP*, vol. 24, no. 7, pp. 7328-7341, July, 2023, DOI: 10.1109/TITS.2023.3262655.

[21] E. Ghiorzi, M. Colledanchise, G. Piquet, S. Bernagozzi, A. Tacchella and L. Natale. "Learning Linear Temporal Properties for Autonomous Robotic Systems," *IEEE ROBOT AUTOM LET*, vol. 8, no. 5, pp. 2930-2937, May, 2023, DOI: 10.1109/LRA.2023.3263368.

[22] I. Stepin, M. Suffian, A. Catala and J. M. Alonso-Moral. "How to Build Self-Explaining Fuzzy Systems: From Interpretability to Explainability [AI-eXplained]," *IEEE COMPUT INTELL M*, vol. 19, no. 1, pp. 81-82, Feb, 2024, DOI: 10.1109/MCI.2023.3328098.

[23] Y. Zhang, Y. Zou, Selpi, Y. Zhang and L. Wu. "Spatiotemporal Interaction Pattern Recognition and Risk Evolution Analysis During Lane Changes," *IEEE T INTELL TRANSP*, vol. 24, no. 6, pp. 6663-6673, June, 2023, DOI: 10.1109/TITS.2022.3233809.

[24] S. Saeidian, G. Cervia, T. J. Oechtering and M. Skoglund. "Pointwise Maximal Leakage," *IEEE T INFORM THEORY*, vol. 69, no. 12, pp. 8054-8080, Dec, 2023, DOI: 10.1109/TIT.2023.3304378.

[25] I. Kim, J. Na, J. P. Yun and S. Lee. "Deep Feature Selection Framework for Quality Prediction in Injection Molding Process," *IEEE T IND INFORM*, vol. 20, no. 1, pp. 503-512, Jan, 2024, DOI: 10.1109/TII.2023.3268421.

[26] G. Leus, A. G. Marques, J. M. F. Moura, A. Ortega and D. I. Shuman. "Graph Signal Processing: History, development, impact, and outlook," *IEEE SIGNAL PROC MAG*, vol. 40, no. 4, pp. 49-60, June, 2023, DOI: 10.1109/MSP.2023.3262906.

[27] Z. He, T. Tian, Q. Wu and X. Jin. "FTW-GAT: An FPGA-Based Accelerator for Graph Attention Networks with Ternary Weights," *IEEE T CIRCUITS-II*, vol. 70, no. 11, pp. 4211-4215, Nov, 2023, DOI: 10.1109/TCSII.2023.3280180.

[28] A. Coscia and A. Endert. "KnowledgeVIS: Interpreting Language Models by Comparing Fill-in-the-Blank Prompts," *IEEE T VIS COMPUT GR*, vol. 30, no. 9, pp. 6520-6532, Sept, 2024, DOI: 10.1109/TVCG.2023.3346713.

[29] Y. Li, Y. Lu, R. Zhang, B. Ai and Z. Zhong. "Deep Learning for Energy Efficient Beamforming in MU-MISO Networks: A GAT-Based Approach," *IEEE WIREL COMMUN LE*, vol. 12, no. 7, pp. 1264-1268, July, 2023, DOI: 10.1109/LWC.2023.3270361.

[30] K. Jang, K. E. S. Pilario, N. Lee, I. Moon and J. Na. "Explainable Artificial Intelligence for Fault Diagnosis of Industrial Processes," *IEEE T IND INFORM*, vol. 21, no. 1, pp. 4-11, Jan, 2025, DOI: 10.1109/TII.2023.3240601.

[31] Liu X, Chen J, Zhang Q T. "Exploration of low-resource language-oriented machine translation system of genetic algorithm-optimized hyper-task network under cloud platform technology," *J SUPERCOMPUT*, vol. 80, no. 3, pp. 3310-3333, Sep, 2024, doi: 10.1007/s11227-023-05604-6.

[32] M. Sajid, A. K. Malik, M. Tanveer and P. N. Suganthan. "Neuro-Fuzzy Random Vector Functional Link Neural Network for Classification and Regression Problems," *IEEE T FUZZY SYST*, vol. 32, no. 5, pp. 2738-2749, May, 2024, DOI: 10.1109/TFUZZ.2024.3359652.