Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105 (Online)
Volume **10**, Number **3**, August **2025**

Machine Translation Model Optimization and Performance Analysis from the Lightweight Deep Neural Network under the Internet of Things

Yan-Ling Hong, Han-Hui Li*

School of Foreign Languages
Fuzhou University of International Studies and Trade, Fuzhou 350202, China
hongyanling@fzfu.edu.cn, lihanhui@fzfu.edu.cn

Cavin Pamintuan

Angeles University Foundation, Angeles City, Philippines
pamintuan.cavin@auf.edu.ph

*Corresponding author: Han-Hui Li

Received June 7, 2024, revised November 7, 2024, accepted March 22, 2025.

Abstract. *With the rapid progress of the economy and society, language exchange has become an essential factor in integrating world cultures and promoting social development. Therefore, Machine Translation (MT) performance is optimized across the board to improve the accuracy and speed of MT models. This paper designs and optimizes MT models based on the Deep Learning Neural Network (DLNN) to improve the translation effect between various languages. Based on this, the current status and characteristics of MT are first summarized. Then, DLNNs are used to design MT algorithms and models. Finally, an Attention Mechanism (AM) is added to the DLNN MT model to optimize the codec program of the model. According to the simulation result, the statistics-based MT model is superior to the traditional MT model. The accuracy rate is generally between 67% and 72%. DLNN effectively improves the accuracy and speed of statistical-based MT models. The accuracy of DLNN machine translation models is generally between 85-88%. In addition, the introduction of AM in the MT model can improve the encoding and decoding rate of the MT model and promote the translation rate and accuracy. This paper provides technical support for optimizing MT's effect and contributes to social culture's integration and development.*
Keywords: Machine translation, Deep learning, Neural networks, Program optimization

1. **Introduction.** Nowadays, deep learning (DL) technology has become a commonly used network technology in society. Because the primary working mechanism of DL is based on biological understanding, it can meet the design needs of human beings to a great extent. Language translation is the leading way of human communication worldwide, and accurate and efficient translation methods have become the universal pursuit of language translation by human beings [1]. Although language translation technology is not mature, many studies have provided a technical basis.

Mahata et al. [2] believed that Machine Translation (MT) was the most active and promising trend in DL technology. They found that MT technology had been continuously updated over the past 60 years, from methods originally based entirely on rules compiled by humans to statistical methods and now Neural Machine Translation (NMT). Incredibly,

in 2012, when DL technology entered people's field of vision, the accuracy of MT continued to refresh. Marco et al. [3] pointed out that Statistical Machine Translation (SMT) and NMT were popular translation modes. They are usually trained on bilingual translation corpora to learn translation rules to generate target translations. They believed that in this mechanism, many factors affected the translation effect, such as the field distribution of training data, the size of sentence pairs, and the quality of translation. These factors are critical. Generally, the closer the training and test data fields, the more pairs of sentences. It helps to learn accurate translation rules and get accurate translations.

In practical applications, the training data is often collected from different sources, topics, and styles to pursue the quality and scale of training data. The data does not exactly match the domain of the target text, leading to the "domain adaptation" problem. In particular, the training corpus of NMT is too large. Also, the number of unregistered words increases due to the limitation of vocabulary. Therefore, the adaptive problem in MT has always been an issue the industry is working to solve. Rajan et al. [4] pointed out that SMT often had the problem of diverse sources and inconsistent text domains being translated. They suggested screening the training corpus according to the text to improve the translation quality of texts facing different fields and achieve the purpose of domain adaptation. They also found that SMT's current domain adaptation methods were based on target data, focusing on domain adaptation of training data or translation models using statistical techniques. Besides, there were no clear domain labels. VanRullen and Kanai constructed a machine translation quality evaluation model based on the DL method to accurately evaluate machine-automatic translation quality. They used DL-based MT language information extraction methods. In the unsupervised and supervised learning stages, the noise reduction automatic coding machine was applied to conduct unsupervised learning of bilingual words. Language vector features in translation samples were reconstructed and obtained. They imported MT information into bilingual words to enhance the feature extraction effect. The language vector features of automatic MT were imported into that evaluation model to realize the quality evaluation of automatic MT. The results showed that the proposed model could precisely assess the quality of MT. There was no negative interference with the translation of sentence patterns and the number of sentences on its evaluation quality. The model performed well. Ali and Yousuf [5] applied DL techniques to Sino-Tibetan MT tasks using an encoder-decoder structure. In the coding stage, every word in a Chinese sentence was first mapped into a fixed-length word vector. All the information in the sentence was compressed via an RNN model. The attention model was applied in the decoding process. Hence, the decoder is more focused on the context-dependent words. The translated word with maximum probability was selected each time to form the target sentence.

To sum up, with the deepening of world connectivity, language translation has become a necessary means of cultural exchange. With the rapid development of science and technology, intelligent language translation has become the main direction of language translation. However, not many intelligent technologies can do this task in ML development. DL technology is currently an advanced intelligent technology. Through this technology, the ML model is designed. Then, the model is optimized to achieve intelligent and accurate language translation to a large extent. Hence, this study applied a Deep Learning Neural Network (DLNN) model to optimize the MT model to realize intelligent MT in various languages.

The above-mentioned scholars have studied the gradual evolution of machine translation technology from rule methods to neural machine translation (NMT), emphasizing the key role of deep learning technology in improving translation accuracy. Researchers have found that NMT and statistical machine translation (SMT) modes have a significant

impact on translation results in factors such as the domain distribution of training data and the number of sentence pairs. In particular, the problem of domain adaptation is still a challenge. Methods such as SMT have made certain progress in the field of translation, but the translation model performs poorly in cross-domain texts, has low accuracy, and the translation process is complex and slow. Unregistered word processing ability is limited, and the effect is not ideal under specific tasks. It shows that the existing machine translation technology needs to be further optimized in terms of translation accuracy and translation speed to improve its domain adaptability.

Based on the existing research, this paper first analyzes the characteristics of machine translation, then explores deep learning technology, designs a neural machine translation model, and finally introduces the attention mechanism (AM) in the neural network machine translation model, which provides technical support for the performance improvement of the machine translation model and also helps to optimize the language translation model. The motivation of this study is based on the shortcomings of existing machine translation technology in accuracy and domain adaptability, especially when dealing with unknown words and cross-domain text translation. With the deepening of globalization, accurate and efficient language translation has become an urgent need. To solve this problem, this paper introduces the deep learning neural network (DLNN) model and attention mechanism to optimize the existing machine translation system, improve its translation accuracy, adaptability, and processing efficiency, and promote the development of intelligent language translation technology.

2. **Research Foundation.**

2.1. **Overview of MT and its applications.**

MT refers to converting a language into another language without changing its meaning through computer technology. Since its origin in 1949, MT has obtained many ideal research results through much research. Besides, a relatively complete framework has been established through the mutual conversion between various languages. Currently, MT systems are classified into rule-based, instance-based, and statistical-based translation methods based on the differences in translation methods [6]. Among them, the rule-based translation method is the earliest language translation method. Its main principle is to design the source language as code and then translate it into the target language by decoding based on the artificially made language dictionary and the summary results of language experts [7]. The rule-based language translation system relies on artificial language dictionaries. It does not need training during the translation process, so the system performance requirements of the computer are relatively low. Therefore, the rule-based translation method must establish basic rules using many manual statistical language pair translation files. However, there will always be specific translation barriers between different languages; in other words, sentences composed of words with the same meaning between different languages occasionally have different meanings. Therefore, it is necessary to establish different rules for translation, which will cause increasing errors. When the number of languages in a rule-based translation system increases to a certain extent, it will also lead to confusion in rules, hindering the improvement of the performance of the translation system [8]. Moreover, a growing number of language translation materials appear due to the popularization of language translation. Nevertheless, the traditional rule-based translation system cannot assimilate and use these updated materials. Therefore, updated technology will eventually replace the rule-based translation system [9]. Figure 1 reveals the basic principle of a rule-based translation system.

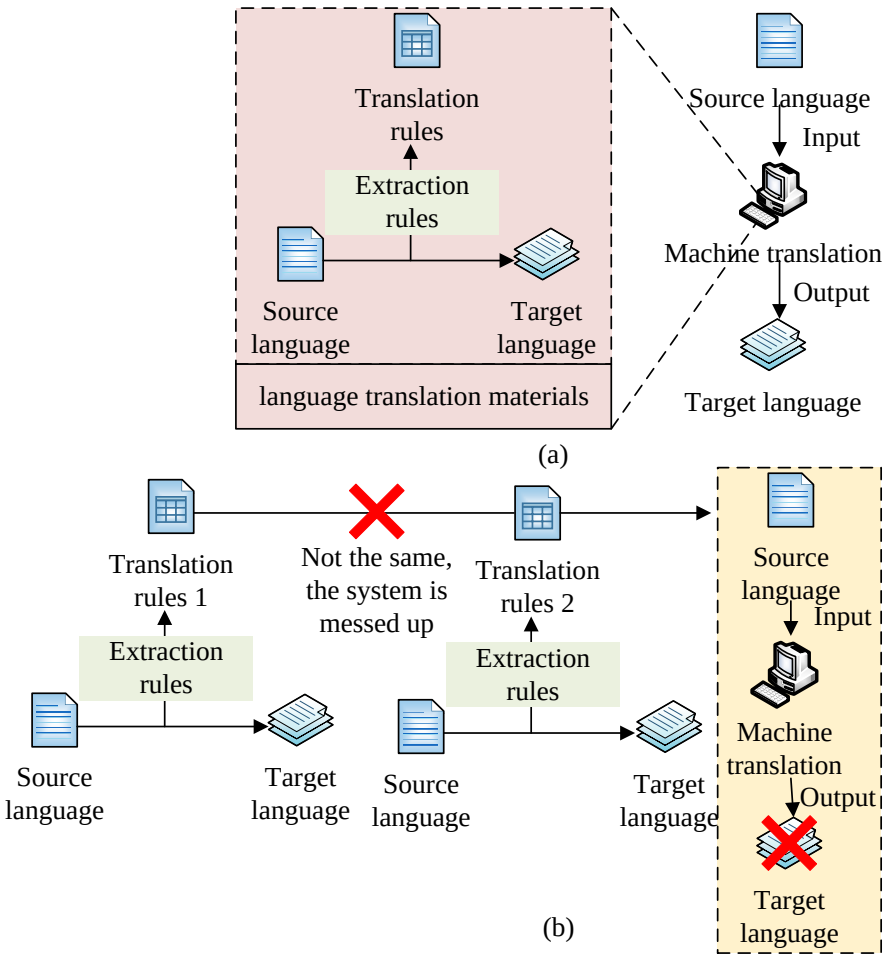

FIGURE 1. Principles of rule-based language translation (a: standard translation principle; b: system confusion translation error)

From Figure 1, rules-based translation systems continuously establish translation rules for language translation. However, with the increase in the number of translations and the differences between words in different languages, confusion about existing rules occurs, resulting in the translation system being unable to improve. An instance-based language translation system refers to a language translation system built by learning the translation method of language translation materials [10]. The basic principle is to translate the fragments of other sentences by analogy after learning the translation method of some sentence fragments. Finally, different sentence fragments are combined to complete the translation of the entire sentence. This method mainly relies on the experience of translating sentences to propose a translation case library for the language. Then, the differences between words in different languages are handled through translation rules proposed in the case library. This approach uses rule-based translation in only a small subset of language translations [11]. Instance-based language translation systems not only use traditional rule-based language translation methods but also create different translation systems. The system builds an example library by learning from translation examples in different languages and provides an advanced and practical method. After the instance-based translation system, a statistical-based translation method is proposed. It refers to the language translation process as a correspondence between the source and target languages. During translation, communication between the two is done from the point of view of probabilistic calculations. Based on the statistical translation system, the

translation results of the target language vocabulary are defined within a specific range. The words with the highest probability values are selected as the corresponding words in the source language by calculating and comparing the translation probabilities of different words. Finally, the vocabulary is matched and integrated based on the probability values [12]. Statistics-based translation systems mainly use statistical methods to learn primary translation models from language translation data. Then, the sentence with the maximum probability value is searched from the target language. Finally, the translation work is completed according to the translation model corresponding to the sentence of the source language. The principle of an instance-based language translation system can be seen in Figure 2.

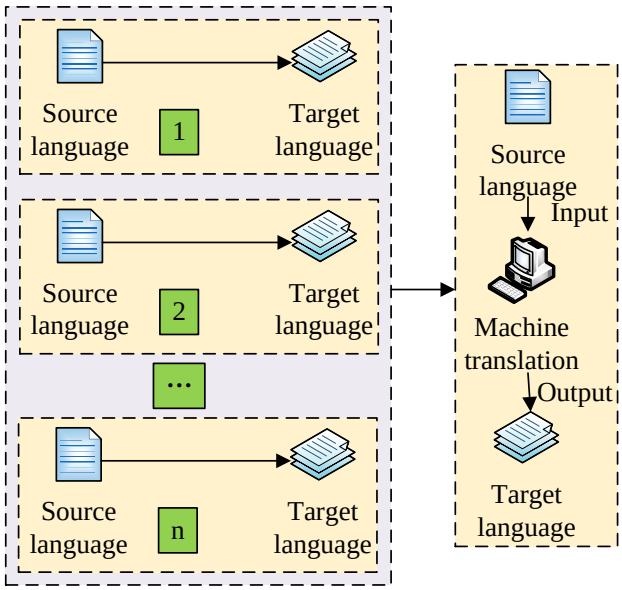

FIGURE 2. Principles of instance-based language translation

From Figure 2, the translation system is superior to the other two language translation systems in terms of robustness and performance improvement. It can handle differences between languages. In addition, in the language translation process, it can quickly absorb translation methods from new language translation materials, establish a language translation system, and improve translation performance in time [13]. Figure 3 shows the rationale for a probability-based language translation system.

The probability-based language translation system can quickly learn the language-translation method in the translation data and selectively translate the language by suggesting a probability method. Its translation performance is more advantageous than other MT methods [14].

2.2. **NMT.** NMT is an MT method that directly uses artificial neural networks to model translations end-to-end. The principle is to use computer technology and AI to mimic brain neurons for language translation. It depends on phrase-based statistical MT. NMT has many characteristics: (1) NMT learns language differently than humans. NMT relies on statistical correlation. NMT's outstanding point is also its handling of correlations in sentences to judge how to give more specific translations according to the context. (2) NMT looks at translated content from different perspectives. Early MT models focused on nouns and segments. However, this method is problematic when treating "long-distance dependency" languages. The NMT, on the other hand, looks at the whole sentence, or even the entire paragraph or article. This makes NMT more context-sensitive. (3)

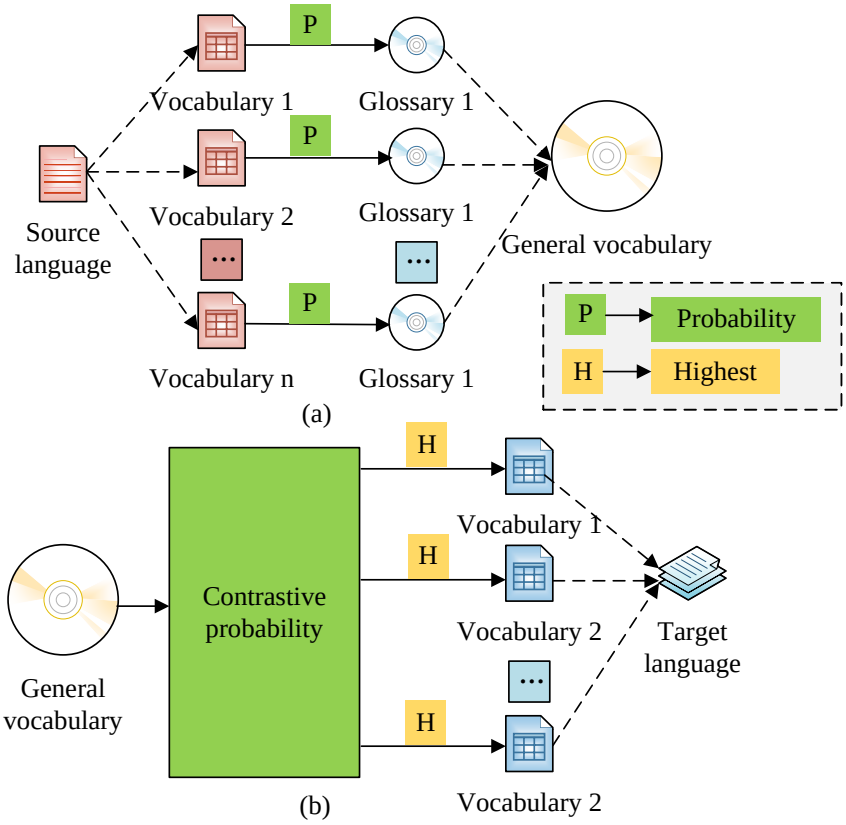

FIGURE 3. Principle of probability-based language translation (a: vocabulary translation; b: sentence translation)

NMT adopts a phrase-based approach. Therefore, it is more conducive to translating rich languages, like Hungarian. Also, it is possible to predict the part of speech by word formation. (4) NMT can be extrapolated across many languages to fill the gap in training data. In MT training, there are already English<>German and English<>Chinese language pairs, but there are no German<>Chinese language pairs. Then, without prior training, NMT can automatically translate new language pairs. Although the results may be worse than the results after training, the improvement from can't translate to being able to translate can be achieved.

In addition, there are still many challenges in NMT: (1) For neural networks engaged in NMT, the requirements for statistical technology of computing power are greatly increased. The efficient support of graphics processing units is required. (2) The translation language obtained by NMT is closer to the natural language. It even eliminates translation accents, making it easier for social interaction. However, it is unclear if this increase in fluency will cause a decrease in translation accuracy. When a translation error appears in ordinary MT, this error is likely obvious because it cannot be understood. However, the results of the NMT translation may be wrong, but it is very smooth. (3) Typical quality evaluation criteria may fail. Current methods for MT like Bilingual Evaluation Understudy may not be proper for NMT. Based on this, this paper will study the MT technology under the neural network to promote the development of this technology to give full play to the above characteristics and make up for the gap.

2.3. **DL techniques and their principles.** With the development of society and the increasing complexity of human needs, intelligent technology has become the main direction of the current social science and technology development. DL technology has become

the leading intelligent technology in the context of the current intelligent development of science and technology. Therefore, this paper studies MT technology by optimizing DL technology to enhance the effect of language MT and is beneficial for cultural exchange. DL is a machine learning method generated under the premise of developing the Internet of Things. It works on the principle of an abstract representation of the automatic hierarchical learning of the input problem through multiple layers of nonlinear changes in features. DL has significant advantages in feature processing, such as speech recognition, image processing and analysis, and AI [15]. As a result, it has also made significant achievements in language processing, including various DLNNs, such as Full Connected Networks (FCNs), Convolutional Neural Networks (CNNs), and RNNs. A Fuzzy Neural Network (FNN) is a hybrid model with mathematical properties. Its primary calculation method is as follows.

$$h_1 = W_1 x + b_1 \tag{1}$$

$$z = \sigma(h_1) \tag{2}$$

$$h_2 = W_2 z + b_2 \tag{3}$$

$$\hat{y} = soft\max(h_2) \tag{4}$$

x represents the value input in the model; W_1, b_1, W_2, and b_2 are the parameters that need to be learned at different levels of the network layer; h_1 and h_2 denote the calculation result of the layer; z and $\hat{y}$ stand for the intermediate result and final prediction result obtained by nonlinear calculation of the neural network. Figure 4 illustrates the structural framework of FNN.

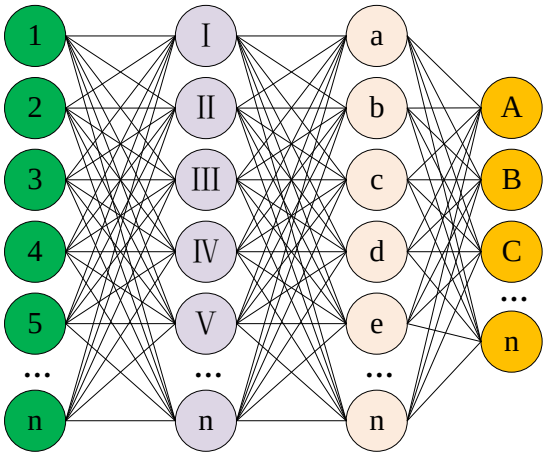

FIGURE 4. The basic structure of an FNN

The layers of FNN are connected, and the nodes at the same layer are isolated. There is a hidden layer consisting of 784 neurons and two output layers composed of 10 neurons, constituting the basic structure of an FNN. CNN is a very special feedforward neural network. As the name suggests, CNN is a DLNN that works through convolution calculation. CNN comprises convolution calculation, nonlinear excitation, and pooling layers. The convolution calculation layer calculates the input value in an instance through the convolution kernel function and obtains the result. The convolution kernel extracts the characteristics of the input value and transmits them to the nonlinear excitation layer for the following calculation. The nonlinear excitation layer converts the nonlinear result of the convolution calculation layer into a linear result. Then, it performs feature extraction, relying on the activation function [16]. The pooling layer reduces the dimensionality of the high-dimensional calculation results of the first two layers, usually using average and

max pooling methods. The dimensionality reduction process can preserve the original numerical features and enhance the numerical calculation effect. The specific structure of CNN and the pooling layer are exhibited in Figure 5.

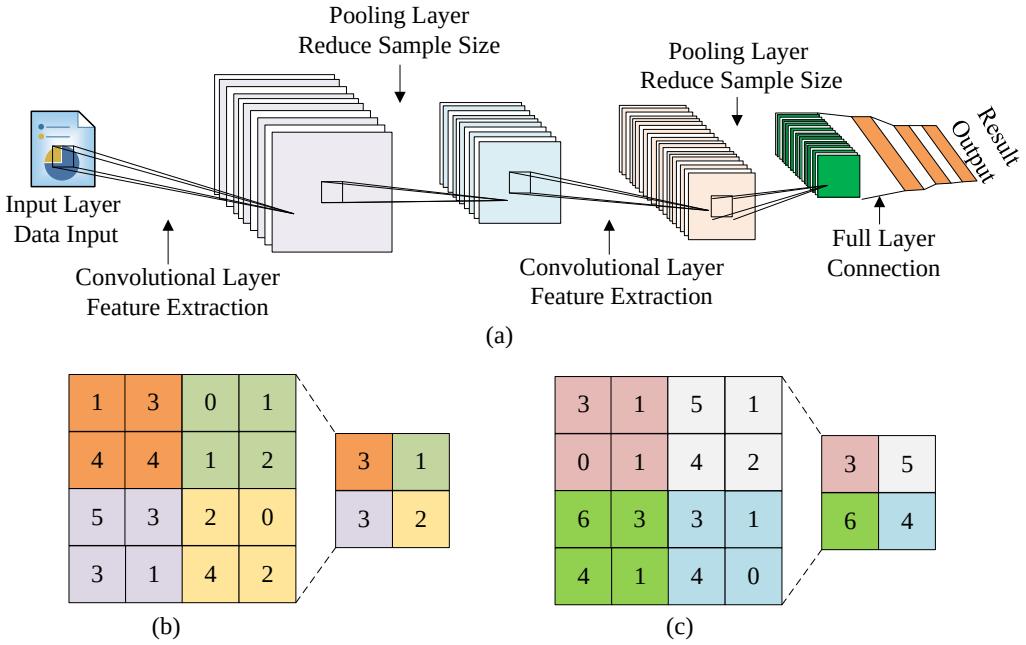

FIGURE 5. Primary structure and principle of CNN (a: the structure of CNN; b: average pooling; c: max pooling)

As presented in Figure 5, the data features will be reflected more intuitively after processing through different layers of CNN. RNN is generally used to process data in sequence, especially long sequences. Therefore, it has advantages in language data processing. It mainly consists of the input, output, and hidden layers. Figure 6 provides the basic structure of an RNN [17].

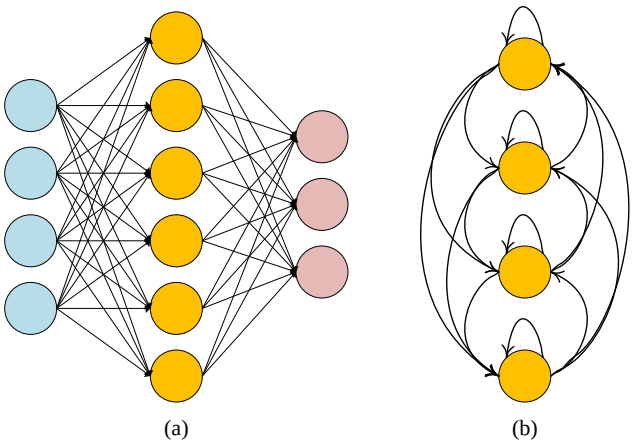

FIGURE 6. The basic structure of RNN (a: the whole network; b: the hidden layer)

According to the above Figure, the neurons in the input and output layers are only connected with those in the hidden layer. The hidden layer neurons exchange data features and perform comprehensive calculations. The hidden layer calculates data according to Equation (5).

$$\boldsymbol{h}^{\langle t \rangle} = \boldsymbol{f}\left(\boldsymbol{h}^{\langle t-1 \rangle}, \boldsymbol{x}^{\langle t \rangle}; \boldsymbol{\theta}\right) \tag{5}$$

In Equation (5), $\boldsymbol{x}^{\langle t \rangle}$ represents the vector of sequence data, $\boldsymbol{h}$ represents the state of the hidden layer, and $\boldsymbol{\theta}$ means the parameters of the RNN model. The overall model's calculation can be described as Equations (6) and (7).

$$\boldsymbol{h}^{\langle t \rangle} = \boldsymbol{f}\left(\boldsymbol{W}\boldsymbol{x}^{\langle t \rangle} + \boldsymbol{U}\boldsymbol{h}^{\langle t-1 \rangle}\right) \tag{6}$$

$$\boldsymbol{o}^{\langle t \rangle} = \boldsymbol{g}\left(\boldsymbol{V}\boldsymbol{h}^{\langle t \rangle}\right) \tag{7}$$

In Equations (6) and (7), $\boldsymbol{U}$, $\boldsymbol{V}$, and $\boldsymbol{W}$ are the parameters to be learned by the model, $\boldsymbol{x}$ indicates the input value, and $\boldsymbol{o}$ refers to the output value. Besides, $\boldsymbol{t}$ denotes the time; $\boldsymbol{f}$ and $\boldsymbol{g}$ mean the activation function. The former is generally the sigmoid function, and the latter is the softmax activation function. The loss function in language processing can be expressed as:

$$\boldsymbol{l}^{\langle t \rangle} = -\log \boldsymbol{P}\left(\boldsymbol{x}^{\langle t \rangle} \mid \boldsymbol{x}^{\langle t-1 \rangle}, \boldsymbol{x}^{\langle -2 \rangle}, \ldots, \boldsymbol{x}^{\langle t \rangle}; \boldsymbol{W}, \boldsymbol{U}, \boldsymbol{V}\right) \tag{8}$$

$$\boldsymbol{l} = \sum_{t=1}^{n} \boldsymbol{l}^{\langle t \rangle} \tag{9}$$

$\boldsymbol{P}$ stands for the calculation probability of the model. When dealing with sequence data, RNN generally adopts the data-sharing method and calculates the data in the hidden layer. The AM is an extensively used method in DL technology. The AM requires learning critical parts of the information and eliminating some invalid parts in the learning process to obtain a better result and improve learning efficiency [18]. The attention model combines the AM with the encoder and the decoder to deal with a specific task. The encoder and decoder are the main functional basis of the AM, which promotes the development of the AM in language processing. The attention model first encodes the source language, subsequently decodes it in the form of the target language, and finally completes the preloaded MT through assembly [19]. Each result of the decoding end is calculated and expressed as a probability, as presented in Equation (10).

$$\boldsymbol{P}\left(\boldsymbol{y}_i \mid \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{i-1}, \boldsymbol{X}\right) = \boldsymbol{g}\left(\boldsymbol{y}_{i-1}, \boldsymbol{s}_i, \boldsymbol{c}_i\right) \tag{10}$$

In Equation (10), $\boldsymbol{X}$, and $\boldsymbol{y}$ stands for sentences in the source and target language, respectively, and $\boldsymbol{g}$ represents a nonlinear transformation function. Figure 7 indicates the encoding-decoding translation principle under the AM.

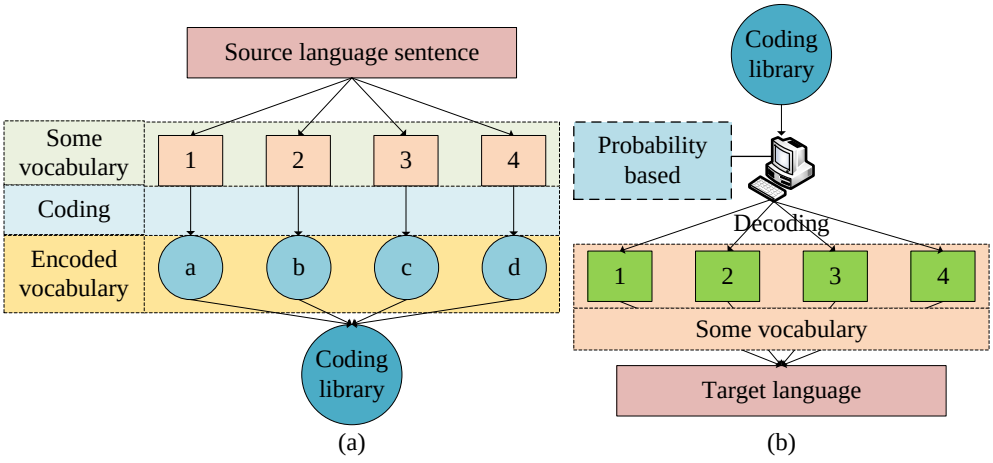

FIGURE 7. Attention model (a: the encoding process; b: the decoding process)

As shown in Figure 7, after applying the attention model, the translation efficiency and quality of the DLNN MT model have been greatly improved. Therefore, this paper uses the coding and decoding framework to study lightweight language translation [20].

3. Method.

3.1. **Method Overview.** The core of DLNN translation is the encoding and decoding program. The source language is translated through model training, model tuning, and translation decoding. The encoding and decoding program is mainly divided into two forms: the encoding and decoding program based on the traditional neural network and that based on the AM. Equation (11) signifies the calculation of the DLNN MT model.

$$\boldsymbol{p}(\boldsymbol{t} \mid \boldsymbol{s}) = \prod_{\boldsymbol{j=1}}^{\boldsymbol{m}} \boldsymbol{p}\left(\boldsymbol{t_j} \mid \boldsymbol{s}, \boldsymbol{t_{<j}}\right) \tag{11}$$

In Equation (11), $\boldsymbol{s}$, $\boldsymbol{t}$ denote the sentence information of the source and target language, and $\boldsymbol{p}$ refers to the probability result by calculation. The model encodes the words in the source language sentence during translation and decodes them based on the target language. Word decoding will be subject to the target language words with the highest probability. Then these phrases will be combined into sentences to make the translation more reasonable and smoother [21]. Taking the RNN as an example, the calculation process of the encoder can be presented as:

$$\vec{\boldsymbol{h}}_{\boldsymbol{t}} = \boldsymbol{RNN}\left(\vec{\boldsymbol{h}}_{\boldsymbol{t-1}}, \boldsymbol{V_s}\right) \tag{12}$$

$$\overleftarrow{\boldsymbol{h}}_{\boldsymbol{t}} = \boldsymbol{RNN}\left(\overleftarrow{\boldsymbol{h}}_{\boldsymbol{t+1}}, \boldsymbol{V_{s^\infty}}\right) \tag{13}$$

$$\boldsymbol{h_t} = \left[\vec{\boldsymbol{h}}_{\boldsymbol{t}}; \overleftarrow{\boldsymbol{h}}_{\boldsymbol{t}}\right] \tag{14}$$

$\vec{\boldsymbol{h}}_{\boldsymbol{t-1}}$ means the hidden state of the encoder at the $\boldsymbol{t-1}$ moment. $\boldsymbol{V}$ represents a vector representation of the source language. The decoder is defined as:

$$\boldsymbol{\eta_j} = \mathbf{RNN}\left(\boldsymbol{h_{j-1}}, \boldsymbol{V_{t^{\langle j-1\rangle}}}\right) \tag{15}$$

$$\boldsymbol{c_j} = \mathbf{Attention}\left(\boldsymbol{\eta_j}, < \boldsymbol{h_i} >_{\boldsymbol{i=1}}^{\boldsymbol{n}}\right) \tag{16}$$

$$\boldsymbol{h_j} = \boldsymbol{RNN}\left(\boldsymbol{\eta_j}, \boldsymbol{c_j}\right) \tag{17}$$

$$\boldsymbol{p}\left(\boldsymbol{t}^{<\boldsymbol{j}>} \mid \boldsymbol{s}, \boldsymbol{t}^{<\boldsymbol{j-1}>}, \boldsymbol{t}^{<\boldsymbol{j-2}>}, \ldots, \boldsymbol{t}^{<\boldsymbol{1}>}\right) = \mathbf{soft\,max}\left(\boldsymbol{g}\left(\boldsymbol{V_{t^{<j-1>}}}, \boldsymbol{h_j}, \boldsymbol{c_j}\right)\right) \tag{18}$$

$\boldsymbol{\eta_j}$ represents the network's output, $\boldsymbol{c_j}$ refers to the vector representation of the AM, and the rest is the same as the above equations. The DLNN MT model needs to be optimized. At present, maximum likelihood function optimization is the most mainstream optimization method, as shown in Equation (19).

$$\boldsymbol{L}\left(\boldsymbol{s}, \boldsymbol{t}; \boldsymbol{\theta}\right) = \log\, \boldsymbol{p}(\boldsymbol{t} \mid \boldsymbol{s}; \boldsymbol{\theta}) = \sum_{\boldsymbol{j=1}}^{\boldsymbol{T_m}} \log\, \boldsymbol{p}\left(\boldsymbol{t_j} \mid \boldsymbol{s}, \boldsymbol{t_{<j}}; \boldsymbol{\theta}\right) \tag{19}$$

In Equation (19), $\boldsymbol{\theta}$ represents the parameter of the attention model, and the rest are the same as the above equations. After the training is completed, the calculation result of the model needs to be checked. Specifically, the translation result needs to satisfy Equation (20).

$$\boldsymbol{t}^* = \underset{\boldsymbol{t \in T}}{\mathbf{argmax}}\, \boldsymbol{p}\left(\boldsymbol{t} \mid \boldsymbol{s}\right) \tag{20}$$

In Equation (20), $\boldsymbol{T}$ represents the overall space for selecting the target language during translation. After translation verification, it can be ensured that the translation result is the optimal result of MT. The MT process is all completed. Figure 8 illustrates the specific translation process of the DLNN model [22].

From Figure 8, when source language information is input to the DLNN translation model, the model first encodes the vocabulary in the source language. Then, the sentence

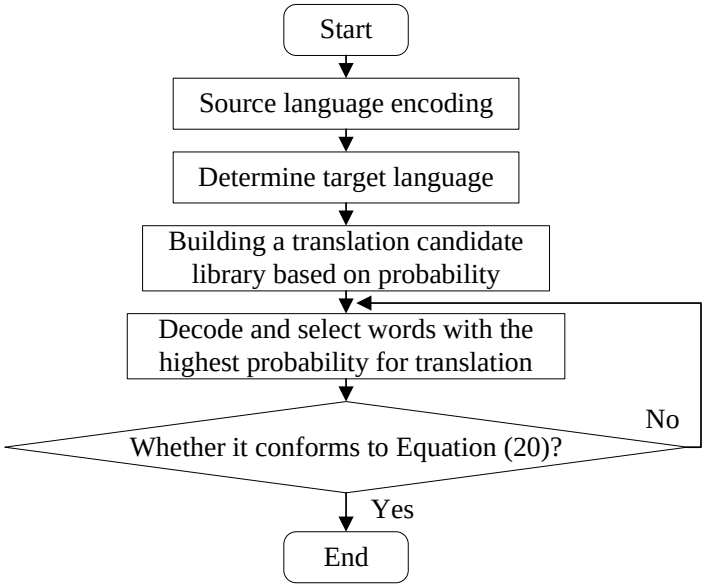

FIGURE 8. MT process of the DLNN translation model

is decoded based on the target language system. Language translation candidates are formed according to the decoding probability. The translation team matches the source and target language vocabulary based on the probability of the target language candidate's outcome. Finally, the model selects the target language vocabulary with the largest probability value to form a complete sentence and determines whether the translation result of the target sentence is in line with the expected target. If there is a match, the conversion ends; otherwise, the model will continue to check until the end [23]. Figure 9 shows the training code used here.

```
EPOCHS = 20
  for epoch in range(EPOCHS):
  start = time.time()
  enc_hidden = encoder.initialize_hidden_state()
  total_loss = 0
  for (batch, (inp, targ)) in enumerate(dataset.take(steps_per_epoch)):
    batch_loss = train_step(inp, targ, enc_hidden)
    total_loss += batch_loss
  if batch % 100 == 0:
      print('Epoch {} Batch {} Loss {:.4f}'.format(epoch + 1,
        batch, batch_loss.numpy()))
  print('Epoch {} Loss {:.4f}'.format(epoch + 1, total_loss / steps_per_epoch))
  print('Time taken for 1 epoch {} sec\n'.format(time.time() - start))
```

FIGURE 9. Training code for MT

In Figure 9, the DLNN MT method used here provides the necessary technical support for language translation. Moreover, this paper greatly adjusts the model through the above training process to make the model better.

3.1.1. *Model architecture.* In the machine translation system of this paper, the DLNN model is the core part. Its architecture is based on the encoder-decoder framework and is optimized in combination with the attention mechanism. In the encoder, the study uses

the RNN network to process the sequential information of the input source language sentence, and by inputting the source language sentence word by word, the encoder generates the hidden state representation of the vocabulary and passes it to the decoder.

The decoder gradually generates the translation of the target language based on the context vector generated by the encoder. The decoder also uses the RNN architecture, but when generating each word, it depends on the words generated in the previous step and the context vector. During the decoding process, the attention mechanism dynamically adjusts the translation of the sentence by calculating the weight of each word in the source language sentence and the target word.

3.1.2. *Hyperparameter settings.* The hyperparameter settings of this experiment are shown in Table 1.

Table 1. Hyperparameter settings

Parameters	Value	Parameters	Value
Word embedding dimension	256	Maximum sequence length	50
Hidden layer size	512	Discard rate	0.3
Batch size	64	Vocabulary size	30,000
Learning rate	0.001	Optimization algorithm	Adam

In Table 1, it can be seen that the word embedding dimension of the DLNN MT model is 256, the hidden layer size is 512, and the batch size is 64. The initial value of the learning rate is 0.001, and it is dynamically adjusted in combination with the Adam optimization algorithm. The maximum sequence length is set to 50, and the dropout rate is set to 0.3 to prevent the model from overfitting. The vocabulary size is 30,000, which indicates the maximum number of words supported by the model.

3.1.3. *Training process.* The specific steps in the DLNN MT model training process are as follows.

(1) The experiment first tokenizes the source language and target language sentences and converts them into indices in the vocabulary. To speed up training, batch training is used to pad fixed-length sentences to ensure that the model input dimensions are consistent.
(2) Initialize model parameters, including word embedding matrix, RNN weight matrix, and attention weight matrix.
(3) For each batch of source language sentences, the encoder generates a context vector, and then the decoder generates the target language sentence.
(4) The experiment uses the cross entropy loss function to evaluate the accuracy of target language generation and uses the Adam optimizer to update the model parameters based on the gradient backpropagation of the loss function. After each epoch, the performance of the model on the validation set is verified, and the learning rate is adjusted dynamically.
(5) During the training process, the translation performance of the model on the validation set is regularly evaluated, and the translation quality is measured using indicators such as the BLEU score. After the model training is completed, the test set is used for final evaluation.

3.2. **Experimental design.** The dataset used is Workshop on Machine Translation 2018. This dataset was applied in shared tasks at the 3rd MT Conference. The conference builds on twelve previous annual workshops about statistical MT. The total size

of the dataset is 1 million sentence pairs, including data from multiple language pairs. The conference has a total of multiple tasks: news and biomedical translation task, multimodal machine translation task, indicator task, quality assessment task, and parallel corpus screening task. The language translations used here cover French-English, Spanish-English, German-English, and Czech-English.

3.2.1. *Data Preprocessing.* For the dataset Workshop on Machine Translation 2018, this paper preprocesses noise data and unregistered words.

(1) Noise data processing

In machine translation tasks, there will be noise data in the dataset, such as typos, incomplete sentences, spelling errors, or abnormal symbols. The experiment removes samples that cannot be parsed normally, such as incomplete sentences and incorrect symbol translation pairs, and normalizes the redundant spaces, punctuation marks, and capitalization issues to ensure the consistency of the model input data.

(2) Unregistered word processing

Unregistered words are words that the model has never seen during the training process. This paper uses BPE (Byte Pair Encoding) technology to decompose rare words and unregistered words into common subwords, reducing the impact of unregistered words on the model, so that the model can better handle vocabulary diversity and improve the coverage of translation.

(3) Word segmentation and vocabulary normalization

To enable the model to better understand the structure and content of sentences, this paper performs word segmentation on the sentences in the dataset. The experiment uses word segmentation based on spaces and treats each word in French, Spanish, German, English, and Czech as an independent translation unit.

After the above data preprocessing and data cleaning, the experiment has a significant impact on the model performance. Eliminating noise data effectively reduces the erroneous input of the model, prevents the model from learning incorrect translation patterns during training, and improves the translation accuracy of the model. Especially on small sample data sets, the impact of noise data will be amplified. The experiment uses BPE technology to process unregistered words, which effectively improves the model's translation ability for low-frequency words and new words. It not only improves the translation coverage of the model but also increases the robustness of the model when translating unfamiliar language pairs. Appropriate word segmentation and vocabulary normalization operations help reduce the diversity of vocabulary, reduce the burden on the model, and improve the accuracy of translation.

3.2.2. *Experimental setup.* The model is first trained during the research process. During the training process, 100 and 200 samples are selected from the above four language translations. Then, the training results are statistically analyzed. The same number of samples of different properties are selected for test evaluation. Moreover, the experimental indicators used here are mainly the accuracy of model translation and the work efficiency of the model.

To ensure that the evaluation of the model is representative, this experiment divides the dataset into a training set and a test set. In the training set, each language pair contains 80,000 pairs of sentences, and in the validation set, each language pair contains 10,000 pairs of sentences for hyperparameter tuning during model training. In the test set, each language pair contains 10,000 pairs of sentences for evaluating the final performance of the model.

4. Verification.

4.1. **Comparison between DLNN MT and the traditional MT.** Traditional MT includes rule-based, instance-based, and statistics-based translation methods. Different MT models can be compared through the translation between different languages. Several language translation types contained in the data set used here are the testing. This paper compares the three traditional MT models regarding translation rate and accuracy. Figure 10 reveals the comparison results of three traditional MT models.

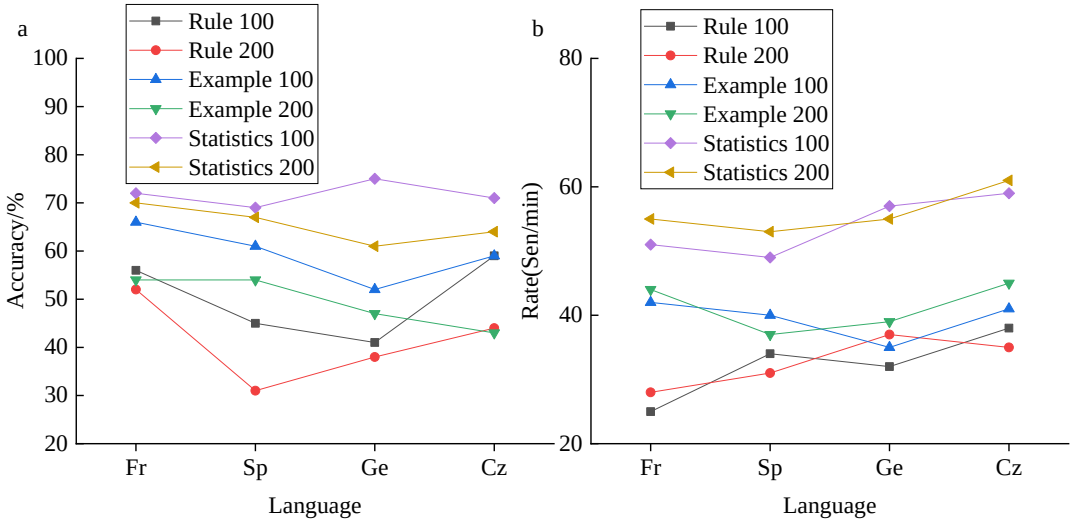

FIGURE 10. Evaluation results of three traditional MT models (a: comparison of the translation accuracy; b: comparison of the translation rate)

In Figure 10, training is first performed, and the proposed here is evaluated against other models. Fr stands for French-English, Sp stands for Spanish-English, Ge stands for German-English, and Cz stands for Czech-English. The rule represents a rule-based translation model, the Example represents an instance-based translation model, and Statistics represents a statistical-based translation method. Moreover, the experiment mainly uses the accuracy and translation efficiency of model translation as the main indicators for training and testing. According to the result, the translation precision of the statistical-based translation model is about 70%. The precision of the instance-based translation model is about 60%. That of the rule-based translation method is around 50%. Statistical-based translation methods are generally superior to the other two in terms of translation rate. Therefore, statistical-based translation methods are the most advantageous among traditional MT methods.

4.2. **DLNN MT Models.** The NMT models designed here include an FNN, a CNN, an RNN, and an AM. These models translate through statistical methods to maximize their performance because the statistics-based MT model has the best performance among the basic MT models. Figure 11 provides the evaluation results of the DLNN MT models designed here.

From Figure 11, the model was tested and assessed based on training, and the translation results of FNN, CNN, and RNN MT models with different sentence counts are compared. Similarly, the test indicators are the translation accuracy and efficiency of the model. It is found that the FCN model has a maximum accuracy rate, of about 84%. The precision of the CNN model is about 83% at most. The RNN model has a maximum accuracy of 88%, which is a significant advantage over the other two models. This

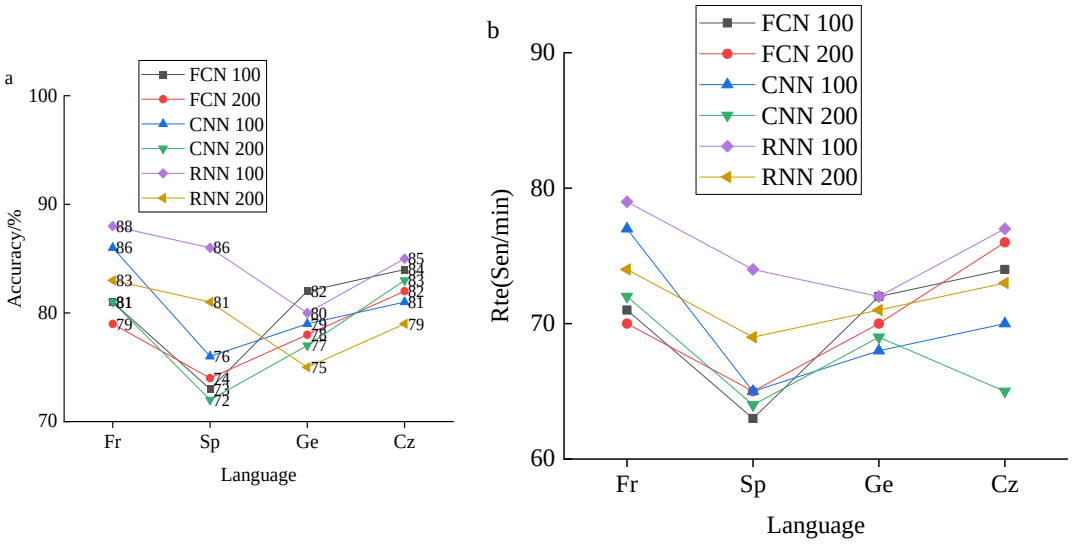

FIGURE 11. Evaluation results of DLNN MT models (a: comparison of the translation accuracy; b: comparison of the translation rate)

phenomenon may be due to data exchange between the hidden layers of the RNN model. Higher accuracy is achieved through exchange and comparison.

4.3. **Evaluation of the encoding-decoding program under the AM.** The encoding-decoding program encodes the source language through the AM and then completes the translation through decoding and probability evaluation to obtain the target language. It increases the comparison between more words to improve the accuracy of language translation and saves translation time through the encoding and decoding process to enhance the efficiency of NMT. Figure 12 shows the encoding and decoding rates of three NMT models under the AM.

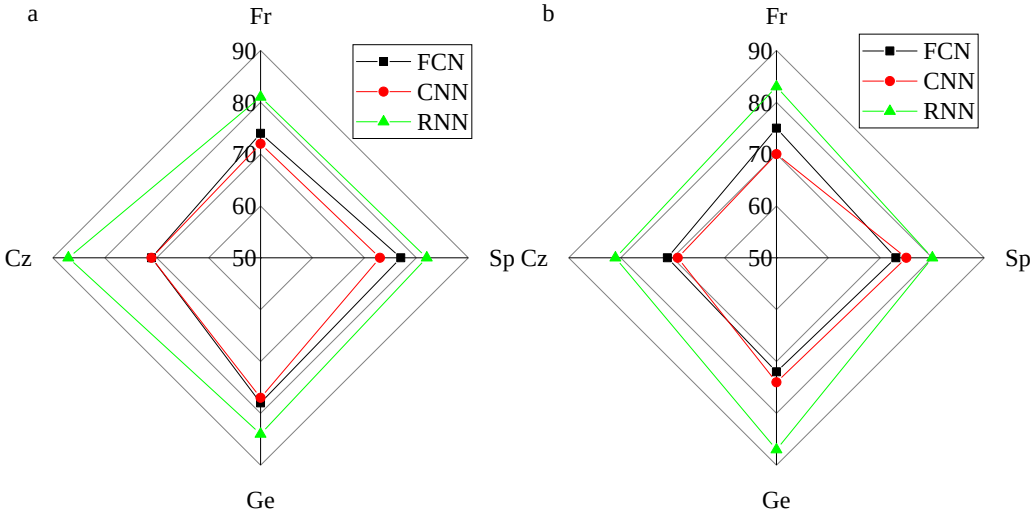

FIGURE 12. Comparison of the encoding-decoding rate of three DLNN MT models under the AM (a: encoding rate; b: decoding rate)

Figure 12 demonstrates that the encoding rate of the NMT model under AM has increased to 70 sentences per minute. In addition, due to language processing advantages, the RNN model's encoding-decoding rates exceed 80 sentences per minute. In summary, optimizing the mechanism of NMT model codes can enhance the quality of NMT models.

Moreover, the mechanism of NMT maintains the principles of statistical translation methods, expands comparative ability, and greatly enhances translation accuracy. Besides, to explore the improvement effect of the model studied in the present research, a comparative analysis is conducted between the traditional MT model and the proposed translation model, thus exploring the effect of the designed model in calculating translation efficiency and precision. The model comparison analysis results are portrayed in Figure 13.

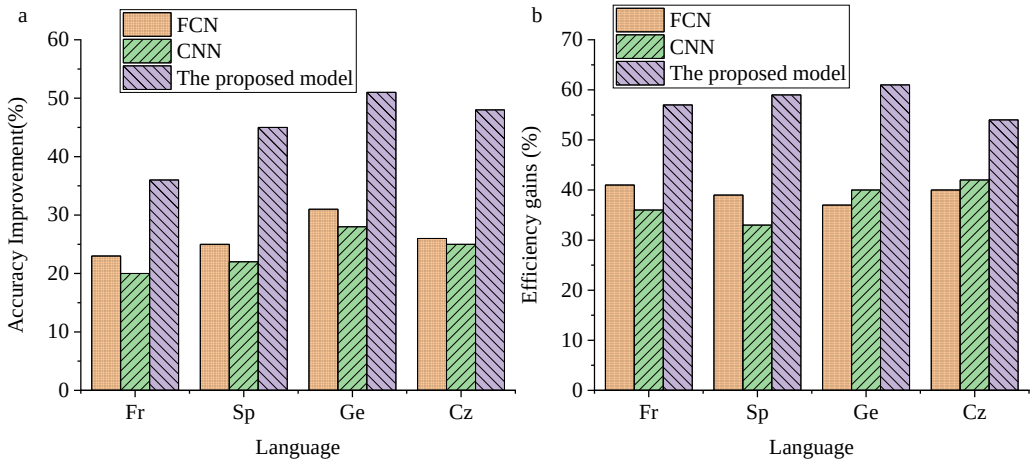

FIGURE 13. Comparison of model improvement effects (a: the comparison of calculation accuracy; b: the comparison of calculation efficiency)

In Figure 13, through comparison, it is found that the highest improvement in calculation precision of the model is about 48%, and the minimum value is about 20%. The model's maximum improvement in computational efficiency is about 61%, and the minimum is about 33%. It can be found that this paper has achieved significant technological breakthroughs, providing important support for the future development of MT technology. Compared to the research by Singh and Mahmood (2021) [24], this paper designs a more advanced and intelligent DL MT model. A comprehensive experimental study is conducted on this model, which improves the application effect of DL technology in MT and provides impetus for the development of MT technology.

4.4. **Performance test of the model on different languages.** To comprehensively explore the impact of different language features on the performance of the DLNN MT model, the experiment studied the performance of the DLNN MT model under different language translations under four language feature translations: Fr, Sp, Ge, and Cz. The results are shown in Table 2.

TABLE 2. Performance test results of the model on different languages

Language	Accuracy (%)	Translation speed (Sen/min)	Computational accuracy (%)	Computational efficiency (%)
Fr	88	79	92	84
Sp	86	74	89	82
Ge	80	72	75	74
Cz	85	77	85	78

In Table 2, French has the highest model accuracy and calculation precision, 88% and 92% respectively. The translation speed and calculation efficiency are also excellent. It

can be seen that the language structures of French and English are relatively similar, and the model can better translate and deal with them. The accuracy and translation speed of Spanish are slightly lower than French, but the overall performance is still better. The translation accuracy and calculation precision of German are relatively low, and the translation speed is also slow, reflecting that the complex grammar and morphological changes of German are more challenging for the model. Although the Czech model performs better than the German, it still lags behind French and Spanish in computational accuracy and efficiency.

Linguistic features have a significant impact on model performance. Compared with English, the vocabulary structure and grammar of French and Spanish are relatively close, and the model's translation accuracy and calculation efficiency are better. However, German and Czech have relatively low translation accuracy and efficiency due to their more complex grammar and frequent lexical morphological changes, which increase the processing difficulty of the model. It shows that characteristics such as language similarity and grammatical complexity will directly affect the performance of the DLNN MT model.

4.5. **Comparison with advanced models.** To comprehensively evaluate the effectiveness of the DLNN MT method, it is now compared with Transformer, BERT, GPT-3, and T5. The results are shown in Figure 14.

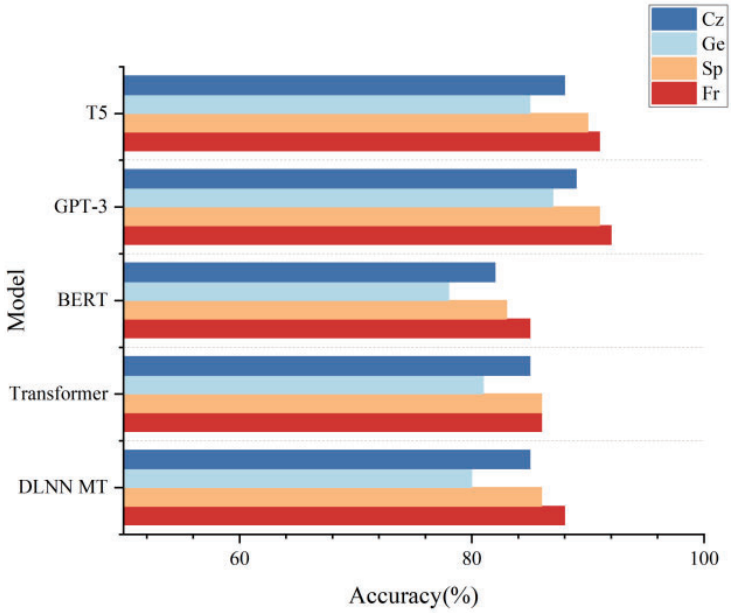

FIGURE 14. Comparison results of translation accuracy with advanced models

In Figure 14, GPT-3 has the highest translation accuracy on all language pairs, which shows its excellent ability in handling language translation tasks, especially on French-English and Spanish-English pairs, with an accuracy of 92% and 91%. T5 also performed very well overall, with an accuracy of 91% and 90% on French-English and Spanish-English pairs, respectively. The DLNN MT model performed relatively stably on the four language pairs, with accuracies of 88%, 86%, 80%, and 85%, respectively, which was comparable to Transformer on French-English and Spanish-English pairs, but was slightly worse when processing other language pairs. Transformer and BERT performed relatively evenly, slightly lower than DLNN MT and T5 on some language pairs. Overall, the DLNN MT model has certain advantages in translation accuracy and coverage of language pairs, indicating its effectiveness.

4.6. **Performance of the attention mechanism in different sentence patterns and semantic scenarios.** To further analyze the specific effects of the attention mechanism in different sentence patterns and semantic scenarios, this paper uses quantitative analysis to analyze the performance of the attention mechanism in translation tasks. The experiment evaluates the effect of the attention mechanism on improving translation accuracy in three types of sentence patterns, simple sentences, complex sentences, and interrogative sentences, and three types of semantic scenarios, narrative scenarios, description scenarios, and reasoning scenarios. The performance of the attention mechanism in different sentence patterns and semantic scenarios is shown in Figure 15.

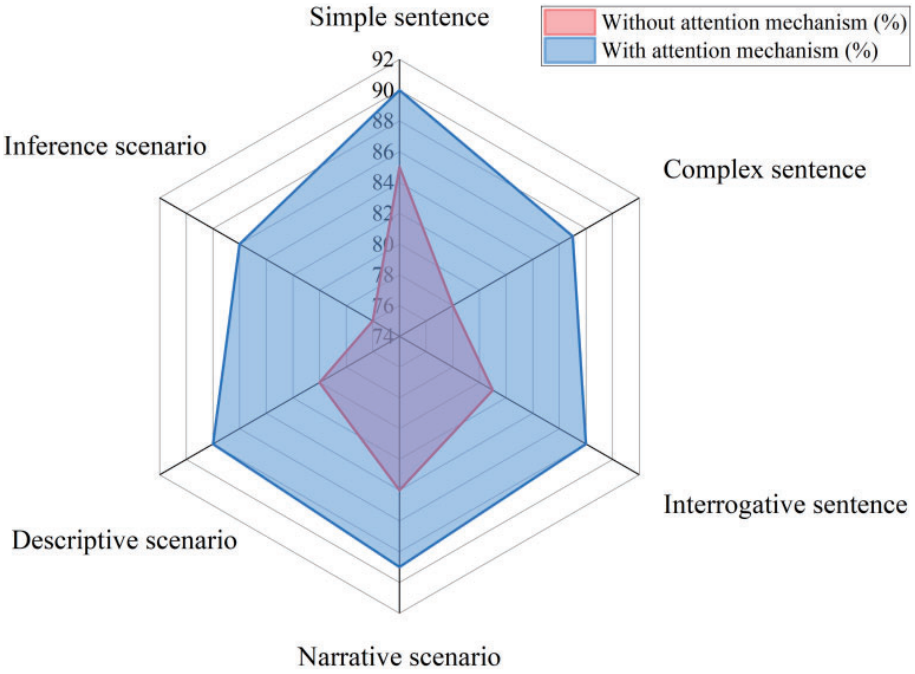

FIGURE 15. Performance of the attention mechanism in different sentence patterns and semantic scenarios

As can be seen in Figure 15, the attention mechanism has the most significant improvement in the translation effect of compound sentences and reasoning scenes, with improvements of 9% and 10% respectively. It shows that the attention mechanism can better handle complex sentence patterns and semantic reasoning tasks, and improve the accuracy of translation. The improvement in simple sentences and narrative scenes is relatively small but still maintains an improvement of about 5%. The results show that there are certain differences in the performance of the attention mechanism in different sentence patterns and semantic scenarios, especially when processing complex sentence patterns and scenarios that require semantic understanding, the effect is more prominent.

4.7. **Relationship between DLNN MT model complexity and translation performance.** To study the relationship between DLNN MT model complexity and translation performance, this paper tests the impact of network layers, number of nodes, and other parameters on translation accuracy. The experiment analyzes the balance between model complexity and performance by adjusting the number of network layers of the model, from shallow to deep layers, and the number of nodes in each layer. The results of the impact of different model complexities on translation performance are shown in Figure 16.

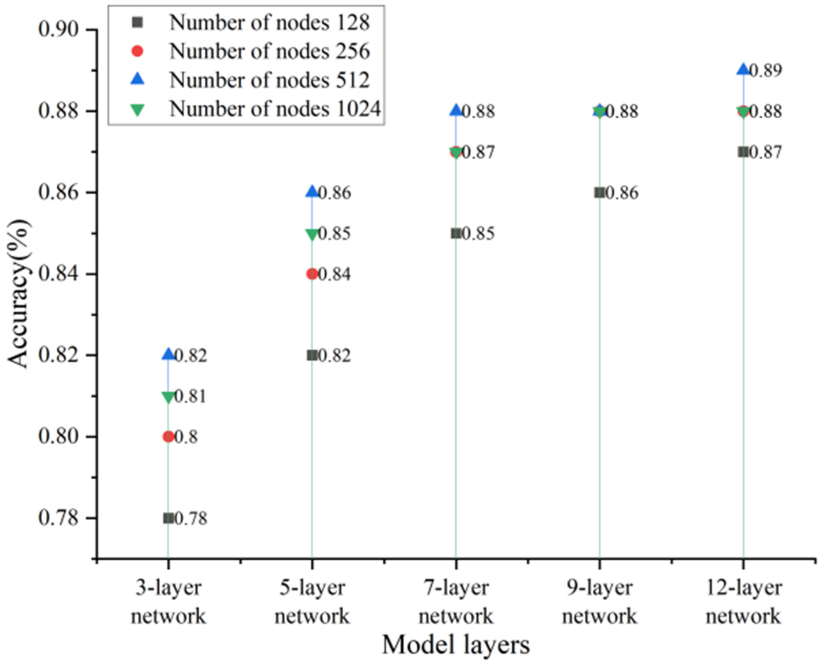

FIGURE 16. Relationship between DLNN MT model complexity and translation performance

In Figure 16, as the number of model layers and nodes increases, the translation accuracy gradually improves. When the number of network layers exceeds 7 and the number of nodes exceeds 512, the performance improvement tends to be flat, showing a slight downward trend. This shows that the DLNN MT model can effectively improve the translation accuracy under a certain complexity, but excessive increase in complexity will increase the computational cost, while the performance improvement is limited. The configuration of a 7-layer network and 512 nodes showed the best balance in this experiment, which can not only ensure high translation performance but also avoid the performance bottleneck caused by excessive complexity.

5. **Experimental Discussion.** Based on the results of this paper, future improvements can start from improving the quality of the corpus. The performance of the DLNN MT model varies between different languages, especially for complex languages. The translation accuracy and efficiency of German and Czech are relatively low. To solve this problem, the training corpus can be further enriched and optimized in the future so that the model can better learn and process these complex language features. At the same time, the processing capabilities of the lexical morphology and grammatical rules of specific languages can be increased, and the accuracy of translation can be improved by introducing more complex word form change analysis modules. Customizing specific decoding strategies for the grammatical features of different languages enables the model to more accurately match the translation needs between language pairs.

Optimizing the model training algorithm is one of the key directions in the future. The DLNN MT model has introduced optimized autoencoders and recurrent neural networks but still encounters efficiency bottlenecks when training on a large scale. In the future, more advanced neural network structures, and self-attention mechanisms such as Transformer and BERT will be explored to further improve the model's processing capabilities for long sentences and complex sentences. At the same time, using multi-task learning and transfer learning methods to allow the model to share parameters between multiple

languages can enhance the generalization ability of the model and further improve the translation effect of low-resource languages. By combining cultural differences and language background information, the model can better cope with implicit semantics in the language and improve the overall translation quality.

6. **Conclusion.** With the continuous development of the concept of global cultural exchange, language translation has become a hot research topic in today's society. To enhance the quality of language translation and improve the performance of MT technology, this study adopts DL technology for research, focusing on optimizing the decoding process to enhance the precision and efficiency of language translation.

Firstly, the principle and application of MT are outlined. Secondly, the application principle of DL technology is analyzed. Finally, by optimizing DLNN technology, a lightweight MT model is designed and evaluated. The main contribution of this research to the field of machine translation is the introduction of deep learning technology and an optimized decoding process, which significantly improves translation accuracy and efficiency. The research results reveal that statistical-based translation models achieve better accuracy and efficiency than traditional translation models based on rules and instances. Especially, the RNN model has advantages in terms of translation accuracy and efficiency due to its unique working principle and mechanism, with a translation accuracy rate generally exceeding 80%. In addition, introducing an automatic encoder can promote the encoding-decoding rate of the DLNN MT model, thereby effectively improving translation accuracy. The optimization strategy in this study significantly improves the performance of machine translation systems and promotes the application of multi-language translation. The research provides strong support for improving translation accuracy and system adaptability by improving corpus quality and optimizing training algorithms.

However, this paper only focuses on optimizing the encoding and decoding program of the model, without an in-depth analysis of other influencing factors in practical applications. Therefore, future research will further explore more aspects of the model in practical applications and comprehensively improve its performance. Possible development directions include improving corpora quality, machine learning training algorithms, and better considering cultural differences. These efforts will help further enhance the development of the MT field and promote global communication and understanding.

Funding. This work was supported by the 2023 Scientific Research Key Program of 14th five-year plan of China Association for Educational Technology "Research on the Implementation Path of Digital Empowerment English Teaching in Local Private Universities" (Grant No: G029) and 2023 Youth Foundation of Wuhan Donghu University "Research on the Implementation Path of Digital Empowerment English Teaching in Local Private Universities" (Grant No: 2023dhsk033).

REFERENCES

[1] Z. Ali, "Research chinese-urdu machine translation based on deep learning," *Journal of Autonomous Intelligence*, vol. 3, no. 09, p. 34, 2021.

[2] S. Mahata, D. Das, and S. Bandyopadhyay, "Mtil2017: Machine translation using recurrent neural network on statistical machine translation," *Journal of Intelligent Systems*, vol. 28, no. 05, pp. 447–453, 2018.

[3] V. Sanz Marco, B. Taylor, Z. Wang, and Y. Elkhatib, "Optimizing deep learning inference on embedded systems through adaptive model selection," *ACM Transactions on Embedded Computing Systems (TECS)*, vol. 19, no. 02, pp. 1–28, 2020.

[4] K. Rajan, A. Zielesny, and C. Steinbeck, "Stout: Smiles to iupac names using neural machine translation," *Journal of Cheminformatics*, vol. 13, no. 04, pp. 1–14, 2021.

[5] M. Ali, M. L. Rahman, J. Chaki, N. Dey, and K. Santosh, "Machine translation using deep learning for universal networking language based on their structure," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 08, pp. 1–12, 2021.
[6] E. Brynjolfsson, X. Hui, and M. Liu, "Does machine translation affect international trade? evidence from a large digital platform," *Management Science*, vol. 65, no. 12, pp. 5449–5460, 2019.
[7] F. Stahlberg, "Neural machine translation: A review," *Journal of Artificial Intelligence Research*, vol. 69, no. 8, pp. 343–418, 2020.
[8] J. Makin, D. Moses, and E. Chang, "Machine translation of cortical activity to text with an encoder-decoder framework," *Nature Neuroscience*, vol. 23, no. 04, pp. 575–582, 2020.
[9] R. Dabre, C. Chu, and A. Kunchukuttan, "A survey of multilingual neural machine translation," *ACM Computing Surveys*, vol. 53, no. 5, pp. 1–38, 2020.
[10] Y. Belinkov, N. Durrani, F. Dalvi, H. Sajjad, and J. Glass, "On the linguistic representational power of neural machine translation models," *Computational Linguistics*, vol. 46, no. 01, pp. 1–57, 2020.
[11] T. Linzen and M. Baroni, "Syntactic structure from deep learning," *Annual Review of Linguistics*, vol. 7, no. 15, pp. 195–212, 2021.
[12] M. Zeeshan, Jawad, M. Zakira, and M. Niaz, "A seq to seq machine translation from urdu to chinese," *Journal of Autonomous Intelligence*, vol. 4, no. 01, pp. 1–5, 2021.
[13] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, "A guide to deep learning in healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 24–29, 2019.
[14] J. Guo, H. He, T. He, L. Lausen, M. Li, H. Lin, X. Shi, C. Wang, J. Xie, S. Zha, A. Zhang, H. Zhang, Z. Zhang, Z. Zhang, and S. Zheng, "Gluoncv and gluonnlp: Deep learning in computer vision and natural language processing," *Journal of Machine Learning Research*, vol. 21, no. 23, pp. 1–7, 2019.
[15] S. Xie, Z. Yu, and Z. Lv, "Multi-disease prediction based on deep learning: a survey," *CMES-Computer Modeling in Engineering and Sciences*, vol. 2, no. 32, pp. 15–17, 2021.
[16] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *IEEE Access*, vol. 9, no. 11, pp. 44 304–44 321, 2021.
[17] J. Yang, A. Li, S. Xiao, W. Lu, and X. Gao, "Mtd-net: Learning to detect deepfakes images by multi-scale texture difference," *IEEE Transactions on Information Forensics and Security*, vol. 16, no. 8, pp. 4234–4245, 2021.
[18] J. M. Stokes, K. Yang, K. Swanson, W. Jin, and J. J. Collins, "A deep learning approach to antibiotic discovery," *Cell*, vol. 180, no. 4, pp. 688–702, 2020.
[19] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, 2019.
[20] S. Ghosh, I. Das, N. Das, and U. Maulik, "Understanding deep learning techniques for image segmentation," *ACM Computing Surveys*, vol. 52, no. 4, pp. 1–35, 2019.
[21] S. Smys, J. I. Z. Chen, and S. Shakya, "Survey on neural network architectures with deep learning," *Journal of Soft Computing Paradigm (JSCP)*, vol. 2, no. 3, pp. 186–194, 2020.
[22] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad Khasmakhi, M. Asgari-Chenaghlu, and J. Gao, "Deep learning–based text classification: A comprehensive review," *ACM Computing Surveys*, vol. 54, pp. 1–40, 04 2021.
[23] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikinen, "Deep learning for generic object detection: A survey," *International journal of computer vision*, vol. 128, no. 2, pp. 261–318, 2020.
[24] S. Singh and A. Mahmood, "The nlp cookbook: Modern recipes for transformer based deep learning architectures," *IEEE Access*, vol. 9, pp. 68 675–68 702, 2021.