# Machine Translation-Based Language Modeling Enables Multi-Scenario Applications of English Language

Shengming Liu*
School of Foreign Languages and Culture, Panzhihua University, Panzhihua 617000, China

***Abstracts*—Traditional machine translation models suffer from problems such as long training time and insufficient adaptability when dealing with multiple English language scenarios. At the same time, some models often struggle to meet practical translation needs in complex language environments. A translation model that combines the feed-forward neural network decoder and the attention mechanism is suggested as a solution to this problem. Additionally, the model analyzes the similarity of the English language to enhance its translation ability. The resulting machine translation model can be applied to different English scenarios. The study's findings showed that the model performs better when the convolutional and attention layers have a higher number of layers relative to one another. The highest average value of the bilingual evaluation study for the research use model was 29.65. The research use model can machine translate different English language application scenarios and also the model performed better. The new model performed better than the traditional model and was able to translate the English language well in a variety of settings. The model used in the study had the maximum parameter data size of 4586, which is 932 higher than the lowest statistical machine translation model of 3654. The metric value was 3.96 higher than the statistical machine translation model. It is evident that investigating the use of the model can enhance the English language scene translation effect, with each scene doing well in translation. This provides new ideas for the direction of multi-scene application of machine translation language model afterwards.**

***Keywords*—Machine translation; decoder; English language; multi-scene; attention mechanism**

## I. INTRODUCTION

Machine translation (MT) refers to the text translation process that uses only machines to translate natural language. MT can translate a large number of English language (EL) text messages more conveniently and quickly, and reduce the communication barriers and other problems that occur in people's work [1]. As science and technology have advanced, machine translation (MT) has grown in importance as a text translation tool. But the traditional MT model can no longer meet the increasingly complex language environment, so how to improve the efficiency of MT, so that it is more suitable for a number of different scenarios has become an important direction of the current research [2]. The traditional MT model is mainly based on the Transformer to carry out multi-level stacked translation to realize the purpose of MT. The high complexity of traditional models makes them take a lot of time in the training process, which makes it difficult to meet the requirements of real-time and high efficiency in multi-scene applications. In addition, the traditional model relies on a multi-level stacked translation mechanism, which causes problems such as dimension explosion when dealing with long text sequences [3]. A type of deep-structure neural network model called a convolutional neural network uses the feed-forward neural network's (FNN) attention mechanism (AM) to shorten training times and enhance training results [4]. FNN is often used in complex data models because it is easy to be trained and can handle high-dimensional data. By processing sequential data more effectively, AM can enhance the accuracy and performance of machine learning models. The AM module can effectively reduce the dimension explosion problem when translating long sequences of text and improve translation accuracy and efficiency. The FNN module can reduce the overall complexity of the model and shorten the training time. The meta-learning module can ensure the translation effect of the model in complex scenes. The encoder-decoder structure enables efficient sequence-to-sequence translation. Due to the addition of the attention mechanism, the encoder-decoder structure can process and translate long sequence texts more accurately. On this basis, the study improves the Transformer based on the traditional MT by adding AM and FNN, and reduces the complexity of the model by extracting the local semantic information of words through AM. The research uses a decoder to decode and analyse linguistic terms and filter out closer results. The new system is able to integrate the attention mechanism and feed-forward neural network, optimising the decoder part of the traditional Transformer model. To improve the system's ability to capture the local semantic information in the language sequence and reduce the complexity of the model. At the same time the model improves the adaptation effect to different English scenarios by introducing a meta-learning approach. The study is broken down into six sections. Introduction is given in Section I. Section II gives detail about the related work. English language model for machine translation is given in Section III. Section IV provides analysis of English language. Discussion is given in Section V and finally, Section VI concludes the paper.

## II. RELATED WORK

MT is a translation model for linguistic text translation through software, which is currently widely used in different translation fields. Using MT to translate independently translated sentences, Maruf S et al. were able to enhance the model's translation quality by adding a neural network model. The study also evaluated the current applicable models and

*Corresponding Author.

analyzed them using both training and test sets. The study's findings demonstrated that applying the approach can improve research in this area [5]. Translation effort is complicated by ambiguity, as discovered by Yang M et al. when certain uncommon words are frequently substituted with tokens in machine translation (MT). Therefore, a hierarchical clustering approach was proposed, by building a dataset of rare words up to be able to be assembled into the MT framework, which can be successfully translated. The results of the study indicated that the quality of MT can be improved by using this method [6]. Dabre R et al. found that the use of multilingual neural MT can improve the quality of translation, and this translation model is more convenient and faster compared with the traditional MT. Therefore the study investigated this MT method in more depth and explored the future research direction of this method by analyzing the current literature and use cases, which provides new ideas for future research in this direction [7]. Khaw L L et al. found a new model for enhancing students' English context through writing. To improve the model's capacity for contextual learning, a RA engineering model was used in the development of the new model. The study's findings demonstrated how well this methodology worked to improve students' contextual learning [8].

According to Ban H et al., as Internet technology has advanced, so too has the necessity for interlanguage communication, and language models used in automatic translation have gotten more efficient. In order to create a novel MT model using an encoder and decoder, the research applied a deep learning MT model. By mapping between languages, the new model can enhance performance by immediately converting data into vectors. The study's findings demonstrated that the new model might outperform the conventional model [9]. Bharti N et al. concluded that the ranking of translated languages can improve the performance of MT and also train the MT effectively. Therefore, the study used a new mechanism for MT by sorting the MT output statements. The study's findings showed that the novel approach can enhance machine translation (MT) performance and is helpful for translating linguistic studies [10]. Wu H found that interactive translation of MT using multimedia is an effective way of translating a large number of utterances and programming languages. Therefore, the study proposed a fuzzy learning-based MT system that uses multimedia software to improve the performance and effectiveness of MT. Finally, the study pointed out the defects and advantages of the current multimedia translation and proposed a new translation program. Additionally, the study's findings demonstrated that employing the new model can improve English translator training [11]. Maimaiti M et al. found that there are many ways to enhance the data in MT, but the traditional methods are difficult to ensure the quality of the data after enhancement. Therefore a new evaluation framework was used in the study for data enhancement and the new approach used a discriminator model in order to reduce syntactic and semantic errors. The outcomes demonstrated that the data enhancement performance of this approach is significantly better than other approaches [12].

In summary, most of the studies in MT are mainly to enhance the effect of translated data and MT performance, followed by the improvement of MT through machine learning, but this approach often fails to analyze and study more complex EL environments. To improve MT's capacity for scene adaption, the study superimposes multi-layer FNN and attention to incorporate AM and FNN into the MT model. Lastly, to improve the model's capacity to feature extract data, the data is feature extracted from the model using conditional random fields (CRF) and bidirectional encoder representations from Transformers.

## III. English Language Modeling for Machine Translation

This section mainly focuses on the model construction of EL based on the MT model, firstly, the AM will be added to the model, and then the structure of the decoder and other structures will be introduced and analyzed, followed by the analysis and elaboration for the purpose of improving the model effect and coping with the translation of EL in different scenarios.

### *A. Language Model for Machine Translation Incorporating the Attention Mechanism*

Language model is a processing model to convert the traditional natural language, the current commonly used language model is mainly using neural networks for language processing and model design. When creating the target language, machine translation (MT) is a crucial model of language processing that can typically convert languages, decode and analyze data from the source language, and treat the previous result as the output of the latter item to obtain the source language model formula for MT, as indicated in Eq. (1) [13].

$$P(Y \mid X,\theta)=\prod_{t=1}^{T^{Y}+1}P(y_t \mid y_{t-1},X;\theta) \tag{1}$$

In Eq. (1), $X$ denotes the source language sequence, $T^{Y}$ is the length of the sequence, and $Y$ is the final language sequence obtained. $\theta$ denotes the set of data parameters and $T$ denotes the length of translated words. Generally the use of parameter estimation to train the machine for translation enables to get the size of the loss function value of the current formula model. Usually in the use of MT it is necessary to convert the linguistic data into dimensional vectors that can be used by the machine, so it is necessary to use an encoder to analyze the MT process. The encoder-decoder's construction is depicted in Fig. 1.

In Fig. 1, the encoder adds word data such as "I", "LOVE", "I", "HOME "The data is converted by the decoder, and then the converted data is expressed in machine language by the decoder, such as "HOME" is expressed as [1,0,1], this kind of data transcoding can solve the problem of semantic conversion, but usually there is also a dimensional explosion in the conversion process problem. Therefore, the study adds the attention module to the traditional model to solve the problem of long language sequences. Because in the translation process of EL, it is impossible to compare and analyze each word, and in fact, the translation only needs to translate the adjacent part of the word, so the mapping of the local word can get the result of all the words translated. For this reason, AM is added to the translation process of EL, which is used to capture the information of global features. As shown in Fig. 2.

In Fig. 2, the attention module mainly includes multiple attention and convolutional attention. Multiple attention includes parameters such as temporal expression of EL sequences, word expression, and so on. Convolutional attention is mainly the linguistic dimension segmentation of the convolutional layer, and the parameter data of the two kinds of attention are segmented and processed to obtain the latest model attention output. The sequence expression of AM is shown in Eq. (2) [14].

$$S = \{W_1, W_2, \cdots, W_n\} \tag{2}$$

In Eq. (2), $S$ denotes the output of the language sequence and $W_n$ denotes the word expression in the sequence. The sequence vector encoding at this point is shown in Eq. (3).

$$E = I\text{nput}(S) \tag{3}$$

In Eq. (3), $E$ belongs to the set of word sequence vectors, and the obtained vectors are subjected to the dimensional splitting operation of the attention input, and the different dimensions of attention are correlated to obtain the correlation shown in Eq. (4) [15].

$$d_S = d_X + d_{X_2} \tag{4}$$

In Eq. (4), $d_S$ is the dimension size of the language sequence and $d_X$ is the dimension size of the number of attention heads. The attention heads is set to $h_T$ and the convolutional AM is set to $h_C$ in Eq. (5) to obtain the AM allocation formula as shown in Eq. (5) [16].

$$h_T * d_T^h + h_C * d_C^h = d_T + d_C = d_S \tag{5}$$

In Eq. (5), $d_T^h$ denotes the dimension assigned to each number of attention for multiple, and $d_C^h$ denotes the dimension assigned to each number of attention for convolution. After analyzing the dimensional information of the attention, the initial attention dimension (AD) is used to build the encoder and decoder as shown in Fig. 3.

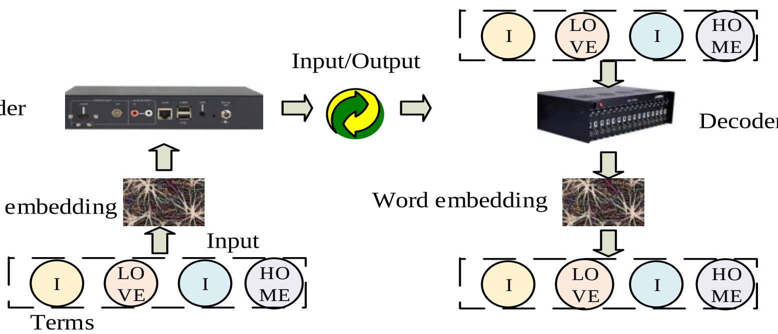


Fig. 1. Encoder decoder structure.

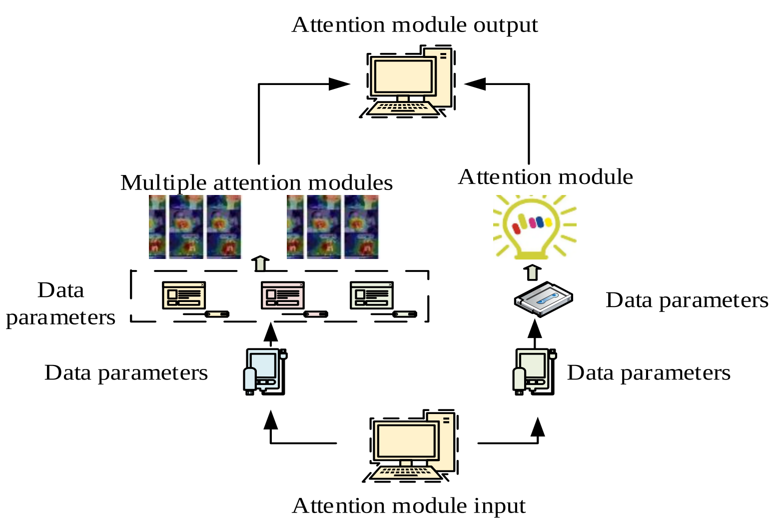


Fig. 2. Information capture process of global features.

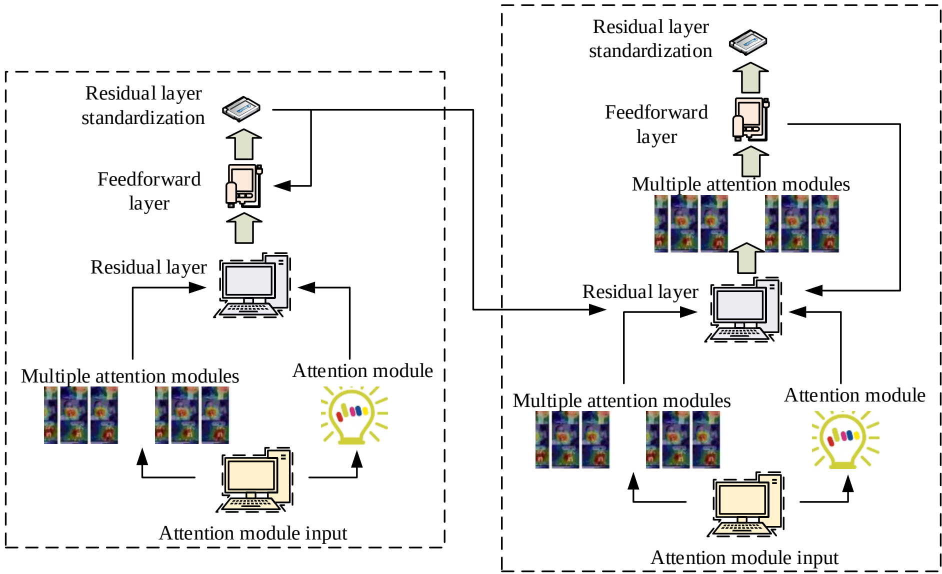


Fig. 3. Initial attention dimension construction process.

Fig. 3 shows the addition of a front-loaded feedback layer and a residual connection layer during the pre-modeling stage of the encoder. The feedback layer primarily modifies the linear data to increase the attention model's computational capacity. The decoder's entire structure is the same as the encoder's, but with additional multiple attention added. The model utilized is the same as the traditional model module, and the EL data obtained after encoding in the encoder is transferred into the encoder for recomputation. In the English data input part of the model, the existing vector positions and inflectional lengths are set, the input data vectors are aligned by means of data complementary zeros, and then the initialization calculation of English word vectors is carried out by the function and the positions are encoded in the data vectors, and the final data obtained are analyzed as the input dimensions.

The encoder stage of the model first splits the data vectors to divide the different dimensions of the attention model, and the assigned dimensions of the two attention models are calculated to obtain the dimension size of the encoder attention convolution [17]. Finally the obtained results and dimensions are merged by computation and later fed into the pre-feedback network to enhance the data. The decoder output stage requires the addition of the same attentional model that enables the current model to obtain more EL text data. The obtained text data is used as input to the attention matrix of the decoder [18]. Again same as the encoder stage the result of the previous layer is augmented by the feed forward network before outputting and the final result obtained is the current desired result.

*B. Machine Translation Model for English Modal Scene Application System*

In the translation of EL scenarios, the main problem is the EL translation and usage problem, which can be solved by analyzing the MT encoder and decoder in the previous section to study the problems such as English translation. The main composition of the model used in the study is composed of AM and Transformer, mainly to enhance the module's feature extraction (FE) ability for nonlinear vectors and words in low latitude space. The upper part of the model used belongs to the feed forward network module as shown in Fig. 4.

In Fig. 4, the feedforward network module is mainly composed of multiple attention models and feedforward network module, while the network module includes data linear change operation, multiple perceptual layers and neural network layer optimization. The inputs and outputs of the feedforward network layer are expressed to obtain the formula for the FNN layer as shown in Eq. (6) [19].

$$FNN(D) = relu(DW_D + b_D)W_R + b_R \tag{6}$$

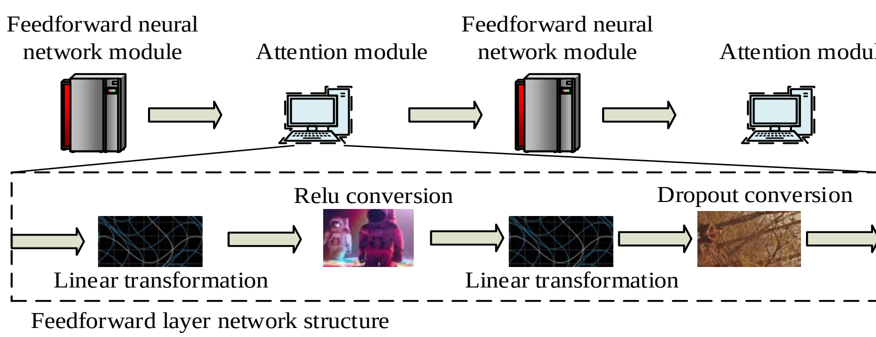


Fig. 4. Feedforward network module.

In Eq. (6), $D$ denotes the input of data and $W_R$ denotes the output of data. $b_R$ is the AD of the output and $b_D$ denotes the AD of the input. The MT model of AM has been able to translate specific ELs, but the information and terminology of the domains are not fixed for different English scenarios, and in order to further make the model adapt to more scenario applications, the model needs to be trained by meta-learning methods. Fig. 5 displays a schematic diagram of the training model.

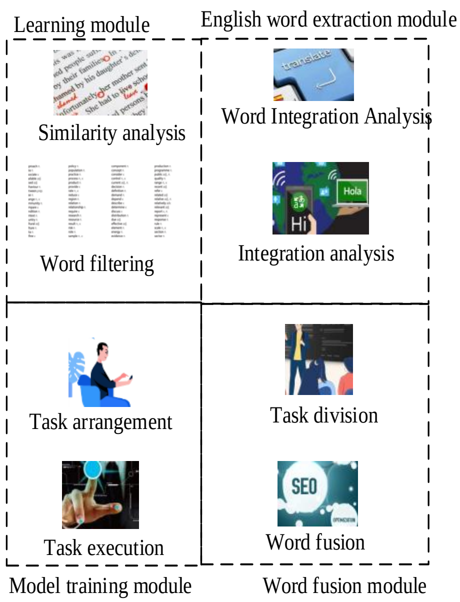


Fig. 5. Schematic diagram of training model.

In Fig. 5, the training for the model is mainly divided into several modules for training, learning module, English word extraction module, word fusion module, and model training module. The learning module is mainly responsible for word similarity analysis of the current application scenarios. The English word extraction module performs word incorporation analysis for different English scenes. The word fusion module divides the tasks for meta-learning. The model training module arranges training tasks for the training process of the model. It is vital to extract words from the language used in the current scene during the model training process in order to integrate additional information from the model training data. This is done by marking the words in the sentence and assessing them in relation to various vocabularies and scenarios for comprehension. Extracting data through the method of extracting words requires FE of data from the feature expression layer and the conditional random distribution layer [20]. Fig. 6 depicts the extracted words FE procedure.

To acquire the feature representation of the EL sentence using the model, the input sequence in Fig. 6 must first be feature extracted. Secondly, the deeper neural network is input as a conditional randomized feature layer. In the stochastic conditional feature layer it is necessary to define the state feature function and map the input feature sequence to the probability distribution of the sequence, and each state function is subjected to positional correlation calculation. If a feature function that can be identified is labeled, the value of the feature function at the current labeled position is used as the current input, and the model FE is able to return a corresponding real number. In addition to defining the function labeled for FE, it is also necessary to define the position of the feature function. The distribution of word features in the model is calculated for the best marking case given by the sequence, and the probability of the sequence is calculated using the function, as shown in Eq. (7) [21].

$$p(y \mid x) = \frac{1}{Z(x)} \exp\left(\sum_{i=1}^{n}\sum_{k=1}^{K} \lambda_k f_k(y_{i-1}, y_i, x_i) + \sum_{i=1}^{n}\sum_{j=1}^{K} \theta_j g(y_{i-1}, y_i)\right) \tag{7}$$

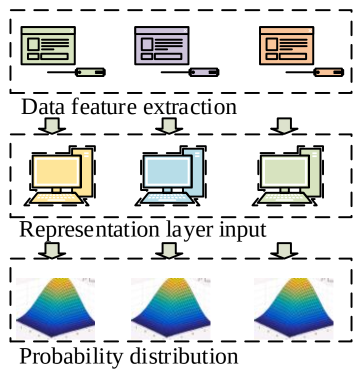


Fig. 6. Extraction process of word features.

In Eq. (7), $Z(x)$ denotes the normalization factor, which is capable of summing up all labeled sequence probabilities, guaranteeing that the distribution of the current probabilities are all integer 1. $\lambda_k$ and $\theta_j$ denote the parameters of the model. $f_k(y_{i-1}, y_i, x_i)$ denotes the current defined feature function. $g(y_{i-1}, y_i)$ denotes the value of another defined feature function. Following the model's training, a comparable function serves as the loss function. The model's parameters are then determined by gradient descent minimization of the loss function, and the best position for the sequences is ultimately determined by decoding the data following training [22].

In addition to the extraction of feature data it is also necessary to translate the sentences in the target context, but since there are already multiple aligned words in the current environment, the study requires automatic acquisition of aligned sentences in the context and translation of the target sentence (TS). As shown in Eq. (8) [23].

$$A = (x_i, y_i) : x_i \in X, y_i \in Y \tag{8}$$

In Eq. (8), $X = x_1, x_2, \cdots, x_n$ denotes a random sentence in the scene and $Y = y_1, y_2, \cdots y_m$ denotes a parallel sentence corresponding to the TS. $(x_i, y_i)$ denotes a word pair where two words are semantically similar in the same sentence in the same context. When extracting word embedding for multiple words, the word embedding need to be obtained through the hidden state of the model, and the similarity of the words is calculated after the word embedded words are obtained. In this way the dot product of contextual principal vectors of word embedding can be computed for each feature of the target word and the aligned word. After that all the features exceeding the threshold set by the model are filtered and then the similarity matrix is obtained through the contextual word embedding as shown in Eq. (9) [24].

$$S = h_x h_y^T \tag{9}$$

In Eq. (9), $S$ denotes the probability distribution of the similarity matrix, and both $h_x$ and $h_y$ denote the word embedding data. The matrix of initial phrases is obtained by this method, and then the final alignment matrix is obtained by ensemble interaction between the initial sentence (IS) and the sentences obtained from the TSs, as shown in Eq. (10) [25].

$$Align = (S_{xy} > c) \cap (S_{yx}^T > c) \tag{10}$$

In Eq. (10), $S_{xy}$ denotes the IS matrix, $S_{yx}^T$ denotes the TS matrix, and $c$ denotes the threshold value. When the threshold value is set to 1 and the rest of the parameter items are set to 0, then the value of the alignment matrix in the matrix is equal to 1 means that at this time the two TSs and the IS are aligned with each other [26]. As some texts are processed for word alignment in different scenarios, multiple subwords need to be aligned [27]. However, in different scenarios where a term includes multiple words and a word contains multiple segmented sub-words, it is only necessary to align a sub-word in the scenario with each other, and then the words can be considered to be aligned with each other. However, this method will greatly increase the error generated during the training of the model, so it is also necessary to filter the aligned sentences. The method used in the model is to represent the two sentences as vectors and calculate the cosine similarity between the two words as shown in Eq. (11) [28].

$$similarity = \cos(\theta) = \frac{A * B}{|A||B|} \tag{11}$$

In Eq. (11), $similarity$ denotes the magnitude of cosine similarity between two sentences and $\cos(\theta)$ denotes the cosine value. $A$ is the IS word and $B$ is the TS word. For the model the loss function is calculated as shown in Eq. (12) [29].

$$L(\phi) = \sum_{t=1}^{T} l^t(\hat{\theta}^t) \tag{12}$$

In Eq. (12), $L(\phi)$ denotes the value of the loss function and $\hat{\theta}^t$ denotes the specific network in the model that performs the initialization operation. $T$ denotes the number of executed tasks, and $l^t(\hat{\theta}^t)$ denotes the loss value of the network performing the task in the model. When the network task is executed at initialization, its execution network is the same as the initial network, then after the model training update, the execution parameters of the current network are different from those of the initial network. At this time, the loss value is calculated, and the loss value is aggregated so that all the loss values of the network model can be obtained [30]. The final flow of the obtained multi-scene translation decoding model is shown in Fig. 7.

In Fig. 7, the initial phase of the model involves first matching the EL file from the system, loading the EL, after which the model is trained through the model. It is determined whether there is a request or not, if yes then the initial EL is obtained from the text, if not then it waits for the instruction to be issued. After obtaining the IS the sentence language is pooled to choose whether to use the decoding pool or to sample the data. After that the EL sentences from different scenarios are modeled and segmented, the resulting ELs are translated into text using MT, and finally the translated results are output as text.

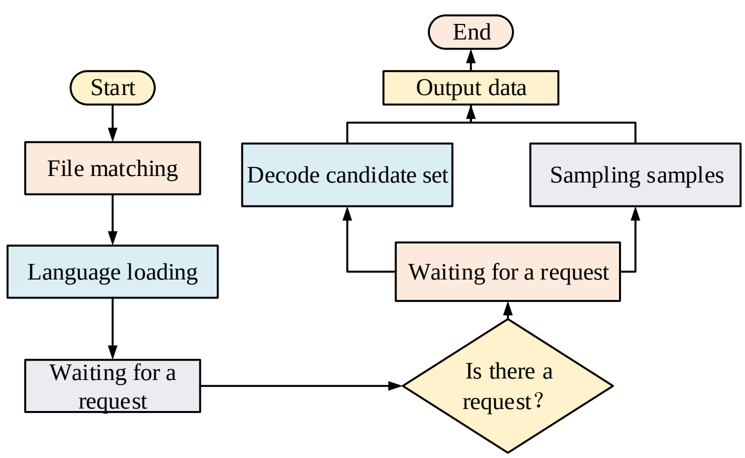


Fig. 7. Multi scene translation decoding model process.

## IV. Application Analysis of English Language Scenarios Based on Machine Translation Language Modeling

This section's primary goal is to validate the advanced nature of the model utilized in the current study by conducting an experimental analysis of the model and comparing and analyzing the impact of model translation. Second, in order to confirm the model application effect even more, experiments are conducted to confirm the translation effect of the model currently in use in various circumstances.

### *A. Machine Translation Model Effect Analysis*

The study uses the WMT-20 English dataset published by MT Association for data processing. The size of the test set in this dataset is 7.9k and the size of the training set is 3.0 M. The processor used for the simulation model of the data used is Intel i9-10920X, the operating system is selected as Ubuntu 20.04, the size of the RAM is 64GB, the graphics card is selected as RTX3090, and the dimensionality of the word embedding is set to 512, and the dimensionality of the hidden layer is set to 2048. The attention and convolutional layers were analyzed in different ratios, and several different ratios were experimentally analyzed to find the best ratio. Eight different model comparison ratios are selected for the study. For example, the convolutional layer comparison attention module is 1:8, 2:7, 3:6, 4:5, 5:4, 6:3, 7:2, and 8:1. These eight attention ratios are compared to get the model ratio training parameter data as shown in Table I. The evaluation index is evaluated with the score of bilingual evaluation understudy (BLEU), which is the evaluation index of MT results, and its score is able to evaluate and analyze the MT model. Comparison of different proportions of the model effect is obtained as shown in Fig. 8.

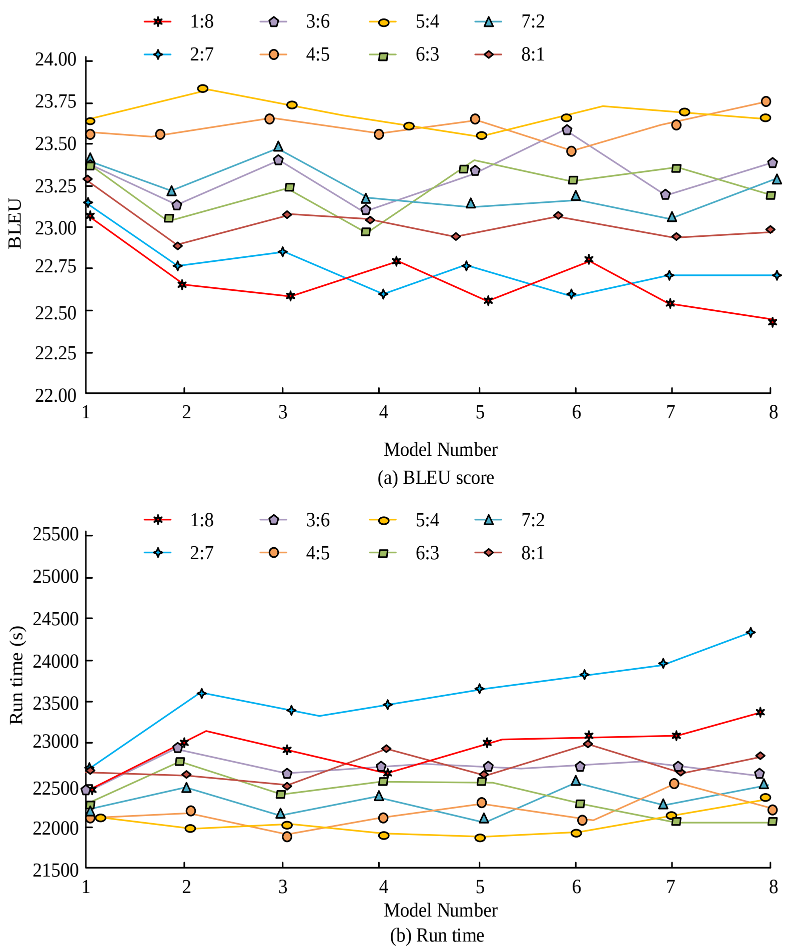


Fig. 8. Comparison of running speed and BLEU values of models with different scales.

In Fig. 8(a), in several model ratios when the number of convolutional and attentional layers are closer to each other, the BLEU value of model performance is larger, which indicates that when the number of layers of convolutional and attentional layers are closer to each other the model performance results are better, so that the highest BLEU mean value of the model with the ratio of 4:5 and 5:4, respectively, is 23.48 and 23.67. Comparing the lowest ratio model 1: 8 mean value of 22.67, it is higher by 0.81 and 1.00 respectively. In Fig. 8(b), in the running time comparison, the closer the ratio is to its performance, the shorter the running time is, the shortest running time is 5:4 ratio at this time the average running time is 22065s, the highest running time is 2:7 running time average is 23587s, the difference between the two ratios of the model running time is 1522s. It can be seen that the closer the ratio of the two modules is to each other, the better the model's effect is. This might be because the attention model runs faster in terms of parameter values and gives the model a higher dimensionality, which translates into a shorter running time. To examine how the number of attention layers affects the model's data processing effect, a variety of distinct attention layers were examined and tested, yielding the comparison parameters displayed in Table I

The model's parameter values, BLEU values, and running time all grow as the number of attention layers does in Table I, suggesting that while adding more attention layers can improve the model's MT index, doing so also lengthens the model's computation time. This suggests that increasing the attention model to boost the model's efficiency is not feasible after the number of layers is chosen and at the same time, the running time and the running efficiency of the model need to be taken into account, so as seen in the figure, the relatively better number of attention layers is 3, and the running time of this layer is 22548s, which has a relatively shorter running time, and at the same time, the BLEU value is at a higher value. To analyze the effect of feedforward network on the model effect, choose the above model in which the attention layer and convolutional layer are close to each other, 3:6, 6:3, 4:5 and 5:4 these four models are compared and tested, and the different feedforward network layers are analyzed to get as shown in Fig. 9.

The BLEU value of the model in Fig. 9(a) increases as the feedforward layers increase, but the four models' BLEU values change relatively little from one another. The model with the ratio of 5:4 has a larger overall change in BLEU value, but the model as a whole joins a feedforward network and its BLEU value does not change significantly. This may be due to the different ratios of the models and has little effect on how many feedforward layers are added. As the feedforward layers is increased, as shown in Fig. 9(b), the model's running speed rises and reaches its maximum when there are four feedforward layers. It can be seen that in the selection of feedforward layers, using higher feedforward layers has a good MT index but at the same time brings slower running speed, so it is more appropriate to choose 3 feedforward layers in the selection of feedforward layers. The parameter data, which shows the number of parameters used in the model during operation, will be compared with the data as shown in Fig. 10 to test the model's current use after adding various feedforward layers and pay attention to the model's effect. The larger the parameter data, the more parameters the model requires during operation, and the better the model's overall performance.

In Fig. 10(a), the parameter data of the model increases as the layers used in the model increases in several models, and the research uses the highest amount of model parameter data, which indicates that the research uses the model with the best results in model processing. The lowest value of the model parameter data for the FNN-only model indicates that the FNN-only model does not enhance the data processing of the model well. In Fig. 10(b), the model increases the running time with the increase of the model layers, and the running time of the research use model has relatively less time in the comparison of the five models, with an average time of 2854 s. The FNN-only model has the shortest running time, which might be because it can significantly cut down on the amount of time it spends using data parameters. In Fig. 10(c), the model used by the study has the highest BLEU value with a mean value at 29.65 and the model using only AM has the lowest BLEU value with a mean value of 21.84. The difference between the mean values of the two models is 7.81. In order to compare the effectiveness of the different algorithmic models with the method used by the study, the traditional statistical machine translation (SMT), rule-based machine translation (RBMT) and neural machine translation (NMT) are compared with the research use model obtained as shown in Table II.

TABLE I. COMPARISON OF DATA WITH DIFFERENT ATTENTION LEVELS

| Model | Parameter values | BLEU | Run time (s) |
|---|---|---|---|
| Attention level 1 | 48756.00 | 20.65 | 19587.00 |
| Attention level 2 | 50325.00 | 22.36 | 20549.00 |
| Attention level 3 | 52364.00 | 23.54 | 22548.00 |
| Attention level 4 | 55368.00 | 25.41 | 24596.00 |
| Attention level 5 | 58642.00 | 27.54 | 25368.00 |

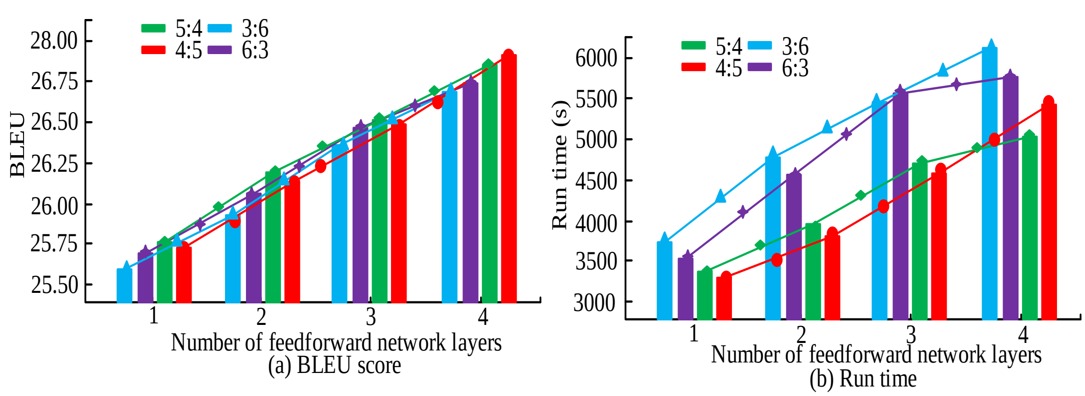


Fig. 9. Comparison of data from different feedforward layers.

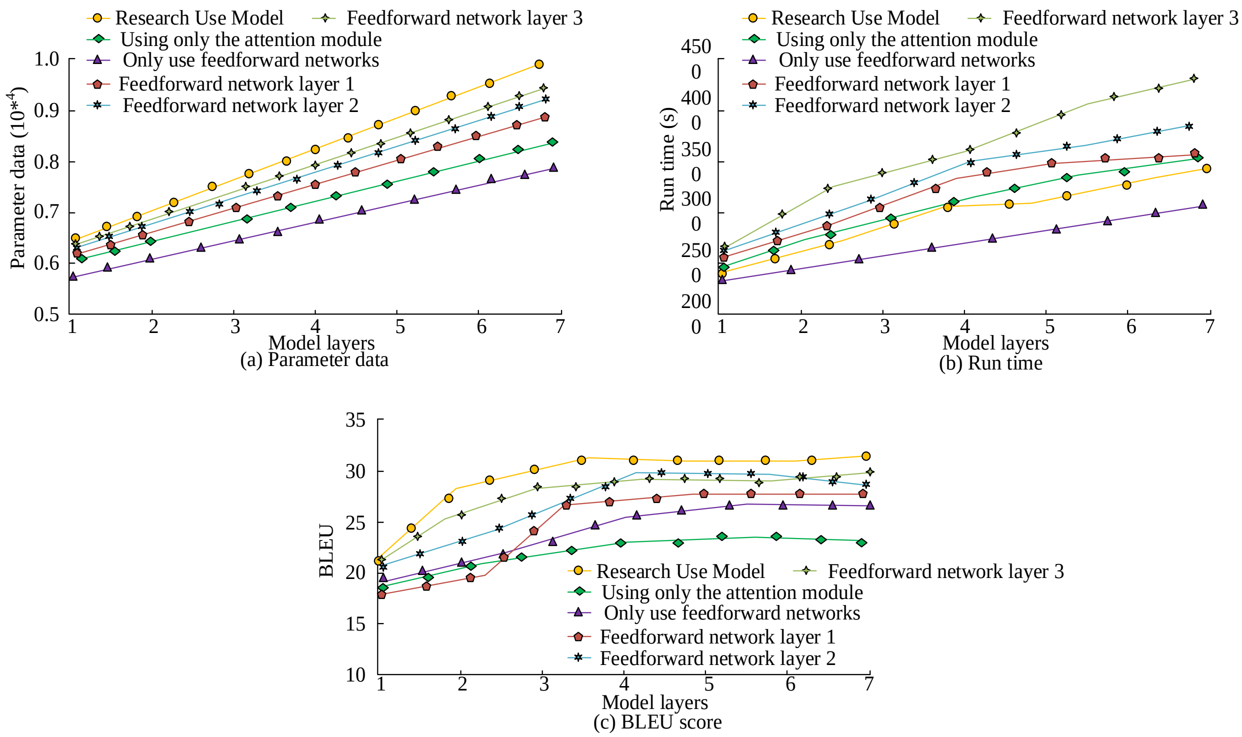


Fig. 10. Comparison of data from different network models.

TABLE II. COMPARISON OF DATA FROM DIFFERENT MODELS

| Data set | Model | Parameter data | BLEU | Run time (s) |
|---|---|---|---|---|
| Dataset 1 | SMT | 3654.00 | 22.58 | 2456.00 |
| | RBMT | 3685.00 | 22.98 | 2635.00 |
| | NMT | 4210.00 | 24.59 | 2678.00 |
| | Transformer | 4326.00 | 24.76 | 2754.00 |
| | Research Use Model | 4586.00 | 26.54 | 2448.00 |
| Dataset 2 | SMT | 4256.00 | 21.65 | 2546.00 |
| | RBMT | 4365.00 | 22.68 | 2635.00 |
| | NMT | 4686.00 | 23.54 | 2485.00 |
| | Transformer | 4758.00 | 24.85 | 2635.00 |
| | Research Use Model | 4962.00 | 25.67 | 2483.00 |

In Table II, in dataset 1, the model used in the study has the largest parameter data size of 4586, which is 932 higher compared to the lowest SMT model of 3654. The BLEU value of the model also has the highest value of 26.54, which is 3.96 higher compared to the lowest SMT model. In the run time comparison, the model used in the study exhibits a lower runtime data compared to several models, as long as 2448s. This suggests that the study's model outperforms other models in terms of overall performance. Additionally, comparing the data in dataset 2, the model employed in the study has parameter data that is 706 times greater than the SMT model, a BLEU value that is 4.02 times higher, and the model with the shortest run time, 2483s, than the SMT model. This shows that the model used in the study has a better performance than the traditional models.

## B. *Effectiveness of Machine Translation Modeling in Practice*

To test the effectiveness of the current research using the model in the application of English translation in multiple scenarios, English data from different scenarios are used, and data from three different scenarios such as legal application scenarios, news application scenarios, and speaking application scenarios are selected. 20,000 of these data are sampled and analyzed. The translation recognition accuracies of several traditional models in different scenarios in the previous subsection are compared and obtained as shown in Fig. 11.

In Fig. 11(a), the change in accuracy of several models used by the research in the legal application scenario increases gradually with the increase in the amount of data, after which it tends to a relatively stable state. The accuracy rate of the research-used model is higher among the five models, with the highest value of 95.6%, which is 3.1% higher relative to the smallest model, RBMT, at 92.5%. In Fig. 11(b), the accuracy of the models varies similarly to Fig. 11(a), with the research use model having a higher accuracy relative to the other models. The accuracy of the research use model is 95.7% higher than the accuracy of the SMT model 93.1% by about 2.6%. Similarly in Fig. 11(c) the accuracy of the research use model 96.5% is higher than the accuracy of the SMT model 93.2% about 3.3%. This shows that the research used model has higher accuracy and better modeling. Many scenarios are modeled and analyzed to produce the model test results displayed in Table III in order to assess the application effect of the research usage model in various English settings.

From Table III, in several models used, the study uses models in different scenarios with better MT effect higher indicators. Among them, the highest indicator value in the scenario of engineering analysis has 27.95, which is 6.90 higher compared to the lowest SMT model. Looking at different scenarios, the MT effect of the model in different scenarios varies, and some scenarios show higher indicator values, which may be due to the fact that the model in that scenario is more suitable for MT. To examine the MT effect of the currently used model in different scenarios, the translation indexes of several scenarios were compared as shown in Fig. 12.

In Fig. 12, the model's BLEU metrics changes in different scenarios all increase with the number of scenario samples, which may be due to the fact that more data samples can improve the model's translation efficiency during the training process. However, the news scene has the lowest metric value among several scene models, which may be due to the fact that the news scene contains more sentences and words about the emotional expression of the scene, which is more challenging for the MT. The model used in the study is able to translate ELs from different scenes and at the same time can achieve better translation results.

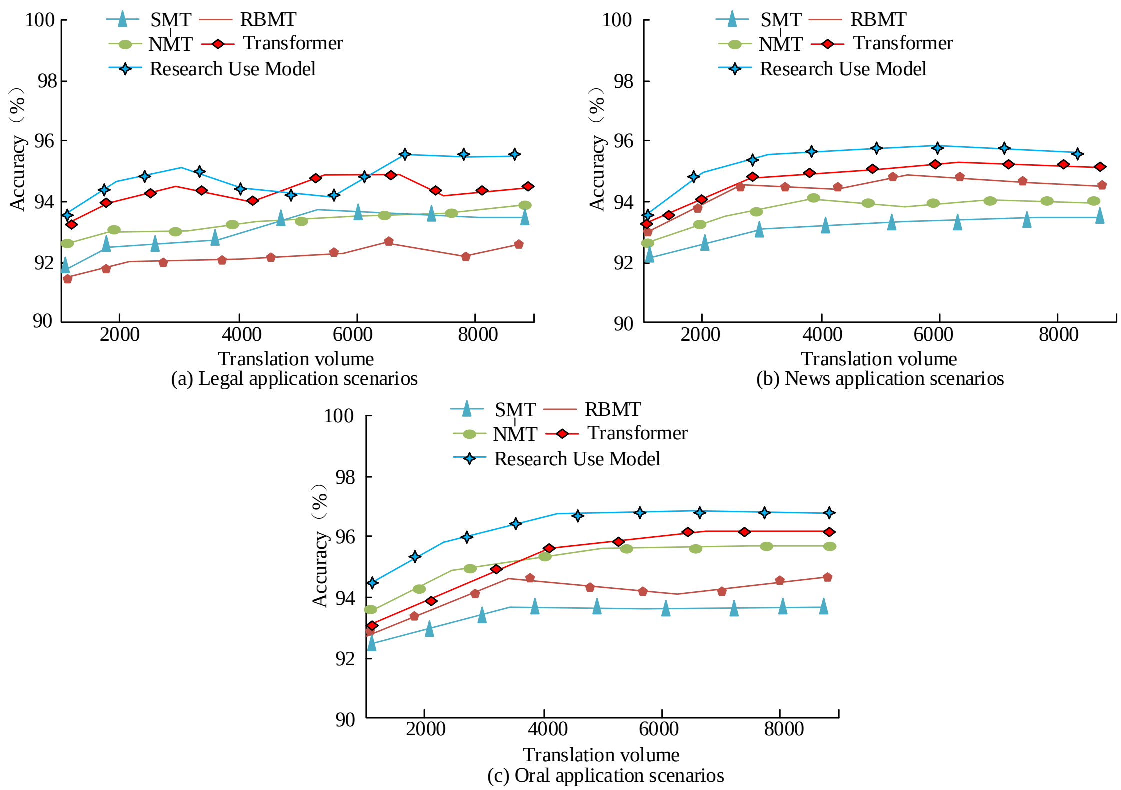


Fig. 11. Comparison of application accuracy in different model scenarios.

TABLE III. COMPARISON OF MACHINE TRANSLATION METRICS FOR DIFFERENT SCENARIO MODELS

| Scene | SMT | RBMT | NMT | Transformer | Research Use Model |
|---|---|---|---|---|---|
| Law | 21.65 | 22.65 | 24.65 | 24.68 | 26.48 |
| News | 22.54 | 23.54 | 24.84 | 24.68 | 27.65 |
| Spoken language | 21.85 | 23.48 | 25.03 | 25.36 | 27.30 |
| Organism | 22.64 | 22.68 | 25.16 | 24.79 | 26.54 |
| Analysis | 21.05 | 23.75 | 24.26 | 25.67 | 27.95 |

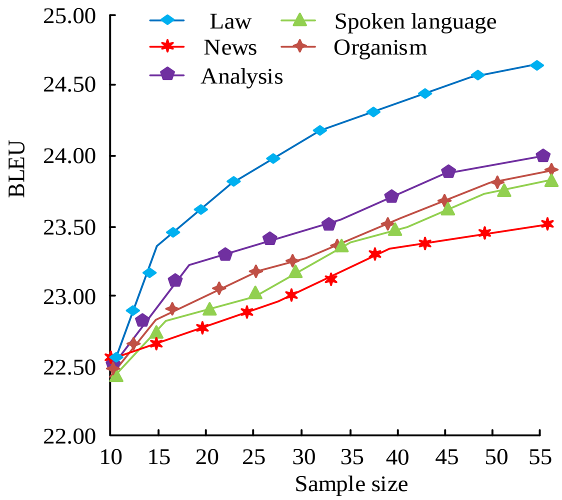


Fig. 12. Machine translation metrics for different scenarios of the model.

## V. Discussion

In the bilingual evaluation scoring test, the actual running effect of the module is best when the ratio of convolutional and attention layers is 4:5 and 5:4, which may be due to the reason that adding more convolutional modules and attention layers can improve the running effect of the model. In the analysis of different attention layer parameters, the more attention layers the better the model runs, which may be due to the fact that adding more attention layers can improve the data processing effect and running efficiency of the model. However, the increase of attention layers also causes the model running speed to decrease, which indicates that the addition of attention layers needs to guarantee the running speed of the model. In the comparison of the different feedforward layers of the model, the higher the number of feedforward layers, the lower the running speed, but at the same time, the higher the running score of the model. This may be due to the fact that the addition of feedforward layers improves the model running effect but reduces the model running speed. Therefore, the study needs to select the appropriate run feedforward layer. In the comparison of the number of layers used in the model, the more the number of layers used in the model the larger the data parameters of the model, which may be due to the fact that the increase in the number of layers used enhances the loading of the model with the used parameters. Also using only a single network model is not effective in enhancing the efficiency of the model, this may be due to the reason that a single module is not effective in enhancing the model. The research use model has the highest BLEU value of 29.65, which indicates the reason that the research use model enhances the model effectiveness with the addition of different modules. In the comparison of the models of different methods, the research use model has the largest data parameter content of 4586, the highest BLEU value of 26.54, and the shortest running time of only 2448 s. This may be due to the fact that the research use model adds more modules to enhance the model running efficiency and parameter loading. This is similar to the results obtained by Shao M et al. [20]. In the practical application, the running time of the research use model can reach up to 95.6%, which may be due to the fact that the research use model adds the attention mechanism to improve the accuracy of the model running. In the application of different scenarios, the research use model was able to run in different scenarios and the model's run-up scores were all at high values, which may be due to the addition of the meta-learning module.

In summary, the research use model showed better model capabilities in model run effectiveness, run time and translation scores. This indicates that the model used in the study has a good prospect for practical application in English translation in multiple scenarios. However, the research can also consider integrating multimodal information, such as pictures and videos, in the model to combine visual and textual information to improve the accuracy of translation and semantic understanding. Meanwhile, on the basis of English multi-scene translation, the model's multi-language translation ability can be further investigated. And the existing models may have the problem of context incoherence when dealing with long text translation. The introduction of a global context mechanism can be explored to enhance the coherence and accuracy of the model in long text translation, thus improving the user experience. Therefore the use of the model in English translation may have good practical application ability and can provide better research value for English translation in multiple scenarios. It also has a better guiding significance for English translation and English learning in multiple scenes.

## VI. Conclusion

The study suggests a new model based on MT and primarily addressed the issue of multi-scene application in English translation today. To enhance the translation effect of the model, the new model used BERT and CRF to extract words from the English scene's starting words. The results of the study showed that the 4:5 and 5:4 models had the highest BLEU mean values of 23.48 and 23.67, respectively. This was 0.81 and 1.00 higher than the lowest model, the 1:8 mean value of 22.67. The shortest run time was the 5:4 ratio at this point in time with a run time mean value of 22065s and the highest run time was the 2:7 run time mean value of 23587s. The number of attention layers where the model works better was 3. The run time for this layer was 22548s and the BLEU value was at a higher value. The highest mean BLEU value for the model used in the study was 29.65. The largest parameter data size for the model used in the study was 4586, which was 932 higher compared to the lowest SMT model 3654. The BLEU value for the model used in the study was also the highest at 26.54, which was 3.96 higher compared to the lowest SMT model. The research use model had a higher accuracy among the five models, with the highest value of 95.6%, which is 3.1% higher compared to the smallest model RBMT with 92.5%. It can be concluded that the model used in the study is able to translate EL for different scenarios very well, and its performance is also better than the traditional MT model. Although the study has achieved a lot of results, there are still many problems, such as the data used in the study is relatively small, more and larger data will be analyzed in the future, and different decoders will be added in the subsequent study to achieve further improvement of the model. The decoder may affect the broad applicability of the system in diverse scenarios when the system is dealing with more complex syntactic structures or rare words. Therefore subsequent research will explore the applicability of the system. Finally the research system may also suffer from computational complexity leading to high deployment and maintenance costs of the system in certain application scenarios. Therefore subsequent research will further reduce the system operating costs.

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

## RESEARCH INVOLVING HUMAN PARTICIPANTS AND / OR ANIMALS

Not applicable.

## INFORMED CONSENT

The author agreed that the paper could be published.

## REFERENCES


[1] Bacca-Acosta J., Fabregat R., Baldiris S., Kinshuk. "Determinants of student performance with mobile-based assessment systems for English as a foreign language courses," J. Comput. Assist. Learn., vol. 38, no. 3, pp. 797-810, 2022, DOI: 10.1111/jcal.12783.

[2] Chen J., Moradi H., Widodo H. P., Wood A., Gupta D. "ASIAN ENGLISH LANGUAGE CLASSROOMS: WHERE THEORY AND PRACTICE MEET," Appl. Linguist., vol. 44, no. 3, pp. 592-595, 2020, DOI: 10.1093/applin/amaa040.

[3] Gonzalez F. A., Elgeti S., Behr M., Auricchio F. "A deforming-mesh finite-element approach applied to the large-translation and free-surface scenario of fused deposition modeling," Int. J. Numer. Methods Fluids, vol. 95, no. 2, pp. 334-351, 2023, DOI: 10.1002/fld.5151.

[4] P. Y. Hao, "Asymmetric Possibility and Necessity Regression by Twin Support Vector Networks," IEEE Transactions on Fuzzy Systems, vol. 29, no. 10, pp. 3028-3042, 2020, DOI: 10.1109/TFUZZ.2020.3011756.

[5] Maruf S., Saleh F., Haffari G. "A Survey on Document-level Neural Machine Translation: Methods and Evaluation," ACM Comput. Surv., vol. 54, no. 2, Article 36, 2021, DOI: 10.1145/3441691.

[6] Yang M., Liu S., Chen K., Zhang H., Zhao E. "A Hierarchical Clustering Approach to Fuzzy Semantic Representation of Rare Words in Neural Machine Translation," IEEE Trans. Fuzzy Syst., vol. 28, no. 5, pp. 992-1002, 2020, DOI: 10.1109/TFUZZ.2020.2969399.

[7] Dabre R., Chu C., Kunchukuttan A. "A Survey of Multilingual Neural Machine Translation," ACM Comput. Surv., vol. 53, no. 5, Article 38, 2020, DOI: 10.1145/3406095.

[8] Khaw L. L., Tan W. W. "Creating Contexts in Engineering Research Writing Using a Problem-Solution-Based Writing Model: Experience of Ph.D. Students," IEEE Trans. Prof. Commun., vol. 63, no. 2, pp. 155-171, 2020, DOI: 10.1109/TPC.2020.2988758.

[9] Ban H., Ning J. "Design of English Automatic Translation System Based on Machine Intelligent Translation and Secure Internet of Things," Mobile Inf. Syst., vol. 2021, no. 4, Article 9, 2021, DOI: 10.1155/2021/8670739.

[10] Bharti N., Joshi N., Mathur I., Katyayan P. "Quality-Based Ranking of Translation Outputs," IT Prof., vol. 22, no. 4, pp. 21-27, 2020, DOI: 10.1109/MITP.2020.2976009.

[11] Wu H. "Multimedia Interaction-Based Computer-Aided Translation Technology in Applied English Teaching," Mobile Inf. Syst., vol. 2021, no. 2, pp. 1-10, 2021, DOI: 10.1155/2021/5578476.

[12] Maimaiti M., Liu Y., Luan H., Sun M. "Data augmentation for low-resource languages NMT guided by constrained sampling," Int. J. Intell. Syst., vol. 37, no. 1, pp. 30-51, 2021, DOI: 10.1002/int.22616.

[13] Bommidi B. S., Teeparthi K., Kosana V. "Hybrid wind speed forecasting using ICEEMDAN and transformer model with novel loss function," Energy, vol. 265, no. 15, Article 120234, 2023, DOI: 10.1016/j.energy.2020.120234.

[14] Beck M. B. D. A. C., Beck D. A. C. "Efficient 3D Molecular Design with an E(3) Invariant Transformer VAE," J. Phys. Chem. A, vol. 127, no. 37, pp. 7844-7852, 2023, DOI: 10.1021/acs.jpca.3c04188.

[15] Nourizadeh H., Mosallanejad A., Setayeshnazar M. "Optimal placement of fixed series compensation and phase shifting transformer in the multi-year generation and transmission expansion planning problem at the pool-based market for maximizing social welfare and reducing the investment costs," IET Gener. Transm. Distrib., vol. 16, no. 15, pp. 2959-2976, 2022, DOI: 10.1049/gtd2.12488.

[16] Sima W., Peng D., Yang M., Sun P. "Reversible Wideband Hybrid Model of Two-Winding Transformer Including the Core Nonlinearity and EMTP Implementation," IEEE Trans. Ind. Electron., vol. 68, no. 4, pp. 3159-3169, 2021, DOI: 10.1109/TIE.2020.2977544.

[17] Yadav S., Mehta R. K. "Modelling of magnetostrictive vibration and acoustics in converter transformer," IET Electr. Power Appl., vol. 15, no. 3, pp. 332-347, 2021, DOI: 10.1049/elp2.12025.

[18] Li Z., Jiao Z., He A. "Knowledge-based artificial neural network for power transformer protection," IET Gener. Transm. Distrib., vol. 14, no. 24, pp. 5782-5791, 2020, DOI: 10.1049/iet-gtd.2020.0542.

[19] Nicolae P. M. T., Nicolae I. D. V. D., Nitu M. C., Nicolae M.-S. P. M. "Analysis and Experiments Concerning Surges Transferred Between Power Transformer Windings Due to Lightning Impulse," IEEE Trans. Electromagn. Compat., vol. 65, no. 5, pp. 1476-1483, 2023, DOI: 10.1109/TEMC.2023.3299630.

[20] Shao M., Qiao Y., Meng D., Wangmeng Z. "Uncertainty-guided hierarchical frequency domain Transformer for image restoration," Knowl.-Based Syst., vol. 263, no. 5, Article 110306, 2023, DOI: 10.1016/j.knosys.2023.110306.

[21] Lu G., Zheng D., Zhang Q., Zhang P. "Effects of Converter Harmonic Voltages on Transformer Insulation Ageing and an Online Monitoring Method for Interlayer Insulation," IEEE Trans. Power Electron., vol. 37, no. 3, pp. 3504-3514, 2022, DOI: 10.1109/TPEL.2021.3118020.

[22] Li H., Li C., Zheng A., Tang J. "MsKAT: Multi-Scale Knowledge-Aware Transformer for Vehicle Re-Identification," IEEE Trans. Intell. Transp. Syst., vol. 23, no. 10, pp. 19557-19568, 2022, DOI: 10.1109/TITS.2022.3166463.

[23] Sun Q., Li Y., Ma D., Zhang Y. "Model Predictive Direct Power Control of Three-port Solid-State Transformer for Hybrid AC/DC Zonal Microgrid Applications," IEEE Trans. Power Deliv., vol. 37, no. 1, pp. 528-538, 2021, DOI: 10.1109/TPWRD.2021.3064418.

[24] Conway T. "An Isolated Active Balancing and Monitoring System for Lithium Ion Battery Stacks Utilizing a Single Transformer Per Cell," IEEE Trans. Power Electron., vol. 36, no. 4, pp. 3727-3734, 2021, DOI: 10.1109/TPEL.2020.3024904.

[25] Qiu H., Wang S., Sun F., Wang Z., Zhang N. "Computational Investigations on the Four-Stage MA-Class Fast Linear Transformer Driver With Sharing Cavity Shell," IEEE Trans. Plasma Sci., vol. 49, no. 9, Article 1, 2021, DOI: 10.1109/TPS.2021.3093561.

[26] Guo Z., Yu R., Xu W., Feng X. "Design and Optimization of a 200-kW Medium-Frequency Transformer for Medium-Voltage SiC PV Inverters," IEEE Trans. Power Electron., vol. 36, no. 9, pp. 10548-10560, 2021, DOI: 10.1109/TPEL.2021.3059879.

[27] Medeiros R. P., Costa F. B., Silva K. M., Muro J. de J. C. "A Clarke-Wavelet-Based Time-Domain Power Transformer Differential Protection," IEEE Trans. Power Deliv., vol. 37, no. 1, pp. 317-328, 2021, DOI: 10.1109/TPWRD.2021.3059732.

[28] Kucka J., Dujic D. "Current Limiting in Overload Conditions of an LLC-Converter-Based DC Transformer," IEEE Trans. Power Electron., vol. 36, no. 9, pp. 10660-10672, 2021, DOI: 10.1109/TPEL.2021.3060106.

[29] Paul A. K. "Structured Protection Measures for Better Use of Nanocrystalline Cores in Air-Cooled Medium-Frequency Transformer for Induction Heating," IEEE Trans. Ind. Electron., vol. 68, no. 5, pp. 3898-3905, 2021, DOI: 10.1109/TIE.2020.2984978.

[30] Usman A. M., Abdullah M. K. "An Assessment of Building Energy Consumption Characteristics Using Analytical Energy and Carbon Footprint Assessment Model," Green Low-Carbon Econ., vol. 1, no. 1, pp. 28-40, 2023, DOI: 10.47852/bonviewGLCE3202545.