Journal of Network Intelligence ©2024 ISSN 2414-8105 (Online)
Taiwan Ubiquitous Information Volume **9**, Number **1**, February **2024**

A Comprehensive Survey and Prospect of Cross-Lingual Summarization Method Research

Jing-Dong Wang, Duo Chang*, Fan-Qi Meng
School of computer science
Northeast Electric Power University
No.169 Changchun Road, Jilin City, Jilin Province, 132012, China
707569380@qq.com, cd2283436411@163.com, 249925066@qq.com

Guangqiang Qu
Academic Affairs Office
Northeast Electric Power University
No.169 Changchun Road, Jilin City, Jilin Province, 132012, China
630665545@qq.com

*Corresponding author: Duo Chang

Received April 6, 2023, revised June 5, 2023, accepted August 3, 2023.

Abstract. *Cross-lingual summarization technology evolved from pipeline-based methods to today's end-to-end approaches, although the problem of erroneous propagation is greatly avoided, there are still problems such as unclear nature of cross–lingual summarization, insufficient translation and summarization unification capabilities, scarcity of large-scale high-quality and multi-type datasets, insufficient research and exploration of low-resource cross-lingual summarization, and lack of multi-angle evaluation indicators. Therefore, according to the development context of Cross-lingual summarization, we first briefly introduce the pipeline-based first-translation-to-summarization method and the first-summarization and post-translation method, and then focus on zero-shot learning, multi-task learning, knowledge distillation method, knowledge enhancement method, pre-training framework and cross-lingual summarization method based on compression ratio, and then sort out the research progress of end-to-end Cross-lingual summarization, as well as the research motivation and content of various methods, and conduct in-depth comparative analysis. At the same time, since most of the world's languages are low-resource, we emphasize and especially sort out the current status of low-resource Cross-lingual summarization research. Finally, we also introduce and analyze the dataset and evaluation indicators of Cross-lingual summarization. At the end we discussed the possible directions of future development and presented our own opinions. Through this comprehensive and in-depth survey, it is hoped that researchers interested in this field, especially in low-resource settings, will be helped to promote the further development of Cross-lingual summarization.*
Keywords: Cross-lingual summarization, low-resource, end-to-end approach, natural language processing

1. **Introduction.** In the context of global economic and cultural prosperity, a large amount of text content in different languages on the Internet (such as user reviews, news, blogs, novels, books, scientific papers, legal documents, biomedical literature, etc.) There is also a rapid increase in the availability of various types of materials in various languages. People who are proficient in a certain language not only rely on information from their own country, but often need global information, which means that the demand for

cross-lingual information processing is increasing dramatically. Users spend a lot of time looking for the information they are looking for, and the information they find contains a lot of repetitive or unimportant content, they can't even read and understand all the text content of the search results, and it is difficult for humans to manually summarize massive text data [1]. Therefore, it is extremely urgent and important to summarize and condense text information in an unfamiliar language that the user is not familiar with into a summary of the language that the user is proficient in. Cross-lingual summarization , as a branch of automatic text summarization , aims to generate summaries in target languages different from the source document language to help solve the above problems. It improves the user's information acquisition speed, saves a lot of manpower and material resources, and has high application and research value in information retrieval, public opinion analysis, content review, cross-border e-commerce and other fields.

Since the advent of text abstracts [2, 3] and machine translation [4] in the 50s of the 20th century, researchers have been working hard to improve abstract generation and machine translation techniques, and so far there are still a large number of researchers improving and perfecting these two tasks. Cross-lingual summarization technology requires both capabilities, so cross-lingual summarization is a promising and challenging task. Cross-lingual abstract research methods can be divided into pipeline-based methods and end-to-end methods. In the early days, due to the difficulty of the CLS model having two capabilities at the same time, the scarcity of the corpus and the limitations of the technology, the research of CLS was basically based on the pipeline method [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15], and researchers often used the existing summarization dataset or translation dataset for experiments, so the method idea of the pipeline is to decompose CLS into monolingual summarization and machine translation subtasks, and then according to the completion order of the subtasks. The pipeline-based approach can be further divided into translation before summarization and summarization before translation type. Although this method is more intuitive and easy to understand, avoiding the difficulty of directly performing CLS, the imperfection of the previous step machine translation (MT) or monolingual abstraction (MS) leads to unsatisfactory results in the latter step, that is, there is a serious problem of error propagation [16, 17]. Therefore, the researchers propose an end-to-end cross-lingual summarization method based on end-to-end [16, 18, 19, 20, 21, 22, 23, 24, 25], which can greatly avoid this problem. In recent years, with the rapid development of neural networks, especially with the popularity and efficiency of the Transformer [26] framework in natural language processing, tasks such as abstracting and translation and cross-lingual representation have achieved good performance [22, 27, 28, 29, 30, 31, 32, 33, 34], which has become a commonly used framework for cross-lingual summarization [16, 18, 19, 20, 21, 22, 23, 24, 25, 31, 32, 35, 36, 37], It also opens up new opportunities for cross-lingual summarization. Nevertheless, CLS still faces many problems, such as: unclear nature of cross-lingual summarization, insufficient translation and abstract unification capabilities, scarcity of large-scale high-quality multi-type datasets, insufficient exploration of low-resource cross-lingual abstracts research, lack of multi-angle evaluation indicators, etc., which urgently require us to comprehensively sort out the development context of CLS to help and promote the development of this field and give researchers new inspiration. Different from the cross-lingual summarization summarized by Wang et al. [38], we divide these research results in more detail according to the techniques generated by cross-lingual summarization, and elaborate the motivations, methods and results of the research, and compare and summarize the advantages and disadvantages in detail. At the same time, we also summarize the research progress of low-resource cross-lingual summarization for the first time, and finally introduce the data sets, evaluation indicators and future development trends in detail.

First of all, we give the definition of cross-lingual summarization, that is, interlingual abstracting is to produce a short, clear and logical summary written in another language containing the main idea in the input document, and clarify its research objectives. Then we analyzed and sorted out the technical development of CLS. In the pipeline method of translation before abstract, it is further divided into extraction method [5, 6, 7], compression method [8, 9, 15] and abstract method [10, 11] according to technical methods. There is less research on abstracting before translation, only Orăsan and Chiorean [12] and Wan et al. [13, 14] have explored it.

Due to the drawbacks of the pipeline method, it is not competent for CLS tasks in the real world [39], so in recent years, the end-to-end method has become more and more popular, and nearly 20 CLS-related articles have been published in top conferences such as ACL, AAAI, SIGIR, EMNLP, etc., gradually becoming the mainstream of current CLS research, and achieving optimistic results. Then, according to the learning type, the research status of end-to-end CLS method is analyzed by focusing on zero-shot learning [18, 19, 40], multi-task learning [16, 20, 21, 22], knowledge distillation method [32], knowledge augmentation method [24, 35], pre-training framework [25, 32, 33, 36] and cross-lingual summarization method based on compression ratio [34]. Since most of the world's languages are low-resource, we particularly emphasize and sort out the research development status and current achievable level of low-resource cross-lingual abstracts [21, 22, 23, 25, 33, 36], and predict the possible development paths in the future.

Secondly, we introduce and analyze the CLS dataset and evaluation indicators. Zhu et al. [16] and Wang et al. [41] have been widely used to construct large-scale CLS datasets using Internet resources using existing MS datasets or Ladhak et al. [42] and Perez-Beltrachini et al. [43], and have demonstrated their feasibility experimentally, laying a data foundation for the further development and research of CLS technology. In view of the shortcomings of the traditional evaluation index ROUGE [44], researchers have recently proposed new evaluation methods based on semantic similarity, such as BERTScore [45] and MoverScore [46]. Finally, we analyze the challenges faced by CLS and their development trends, and give possible future development directions.

Our contribution:

(1) We have divided these research results in more detail according to the techniques generated by cross-language abstracts, and elaborated the motivation, experimental methods and results of the research, and compared and summarized the advantages and disadvantages in detail, which is more conducive to the understanding and learning of researchers and beginners;

(2) At the same time, we summarize the research progress of low-resource cross-lingual abstracts for the first time to help scholars better understand the development status and development problems of low-resource cross-lingual abstracts;

(3) We have made a unique analysis of the challenges faced by CLS and its development trend, and given possible future development directions.

2. **Cross-Lingual Summarization Definitions.** The Cambridge English Dictionary explanatory summary is "A short, clear description that gives the main factsor ideas about something". In the field of computing, Maybury [47] defines automatic summarization as follows: "An effective summarizationt extracts the most important information from one (or more) sources, generating an abridged version of the original information for specific users and tasks." Radev et al. [48] define a summarization as follows: "A summarization can be loosely defined as a text produced from one or more texts that conveys important information in the original text and is not more than half of the original text, usually significantly less." Text here is used rather loosely and can refer to speech, multimedia

documents, hypertext, etc. The resulting abstract should be shorter than the input text and include the most important information from the input text [?]. cross-lingual research mainly maps two different languages into a shared space in order to carry out mutual knowledge transfer, so as to promote the research and development of cross-lingual sentiment analysis, cross-lingual summarization, cross-lingual retrieval and cross-lingual question answering [50]. Therefore, an interlanguage summary is a short, clear, logical summary written in another language that contains the main ideas in the input document. Huang et al. [51] measured the quality of summarizations from four aspects: information coverage, information meaning, information redundancy and discourse cohesion.

The formal definition of a cross-lingual summarization is that, in language A, given a collection of original documents $D^A = \left\{x_1^A, x_2^A, ..., x_m^A\right\}$, m represents the total number of words, the cross-lingual summarization system generates summaries from the original collection of documents in language B $S^B = \left\{y_1^B, y_2^B, ..., Y_n^B\right\}$, n represents the number of abstract words, where $m \gg n$. In CLS, the main goal of research is to learn a model that can generate a summary of target language B $S^B = \{y_i\}_{i=1}^n$ for a given collection of articles in source language A $D^A = \{x_i\}_{i=1}^m$, which can be formally expressed as :

$$p_\theta(S^B|D^A) = \prod_{i=1}^{n} p_\theta(y_t|D^A, y_{1:t-1}) \tag{1}$$

where θ represents the model parameters and $y_{1:t-1}$ represents the true summary of the part.

3. **cross-lingual summarization techniques and methods.**

With the continuous acceleration and prosperity of global informatization, the research value and application value of cross-lingual summarization are becoming more and more extensive, which has attracted extensive attention from scholars at home and abroad, and the boom of deep learning has brought new opportunities for cross-lingual summarization research. Cross-lingual summarization techniques are mainly divided into pipe-based (pipeline) methods and end-to-end-based methods. The traditional CLS approach is based on the pipeline paradigm, that is, first translate the original document into the target language, and then summarize the translated document [5, 6, 7, 8, 9, 10, 11] or first summarize the original document, and then translate the summary into the target language [12, 13], and adopt different strategies to incorporate bilingual features into the pipeline model. Although the pipeline-based method is intuitive and straightforward, it has problems of error propagation [16, 17] and content duplication, which is not suitable for the application of practical scenarios. Judging from the research and publications of cross-lingual abstracts in recent years, most of the current researchers are interested in end-to-end methods, that is, aiming to train an end-to-end model for CLS, which can greatly avoid the problem of error propagation accumulation [52].

3.1. **Pipeline-Based Approach.**

Since CLS can be understood to some extent as a combination of machine translation (MT) and monolingual summarization (MS), the main idea is to break down CLS into monolingual abstracts and machine translation subtasks, and then complete them step by step. Depending on the order in which the subtasks are completed, these methods can be further divided into translation before summary and summary before translation type. For each type, the following is systematically detailed, comparing both types to provide a comprehensive and in-depth analysis.

3.1.1. *Translation-Summarization.*

The pipeline-based CLS method of Translation-Summarization can be further divided into three categories, namely the extraction method, the compression method and the abstract method. The extraction method is to select one or more complete sentences from the translation as a summary; The compression method is to first extract key sentences from the translation, and then remove irrelevant or informational words from the key sentences to obtain a summary; The abstract method is based on the content of the article to learn by yourself to generate new sentences as abstracts, not limited to the original words and phrases.

(1) Extraction method: Leuski et al. [5] constructed a cross-lingual information transfer system C*ST*RD, which translates Hindi documents into English through a statistical MT model, and then extracts important English sentences to form abstracts. The abstracter mainly relies on the translated document, so the accuracy of the machine translation result has a great impact on the final abstract, which may lead to defects in the abstract, and the system lacks semantic information to consider both sides of the language.

In 2011 , Wan [6] considered that some previous studies only used one language information, resulting in low reliability of abstracts, so two graph-based summarization methods (SimFusion and CoRank) were proposed for English-Chinese cross-lingual summarization, using both English-Chinese information in English-Chinese cross-lingual summarization. SimFusion method makes full use of the information of both sides through linear fusion of Chinese and English similarity, and then measures the significance score of Chinese sentences based on the framework of the graph, uses the greedy algorithm to punish sentences with high overlap with other high-scoring sentences, and finally selects prominent and novel Chinese sentences as concluding sentences. The CoRank method adopts a joint ranking algorithm, combined with the interaction between English and Chinese sentences, sorts the two sentences at the same time, and after obtaining the significance score of the Chinese sentence, the same greedy algorithm is used to redundantly remove, and finally several top sentences are selected as the concluding sentences. The experimental results show that the Chinese information is more beneficial than the English information, the Chinese information and the British information complement each other, and the proposed CoRank method is more reliable and robust than the proposed SimFusion method.

Later, Boudin et al. [7] proposed a graph-based multi-document cross-lingual summarization method, focusing on English and French multi-document summarizations, the main motivation of which is to enable French readers to learn more about the news through English news resources. Considering that the machine translation (MT) system may not translate fluently or incomprehensibly, it translates the document from English to French, and then uses the ϵ-Support vector regression method (ϵ-SVR) to predict the translation quality of each sentence based on bilingual features. Then, based on the quality of the translation, key translated sentences are selected based on Google's PageRank sorting algorithm. Finally, redundant sentences are removed from the selected sentences to form a final summary and guarantee maximum temporal coherence. The summary process is shown in Figure 1.

(2) Compression method: Yao et al. [8] were inspired by phrase-based machine translation (Phrase-based MT) and proposed a model of phrase-based compressed summarization that can score, extract, and compress sentences simultaneously. In detail, this study is not limited to producing extractive abstracts, but can be used to perform joint sentence selection and compression by using scoring schemes for cross-lingual document summarization tasks designed based on phrase alignment information, and remove redundant and poorly translated phrases. Unlike typical sentence compression methods, this model does not require additional syntactic preprocessing, such as part-of-speech tagging or syntactic analysis, and only utilizes information from the translated text with phrase alignment. Finally, an efficient greedy algorithm is designed to obtain an approximate solution to generate an abstract.

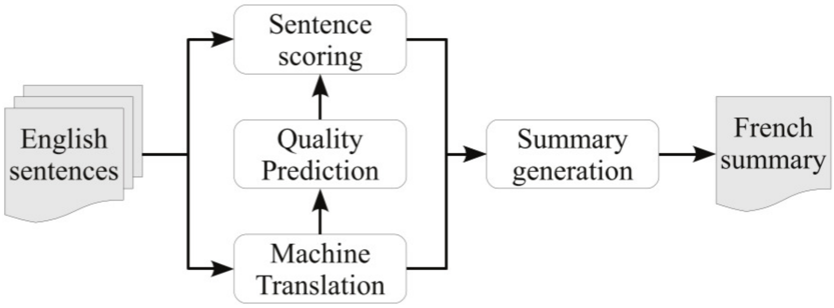

FIGURE 1. Architecture of the summarization system proposed by Boudin et al. [7]

Since the CoRank method [7] considers the similarity between the original text and the translation and its effectiveness and reliability, Linhares Pontes et al. [9] improve it according to this method and propose a French-to-English cross-lingual summarization framework. In sentence similarity measurement, the block is considered first rather than just the word, and then the sentence is compressed, using sentence and multi-sentence compression methods to retain the main information. At the same time, it establishes a long short-term memory (LSTM) model to analyze sentences and decide which words should remain in compressed sentences. Integer linear programming (ILP) formulas are also used to compress similar sentences while analyzing grammar and amount of information.

Building on Linhares Pontes et al. [9] , Pontes et al. [15] expanded on their previous work in three main ways: (1) They improved the previous sentence compression (SC) model and added an attention mechanism to the previous long short-term memory (LSTM) model to identify the main points of sentences. The model compresses sentences by removing irrelevant words. (2) They tweaked the Multiple Sentence Compression (MSC) method to simplify the integer linear programming (ILP) formula to optimize complexity, focusing only on the cohesion of word and keyword lists. (3) They conduct automated and manual evaluations to compare the quality of our compression cross-lingual approach with state-of-the-art extraction methods. They used MSC's system not only to generate more informative cross-lingual summaries, but also to obtain grammatical scores similar to extractive cross-lingual summaries. Unfortunately, adding attention mechanisms to sentence compression methods is not enough to produce smooth compression for long and complex source sentences. The low performance of the SC method also reduces the quality of SC+MSC compression.

(3) Abstract method: In order to solve the problem that previous summary models cannot make full use of similar sentences with complementary information and the knowledge of translation models is not comprehensively studied, Zhang et al. [10] propose an abstract English-Chinese cross-lingual summarization framework. First, a machine translation system is used to translate the source language document into the target language. The method then constructs a pool of bilingual concepts and facts represented by source-side Predicate-Argument Structures (PAS) and their target counterparts. Finally, bilingual PAS elements are fused with integer linear programming (ILP) algorithms to generate new concluding sentences to maximize the prominence and translation quality of PAS elements.

Ouyang et al. [11] argue that most of the world's languages are low-resource and have no digest corpus available, which leads to the only possible cross-lingual summarization method for translating and summarizing later. To this end, they developed an abstract summarizer based on the Pointer-Generator Network [53], proposed A method for generating cross-lingual summarization systems for low-resource languages that currently do not have an abstract corpus offers a potential text summarization solution for thousands of such languages. Specifically: on the New York Times document/abstract pair annotation corpus, a digest corpus was created using neural machine translation for Somali, Kiswahili, and Tagalog documents. They then trained noisy English input documents with clean English reference abstracts using a translation before abstract, and they also evaluated the low-resource summarization system on Arabic, a fourth source language, demonstrating that it could produce more fluent English abstracts from automatically translated documents.

3.1.2. *Summarization-Translation.* Orăsan and Chiorean [12] proposed a cross-lingual , multi-document digest that uses a maximum marginal correlation (MMR) algorithm to generate summaries from Romanian-language news articles, which are then automatically translated into English by eTranslator, a free Romanian-to-English translation engine. Due to the poor quality of machine translation at the time, the effect of abstracts was not very satisfactory.

It is precisely in the face of the unsatisfactory machine translation service that Wan et al. [13] proposed to consider the translation quality of each sentence in the English-Chinese cross-lingual summarization process. Specifically: firstly, the support vector machine regression method (SVM) [54] is used to predict the translation quality of each English sentence in the document set, and then the MT quality score of each predicted sentence is incorporated into the sentence evaluation process, and finally the English sentences with high translation quality and large amount of information are selected for translation to form a Chinese abstract.

3.1.3. *cross-lingual document summarization based on multiple summarization extraction and sorting.* Considering that the overall quality of the abstract is determined by many different factors, not only depends on the ranking results of the sentence, including a large number of part-level, sentence-level, and summary-level factors, and the summary model is not universal, different models produce different abstract quality. To this end, Wan et al. [14] propose a framework to solve the cross-lingual document summarization task by extracting and sorting multiple abstracts in the target language, as shown in Figure 2. Specifically: First, different summarization models (including digest-to-translate and post-translation methods) are used to extract multiple candidate abstracts for each document set, and several strategies are proposed to make the candidate abstracts of each document set contain some high-quality abstracts. Then, based on the candidate abstracts, they further propose a top-K integrated sorting method that makes full use of multiple characteristics at different levels to sort the candidate abstracts of each document set, and finally the top abstract is used as the final cross-lingual summarization of each document set.

3.1.4. *Discussion.* Many of the more effective methods basically belong to the extraction and compression class. They differ in how the similarity of sentences is calculated, and how the introduction summary that reduces the impact of translation errors is achieved. Summarization models that translate before summarizing often benefit from bilingual documents and make better use of bilingual information, while summarizing before translating methods tend to only use source language documents. Therefore, Translation-Summarization method usually has better results and performance than Summarization-Translation method. Wan et al. [13] believe that Summarization-Translation is a better

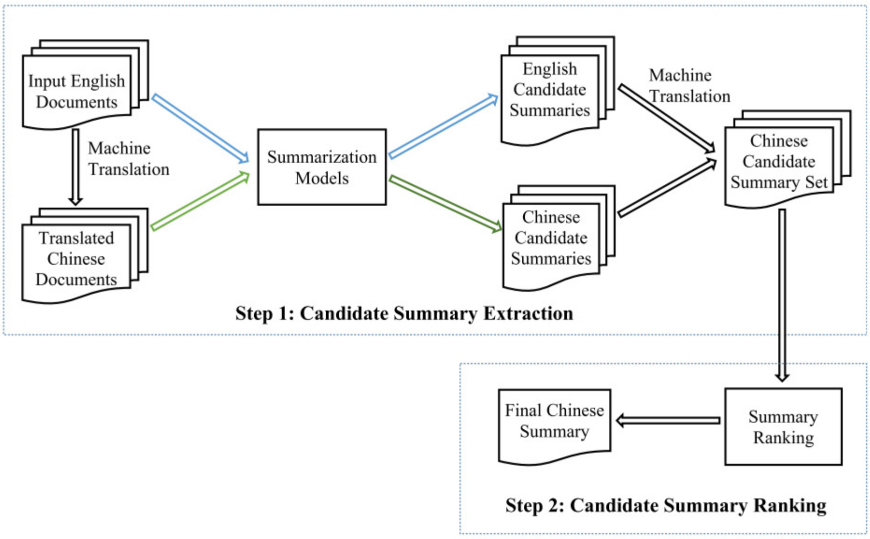

FIGURE 2. A cross-lingual summarization framework proposed by Wan et al.[14]

method, because it avoids both the computational overhead of translating more sentences and the error of sentence extraction caused by incorrect translation. Therefore, the method of Translation-Summarization has the disadvantage of low efficiency. However, the method of Summarization-Translation itself has the problem of poor effect or inability to do anything when the source language resources are scarce and the low-resource language is a low-resource language, and the cross-lingual summary dataset that can be used is inherently scarce in low-resource languages, so if the source language is low-resource, then the method of Translation-Summarization is of great help to it. At the same time, if the target language is a low-resource language, Summarization-Translation method is more suitable for it.

3.2. **End-to-End Based Approach.** The pipeline-based cross-lingual summarization method was introduced earlier, and it can be seen that although it is easy to understand, the imperfection of the previous step machine translation (MT) or monolingual summarization (MS) leads to unsatisfactory results in the latter step, that is, there is a serious problem of error propagation [16, 17]. An end-to-end-to-end cross-lingual summarization approach avoids this problem. With the popularity and efficiency of the Transformer [26] framework in natural language processing, it has become a commonly adopted framework for cross-lingual summarization, and it has also brought new opportunities for cross-lingual summarization. At the same time, the end-to-end cross-lingual summary model also has certain challenges in directly learning the ability to translate and summarize [58]. Based on the occurrence time, categories, and connections between previous end-to-end CLS models, the following will be introduced in the order of zero-shot learning, multi-task learning, knowledge distillation methods, knowledge augmentation methods, pre-training frameworks, and cross-lingual summarization methods based on compression ratio, and the advantages and disadvantages of each framework will be discussed at the end.

3.2.1. *Zero-shot learning framework.* Zero-shot learning (ZSL) [55] is one of the important cutting-edge research branches of academia. Its purpose is to solve the classification problem of unknown classes (test classes), that is, during the training process of the model, these unknown classes are not visible, and there are no relevant labeled training samples. Simply put, it is to use existing relevant data to train a model to recognize or learn new things that have not been seen. This solves the problem of lack of datasets in related fields. A comparison of the studies was analyzed in Table 1.

Ayana et al. [40] proposed an end-to-end cross-lingual neural title generation model CNHG for cross-lingual title generation due to the lack of cross-lingual summary datasets at that time and the problems of model differences and error propagation in pipeline-based methods. Based on zero-sample learning, the English-Chinese cross-lingual title generation task is trained on the basis of the existing parallel corpus (English title generation, Chinese title generation and English-Chinese translation corpus). The specific framework is shown in Figure 3. Specifically, it utilizes the teacher-student framework to solve cross-lingual title generation, that is, to teach the student model (CNHG model) with a pre-trained teacher model (translation or title generation model). The first teacher model is a pre-trained English-Chinese neural machine translation model (NMT model), the second teacher model is a pre-trained Chinese title generation model (NHG model) on the pseudo-Chinese article corpus established by the NMT model, and the third teacher model is a combination of NMT model and NHG model (NMT+NHG model). They used these three teacher models to teach the student model CNHG, in which all the encoders used bidirectional GRU recurrent neural networks, the encoders used attention-based GRU recurrent neural networks, and the Adadelta algorithm was used to optimize the model parameters. Finally, the evaluation data set was established with DUC2003 and DUC2004 datasets to test them. Experimental results show that the proposed method is a great improvement over the baseline method.

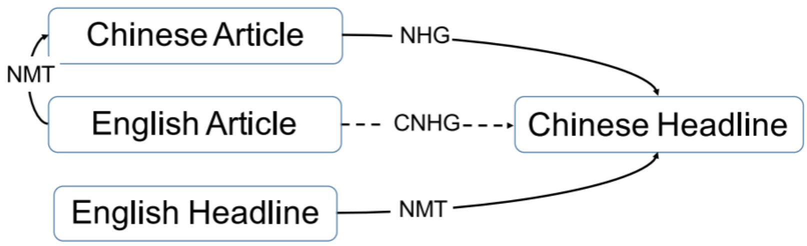

Figure 3. Ayana et al. proposed a model overview of the CNHG method [40]

Similarly, in the face of the lack of cross-lingual summarization datasets, compared with Ayana et al. [40], Duan et al. [18] applied reverse translation attention pseudo-source on large-scale monolingual digest datasets to form pseudo-source-true abstract pairs, rather than true source-pseudo-abstract pairs, and proposed a teacher-student framework based on Transformer [26] to train the cross-lingual abstract sentence summarization system ASSUM. As shown in Figure 4, Ayana et al. [40] and Duan et al. [18] propose the difference between training frameworks. The purpose of this teacher model is to teach students how to generate summary words under a suitable distribution by using cross-entropy loss to encourage similarities between the two distributions, and to encourage consistency by using cross-entropy loss to encourage similarities between two distributions, and by using the Euclidean distance between the attention weights of the two models as a loss, to encourage their consistency. Experimental results show that compared with some baseline and Ayana et al. [40] methods, it is greatly improved, and the performance gap between cross-lingual sentence summarization and monolingual sentence digest on the benchmark dataset can be significantly narrowed.

Unlike Ayana et al. [40] and Duan et al. [18] above to synthesize pseudo-training data through machine translation (MT), Dou et al. [19] proposed a zero-sample cross-lingual digest deep reinforcement model based on bilingual semantic similarity reward.

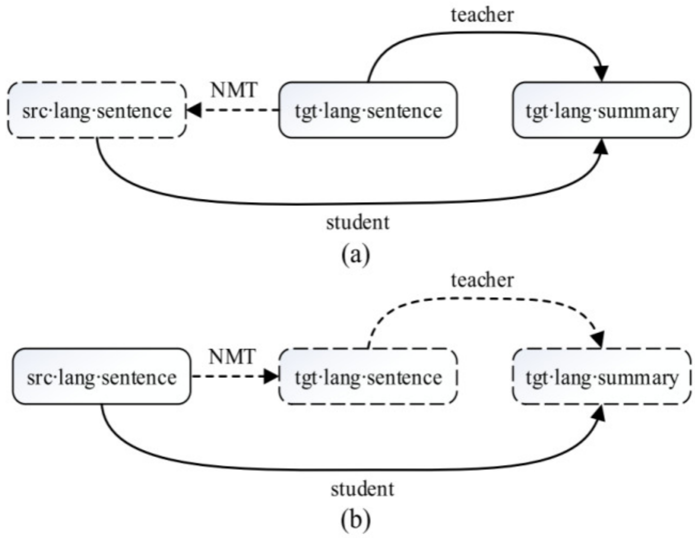

FIGURE 4. A comparison diagram of two frameworks by Duan et al. (a) represents the overall framework of Duan et al. and (b) represents the framework of Ayana et al. [18]

Specifically, the multi-task learning framework of this model with reference to Zhu et al. [16] consists of an encoder and two decoders, which first uses MT and CLS tasks to pre-train the model, and introduces a extractive monolingual summary model acting on the encoder to predict the probability that each sentence or keyword in the input article is related to the abstract. For the MT pre-training task, they used the WMT2014 English-German dataset and the WMT2017 Chinese-English dataset, and for the CLS pre-training task, they used the En2ZhSum dataset [16] and the English-German dataset constructed by Dou et al. [19] from the Gigaword dataset [56] Based on this, they use the reinforcement learning computational model to generate cross-lingual similarity between cross-lingual digests and cross-lingual digests annotated by source humans, and reward the encoder and the decoder responsible for the CLS task to fine-tune the model. Evaluated on the CLS test sets in both English and English, the model using reinforcement learning is better than the model before fine-tuning, and it can be more accurate, fluent, and relevant than the summary generated by the baseline.

3.2.2. *Multi-task learning framework.* Multi-task learning [57], as the name suggests, is to put multiple related tasks together and learn multiple tasks at the same time. Since cross-lingual summarization can be understood to a certain extent as the combination of machine translation (MT) and monolingual abstracts (MS), most researchers resort to combining the two related tasks of MT and MS with CLS tasks to train together, that is, through multi-task learning to solve the difficulty of generating cross-lingual summarization. A comparison of the studies was analyzed in Table 2.

Zhu et al. [16] proposed for the first time an end-to-end CLS framework called Neural Cross-Lingual Summarization (NCLS), and the basic structure is implemented entirely on the Transformer [26] framework. To further improve NCLS performance, they adopted a one-to-many multi-task learning approach that integrates MT and MS into the CLS training process, as shown in Figure 5. It can be seen from the figure that both CLS+MS and CLS+MT consist of a shared encoder and two decoders. The training data is that they propose large-scale Chinese-English and English-Chinese abstract datasets En2ZhSum [16] and Zh2EnSum [16] according to the bidirectional translation strategy. Experimental

Table 1. Comparison table of zero-shot learning studies

Author or model	Framework	Dataset	Baseline	Evaluation
Ayana et al. [40] CNHG	GRU	DUC2003 DUC2004	Baseline-TS Baseline-ST Baseline-PSEUDO	ROUGE
Duan et al. [12] ASSUM	Transformer	Gigaword DUC2004	Pipeline-TS Pipeline-ST Pseudo-Summary Pivot-based	ROUGFE
Dou et al. [19]	Transformer	En2ZhSum English-German	Pipeline-TS Pipeline-ST MLE-XLS+MT[16] MLE-XLS+MT+DIS[19] RL-ROUGE[19]	ROUGFE

results show that compared with the most advanced pipeline-based methods, NCLS can achieve +4.87 ROUGE-2 on En2ZhSum and +5.07 ROUGE-2 on Zh2EnSum. In addition, NCLS with multi-task learning can further achieve +3.60 ROUGE-2 on En2ZhSum and +0.72 ROUGE-2 on Zh2EnSum [16]. Their research method can serve as a baseline for CLS.

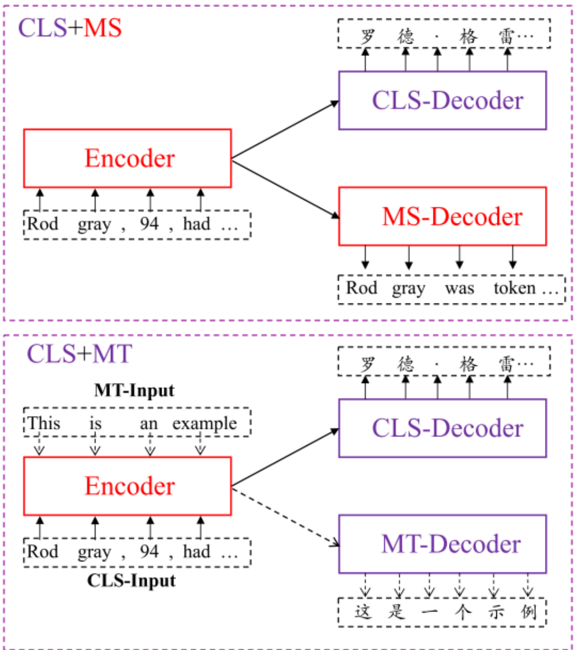

Figure 5. Multi-task learning NCLS framework by Zhu et al. [16]

Cao et al. [20] proposed a multitasking framework for joint learning summarization and alignment of context-level representations. Specifically, the framework consists of two encoders, two decoders, two linear mappers, and two discriminators. The method first integrates a monolingual summary model and a cross-lingual summary model into a unified model, and then builds two linear mappings that use cross-lingual word alignment techniques to project the context representation from one language to another. Among them, the adversarial training [58] technique is used, and two discriminators are used to make the mapper generative learning and generate better maps. The discriminator is responsible for distinguishing between coded notation and mapping notation. The model generates a cross-lingual summarization (taking the English-Chinese summarization as an example) by first using an English encoder, then using a mapper to map the English representation to Chinese space, and finally using a Chinese decoder to generate a Chinese abstract. At the same time, unsupervised training methods and supervised training methods are proposed. The unsupervised training method shows that the model can generate cross-lingual summarization even without a cross-lingual corpus, but it is still far behind the supervised training method. At the same time, it is shown that cross-lingual representation is beneficial for cross-lingual summarization. Specific cross-lingual word embedding and cross-lingual representation techniques can be found in the literature [59, 60, 61, 62, 63] discussion.

Bai et al. [21], faced with the lack and high cost of large parallel datasets that are low-resource and cross-lingual in most languages, proposed a multi-tasking framework MCLAS for cross-lingual summarization abstraction in low-resource environment, and explored and studied cross-lingual summarizationion in low-resource environment. Unlike Zhu et al. [16], the framework has shared parameters for the decoder, using a unified encoder and decoder. Previous work, due to the independence of the decoder, the abstract knowledge cannot be well transmitted to cross-lingual summarization for knowledge sharing, especially in the face of summary knowledge transfer from high-resource languages to low-resource languages. As a result, MCLAS uses a unified decoder to sequentially connect MS and CLS generation, making MT a prerequisite for CLS generation, so that shared decoders can learn interactions involving cross-lingual alignment and summarization patterns, facilitating knowledge transfer. MCLAS is also based on the Transformer [26] framework. Experiments on the En2ZhSum[16], Zh2EnSum[16] and their self-constructed En2DeSum [21] datasets show that the MCLAS model is significantly better than the baseline model of Zhu et al. [16] in both low-resource and full-dataset scenarios. At the same time, the interpretability of the multi-task structure proposed in this study is visually analyzed, which proves the effectiveness of the learning mode of the MCLAS model.

Liang et al. [22] considered that the previous study did not fully consider the hierarchical relationship between MT, MS and CLS, so they tried to use the Conditional Variational Auto-Encoder (CVAE) [64] to simulate the hierarchical relationship between MT, MS and CLS, and proposed a variational hierarchical model VHM that uses both translation and abstracting. CVAE is superior in learning hierarchies with hierarchical latent variables [65, 66, 67]. Specifically, VHM learns hierarchical relationships between MT&MS and CLS using CVAE-based hierarchical hidden variables, which consists of a priori network and an identification (posteriori) network, the latter is responsible for guiding the learning of the prior network by Kullback-Leibler (KL) divergence [68], with the aim of learning two local variables of global variables in CLS for translation and summarization. These two local variables are constrained to reconstruct the translation and source language digest, and then the two local variables are explicitly leveraged using global variables to obtain a better CLS. The CLS task was fine-tuned using mBART [69] as model initialization, and after experiments with En2ZhSum [16] and Zh2EnSum [16], it significantly outperformed most of the previous state-of-the-art methods Zhu et al. [16] and Cao et al. [20]. At the same time, the framework also takes into account the limited CLS data in low-resource languages, and still achieves better performance than existing methods.

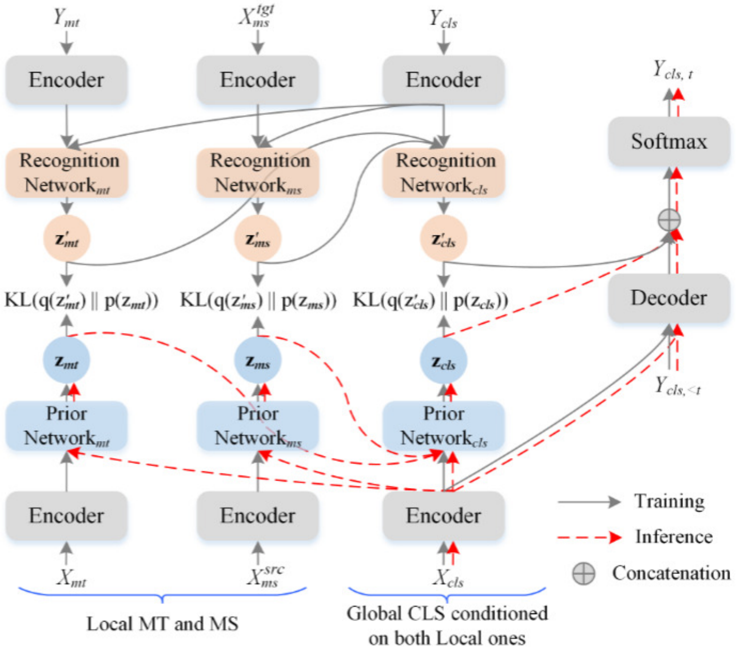

FIGURE 6. VHM framework by Liang et al. [22]

TABLE 2. Sample data

Author or model	Framework	Dataset	Baseline	Evaluation
Zhu et al. [16] TNCLS CLS+MS CLS+MT	Transformer	En2ZhSum Zh2EnSum	TETran GETran TLTran GLTran	ROUGE
Cao et al. [54]	Transformer	Gigaword DUC2004 En2ZhSum Zh2EnSum	Unsupervised:Unified Unified+CLWE Supervised:Pipe-TS Pipe-TS* Pipe-ST Pipe-ST* Pseudo XLM Pretraining	ROUGE
Bai et al. [21] MCLAS	Transformer	En2ZhSum Zh2EnSum En2DeSum	TNCLS CLS+MS	ROUGE BERTScore
Liang et al. [22] VHM	Transformer	En2ZhSum Zh2EnSum	GETran [16] GLTran [16] TNCLS ATS-A [24] CLS+MS CLS+MT MS-CLS-Rec [20]	ROUGE MoverScore

3.2.3. *Knowledge Distillation Framework.* Knowledge distillation, which has attracted more and more attention from the research community in recent years, has become a research hotspot and focus in the field of deep learning. At first, in order to solve complex problems and improve the training effect of the model, the network structure of the model was gradually designed deep and complex, which was difficult to adapt to the needs of mobile computing development for low resources and low power consumption. As a result, the knowledge distillation method was proposed, and knowledge distillation was originally used as a learning paradigm to transfer knowledge from a large teacher model to a shallow student model and improve performance. However, with the development of knowledge distillation, its teacher-student architecture (Teacher-Student model) as a special transfer learning method, has evolved a rich variety of variants and architectures. Its main idea is to transfer the "knowledge" in the complex teacher model with strong learning ability to the simple student model, that is, the student model imitates the teacher model, and the teacher model transfers the learned knowledge to the student model, so that it has a strong learning generalization ability [70, 71, 72]. A comparison of the studies was analyzed in Table 3.

The above-introduced Ayana et al. [40] and Duan et al. [18] also use the teacher-student architecture in the framework of their model, and in order to show their handling of zero-sample learning, they are introduced in the zero-sample learning framework. On this basis, Nguyen et al. [23] proposed a cross-lingual abstract generation framework based on knowledge extraction in view of the inefficiency of the current CLS multi-tasking framework in capturing key cross-lingual representations between languages, especially between low-resource and long-distance languages (languages are far apart in structure and morphology) [73, 74]. Specifically, it is also a teacher-student framework that inherits the architecture of the Transformer [26] model. Build cross-lingual relevance by extracting knowledge from the monolingual summary teacher model onto a student model of cross-lingual digests. At the same time, since the implicit representations of the teacher and student models exist on two completely different vector spaces (because they represent two different languages), they propose a knowledge distillation loss based on Sinkhorn divergence [75, 76] between the teacher-student framework to evaluate the difference between teacher and student representations. Due to the intuitive geometric nature of Sinkhorn divergence, it has been shown to be beneficial for cross-lingual and multilingual representation learning, making alignment between distant languages easier [77]. Experiments on the cross-lingual abstract digest datasets En2ZhSum [16] and Zh2EnSum [16] in two different languages show that the performance of this method is better than the existing abstract model at that time, whether in the case of high or low resources. The effectiveness of this framework was also evaluated on the distant languages English-to-Arabic (EN2AR), English-to-Japanese (EN2JA), Japanese-to-English (JA2EN), and English-to-Vietnamese (EN2VI) (preprocessed by the Wikilingua dataset [39]). It is worth emphasizing that following the low-resource setting of Bai et al. [21], the experimental results of this model show that the quality of cross-lingual digest generation is better than that of the MCLAS model.

Table 3. Knowledge Distillation Method Research Comparison Table

Author or model	Framework	Dataset	Baseline	Evaluation
Nguyen et al. [23]	Transformer	En2ZhSum Zh2EnSum WikiLingua	TLTran [21] TNCLS CLS+MS MCLAS	ROUGE

3.2.4. *Knowledge Enhancement Framework.* The Knowledge Enhancement Framework, as the name suggests, is to enhance the quality of CLS generation by utilizing external or internal knowledge that is not limited to keywords, topics, linguistic features, knowledge bases, etc. A comparison of the studies was analyzed in Table 4.

Zhu et al. [24] focus on the use of some words in the source language document from the perspective of translation mode, and use the Fast-Align Tool [78] to extract word alignment from source-destination and target-source directions on the bilingual parallel LDC corpus for machine translation, forming a probabilistic bilingual dictionary to assist in the generation of abstracts. Specifically, the structure of Transformer [26] is first used to focus on the neural probability distribution of the source word (some specific words), and then according to the three translation strategies they propose, combined with the translation distribution of the probabilistic bilingual dictionary, the final word probability is selected from the final candidate translations. It leverages both internal and external knowledge from CLS datasets (probabilistic bilingual dictionaries) to enhance cross-lingual summarization effects. Compared with the existing method Zhu et al. [16], the model ATS proposed by Zhu et al. [24] has a smaller model size and faster training speed, specifically, it only adopts an additional probabilistic bilingual dictionary instead of a large-scale parallel machine translation dataset, which greatly reduces the dependence of the model on data.

Inspired by the thinking steps of professionals writing cross-lingual abstracts (usually selecting and translating keywords, named entities, etc. as key clues for abstracts, and then composing abstracts based on understanding the text based on these key clues), Jiang et al. [35] propose a clue-guided cross-lingual abstract abstractor Cluegraphsum. This model, Cluegraphsum, consists of a graphics encoder, clue encoder and decoder, and is mainly based on the Transformer [26] structure. Specifically, TextRank[79] is first used to extract the keywords and named entities of the sentences in the input article as its key clues. Then, according to these clues, an article diagram with less noise sentences is constructed to capture the relationship and importance between sentences; Finally, the article graph and key clue distribution are fed into the graphics encoder and clue encoder, and finally the decoder generates a summary of the target language. At the same time, the Cluegraphsum model also uses the Naive strategy proposed by Zhu et al. [24] to select candidate translation words for some keywords, and play a role in abstract. In doing so, the model uses internal knowledge to enhance the effectiveness of the summary in the form of graphs and key clues, especially if the document is long. Experimental results show that the abstract score of this method is significantly improved compared with other models [16, 24].

3.2.5. *Pre-Training framework.* The pre-trained model is an application of transfer learning, using a large-scale corpus from the open domain to learn the context-sensitive representation of each member of the input sentence, implicitly learning the general grammatical semantic knowledge, and then transferring it to downstream tasks to improve low-resource tasks, which is also very beneficial for low-resource language processing. And the pre-trained model + fine-tuning mechanism has good scalability and can be adapted to specific new tasks. A comparison of the studies was analyzed in Table 5.

Faced with the problem of the relative inadequacy of natural language processing (NLP) research on cross-lingual pre-training models, that is, the insufficient ability of pre-trained models to represent different languages in cross-lingual shared space, Chi et al. [36] proposed a cross-lingual pre-training model XNLG based on Transformer [26] sequence-to-sequence model. It pre-trains monolingual and cross-lingual targets, and then based on several fine-tuning strategies to make XNLG capable of solving specific problems, and

Table 4. Comparative table of research on knowledge enhancement methods

Author or model	Framework	Dataset	Baseline	Evaluation
Zhu et al. [24] ATS-NE ATS-A	Transformer	En2ZhSum Zh2EnSum	GETran [16] GLTran [16] TNCLS CLS+MS CLS+MT	ROUGE MoverScore
Jiang et al. [35] Cluegraphsum	Transformer	En2ZhSum Zh2EnSum CyEn2ZhSum [35]	Pipe-TS Pipe-ST TNCLS ATS-NE ATS-A	ROUGE MoverScore METEOR

evaluates XNLG on zero-sample language generation, cross-lingual digest and problem generation. Dou et al. [19] proposed a zero-sample cross-lingual digest deep reinforcement model based on bilingual semantic similarity reward, which first used MT and CLS tasks to pre-train the model, which was introduced above. Xu et al. [25] also share encoders and decoders for pre-training and task optimization on the architecture of Transformer [26], but their model is pre-trained on single-language tasks (including mask language model (MLM), denoising autoencoder (DAE) and monolingual summary (MS), as well as cross-lingual tasks (such as cross-lingual mask language model (CMLM) and machine translation (MT)) to achieve the use of a large amount of unlabeled monolingual data to enhance the model's language modeling capabilities. On this basis, the downstream specific tasks are fine-tuned and optimized, and CLS experiments are carried out on the dataset proposed by Zhu et al. [16]. Compared with Chi et al.'s XNLG [36], CLS has made great progress.

Due to the recent general-purpose multilingual pre-training model in NLP tasks [69, 80, 81, 82], and effectively improve cross-lingual portability [27, 28, 29]. For example, the multilingual pre-training model mBART [69] is pre-trained on a large number of unlabeled multilingual data, inheriting the denoising and other characteristics of the BART [30] model. Like Liang et al. [22], they fine-tune CLS tasks using mBART [69] as model initialization, surpassing the initial model and many types of CLS models on CLS tasks. The mT5 [31] is a multilingual pre-trained text-to-text converter that inherits the advantages of the T5 [83] model, which is pre-trained in 101 languages. Chi et al. [32] improved the problem of mT5 model in the lack of utilization of translation data, and proposed MT6, a pre-trained model for improving multilingual text-to-text with translated data. Specifically, it explores three cross-lingual text-to-text pre-training tasks, namely Machine Translation (MT), Translation Pair Span Corruption (TPSC), and Translation Span Corruption TSC. At the same time, it also proposes a partially non-autoregressive target (PNAT). Finally, these methods were pre-trained on multiple multilingual benchmark datasets (CC-Net [84], Multiun [85], IIT Bombay [86], OPUS, and Wikimatrix [87]), including a total of 94 languages. Evaluated on sentence classification, named entity recognition, question answering, and abstract summarization tasks. Experimental results show that MT6 improves cross-lingual portability compared to MT5. Considering that the natural language generation (NLG) task is based on the encoder decoder structure, and the pre-trained encoder can only partially benefit it, Ma et al. [33], unlike previous work such as BART [30] and T5 [83], regard the decoder as a task layer of ready-made pre-trained encoders, and propose a pre-trained multilingual encoder-decoder model

DeltaLM (ΔLM), whose encoder and decoder are initialized by pre-trained multilingual encoders. and train in a self-supervised manner. DeltaLM used parameters from Chi et al. [29] InfoXLM and trained on Span Corruption (SC) and Translation Span Corruption (TSC) tasks on CC100 [80], CC-Net, Wikipedia Dump, CCAlligend [88], and Opus Corora, including 100 languages. It was also evaluated on cross-lingual summarization tasks. Instead of relying only on shared vocabulary and bilingual contexts to encourage cross-lingual correlation, Luo et al. [37] inserted cross-note [89, 90]modules in the Transformer encoder, explicitly constructed interdependencies between languages, proposed a cross-lingual model VECO, and conducted experiments on 9 cross-lingual understanding tasks on the XTREME benchmark platform. Includes text classification, sequence labeling, question answering, and sentence retrieval.

Table 5. Pre-training method research comparison table

Author or model	Framework	Dataset	Baseline	Evaluation
Xu et al. [25]	Transformer	En2ZhSum Zh2EnSum	TETran GETran TLTran GLTran TNCLS CLS+MS CLS+MT XNLG [36] ATS	ROUGE
Chi et al. [32] MT6	Transformer	WikiLingua	MT5	ROUGE
Ma et al. [33] DeltaLM(δLM)	Transformer	WikiLingua	mBART MT5	ROUGE

3.2.6. *cross-lingual summarization based on compression ratio.* In the face of the current CLS task, people's research is basically on the existing machine translation or monolingual summary dataset with more advanced methods to generate cross-lingual summary corpora, based on multi-tasking, knowledge-based augmentation, based on teacher-student framework or pre-training based methods, lack of deeper exploration, and further application of large-scale machine translation corpus. To this end, Bai et al. [34] propose a new cross-lingual summarization task-compression ratio cross-lingual summarization (CSC). By introducing the compression ratio, which is the ratio of information between the source text and the target text, the MT task is treated as a special CLS task with a compression rate of 100%. In this way, cross-lingual Compressed Digest (CSC) unifies MT and CLS. However, there is a huge gap between MT tasks and CLS tasks, and samples with compression ratios between 30%-90% are extremely rare. Therefore, in order to smoothly connect these two tasks, an effective data augmentation method is proposed (given a well-annotated CLS data sample, iteratively remove less important sentences and words to shorten the source document, and generate a document-digest pair with a larger compression ratio.) to produce document-digest pairs with different compression ratios. This method not only improves the performance of CLS tasks, but also provides control over generating summaries of the required length. Experiments show that the proposed method outperforms various strong baselines on three cross-lingual summary datasets.

3.2.7. *Discussion.* By summarizing all existing end-to-end cross-lingual summarization CLS models, we find that with the emergence of large-scale cross-lingual summarization datasets, there are more and more research methods for CLS tasks, but they are basically based on the Transformer framework and using MT and MS tasks for research and experimentation. The zero-sample learning approach is basically done using the teacher-student framework; The multitasking framework relies on simultaneous training with MT, MS and CLS; The knowledge augmentation framework uses knowledge inside or outside the dataset for enhancement, but when the knowledge is extracted incorrectly, it will cause the impact of mispropagation; The pre-training framework can benefit from unlabeled corpus and labeled corpus, specifically, the pre-trained model learns common language knowledge from large-scale unlabeled data with self-supervised goals, and often achieves good results on CLS tasks. The emergence of the new research mission CSC also made possible the unification of the CLS mission and the MT mission. Faced with the problem of low-resource settings and low-resource languages, multi-task learning, knowledge distillation frameworks for teachers and students, cross-lingual representation, and pre-training methods can provide research ideas.

4. **Low-resource cross-lingual summarization research development.** Previous cross-lingual summarization studies on low-resource languages have been achieved using translation-first-digest methods on existing datasets, such as Leuski et al. [5], Orăsan and Chiorean [12] and Ouyang et al. [11]. However, there is an impact on the quality of machine translation on the performance of the abstract. Later, after the end-to-end model was proposed, although the problem of error propagation was effectively avoided, some researchers [18, 19, 40] carried out zero-sample learning by synthesizing pseudo-training data in the absence of datasets, which solved the problem of scarcity of corresponding datasets to a certain extent, but there was still an impact of translation inaccuracy on the abstract effect when using pseudo-datasets for training. Zhu et al. [16] used a round-trip translation strategy to construct a cross-lingual digest dataset by filtering out low-quality document summary pairs by ROUGE [44] on the existing CNN/Dailymail [91], MSMO [92] datasets and LCSTS [93] datasets. Since then, the problem of false propagation has been greatly avoided, and it has also stimulated a wave of research on CLS. Bai et al. [21] were the first researchers to propose CLS in a low-resource environment, due to the lack of low-resource language datasets, by setting up fewer resources on existing large-scale CLS datasets, as shown in Table 6. They also performed the same low-resource training and comparison on the baseline models of NCLS and NCLS+MS of Zhu et al. [16]. Nguyen et al. [23] also experimented with the low-resource setup of Bai et al. [21]. Table 7 and Table 8 show the summary performance scores of these four methods in a low-resource environment. It can be seen that achieving good abstract results in low-resource environments is a great challenge, requiring continuous innovation by researchers, and this field is still in its infancy. Liang et al. [22] also conducted low-resource experiments. Specifically, they randomly selected 0.1%, 1%, 10%, and 50% of CLS training datasets (Zh2EnSum, En2ZhSum) for experimentation, and found that when the CLS training data became less and less, the performance gap between the comparison model and VHM became larger and larger. This is due to more translation and summary data, which makes the impact of MT and MS greater, effectively enhancing the CLS model. Their specific research methods and contents are described above.

The above research is basically based on multi-task learning, which can borrow useful knowledge from other related tasks for target tasks to help learning tasks under low-resource conditions [94, 95]. Similarly, there are researchers who resort to pre-training methods. Chi et al. [36] fine-tune based on the proposed pre-training model XNLG, when

Table 6. Low resource setup table for Bai et al. [21].

Scene	Zh2EnSum	En2DeSum	En2ZhSum
Minimum	5,000 (0.3%)	2,619 (0.6%)	1,500 (0.4%)
Medium	25,000 (1.5%)	12,925(3.0%)	7,500 (2.0%)
Maximum	50,000 (3.0%)	25,832 (6.0%)	15,000 (4.0%)
Full	1,693,713	429,393	364,687

Table 7. Zhu et al. [16], Bai et al. [21] and Nguyen et al. [23] CLS evaluation results of Zh2EnSum in low-resource scenarios.

Dataset size Evaluation	Minimum R-1/R-2/R-L	Medium R-1/R-2/R-L	Maximum R-1/R-2/R-L	Full R-1/R-2/R-L
NCLS [16]	20.93/5.88/17.58	26.42/8.90/22.05	29.05/10.88/24.32	35.60/16.78/30.27
NCLS +MS [16]	20.50/5.45/17.25	26.86/9.06/22.47	28.63/10.63/24.00	34.84/16.05/29.47
MCLAS [21]	21.03/6.03/18.16	26.86/9.06/22.47	30.73/12.26/26.51	35.65/16.97/31.14
Nguyen et al.'s approach [23]	22.37/6.50/18.47	27.97/11.51/27.16	31.08/12.70/27.16	36.93/20.99/32.33

Table 8. Zhu et al. [16], Bai et al. [21] and Nguyen et al. [23] CLS assessment results of En2ZhSum in low-resource scenarios.

Dataset size Evaluation	Minimum R-1/R-2/R-L	Medium R-1/R-2/R-L	Maximum R-1/R-2/R-L	Full R-1/R-2/R-L
NCLS [16]	34.14/12.45/21.20	35.98/15.88/23.79	40.18/19.86/26.52	44.16/24.28/30.23
NCLS +MS [16]	33.96/12.38/21.07	38.95/18.09/25.39	39.86/19.87/26.64	42.68/23.51/29.24
MCLAS [21]	21.03/6.03/18.16	37.28/18.10/25.26	38.35/19.75/26.41	42.27/24.60/30.09
Nguyen et al.'s approach [23]	22.37/6.50/18.47	40.30/20.01/25.79	41.24/20.01/27.06	44.75/25.76/31.05

the target language is the same as the language of the training data, they fine-tune all parameters, and when the target language is different from the language of the training data, they fine-tune the parameters of the encoder to adapt to the downstream NLP task. In their summarization generation experiments, they trained only on English data in a zero-shot setting, and then directly evaluated the model on other languages and found that they were better than their baseline task, and the problem of error propagation was more serious when the two languages were far apart. Xu et al. [25] also fine-tune downstream specific tasks based on their pre-trained model, and they explore the low-resource scheme of CLS. Specifically, they extracted subsets of size 1K and 10K from the cross-lingual summarization training data proposed by Zhu et al. [16], and optimized the pre-training model on these subsets. The results show that the pre-trained model has a greater gain than the zero-based model on the same subset compared to the pre-trained model. Similarly, the pre-trained multilingual model DeltaLM (ΔLM) proposed by Ma et al. [33] also explores the zero-sample cross-lingual transfer capability of the pre-trained model. In short, in the case of zero sample or low resources, with the blessing of pre-training, the ability of CLS has been improved.

At present, due to the scarcity of CLS datasets in low-resource languages, the main research focus is basically on low-resource settings (few samples) or zero-samples based on pre-training. However, a recent systematic review of cue methods in natural language processing by Liu et al. [96] systematically investigated a new paradigm in natural language processing, called "prompt-based learning". It allows language models to be pre-trained on large amounts of raw text, and by defining a new prompt function, the model is able to learn with few or even zero samples to adapt to new scenarios with little or no labeled data, which were explored in low-resource settings in the field of text generation by Schick and Schütze [97], Li and Liang [98].

5. Cross-lingual summarization datasets.

5.1. En2ZhSum, Zh2EnSum, and En2DeSum. Zhu et al. [16] proposed a new round-trip translation strategy to obtain large-scale CLS datasets from existing large-scale MS datasets. They built a 370K English-Chinese translation corpus En2ZhSum and a 169M Chinese-English translation corpus Zh2EnSum. Specifically, cross-lingual document summary pairs were obtained from the existing monolingual summary dataset [99, 100, 101]. According to the method of Zhu et al. [16], Bai et al. [21] constructed the En2DeSum dataset.

5.2. WikiLingua. Ladhak et al. [42] proposed a new benchmark dataset, Wikilingua, which consists of gold-standard abstracts and collaboratively written how-to guides in 18 languages. Each article and abstract is written and edited by 23 people, with an average of 16 further reviewed, which ensures the high quality of the content. These articles describe a variety of methods and steps from different topic sets to accomplish a procedural task, such as "how to make cream coffee", "how to exercise to reduce back pain". Each step contains a sentence summary, followed by a paragraph detailing the instruction, and an image illustrating the given instruction. Nguyen et al. [32] preprocessed the dataset of Wikilingua and selected 4 variants of Wikilingua for evaluation, namely English to Arabic (EN2AR), English to Japanese (EN2JA), Japanese to English (JA2EN), and English to Vietnamese (EN2VI).

5.3. Long text cross-lingual summarization dataset Perseus. To facilitate the study of long-document CLS, Zheng et al. [101] constructed the first long-document CLS dataset Perseus, which collected about 94K Chinese of scientific literature and English abstracts. In total, the dataset covers four disciplines, including engineering applications, natural sciences, agricultural sciences, and medical sciences. The average length of its source file is 2872.9 Chinese characters. At the same time, in order to evaluate the versatility of the long document CLS model, it also provides an extraterritorial test set containing 500 sports field Chinese documents and their English summary pairs.

5.4. **XWikis dataset.** Perez-Beltrachini et al. [43] proposed a cross-lingual summary big dataset XWikis containing document-abstract pairs in four languages, which contains long documents in the source language and multiple sentence abstracts in the target language, including 12 language pairs and directions for the four European languages of Czech, English, French and German. And different summarization scenarios are supported: degree of supervision (supervision, zero and few shots), language combination (cross-lingual and multilingual), and language resources (high and low resources). Specifically, inspired by past research on monolingual and multilingual descriptive abstracts [102, 103, 104], they derived examples of cross-lingual document summaries from Wikipedia by combining citation paragraphs and article bodies from language-aligned Wikipedia titles.

5.5. **cross-lingual conversation summarization ClidSum dataset.** Wang et al. [41] proposed ClidSum, a benchmark dataset for building cross-lingual summarization systems on conversational documents. It consists of 67K+ dialogue documents and 112K+ annotated summaries in different target languages, and introduces two baseline settings for supervised and semi-supervised scenarios. At the same time, they introduced the cross-lingual Conversation Summary (XLDS) task, which aims to summarize the dialogue in the source language into different languages, which is the first large-scale XLDS benchmark dataset. Clidsum was generated based on two existing monolingual dialogue digest datasets, Samsum [105] and Mediasum [106], and hired professional translators to translate Samsum and Mediasum's original English abstracts into German and Chinese.

6. cross-lingual summarization evaluation methods.

6.1. **ROUGE.** ROUGE is an automatic abstract evaluation method proposed by Lin [44], which is widely used to evaluate the performance of automatic abstract models. The basic idea is to compare the system summary and reference summary produced by the model, and evaluate the quality of the system summary by counting the number of basic units that overlap between them. Commonly used evaluation indicators are ROUGE-1, ROUGE-2, ROUGE-L, etc., where 1, 2, L represent based on 1 unary word, 2 unary word and longest substring, respectively. This method is one of the general criteria for abstract evaluation systems, but this method can only evaluate the surface information of reference abstracts and system abstracts, and does not involve semantic evaluation. The calculation formula is

$$R_{ROUGE-N} = \frac{\sum_{S\in\{Ref\}}\sum_{N_{n-gram}\in S} Count_{match}(N_{n-gram})}{\sum_{S\in\{Ref\}}\sum_{N_{n-gram}\in S} Count(N_{n-gram})} \tag{2}$$

where $n-gram$ represents n-gram words and Ref represents reference summarization, $Count_{match}(N_{n-gram})$ represents the number of n-grams in both the system summarization and the reference summarization, $Count(N_{n-gram})$ indicates the number of n-grams appearing in the reference summarization. ROUGE also has 3 evaluation indicators: accuracy P (precision), recall R (recall) and F value. The ROUGE formula is derived from the formula for calculating recall. During the evaluation phase, researchers often use the toolkit pyrouge to calculate the ROUGE score of the model.

6.2. **MoverScore.** The ideal metric is to compare system output and references based on the semantics of the reference rather than the surface form. Zhao et al. [46] studied the encoding system and the strategy of reference text, and designed a measurement method MoverScore, which is highly correlated with human judgment of text quality. It is based on a combination of contextualized representations of the system and reference text and the distance between these representations, and the semantic distance between the system output and the reference. Finally, a single overall score is assigned to the system-generated text by matching it to human references. It is especially important that this measure not only captures the amount of content shared between two texts, but also accurately reflects how much the system's text deviates from the reference.

6.3. **BERTScore.** BERTScore, Zhang et al. [**?**]proposed a language generation evaluation index based on pre-trained BERT context embedding. Similar to common measures, BERTScore calculates the similarity score for each mark in the candidate sentence versus each mark in the reference sentence. Specifically, BERTScore calculates the similarity of two sentences as the sum of cosine similarities between their token embeddings, and uses context embeddings to calculate token similarity instead of exact matching. BertScore addresses two common flaws in n-gram-based metrics [107, 108]: first, such methods tend not to be a strong match with interpretations, and second, their inability to capture distant dependencies and penalize semantically critical order changes.

6.4. **Human evaluation.** Because the automatic evaluation method at this stage can only depict the surface relationship between sentences, and cannot distinguish the quality of the abstract through semantics, the emergence of manual evaluation makes up for the shortcomings of the automatic evaluation method to some extent. However, the manual evaluation method is greatly affected by factors such as mother tongue and education level, which is slightly subjective and too inefficient. Depending on the problem, the focus of manual evaluation is also different. Abstracts are usually artificially scored based on attributes such as readability, relevance to the original text, fluency, and whether grammatical restrictions are met.

(1) Readability. The abstract should be written fluently and the spelling should be correct.

(2) Correlation. The abstract should be closely related to the subject information of the original text and should not deviate from the original meaning.

(3) Informative. The abstract should contain most of the important information of the original text, and if little information is obtained from the abstract, then the abstract is likely to be substandard.

(4) Coherence. The logic and syntax of the abstract should be correct.

(5) Simplicity. The length of the summary should be as concise as possible, not too much duplication to promote other indicators, and as little redundant information as possible.

7. **Challenges and trends in cross-lingual summarization.** With the emergence of a large number of cross-lingual summary datasets, research methods are basically presented in an end-to-end manner, although they avoid the problem of mispropagation based on pipelined methods, but lack of targeted leapfrog progress, and breakthrough innovation work is needed to improve performance to improve the quality and performance of CLS tasks. The following section explores the issues and future directions for cross-lingual summarization:

(1) Datasets: Although large-scale CLS datasets have emerged and presented multi-typed, they are basically generated from existing monolingual abstracts and machine translation datasets, most of which are concentrated in a single field and lack of real and comprehensive diversity of large-scale CLS datasets. At the same time, CLS datasets in low-resource languages also urgently need to be built, because there are thousands of languages in the world, except for several general-purpose languages in the world, most of which are low-resources, and all the construction of high-quality, large-scale CLS datasets

requires further development by future researchers. This is of great benefit in helping to get a complete picture of the world.

(2) Method innovation: At present, CLS tasks are basically implemented based on the Transformer framework, although some researchers use sentence maps, keywords, bilingual dictionaries and other information or learn general knowledge based on pre-training to enhance CLS tasks, or researchers use MT and MS to multi-task learning CLS, but the full use of dataset knowledge and explore the relationship between the essence of CLS tasks and MT and MS still need to be further improved. Therefore, it is also necessary to inject more strength in the innovation of exploration methods. For example, the new cross-lingual abstracts task-compression ratio cross-lingual abstracts (CSC) proposed by Bai et al. [34] provides ideas for the unification of CLS and MT. and Chi et al. [32] who used multiple tasks to pre-train the model to learn its general knowledge, which also opened up a new path. At the same time, weak supervision or no supervision, zero or small sample learning development should also be given special attention. Due to the scarcity of high-quality datasets and the fact that most of the world's languages are low-resource languages, it is also worth exploring the direction of using efficient algorithms to process CLS tasks with a small amount of training data or no training data. For example, researchers currently learn the general knowledge of cross-lingual representations through large-scale pre-trained models, and then perform cross-lingual transfer learning.

(3) Content expansion: Similarly, we should also pay attention to the development of multimodal cross-lingual summarization and multi-document cross-lingual summarization. With the diversification of information forms on the Internet, researchers have explored multimodal monolingual abstracts, but ignore cross-lingual and multilingual scenarios, so the future multimodal cross-lingual or multilingual abstracts are a promising research direction. In monolingual abstracts research, multi-document abstracts have been explored very early, so cross-lingual multi-document abstracts are worth discussing.

(4) Evaluation indicators: Although the current evaluation indicators of MoverScore and BERTScore are based on the language understanding level and the comparison of target reference abstracts, there is a big gap compared with humans, so the research on the semantic level of CLS task evaluation indicators and the research on how to use other information to evaluate the quality of abstracts are conducive to promoting the rapid development of CLS.

8. **Conclusions.** This paper comprehensively introduces the research development and current status of cross-lingual summarization CLS, systematically classifies and introduces the existing methods, analyzes their research characteristics and motivations, and conducts comparative analysis. In particular, we sorted out the development of low-resource CLS, introduced a variety of CLS datasets and evaluation indicators, and looked forward to and analyzed the future development to promote the further development of CLS.

Acknowledgment. This work was supported by the Jilin Provincial Science and Technology Development Program (Name: Research on Large-scale Group Decision Optimization Algorithm for Incomplete Hesitation Fuzzy Information, Project No.: 20230101176JC) and the Jilin Provincial Science and Technology Development Program (Project Name: Research on Knowledge Graph Construction Technology for Electrical Primary Equipment Fault Diagnosis, No. 20220101129JC). The authors also thank the reviewers for their helpful comments and suggestions, which improved the presentation.

REFERENCES

[1] W.S. El-Kassas, C.R. Salama, A.A. Rafea, and H.K. Mohamed, "Automatic text summarization: A comprehensive survey," *Expert Systems with Applications*, vol. 165, 113679, 2021.

[2] H. P. Luhn, "The automatic creation of literature abstracts," *IBM Journal of Research and Development*, vol. 2, no. 2, pp. 159–165, 1958.

[3] M. Gambhir, V. Gupta,"Recent automatic text summarization techniques: a survey," *Artificial Intelligence Review*, vol. 47, no. 1, pp. 1–66, 2017.

[4] L. Han, "An overview on machine translation evaluation," *Expert Systems with Applications*, vol. 2202. 11027, 2022.

[5] A. Leuski, C. Lin, L. Zhou, U. Germann, F.-J. Och, and E.-H. Hovy, "Cross-lingual C*ST*RD: English access to Hindi information," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 2, no. 3, pp. 245–269, 2003.

[6] X.-J. Wan, "Using Bilingual Information for cross-lingual Document Summarization," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL 2011)*. 2011, pp. 1546–1555.

[7] F. Boudin, S. Huet, J.-M. Torres-Moreno, "A graph-based approach to cross-lingual multi-document summarization," *Polibits*, vol. 2011, no. 43, pp. 113–118, 2011.

[8] J. Yao, X. Wan, J. Xiao, "Phrase-based compressive cross-lingual summarization," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing(EMNLP)*. 2015, pp. 118–127.

[9] E.L. Pontes, S. Huet, J. Torres-Moreno, and A.C. Linhares, "cross-lingual Text Summarization Using Sentence and Multi-Sentence Compression," in *Natural Language Processing and Information Systems: 23rd International Conference on Applications of Natural Language to Information Systems, (NLDB 2018)*. Springer, 2018, pp. 467–479.

[10] J. Zhang, Y. Zhou, C. Zong, "Abstractive cross-lingual summarization via translation model enhanced predicate argument structure fusing," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 10, pp. 1842–1853, 2016.

[11] J. Ouyang, B. Song, K. McKeown, "A robust abstractive system for cross-lingual summarization," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies(ACL)*. 2019, pp. 2025–2031.

[12] C. Orăsan, and O.-A. Chiorean, "Evaluation of a cross-lingual Romanian-English multi-document summariser," in *Proceedings of the 6th International Conference on Language Resources and Evaluation(LREC)*. 2008, pp. 2114–2119.

[13] Wan X, Li H, Xiao J., "cross-lingual document summarization based on machine translation quality prediction," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics(ACL)*. 2010, pp. 917–926

[14] X. Wan, F. Luo, X. Sun, Huang, S., and J. Yao, "cross-lingual document summarization via extraction and ranking of multiple summaries," *Knowledge and Information Systems*, vol. 58, pp. 481–499, 2019.

[15] E.-L. Pontes, S. Huet, J.-M. Torres-Moreno, and A.-C. Linhares, "Compressive approaches for cross-lingual multi-document summarization," *Data & Knowledge Engineering*, vol. 125, 101763, 2020.

[16] J.-N. Zhu, Q. Wang, Y.-N. Wang, Y. Zhou, J.-J. Zhang, S.-N Wang, and C.-Q. Zong, "NCLS: Neural Cross-Lingual Summarization," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing(EMNLP-IJCNLP)*. 2019, pp. 3054–3064.

[17] Y. Cheng, Q. Yang, Y. Liu, M.-S. Sun, and W. Xu, "Joint training for pivot-based neural machine translation," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI'17)*. AAAI Press, 2017, pp. 3974–3980.

[18] X.-Y. Duan, M.-M. Yin, M. Zhang, B.-X. Chen, and W.-H. Luo, "Zero-Shot Cross-Lingual Abstractive Sentence Summarization through Teaching Generation and Attention," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics(ACL)*. 2019, pp. 3162–3172.

[19] Z.-Y. Dou, S. Kumar, and Y. Tsvetkov , "A Deep Reinforced Model for Zero-Shot Cross-Lingual Summarization with Bilingual Semantic Similarity Rewards," in *Proceedings of the Fourth Workshop on Neural Generation and Translation, Online. Association for Computational Linguistics*. 2020, pp. 60–68.

[20] Z.-Y. Dou, S. Kumar, and Y. Tsvetkov, "Multi-task sequence to sequence learning," in *Proceedings of the International Conference on Learning Representations (ICLR)*. 2016, 1511.06144.

[21] Y. Bai, Y. Gao, H. Huang, "Cross-Lingual Abstractive Summarization with Limited Parallel Resources," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL)*. 2021, pp. 6910–6924.

[22] Y. Liang, F. Meng, C. Zhou, J. Xu, Y. Chen, J. Su, and J. Zhou, " A Variational Hierarchical Model for Neural Cross-Lingual Summarization," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2022, pp. 2088–2099

[23] T.-T. Nguyen, A.-T. Luu, " Improving neural cross-lingual abstractive summarization via employing optimal transport distance for knowledge distillation," in *Proceedings of the AAAI Conference on Artificial Intelligence*. 2022, 36(10), pp. 11103–11111

[24] J. Zhu, Y. Zhou, J. Zhang, and C. Zong, " Attend, translate and summarize: An efficient method for neural cross-lingual summarization," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020, pp. 1309–1321.

[25] R. Xu, C. Zhu, Y. Shi, M. Zeng, and X. Huang, "Mixed-Lingual Pre-training for Cross-lingual Summarization," in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*. 2020, pp. 536–541.

[26] A. Vaswani, N.-M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.-N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is All you Need," *Advances in Neural Information Processing Systems*, vol. 1706, 03762, 2017.

[27] K. Karthikeyan, Z. Wang, S. Mayhew, and D. Roth, "Cross-lingual ability of multilingual bert: An empirical study," in *International Conference on Learning Representations.*. 2020.

[28] S. Wu, M. Dredze. Beto, Bentz, "Becas: The Surprising Cross-Lingual Effectiveness of BERT," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 2019, pp. 833–844.

[29] Z.-W. Chi, L. Dong, F.-R. Wei, N. Yang, S. Singhal, W.-H. Wang, X. Song, X.-L. Mao, H.-Y.Huang, and M. Zhou, "InfoXLM: An Information-Theoretic Framework for Cross-Lingual Language Model Pre-Training," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2021, pp. 3576–3588.

[30] M. Lewis, Y.-H. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020, pp. 7871–7880.

[31] L.-T. Xue, N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, and C. Raffel, "mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2021, pp. 483–498.

[32] Z.-W. Chi, L. Dong, S.-M. Ma, S.-H. Huang, S. Singhal, X.-L. Mao, H.-Y. Huang, X. Song, and F.-R. Wei, "mT6: Multilingual Pretrained Text-to-Text Transformer with Translation Pairs," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 2021, pp. 1671–1683.

[33] S. Ma, L. Dong, S. Huang, D. Zhang, A. Muzio, S. Singhal, H.-H. Awadalla, X. Song, and F. Wei, "Deltalm: Encoder-decoder pre-training for language generation and translation by augmenting pretrained multilingual encoders," *arXiv preprint*, vol. 2106, 13736, 2021.

[34] Y. Bai, H. Huang, K. Fan, Y. Gao, Y. Zhu, J. Zhan, Z. Chi, and B. Chen, "Unifying Cross-lingual Summarization and Machine Translation with Compression Rate," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2022, pp. 1087–1097.

[35] S. Jiang, D. Tu, X. Chen, R. Tang, W. Wang, and H. Wang, "ClueGraphSum: Let Key Clues Guide the Cross-Lingual Abstractive Summarization," *arXiv preprint*, vol. 2203, 02797, 2022.

[36] Z. Chi, L. Dong, F. Wei, W. Wang, X. Mao, and H. Huang, "Cross-Lingual Natural Language Generation via Pre-Training," in *Proceedings of the AAAI conference on artificial intelligence*. 2020, 34(05), pp. 7570–7577.

[37] F.-L. Luo, W. Wei, J.-H. Liu, Y.-J. Liu, B. Bi, S.-F. Huang, F. Huang, and L. Si, "VECO: Variable and Flexible Cross-lingual Pre-training for Language Understanding and Generation," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. 2021, pp. 3980–3994.

[38] J. Wang , F.-D. Meng, D. Zheng, Y.-L. Liang, Z.-X. Li, J.-F. Qu, and J. Zhou, " A survey on cross-lingual summarization," *Transactions of the Association for Computational Linguistics*, vol. 10, pp. 1304-1323, 2022.
[39] F. Ladhak, E. Durmus, C. Cardie, and K. McKeown, "WikiLingua: A New Benchmark Dataset for Cross-Lingual Abstractive Summarization," in *Findings of the Association for Computational Linguistics(EMNLP)*. 2020, pp. 4034–4048.
[40] Ayana, S.-Q. Shen, Y. Chen, C. Yang, Z.-Y. Liu and M.-S. Sun, "Zero-Shot Cross-Lingual Neural Headline Generation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. vol. 26, no. 12, pp. 2319-2327
[41] J. Wang, F. Meng, Z. Lu, D. Zheng, Z. Li, J. Qu, and J. Zhou, "ClidSum: A Benchmark Dataset for Cross-Lingual Dialogue Summarization," *arXiv preprint*, vol. 2202, 05599, 2022.
[42] F. Ladhak, E. Durmus, C. Cardie, and K. McKeown, "WikiLingua: A New Benchmark Dataset for Cross-Lingual Abstractive Summarization," in *Findings of the Association for Computational Linguistics(EMNLP)*. 2020, pp. 4034–4048.
[43] L. Perez-Beltrachini, M. Lapata, "Models and Datasets for Cross-Lingual Summarisation," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021, pp. 9408–9423.
[44] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," in *Text summarization Branches Out.*, 2004, pp. 74–81.
[45] T. Zhang, V. Kishore, F. Wu, K.-Q. Weinberger, and Y. Artzi, "Bertscore: Evaluating text generation with bert," *arXiv preprint*, vol. 1904, 09675, 2019.
[46] W. Zhao, M. Peyrard, F. Liu, Y. Gao, M. Christian M., and S. Eger., "MoverScore: Text Generation Evaluating with Contextualized Embeddings and Earth Mover Distance," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 563–578.
[47] M.-T. Maybury, "Generating summaries from event data," *Information Processing & Management*, vol. 31, no. 5, pp. 735–751, 1995.
[48] D. Radev, E. Hovy, K. McKeown, "Introduction to the special issue on summarization," *Computational linguistics*, vol. 28, no. 4, pp. 399–408, 2002.
[49] M. Gambhir, V. Gupta, "Recent automatic text summarization techniques: a survey," *Artificial Intelligence Review*, vol. 47, no. 1, pp. 1–66, 2017.
[50] A. Søgaard, I. Vulić, S. Ruder, et al., "Cross-lingual word embeddings," *Synthesis Lectures on Human Language Technologies*, vol. 12, no. 2, pp. 1–132, 2019.
[51] L. Huang, Y. He, F. Wei and W. Li, "Modeling Document Summarization as Multi-objective Optimization," in *2010 Third International Symposium on Intelligent Information Technology and Security Informatics*, 2010, pp. 382–386.
[52] M.-M. Yin, "Cross-lingual sentence summarization research," Master. dissertation, Soochow University, 2020
[53] A. See, P.-J. Liu, C.-D. Manning, "Get to the point: Summarization with pointer-generator networks," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 2017, pp. 1073–1083.
[54] C. Cortes, V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995.
[55] Y. Xian, C.-H. Lampert, B. Schiele and Z. Akata, "Zero-Shot Learning—A Comprehensive Evaluation of the Good, the Bad and the Ugly," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 9, pp. 2251–2265, 2019.
[56] A.-M. Rush, S. Chopra, J.-A. Weston, "Neural Attention Model for Abstractive Sentence Summarization," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015, pp. 379–389.
[57] M. Luong, Q.-V. Le, I. Sutskever, O. Vinyals, and L. Kaiser, "Multi-task sequence to sequence learning," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2016.
[58] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta and A.-A. Bharath, "Generative Adversarial Networks: An Overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53-65, 2018.
[59] C. Xing, D. Wang, C. Liu, and Y.-Y. Lin, "Normalized word embedding and orthogonal transform for bilingual word translation," in *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2015, pp. 1006–1011.

[60] M. Artetxe, G. Labaka, E. Agirre, "A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* , 2018, pp. 789–798.

[61] X. Peng, C. Lin, M. Stevenson, "Cross-Lingual Word Embedding Refinement by $\ell 1$ Norm Optimisation Norm Optimisation," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2021, pp. 2690–2701.

[62] Y.-X. Wang, Y. Hou, W. Che, and T. Liu, "From static to dynamic word representations: a survey," *International Journal of Machine Learning and Cybernetics*, vol. 11, pp. 1611-1630, 2020.

[63] H. Yang, H. Chen, H. Zhou, and L. Li, "Enhancing cross-lingual transfer by manifold mixup," *arXiv preprint*, vol. 2205, 04182, 2022.

[64] K. Sohn, X. Yan, H. Lee, "Learning structured output representation using deep conditional generative models," in *Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2015, pp. 3483–3491.

[65] I.-V. Serban, A. Sordoni, R. Lowe, L. Charlin, J. Pineau, A.-C. Courville, and Y. Bengio, "A hierarchical latent variable encoder-decoder model for generating dialogues," in *Proceedings of the 31th AAAI Conference on Artificial Intelligence*, 2017, pp. 3295–3301.

[66] L. Shen, Y. Feng, H. Zhan, "Modeling Semantic Relationship in Multi-turn Conversations with Hierarchical Latent Variables," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 5497–5502.

[67] L. Shen, F.-D. Meng, J.-C. Zhang, Y. Feng, and J. Zhou, "GTM: A Generative Triple-wise Model for Conversational Question Generation," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 2021, pp. 3495–3506.

[68] D.-P. Kingma, M. Welling, "Auto-encoding variational bayes," *arXiv preprint*, vol. 1312, 6114, 2013.

[69] Y.-H. Liu, J.-T. Gu, N.-M. Goyal, X. Li, S. Edunov, M. Ghazvininejad, M. Lewis, and L. Zettlemoyer, "Multilingual denoising pre-training for neural machine translation," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 726–742, 2020.

[70] G. Hinton, O. Vinyals, J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint*, vol. 1503, 02531, 2015.

[71] Z.-H. Huang, S.-Z. Yang, W. Lin, J. Ni, S.-L. Sun, Y.-W. Chen, Y. Tang, "Review of knowledge distillation research," *Chinese Journal of Computers*, vol. 45, no. 03, pp. 624–653, 2022.

[72] R.-R. Shao, Y. Liu, W. Zhang, J. Wang, "Review of knowledge distillation in deep learning," *Chinese Journal of Computers*, vol. 45, no. 08, pp. 1638–1673, 2022.

[73] A. Søgaard, S. Ruder, I. Vulić, "On the Limitations of Unsupervised Bilingual Dictionary Induction," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* , 2018, pp. 778–788.

[74] B. Patra, J.-R.-A. Moniz, S. Garg, M.-R. Gormley, G. Neubig, "Bilingual Lexicon Induction with Semisupervision in Non-Isometric Embedding Spaces," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 184–193.

[75] T. Séjourné, J. Feydy, F.-X. Vialard, A. Trouvé, and G. Peyré, "Sinkhorn divergences for unbalanced optimal transport," *arXiv preprint*, vol. 1910, 12958, 2019.

[76] J. Feydy, T. Séjourné, F. Vialard, S. Amari, A. Trouvé, and G. Peyré, "Interpolating between optimal transport and mmd using sinkhorn divergences," in *the 22nd International Conference on Artificial Intelligence and Statistics(PMLR)*, 2019, pp. 2681–2690.

[77] K.-H. Huang, W. Ahmad, N.-Y. Peng, and K.-W. Chang, "Improving Zero-Shot Cross-Lingual Transfer Learning via Robust Training," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021, pp. 1684–1697.

[78] C. Dyer, V. Chahuneau, N.-A. Smith, "A simple, fast, and effective reparameterization of IBM model 2," in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2013, pp. 644–648.

[79] R. Mihalcea, P. Tarau, "Textrank: Bringing order into text," in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 2004, pp. 404–411.

[80] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L, Zettlemoyer, and V. Stoyanov, "Unsupervised Cross-lingual Representation Learning at Scale," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 8440–8451.

[81] A. Conneau A, G. Lample, "Cross-lingual language model pretraining," *Advances in Neural Information Processing Systems*, vol. 634, pp. 7059–7069, 2019.

[82] Z.-W. Chi, S.-H. Huang, D. Li, S.-M. Ma, B. Zheng, S. Singhal, P. Bajaj, X. Song, X.-L. Mao, H.-Y. Huang, and F.-R. Wei, “XLM-E: Cross-lingual Language Model Pre-training via ELECTRA,” in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, 2022, pp. 6170–6182.

[83] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y.-Q. Zhou, W. Li, and L. Peter J., “Exploring the limits of transfer learning with a unified text-to-text transformer,” *The Journal of Machine Learning Research*, vol. 21, no. 1, pp. 5485–5551, 2020.

[84] G. Wenzek, M.-A. Lachaux, A. Conneau, V. Chaudhary, F. Guzmán, A. Joulin, and E. Grave, “CCNet: Extracting High Quality Monolingual Datasets from Web Crawl Data,” in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 2020, pp. 4003–4012.

[85] M. Ziemski, M. Junczys-Dowmunt, B. Pouliquen, “The united nations parallel corpus v1. 0,” in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, 2016, pp. 3530–3534.

[86] A. Kunchukuttan, P. Mehta, P. Bhattacharyya, “The IIT Bombay English-Hindi Parallel Corpus,” in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

[87] H. Schwenk, V. Chaudhary, S. Sun, H.-Y. Gong, and F. Guzmán, “WikiMatrix: Mining 135M Parallel Sentences in 1620 Language Pairs from Wikipedia,” in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, 2021, pp. 1351–1361.

[88] A. El-Kishky, V. Chaudhary, F. Guzmán, and P. Koehn, “CCAligned: A Massive Collection of Cross-Lingual Web-Document Pairs,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 5960–5969.

[89] A. Conneau, G. Lample, “ Cross-lingual language model pretraining,” *Advances in neural information processing systems*, vol. 634, pp. 7059–7069, 2019.

[90] E. K. Wang, X. Zhang, F. Wang, T.-Y. Wu, “ Multilayer dense attention model for image caption,” *IEEE Access*, vol. 7, pp. 66358-66368, 2019

[91] K.-M. Hermann, T. Kočiský, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom, “Teaching machines to read and comprehend,” in *Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS’15)*, MIT Press, 2015, pp. 1693–1701.

[92] J.-N. Zhu, H.R. Li, T.-S. Liu, Y. Zhou, J.-J. Zhang, and C.-Q. Zong, “MSMO: Multimodal summarization with multimodal output,” in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 4154–4164.

[93] B. Hu, Q. Chen, F. Zhu, “LCSTS: A Large Scale Chinese Short Text Summarization Dataset,” in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015, pp. 1967–1972.

[94] Y.-Q. Wang, Q.-M. Yao, J.-T. Kwok, and L.-M. Ni, “Generalizing from a few examples: A survey on few-shot learning,” *ACM computing surveys*, vol. 53, no. 3, pp. 1–34, 2020.

[95] S. Motiian, Q. Jones, S.-M. Iranmanesh, and G. Doretto, “Few-shot adversarial domain adaptation,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 6673–6683.

[96] P.-F. Liu, W.-Z. Yuan, J.-L. Fu, Z.-B. Jiang, H. Hayashi, and G. Neubig, “Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing,” *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–35, 2023.

[97] T. Schick, H. Schütze, “Few-shot text generation with pattern-exploiting training,” *arXiv preprint*, vol. 2012, 11926, 2020.

[98] X.-L. Li, P. Liang, “Prefix-Tuning: Optimizing Continuous Prompts for Generation,” in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 2021, pp. 4582–4597.

[99] J.-N. Zhu, H.R. Li, T.-S. Liu, Y. Zhou, J.-J. Zhang, and C.-Q. Zong, “MSMO: Multimodal summarization with multimodal output,” in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 4154–4164.

[100] B. Hu, Q. Chen, F. Zhu, “LCSTS: A Large Scale Chinese Short Text Summarization Dataset,” in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015, pp. 1967–1972.

[101] S.-H. Zheng, Z.-X. Li, J.-A. Wang, J.-F. Qu, A. Liu, L. Zhao, and Z.-G. Chen, “Long-Document Cross-Lingual Summarization,” in *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining (WSDM ’23)*, 2023, pp. 1084–1092.

[102] Y. Liu, M. Lapata, "Hierarchical Transformers for Multi-Document Summarization," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 5070–5081.

[103] H. Hayashi, P. Budania, P. Wang, C. Ackerson, R. Neervannan, and G. Neubig, "WikiAsp: A dataset for multi-domain aspect-based summarization," *Transactions of the Association for Computational Linguistics*, vol. 9, pp. 211–225, 2021.

[104] P.-J. Liu, M. Saleh, E. Pot, " Generating wikipedia by summarizing long sequences," *arXiv preprint*, vol. 1801, 10198, 2018.

[105] B. Gliwa, I. Mochol, M. Biesek, and A. Wawer, "SAMSum Corpus: A Human-annotated Dialogue Dataset for Abstractive Summarization," in *Proceedings of the 2nd Workshop on New Frontiers in Summarization*, 2019, pp. 70–79.

[106] C.-G. Zhu, Y. Liu, J. Mei, and M. Zeng, "MediaSum: A Large-scale Media Interview Dataset for Dialogue Summarization," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2021, pp. 5927–5934.

[107] S. Banerjee, A. Lavie, "METEOR: An automatic metric for MT evaluation with improved correlation with human judgments," in *Proceedings of the Acl Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, 2005, pp. 65–72.

[108] H. Isozaki, T. Hirao, K. Duh, K. Sudoh, and H. Tsukada, "Automatic Evaluation of Translation Quality for Distant Language Pairs," in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 2010, pp. 944–952.

[109] Y. Cheng, Y. Cheng, "Joint training for pivot-based neural machine translation," *Joint Training for Neural Machine Translation*, vol. 2019, pp. 41–45, 2019.