applied sciences

MDPI

Article

T5-Based Model for Abstractive Summarization: A Semi-Supervised Learning Approach with Consistency Loss Functions

Mingye Wang [1,*], Pan Xie [1], Yao Du [1] and Xiaohui Hu [2]

1 School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China; panxie@buaa.edu.cn (P.X.); by1903066@buaa.edu.cn (Y.D.)
2 Science and Technology on Integrated Information System Laboratory, Institute of Software, Chinese Academy of Sciences, Beijing 100045, China; hxh@iscas.ac.cn
* Correspondence: marcel0829@buaa.edu.cn

Abstract: Text summarization is a prominent task in natural language processing (NLP) that condenses lengthy texts into concise summaries. Despite the success of existing supervised models, they often rely on datasets of well-constructed text pairs, which can be insufficient for languages with limited annotated data, such as Chinese. To address this issue, we propose a semi-supervised learning method for text summarization. Our method is inspired by the cycle-consistent adversarial network (CycleGAN) and considers text summarization as a style transfer task. The model is trained by using a similar procedure and loss function to those of CycleGAN and learns to transfer the style of a document to its summary and vice versa. Our method can be applied to multiple languages, but this paper focuses on its performance on Chinese documents. We trained a T5-based model and evaluated it on two datasets, CSL and LCSTS, and the results demonstrate the effectiveness of the proposed method.

Keywords: natural language processing; automatic text summarization; abstractive summarization; semi-supervised learning; consistency loss function

check for updates

Citation: Wang, M.; Xie, P.; Du, Y.; Hu, X. T5-Based Model for Abstractive Summarization: A Semi-Supervised Learning Approach with Consistency Loss Functions. *Appl. Sci.* **2023**, *13*, 7111. https://doi.org/10.3390/app13127111

Academic Editor: Alessandro Di Nuovo

Received: 20 April 2023
Revised: 1 June 2023
Accepted: 9 June 2023
Published: 14 June 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Automatic text summarization is a crucial task in natural language processing (NLP) that aims to condense the core information of a given corpus into a brief summary. With the exponential growth of textual data, including documents, articles, and news, automatic summarization has become increasingly important.

Text summarization methods can be classified into two categories: extractive and abstractive. Extractive summarization selects the most important sentences from the original corpus based on statistical or linguistic features, whereas abstractive summarization generates a summary by semantically understanding the text and expressing it in a new way [1]. Abstractive summarization is more challenging than extractive summarization, but it is also considered superior, as it avoids the issues of coherence and consistency in the summaries generated with extractive methods.

Deep learning has achieved state-of-the-art results in NLP, and more researchers have shifted their focus to abstractive summarization. The sequence-to-sequence (seq2seq) model [2] combined with an attention mechanism has become a benchmark in abstractive summarization [3–5]. However, these methods require well-constructed datasets, which can be difficult and costly to build.

In this paper, we propose a semi-supervised learning method for text summarization that treats summarization as a style transfer task. Our approach uses a transfer text-to-text transformer (T5) model as the text generator and trains it with loss functions from the cycle-consistent adversarial network (CycleGAN) for semantic transfer.

The remainder of this paper is structured as follows. In Section 2, we review previous research related to our work. Section 3 describes our method of text summarization in detail. Section 4 presents the experimental results of our proposed model. In Section 5, we perform an extensive ablation study to validate the effectiveness of our model. Finally, we summarize our work in Section 6.

2. Related Works

2.1. Automatic Text Summarization

Automatic text summarization is a crucial task in the field of natural language processing (NLP), and it has received a significant amount of attention from researchers in recent years. Over the years, a range of methods and models have been proposed to improve the quality of automatic text summaries. In the early days of NLP research, traditional approaches to text summarization were based on sentence ranking algorithms that evaluated the importance of sentences in a given text. These methods used statistical features, such as frequency and centrality, to rank sentences and select the most important ones to form a summary [6–8].

With the advent of machine learning techniques in the 1990s, researchers have applied these methods to NLP to improve the quality of summaries. In automatic text summarization, this is mostly considered a sequence classification problem. Models are trained to differentiate summary sentences from non-summary sentences [9–12]. These methods are referred to as extractive, as they essentially extract important phrases or sentences from the text without fully understanding their meaning. Thanks to the tremendous success of deep learning techniques, many extractive summarization studies have been proposed based on techniques including the encoder–decoder classifier [13], recurrent neural network (RNN) [14], sentence embeddings [15], reinforcement learning, and long short-term memory (LSTM) network [16].

Moreover, the development of deep learning has given rise to a method called abstract summarization. Abstract summarization has improved significantly and has become a crucial area of research in the NLP field. Researchers have made remarkable progress in this field by leveraging deep learning techniques, such as RNN [3], LSTM [17], and classic seq2seq models [4,5].

With the introduction of the transformer architecture in 2017 [18], transformer-based models have significantly outperformed other models in many NLP tasks. This architecture has been naturally applied to the text summarization task, leading to the development of several models based on pre-trained language models, including BERT [19], BART [20], and T5 [21]. These models have demonstrated remarkable performance on various NLP tasks, including text summarization.

2.2. Text Style Transfer

Text style transfer is a task in the field of NLP that focuses on modifying the style of a text without altering its content. This task has received considerable attention from researchers due to its potential applications in many areas, such as creative writing, machine translation, and sentiment analysis.

The early methods for text style transfer mainly focused on rule-based approaches, where linguistic patterns and attributes were manually defined and applied to modify the style of text [22]. These methods, though simple and effective, are limited by the fixed set of rules that they rely on, which may not adapt well to changing styles and genres.

With the advent of deep learning, several machine-learning-based approaches have been proposed. The most well-known method is the sequence-to-sequence (seq2seq) model [2]. Seq2seq models have been used in various NLP tasks, such as text summarization and machine translation, due to their ability to encode the source text and generate a target text.

Recently, generative adversarial networks (GANs) [23] were applied to the task of text style transfer. The idea of GANs is to train two neural networks: a generator and a discriminator. The generator tries to generate text that is indistinguishable from the target style, while the discriminator tries to differentiate between the generated text and the real target text.

2.3. Cycle-Consistent Adversarial Network

The cycle-consistent adversarial network (CycleGAN) is a generative adversarial network (GAN) architecture for image-to-image translation tasks. This approach has been widely used in various domains, including but not limited to image style transfer, domain adaptation, and super-resolution. The key idea of CycleGAN is to train two generator–discriminator pairs, with each pair consisting of a generator and a discriminator. One generator aims to translate an image from the source domain to the target domain, while the other generator aims to translate an image from the target domain back to the source domain. The discriminator in each pair is trained to distinguish the translated images from the real images in the corresponding domain. The cycle consistency loss is introduced to force the translated image to be transformed back into the original image.

Figure 1 illustrates how CycleGAN works in one direction.

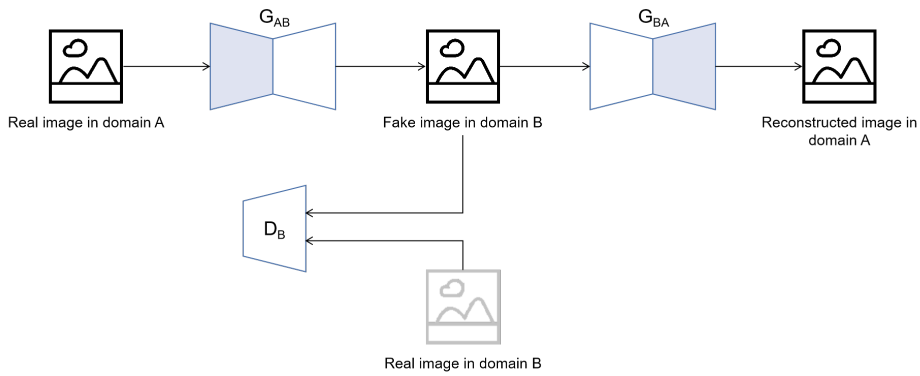


Figure 1. Working principle of CycleGAN.

CycleGAN is focused on the application of style transfer in computer vision. For example, Zhu et al. [24] originally proposed CycleGAN for unpaired image-to-image translation, where there was no one-to-one mapping between the source and target domains. This method has been widely used in tasks such as colorization, super-resolution, and style transfer. Based on CycleGAN, different models have been proposed for face transfer [25], Chinese handwritten character generation [26], image generation from text [27], image correction [28], and tasks in the audio field [29–31].

One of the highlights of CycleGAN is the implementation of two consistency losses in addition to the original GAN loss: identity mapping loss and cycle consistency loss. The identity mapping loss implies that the source data should not be changed during transformation if they are already in the target domain. The cycle consistency loss comes with the idea of back translation: The result of back translation should be the same as the original source. These two loss functions cause the CycleGAN model to keep great consistency during its transfer procedure; thus, it is possible to handle unpaired images and achieve outstanding results.

2.4. Transfer Text-to-Text Transformer

The transfer text-to-text transformer (T5) [21] is a state-of-the-art pre-trained language model based on the transformer architecture. It adopts a unified text-to-text framework that can handle any natural language processing (NLP) task by converting both the input and output into natural language texts. T5 can be easily scaled up by varying the number of parameters (from 60M to 11B), which enables it to achieve superior performance on various NLP benchmarks. Moreover, T5 employs a full-attention mechanism that allows it to capture long-range dependencies and complex semantic relations in natural language texts. T5 has been successfully applied to many NLP tasks, such as machine translation, text summarization, question answering, and sentiment analysis [21].

The T5 model follows the typical encoder–decoder structure, and its architecture is shown in Figure 2.

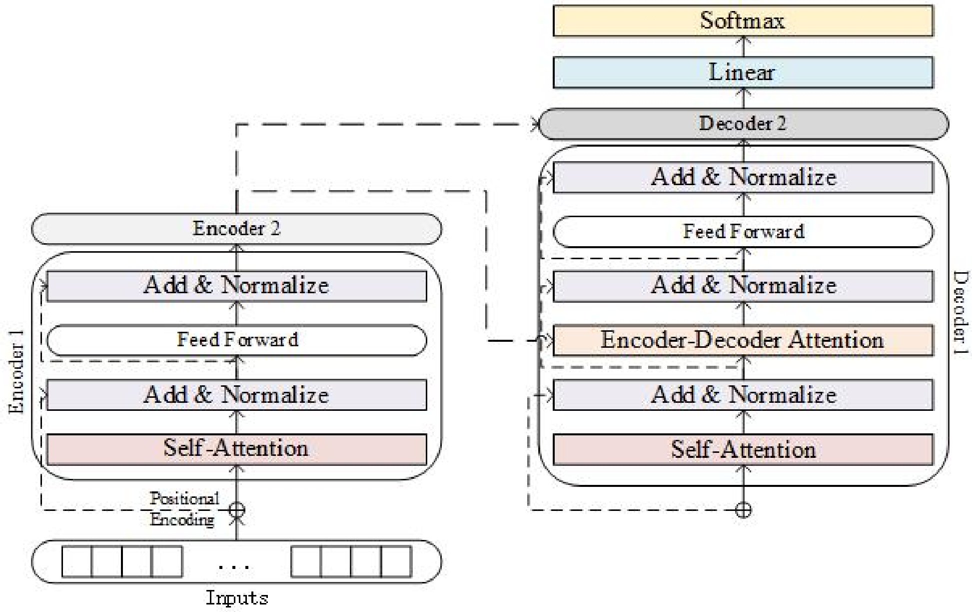


Figure 2. Architecture of the T5 model.

One of the key features of T5's text-to-text framework is the use of different prefixes to indicate different tasks, thus transforming all NLP problems into text generation problems. For example, to perform sentiment analysis on a given sentence, T5 simply adds the prefix "sentiment:" before the sentence and generates either "positive" or "negative" as the output. This feature makes it possible to train a single model that can perform multiple tasks without changing its architecture or objective function.

3. Proposed Methodology

3.1. Overall

This section presents the foundation of our semi-supervised method for automatic text summarization. Unlike existing models, which rely heavily on paired text for supervised training, our approach leverages a small paired dataset followed by a semi-supervised training process with unpaired corpora. The algorithm used in our method is illustrated in Algorithm 1, where L denotes the loss incurred by comparing two texts.

Our approach is inspired by the CycleGAN architecture, which uses two generators to facilitate style transfer in two respective directions. The first part of our method comprises a warm-up step that employs real text pairs to clarify the tasks of the style transferers T_{a2s} and T_{s2a} and generate basic outputs. The subscripts $a2s$ and $s2a$, which represent "article-to-summary" and vice versa, are employed to clarify the transfer direction. The second part adopts a similar training procedure to that of CycleGAN with consistency loss functions to further train the models without supervision.

Specifically, the identity mapping loss ensures that a text should not be summarized if it is already a summary and vice versa. The corresponding training procedure is based on calling the model to re-generate an *identity* of the input text. The loss is then calculated by measuring the difference between the original text and the generated identity. This part is designed to train the model to be capable of identifying the characteristics of two distinct text domains. In the following sections of the paper, a superscript *idt* is used to indicate re-generated identity texts.

In contrast, the cycle consistency loss trains the model to reconstruct a summary after expanding it or vice versa. The corresponding training procedure follows a cyclical process: For a real summary s, the model T_{s2a} first expands it and generates a fake article. The term "fake" indicates that it is generated by our model, rather than a real example from datasets. Next, the fake article is sent to T_{a2s} to re-generate its summary. For real articles, the same cycle steps are utilized. This part is designed to train the model to be capable of transferring texts between two domains. In the following, a superscript *fake* is used to indicate the fake texts generated by the models, and a superscript *cyc* is used to indicate the final outputs after such a cycle procedure.

Algorithm 1 Semi-supervised automatic text summarization.

for each $batch \in gold\_batches$ **do**
  fine-tune T_{a2s} and T_{s2a} with $batch$ ▷ Finetune with real text pairs
end for
for $epoch \in [1, nb\_epochs]$ **do**
  for all (a_i, s_i) such that $a_i \in Articles$ and $s_i \in Summaries$ **do**
    $(a_i^{idt},\ s_i^{idt}) \leftarrow (T_{s2a}(a_i),\ T_{a2s}(s_i))$ ▷ Re-expand and re-summary
    $(L_a^{idt},\ L_s^{idt}) \leftarrow (L(a_i, a_i^{idt}),\ L(s_i, s_i^{idt}))$ ▷ identity mapping loss
    $(s_i^{fake},\ a_i^{fake}) \leftarrow (T_{a2s}(a_i),\ T_{s2a}(s_i))$ ▷ Generate fake summary and article
    $(a_i^{cyc},\ s_i^{cyc}) \leftarrow (T_{s2a}(s_i^{fake}),\ T_{a2s}(a_i^{fake}))$ ▷ Restore article and summary
    $(L_a^{cyc},\ L_s^{cyc}) \leftarrow (L(a_i, a_i^{cyc}),\ L(s_i, s_i^{cyc}))$ ▷ cycle consistency loss
    $Loss \leftarrow L_a^{idt} + L_s^{idt} + L_a^{cyc} + L_s^{cyc}$ ▷ Total loss
    Back-propagation of $Loss$
  end for
end for

As observed, despite the integration of the CycleGAN loss functions, we refrain from constructing a GAN architecture for our task. This decision arises from two factors: firstly, the challenge involved in the back-propagation phase of discrete sampling during text generation; secondly, the lack of discernible improvement vis-à-vis our method during development and the inherent instability in the training process.

The back-propagation of gradients for text generation in a GAN framework presents an arduous problem, which is primarily due to the discrete nature of text data. Consequently, the GAN model for text generation often entails the adoption of reinforcement learning or the use of Gumbel–softmax approximation. These techniques are complicated and may render the training process unstable, leading to the production of sub-optimal summaries.

Moreover, we found no clear evidence of improved performance through the use of GAN-based models in our task in comparison with our semi-supervised method with CycleGAN loss functions. Therefore, we conclude that our approach presents a promising solution for automatic text summarization and is better suited for our task given its simplicity and effectiveness.

3.2. Style Transfer Model

As mentioned previously, we view the summarization task as a style transfer problem. To accomplish this, we employ a T5 model, which offers several advantages over alternative models. Firstly, the native tasks of the T5 model align well with the requirements of the style transfer task. Secondly, by modifying the prefix of the input text, a T5 model can perform tasks in both directions, i.e., from text to summary and vice versa.

As illustrated in Figure 3, a single T5 model can perform the tasks of T_{a2s} and T_{s2a} outlined in Algorithm 1 by changing the prefix of the input text. Therefore, we only require one generator for both directions, unlike in the original CycleGAN architecture.

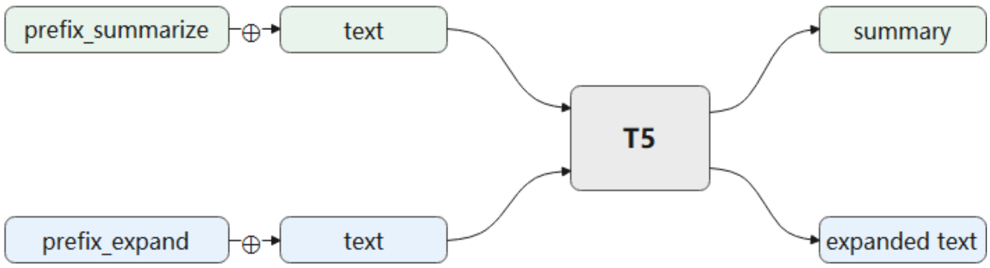


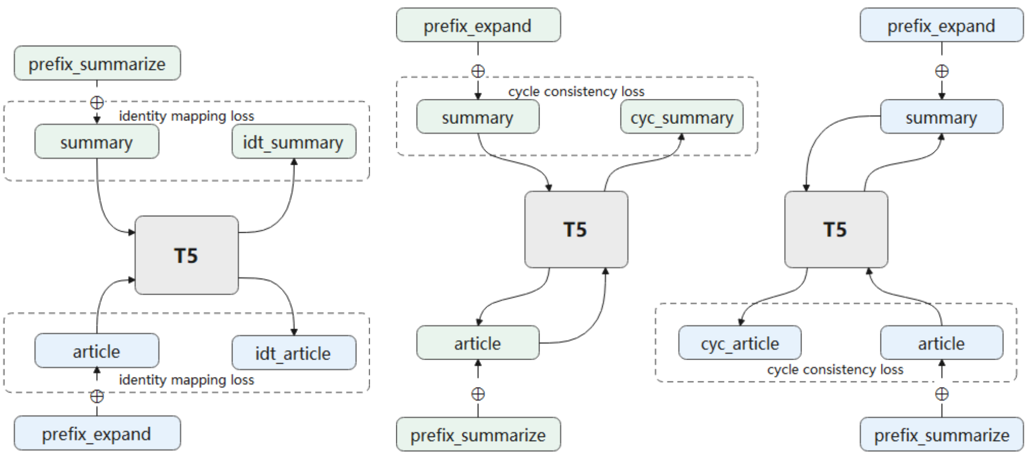
Figure 3. T5 model with different prefixes.

The versatility of the T5 model in undertaking various natural language processing tasks has been well documented in recent research. The model's pre-training process enables it to perform a wide range of tasks, including question answering, text classification, and text generation. By leveraging the strengths of the T5 model, our approach provides an effective solution to the problem of automatic text summarization.

3.3. Training with the T5 Model

Our training procedure consists of two parts: a supervised part and an unsupervised part. In the supervised part, we use small labeled data for warm-up while following the same procedure as that in the original T5 model. In this part, we fine-tune the T5 model with pairs of articles and summaries using different prefixes to indicate the generation direction. The loss function for the supervised part is cross-entropy, which is the same loss as that used in the original T5 model.

In the unsupervised part, we adopt a training procedure inspired by the CycleGAN architecture, thus incorporating identity mapping loss and cycle consistency loss. The identity mapping loss deters the model from re-summarizing a summary or expanding a full article by minimizing the difference between the input and output texts. Meanwhile, the cycle consistency loss ensures that the model preserves the source text after a cyclical transfer by minimizing the difference between the input and reconstructed texts. Figure 4 illustrates these two processes.



(**a**) Identity mapping loss (**b**) Cycle consistency loss

Figure 4. CycleGAN losses of the proposed model.

We propose a novel training procedure that uses a single T5 model for both generation tasks with different prefixes. Given an article a and its summary s, we use the T5 model to generate a fake summary s^{fake} from a and a fake article a^{fake} from s. To indicate the desired task, we prepend a prefix string to the input text. The generation process can be formulated as follows:

$$\begin{aligned} s^{fake} &= T_s(a) = T(P_s \oplus a) \\ a^{fake} &= T_e(s) = T(P_e \oplus s) \end{aligned} \tag{1}$$

where $T_s()$ and $T_e()$ denote the T5 model with the summary prefix and the expansion prefix, respectively.

The training process follows a typical supervised paradigm, a cross-entropy loss [32] is calculated to measure the difference between two texts, and the model is trained via back-propagation.

$$L(x, x^{fake}) = -\sum_{i=1}^{C} p_i(x) \log p_i(x^{fake}) \tag{2}$$

where C is the vocabulary size, and $p_i()$ is the probability of i-th word in the vocabulary.

For the rest of the dataset, where an article a and a summary s are not paired, we calculate the two consistency losses. The identity mapping loss is calculated by re-summarizing a summary or re-expanding an article as follows:

$$\begin{aligned} a^{idt} &= T_e(a) & s^{idt} &= T_s(s) \\ L_a^{idt} &= L(a, a^{idt}) & L_s^{idt} &= L(s, s^{idt}) \end{aligned} \tag{3}$$

As for the cycle consistency loss, the model first generates s^{fake} and a^{fake} as stated before; then, it regenerates a^{cycle} and s^{cycle} based on s^{fake} and a^{fake}. After such a cycle, the losses are calculated as follows:

$$\begin{aligned} a^{fake} &= T_s(a) & s^{fake} &= T_e(s) \\ a^{cyc} &= T_e(s^{fake}) & s^{cyc} &= T_s(a^{fake}) \\ L_a^{cyc} &= L(a, a^{cyc}) & L_s^{cyc} &= L(s, s^{cyc}) \end{aligned} \tag{4}$$

The training algorithm is, thus, adapted as in Algorithm 2 (T for T5 model, $\oplus$ for concatenation of texts). We use P_s and P_e to denote $prefix\_summarize$ and $prefix\_expand$, respectively.

Algorithm 2 Semi-supervised automatic text summarization with T5.

Set $prefix\_summarize$ and $prefix\_expand$ as P_s and P_e
for each $batch \in gold\_batches$ **do**
    $(article, summary) \leftarrow batch;$
    fine-tune T with $(P_s \oplus article, summary)$ and $(P_e \oplus summary, article)$
        ▷ Fine-tune with real text pairs
end for
for $epoch \in [1, nb\_epochs]$ **do**
    for all (a_i, s_i) such that $a_i \in Articles$ and $s_i \in Summaries$ **do**
        $(a_i^{idt},\ s_i^{idt}) \leftarrow (T(P_e \oplus a_i),\ T(P_s \oplus s_i))$ ▷ Re-expand and re-summarize
        $(L_a^{idt},\ L_s^{idt}) \leftarrow (L(a_i, a_i^{idt}),\ L(s_i, s_i^{idt}))$ ▷ identity mapping loss
        $(s_i^{fake},\ a_i^{fake}) \leftarrow (T(P_s \oplus a_i),\ T(P_e \oplus s_i))$ ▷ Generate fake summary and article
        $(a_i^{cyc},\ s_i^{cyc}) \leftarrow (T(P_e \oplus s_i^{fake}),\ T(P_s \oplus a_i^{fake}))$ ▷ Restore article and summary
        $(L_a^{cyc},\ L_s^{cyc}) \leftarrow (L(a_i, a_i^{cyc}),\ L(s_i, s_i^{cyc}))$ ▷ cycle consistency loss
        $Loss \leftarrow \lambda_{idt} L_{idt}(a_i, a_i^{idt}) + \lambda_{idt} L_{idt}(s_i, s_i^{idt}) + \lambda_{cyc} L_{cyc}(a_i, a_i^{cyc}) + \lambda_{cyc} L_{cyc}(s_i, s_i^{cyc})$
            ▷ Total loss
        Back-propagation of $Loss$
    end for
end for

Here, the hyperparameters λ_{idt} and λ_{cyc} control the weights of the two types of losses.

4. Experiments

This section presents the experimental details for evaluating the performance of our method.

4.1. Datasets

We conducted experiments on two datasets: CSL (Chinese Scientific Literature Dataset) [33] and LCSTS (Large Scale Chinese Short Text Summarization Dataset) [34].

The CSL is the first scientific document dataset in Chinese consisting of 396,209 papers' meta-information obtained from the National Engineering Research Center for Science and Technology Resources Sharing Service (NSTR) and spanning from 2010 to 2020. In our experiments, we used the paper titles and abstracts to generate summary–article pairs for training and evaluation purposes. To facilitate evaluation and comparison, we chose the subset of CSL used in the Chinese Language Generation Evaluation (CLGE) [35] for our experiments. This sub-dataset comprised 3500 computer science papers.

The LCSTS is a large dataset collecting 2,108,915 Chinese news articles published on Weibo, the most popular Chinese microblogging website. The data in LCSTS include news titles and contents posted by verified media accounts. Similarly to with CSL, we used the news titles and contents to create summary–article pairs for our experiments.

Examples from these datasets can be viewed in Figures A1 and A2.

For the unsupervised training part, our model did not have access to the matched summary–article pairs. Instead, we intentionally broke the pairs and randomly shuffled the data, ensuring that the model did not receive matched data during this part of the training.

4.2. Implementation Details

The original datasets contained well-paired texts. We used only a fraction of the paired data during the warm-up stage. The unsupervised part used text samples of the corresponding dataset without pair information.

Since the original T5 model does not support the Chinese language, we chose Mengzi [36], a high-performing lightweight (103M parameters) pre-trained language model for Chinese in our experiments (Mengzi includes a family of pre-trained models, among which we used the T5-based one).

We used the AdamW optimizer to train the model with the learning rate, $\beta 1$, $\beta 2$, ϵ, and weight decay as 5×10^{-5}, 0.9, 0.999, 1×10^{-6}, and 0.01, respectively. Moreover, we set the learning rate with a cosine decay schedule. We restricted the length of sentences in each batch to a maximum of 512 tokens, and we set the batch size to 8. The two consistency losses were weighted with factors of 0.1 for the identity mapping loss and 0.2 for the cycle consistency loss. The higher weight for the cycle consistency loss was due to its direct contribution to the model's ability to transfer texts, which was the primary objective of the task. In contrast, the identity mapping loss helped preserve the characteristics of the input texts, but it did not directly contribute to the summarization process. All of the experiments were conducted by using Python 3.7.12 with PaddlePaddle 2.3 and PyTorch 1.11 while running on an NVIDIA Tesla 32GB V100 GPU. For clarity, the hyperparameter settings used in our experiments are presented in Table 1.

Table 1. Hyperparameters used to train the model.

Hyperparameter	Value
Optimizer	AdamW
Learning rate	5×10^{-5}
$\beta 1$	0.9
$\beta 2$	0.999
ϵ	1×10^{-6}
Weight decay	0.01
Learning rate schedule	Cosine decay
Sentence length	512 tokens
Batch size	8
Identity mapping loss weight	0.1
Cycle consistency loss weight	0.2

4.3. Results

In this section, we present the results of our proposed approach for automatic text summarization and compare its performance with baselines on four commonly used evaluation metrics: the ROUGE-1, ROUGE-2, ROUGE-L [37], and BLEU [38] scores. ROUGE is the acronym for Recall-Oriented Understudy for Gisting Evaluation, and BLEU is the acronym for BiLingual Evaluation Understudy.

The evaluation metrics play a critical role in assessing the effectiveness of a summarization model. The ROUGE and BLEU scores are widely used to evaluate the quality of generated summaries. ROUGE measures the overlap between the generated summary and the reference summary at the n-gram level, whereas BLEU assesses the quality of the summary by computing the n-gram precision between the generated summary and the reference summary. By comparing the performance of our proposed model with the baselines on these four metrics, we can determine the effectiveness of our approach in automatic text summarization. To provide clarity, we present the formal definitions of these metrics as follows:

$$ROUGE\text{-}N = \frac{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count(gram_n)} \tag{5}$$

where n stands for the length of the n-gram, $gram_n$, and $Count_{match}(gram_n)$ is the maximum number of n-grams co-occurring in a candidate summary and a set of reference summaries. By switching the reference and summary, we get the precision and recall values. The final ROUGE-N score is, hence, the F1 score. We used ROUGE-1 and ROUGE-2 in our experiments. ROUGE-L is based on the longest common subsequence (LCS). It is calculated in the same way as ROUGE-N, but by replacing the n-gram match with the LCS.

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right) \tag{6}$$

where p_n is the proportion of correctly predicted n-grams within all predicted n-grams. Typically, we use $N = 4$ kinds of grams and uniform weights $w_n = N/4$. BP is the brevity penalty, which penalizes sentences that are too short:

$$\text{Brevity Penalty } = \begin{cases} 1, & \text{if } c > r \\ e^{(1-r/c)}, & \text{if } c <= r \end{cases} \tag{7}$$

where c is the predicted length and r is the target length.

We conducted experiments on two Chinese datasets: CSL [33], which consists of abstracts from the scientific literature and their corresponding titles, and LCSTS [34], which consists of Chinese news articles and their corresponding human-written summaries. Due to the lack of research on semi-supervised Chinese summarization, all baselines used in this study were fully supervised models and were proposed by the organizers of the original corresponding datasets. For the CSL dataset, we conducted the supervised part of the experiment with two fractions of the original dataset: one using 50 paired samples, and the other using 250, while the remaining data were used for the unsupervised part of our method. For the LCSTS dataset, which was larger than CSL, we conducted the experiments with 200 and 1000 paired samples.

We also performed an ablation study in comparison with the T5 model trained with labeled data only and without our proposed loss functions. The T5 models in Table 2 refer to the results obtained in these cases.

Table 2 illustrates the performance of the baselines and our proposed approach on the CSL dataset, while Table 3 shows the results on the LCSTS dataset.

Table 2. CSL results.

Models	ROUGE-1	ROUGE-2	ROUGE-L	BLEU
ALBERT-tiny	52.75	37.96	48.11	21.63
BERT-base	63.83	51.29	59.76	41.45
BERT-wwm-ext	63.44	51	59.4	41.19
RoBERTa-wwm-ext	63.23	50.74	58.99	41.31
LSTM-seq2seq	46.48	30.48	41.8	22
Original T5 50	34.82	19.93	32.62	3.85
T5 50 with CL (ours)	53.13	41.03	50.85	33.95
Original T5 250	56.45	45.01	53.96	37.48
T5 250 with CL (ours)	59.41	47.93	56.16	38.91

Table 3. LCSTS results.

Models	ROUGE-1	ROUGE-2	ROUGE-L	BLEU
RNN-Word	17.7	8.5	15.8	-
RNN-Char	21.5	8.9	18.6	-
RNN-context-Word	26.8	16.1	24.1	-
RNN-context-Char	29.9	17.4	27.2	-
mT5	-	-	34.8	-
CPM-2	-	-	35.88	-
Original T5 200	23.61	12.00	21.80	3.99
T5 200 with CL (ours)	28.28	15.48	25.84	10.56
Original T5 1000	28.01	15.59	25.66	9.51
T5 1000 with CL (ours)	30.09	18.59	29.00	14.74

The results presented in Tables 2 and 3 demonstrate that our method achieved comparable performance to that of early supervised large models and even outperformed them in several metrics, despite using only a lightweight model and a limited amount of data. However, the performance of recent supervised models was still better than that of our semi-supervised method. For instance, on CSL, our best results achieved over 93% of the fully supervised BERT-base's performance on every metric, significantly outperforming LSTM-seq2seq and ALBERT-tiny. Regarding LCSTS, our model achieved better results than the best early fully supervised model, RNN-context-Char, by about 6%, and it had a score that was approximately 81% of the ROUGE-L of recent models, such as mT5 and CPM2. The experimental results confirm the effectiveness of our proposed approach in automatic text summarization.

In addition to comparing our results with those of other models, it is important to highlight the comparison between the results of our models and that of the original T5 models without unsupervised learning. This comparison sheds light on the effectiveness of incorporating unsupervised learning techniques in our approach, as evidenced by the improved summarization performance, particularly when well-paired data or "gold batches" were limited. Our semi-supervised method notably improved the performance across every metric compared to the fully supervised T5 model trained on a limited amount of labeled data. When labeled text pairs were extremely rare, the proposed method significantly improved the performance on every metric, especially the BLEU score (from 3.85 to 33.95 on SCL and from 3.99 to 10.56 on LCSTS). As the number of golden batches increased, the original T5 achieved better results, while our method still ameliorated its performance. This demonstrates the effectiveness of our approach in leveraging the information contained in unlabeled data.

The present study showcases a portion of the experimental findings, which are visually presented in Figures A1 and A2.

5. Conclusions

This study presents a novel semi-supervised learning method for abstractive summarization. To achieve this, we employed a T5-based model to process texts and utilized an identity mapping constraint and a cycle consistency constraint to exploit the information contained in unlabeled data. The identity mapping constraint ensures that the input and output of the model have a similar representation, whereas the cycle consistency constraint ensures that the input text can be reconstructed from the output summary. Through this approach, we aim to improve the generalization ability of the model by leveraging unlabeled data while requiring only a limited number of labeled examples.

A key contribution of this study is the successful application of CycleGAN's training process and loss functions to NLP tasks, particularly text summarization. Our method demonstrates significant advantages in addressing the problem of limited annotated data and showcases its potential for wide applicability in a multilingual context, especially when handling Chinese documents. Despite not modifying the model architecture, our approach effectively leverages the strengths of the original T5 model while incorporating the benefits of semi-supervised learning.

Our proposed method was evaluated on various datasets, and the experimental results demonstrate its effectiveness in generating high-quality summaries with a limited number of labeled examples. In addition, our method employs lightweight models, making it computationally efficient and practical for real-world applications.

Our approach can be particularly useful in scenarios where obtaining large amounts of labeled data is challenging, such as when working with rare languages or specialized domains.

It is worth noting that our proposed method can be further improved by using more advanced pre-training techniques or by fine-tuning on larger datasets. Additionally, exploring different loss functions and architectures could also lead to better performance.

In summary, our study introduces a novel semi-supervised learning approach for abstractive summarization, which leverages the information contained in unlabeled data and requires only a few labeled examples. The proposed approach offers a practical and efficient method for generating high-quality summaries, and the experimental results demonstrate its effectiveness on various datasets.

6. Limitations and Future Work

In this section, we discuss the limitations of our proposed T5-based abstractive summarization method and suggest directions for future work to address these limitations.

Semi-supervised training requirement: Our model cannot be trained entirely in an unsupervised manner. Instead, it requires a small amount of labeled data for a "warm-up" in a semi-supervised training setting. In our experiments, we found that the performance of the model trained in a completely unsupervised fashion was inferior to that of the semi-supervised approach. Future work could explore ways to reduce the reliance on labeled data or investigate alternative unsupervised training techniques to improve the model's performance.

Room for improvement in model performance: Although our model can match the performance of some earlier supervised training models, there is still a gap between its performance and that of more recent state-of-the-art models. Future research could focus on refining the model architecture, incorporating additional contextual information, or exploring novel training strategies to further enhance the performance of our proposed method.

Domain adaptability: The adaptability of our model to other domains remains to be tested through further experimentation. Our current results demonstrate the model's effectiveness on specific datasets, but its generalizability to different contexts and domains is still an open question. Future work could involve testing the model on a diverse range of datasets and languages, as well as developing techniques for domain adaptation to improve its applicability across various settings.

Author Contributions: Conceptualization, M.W.; methodology, M.W.; software, M.W.; validation, P.X. and Y.D.; formal analysis, M.W.; investigation, M.W.; resources, P.X.; data curation, Y.D.; writing—original draft preparation, M.W.; writing—review and editing, M.W. and X.H.; visualization, M.W.; supervision, X.H.; project administration, X.H.; funding acquisition, X.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the key R&D project of the Ministry of Science and Technology of the People's Republic of China with grant number 2020-JCJQ-ZD-079-00.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The datasets and baselines utilized in our experiments are available at the following URLs: https://github.com/ydli-ai/CSL and http://icrc.hitsz.edu.cn/Article/show/139.html. The codes and outputs of our proposed model can also be accessed at https://github.com/StarsMoon/ATS (20 April 2023).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Original text	运用传统体系结构描述语言描述的软件体系结构(SA)方案始终存在着一些横切行为和特征,它们混杂和散列在不同的SA设计单元中,使得SA难以理解、难以演化和难以重用。针对这一问题,基于时序逻辑语言XYZ/E,在统一的时序逻辑框架下设计出一种面向方面体系结构描述语言AC2-ADL。系统地阐述了AC2-ADL的概念框架并用XYZ/E进行语义解释,最后结合案例介绍了如何用AC2-ADL对SA进行描述。 The software architecture (SA) schemes described using traditional architectural description languages always have some cross-cutting behaviors and characteristics, which are mixed and scattered in different SA design units, making it difficult to understand, evolve and reuse the SA. To address this issue, a new aspect-oriented architecture description language (AC2-ADL) was designed based on the temporal logic language XYZ/E, which provides a unified temporal logic framework. The concept framework of AC2-ADL is systematically explained and semantically interpreted using XYZ/E. Finally, a case study is presented to demonstrate how to describe SA using AC2-ADL.
Ground truth	基于时序逻辑的面向方面体系结构描述语言 A Aspect-Oriented Architecture Description Language Based on Temporal Logic
Prediction	基于时序逻辑结构描述的软件体系结构描述 Software architecture description based on temporal logic structure description
Original text	在外包数据库运行模式下,由第三方提供的数据库服务器处于非信任域,存在数据文件盗版、数据内容篡改等安全风险。构建了一个基于信息隐藏技术的外包数据库版权保护系统,综合运用数字水印、PKI机制与数字证书、数字签名及可信硬件模块USB Key等技术,设计相应的水印协议,实现对外包数据库的版权保护。与传统的以数据加密、数字签名等方法为主要技术手段的解决方案相比,它具有冗余存储量及网络附加流量小、隐蔽性好、验证信息难以删除等优点。 In the outsourcing database operation mode, the database servers provided by third parties are in non-trusted domains, and there are security risks such as data file piracy and data content tampering. A copyright protection system based on information hiding technology has been constructed for outsourced databases. It comprehensively uses technologies such as digital watermarking, PKI mechanism with digital certificates, digital signatures, and trusted hardware modules such as USB Key. Corresponding watermarking protocols have been designed to achieve copyright protection for outsourced databases. Compared with traditional solutions that mainly rely on techniques such as data encryption and digital signatures, it has the advantages of small network additional traffic, good concealment, difficult-to-delete verification information, and redundant storage capacity.
Ground truth	基于信息隐藏的外包数据库版权保护系统 An outsourced database copyright protection system based on information hiding
Prediction	基于信息隐藏的外包数据库版权保护 Copyright Protection of Outsourced Databases Based on Information Hiding
Original text	川化股份近日发布公告称，公司董事长、总经理、董秘、财务负责人集体辞职。雪上加霜的是，川化股份控股股东四川化工控股（集团）有限责任公司从12月18日开始，通过多次减持，共减持公司2500万股股票，套现8156万元左右。 Recently, Chuanhua Group announced that the company's chairman, general manager, secretary of the board of directors, and finance director have all resigned. To make matters worse, Chuanhua Group's controlling shareholder, Sichuan Chemical Holding (Group) Co., Ltd., began reducing its stake in the company on December 18th through multiple sales, selling a total of 25 million shares and cashing in approximately 81.56 million yuan.
Ground truth	川化股份四高管集体辞职控股股东大手笔抛售股票 Four senior executives of Chuanhua Group have collectively resigned, and the controlling shareholder has sold a large amount of stocks
Prediction	川化股份高管集体辞职 The senior executives of Chuanhua Group have resigned collectively
Original text	为了精确地对体数据进行可以交互的切割模拟,提出了切割距离场的概念和一种基于切割距离场的三维体数据切割算法,该算法利用切割距离场作为切割体的数据表示形式,在图形处理单元(GPU)的着色器中同时实现体重建和实时切割两种功能,实验结果表明该切割算法可以达到可交互的速度,可以处理任意形状的切割体。 In order to accurately simulate interactive cutting of volumetric data, the concept of a cutting distance field and a three-dimensional volumetric data cutting algorithm based on the cutting distance field are proposed. The algorithm uses the cutting distance field as the representation of the cutting volume, and simultaneously implements both volumetric reconstruction and real-time cutting in the graphics processing unit (GPU) shader. The experimental results show that the cutting algorithm can achieve interactive speed and handle cutting volumes of arbitrary shapes.
Ground truth	基于切割距离场的体数据切割算法研究 Research on volumetric data cutting algorithm based on cutting distance field
Prediction	基于切割距离场的三维体数据切割算法 3D volumetric data cutting algorithm based on cutting distance field
Original text	普适计算环境中的服务推荐需要满足系统异构性和移动性的要求。提出了一种基于贝叶斯网络的多Agent服务推荐机制并进行实现,将贝叶斯网络和聚类方法应用于服务推荐中,并设计了推荐模型自学习机制,充分考虑了上下文对服务推荐的影响及改进。实现系统由完成历史上下文汇集、知识训练、决策推荐和自学习功能的多个Agent构成,通过Agent之间的通信内容设计,在Agent之间建立流程控制和数据共享通道。 Service recommendation in a ubiquitous computing environment needs to meet the requirements of system heterogeneity and mobility. A multi-agent service recommendation mechanism based on Bayesian networks is proposed and implemented, in which Bayesian networks and clustering methods are applied to service recommendation, and a self-learning mechanism for recommendation models is designed to fully consider the impact of context on service recommendation and improvement. The implemented system is composed of multiple agents that complete the collection of historical context, knowledge training, decision recommendation, and self-learning functions. Through the design of communication content between agents, a process control and data sharing channel is established among them.
Ground truth	基于贝叶斯网络的多Agent服务推荐机制研究 Research on multi-agent service recommendation mechanism based on bayesian network
Prediction	基于贝叶斯网络的多Agent服务推荐 Multi-agent service recommendation based on bayesian networks

Figure A1. Some experimental results on CSL with human translation.

Original text	84岁高龄的“中国杂交水稻之父”袁隆平近日接受媒体采访时表示，他正在研究“转基因水稻”，并力挺“转基因”，称其为“今后的发展方向”。袁隆平说，“我们也不能听到转基因就害怕，要谨慎对待转基因，而很多转基因还是好的”。 In a recent interview with the media, Yuan Longping, the 84-year-old "father of hybrid rice in China", stated that he is currently researching "genetically modified rice" and supports GMOs, calling them "the direction for future development." Yuan Longping said, "We cannot be afraid of GMOs just because we hear the term. We should treat GMOs with caution, and many GMOs are actually good."
Ground truth	袁隆平力挺转基因称对其不能一概而论 Yuan Longping supports genetically modified organisms (GMOs) and believes that they cannot be judged as a whole
Prediction	袁隆平:不能听到转基因就害怕 Yuan Longping: Don't be afraid of GMOs just because you hear the term.
Original text	从2010年至2013年10月东阿阿胶价格上调6次，如今每公斤已涨至1098元。其中五次为提高出厂价，一次为提高零售指导价。东阿阿胶相关人士表示，调价的原因，主要是阿胶的价值所决定的，同时也是保护阿胶产业，涵养整个产业链的客观需求。 From 2010 to October 2013, the price of Dong-E E-Jiao, a traditional Chinese medicine made from donkey hide, was increased six times, with the current price per kilogram rising to 1098 yuan. Five of these price hikes were to raise the factory price, and one was to raise the retail price. Officials from Dong-E E-Jiao stated that the reason for the price adjustments was mainly due to the value of E-Jiao and to protect the E-Jiao industry, as well as to meet the objective needs of the entire industry chain.
Ground truth	东阿阿胶四年提价181%每公斤超千元 Over four years, the price of Dong-E E-Jiao increased by 181%, with the current price exceeding one thousand yuan per kilogram.
Prediction	东阿阿胶涨价6次涨至1098元 The price of Dong-E E-Jiao was increased six times, with the current price rising to 1098 yuan.
Original text	川化股份近日发布公告称，公司董事长、总经理、董秘、财务负责人集体辞职。雪上加霜的是，川化股份控股股东四川化工控股（集团）有限责任公司从12月18日开始，通过多次减持，共减持公司2500万股股票，套现8156万元左右。 Recently, Chuanhua Group announced that the company's chairman, general manager, secretary of the board of directors, and finance director have all resigned. To make matters worse, Chuanhua Group's controlling shareholder, Sichuan Chemical Holding (Group) Co., Ltd., began reducing its stake in the company on December 18th through multiple sales, selling a total of 25 million shares and cashing in approximately 81.56 million yuan.
Ground truth	川化股份四高管集体辞职控股股东大手笔抛售股票 Four senior executives of Chuanhua Group have collectively resigned, and the controlling shareholder has sold a large amount of stocks
Prediction	川化股份高管集体辞职 The senior executives of Chuanhua Group have resigned collectively
Original text	未来的房价走势，在城市资源配置没有根本改观之前，各城市的差异化仍将继续，马太效应仍将加剧。正因如此，一线城市居民的住房资产不会严重缩水。挣钱不易，投资不易，且行且珍惜。 The future trend of housing prices will continue to vary among cities until there is a fundamental improvement in the allocation of urban resources, and the Matthew Effect will continue to intensify. Therefore, residents in first-tier cities will not suffer a significant decline in their housing assets. It is not easy to earn money or invest, so cherish every opportunity.
Ground truth	一线城市的住房资产会严重缩水吗? Will the real estate assets in first-tier cities experience a significant drop in value?
Prediction	一线城市房价不会严重缩水? Will the housing prices in first-tier cities not experience a significant decline?
Original text	证监会新闻发言人表示，证监会对各类市场操纵行为始终保持高压态势，发现一起、查处一起，资金量小也难逃法眼。证监会新闻发言人张晓军22日通报了4起市场操纵典型案例，并对其中的新型操纵行为进行了分析。 The spokesperson of China Securities Regulatory Commission (CSRC) stated that the CSRC has always maintained a high-pressure situation against various market manipulation behaviors. Any discovered cases will be investigated and punished, even if the amount of funds involved is small. On the 22nd, the CSRC spokesperson, Zhang Xiaojun, announced four typical cases of market manipulation and analyzed the new types of manipulation behaviors involved.
Ground truth	证监会通报4起市场操纵典型案例 CSRC releases four typical cases of market manipulation
Prediction	证监会通报4起市场操纵典型案例 CSRC releases four typical cases of market manipulation
Original text	截至今年5月1日，一次性发泡塑料餐具“重出江湖”一年整。因“白色污染”被禁用十余年，该产品于去年恢复生产和使用。今年4月国际食品包装协会对各地的生产企业进行调查显示，所用原料不合格的高达70%，按照国家要求进行产品包装标识的仅有1家。 As of May 1st this year, disposable foam plastic tableware has been back in use for a whole year. It had been banned for over a decade due to "white pollution," but production and use of the product resumed last year. In April of this year, an investigation by the International Food Packaging Association showed that up to 70% of the raw materials used by production companies in various regions were unqualified, and only one company complied with national requirements for product packaging labeling.
Ground truth	一次性发泡塑料餐具“归来” The return of disposable foam plastic tableware
Prediction	一次性发泡塑料餐具“重出江湖” The resurgence of disposable foam plastic tableware

Figure A2. Some experimental results on LCSTS with human translation.

References

1. Yao, K.; Zhang, L.; Luo, T.; Wu, Y. Deep reinforcement learning for extractive document summarization. *Neurocomputing* **2018**, *284*, 52–62. [CrossRef]
2. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to sequence learning with neural networks. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 3104–3112.

3. Chopra, S.; Auli, M.; Rush, A.M. Abstractive sentence summarization with attentive recurrent neural networks. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; pp. 93–98.
4. Hou, L.; Hu, P.; Bei, C. Abstractive document summarization via neural model with joint attention. In Proceedings of the National CCF Conference on Natural Language Processing and Chinese Computing, Dalian, China, 8–12 November 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 329–338.
5. Nayeem, M.T.; Fuad, T.A.; Chali, Y. Neural diverse abstractive sentence compression generation. In Proceedings of the European Conference on Information Retrieval, Cologne, Germany, 14–18 April 2019; pp. 109–116.
6. Ferreira, R.; Cabral, L.; Lins, R.D.; Silva, G.; Favaro, L. Assessing sentence scoring techniques for extractive text summarization. *Expert Syst. Appl.* **2013**, *40*, 5755–5764. [CrossRef]
7. Radev, D.R. LexRank: Graph-based Lexical Centrality as Salience in Text Summarization. *J. Qiqihar Jr. Teach. Coll.* **2004**, *22*, 2004.
8. Alguliev, R.M.; Aliguliyev, R.M.; Isazade, N.R. Multiple documents summarization based on evolutionary optimization algorithm. *Expert Syst. Appl.* **2013**, *40*, 1675–1689. [CrossRef]
9. Conroy, J.M.; O'Leary, D.P. Text summarization via hidden Markov models. In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New Orleans, LA, USA, 13 September 2001.
10. Mihalcea, R.; Tarau, P. TextRank: Bringing Order into Texts. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, 20 October 2004.
11. Bollegala, D.T.; Okazaki, N.; Ishizuka, M. A machine learning approach to sentence ordering for multidocument summarization and its evaluation. In Proceedings of the International Conference on Natural Language Processing, Jeju Island, Republic of Korea, 11–13 October 2005.
12. Baralis, E.; Cagliero, L.; Mahoto, N.; Fiori, A. GRAPHSUM: Discovering correlations among multiple terms for graph-based summarization. *Inf. Sci.* **2013**, *249*, 96–109. [CrossRef]
13. Cheng, J.; Lapata, M. Neural Summarization by Extracting Sentences and Words. *arXiv* **2016**, arXiv:1603.07252.
14. Nallapati, R.; Zhai, F.; Zhou, B. SummaRuNNer: A Recurrent Neural Network based Sequence Model for Extractive Summarization of Documents. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.
15. Anand, D.; Wagh, R. Effective Deep Learning Approaches for Summarization of Legal Texts. *J. King Saud Univ.-Comput. Inf. Sci.* **2019**, *34*, 2141–2150. [CrossRef]
16. Mohsen, F.; Wang, J.; Al-Sabahi, K. A hierarchical self-attentive neural extractive summarizer via reinforcement learning (HSASRL). *Appl. Intell.* **2020**, *50*, 2633–2646. [CrossRef]
17. Rush, A.M.; Chopra, S.; Weston, J. A Neural Attention Model for Abstractive Sentence Summarization. *arXiv* **2015**, arXiv:1509.00685.
18. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, *30*, 5998–6008.
19. Zhang, H.; Gong, Y.; Yan, Y.; Duan, N.; Xu, J.; Wang, J.; Gong, M.; Zhou, M. Pretraining-Based Natural Language Generation for Text Summarization. In Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL), Hong Kong, China, 21 November 2019.
20. Lewis, M.; Liu, Y.; Goyal, N.; Ghazvininejad, M.; Zettlemoyer, L. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. *arXiv* **2019**, arXiv:1910.13461.
21. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *J. Mach. Learn. Res.* **2020**, *21*, 5485–5551.
22. Ban, H. Stylistic Characteristics of English News. In Proceedings of the Japan-Korea Joint Symposium on Emotion & Sensibility, Daejeon, Republic of Korea, 4–5 June 2004.
23. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014.
24. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
25. Wu, R.; Gu, X.; Tao, X.; Shen, X.; Tai, Y.W.; Jia, J.I. Landmark Assisted CycleGAN for Cartoon Face Generation. *arXiv* **2019**, arXiv:1907.01424.
26. Bo, C.; Zhang, Q.; Pan, S.; Meng, L. Generating Handwritten Chinese Characters using CycleGAN. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018.
27. Gorti, S.K.; Ma, J. Text-to-Image-to-Text Translation using Cycle Consistent Adversarial Networks. *arXiv* **2018**, arXiv:1808.04538.
28. Harms, J.; Lei, Y.; Wang, T.; Zhang, R.; Zhou, J.; Tang, X.; Curran, W.J.; Liu, T.; Yang, X. Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. *Med. Phys.* **2019**, *46*, 3998–4009. [CrossRef] [PubMed]
29. Kaneko, T.; Kameoka, H. CycleGAN-VC: Non-parallel Voice Conversion Using Cycle-Consistent Adversarial Networks. In Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), Roma, Italy, 3–7 September 2018.
30. Kaneko, T.; Kameoka, H.; Tanaka, K.; Hojo, N. CycleGAN-VC2: Improved CycleGAN-based Non-parallel Voice Conversion. In ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 9 April 2019.

31. Kaneko, T.; Kameoka, H.; Tanaka, K.; Hojo, N. CycleGAN-VC3: Examining and Improving CycleGAN-VCs for Mel-spectrogram Conversion. *arXiv* **2020**, arXiv:2010.11672.
32. Bishop, C. *Pattern Recognition and Machine Learning*; Stat Sci; Springer: Berlin/Heidelberg, Germany, 2006.
33. Li, Y.; Zhang, Y.; Zhao, Z.; Shen, L.; Liu, W.; Mao, W.; Zhang, H. CSL: A Large-scale Chinese Scientific Literature Dataset. In Proceedings of the 29th International Conference on Computational Linguistics, Gyeongju, Republic of Korea, 12–17 October 2022; pp. 3917–3923.
34. Hu, B.; Chen, Q.; Zhu, F. LCSTS: A Large Scale Chinese Short Text Summarization Dataset. *arXiv* **2015**, arXiv:1506.05865.
35. CLUEbenchmark. Chinese Language Generation Evaluation. 2020. Available online: https://github.com/CLUEbenchmark/CLGE (accessed on 8 June 2023).
36. Zhang, Z.; Zhang, H.; Chen, K.; Guo, Y.; Hua, J.; Wang, Y.; Zhou, M. Mengzi: Towards Lightweight Yet Ingenious Pre-Trained Models for Chinese. 2021. Available online: http://xxx.lanl.gov/abs/2110.06696 (accessed on 8 June 2023).
37. Lin, C.Y. Rouge: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*; Association for Computational Linguistics: Barcelona, Spain, 2004; pp. 74–81.
38. Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.J. Bleu: A Method for Automatic Evaluation of Machine Translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 7–12 July 2002; pp. 311–318. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.