English Novel Summarization and Visualizations

1st Fatemeh Pesaran zadeh
Department of Computer Science and Engineering
Seoul National University
Seoul, Republic of Korea
fatemehpesaran@snu.ac.kr

1st Sehee Kim
Department of Computer Science and Engineering
Seoul National University
Seoul, Republic of Korea
mihee90@snu.ac.kr

I. Abstract

Our paper aims to enhance language learning by leveraging visual aids to bring English novels to life. We focus on the main characters within these novels, extracting key details such as their appearance, personality traits, and relationships. From the collected information, we create captions, and create illustrations through image generation. We also provide a user interface , which is used to further ensure that the captions are true to the original text. Through these techniques, we aim to create a captivating visual experience that immerses learners in the world of the main characters, fostering a deeper connection and understanding of the narrative.

II. Introduction

In today's increasingly interconnected world, language acquisition has become a paramount endeavor, with individuals eagerly seeking to learn languages beyond their mother tongue. While numerous methodologies exist to facilitate foreign language learning, one particularly potent approach involves the integration of visual aids. In this context, our project aims to use technologies to visualize English novels.

In this paper we used Natural Language Processing (NLP) techniques, such as Named Entity Recognition (NER), to identify the central figures within the narrative.

Subsequently, we delve into the realm of prompt engineering, deftly crafting a series of targeted queries meticulously designed to extract key facets of the main characters. These thought-provoking prompts encompass a broad spectrum, encompassing inquiries into appearance features, personality traits, and pivotal roles in propelling the plot forward. Through this intricate interplay of language and context, we endeavor to unravel the tapestry of characteristics and relationships that define the central figure's captivating journey.

Once we have our main role description, we move forward to visualization. Each sentence in the description acts as a starting point to create new images. We input these sentence prompts into a special model that generates images related to the story. This process brings the story to life by turning words into vibrant pictures. This combination of language and imagery creates an exceptional language-learning experience. Learners can truly dive into the main character's world, forming a stronger connection and deeper understanding of the story.

While the previous processes do provide visualized results, there is no guarantee that the character descriptions and the subsequently created images are completely accurate to the provided text. Therefore, we use human involvement to complement these inaccuracies; by creating a user console, we allow users to examine created captions and images, then modify the captions in order to create images that are more faithful to the original text.

III. background

For language learning, the integration of visual aids has emerged as a powerful technique to enhance comprehension and engagement. In this context, our project endeavors to revolutionize language acquisition by visualizing English novels. By focusing on the main characters within these literary works, we extract key details such as their appearance and personality traits. Through innovative techniques in language processing and image generation, our aim is to create a captivating visual experience that immerses learners in the world of these characters, enabling a deeper connection and understanding of the narrative.

A. Automated Summarization Techniques

Automated text summarization is a field of study within natural language processing (NLP) and information retrieval that aims to generate concise and coherent summaries from longer textual documents. This technique has gained significant attention in computer science due to its potential for extracting key information and facilitating efficient information consumption. In the context of visualizing English novels for language learners, automated summarization techniques play a crucial role in distilling the essence of the narrative and main characters.

There are two primary approaches to automated summarization: extractive and abstractive summarization. Extractive summarization involves identifying important sentences or passages from the source text and assembling them to form a summary. This approach relies on the extraction of salient information and maintains the original wording and structure of the text. On the other hand, abstractive summarization aims to generate a summary by paraphrasing and rephrasing the source text, potentially introducing new words or phrases that capture the key concepts. This approach involves a higher level of language understanding and generation.

Several algorithms and techniques have been developed to improve the effectiveness of automated summarization. These include graph-based methods, such as TextRank [1], which applies graph algorithms to identify the most important sentences based on their connectivity within the document. Another notable approach is the use of machine learning techniques, such as Support Vector Machines (SVM) or Recurrent Neural Networks (RNN), to train models that can predict the salience or relevance of sentences for summarization [2].

Moreover, recent advancements in deep learning have led to the development of neural network-based models for abstractive summarization. These models, such as the Transformer architecture introduced by Vaswani et al. (2017), have shown promising results in generating coherent and contextually appropriate summaries. They leverage self-attention mechanisms and sequence-to-sequence learning to capture dependencies and generate abstractive summaries that go beyond simple sentence extraction.

The application of automated summarization techniques in the visualization of English novels for language learners offers the potential to distill the main storyline, character descriptions, and key events into concise and informative summaries. By incorporating these summarization techniques into the visualization pipeline, learners can grasp the core aspects of the narrative and main characters, aiding in their understanding and immersion in the story world.

B. Image Generation and Visualization

Image generation and visualization are key areas of research in computer science that focus on creating and representing visual content using computational methods. In the context of visualizing English novels for language learners, image generation techniques play a vital role in bringing the main characters and story settings to life, enhancing the immersive experience.

One prominent approach in image generation is the use of generative adversarial networks (GANs). GANs, introduced by Goodfellow et al. (2014), consist of a generator network and a discriminator network that work in tandem to produce realistic images. The generator generates synthetic images, while the discriminator evaluates the authenticity of the generated images. Through an adversarial training process, GANs can learn to generate visually convincing images that resemble real-world objects or scenes.

Furthermore, deep learning models, such as convolutional neural networks (CNNs), have been widely employed in image generation and visualization tasks. CNNs excel at capturing intricate patterns and extracting high-level features from images. Variants of CNNs, including autoencoders and variational autoencoders (VAEs), have been applied to generate and manipulate images by encoding and decoding their latent representations [3].

These image-generation techniques can be leveraged to create visual representations of the main characters, their appearances, and the story settings in English novels. By feeding textual descriptions or other relevant inputs into these models, realistic and contextually appropriate images can be generated to enrich the language learning experience.

IV. Methodology

A. Main Character Description

1) Name Entity Recognition: Despite significant progress in text summarization techniques, extracting relevant information about the main character remains a formidable challenge. However, our project aims to overcome this hurdle by harnessing the power of named entity recognition (NER) models. To achieve this, we employed the highly effective bert-base-NER model, which is a fine-tuned BERT (Bidirectional Encoder Representations from Transformers) model specifically designed for Named Entity Recognition tasks. This model achieves state-of-the-art performance for NER and has been trained to recognize four types of entities: location (LOC), organizations (ORG), person (PER), and miscellaneous (MISC).

The bert-base-NER model used in our project is a bert-base-cased model that was fine-tuned on the English version of the standard CoNLL-2003 Named Entity Recognition dataset. By leveraging the capabilities of these advanced NER models, we were able to accurately identify and extract the name of the main character, marking a crucial initial step in our process of generating character-specific captions and visually engaging images that capture the main character's journey throughout the narrative. Through this innovative approach, we successfully bridge the gap between textual information and visual representation, providing learners with an immersive language learning experience that brings the main character to life.

2) Main Character feature Extraction: In order to extract pertinent features of the main character from the novel, we employed question-answering models. By asking specific questions such as:

- "Can you describe the **main character's** appearance?"
- "Tell me about the **main character's** physical features."
- "What are the distinguishing characteristics of the **main character**?"

We were able to obtain valuable information about the main character's appearance. These carefully crafted questions served as prompts to guide the question-answering models toward extracting relevant details. The extracted information provided us with key insights into the main character's physical attributes, enabling us to create a comprehensive portrayal. Subsequently, we utilized this gathered information to recreate a concise sentence that captures the essence of the main character's appearance. This caption, derived from the extracted prompts, served as the initial sentence for our image-generative models. By leveraging this integrated approach, we seamlessly merged textual and visual elements, resulting in the generation of captivating images that authentically represent the main character's physical appearance throughout the narrative. This innovative process empowers language learners to forge a deeper connection with the main character, enhancing their understanding and engagement with the story.

B. Main Character Visualization

To generate visually coherent images related to the main character throughout the novel, we drew inspiration from the groundbreaking paper titled "Prompt-to-Prompt Image Editing with Cross Attention" [4]. This paper introduced an innovative diffusion model that aligned seamlessly with our project's objectives. By leveraging the concepts presented in this paper, we were able to create a robust framework for generating corresponding images that maintain consistency while accommodating minor prompt modifications.

The diffusion model described in the paper allowed us to explore the potential of prompt-based image editing and cross-attention mechanisms. This enabled us to make precise adjustments to the generated captions, whether by modifying prompts, adding new prompts, or refining existing ones. By iteratively refining the captions, we could intricately shape the desired narrative and visual representation of the main character's journey. This iterative process fostered a dynamic interplay between textual prompts and the image generation model, ensuring a seamless connection between the generated captions and the resulting images.

Through this approach, we facilitated a nuanced exploration of the main character's development and interactions within the novel's storyline. The generated images served as visual representations that reflected the evolving nature of the main character, providing learners with an immersive and engaging language learning experience. By witnessing the visual manifestations of the main character's progression, learners could forge a deeper connection with the narrative and develop a comprehensive understanding of the character's role and growth.

Fig. 1. Generated image with the first generated caption. Caption: **"Cinderella is Beautiful Kind-hearted, gentle, and hardworking"**

C. User interface

The previously mentioned tasks provide the user with captions and images, but it is highly unlikely that they will be completely accurate to the input text in every situation. Therefore, we proposed the idea of adding a user console that allows human involvement. The previous tasks will be run on a cloud, whereas the user console will be on a mobile device.

The process is as follows. The caption and image data that was created on the cloud will be sent to the user console, where it will display that data, as shown in Fig. 2. After the data is received, the user may examine them to see if their accuracy is satisfactory. If not, then the user may input a modified version of that caption to back to the cloud, where it will be used to create a new image. This process may be repeated until the user is satisfied with the results.

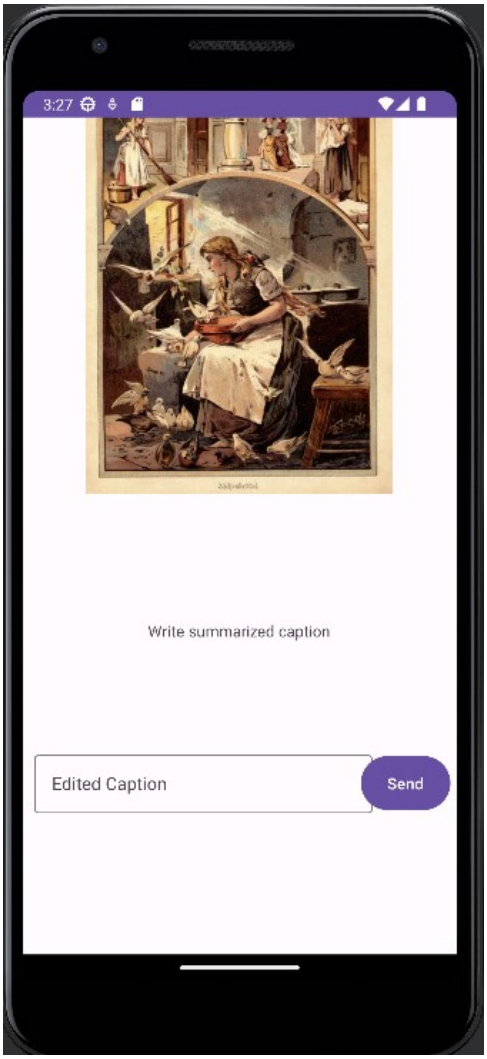

Fig. 2. The user console display

V. Conclusion

Our project aimed to enhance language learning experiences by employing a novel approach that integrates text summarization, named entity recognition, and image generation techniques. By harnessing the power of advanced NER models, we successfully extracted relevant information about the main character from the novel. Leveraging this information, we generated a caption that encapsulates the main character. Through

the utilization of the ”Prompt-to-Prompt Image Editing with Cross Attention” paper's diffusion model, we transformed this caption into an image related to the main character. We also provided a user interface so that the caption need not fully rely on the models; it could be further improved by user involvement.

A. Limitation and Future Work

Despite our efforts to create a coherent series of captions about the main character in the novel, we encountered limitations that affected the extent of our progress. Due to time constraints, we were only able to generate the first caption, limiting the completeness of the series. We were also only able to create the design of the user interface, and were not able to connect it to a cloud.

Additionally, the diffusion model we used for image generation was not fine-tuned on a specific dataset related to our novel, resulting in occasional discrepancies between the generated images and our expectations. The quality of the generated images also posed some concerns, as they did not consistently meet the desired standards. These limitations highlight the need for further refinement and fine-tuning of the models to improve the coherence and quality of the generated captions and images. Future work should focus on addressing these limitations to enhance the overall effectiveness and accuracy of the image visualization process.

REFERENCES

[1] R. Mihalcea and P. Tarau, “Textrank: Bringing order into text,” in *Proceedings of the 2004 conference on empirical methods in natural language processing*, 2004, pp. 404–411.

[2] S. Chopra, M. Auli, and A. M. Rush, “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*, 2016, pp. 93–98.

[3] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” 2022.

[4] A. Hertz, R. Mokady, J. Tenenbaum, K. Aberman, Y. Pritch, and D. Cohen-Or, “Prompt-to-prompt image editing with cross attention control,” 2022.