Interactive System for Facilitating Digital Story Illustration

Maryam Kermanshahani
Amirkabir University of Technology
Tehran, Iran
maryam2001km@gmail.com

Abstract
Several text-to-image generative models have developed in recent years, all of which can generate high-quality images based on the given query. However, most of them could just generate an image based on a phrase, and not a paragraph. Furthermore, giving a result which can align with user real intent is challenging, since sometimes users cannot justify their intentions, or the models interpret their query in another way. We are developing a system named **"WordWeaver"** that can ease the procedure of generating story images for users, aesthetically appealing to them, and making working with this interface enjoyable.

Keywords: Human Computer Interaction, Text-to-Image, Storytelling

1. Introduction

The advancements in artificial intelligence research, particularly in computer vision, have led to the development of previously unimaginable applications, such as generating new contents based on text description (Berrahal & Azizi, 2022). Nowadays AI algorithms can create images or animated videos that look very realistic based on a set of parameters to create new images. With the development of GANs, they have been widely used for this purpose (Li et al., 2020). These image-generator tools are used in different areas including: art (Xu et al., 2018), architecture (Beyan & Rossy, 2023), achieving defined learning goals (Azman et al., 2015), etc. Another application of these tools is in generating images for stories for either illustrating story content (Fotedar et al., 2020) or producing story book covers (Haque et al., 2022, Zhang et al., 2021, Iwana et al., 2016).

Photographs plays a central role in many types of informal storytelling (Balabanović et al., 2000). The use of illustrations to depict key elements of a story is essential. These images can be employed to illustrate the content within a narrative or even for a book cover. This is because, alongside the importance of images within the story's content, a book cover makes an initial impression on potential readers, serving as a salesman to its audience (Haque et al., 2022). Book covers are designed to give potential readers clues about the contents of a book, serving as a form of communication between the author and the reader (Zhang et al., 2021), as human perception is known to be predominantly visual (Nag Chowdhury et al., 2020).

There are some generative models such as Stable Diffusion (SD) and DALL-E-2 that can generate high-quality images. However, they are mainly designed for short queries. Moreover, learning to generate a meaningful and coherent sequences of images from a natural language story is a challenging task that requires understanding and reasoning on both natural language and images (Li et al., 2019). Queries that, when processed by these models, yield satisfactory results, often requires experimentations with multiple queries to achieve the desired outcome. Furthermore, for a story there should be a consistency between images, for example the story should maintain a constant theme, or the whole appearance of a character should not change within the story.

In this work, our goal is to create a user interface that utilizes some existing generative models, such as Dall-E-2 or SD, for generating images for a story. This system, which is named **"WordWeaver"**, provides users with a user-friendly interface that will make story illustration more enjoyable, as well as some tools that make the process easier and faster. With tools like labeling, users can label their characters and backgrounds, ensuring a consistent style and constant images for characters in the story. Furthermore, we are working on query engineering to facilitate the process of crafting queries by investigating how users justify their intentions and how the queries are interpreted in such generative models.

In the second phase of the project, this system will undergo evaluation, consisting of two stages: defining evaluation criteria and implementing it. In other words, the system will be tested in real-world scenarios using these criteria. Such systems can greatly help authors in speeding up the storytelling process and reducing costs. Additionally, beyond their benefits to storytellers and writers, they can find applications in various other fields. For example, they can be used to enhance education for children and students, or they can provide social media users with the ability to create relevant images corresponding to a given text or story and share them on their pages.

Despite the promising possibilities, there are various challenges along the way, including finding appropriate tools to aid users in searching for content, ensuring consistency in the style of generated images for the story, designing a user-friendly interface, and helping users improve their queries.

2. Related Work

DALL-E-2, Stable Diffusion, and Midjourney are three popular image generative models (Borji, 2022). DALL-E-2, created by OpenAI lab, is an AI system that was launched in 2021. It has a search box where users can input their queries and see the resulted images. It also has an API that people can connect to. Stable Diffusion, released by the Stability.ai company, in 2022, is an open-source model, and can be used locally. Midjourney has a similar functionality; It can be used via a Discord bot. Each of these three has its own advantages and disadvantages, and there are even articles available for comparing them. You can refer to those for more details (Borji, 2022).

Besides the mentioned large and successful projects, research has also been conducted in the field of generating book cover images. For example, in one of the articles, they created a new dataset, trained their model, and ultimately obtained results. According to their own statement, “even though these photos can’t be used as a professional book cover, they can certainly be used as creative inspiration or book cover design suggestions, which will benefit the artists when designing book covers” (Haque et al., 2022). In another example, the website Booksby.ai uses progressive GANs to create book covers (Haque et al., 2022). None of these methods, however, resulted in producing flawless book covers (Haque et al., 2022).

In the field of generating images for story content, there is also research available. For example, one of these studies demonstrates a Storytelling AI system which is capable of generating short stories and complementary illustrated images with minimal input from the user, using a text generation model (Fotedar et al., 2020). Several researches have also been done in the field of generating images for book covers (Haque et al., 2022, Zhang et al., 2021, Iwana et al., 2016).

For prompt engineering, researchers have also developed various systems. For example, in one research project, an interactive system named ‘Promptify’ was created to facilitate iterative prompt exploration and refinement for SD (Brade et al., 2023). The goal of prompt engineering is to help users to craft queries that would result in more related images. This can be done by helping users to justify their intentions clearly, as well as paying more attention to some parts or keywords that are more important to LLMs.

However, developing an interactive system with all these features that is aesthetically appealing to users and can assist them in story illustration, would be an innovation in this area. Users will be able to generate a sequence of images for book cover or story content, while ensuring consistency with previously produced images, whereas in previous systems, the focus was mainly on generating an image for a single prompt. This improvement will make the entire process more enjoyable, user-friendly, and efficient for users.

3. WordWeaver: Crafting Covers and Images for Stories

The goal of "WordWeaver” is to allow people to generate images for their stories. The project consists of two main phases. “Design and Implementation” and “Evaluation”, which we will explain in the following.

3.1 Design and Implementation

In the first phase, we work on designing and implementing the system. User can write their story and generate images for it. The user can also use the available tools to specify the elements of their story like the characters, background, places, the style of images, etc. Then, the query for generating images should be crafted based on the keywords, with editing based on the knowledge gained by studying query engineering. The user can then choose to combine these images to create their desired image. This user interface uses existing image generation models like DALL-E-2 or SD, or any other suitable models, and send the user’s modified requests to these models, displaying the results to the user. For front-end, we firstly design a user interface in Figma (Figure 1). Then, for implementation of front-end we decided to use **React** and **Tailwind CSS**, and currently we are working on that.

3.2 Evaluation

In the second phase, we evaluate the system. This evaluation process consists of two stages:

1. Designing Appropriate Metrics
2. Implementation of Metrics

3.2.1 Designing Appropriate Metrics

Firstly, it is necessary to define suitable evaluation metrics, due to the type of user. Some metrics such as user satisfaction with the user interface, their satisfaction with the generated images in terms of aesthetics and alignment with the text, and the frequency of future visits, and reading or writing time (based on type of user) could be defined.

3.2.2 Implementing of Metrics

The implementation of the metrics will be carried out. A number of users, both as authors and readers will be selected based on some factors like age, education, etc. Finally, user data will be collected for assessment.

4. Conclusion

We are currently developing “WordWeaver”, an interactive system designed to facilitate story illustration, and book cover image generation. This system provides users with a user-friendly interface aims to make storytelling easier, faster and more enjoyable. This satisfaction with this system will also be assessed in the future.

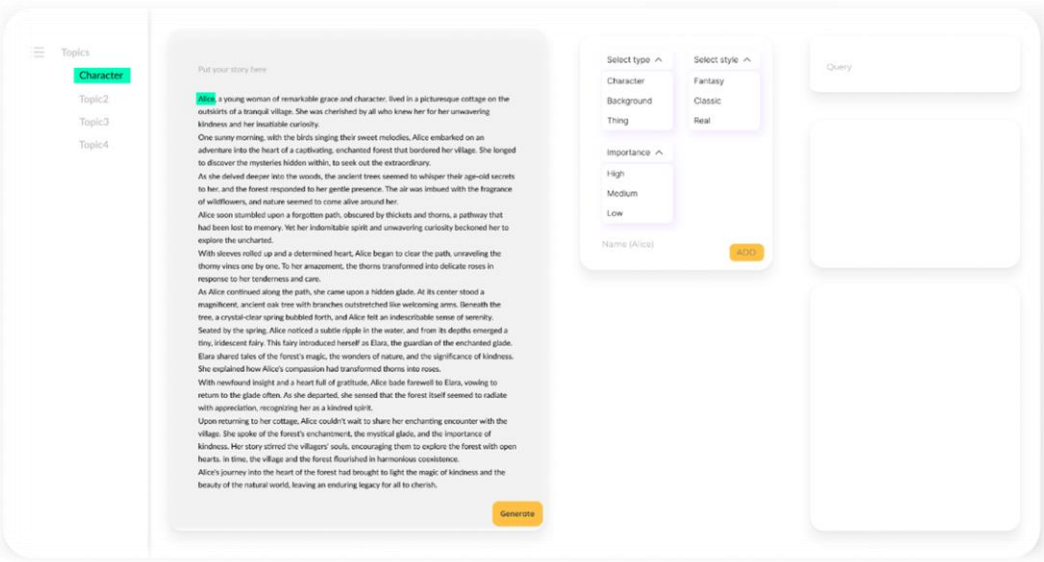

Figure 1: Initial proposed design by Figma.

5. References

Azman, F. N., Zaibon, S. B., & Shiratuddin, N. (2015). Digital storytelling tool for education: An analysis of comic authoring environments. Advances in Visual Informatics: 4th International Visual Informatics Conference, IVIC 2015, Bangi, Malaysia, November 17-19, 2015, Proceedings 4,

Balabanović, M., Chu, L. L., & Wolff, G. J. (2000). Storytelling with digital photographs. Proceedings of the SIGCHI conference on Human factors in computing systems,

Berrahal, M., & Azizi, M. (2022). Optimal text-to-image synthesis model for generating portrait images using generative adversarial network techniques. *Indones. J. Electr. Eng. Comput. Sci*, *25*(2), 972-979.

Beyan, E. V. P., & Rossy, A. G. C. (2023). A Review of AI Image Generator: Influences, Challenges, and Future Prospects for Architectural Field. *Journal of Artificial Intelligence in Architecture*, *2*(1), 53-65.

Borji, A. (2022). Generated faces in the wild: Quantitative comparison of stable diffusion, midjourney and dall-e 2. *arXiv preprint arXiv:2210.00586*.

Brade, S., Wang, B., Sousa, M., Oore, S., & Grossman, T. (2023). Promptify: Text-to-Image Generation through Interactive Prompt Exploration with Large Language Models. *arXiv preprint arXiv:2304.09337*.

Fotedar, S., Vannisselroij, K., Khalil, S., & Ploeg, B. (2020). Storytelling AI: A Generative Approach to Story Narration. AI4Narratives@ IJCAI,

Haque, E., Khan, M. F. K., Jubair, M. I., Anjum, J., & Niloy, A. Z. (2022). Book Cover Synthesis from the Summary. 2022 IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA),

Iwana, B. K., Rizvi, S. T. R., Ahmed, S., Dengel, A., & Uchida, S. (2016). Judging a book by its cover. *arXiv preprint arXiv:1610.09204*.

Li, L., Sun, Y., Hu, F., Zhou, T., Xi, X., & Ren, J. (2020). Text to realistic image generation with attentional concatenation generative adversarial networks. *Discrete Dynamics in Nature and Society*, *2020*, 1-10.

Li, Y., Gan, Z., Shen, Y., Liu, J., Cheng, Y., Wu, Y., Carin, L., Carlson, D., & Gao, J. (2019). Storygan: A sequential conditional gan for story visualization. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,

Nag Chowdhury, S., Cheng, W., De Melo, G., Razniewski, S., & Weikum, G. (2020). Illustrate your story: Enriching text with images. Proceedings of the 13th International Conference on Web Search and Data Mining,

Xu, T., Zhang, P., Huang, Q., Zhang, H., Gan, Z., Huang, X., & He, X. (2018). Attngan: Fine-grained text to image generation with attentional generative adversarial networks. Proceedings of the IEEE conference on computer vision and pattern recognition,

Zhang, W., Zheng, Y., Miyazono, T., Uchida, S., & Iwana, B. K. (2021). Towards Book Cover Design via Layout Graphs. Document Analysis and Recognition–ICDAR 2021: 16th International Conference, Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part III 16,