Project Proposal

Generative Image Models for Artistic Creation

Jianwen LYU, 20320978

[scyjl14@nottingham.ac.uk](mailto:scyjl14@nottingham.ac.uk)

Background and Motivations

Generative image models with novel training techniques such as Stable Diffusion or Disco Diffusion is gradually grabbing more attentions from public. The most typical task for these image models is text-to-image generation, which means to generate images that can semantically match the prompt (a given text descriptions from user) with high precisions [1]. With text-to-image generation models one can easily create digital images or artworks simply by writing prompts in natural language [2]. This brings to the model endless possibilities for future applications, such as s image editing, video games, and computer-aided design [1].

However, finding precise and suitable prompts for the model to complete specific visual tasks, which is also referred as prompt engineering, has emerged as a meaningful research direction [3]. During the process of writing prompts for generative AI, many problems can occur. First, because of the complexity and ambiguity of natural language, it is often difficult to describe the desired image features and develop effective prompts [4]. Second, user may confuse about how to adjust their prompts after getting unwanted results, since the possibilities for image generation from text are uncountable, with this characteristic, user’s prompting process can easily become simple try-and-error cycles, which is likely random and unprincipled [5]. Hence, the concept of prompt engineering support tools to help people develop suitable prompts comes into the view of public, and quickly forms a relatively urgent need.

Previous works of prompt engineering refiners usually combine language processing methods to discover semantic features under images and prompts in order to better identify patterns under user prompts and similar prompts, and hence provide prompting support such as prompt recommendation. For example, CLIP model can be used to detect similarity between image and text prompt for prompt recommendation [4], also, large deep language models, such as ChatGPT, can be used to classify catalogs for prompt patterns [6]. However, many of these works have not considered the scenario of both text prompt and a “template images” for model to reference as inputs. Prompt recommendation in this scenario, which appears in generating process of image generator Midiourney (in Midjourney, user can generate image based on prompt and a user-provided image to base on), can be the focus of this project.

Aim and Objectives

This project mainly aims to design a prompt engineering supporter, which can perform prompt recommendation both in the condition of receiving pure text prompt and receiving text prompt and template image. A jupyter-notebook extension that can support prompt engineering is planned to be the primitive form of the supporter.

Here are the key objectives of this project:

* Study previous researches. From previous studies and prompt engineering works, general pattern of image retrieving, key-word mining and matching can be learned and can probably pave the way for this project.
* Design an interface for prompt engineering supporter, which currently planned to be a jupyter-notebook extension, this will be done by python and javascript.
* Design and implement the algorithm part that can perform image generation, image evaluation and prompt recommendation based on both pure-text input and text-template image input. This part will be done by python, since model training and evaluating should be envolved.
* Test the algorithm and collect user stories for future evaluation, optimization, and UI improvements.

Project Plan

At early stage of the project, a visualization tool embedded in jupyter-notebook extension for image generation model will be designed and programmed. Then, based on the style and arranging of visualization process, design prompt recommendation algorithm. Finally, add the recommendation function to visualization, and test algorithm.

The project will be executed over a timeline with each phase contributing to the development of a user-friendly Prompt Engineering system, taking the GRE test I need to take in November and graduate-project application affairs into account.



References

[1] Oppenlaender, J. (2022). The Creativity of Text-to-Image Generation. 25th International Academic Mindtrek conference, [online] pp.192–202. doi:https://doi.org/10.1145/3569219.3569352.

[2] Oppenlaender, J. (2022). The Creativity of Text-to-Image Generation. *25th International Academic Mindtrek conference*, [online] pp.192–202. doi:https://doi.org/10.1145/3569219.3569352.

[3] Wang J, Liu Z, Zhao L, et al. Review of large vision models and visual prompt engineering[J]. arXiv preprint arXiv:2307.00855, 2023.

[4] Feng, Y., Wang, X., Wong, K.K., Wang, S., Lu, Y., Zhu, M., Wang, B. and Chen, W. (2023). PromptMagician: Interactive Prompt Engineering for Text-to-Image Creation. [online] arXiv.org. doi:https://doi.org/10.48550/arXiv.2307.09036.

[5] Liu, V. and Chilton, L.B. (2022) ‘Design guidelines for prompt engineering text-to-image generative models’, CHI Conference on Human Factors in Computing Systems [Preprint]. doi:10.1145/3491102.3501825.

[6] White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J. and Schmidt, D.C. (2023). A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. doi:https://doi.org/10.48550/arxiv.2302.11382.