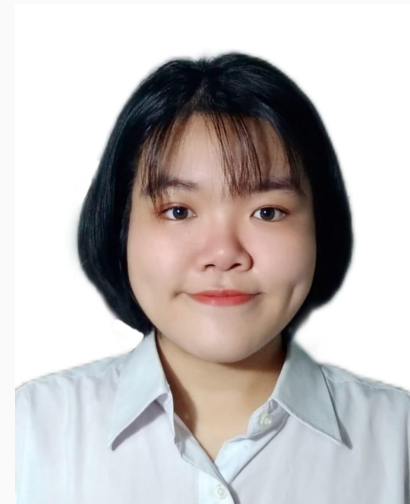


TẠO SINH HÌNH ẢNH TỪ VĂN BẢN VỚI QUY TRÌNH: GHI LẠI, LẬP KẾ HOẠCH VÀ TẠO SINH VỚI MLLMS

Lê Thị Như Ý
MSSV: 21522818

Tóm tắt

- Lớp: CS519.O11
- Link Github của nhóm:
[lethinhuy203/CS519.O11 \(github.com\)](https://github.com/lethinhuy203/CS519.O11)
- Link YouTube video:
https://youtu.be/oZEtnYW1tz0?si=GD63Cgz-P0I_JzLc



Lê Thị Như Ý

Giới thiệu

- Input

Prompt: A beautiful landscape with a river in the middle, the left of the river is **in the evening** and **in the winter** with a big iceberg and a small village while some people are skiing on the river and some people are skating, the right of the river is **in the summer** with a **volcano** **in the morning** and a small village while some people are playing.

- Output

DALL-E 3



Giới thiệu

- Input

Prompt: A **green twintail** girl in **orange dress** is sitting on the sofa while **a messy desk** is under a big window on the left, while a **lively aquarium** is on the **top right of the sofa**, realistic style.

- Output

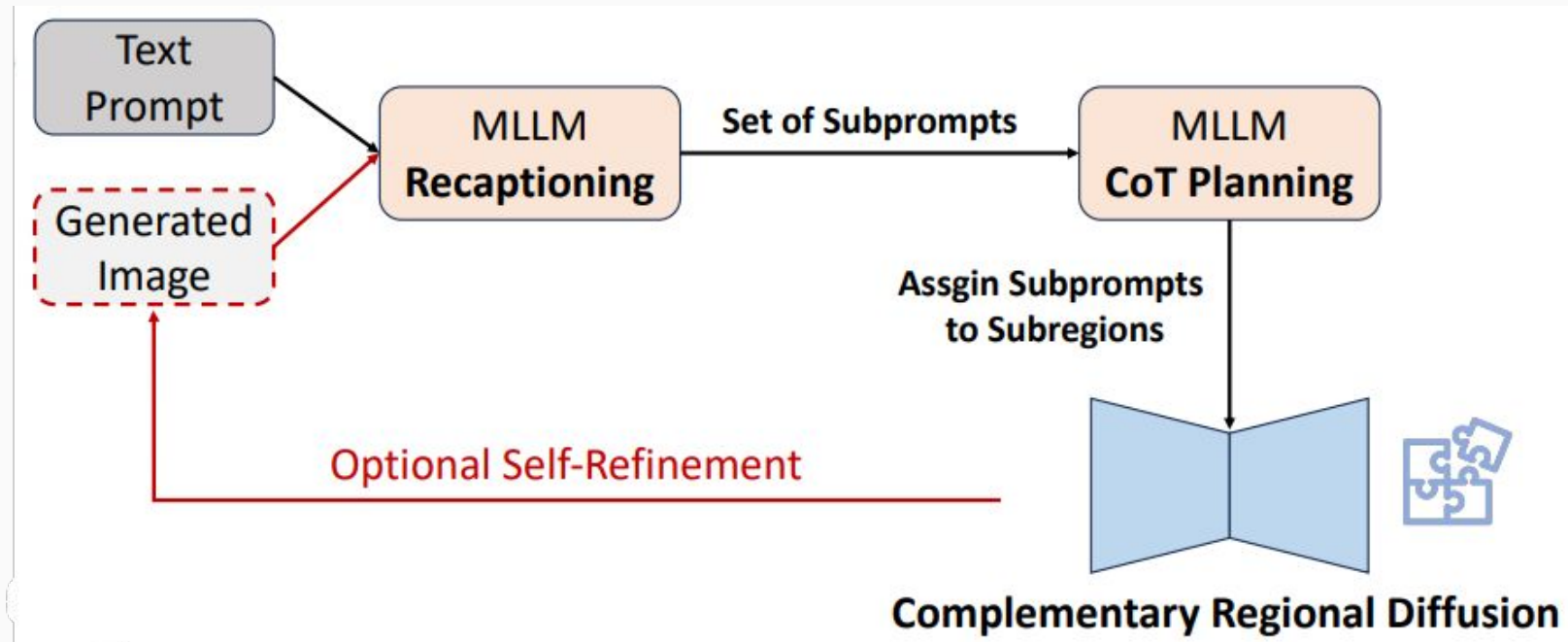
SDXL



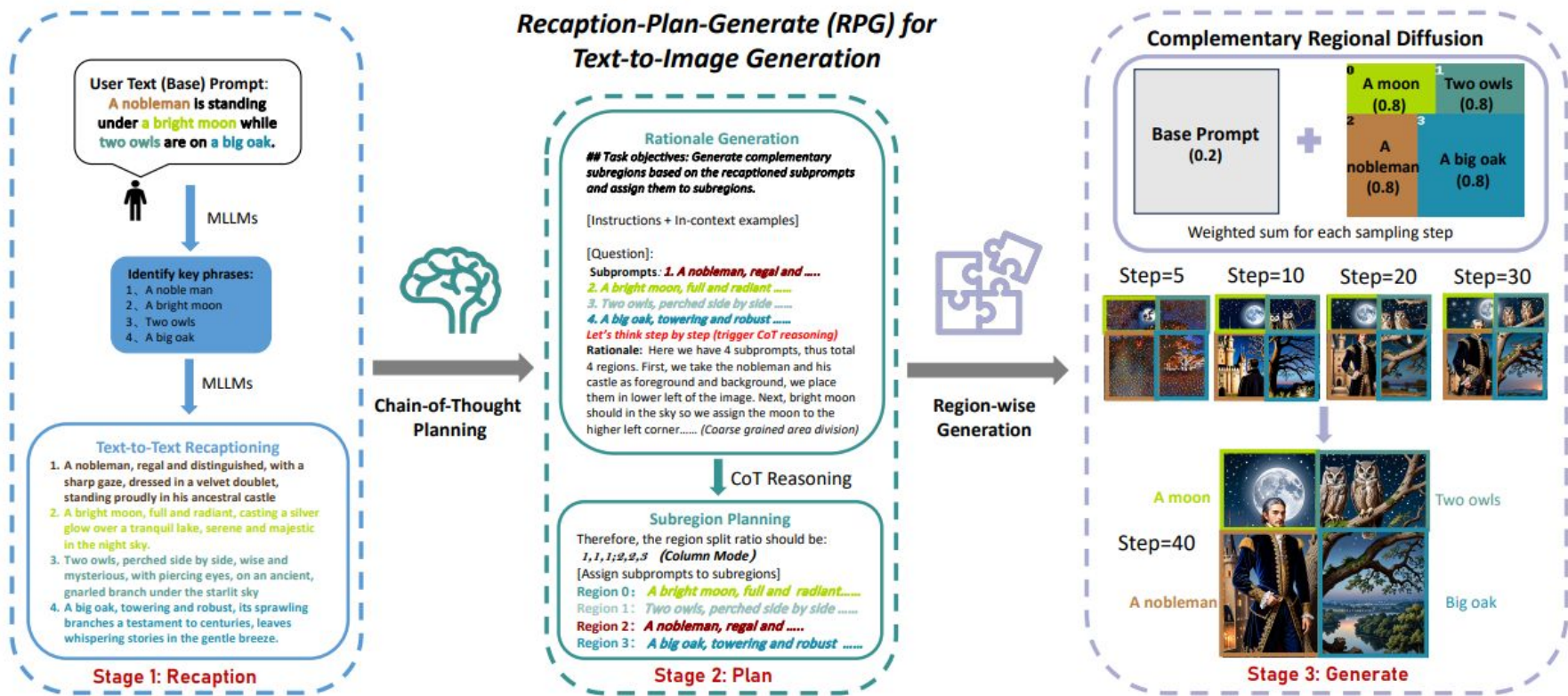
Mục tiêu

- Việc hiểu đầy đủ ý nghĩa văn bản phức tạp.
- Tạo hình ảnh có tính thống nhất và tự nhiên đối với văn bản phức tạp.

Nội dung và Phương pháp



Nội dung và Phương pháp



Kết quả dự kiến

Dựa trên chuẩn đánh giá tạo sinh hình ảnh từ văn bản T2I-CompBench, chúng tôi mong muốn kết quả của phương pháp này theo các tiêu chí như sau:

- Color: > 0.8
- Shape: > 0.5
- Texture: > 0.8
- Spatial: > 0.2
- Non-spatial: > 0.3
- Complex: > 0.4



a brown bear? → 0.9925
a blue boat? → 0.9878

Score: **0.9804**

Tài liệu tham khảo

- [1]. Bar-Tal, O., Yariv, L., Lipman, Y., and Dekel, T. Multidiffusion: Fusing diffusion paths for controlled image generation. arXiv preprint arXiv:2302.08113, 2023.
- [2]. Avrahami, O., Hayes, T., Gafni, O., Gupta, S., Taigman, Y., Parikh, D., Lischinski, D., Fried, O., and Yin, X. Spatext: Spatio-textual representation for controllable image generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18370–18380, 2023.
- [3]. Betker, J., Goh, G., Jing, L., Brooks, T., Wang, J., Li, L., Ouyang, L., Zhuang, J., Lee, J., Guo, Y., et al. Improving image generation with better captions. Computer Science. <https://cdn.openai.com/papers/dall-e-3.pdf>, 2023.
- [4]. Brooks, T., Holynski, A., and Efros, A. A. Instructpix2pix: Learning to follow image editing instructions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18392–18402, 2023.
- [5]. Cao, M., Wang, X., Qi, Z., Shan, Y., Qie, X., and Zheng, Y. Masactrl: Tuning-free mutual self-attention control for consistent image synthesis and editing. arXiv preprint arXiv:2304.08465, 2023.