PHƯƠNG PHÁP CHỈNH SỬA HÌNH ẢNH DỰA TRÊN VĂN BẢN VỚI CẢI TIẾN VÙNG CHỈNH SỬA DỰA TRÊN PATCHES VÀ TINH CHỈNH ĐA CẤP

Cao Quyet Chien - 240101005

Tóm tắt

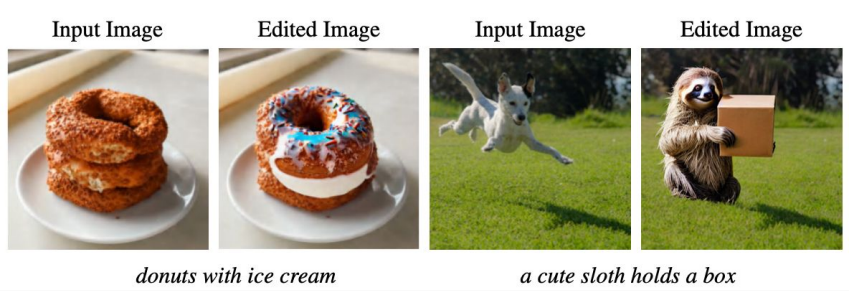
- Lớp: CS2205.NOV2024
- Link Github của nhóm: Github Document
- Link YouTube video:
- Cao Quyet Chien

Giới thiệu

Text-driven image editing: Là phương pháp chỉnh sửa hình ảnh trực quan thông qua câu lệnh ngôn ngữ tự nhiên, ứng dụng các mô hình vision-language quy mô lớn.

- **Hạn chế hiện tại:**
  - Định vị vùng chỉnh sửa không chính xác.
  - Kết quả chỉnh sửa thiếu tự nhiên, xuất hiện artifacts.
  - Khó xử lý các prompts phức tạp hoặc yêu cầu chi tiết.

- **Giải pháp đề xuất:**
  - Patch-based region optimization: Định vị chính xác vùng chỉnh sửa.
  - Multi-scale refinement: Hòa trộn chỉnh sửa mượt mà, duy trì tính chân thực.



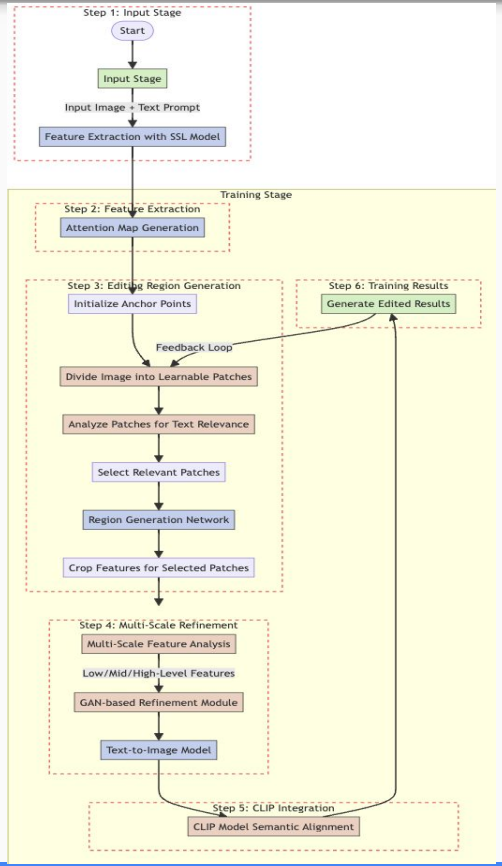
Mục tiêu

- Tăng độ chính xác vùng chỉnh sửa:
  - Tích hợp kỹ thuật patch-based region optimization để định vị chính xác khu vực cần chỉnh sửa, đảm bảo chỉ tác động lên vùng mục tiêu.
- Duy trì tính tự nhiên và hòa trộn mượt mà:
  - Áp dụng kỹ thuật multi-scale refinement để chỉnh sửa phù hợp ngữ cảnh tổng thể của hình ảnh.
- Đánh giá hiệu quả hệ thống:
  - Dựa trên các tiêu chí: độ chính xác chỉnh sửa cục bộ, mức độ phù hợp với prompt, và tính chân thực tổng thể.

Nội dung và Phương pháp

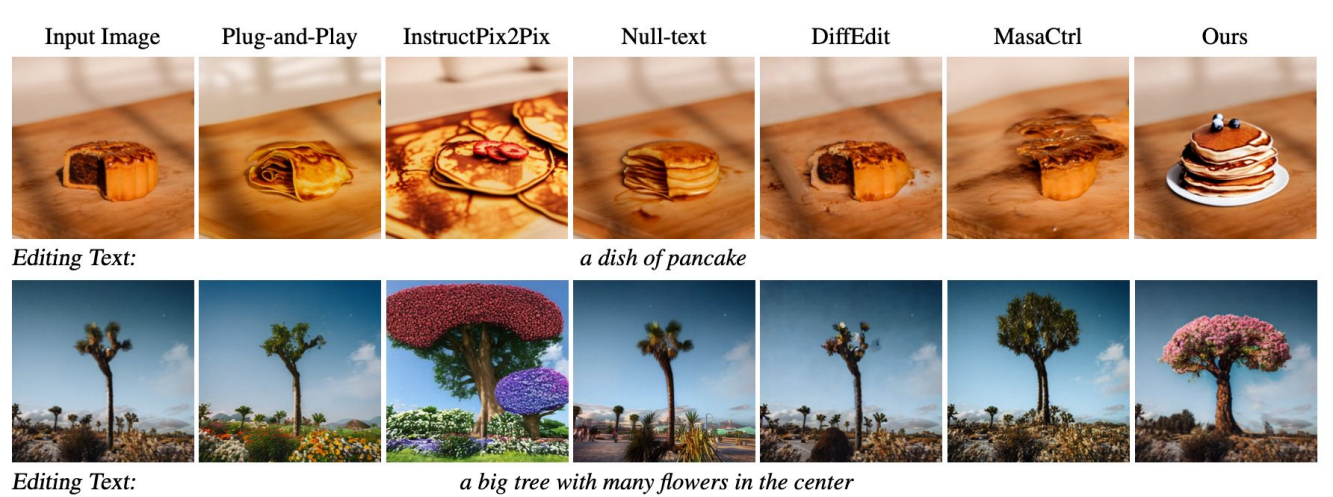
- Patch-based Region Optimization:
  - Chia hình ảnh thành các vùng nhỏ (patches) có thể học được.
  - Sử dụng mô hình vision-language như CLIP để ánh xạ prompt vào không gian hình ảnh.

- Multi-scale Refinement:
  - Phân tích hình ảnh ở nhiều cấp độ: từ đặc trưng cấp thấp (màu sắc, kết cấu) đến cấp cao (ánh sáng, bố cục).
  - Tích hợp mạng GAN (Generative Adversarial Network) để đảm bảo tính tự nhiên và loại bỏ artifacts.



Nội dung và Phương pháp

- Đánh giá hiệu quả hệ thống:
  - Đánh giá tự động: Sử dụng các chỉ số như SSIM để đo độ tương đồng giữa hình ảnh chỉnh sửa và hình ảnh mong muốn.
  - Đánh giá thủ công: Khảo sát người dùng về mức độ phù hợp với prompt và tính tự nhiên của hình ảnh.
  - So sánh với các phương pháp hiện tại như DALL-E và Stable Diffusion.



Kết quả dự kiến

- Hệ thống chỉnh sửa hình ảnh dựa trên văn bản:
  - Định vị chính xác vùng cần chỉnh sửa.
  - Thực hiện chỉnh sửa tự nhiên, chân thực.
  - Đáp ứng các yêu cầu từ đơn giản đến phức tạp.
- Giao diện ứng dụng web:
  - Hỗ trợ tải lên hình ảnh, nhập lệnh chỉnh sửa, xem trước kết quả và tải xuống hình ảnh chỉnh sửa.
  - Thân thiện với người dùng, hỗ trợ đa dạng loại hình ảnh và tình huống chỉnh sửa.
- Đóng góp:
  - Nâng cao chất lượng và tính ứng dụng của text-driven image editing.
  - Mở rộng khả năng ứng dụng trong thiết kế sáng tạo, sản xuất nội dung số, truyền thông, quảng cáo.

Tài liệu tham khảo

- [1] Yuanze Lin, Yi-Wen Chen, Yi-Hsuan Tsai, Lu Jiang, Ming-Hsuan Yang: Text-Driven Image Editing via Learnable Regions. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024, pp. 1–13.
- [2] Patrick Esser, Robin Rombach, Björn Ommer: Taming Transformers for High-Resolution Image Synthesis. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 12873–12883.
- [3] Jonathan Ho, Ajay Jain, Pieter Abbeel: Denoising Diffusion Probabilistic Models. Advances in Neural Information Processing Systems (NeurIPS), 2020, pp. 6840–6851.
- [4] Ruiqi Gao, Xiaoyong Shen, Jiaya Jia: MaskGIT: Masked Generative Image Transformer. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 11315–11325.
- [5] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever: Learning Transferable Visual Models From Natural Language Supervision. Proceedings of the 38th International Conference on Machine Learning (ICML), 2021, pp. 8748–8763.