

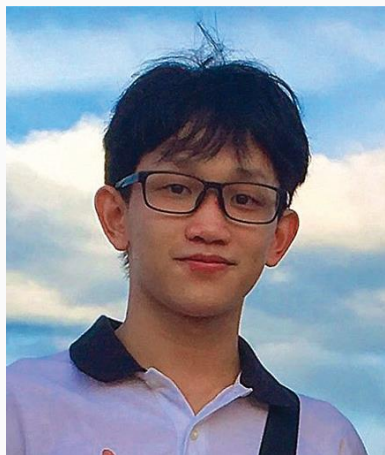
CƠ CHẾ CHÚ Ý DÀI KHÔNG CẦN TINH CHỈNH CHO BÀI TOÁN TỔNG HỢP VÀ CHỈNH SỬA HÌNH ẢNH NHẤT QUÁN

Phạm Văn Hùng - 21522124

Nguyễn Tiến Đạt - 21520707

Tóm tắt

- Lớp: CS519.011
- Link Github của nhóm: <https://github.com/HungStark/CS519.011/>
- Link YouTube video: https://youtu.be/7qQL_j7FIKM



Phạm Văn Hùng



Nguyễn Tiến Đạt

Giới thiệu

Sinh ảnh từ
văn bản phát
triển mạnh

Vấn đề về tổng hợp và
chỉnh sửa hình ảnh
nhất quán

Mô hình Tuning-Free Mutual
Self-Attention ra đời nhưng độ
phức tạp còn lớn

Làm thế nào để giảm độ
phức tạp của mô hình?



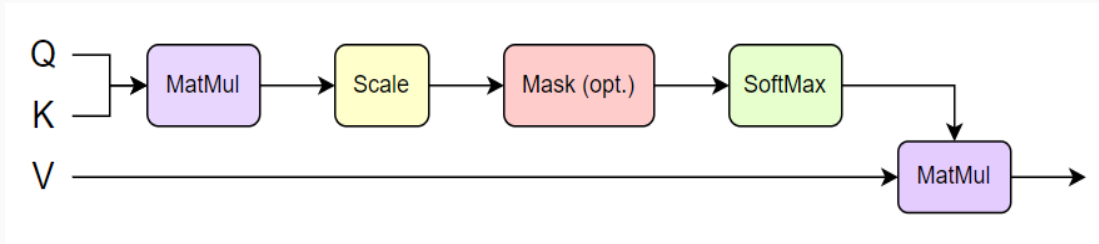
Sử dụng **Cơ chế chú ý dải (Strip Attention)**
thay cho Cơ chế tự chú ý (Self-Attention)

Mục tiêu

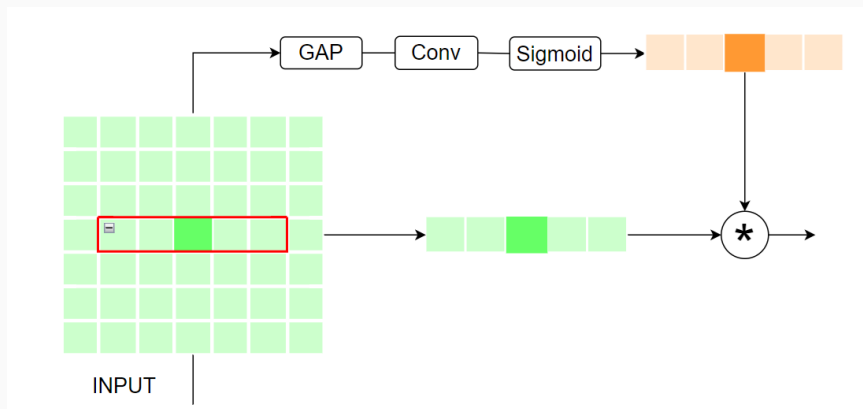
- Xây dựng được mô hình mới cho bài toán tổng hợp và chỉnh sửa hình ảnh nhất quán sử dụng Tuning-Free Strip Attention Control.
- Mô hình mới cải thiện về tốc độ nhưng không quá suy giảm về tính chính xác.
- Đánh giá được mô hình dựa trên các yếu tố tốc độ, độ chân thực, sự ổn định và khả năng tinh chỉnh.

Nội dung và Phương pháp

1. Nội dung



Cơ chế Self-Attention



Cơ chế Strip Attention

Nội dung và Phương pháp

2. Phương pháp

- Tìm hiểu về mô hình MasaCtrl và cơ chế chú ý dài.
- Fine-tuning các siêu tham số của cơ chế chú ý dài.
- Tìm hiểu thuật toán tuning-free và cơ chế tự chú ý chung.
- Nghiên cứu một thuật toán tuning-free mới và cơ chế chú ý chung mới đối với cơ chế chú ý dài.

Nội dung và Phương pháp

2. Phương pháp

- Chuẩn bị và thu thập các tập dữ liệu.
- Tiến hành cài đặt các thuật toán và cơ chế đã được nghiên cứu
- Đánh giá và so sánh với các mô hình trước đó về thời gian chạy, độ chân thực, sự ổn định, khả năng tinh chỉnh.

Kết quả dự kiến

- Mô hình mới chạy nhanh hơn ít nhất 10 lần so với mô hình MasaCtrl gốc.
- Độ chính xác và nhất quán suy giảm không quá 5%.
- Báo cáo về kĩ thuật, các bảng so sánh, đánh giá kết quả thực nghiệm.
- Bảng thống kê các trường hợp lỗi:
 - Trạng thái của đối tượng không chính xác.
 - Ngoại trang của đối tượng bị thay đổi.
 - Hình ảnh của đối tượng được sinh ra không chân thực.

Tài liệu tham khảo

- [1] Mingdeng Cao, Xintao Wang, Zhongang Qi, Ying Shan, Xiaohu Qie, Yinqiang Zheng: MasaCtrl: Tuning-Free Mutual Self-Attention Control for Consistent Image Synthesis and Editing. ICCV 2023: 22503-22513
- [2] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaolei Huang, Xiaogang Wang, Dimitris N. Metaxas: StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks. CoRR abs/1612.03242 (2016)
- [3] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer: High-Resolution Image Synthesis with Latent Diffusion Models. CoRR abs/2112.10752 (2021)
- [4] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, Daniel Cohen-Or: Prompt-to-Prompt Image Editing with Cross-Attention Control. ICLR 2023
- [5] Yuning Cui, Yi Tao, Luoxi Jing, Alois Knoll: Strip Attention for Image Restoration. IJCAI 2023: 645-653
- [6] Chong Mou, Xintao Wang, Liangbin Xie, Jian Zhang, Zhongang Qi, Ying Shan, Xiaohu Qie: T2I-Adapter: Learning Adapters to Dig out More Controllable Ability for Text-to-Image Diffusion Models. CoRR abs/2302.08453 (2023)