

Cơ chế chú ý dải không cần tinh chỉnh cho bài toán tổng hợp và chỉnh sửa hình ảnh nhất quán

Nguyễn Tiến Đạt¹ and Phạm Văn Hùng¹

¹ Trường ĐH Công nghệ Thông tin, ĐHQG TP.HCM

What ?

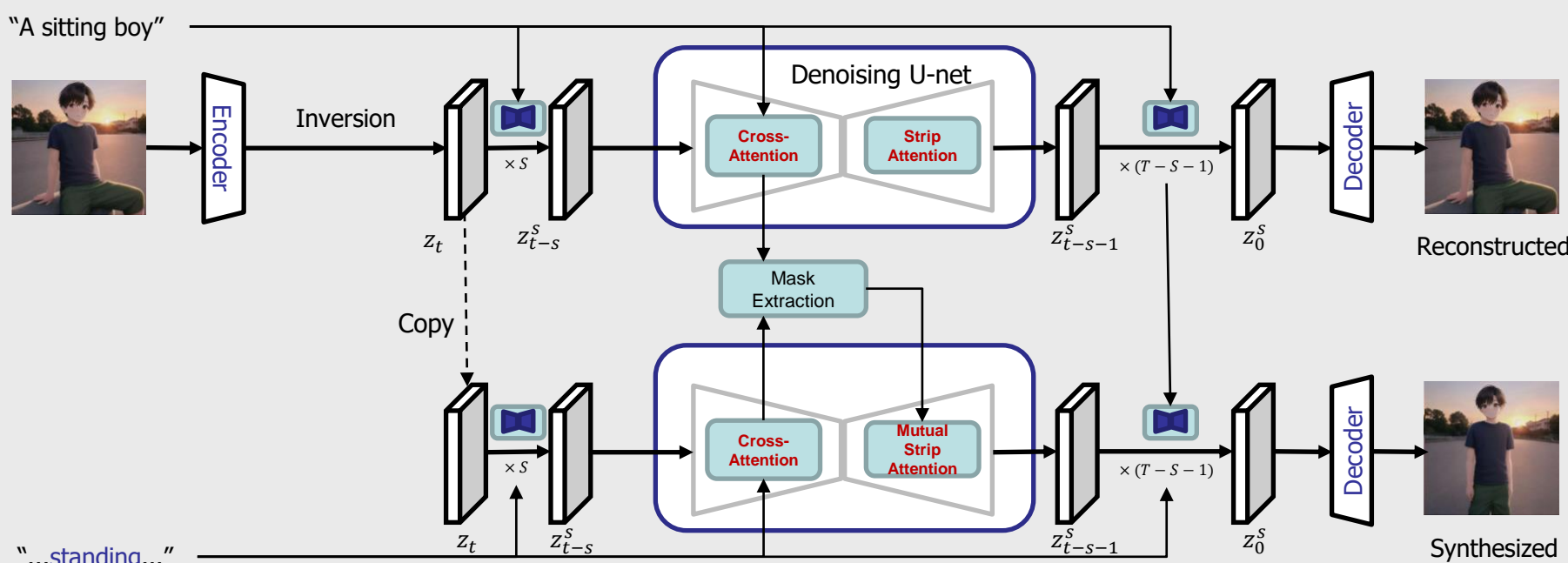
Bài toán **tổng hợp và chỉnh sửa hình ảnh nhất quán** (consistent image synthesis and editing) là bài toán sinh ra nhiều ảnh của cùng một đối tượng/nhân vật/... nhưng với những tư thế, trạng thái khác nhau.

Sử dụng **cơ chế chú ý dải** (Strip Attention) cho phép cải tiến mô hình tổng hợp và chỉnh sửa hình ảnh nhất quán có sẵn hiện nay.

Why ?

- Các mô hình sinh ảnh từ văn bản đã đạt được những bước tiến lớn, tuy nhiên **vẫn còn khoảng cách rất xa giữa nhu cầu và phương pháp** để sinh ra và chỉnh sửa hình ảnh nhất quán.
- Để giải quyết vấn đề trên, người ta đề xuất mô hình MasaCtrl nhưng **độ phức tạp bậc hai của cơ chế tự chú ý** đã khiến cho mô hình này gặp khó khăn đối với ảnh có kích thước lớn.

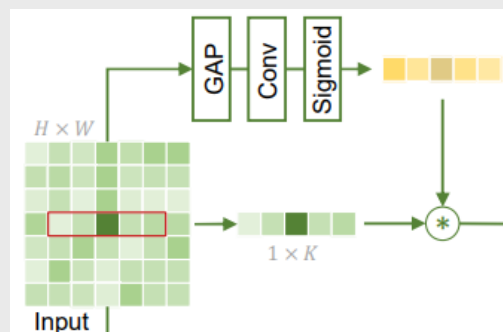
Overview: MasaCtrl w/ Strip Attention



Description

1. Nội dung nghiên cứu

- Đặc điểm của Strip Attention: tính mức Attention bằng cách cho dải pixel đi qua các lớp Global Average Pooling, lớp tích chập 1x1 và hàm Sigmoid thay vì tính các ma trận Query, Key, Value



- Cần thay đổi cơ chế chú ý trên MasaCtrl thành cơ chế chú ý dải, tuy nhiên cần phải nghiên cứu và điều chỉnh các thuật toán tuning-free và cơ chế chú ý chung (mutual attention).

2. Phương pháp nghiên cứu

- Tìm hiểu và cài đặt lại mô hình MasaCtrl.
- Tìm hiểu và cài đặt cơ chế chú ý dải trên mô hình MasaCtrl.
- Fine-tuning các siêu tham số của cơ chế chú ý dải trên MasaCtrl.
- Nghiên cứu một thuật toán tuning-free mới và cơ chế chú ý chung mới đối với cơ chế chú ý dải trên MasaCtrl.
- Chuẩn bị và thu thập các tập dữ liệu để huấn luyện và đánh giá mô hình.
- Tiến hành cài đặt các thuật toán và cơ chế đã được nghiên cứu, sau đó đánh giá và so sánh với các mô hình khác về độ chính xác của mô hình, thời gian chạy mô hình và phân tích các trường hợp lỗi từ đó đề xuất hướng phát triển cho tương lai.

3. Kết quả mong đợi

- Mô hình xây dựng được giúp khắc phục nhược điểm của các mô hình trước đó về độ phức tạp, đồng thời chạy nhanh hơn ít nhất 10 lần so với mô hình MasaCtrl và các mô hình khác.
- Mô hình đảm bảo được độ chính xác và độ nhất quán của mô hình không bị suy giảm đáng kể (dưới 5%).
- Báo cáo về kĩ thuật, các bảng so sánh, đánh giá kết quả thực nghiệm giữa mô hình xây dựng được và các mô hình khác trên cùng bài toán.
- Bảng thống kê các trường hợp lỗi.