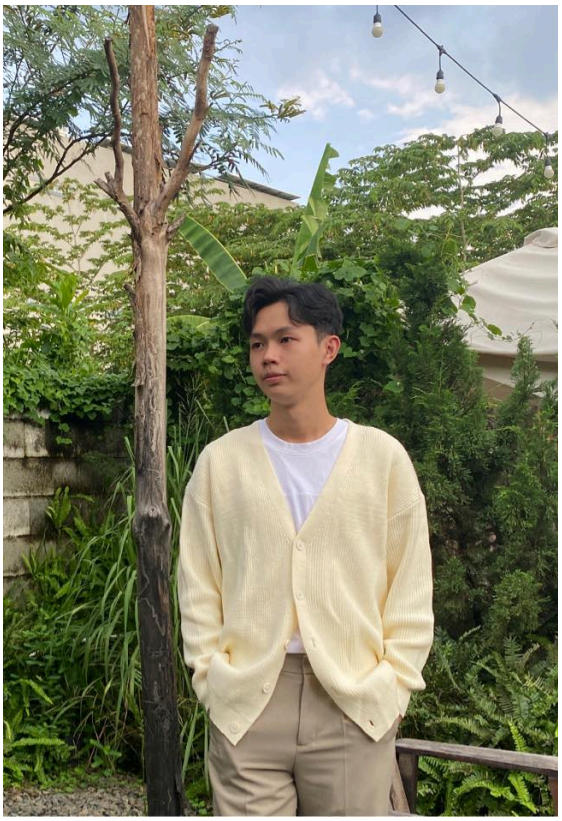
THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút):
  (ví dụ: https://www.youtube.com/watch?v=AWq7uw-36Ng)
- Link slides (dạng .pdf đặt trên Github của nhóm):
  https://github.com/Taidvt/CS2205.CH183/blob/main/DamVuTrongTai%20-%20CS2205.NOV2024.DeCuong.FinalReport.Template.Slide.pdf

- Họ và Tên: Đàm Vũ Trọng Tài
- MSSV: 240101023



- Lớp: CS2205.CH183
- Tự đánh giá (điểm tổng kết môn): 9.5/10
- Số buổi vắng: 0
- Số câu hỏi QT cá nhân: 10
- Số câu hỏi QT của cả nhóm: 10
- Link Github:
  https://github.com/Taidvt/CS2205.CH183/

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

BẢO TOÀN ĐỊNH DANH TRONG BÀI TOÁN TẠO SINH VIDEO NHÂN VẬT DỰA TRÊN TƯ THẾ

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

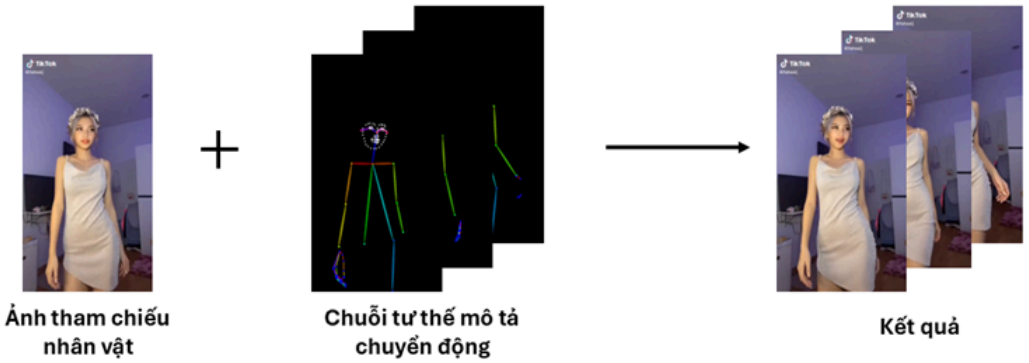
IDENTITY PRESERVATION FOR HUMAN VIDEO GENERATION

TÓM TẮT *(Tối đa 400 từ)*

Đề tài "Bảo Toàn Định Danh Trong Bài Toán Tạo Sinh Video Nhân Vật Dựa Trên Tư Thế" tập trung vào việc giải quyết vấn đề bảo toàn định danh trong quá trình tạo sinh video nhân vật từ ảnh tham chiếu và chuỗi tư thế, đặc biệt là trong bối cảnh các chuyển động phức tạp và thay đổi góc nhìn. Mặc dù các phương pháp hiện tại, như MagicPose, AnimateAnyone, DisCo và MagicAnimate, đã đạt được những thành tựu nhất định trong việc tạo sinh video nhân vật, nhưng vẫn gặp phải hạn chế lớn trong việc duy trì tính nhất quán về đặc trưng nhân vật qua các khung hình. Các mô hình này thường chỉ chú trọng đến các đặc trưng lớn của cơ thể, bỏ qua những chi tiết quan trọng như khuôn mặt, trang phục hay phụ kiện. Để khắc phục những hạn chế này, nghiên cứu đề xuất phương pháp kết hợp mô hình khuếch tán tiềm ẩn đã được tinh chỉnh từ ảnh, kết hợp với các cơ chế hướng dẫn bổ sung nhằm tăng cường khả năng bảo toàn định danh trong suốt quá trình tạo sinh. Phương pháp mới này sẽ bao gồm việc khảo sát và phân tích các phương pháp hiện tại, đề xuất các cải tiến về mô hình để duy trì các đặc trưng nhận diện quan trọng như khuôn mặt, tỉ lệ cơ thể, họa tiết trang phục, và thực hiện các thử nghiệm trên các bộ dữ liệu đa dạng như video từ TikTok để đánh giá hiệu quả. Kết quả dự kiến sẽ là một phương pháp cải thiện đáng kể khả năng bảo toàn định danh so với các mô hình hiện nay, từ đó mở ra khả năng ứng dụng trong các lĩnh vực như sản xuất điện ảnh, trò chơi điện tử và giải trí. Nghiên cứu này cũng dự kiến công bố ít nhất một bài báo khoa học quốc tế về phương pháp đề xuất.

GIỚI THIỆU *(Tối đa 1 trang A4)*

Lĩnh vực tạo sinh video nhân vật đang phát triển mạnh nhờ tiềm năng ứng dụng trong điện ảnh, AR/VR và giải trí. Các phương pháp hiện tại thường tách biệt hai mô đun: (1) trích xuất đặc trưng ngoại hình từ ảnh tham chiếu và (2) trích xuất đặc trưng chuyển động từ chuỗi tư thế, sau đó kết hợp chúng qua mô hình khuếch tán tiềm ẩn (LDM) để sinh video.



Định nghĩa bài toán:

- Đầu vào: Ảnh tham chiếu của nhân vật và chuỗi tư thế mô tả chuyển động.
- Đầu ra: Video nhân vật thực hiện theo chuyển động cho trước.

Lý do chọn đề tài: Công nghệ tạo sinh video nhân vật đang đóng vai trò quan trọng trong việc nâng cao trải nghiệm người dùng nhờ khả năng tái tạo chuyển động và ngoại hình của chủ thể một cách trực quan. Tuy nhiên, một trong những thách thức lớn hiện nay là đảm bảo tính định danh của nhân vật theo ảnh tham chiếu và duy trì sự nhất quán theo thời gian trong suốt quá trình tạo sinh. Hiện đã có một số phương pháp tiếp cận nhằm giải quyết vấn đề này [1, 3], nhưng các phương pháp hiện tại vẫn chưa mang lại kết quả ổn định và triệt để khi đối mặt với các chuyển động phức tạp. Việc giải quyết vấn đề này sẽ mở ra khả năng ứng dụng vào thế giới thực khi có thể sinh ra các kết quả thực tế.

MỤC TIÊU *(Viết trong vòng 3 mục tiêu)*

1. Khảo sát và phân tích các phương pháp tiên tiến cho tạo sinh video nhân vật dựa trên tư thế.

2. Đề xuất và đánh giá phương pháp cải thiện sự bảo toàn định danh trong bài toán tạo sinh video nhân vật dựa trên tư thế.
3. Mở rộng thực nghiệm trên các miền dữ liệu khác nhau

NỘI DUNG VÀ PHƯƠNG PHÁP

Nội dung 1: Khảo sát phương pháp tạo sinh video nhân vật dựa trên tư thế

Mục tiêu: Đánh giá các mô hình khuếch tán hiện đại trong tạo sinh video [1-4], tập trung vào kiến trúc, khả năng biểu đạt và hạn chế về tính nhất quán thời gian/bảo toàn định danh.

Phương pháp:

- Triển khai lại các mô hình tiên tiến (tinh chỉnh kiến trúc, cập nhật khái niệm mới).
- Thử nghiệm trên dữ liệu đa dạng, so sánh hiệu suất qua chất lượng hình ảnh, độ đồng nhất thời gian, và bảo toàn định danh

Nội dung 2: Đề xuất phương pháp bảo toàn định danh

Phát hiện: Các phương pháp dựa trên hàm khử nhiễu [1-4] thiếu cơ chế rõ ràng để duy trì đặc trưng định danh (nét mặt, trang phục), gây biến đổi không nhất quán.

Giải pháp:

Tích hợp cơ chế hướng dẫn bổ sung (ví dụ: hàm mục tiêu phụ) vào mô hình khuếch tán, tập trung vào đặc trưng quan trọng.

Thực nghiệm:

Đánh giá trên bộ dữ liệu TikTok, đo lường qua chất lượng hình ảnh, tính nhất quán, và khả năng bảo toàn định danh.

Nội dung 3: Mở rộng thực nghiệm đa miền dữ liệu

Mục tiêu: Đánh giá khả năng thích ứng và tổng quát hóa của mô hình trên các miền dữ liệu khác nhau (ví dụ: hoạt hình).

Phương pháp:

- Phân tích đặc thù từng miền (hình thái, màu sắc, chi tiết).

- Tinh chỉnh mô hình để tăng khả năng mô phỏng đa dạng ngoại hình.

Thực nghiệm:

Kiểm tra tính ổn định của định danh nhân vật trong nhiều ngữ cảnh/chuyển động.

KẾT QUẢ MONG ĐỢI

- Kết quả định tính và định lượng so sánh phương pháp đề xuất và các phương pháp hiện có.
- Công bố một bài báo khoa học được chấp nhận trên tạp chí/hội nghị quốc tế uy tín thuộc danh mục Scopus.
- Bản báo cáo luận văn Thạc sĩ với bài báo đã được công bố.

TÀI LIỆU THAM KHẢO *(Định dạng DBLP)*

[1] Chang, Di, et al. "MagicPose: Realistic Human Poses and Facial Expressions Retargeting with Identity-aware Diffusion." Forty-first International Conference on Machine Learning. 2023.

[2] Hu, Li. "Animate anyone: Consistent and controllable image-to-video synthesis for character animation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.

[3] Wang, Tan, et al. "Disco: Disentangled control for referring human dance generation in real world." arXiv e-prints (2023): arXiv-2307.

[4] Xu, Zhongcong, et al. "Magicanimate: Temporally consistent human image animation using diffusion model." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.