

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

Đàm Vũ Trọng Tài - 24010123

Tóm tắt

- Lớp: CS2205.CH183
- Link Github của nhóm: [github](#)
- Link YouTube video: [presentation](#)
- Họ tên: Đàm Vũ Trọng Tài

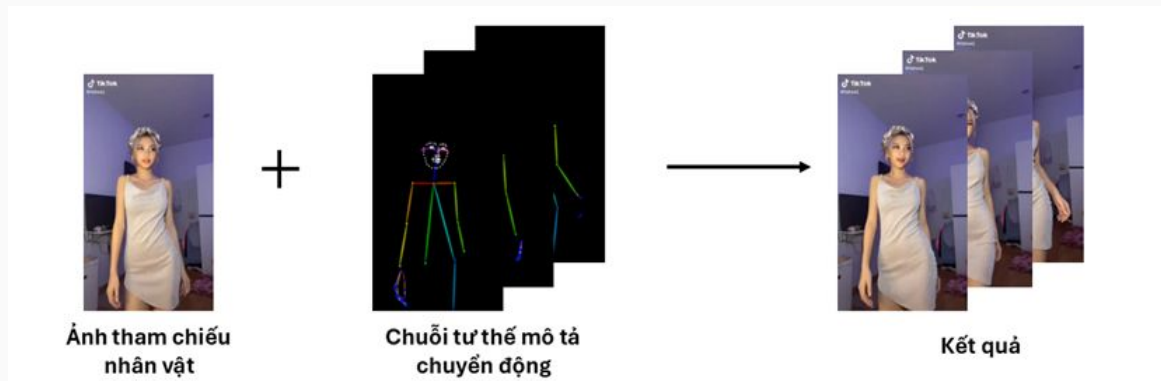


Giới thiệu

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 tả ngoại hình nhân vật + chuỗi tư thế mô tả chuyển động.
- **Đầu ra:** Video nhân vật thực hiện chuyển động mong muốn.



Giới thiệu

Lý do chọn đề tài:

Dù đã có nhiều nỗ lực cải thiện, các phương pháp hiện tại vẫn chưa giải quyết triệt để thách thức về bảo toàn định danh (ID) của nhân vật và tính nhất quán khi xử lý chuyển động phức tạp. Nghiên cứu này tập trung nâng cao khả năng bảo toàn ID từ ảnh tham chiếu và ổn định kết quả theo thời gian, hướng đến ứng dụng thực tế như sản xuất nội dung số hay tương tác đa phương tiện.

Mục tiêu

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

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 và Phương pháp

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 dự kiến

- 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

- [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.