TẠO SINH VIDEO VẬT THỂ THỰC HIỆN HÀNH ĐỘNG CỦA CON NGƯỜI DỰA TRÊN CHỈ DẪN VĂN BẢN VỚI ĐA RÀNG BUỘC

Phạm Thị Bích Nga - 240101018

Tóm tắt

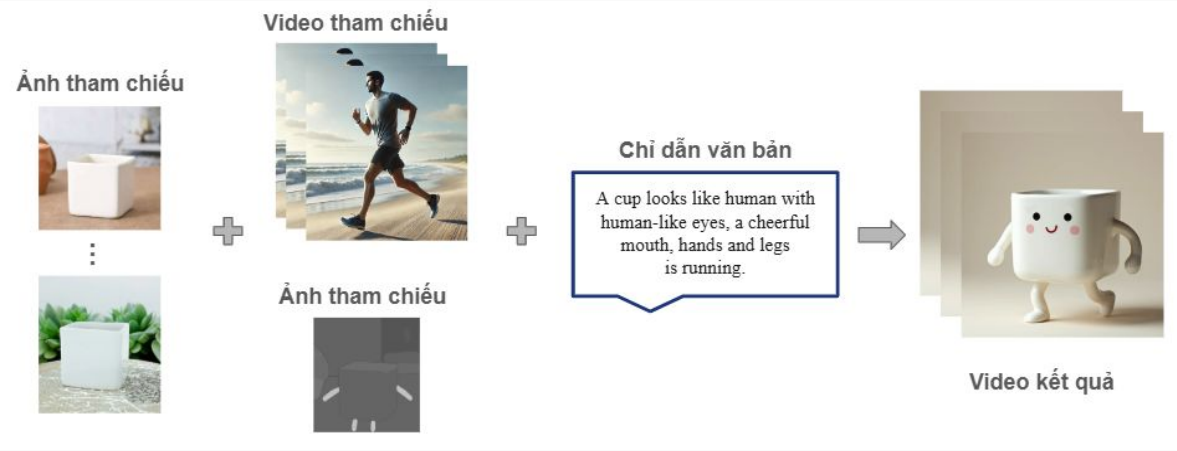
- Lớp: CS2205.CH183
- Link Youtube: youtube_vid
- Link Github: github_link
- Họ tên: Phạm Thị Bích Nga

Giới thiệu

- Bài toán **tạo sinh video nhân hóa vật thể thực hiện hành động của con người** nhằm mô phỏng vật không phải người nhưng có hình dáng, cử chỉ và chuyển động giống con người
- Đây là một bài toán mới, ứng dụng tiềm năng trong các lĩnh vực phim ảnh, quảng cáo và hoạt hình.
- Thách thức: mô hình tạo sinh hiện tại chưa diễn giải tốt các loại chỉ dẫn văn bản sáng tạo như “nhân hóa”

  => Bổ sung thêm các ràng buộc để cung cấp thêm thông tin ngữ nghĩa cho mô hình

  => Tạo sinh video vật thể thực hiện hành động của con người dựa trên chỉ dẫn văn bản và đa ràng buộc.



Mục tiêu

- Khảo sát & phân tích
  - Nghiên cứu các mô hình tạo sinh ảnh/video hiện có.
  - Đánh giá khả năng nhân hóa vật thể & tích hợp đa ràng buộc.
- Đề xuất phương pháp mới
  - Mô hình tạo sinh video nhân hóa vật thể kết hợp đa ràng buộc
- Thực nghiệm & đánh giá trên các tiêu chí
  - Chất lượng hình ảnh.
  - Tuân thủ ràng buộc đầu vào.
  - Độ mượt mà của chuyển động.
  - So sánh với các phương pháp hiện có.

Nội dung và Phương pháp

Nội dung 1: Khảo sát và phân tích ưu/nhược điểm của các mô hình sinh ảnh và video kết hợp đa ràng buộc

Mục tiêu:

- Nắm được kiến trúc và nguyên lý hoạt động của Diffusion Models và các mô hình sinh video dựa trên nó.
- Đánh giá các phương pháp tạo sinh video với đa ràng buộc.
- Phân tích mô-đun học đặc trưng và tích hợp ràng buộc.

Phương pháp:

- **Nghiên cứu lý thuyết** về Diffusion Models, ứng dụng trong sinh ảnh & video.
- **Khảo sát và đánh giá ưu/nhược điểm của các phương pháp:** Tinh chỉnh mô hình sinh ảnh để tạo video [1-2] và mô hình sinh video [3] trực tiếp từ chỉ dẫn văn bản.
- **Phân tích mô hình học đặc trưng và tích hợp ràng buộc**

Nội dung và Phương pháp

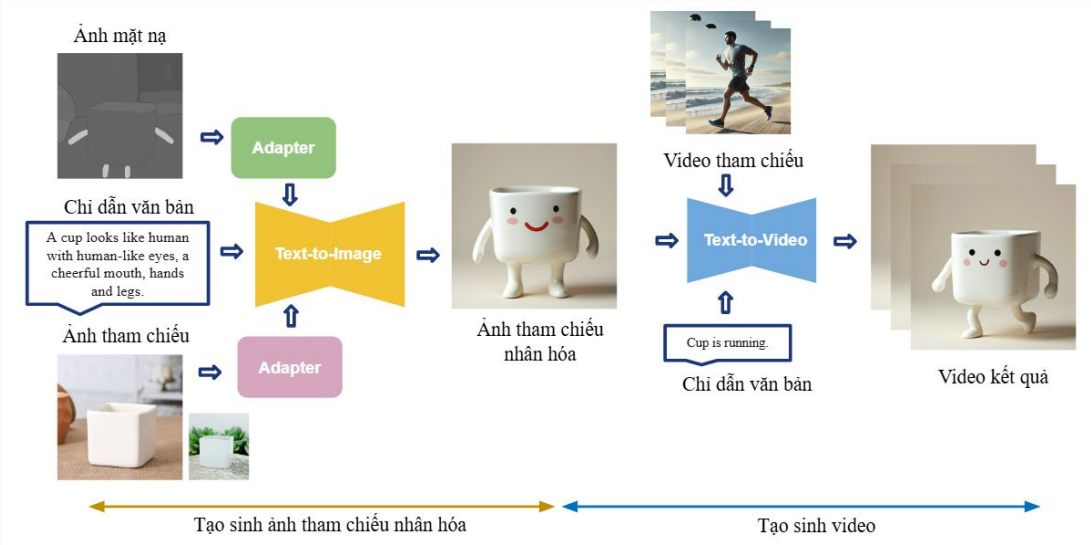
Nội dung 2: Đề xuất phương pháp tạo sinh video nhân hóa vật thể

Mục tiêu:

- Phát triển mô hình tạo sinh video nhân hóa vật thể kết hợp đa ràng buộc nhằm cung cấp thông tin cải thiện khả năng sinh ảnh và chuyển động của vật thể.

Phương pháp:

- Thiết kế kiến trúc mô hình tạo sinh có thể tích hợp các ràng buộc đa dạng
- Xây dựng mã nguồn và chuẩn bị dữ liệu thử nghiệm



Nội dung và Phương pháp

Nội dung 3: Thực nghiệm & đánh giá phương pháp đề xuất

Mục tiêu:

- Đánh giá hiệu quả dựa trên các tiêu chí:
  - chất lượng hình ảnh
  - mức độ tuân thủ ràng buộc đầu vào,
  - độ mượt mà của chuyển động.
- So sánh với các phương pháp hiện có.

Phương pháp:

- Thực nghiệm trên bộ dữ liệu đa dạng
- Đánh giá chất lượng và mức độ tuân thủ ràng buộc
- So sánh với phương pháp hiện có

Kết quả dự kiến

- **Tổng hợp đánh giá & phân tích**
  - Ưu/nhược điểm của các mô hình tạo sinh ảnh/video tích hợp đa ràng buộc.
- **Mô hình tạo sinh video nhân hóa vật thể**
  - Tích hợp hiệu quả các ràng buộc, cải thiện khả năng tạo sinh, video chuyển động mượt mà.
  - Đảm bảo tuân thủ ràng buộc đầu vào.
- **Kết quả đánh giá và phân tích mô hình được đề xuất**
- **Xây dựng bộ dữ liệu thử nghiệm & mã nguồn**
  - Hỗ trợ nghiên cứu và phát triển tiếp theo.
- **Công bố kết quả nghiên cứu**

Tài liệu tham khảo

[1]. Jay Zhangjie Wu, Yixiao Ge, Xintao Wang, Stan Weixian Lei, Yuchao Gu, Yufei Shi, Wynne Hsu, Ying Shan, Xiaohu Qie, Mike Zheng Shou: Tune-A-Video: One-Shot Tuning of Image Diffusion Models for Text-to-Video Generation. ICCV 2023: 7589-7599

[2]. Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyuan Hu, Harry Yang, Oron Ashual, Oran Gafni, Devi Parikh, Sonal Gupta, Yaniv Taigman: Make-A-Video: Text-to-Video Generation without Text-Video Data. ICLR 2023

[3]. Spencer Sterling. Zeroscope. https://huggingface.co/cerspense/zeroscope_v2_576w, 2023

[4]. Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit Haim Bermano, Gal Chechik, Daniel Cohen-Or: An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion. ICLR 2023

[5]. Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, Kfir Aberman: DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. CVPR 2023: 22500-22510