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 [1,2]

[1] Trường Đại Học Công Nghệ Thông Tin, ĐHQG HCM
[2] Đại học Quốc gia TPHCM, Việt Nam

What ?

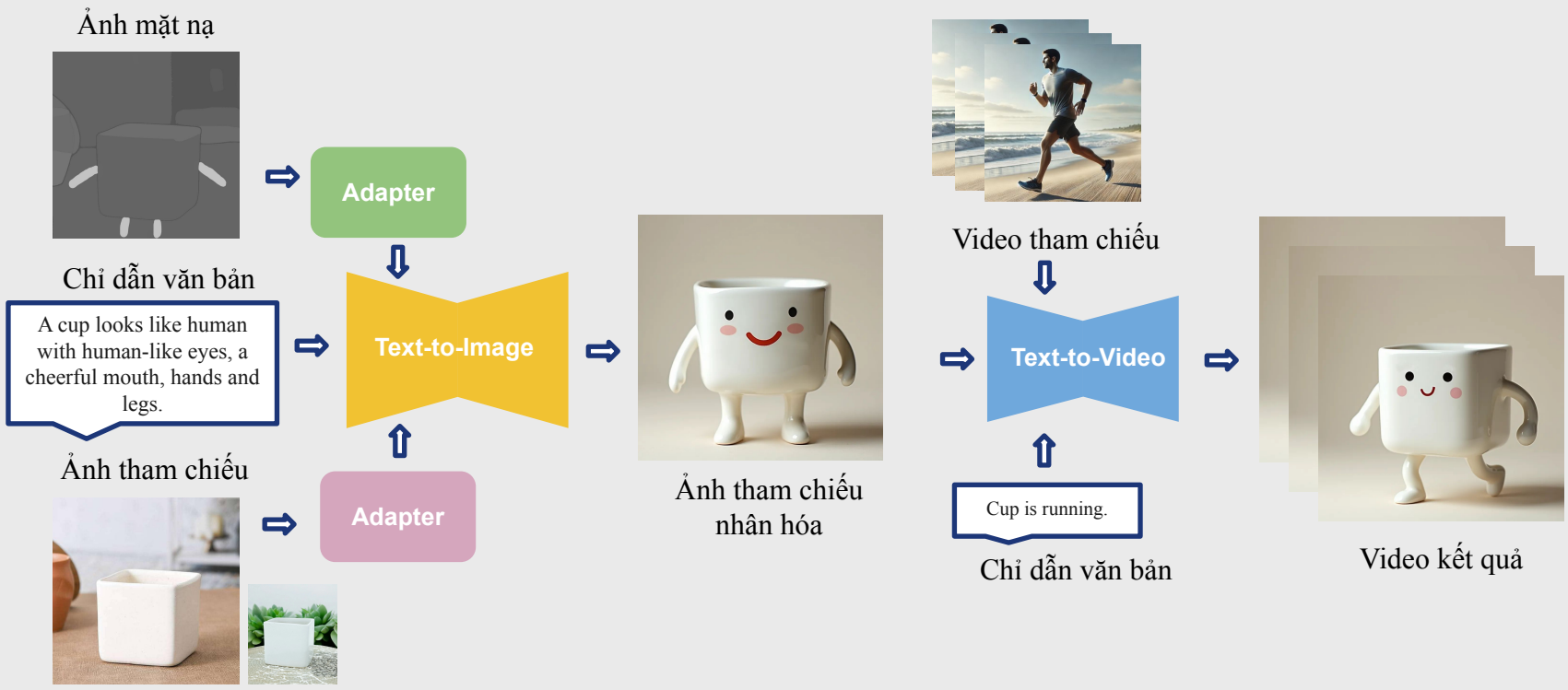
Đề xuất phương pháp tạo sinh video nhân hóa vật thể kết hợp đa ràng buộc:

- Tích hợp ràng buộc mặt nạ phân đoạn ngữ nghĩa để mô hình hiểu và định vị chính xác các bộ phận như tay, chân, bổ trợ thông tin cho chỉ dẫn văn bản.
- Mô hình hai giai đoạn tạo ảnh tham chiếu nhân hóa trước khi sinh video, giúp cải thiện tính chính xác và nhất quán của hình dạng vật thể.

Why ?

- Do mô hình sinh ảnh và video hiện tại chưa thể diễn giải tốt các chỉ dẫn sáng tạo, đặc biệt khi nhân hóa vật thể, nên mặt nạ phân đoạn giúp xác định rõ tay, chân, cải thiện mô phỏng hình dáng và chuyển động.
- Do mô hình sinh ảnh và video hiện tại chưa thể diễn giải tốt các chỉ dẫn sáng tạo, đặc biệt khi nhân hóa vật thể, nên mặt nạ phân đoạn giúp xác định rõ tay, chân, cải thiện mô phỏng hình dáng và chuyển động.

Overview



Description

Giai đoạn 1. Tạo sinh ảnh tham chiếu nhân hóa

Bước 1: Chuẩn bị dữ liệu đầu vào

- *Chỉ dẫn văn bản* mô tả đặc điểm nhân hóa mong muốn.
- *Ảnh tham chiếu của vật thể*
- *Ảnh mặt nạ* phân đoạn ngữ nghĩa giúp xác định vị trí của các bộ phận nhân hóa như tay, chân.

Bước 2: Xử lý đầu vào qua Adapter

- Các thông tin đầu vào được đưa qua *bộ điều hợp (Adapter)* để chuẩn hóa và ánh xạ vào không gian đặc trưng phù hợp.

Bước 3: Sinh ảnh nhân hóa bằng mô hình Text-to-Image

- Mô hình Text-to-Image nhận đầu vào từ Adapter, sử dụng cơ chế kết hợp thông tin từ văn bản, ảnh gốc và ảnh mặt nạ để tạo ảnh nhân hóa có độ chính xác cao.
- *Đầu ra là một ảnh nhân hóa của vật thể* với đặc điểm người như mắt, miệng, tay, chân.

Giai đoạn 2. Tạo sinh video

Bước 4: Chuẩn bị dữ liệu đầu vào cho mô hình sinh video

- *Ảnh tham chiếu nhân hóa* từ giai đoạn 1 giúp đảm bảo sự nhất quán về hình dạng trong suốt video.
- *Chỉ dẫn văn bản* mô tả hành động
- *Video tham chiếu* về hành động mong muốn

Bước 5: Sinh video bằng mô hình Text-to-Video

- Mô hình Text-to-Video sử dụng thông tin từ ảnh tham chiếu, video tham chiếu và chỉ dẫn văn bản để tạo chuỗi khung hình thể hiện chuyển động của vật thể.

Bước 6: Xuất video kết quả

- *Video đầu ra* hiển thị vật thể nhân hóa thực hiện hành động một cách mượt mà và nhất quán với các ràng buộc đầu vào.

NII

Bich-Nga Pham – Trường Đại học Công Nghệ Thông Tin
TEL : 0393942888 Email : ngaptb.19@grad.uit.edu.vn