

TỪ VIDEO SANG PROMPT, KỸ THUẬT CHUYỂN ĐỔI NGƯỢC DÀNH CHO VIỆC XÂY DỰNG CONTEXT TRONG STABLE DIFFUSION

Vũ Ngọc Thạch¹ - 230201054

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

What ?

Đề xuất một mô hình máy học cho phép thực hiện tác vụ chuyển đổi ngược phương pháp nội sinh AI:

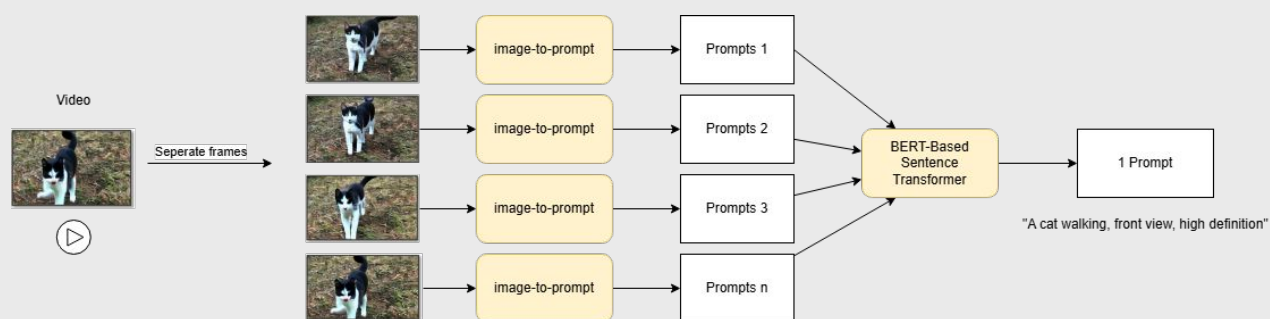
- Dự đoán prompt từ video được sinh ra từ Stable Video Diffusion
- Xây dựng một ứng dụng từ mô hình để xem xét khả năng trong thực tế

Why ?

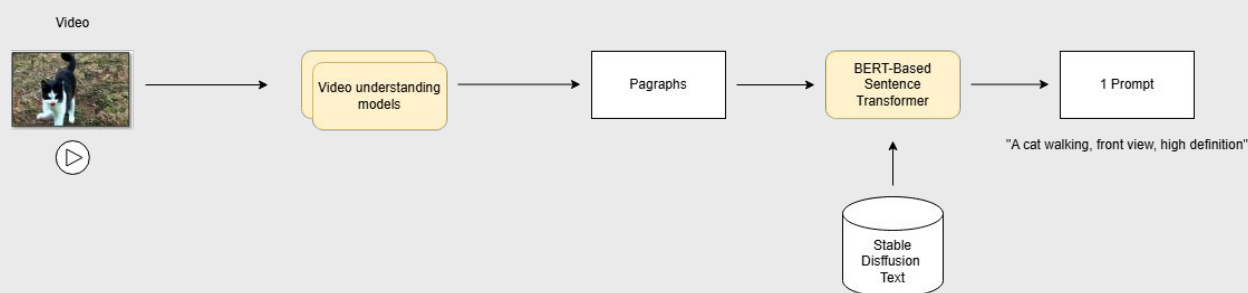
- Người dùng sử dụng AI chuyển đổi từ prompt sang video cần phải học về cách viết prompt và thử rất nhiều lần trước khi ra được 1 video ưng ý
- Nhiều video người dùng không nắm bắt được các prompt chuẩn để tạo nên nó
- Tốn thời gian và chi phí cho mỗi lần thực hiện việc lặp lại như vậy

Overview

Task 1: Train dựa trên mô hình text-to-image hiện có



Task 2: Sử dụng các thuật toán đọc hiểu video



Description

Task 1

1. Phân tích video

- Tách lớp video thành từng khung hình nhỏ hơn dùng cho việc huấn luyện từng khung hình này

2. Dự đoán prompt trong từng khung hình

- Sử dụng mô hình image-to-prompt để dự đoán trên từng khung hình này

3. Tổng hợp kết nối các prompt

- Tổng hợp các prompt theo từng khung hình và sử dụng các mô hình xử lý ngôn ngữ tự nhiên để phân tích, chuẩn hoá lại thành 1 prompt duy nhất thể hiện sự di chuyển của toàn bộ khung hình

Task 2

1. Phân tích nội dung video

- Sử dụng các mô hình đọc hiểu video hiện nay để phân tích, đánh giá nội dung có trong video

2. Chuyển đổi ngữ nghĩa thành prompt

- Kết hợp nội dung ở trên và sử dụng các mô hình ngôn ngữ tự nhiên + tập cơ sở dữ liệu prompt của Stable Diffusion cung cấp để tạo thành 1 prompt hoàn chỉnh