

TEXT-TO-VIDEO WITH DIFFUSION MODEL

Huỳnh Khánh Hòa¹

¹ Trường Đại học Công Nghệ Thông Tin - ĐHQG-TPHCM, Việt Nam

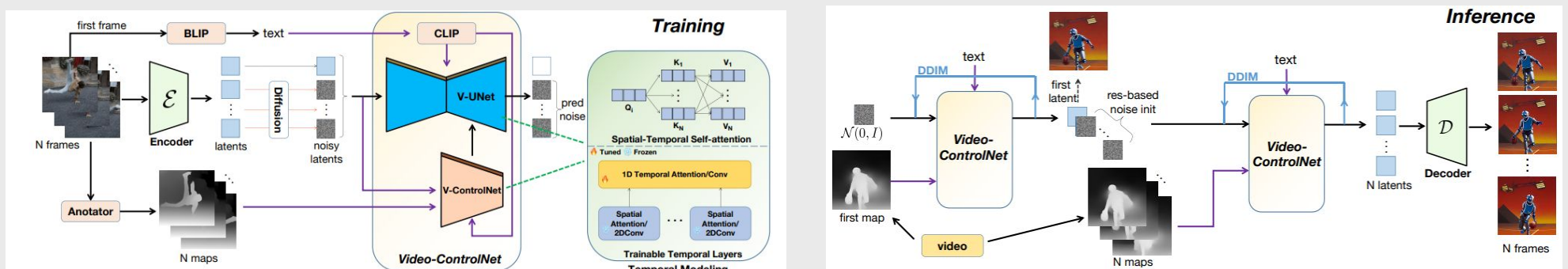
What ?

- Tìm hiểu về mô hình sinh video và diffusion model, thực nghiệm đề xuất phương hướng mới.
- Xây dựng crawler, quy định về tiêu chuẩn dữ liệu, crawl bộ dữ liệu và làm sạch.
- Đánh giá mô hình bằng cả 2 phương pháp định tính và định lượng cùng với các độ đo khác nhau.

Why ?

- Text-to-Video là công nghệ đột phá giúp chuyển đổi văn bản thành video tự động. Nó cải thiện trải nghiệm người dùng, tiết kiệm thời gian và chi phí. Tạo video tự động hấp dẫn trong giáo dục, truyền thông, quảng cáo và giải trí, định hình tương lai của đa phương tiện số.
- Đây là phương hướng có nhiều triển vọng kể từ khi diffusion model thành công trong lĩnh vực sinh ảnh, khó khăn cả về mặt tài nguyên tính toán lẫn lý thuyết.

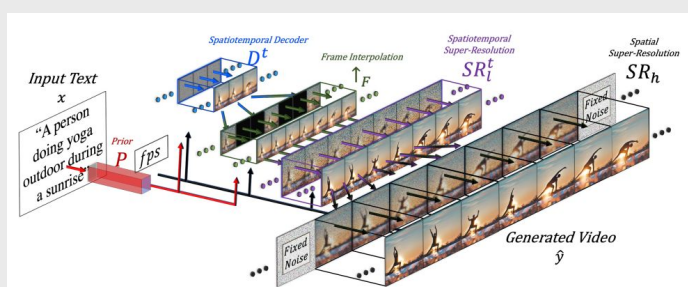
Overview



Description

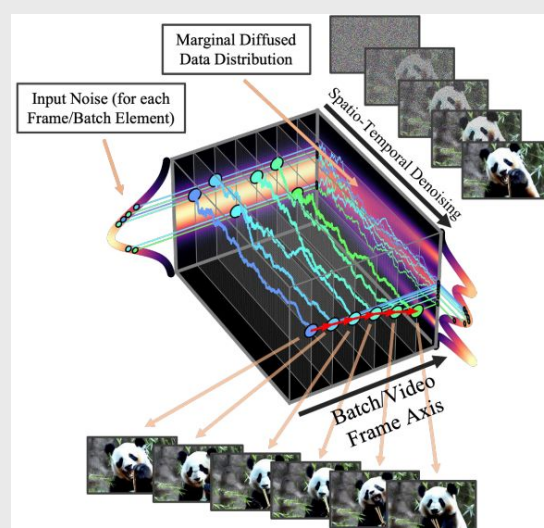
1. Text to Video

- Text to Video (T2V) là một công nghệ trong lĩnh vực Trí tuệ nhân tạo và xử lý ngôn ngữ tự nhiên, cho phép tự động chuyển đổi văn bản thành video. T2V sử dụng các thuật toán và mô hình học máy để tạo video động từ nội dung văn bản, cung cấp nội dung đa phương tiện sinh động và hấp dẫn trong các lĩnh vực như giáo dục, truyền thông, quảng cáo và giải trí.



2. Diffusion model

- Diffusion model là một mô hình toán học và thống kê được sử dụng để mô tả sự lan truyền hoặc phân phối của các hiện tượng, thông tin, hay dữ liệu trong một hệ thống. Mô hình này dựa trên nguyên tắc rằng các yếu tố hoặc đối tượng có xu hướng di chuyển, lan truyền, hoặc phân phối từ vị trí có năng lượng cao tới vị trí có năng lượng thấp, tạo thành sự chênh lệch năng lượng giữa các vùng trong hệ thống.



3. Vài kết quả thực nghiệm trên các paper

