TÊN ĐỀ TÀI - TEXT2VIDEO-ZERO: TEXT TO IMAGE DIFFUSION MODEL ARE ZERO-SHOT VIDEO GENERATORS

Thi Vĩnh Huy - 21252165

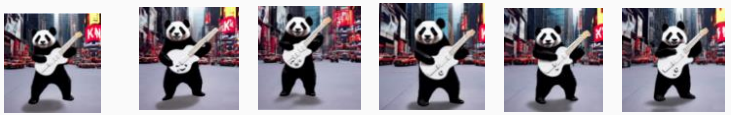
Tóm tắt

- Lớp: CS519.O11
- Link Github : https://github.com/thivinhhuy/CS519.O11
- Link YouTube video:
- Thi Vĩnh Huy - 21522165

Giới thiệu

- Video là một phương tiện truyền thông ngày càng quan trọng trong cuộc sống hiện đại.
- Video được sử dụng trong nhiều ứng dụng khác nhau, chẳng hạn như tin tức, giáo dục, giải trí, tiếp thị và truyền thông xã hội.
- Tuy nhiên, tạo video là một quá trình tốn nhiều thời gian và công sức.

A panda is playing guitar on Times square



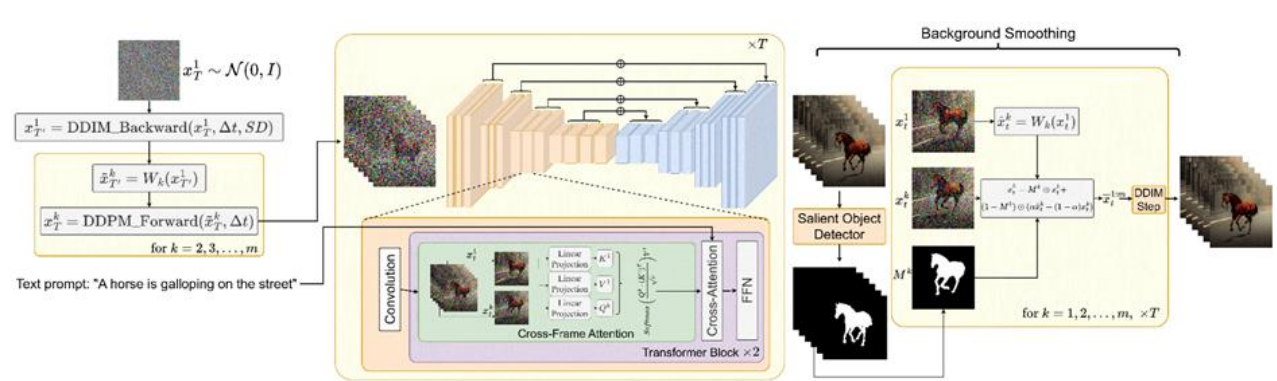
Mục tiêu

- Tạo video từ văn bản không huấn luyện (zero-shot text-to-video).
- Đề xuất một cách tiếp cận tiết kiệm chi phí để giải quyết nhiệm vụ này.
- Đánh giá hiệu suất của cách tiếp cận được đề xuất trên một số bộ dữ liệu video.

Nội dung và Phương pháp

- Sử dụng một mô hình khuếch tán văn bản thành hình ảnh đã được huấn luyện trước để tạo ra một loạt các khung hình video từ mô tả bằng văn bản.
- Áp dụng hai kỹ thuật để đảm bảo tính nhất quán về thời gian của các khung hình:
  - Thêm thông tin chuyển động vào mã tiềm ẩn của các khung hình:
  - Sử dụng sự chú ý giữa các khung hình:

Nội dung và Phương pháp



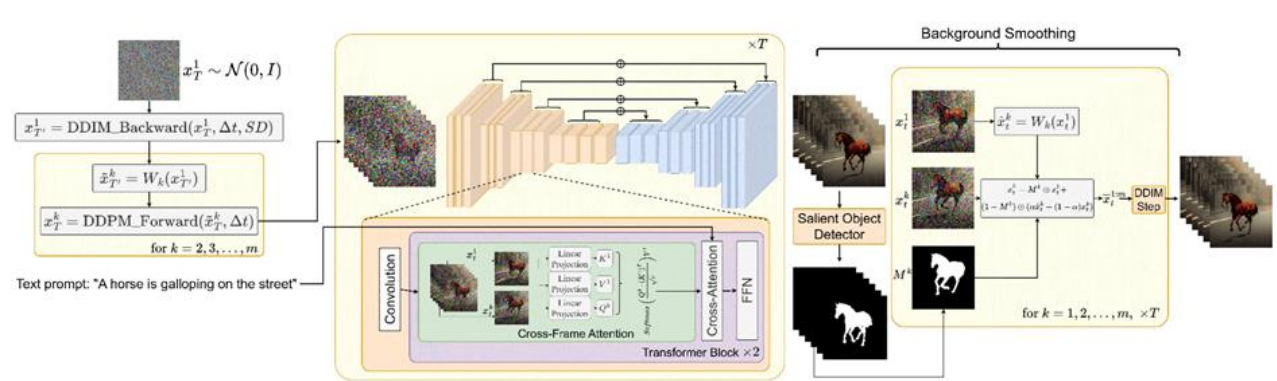
•Đào tạo mô hình khuếch tán hình ảnh tới văn bản:

Đào tạo một mô hình khuếch tán hình ảnh tới văn bản trên bộ dữ liệu đã thu thập được.

•Tạo latent codes cho video:

Sử dụng mô hình khuếch tán hình ảnh tới văn bản để tạo ra một chuỗi các latent codes cho video.

Nội dung và Phương pháp



•Mô hình hóa motion dynamics in latent codes:

Sử dụng một mô hình hồi quy hoặc mô hình động để mô hình hóa motion dynamics in latent codes.

•Tạo video từ latent codes:

Sử dụng các latent codes đã được mô hình hóa để tạo ra một chuỗi các hình ảnh.

•Liên kết các hình ảnh lại với nhau để tạo thành video:

Kết quả dự kiến

- Tạo ra video chất lượng cao và nhất quán từ văn bản.
- Video được tạo ra phải có nội dung phù hợp với văn bản.
- Video phải có thể được tạo ra một cách nhanh chóng và dễ dàng.
- Video phải có thể được tạo ra với chi phí thấp.

Tài liệu tham khảo

- [1]. KHACHATRYAN, Levon, et al. Text2video-zero: Text-to-image diffusion models are zero-shot video generators. 2023.
- [2]. Jonathan Ho, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P Kingma, Ben Poole, Mohammad Norouzi, David J Fleet, et al. Imagen video: High definition video generation with diffusion models. arXiv preprint arXiv:2210.02303, 2022.
- [3]. Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyuan Hu, Harry Yang, Oron Ashual, Oran Gafni, et al. Make-a-video: Text-to-video generation without text-video data. arXiv preprint arXiv:2209.14792, 2022.
- [4]. Jay Zhangjie Wu, Yixiao Ge, Xintao Wang, Weixian Lei, Yuchao Gu, Wynne Hsu, Ying Shan, Xiaohu Qie, and Mike Zheng Shou. Tune-a-video: One-shot tuning of image diffusion models for text-to-video generation. arXiv preprint arXiv:2212.11565, 2022.
- [5]. Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyuan Hu, Harry Yang, Oron Ashual, Oran Gafni, et al. Make-a-video: Text-to-video generation without text-video data. arXiv preprint arXiv:2209.14792, 2022.
- [6]. Eyal Molad, Eliahu Horwitz, Dani Valevski, Alex Rav Acha, Yossi Matias, Yael Pritch, Yaniv Leviathan, and Yedid Hoshen. Dreamix: Video diffusion models are general video editors. arXiv preprint arXiv:2302.01329, 2023.