

BÁO CÁO ĐỒ ÁN CUỐI KỲ

Môn học: CS519 - PHƯƠNG PHÁP LUẬN NCKH

Lớp: CS519.O11

GV: PGS.TS. Lê Đình Duy

Trường ĐH Công Nghệ Thông Tin, ĐHQG-HCM



CONTROLNET: ĐIỀU KHIỂN BỐ CỤC KHÔNG GIAN TRONG MÔ HÌNH KHUẾCH TÁN VĂN BẢN-SANG-ẢNH

Trương Văn Chinh - 20521137

Tóm tắt

- Lớp: CS519.011
- Link Github: <https://github.com/truongvanchinh/CS519.011>
- Link YouTube video:
<https://www.youtube.com/watch?v=HwZd5E57zPA>



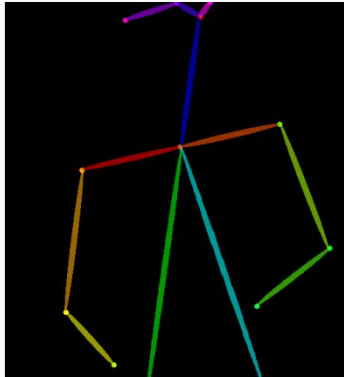
Trương Văn Chinh

Giới thiệu

Input

Chef in kitchen

Output



+ Chef in kitchen



Mục tiêu

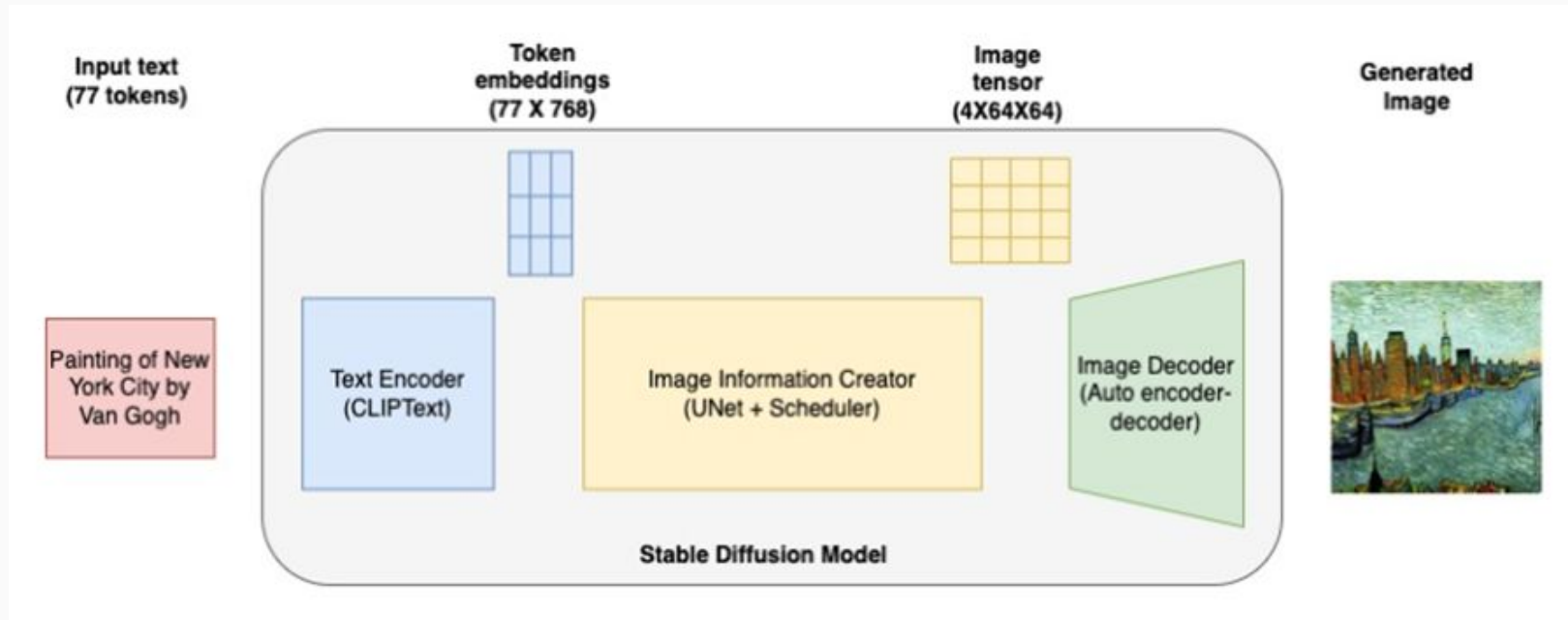
- Tìm hiểu tổng quan về Latent Diffusion Model trong bài toán sinh ảnh từ văn bản.
- Cải tiến Latent Diffusion Model thêm điều kiện như bố cục, vị trí, khung xương, nét vẽ.

Nội dung và Phương pháp 1

- Tìm hiểu thông tin cơ bản của Latent Diffusion Models, GANs, Diffusion Models và so sánh chúng.
- Không gian tiềm ẩn.
- Giai đoạn huấn luyện của LDMs.
- Ưu điểm: tốc độ đánh giá, khả năng huấn luyện, tiết kiệm tài nguyên, chất lượng ảnh.
- Cách hoạt động.

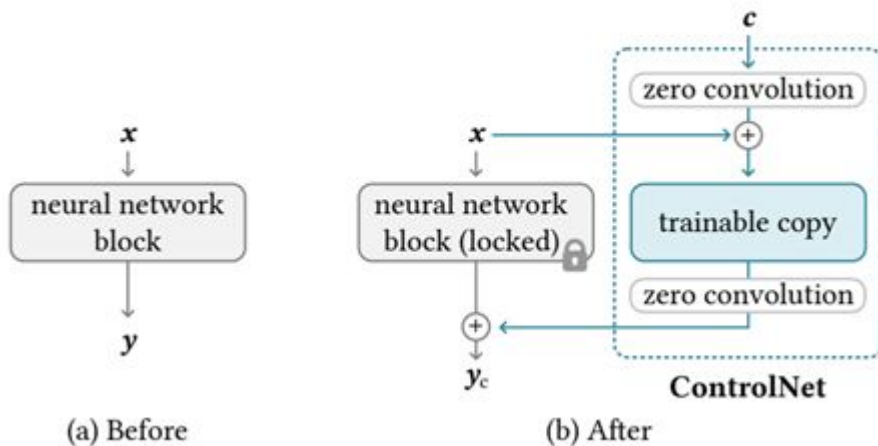
Nội dung và Phương pháp 1

- Cách hoạt động.



Nội dung và Phương pháp 2

- ControlNet giúp cải thiện mô hình tạo ảnh bằng cách cung cấp thông tin hình ảnh cụ thể.
- Bản sao có thể sử dụng lại mô hình huấn luyện trước giúp nó nhanh chóng học được các điều kiện mới và xử lý đa dạng đầu vào



Nội dung và Phương pháp 2

- Tìm hiểu cách áp dụng ControlNet vào Stable Diffusion.
- Kiểm tra độ hiệu quả tính toán sau khi áp dụng khối kết nối ControlNet.

Kết quả dự kiến

- Hiểu rõ thông tin cơ bản về cấu trúc và cách hoạt động của LDMs với bài toán sinh ảnh từ văn bản.
- Ảnh được sinh ra đúng với yêu cầu hơn so với phương pháp chỉ dùng các đoạn text thông thường, tăng tính chính xác của kết quả.

Tài liệu tham khảo

- [1]. Lvmin Zhang, Maneesh Agrawala: Adding Conditional Control to Text-to-Image Diffusion Models. CoRR abs/2302.05543 (2023).
- [2]. Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, Surya Ganguli: Deep Unsupervised Learning using Nonequilibrium Thermodynamics. CoRR abs/1503.03585 (2015).
- [3]. Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer: High-Resolution Image Synthesis with Latent Diffusion Models. CVPR 2022: 10674-10685.
- [4]. Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, Yoshua Bengio: Generative Adversarial Nets. NIPS 2014: 2672-2680.
- [5]. Olaf Ronneberger, Philipp Fischer, Thomas Brox: U-Net: Convolutional Networks for Biomedical Image Segmentation. CoRR abs/1505.04597 (2015).
- [6]. Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever: Learning Transferable Visual Models From Natural Language Supervision. CoRR abs/2103.00020 (2021).