

Controlnet: Spatial Layout Control In Text-To-Image Diffusion Models

Truong Van Chinh

Trường ĐH Công nghệ Thông tin -
TP.HCM

What ?

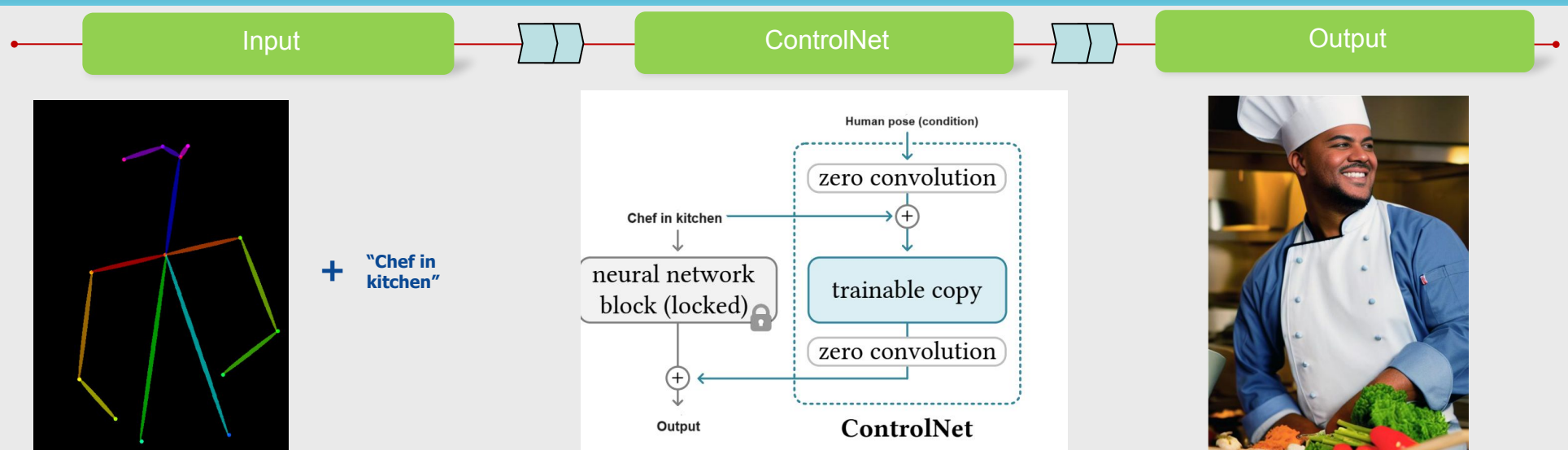
Tìm hiểu và cải thiện Stable Diffusion (Latent Diffusion Models):

- Tìm hiểu tổng quan về Latent Diffusion Models .
- Cải thiện bằng cách thêm bố cục không gian của hình ảnh

Why ?

- Các mô hình sinh ảnh với đầu vào là văn bản chưa thể hiện đủ yêu cầu của người dùng vì thế kết quả còn giới hạn chưa đạt hiệu quả người dùng mong muốn. Nên hầu hết mọi người phải sửa lại văn bản nhiều lần. Vậy có cách nào giải quyết vấn đề? Thì một giải pháp là ngoài đầu vào là văn bản thì chúng tôi thêm điều kiện khác như bố cục, tư thế, vị trí.
- Do đó phương pháp với tên gọi ControlNet đã ra đời.

Overview



Description

1. Nội dung

- Tìm hiểu tổng quan, cách thức hoạt động và các ưu điểm của Latent Diffusion Models so với các phương pháp trước đó (GANs, Diffusion Models)
- Cải tiến Latent Diffusion Model bằng cách thêm điều kiện điều khiển bổ sung như bố cục, hình ảnh khung xương, vị trí các đối tượng với ControlNet.

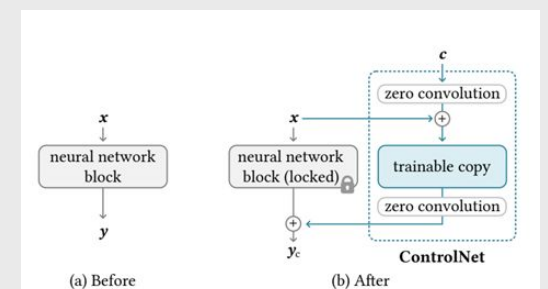
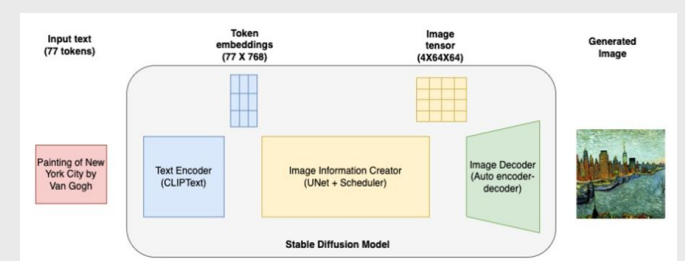
2. Phương Pháp

Nội dung 1

- Tìm hiểu các thông tin tổng quan của LDMs, GANs, Diffusion Models.
- Cách mà LDMs sử dụng không gian tiềm ẩn (Departure to Latent Space). Không gian tiềm ẩn là một không gian chiều thấp hơn không gian dữ liệu gốc, nhưng vẫn giữ lại thông tin quan trọng về dữ liệu gốc.
- LDMs chia quá trình huấn luyện thành 2 giai đoạn.
- Xem xét một vài ưu điểm của mô hình Latent Diffusion Models.
- Tìm hiểu về cấu trúc và cách hoạt động của LDMs

Nội dung 2

- ControlNet giúp cải thiện mô hình tạo ảnh bằng cách cung cấp thông tin hình ảnh cụ thể.
- ControlNet chèn các điều kiện bổ sung vào các khối của một mạng nơ-ron bằng cách đóng băng lại khối gốc sau đó nhân bản nó ra và khối nhân bản sẽ sử dụng vector điều kiện đầu vào làm input.
- ControlNet sử dụng 2 lớp tích chập không (zero convolution layers) để kết nối bản sao và khối mạng gốc.
- 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 mà không cần phải đào tạo lại toàn bộ mạng từ đầu.
- Tìm hiểu cách chuyển đổi từng khối trong mô hình ban đầu thành từng khối mới cho mỗi tầng mã hóa của U-Net.
- Tạo ra các kết nối đi từ các khối encoder đến zero convolution, các kết nối này có tác dụng truyền thông tin từ các khối mã hóa đến các lớp zero convolution của ControlNet.
- Kiểm tra độ hiệu quả tính toán sau khi áp dụng khối kết nối ControlNet.



3. 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.