TRƯỜNG ĐẠI HỌC SÀI GÒN

**KHOA CÔNG NGHỆ THÔNG TIN**

****----- 🙡 🕮 🙣 -----

**MÔN PHƯƠNG PHÁP NGHIÊN CỨU KHOA HỌC**

**Tên Bài: Khảo sát đề tài**

Giảng viên hướng dẫn: **TS. Đỗ Như Tài**

Sinh viên thực hiện:

Tạ Hồng Quí – 3122410348

TP. THCM

Nội dung

[1. Lý do chọn đề tài 2](#_Toc194009386)

[2. Tình hình nghiên cứu hiện tại. 3](#_Toc194009387)

[3. Thách thức trong bài toán Text-to-Image 3](#_Toc194009388)

[4. Lý do cần nghiên cứu và cải tiến 4](#_Toc194009389)

# Lý do chọn đề tài

Trong thời đại công nghiệp 4.0, trí tuệ nhân tạo (AI) ngày càng phát triển và đạt được những thành tựu nhất định. Đặc biệt là trong lĩnh vực dữ liệu và hình ảnh. Các phương pháp sinh ảnh từ văn bản (text-to-images Generation) đang phát triển mạnh mẽ và mở ra nhiều ứng dụng trong trong việc tạo ra hình ảnh phục vụ cho nội dung số, ý tế.

Trước đây, đã có một vài mô hình sinh ảnh đã được phát triển như GANs (Generative Adversarial Networks) và VAEs (Variational Autoencoders), những vẫn tồn tại những mặt hạn chế về sự ổn định và chi tiết của ảnh. Gần đây, mô hình khuêch tán (Diffusion Models) đã trở thành bước đi đột phá trong bài toán text-to-Image, cho phép tạo ra ảnh sắc nét, chân thực hơn.

Một trong những mô hình tiêu biểu là Denoising Diffusion Probabilistic Models (DDPM), được giới thiệu bởi Ho et al. (2020) [1]. DDPM hoạt động dựa trên nguyên lý khử nhiễu dần dần từ một phân phối Gaussian thuần túy, giúp tạo ra ảnh có chất lượng cao hơn so với các phương pháp truyền thống. Tuy nhiên, dù đạt được nhiều thành tựu, DDPM vẫn gặp một số thách thức quan trọng, đặc biệt là về chất lượng ảnh:

* Tốn tài nguyên tính toán, thời gian huấn luyện lâu hơn.[1]
* Mô hình đang hoạt động ở trạng thái không có điều kiện [1]
* Chi tiết ảnh bị mờ hoặc thiếu sắc nét khi mô hình chưa hội tụ tốt.
* Hình ảnh không phản ánh chính xác nội dung văn bản đầu vào, gây ra lỗi không mong muốn.
* Các vùng ảnh phức tạp (như khuôn mặt, bàn tay, chữ viết) có thể bị biến dạng hoặc mất thông tin quan trọng.

Để khắc phục những vấn đề trên, nghiên cứu này tập trung vào cải thiện chất lượng ảnh sinh ra từ mô hình DDPM, cụ thể là:

* Tối ưu hóa quá trình huấn luyện, giúp mô hình học được các đặc trưng hình ảnh tốt hơn.
* Kiểm tra hiệu suất của mô hình có điều kiện
* **Ứng dụng cơ chế Attention hoặc biến thể của Transformer** để cải thiện khả năng tập trung vào chi tiết quan trọng.
* **Sử dụng Cross-attention để tăng trích xuất đặc trưng từ ảnh.**
* **Đánh giá chất lượng ảnh dựa trên các chỉ số khách quan** như FID (Fréchet Inception Distance) và CLIP Score.

# Tình hình nghiên cứu hiện tại.

GANs (Generative Adversarial Networks): Được giới thiệu lần đầu tiên trong bài báo khoa học có tựa đề "Generative Adversarial Nets" vào năm 2014 bởi Ian Goodfellow và các cộng sự.

VAEs (Variational Autoencoders): Được giới thiệu vào năm 2013 với hai bài báo khoa học độc lập:

* + "Auto-Encoding Variational Bayes" của Diederik P. Kingma và Max Welling.
  + "Stochastic Backpropagation and Approximate Inference in Deep Generative Models" của Danilo Jimenez Rezende, Shakir Mohamed, và Daan Wierstra.

DDPMs (Denoising Diffusion Probabilistic Models): Mặc dù các khái niệm cơ bản đã được phát triển từ lâu, nhưng sự phổ biến của DDPMs tăng mạnh sau bài báo "Denoising Diffusion Probabilistic Models" của Jonathan Ho, Ajay Jain, và Pieter Abbeel, được công bố vào năm 2020.

# Thách thức trong bài toán Text-to-Image

Mặc dù các mô hình Text-to-Image đã đạt được nhiều tiến bộ, nhưng vẫn còn nhiều thách thức cần giải quyết:

* Chất lượng ảnh: Mô hình có thể tạo ra ảnh không đúng với mô tả hoặc bị méo mó.
* Tính nhất quán: Một số mô hình có thể tạo ra hình ảnh không đúng với ngữ cảnh của văn bản.
* Tốc độ sinh ảnh: Các mô hình khuếch tán như Stable Diffusion hoặc DDPM thường có thời gian suy luận lâu, gây khó khăn cho các ứng dụng thời gian thực.

Ví dụ, khi mô tả "một con mèo ngồi trên ghế đỏ", mô hình có thể tạo ra hình ảnh trong đó:

* Màu sắc của ghế không chính xác.
* Hình dạng con mèo bị biến dạng.
* Chi tiết trong ảnh không tự nhiên hoặc thiếu logic.

# Lý do cần nghiên cứu và cải tiến

Để khắc phục những hạn chế này, nghiên cứu này tập trung vào việc:

* Cải thiện độ chính xác của ảnh sinh ra từ văn bản mô tả.
* Tối ưu hóa tốc độ suy luận của mô hình khuếch tán, giúp tạo ảnh nhanh hơn.
* So sánh và đánh giá hiệu suất giữa các phương pháp khác nhau nhằm đưa ra giải pháp phù hợp.

Tài liệu tham khảo:

[1] Jonathan Ho, Ajay Jain, và Pieter Abbeel “Denoising Diffusion Probabilistic MModels” in Advances in Neural Information Processing Systems 33 (NeurIPS 2020)

[2] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, và Björn Ommer in ​High-Resolution Image Synthesis with Latent Diffusion Models in CVPR 2022