TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**TÔ NGỌC HUYỀN - 242805005**

**TÌM HIỂU CÁC MÔ HÌNH AI TẠO SINH TRONG LĨNH VỰC**

**THỊ GIÁC MÁY TÍNH**

**BÁO CÁO**

**CHUYÊN ĐỀ NGHIÊN CỨU 2**

**KHOA HỌC MÁY TÍNH**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2025**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**TÔ NGỌC HUYỀN - 242805005**

**TÌM HIỂU CÁC MÔ HÌNH AI TẠO SINH TRONG LĨNH VỰC**

**THỊ GIÁC MÁY TÍNH**

**BÁO CÁO**

**CHUYÊN ĐỀ NGHIÊN CỨU 2**

**KHOA HỌC MÁY TÍNH**

Người hướng dẫn

**TS. Phạm Văn Huy**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2025**

**LỜI CẢM ƠN**

Em xin gửi lời cảm ơn chân thành đến trường Đại học Tôn Đức Thắng đã tạo một môi trường học tập hiện đại, năng động, giúp em có điều kiện thuận lợi để học tập và phát triển trong suốt quá trình nghiên cứu.

Em cũng xin cảm ơn khoa Công nghệ thông tin đã tổ chức và triển khai môn học Chuyên đề nghiên cứu 2, đồng thời luôn đồng hành, định hướng và hỗ trợ chúng em tiếp cận các kiến thức chuyên sâu và tài liệu hữu ích.

Đặc biệt, em xin bày tỏ lòng biết ơn sâu sắc đến thầy Phạm Văn Huy – giảng viên hướng dẫn chuyên đề đã tận tình định hướng, hỗ trợ và góp ý chuyên môn trong suốt quá trình em thực hiện đề tài. Sự hướng dẫn tận tâm của thầy là nguồn động lực quý báu giúp em hoàn thành tốt báo cáo này.

Do còn hạn chế về kiến thức và kinh nghiệm, báo cáo không tránh khỏi thiếu sót. Em rất mong nhận được những ý kiến đóng góp quý báu từ quý Thầy Cô để hoàn thiện hơn trong các nghiên cứu tiếp theo.

*TP. Hồ Chí Minh, ngày15 tháng 07 năm 2025*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*

*Huyền*

*Tô Ngọc Huyền*

**CÔNG TRÌNH ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi và được sự hướng dẫn khoa học của TS. Phạm Văn Huy. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong Dự án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung Dự án của mình**. Trường Đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày 15 tháng 07 năm 2025*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*

*Huyền*

*Tô Ngọc Huyền*

**TÌM HIỂU CÁC MÔ HÌNH AI TẠO SINH TRONG LĨNH VỰC THỊ GIÁC MÁY TÍNH**

**TÓM TẮT**

Báo cáo này tập trung nghiên cứu các mô hình Trí tuệ nhân tạo tạo sinh (Generative AI) trong lĩnh vực Thị giác máy tính (Computer Vision), một hướng tiếp cận tiên tiến cho phép máy tính không chỉ hiểu mà còn tạo ra dữ liệu hình ảnh mới. Khác với các mô hình AI truyền thống chỉ dừng lại ở phân tích hoặc nhận dạng, AI tạo sinh có khả năng học các đặc trưng ẩn và sinh dữ liệu mới có tính chân thực cao, mở ra nhiều tiềm năng trong y tế, nghệ thuật số, thiết kế và nhiều lĩnh vực khác.

Nội dung báo cáo đi sâu vào ba mô hình tiêu biểu: Generative Adversarial Networks (GANs), Variational Autoencoders (VAE) và Diffusion Models. Mỗi mô hình được phân tích toàn diện từ kiến trúc, nguyên lý hoạt động, ưu nhược điểm cho đến khả năng ứng dụng trong các bài toán như tạo ảnh từ văn bản, phục hồi ảnh, chuyển đổi phong cách, siêu phân giải và phát hiện bất thường.

Các thực nghiệm trên bộ dữ liệu MNIST được thực hiện nhằm minh họa trực quan hoạt động của từng mô hình, từ việc huấn luyện đến quá trình sinh ảnh. Ngoài ra, báo cáo cũng trình bày so sánh các mô hình trên nhiều tiêu chí: chất lượng ảnh, độ ổn định huấn luyện, tốc độ suy luận và khả năng kiểm soát quá trình tạo sinh.

Kết quả nghiên cứu cho thấy mỗi mô hình đều có ưu thế riêng: GANs vượt trội về độ chân thực của hình ảnh, VAE mạnh về không gian tiềm ẩn có cấu trúc, trong khi Diffusion Models nổi bật với chất lượng ảnh siêu thực và ổn định huấn luyện. Báo cáo hướng đến việc cung cấp nền tảng lý thuyết vững chắc, đồng thời định hướng cho những nghiên cứu và ứng dụng chuyên sâu hơn trong tương lai của AI tạo sinh trong thị giác máy tính.

**AN INVESTIGATION OF GENERATIVE AI MODELS FOR COMPUTER VISION APPLICATIONS**

**ABSTRACT**

This report focuses on the study of Generative Artificial Intelligence (Generative AI) models in the field of Computer Vision, a cutting-edge approach that enables machines not only to understand but also to generate new visual data. Unlike traditional AI models that primarily focus on analysis or recognition, generative AI is capable of learning latent features and producing highly realistic synthetic data, unlocking significant potential across various domains such as healthcare, digital art, design, and more.

The report delves into three prominent generative models: Generative Adversarial Networks (GANs), Variational Autoencoders (VAE), and Diffusion Models. Each model is comprehensively analyzed in terms of architecture, working principles, strengths and limitations, as well as real-world applications in tasks like text-to-image generation, image restoration, style transfer, super-resolution, and anomaly detection.

Experiments conducted on the MNIST dataset illustrate how these models operate, from training processes to image generation. Additionally, the report presents a comparative analysis of the models across several criteria including image quality, training stability, inference speed, and controllability of the generation process.

Findings indicate that each model has its own advantages: GANs excel in generating highly realistic images, VAE offer structured latent spaces, and Diffusion Models achieve superior image quality and training stability. This report aims to provide a solid theoretical foundation while also guiding further research and advanced applications of generative AI in computer vision.

**MỤC LỤC**

[DANH MỤC HÌNH VẼ vii](#_Toc203426542)

[DANH MỤC BẢNG BIỂU viii](#_Toc203426543)

[DANH MỤC CÁC CHỮ VIẾT TẮT ix](#_Toc203426544)

[CHƯƠNG 1. MỞ ĐẦU VÀ TỔNG QUAN ĐỀ TÀI 1](#_Toc203426545)

[1.1 Bối cảnh và lý do chọn đề tài 1](#_Toc203426546)

[1.2 Mục tiêu thực hiện đề tài 2](#_Toc203426547)

[1.3 Đối tượng và phạm vi nghiên cứu 2](#_Toc203426548)

[1.4 Phương pháp nghiên cứu 2](#_Toc203426549)

[CHƯƠNG 2. TỔNG QUAN VỀ AI TẠO SINH VÀ THỊ GIÁC MÁY TÍNH 4](#_Toc203426550)

[2.1 Trí tuệ nhân tạo tạo sinh (Generative Artificial Intelligence) 4](#_Toc203426551)

[2.2 Thị giác máy tính (Computer Vision) 4](#_Toc203426552)

[2.3 Mô hình AI tạo sinh và vai trò trong thị giác máy tính 4](#_Toc203426553)

[CHƯƠNG 3. CÁC MÔ HÌNH TẠO SINH TIÊU BIỂU 7](#_Toc203426554)

[3.1 Generative Adversarial Networks (GANs) 7](#_Toc203426555)

[3.1.1 Kiến trúc tổng thể của GANs 7](#_Toc203426556)

[3.1.2 Nguyên lý hoạt động 8](#_Toc203426557)

[3.1.3 Các biến thể phổ biến và cải tiến 9](#_Toc203426558)

[3.1.4 Ưu điểm và nhược điểm 9](#_Toc203426559)

[3.1.5 Ứng dụng trong Thị giác máy tính 11](#_Toc203426560)

[3.2 Variational Autoencoders (VAE) 12](#_Toc203426561)

[3.2.1 Kiến trúc tổng thể của VAE 12](#_Toc203426562)

[3.2.2 Hàm mất mát – ELBO 13](#_Toc203426563)

[3.2.3 Reparameterization Trick 14](#_Toc203426564)

[3.2.4 Ưu điểm và nhược điểm 14](#_Toc203426565)

[3.2.5 Ứng dụng trong Thị giác máy tính 16](#_Toc203426566)

[3.3 Diffusion Models 17](#_Toc203426567)

[3.3.1 Nguyên lý hoạt động 18](#_Toc203426568)

[3.3.2 Kiến trúc U-Net và Hàm mất mát 19](#_Toc203426569)

[3.3.3 Ưu điểm và Nhược điểm 20](#_Toc203426570)

[3.3.4 Ứng dụng trong Thị giác Máy tính 21](#_Toc203426571)

[3.4 So sánh các mô hình AI tạo sinh 23](#_Toc203426572)

[CHƯƠNG 4. THỰC NGHIỆM 25](#_Toc203426573)

[4.1 Dữ liệu thực nghiệm 25](#_Toc203426574)

[4.2 Môi trường và công cụ thực nghiệm 25](#_Toc203426575)

[4.3 Demo 1: Tạo ảnh chữ số viết tay bằng GANs 25](#_Toc203426576)

[4.4 Demo 2: Tạo ảnh chữ số viết tay bằng VAE 27](#_Toc203426577)

[4.5 Demo 3: Tạo ảnh chữ số viết tay bằng Diffusion Model 29](#_Toc203426578)

[CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 31](#_Toc203426579)

[5.1 Kết luận 31](#_Toc203426580)

[5.2 Hướng phát triển 31](#_Toc203426581)

[TÀI LIỆU THAM KHẢO 33](#_Toc203426582)

# DANH MỤC HÌNH VẼ

[*Hình 3.1 Nguyên lý hoạt động vủa GANs* 7](#_Toc203417285)

[*Hình 4.1 Kết quả tạo ảnh chữ số viết tay bằng GANs* 27](#_Toc203460371)

[*Hình 4.2 Kết quả tạo ảnh chữ số viết tay bằng VAE* 29](#_Toc203460372)

[*Hình 4.3 Kết quả tạo ảnh chữ số viết tay bằng Diffusion Model* 30](#_Toc203460373)

# DANH MỤC BẢNG BIỂU

[*Bảng 3.1 Bảng so sánh mô hình GANs, VAE và Diffusion Models* 23](#_Toc203417267)

# DANH MỤC CÁC CHỮ VIẾT TẮT

|  |  |
| --- | --- |
| AI | Artificial Intelligence |
| CV | Computer Vision |
| GAN | Generative Adversarial Network |
| DCGAN | Deep Convolutional GAN |
| WGAN | Wasserstein GAN |
| VAE | Variational Autoencoders |
| ELBO | Evidence Lower Bound |
| KL | Kullback-Leibler Divergence |
| CNN | Convolutional Neural Network |
| MSE | Mean Squared Error |
| LDM | Latent Diffusion Model |
| DDPM | Denoising Diffusion Probabilistic Model |
| U-Net | U-shaped Convolutional Neural Network |
| MNIST | Modified National Institute of Standards and Technology |
| VR | Virtual Reality |
| AR | Augmented Reality |
| NLP | Natural Language Processing |
| BERT | Bidirectional Encoder Representations from Transformers |

# MỞ ĐẦU VÀ TỔNG QUAN ĐỀ TÀI

## Bối cảnh và lý do chọn đề tài

Trong những năm gần đây, sự phát triển vượt bậc của Trí tuệ nhân tạo (AI) đã tạo ra những thay đổi sâu rộng trong nhiều lĩnh vực, đặc biệt là cách con người tương tác, xử lý và tạo ra dữ liệu số. Trong số đó, AI tạo sinh (Generative AI) nổi lên như một hướng nghiên cứu mang tính đột phá, với khả năng tạo ra dữ liệu mới có tính chân thực và giá trị ứng dụng cao. Khác với các mô hình AI truyền thống vốn chủ yếu tập trung vào việc phân tích hoặc nhận dạng dữ liệu hiện có, AI tạo sinh có khả năng học được các đặc trưng tiềm ẩn trong dữ liệu và từ đó tổng hợp nội dung hoàn toàn mới từ hình ảnh, video cho đến văn bản và âm thanh. Sự chuyển dịch từ khả năng phân tích sang khả năng tạo sinh này đại diện cho một bước tiến cơ bản trong năng lực của AI. Khi AI không chỉ có thể hiểu dữ liệu mà còn có thể tổng hợp dữ liệu mới, chân thực, điều này mở ra những ứng dụng và mô hình mới chưa từng có. Điều này cho thấy AI đang tiến tới việc trở thành một hệ thống "sáng tạo" và "chủ động" hơn. Sự thay đổi này về cơ bản thay đổi mô hình tương tác giữa con người và máy tính, chuyển từ việc con người cung cấp đầu vào để AI phân tích, sang việc AI chủ động tham gia vào quá trình tạo nội dung, thiết kế và giải quyết vấn đề. Điều này có những tác động sâu sắc đến các ngành công nghiệp dựa vào nội dung, thiết kế và tăng cường dữ liệu.

Thị giác máy tính (Computer Vision – CV) là một trong những lĩnh vực chịu ảnh hưởng sâu sắc nhất từ sự phát triển của AI tạo sinh, khi các mô hình mới không chỉ nâng cao khả năng hiểu hình ảnh mà còn mở ra tiềm năng sáng tạo và tái cấu trúc nội dung hình ảnh một cách linh hoạt và chính xác. Thị giác máy tính giúp máy tính có thể “nhìn thấy” và có thể hiểu được thế giới thông qua dữ liệu hình ảnh. Việc tích hợp các mô hình tạo sinh vào thị giác máy tính không chỉ mở rộng khả năng xử lý và hiểu hình ảnh, mà nó còn mang lại những tiến bộ vượt bậc trong các bài toán như phục hồi ảnh, tạo ảnh mới, chuyển đổi phong cách, hoặc tái cấu trúc cảnh vật trong không gian ba chiều.

Xuất phát từ tầm quan trọng và tiềm năng ứng dụng to lớn của chủ đề nêu trên, báo cáo chuyên đề này được thực hiện với mục tiêu tìm hiểu, phân tích và hệ thống hóa kiến thức về các mô hình AI tạo sinh đang được ứng dụng trong lĩnh vực thị giác máy tính. Báo cáo sẽ tập trung vào các mô hình tiêu biểu như Mạng đối kháng tạo sinh (Generative Adversarial Networks - GANs), Bộ tự mã hóa biến phân (Variational Autoencoders - VAE), Mô hình khuếch tán (Diffusion Models), và một số mô hình tiên tiến khác.

## Mục tiêu thực hiện đề tài

Đề tài hướng tới các mục tiêu sau:

* Nghiên cứu, tổng hợp và hệ thống hóa kiến thức về các mô hình AI tạo sinh trong lĩnh vực thị giác máy tính.
* Phân tích sâu về kiến trúc, nguyên lý hoạt động, ưu nhược điểm và ứng dụng của các mô hình tiêu biểu (GANs, VAE, Diffusion Models).
* Khảo sát các mô hình tiên tiến và ứng dụng đặc thù đang nổi lên.
* Đánh giá tiềm năng, nhận diện thách thức và dự báo xu hướng phát triển tương lai của AI tạo sinh trong thị giác máy tính

## Đối tượng và phạm vi nghiên cứu

Đối tượng: Các mô hình AI tạo sinh, đặc biệt là các mô hình được ứng dụng trong thị giác máy tính.

Phạm vi: Đề tài giới hạn trong việc phân tích và thực nghiệm với ba nhóm mô hình chính gồm: GANs, VAE và Diffusion Models. Ngoài ra, báo cáo cũng đề cập đến một số mô hình tiên tiến hoặc biến thể có liên quan, phục vụ cho việc mở rộng và so sánh.

## Phương pháp nghiên cứu

Để đạt được các mục tiêu đã đề ra, luận văn sử dụng các phương pháp nghiên cứu sau:

* Phương pháp tổng hợp tài liệu: Thu thập, chọn lọc và phân tích các tài liệu học thuật từ các công trình nghiên cứu, bài báo khoa học và tài nguyên chính thống liên quan đến Generative AI và thị giác máy tính.
* Phân tích so sánh: Đánh giá các mô hình thông qua phân tích lý thuyết và thực nghiệm, dựa trên các tiêu chí như chất lượng ảnh, độ ổn định huấn luyện, khả năng kiểm soát và tốc độ suy luận.
* Thực nghiệm mô phỏng: Triển khai các mô hình GAN, VAE và Diffusion Model trên tập dữ liệu MNIST để minh họa trực quan nguyên lý hoạt động và khả năng tạo ảnh.
* Phương pháp dự báo và đánh giá xu hướng: Tổng hợp kết quả nghiên cứu, từ đó đưa ra những nhận định mang tính định hướng cho sự phát triển của các mô hình AI tạo sinh trong tương lai.

# TỔNG QUAN VỀ AI TẠO SINH VÀ THỊ GIÁC MÁY TÍNH

## Trí tuệ nhân tạo tạo sinh (Generative Artificial Intelligence)

Trí tuệ nhân tạo tạo sinh (Generative AI) là một nhánh của học máy (machine learning) chuyên nghiên cứu các mô hình có khả năng học từ dữ liệu đầu vào và sinh ra các mẫu dữ liệu mới có tính tương đồng thống kê với dữ liệu huấn luyện. Về mặt bản chất, mô hình AI tạo sinh không chỉ học quy luật biểu hiện bên ngoài mà còn khai thác các biểu diễn tiềm ẩn trong không gian dữ liệu, từ đó tái tạo hoặc tổng hợp dữ liệu mới có tính sáng tạo và logic.

Khác với mô hình học có giám sát (supervised learning), vốn học một hàm ánh xạ đầu vào và đầu ra, các mô hình tạo sinh chủ yếu được xây dựng dựa trên các phương pháp học không giám sát (unsupervised learning) hoặc học bán giám sát (semi-supervised learning). Mục tiêu chính là ước lượng phân phối xác suất tiềm ẩn của dữ liệu , hoặc điều kiện nơi là biến ẩn sinh ra từ một phân phối chuẩn hoá (thường là Gaussian).

## Thị giác máy tính (Computer Vision)

Thị giác máy tính là lĩnh vực nghiên cứu liên quan đến việc tự động hóa các nhiệm vụ mà hệ thị giác con người có thể thực hiện, bao gồm nhận dạng đối tượng, phân đoạn ảnh, ước lượng độ sâu, theo dõi chuyển động, và hiểu nội dung hình ảnh. Mục tiêu của thị giác máy tính là trích xuất thông tin có cấu trúc từ dữ liệu thị giác (hình ảnh hoặc video) và sử dụng thông tin đó cho các tác vụ thông minh.

Trong bối cảnh truyền thống, Thị giác máy tính sử dụng các kỹ thuật xử lý ảnh (image processing), học đặc trưng thủ công (hand-crafted features) và các bộ phân loại cổ điển. Tuy nhiên, với sự phát triển của học sâu (deep learning), đặc biệt là mạng nơ-ron tích chập (Convolutional Neural Networks – CNNs), thị giác máy tính đã đạt được những bước tiến vượt bậc trong độ chính xác và khả năng khái quát hóa.

## Mô hình AI tạo sinh và vai trò trong thị giác máy tính

Trong những năm gần đây, sự kết hợp giữa các mô hình Trí tuệ nhân tạo tạo sinh (Generative AI) và lĩnh vực Thị giác máy tính (Computer Vision) đã tạo ra những bước tiến đáng kể trong cả nghiên cứu học thuật lẫn ứng dụng thực tiễn. Các mô hình AI tạo sinh không chỉ đơn thuần hỗ trợ máy tính hiểu và phân tích dữ liệu hình ảnh, mà còn mang lại khả năng tổng hợp và tạo mới thông tin thị giác. Việc tích hợp các mô hình này đã mở rộng phạm vi xử lý hình ảnh từ thuần túy phân tích sang khả năng sáng tạo nội dung, từ đó nâng cao chất lượng và hiệu quả trong nhiều tác vụ quan trọng. Dưới đây là một số vai trò nổi bật của AI tạo sinh trong thị giác máy tính:

* Tái tạo và phục hồi hình ảnh (Image Reconstruction & Inpainting): Một trong những ứng dụng quan trọng nhất của Generative AI trong thị giác máy tính là khả năng tái tạo những hình ảnh bị hư hỏng, bị nhiễu hoặc thiếu dữ liệu. Các mô hình như GANs hoặc Diffusion Models có thể được huấn luyện để lấp đầy những vùng bị mất mát thông tin trong ảnh, khôi phục các chi tiết bị mờ, hoặc tái tạo lại hình ảnh gốc với độ chính xác cao. Điều này đặc biệt hữu ích trong y tế (phục hồi ảnh MRI, CT), bảo tồn nghệ thuật (sửa chữa tranh cũ), và giám sát an ninh (khôi phục video chất lượng thấp).
* Sinh ảnh từ dữ liệu không phải hình ảnh (Cross-modal Generation): Generative AI có thể tạo ra hình ảnh từ các dạng dữ liệu khác như văn bản hoặc âm thanh. Mô hình text-to-image (ví dụ: DALL·E, Stable Diffusion) là một ví dụ tiêu biểu, trong đó AI có thể sinh ảnh có nội dung và ngữ cảnh phù hợp với mô tả bằng ngôn ngữ tự nhiên. Điều này mở ra tiềm năng to lớn trong thiết kế sáng tạo, giáo dục trực quan, và phát triển nội dung số. Ngoài ra, một số mô hình còn cho phép chuyển đổi từ tín hiệu âm thanh (như giọng nói hoặc tiếng động) thành ảnh tương ứng, hỗ trợ các ứng dụng trong nhận dạng ngữ cảnh hoặc phân tích hành vi.
* Tăng cường dữ liệu huấn luyện (Data Augmentation): Trong các tình huống dữ liệu huấn luyện khan hiếm hoặc mất cân bằng, các mô hình tạo sinh có thể được sử dụng để tạo ra dữ liệu tổng hợp nhằm cải thiện độ bao phủ và độ chính xác của các mô hình học máy. Việc sinh thêm ảnh huấn luyện từ các lớp ít dữ liệu không chỉ giúp giảm overfitting mà còn tăng khả năng khái quát của mô hình. Đây là một chiến lược quan trọng trong các lĩnh vực như thị giác y học, nơi dữ liệu có nhãn thường đắt đỏ và khó thu thập.
* Chuyển đổi miền ảnh (Image Domain Transfer): AI tạo sinh cho phép chuyển đổi ảnh từ một miền (domain) này sang một miền khác mà vẫn bảo toàn cấu trúc nội dung cơ bản. Một ví dụ điển hình là CycleGAN, cho phép chuyển đổi ảnh ngày thành ảnh đêm, ảnh phong cách hoạt hình thành ảnh thật, hoặc ảnh thời tiết mùa hè thành mùa đông. Ứng dụng này có vai trò thiết thực trong mô phỏng, thiết kế trò chơi, thị giác robot, và các hệ thống thị giác trong điều kiện thay đổi môi trường.
* Sinh ảnh 3D từ ảnh 2D (3D Scene Generation): Với sự phát triển của các mô hình như Neural Radiance Fields (NeRF), AI tạo sinh có khả năng tái cấu trúc không gian ba chiều từ một hoặc nhiều ảnh hai chiều. Điều này giúp tạo ra các mô hình 3D chân thực, phục vụ các ứng dụng như thực tế ảo (VR), thực tế tăng cường (AR), kiến trúc – xây dựng, và robot tự hành. Việc học biểu diễn không gian liên tục trong NeRF không chỉ giúp tái tạo hình học chính xác mà còn giữ lại các chi tiết hình ảnh sắc nét.

# CÁC MÔ HÌNH TẠO SINH TIÊU BIỂU

Chương này đi sâu vào ba mô hình AI tạo sinh nền tảng nhất trong lĩnh vực thị giác máy tính: Generative Adversarial Networks (GANs), Variational Autoencoders (VAE) và Diffusion Models. Mỗi mô hình có kiến trúc, nguyên lý hoạt động, ưu nhược điểm và ứng dụng riêng biệt, đóng góp vào sự phát triển đa dạng của AI tạo sinh.

## Generative Adversarial Networks (GANs)

Mạng Đối kháng tạo sinh (Generative Adversarial Networks - GANs) là một trong những đột phá quan trọng nhất trong lĩnh vực AI tạo sinh, được Ian Goodfellow và các cộng sự giới thiệu vào tháng 6 năm 2014.

GANs được thiết kế để sinh ra dữ liệu mới có phân phối thống kê tương tự như dữ liệu huấn luyện. Trong lĩnh vực thị giác máy tính (computer vision), GANs đã mở ra nhiều hướng phát triển đột phá như: tạo ảnh chân dung, biến đổi phong cách, phục hồi ảnh, và nhiều ứng dụng khác.

### Kiến trúc tổng thể của GANs

Trong mô hình GANs (Generative Adversarial Networks), hai thành phần chính Generator (G) và Discriminator (D) hoạt động theo nguyên lý đối kháng nhằm học và sinh ra dữ liệu mới có phân phối tương tự dữ liệu thật. Mỗi thành phần đảm nhiệm một vai trò riêng biệt nhưng bổ trợ cho nhau trong suốt quá trình huấn luyện.

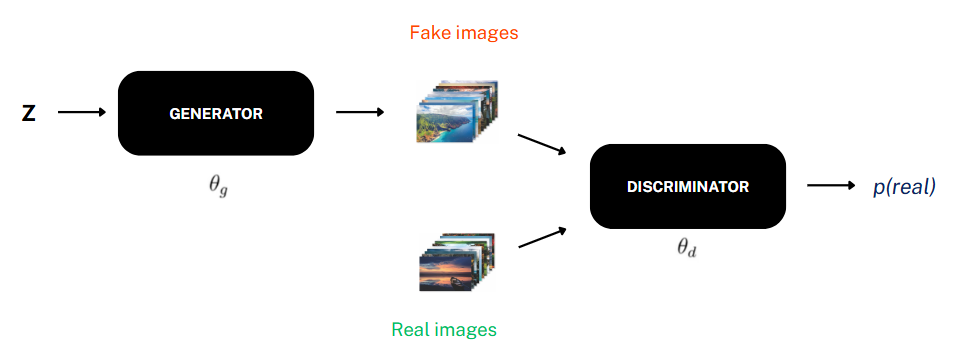
Generator (G) – Bộ sinh dữ liệu

* Generator là một mạng nơ-ron có chức năng sinh dữ liệu mới từ một vector nhiễu ngẫu nhiên (phân phối chuẩn hoặc đồng đều)
* Mục tiêu của Generator là tạo ra dữ liệu giả sao cho không thể phân biệt được với dữ liệu thật bởi bộ phân biệt D.

Discriminator (D) - Bộ phân biệt thật/giả

* Discriminator là một mạng nơ-ron phân loại nhị phân có nhiệm vụ đánh giá xem một mẫu dữ liệu đầu vào là thật (từ tập huấn luyện) hay giả (do Generator sinh ra)
* D xuất ra một xác suất đại diện cho độ tin cậy của ảnh đầu vào là ảnh thật.

### Nguyên lý hoạt động



*Hình 3.1 Nguyên lý hoạt động của GANs*

Trong quá trình huấn luyện, hai mạng G và D được tối ưu với các mục tiêu trái ngược:

* Discriminator cố gắng tối đa hóa khả năng phân biệt chính xác giữa dữ liệu thật và dữ liệu giả
* Generator cố gắng tối thiểu hóa khả năng bị D phát hiện, bằng cách sinh ra các mẫu giả có phân phối gần với phân phối thật.

Quá trình huấn luyện trong mô hình GAN có thể được mô hình hóa như một bài toán tối ưu đối kháng giữa hai thành phần. Quá trình này được mô hình hóa dưới dạng một bài toán tối ưu min-max với hàm mục tiêu như sau:

Trong đó:

* là xác suất mà D cho rằng x là thật
* là dữ liệu giả do G tạo ra từ nhiễu z
* là phân phối dữ liệu thật
* là phân phối nhiễu đầu vào

### Các biến thể phổ biến và cải tiến

Mặc dù kiến trúc GANs cơ bản đã mang lại những đột phá đáng kể trong lĩnh vực tạo sinh dữ liệu, nó vẫn tồn tại nhiều hạn chế như hiện tượng mode collapse, khó khăn trong việc kiểm soát chất lượng đầu ra. Nhằm khắc phục những vấn đề này và mở rộng khả năng ứng dụng, nhiều biến thể và cải tiến đã được đề xuất. Dưới đây là một số biến thể tiêu biểu:

* DCGAN (Deep Convolutional GAN): Ứng dụng mạng tích chập thay cho fully-connected, giúp mô hình học đặc trưng tốt hơn và ổn định hơn trong huấn luyện.
* WGAN (Wasserstein GAN): Thay hàm mất mát bằng khoảng cách Wasserstein để cải thiện độ hội tụ và ổn định, giảm hiện tượng mode collapse. Phiên bản WGAN-GP thêm gradient penalty để tăng hiệu quả huấn luyện.
* CycleGAN: Hỗ trợ chuyển đổi hình ảnh giữa hai miền không có cặp đối ứng bằng cơ chế ràng buộc vòng lặp (cycle consistency), hữu ích trong chuyển đổi phong cách hoặc ánh sáng.
* StyleGAN: Cải tiến mạnh mẽ trong điều khiển không gian tiềm ẩn và tạo ảnh độ phân giải cao. Các phiên bản sau như StyleGAN2, StyleGAN3 tiếp tục nâng chất lượng ảnh và tính liên tục.

### Ưu điểm và nhược điểm

#### Ưu điểm

Chất lượng hình ảnh chân thực cao: GANs có khả năng tạo ra hình ảnh cực kỳ chân thực, khó phân biệt với ảnh thật bằng mắt người.

Học không giám sát hiệu quả: Cơ chế huấn luyện đối kháng cho phép GANs học từ dữ liệu không có nhãn một cách hiệu quả hơn so với nhiều phương pháp truyền thống.

Tạo mẫu hoàn chỉnh trong một lần chạy: Không giống như các mô hình tự hồi quy, GANs có thể tạo ra một mẫu hoàn chỉnh trong một lần chạy qua mạng, giúp tốc độ suy luận nhanh hơn. GigaGAN là một ví dụ điển hình với tốc độ suy luận cực nhanh.

Kiểm soát cao đối với biến thể hình ảnh: Các biến thể như StyleGAN, WGAN và BigGAN cho phép kiểm soát tốt hơn đối với các thuộc tính và biến thể của hình ảnh được tạo ra.

Tăng cường dữ liệu (Data Augmentation): GANs có thể tạo ra dữ liệu tổng hợp để mở rộng tập dữ liệu huấn luyện cho các mô hình học sâu, đặc biệt hữu ích trong các kịch bản dữ liệu hạn chế.

Tạo mô hình 2D và 3D: Có khả năng tổng hợp các mô hình 2D và 3D chi tiết và chân thực, ứng dụng trong VR, mô phỏng và kiến trúc.

Khả năng chỉnh sửa không gian tiềm ẩn: Các mô hình như StyleGAN và GigaGAN cho phép nội suy, trộn phong cách và thực hiện các phép toán vector trong không gian tiềm ẩn để chỉnh sửa hình ảnh.

#### Nhược điểm

Khó khăn trong huấn luyện (Training Instability): GANs nổi tiếng là khó huấn luyện và dễ gặp vấn đề không hội tụ (non-convergence). Việc đạt được trạng thái cân bằng giữa Generator và Discriminator là một thách thức lớn.

Mode Collapse: Đây là một vấn đề phổ biến khi Generator chỉ học cách tạo ra một số ít loại mẫu, làm giảm sự đa dạng của đầu ra và không thể bao phủ toàn bộ phân phối dữ liệu huấn luyện.

Khó khăn trong đánh giá: Việc đánh giá chất lượng và sự đa dạng của các mẫu được tạo ra bởi GANs là một thách thức, vì không có thước đo khách quan nào hoàn hảo.

Phụ thuộc vào chất lượng dữ liệu huấn luyện: Chất lượng đầu ra của GANs phụ thuộc trực tiếp vào chất lượng và sự đa dạng của dữ liệu mà chúng được huấn luyện. Nếu dữ liệu huấn luyện có thiên vị hoặc không đầy đủ, đầu ra cũng sẽ phản ánh những sai sót đó.

Thiếu khả năng "hiểu": GANs không thực sự "hiểu" những gì chúng tạo ra; chúng đơn giản chỉ dự đoán những gì có thể xảy ra tiếp theo dựa trên dữ liệu huấn luyện.

Độ phân giải giới hạn: Các phiên bản GANs ban đầu (như DCGAN) thường chỉ tạo ra hình ảnh ở độ phân giải tương đối thấp. Mặc dù các biến thể sau này đã cải thiện đáng kể, việc tạo hình ảnh độ phân giải cực cao vẫn là một thách thức.

Chi phí tính toán: Huấn luyện các mô hình GANs phức tạp, đặc biệt là các mô hình tạo hình ảnh chất lượng cao, yêu cầu sức mạnh tính toán đáng kể.

### Ứng dụng trong Thị giác máy tính

Mạng đối sinh (GANs) đã phát triển từ một hướng nghiên cứu học thuật thành một công nghệ nền tảng, với khả năng tạo dữ liệu tổng hợp có tính chân thực cao, phục vụ nhiều ứng dụng trong thị giác máy tính và các ngành công nghiệp liên quan đến nội dung trực quan.

Những ứng dụng tiêu biểu bao gồm:

* Tổng hợp hình ảnh và video chân thực: Tạo ra nội dung mới chất lượng cao phục vụ nghệ thuật số, truyền thông và thực tế ảo.
* Tăng cường dữ liệu: Sinh thêm dữ liệu huấn luyện trong các lĩnh vực có dữ liệu hạn chế (như y tế), góp phần cải thiện hiệu suất mô hình và bảo vệ quyền riêng tư.
* Chuyển đổi miền ảnh (image-to-image translation): Biến đổi ảnh từ miền này sang miền khác (ví dụ: ngày - đêm, phác thảo - thật, chuyển phong cách).
* Thao tác đặc trưng khuôn mặt: Thay đổi các thuộc tính khuôn mặt như độ tuổi, cảm xúc, hoặc giới tính.
* Cải thiện chất lượng hình ảnh/video: Hỗ trợ nâng độ phân giải, khử nhiễu và phục hồi ảnh.
* Tạo dữ liệu cho tài chính và y tế: Mô phỏng dữ liệu tài chính, phát hiện gian lận, và tạo ảnh y tế tổng hợp để huấn luyện mô hình chẩn đoán mà không vi phạm quyền riêng tư.
* Tạo mô hình 2D/3D và thiết kế nội dung số: Phục vụ thiết kế kiến trúc, game, VR, và các hệ thống mô phỏng.
* Tạo ảnh từ văn bản (text-to-image): Các mô hình tiên tiến như GigaGAN có thể sinh ảnh từ mô tả bằng ngôn ngữ tự nhiên.

## Variational Autoencoders (VAE)

Variational Autoencoders (VAE) là một trong những kiến trúc mô hình tạo sinh phổ biến, được đề xuất bởi Kingma và Welling vào năm 2013. Mô hình này mở rộng kiến trúc Autoencoder truyền thống bằng cách áp dụng phương pháp suy luận biến phân (variational inference) để ước lượng phân phối xác suất của dữ liệu tiềm ẩn, từ đó cho phép sinh dữ liệu mới một cách khả kiểm soát.

Không giống như GANs, VAE xây dựng một mô hình xác suất rõ ràng, trong đó các biến tiềm ẩn (latent variables) được giả định tuân theo một phân phối chuẩn hoá, thường là phân phối Gaussian đa biến.

### Kiến trúc tổng thể của VAE

Kiến trúc của một VAE bao gồm hai phần chính:

* Encoder (hay còn gọi là Recognition Model hoặc Inference Network): ánh xạ dữ liệu đầu vào sang một phân phối xác suất trong không gian tiềm ẩn . Thường thì encoder không sinh ra giá trị cụ thể mà dự đoán các tham số của phân phối Gaussian đa biến: trung bình μ và độ lệch chuẩn σ.
* Decoder (hay còn gọi là Generative Model): nhận mẫu từ không gian tiềm ẩn và tái tạo lại dữ liệu gốc, tức là sinh ra

VAE hoạt động thông qua sự kết hợp của mã hóa, lấy mẫu, giải mã và tối ưu hóa hàm mất mát. Bằng cách học một biểu diễn tiềm ẩn tối ưu, VAE có thể tái tạo chính xác dữ liệu.

Điểm đổi mới quan trọng nhất của VAE so với autoencoder truyền thống là cách mã hóa thông tin. Thay vì mã hóa một ảnh thành một điểm duy nhất trong không gian tiềm ẩn, VAE mã hóa nó thành một phân phối xác suất (thường là phân phối chuẩn, với giá trị trung bình và độ lệch chuẩn).

Nhờ cách tiếp cận này, mô hình có thể lấy mẫu ngẫu nhiên từ không gian tiềm ẩn, giúp quá trình sinh dữ liệu trở nên mượt mà và linh hoạt hơn. Đặc biệt, kỹ thuật “reparameterization trick” cho phép mô hình học được quá trình lấy mẫu này một cách hiệu quả trong quá trình huấn luyện.

Nếu không có phần xử lý xác suất này, VAE sẽ chỉ giống như một autoencoder thông thường – tức là chỉ nén và giải nén dữ liệu, không thật sự có khả năng tạo ra dữ liệu mới một cách tự nhiên và liên tục.

### Hàm mất mát – ELBO

Thay vì tối đa hóa trực tiếp xác suất log của dữ liệu ta tối ưu một cận dưới gọi là Evidence Lower Bound (ELBO):

Trong đó:

* là phân phối xấp xỉ hậu nghiệm do encoder học.
* là phân phối có điều kiện do decoder sinh ra.
* là độ đo phân kỳ Kullback-Leibler giữa phân phối xấp xỉ và phân phối chuẩn

### Reparameterization Trick

Một thách thức lớn trong việc huấn luyện VAE là quá trình lấy mẫu từ phân phối tiềm ẩn. Việc lấy mẫu từ một phân phối là một hoạt động không khả vi (non-differentiable) , điều này ngăn cản việc áp dụng lan truyền ngược (back-propagation) trực tiếp để tối ưu hóa các tham số của bộ mã hóa.

Để giải quyết vấn đề này, kỹ thuật Reparameterization Trick đã được giới thiệu. Kỹ thuật này cho phép lan truyền ngược thông qua quá trình lấy mẫu bằng cách biểu diễn biến tiềm ẩn dưới dạng

Trong đó

* μ là trung bình và σ là độ lệch chuẩn (hoặc exp(0.5 \* log\_var)) được xuất ra bởi bộ mã hóa.
* ε là nhiễu ngẫu nhiên được lấy mẫu từ phân phối Gaussian chuẩn

Kỹ thuật reparameterization trick không chỉ là một cách tối ưu hóa mà nó là yếu tố cơ bản cho phép huấn luyện VAE một cách hiệu quả. Nếu không có nó, bản chất xác suất của VAE (mã hóa thành một phân phối) sẽ khiến việc tối ưu hóa dựa trên gradient trở nên bất khả thi, từ đó ngăn mô hình học một không gian tiềm ẩn mượt mà và có ý nghĩa, điều cần thiết cho việc tạo sinh mạch lạc. Sự tương tác giữa mất mát tái tạo (tính trung thực) và phân kỳ KL (điều hòa/độ mượt) là một sự cân bằng tinh tế, nơi kỹ thuật reparameterization trick cho phép sự cân bằng này được học thông qua các kỹ thuật tối ưu hóa học sâu tiêu chuẩn.

### Ưu điểm và nhược điểm

#### Ưu điểm:

Khả năng tạo dữ liệu mới: VAE có khả năng tạo ra các mẫu dữ liệu mới tương tự dữ liệu huấn luyện nhưng không giống hệt bất kỳ trường hợp cụ thể nào, cho phép tạo ra nội dung sáng tạo.

Biểu diễn không gian tiềm ẩn liên tục và có cấu trúc: VAE học một không gian tiềm ẩn có cấu trúc và liên tục, giúp dễ dàng nội suy giữa các mẫu và thao tác các đặc trưng dữ liệu cụ thể (ví dụ: thay đổi ánh sáng hoặc biểu cảm khuôn mặt). Điều này thúc đẩy khả năng diễn giải của mô hình.

Học không giám sát: VAE có thể được huấn luyện trên dữ liệu không có nhãn, làm cho chúng phù hợp cho các kịch bản mà dữ liệu có nhãn khan hiếm hoặc đắt tiền để thu thập.

Điền dữ liệu và khử nhiễu: VAE đặc biệt mạnh mẽ trong việc tái tạo dữ liệu có phần bị thiếu hoặc nhiễu, ví dụ như khử nhiễu hình ảnh y tế.

Huấn luyện ổn định hơn: So với GANs, VAE thường ổn định hơn trong quá trình huấn luyện, ít gặp phải các vấn đề như mode collapse.

Đo lường khả năng xảy ra: VAE có thể được sử dụng để đo lường khả năng xảy ra của một hình ảnh dưới phân phối hình ảnh huấn luyện, hữu ích cho việc tìm kiếm dữ liệu ngoài phân phối hoặc các ví dụ điển hình/không điển hình.

#### Nhược điểm

Hình ảnh tạo ra thường bị mờ: Một trong những nhược điểm lớn nhất của VAE là hình ảnh tạo ra thường có xu hướng bị mờ (blurriness) và kém chân thực hơn so với GANs. Điều này là do bản chất xác suất của VAE và sự đánh đổi giữa việc giảm thiểu KL divergence và lỗi tái tạo. VAE ưu tiên bao phủ phân phối hơn là tái tạo từng pixel một cách hoàn hảo.

Thách thức trong huấn luyện: Mặc dù ổn định hơn GANs, việc huấn luyện VAE vẫn có thể phức tạp do hàm mục tiêu phức tạp và các vấn đề tiềm ẩn như mode collapse (mặc dù ít phổ biến hơn).

Chi phí tính toán cao: Huấn luyện VAE, đặc biệt với các kiến trúc phức tạp và tập dữ liệu lớn, có thể tốn kém về mặt tính toán.

Khó khăn trong diễn giải không gian tiềm ẩn: Mặc dù không gian tiềm ẩn của VAE có cấu trúc, việc diễn giải ý nghĩa cụ thể của từng chiều và mối quan hệ của chúng với các đặc trưng dữ liệu vẫn có thể là một thách thức.

Tốc độ lấy mẫu chậm hơn: Quá trình lấy mẫu để tạo ra dữ liệu mới trong VAE có thể chậm hơn so với GANs.

Giả định Gaussianity: Giả định rằng phân phối tiềm ẩn tuân theo phân phối Gaussian thường bị chỉ trích là một hạn chế, vì nó có thể không đủ biểu cảm để nắm bắt các phân phối dữ liệu phức tạp.

### Ứng dụng trong Thị giác máy tính

VAE, với khung xác suất và không gian tiềm ẩn có cấu trúc, đặc biệt có giá trị cho các ứng dụng đòi hỏi sự hiểu biết, thao tác dữ liệu và tổng hợp bảo vệ quyền riêng tư. Điều này định vị VAE như một công cụ mạnh mẽ cho tiện ích dữ liệu (khử nhiễu, điền dữ liệu, phát hiện bất thường) và tạo dữ liệu nhạy cảm với quyền riêng tư (hình ảnh y tế tổng hợp), thay vì chỉ tạo nội dung quang học chân thực.

Các ứng dụng quan trọng của VAE trong thị giác máy tính bao gồm:

* Mô hình tạo sinh (Generative Modeling): Khả năng cốt lõi của VAE là tạo ra các mẫu dữ liệu mới tương tự dữ liệu huấn luyện nhưng không giống hệt bất kỳ trường hợp cụ thể nào. Điều này hữu ích trong việc tạo ra các tác phẩm nghệ thuật mới, khuôn mặt chân thực hoặc các thiết kế mới trong thời trang và kiến trúc.
* Phát hiện bất thường (Anomaly Detection): Bằng cách học phân phối dữ liệu bình thường, VAE có thể xác định các điểm dữ liệu lệch đáng kể so với phân phối này là bất thường. Ứng dụng trong phát hiện gian lận tài chính, dự đoán lỗi thiết bị và phát hiện bất thường trong hình ảnh y tế.
* Điền dữ liệu và khử nhiễu (Data Imputation and Denoising): VAE có khả năng tái tạo dữ liệu có phần bị thiếu hoặc nhiễu. Ví dụ, chúng có thể khử nhiễu hình ảnh bị hỏng hoặc khôi phục dữ liệu y tế bị nhiễu, điều này rất cần thiết cho chẩn đoán chính xác.
* Học bán giám sát (Semi-supervised Learning): Trong các kịch bản học bán giám sát, VAE có thể cải thiện hiệu suất của bộ phân loại bằng cách sử dụng không gian tiềm ẩn để nắm bắt cấu trúc dữ liệu cơ bản, từ đó nâng cao quá trình học với dữ liệu có nhãn hạn chế.
* Thao tác không gian tiềm ẩn (Latent Space Manipulation): Không gian tiềm ẩn có cấu trúc của VAE cho phép điều chỉnh các đặc trưng cụ thể của hình ảnh (ví dụ: ánh sáng, biểu cảm khuôn mặt) bằng cách điều hướng trong không gian đó. Tính năng này đặc biệt hữu ích trong các ngành công nghiệp sáng tạo để chỉnh sửa và nâng cao hình ảnh và video.
* Tạo hình ảnh và thao tác: VAE đã được áp dụng rộng rãi cho các tác vụ tạo hình ảnh như dịch ảnh-sang-ảnh, siêu phân giải và tổng hợp hình ảnh.
* Tạo dữ liệu tổng hợp: Đặc biệt quan trọng là khả năng tạo ra hình ảnh y tế tổng hợp (ví dụ: X - quang, MRI scans) cho mục đích huấn luyện và nghiên cứu mà không ảnh hưởng đến quyền riêng tư của bệnh nhân.
* Nén hình ảnh và giảm chiều dữ liệu: VAE có thể học các biểu diễn nén của dữ liệu, giúp giảm chiều dữ liệu và nén hình ảnh hiệu quả.

## Diffusion Models

Diffusion Models (mô hình khuếch tán) là một trong những hướng tiếp cận tiên tiến nhất trong lĩnh vực AI tạo sinh, đặc biệt nổi bật trong việc sinh ảnh chất lượng cao. Các mô hình này dựa trên cơ sở lý thuyết của các quá trình Markov ngược và đã chứng minh hiệu quả vượt trội trong các tác vụ như tạo ảnh chân thực, phục hồi ảnh và chuyển đổi miền. Khác với GANs hoặc VAE, mô hình khuếch tán hoạt động bằng cách học cách đảo ngược một quá trình "nhiễu hóa" dữ liệu, từ đó tái tạo ảnh từ nhiễu ngẫu nhiên.

### Nguyên lý hoạt động

Mô hình khuếch tán (Diffusion Models), còn được gọi là mô hình tạo sinh dựa trên khuếch tán hoặc mô hình tạo sinh dựa trên điểm số, là một lớp mô hình tạo sinh biến tiềm ẩn đã đạt được thành công đáng kể trong việc tạo ra các mẫu dữ liệu chân thực.

Nguyên lý hoạt động của mô hình khuếch tán dựa trên hai quá trình chính:

* Quá trình khuếch tán thuận (Forward Diffusion Process):
* Quá trình này bắt đầu với một điểm dữ liệu gốc (ví dụ: một hình ảnh sạch).
* Nó dần dần thêm nhiễu Gaussian vào dữ liệu qua một chuỗi các bước thời gian T (được mô hình hóa như một chuỗi Markov).
* Ở mỗi bước, một lượng nhỏ nhiễu được thêm vào, làm cho dữ liệu ngày càng ngẫu nhiên và bị hỏng, cho đến khi ở bước cuối cùng , hình ảnh gần như trở thành nhiễu thuần túy, không thể phân biệt được với một phân phối Gaussian chuẩn.
* Điều quan trọng là quá trình khuếch tán thuận này là một quá trình cố định và không liên quan đến mô hình học máy nào.
* Quá trình lấy mẫu ngược (Reverse Sampling Process / Backward Diffusion):
* Đây là phần học được của mô hình. Mục tiêu là học cách đảo ngược quá trình thêm nhiễu đã xảy ra trong quá trình thuận.
* Bắt đầu từ nhiễu thuần túy (phân phối Gaussian) ở bước T, mô hình lặp đi lặp lại loại bỏ nhiễu, từng bước một, để tái tạo lại dữ liệu gốc .
* Quá trình này được thực hiện bởi một mạng nơ-ron (thường là U-Net, sẽ được thảo luận chi tiết hơn ở phần sau) được huấn luyện để dự đoán nhiễu được thêm vào ở mỗi bước của quá trình thuận.
* Bằng cách học cách dự đoán và loại bỏ nhiễu một cách dần dần, mô hình học được cách tái tạo dữ liệu gốc.

### Kiến trúc U-Net và Hàm mất mát

Để thực hiện quá trình khử nhiễu ngược một cách hiệu quả, các mô hình khuếch tán thường sử dụng một kiến trúc mạng nơ-ron cụ thể và một hàm mất mát được thiết kế để tối ưu hóa việc dự đoán nhiễu.

#### Kiến trúc U-Net

U-Net là một mạng nơ-ron tích chập (CNN) ban đầu được phát triển cho phân đoạn hình ảnh. Tuy nhiên, kiến trúc này đã được điều chỉnh và mở rộng để hoạt động hiệu quả trong các mô hình khuếch tán cho việc khử nhiễu hình ảnh lặp lại. U-Net được đặt tên theo hình dạng chữ "U" đặc trưng trong sơ đồ mô hình.

Kiến trúc U-Net bao gồm hai đường dẫn chính

* Đường co lại (Contracting Path / Encoder): Đây là một mạng tích chập điển hình bao gồm các lớp tích chập lặp lại, mỗi lớp theo sau là một hàm kích hoạt ReLU và một phép toán max pooling. Trong quá trình co lại, thông tin không gian của hình ảnh giảm đi (độ phân giải giảm), trong khi thông tin đặc trưng (số lượng kênh) tăng lên.
* Đường mở rộng (Expansive Path / Decoder): Đường dẫn này kết hợp thông tin đặc trưng và không gian thông qua một chuỗi các up-convolution (tích chập ngược) và nối (concatenation) với các đặc trưng độ phân giải cao từ đường co lại (thường được gọi là "skip connections"). Các skip connections này cho phép mạng truyền thông tin ngữ cảnh từ các lớp đầu tiên của encoder đến các lớp độ phân giải cao hơn trong decoder, giúp tái tạo lại các chi tiết tinh tế của hình ảnh.

#### Hàm mất mát (Loss Function)

Mục tiêu chính khi huấn luyện mô hình khuếch tán là đo lường sự khác biệt giữa nhiễu dự đoán của mô hình và nhiễu thực tế được thêm vào trong quá trình khuếch tán.

Hàm mất mát phổ biến nhất được sử dụng là Mean Squared Error (MSE) loss giữa nhiễu dự đoán của mô hình và nhiễu thực tế được áp dụng cho dữ liệu. Hàm này khuyến khích mô hình dự đoán chính xác nhiễu được thêm vào ở mỗi bước, dẫn đến việc tạo ra hình ảnh chất lượng cao.

Một số biến thể của mô hình khuếch tán sử dụng phiên bản có trọng số của MSE loss để tính đến các mức độ nhiễu khác nhau qua các bước thời gian, giúp mô hình tập trung vào việc khử nhiễu ở các trạng thái khó hơn.

Mặc dù mục tiêu cơ bản của mô hình khuếch tán được dựa trên giới hạn dưới biến phân (Variational Lower Bound - ELBO) , trong thực tế, MSE loss thường được sử dụng do hiệu quả tính toán và ổn định huấn luyện.

### Ưu điểm và Nhược điểm

#### Ưu điểm

Chất lượng đầu ra cao và chân thực: Mô hình khuếch tán nổi tiếng với khả năng tạo ra hình ảnh chất lượng cao, có độ chân thực vượt trội. Bằng cách tinh chỉnh nhiễu ngẫu nhiên một cách lặp đi lặp lại thông qua các quy trình khử nhiễu đã học, chúng thể hiện tính mạnh mẽ và tính linh hoạt đặc biệt.

Huấn luyện ổn định hơn: So với GANs, mô hình khuếch tán thường ổn định hơn trong huấn luyện và ít gặp phải các vấn đề như mode collapse hoặc động lực huấn luyện không ổn định.

Nền tảng lý thuyết vững chắc: Chúng được xây dựng trên các nguyên tắc được hiểu rõ từ các quá trình ngẫu nhiên và cơ học thống kê.

Khả năng tạo dữ liệu tổng hợp: Mô hình khuếch tán có thể được sử dụng hiệu quả để tạo ra dữ liệu tổng hợp, giúp tăng cường tập dữ liệu huấn luyện cho các mô hình học máy khác.

Phát hiện bất thường: Bằng cách mô hình hóa phân phối dữ liệu bình thường, mô hình khuếch tán có thể giúp xác định các điểm bất thường lệch khỏi phân phối này.

Tạo chuỗi token song song: Không giống như các mô hình tự hồi quy truyền thống, mô hình khuếch tán có thể tạo ra các phần chính xác của chuỗi token cuối cùng song song, có thể dẫn đến tốc độ nhanh hơn trong một số trường hợp.

Linh hoạt trong số lượt chạy: Có thể được huấn luyện để thực hiện ít lượt chạy hơn (đổi lấy chất lượng đầu ra thấp hơn), cho phép linh hoạt trong việc cân bằng giữa tốc độ và chất lượng.

#### Nhược điểm

Quá trình lấy mẫu/suy luận chậm: Ưu điểm về chất lượng của mô hình khuếch tán đến từ bản chất lặp lại của quá trình khử nhiễu. Điều này dẫn đến thời gian suy luận chậm hơn và tốn thời gian hơn đáng kể so với các mô hình một lần chạy như GANs và VAE.

Đầu ra có độ dài cố định: Mô hình khuếch tán thường tạo ra đầu ra có độ dài cố định, kém linh hoạt hơn các mô hình tự hồi quy trong việc tạo chuỗi có độ dài thay đổi.

Xử lý ngữ cảnh dài chậm: Chúng chậm hơn trong việc xử lý các cửa sổ ngữ cảnh dài, vì phải tính toán lại cơ chế chú ý cho toàn bộ cửa sổ ngữ cảnh ở mỗi lượt khử nhiễu, dẫn đến chi phí tính toán cao hơn.

Chi phí tính toán cao: Huấn luyện và chạy các mô hình khuếch tán, đặc biệt là các mô hình lớn và phức tạp, đòi hỏi sức mạnh tính toán đáng kể.

### Ứng dụng trong Thị giác máy tính

Mô hình khuếch tán đã trở thành kiến trúc thống trị cho việc tạo văn bản thành hình ảnh tiên tiến , tác động trực tiếp đến các ngành công nghiệp sáng tạo và nhận thức của công chúng về AI. Khả năng tạo ra hình ảnh có tính nhất quán cao và trực quan tuyệt đẹp từ các yêu cầu văn bản đơn giản đã dân chủ hóa nghệ thuật và thiết kế AI, giúp nhiều người tiếp cận được việc tổng hợp hình ảnh phức tạp.

Các ứng dụng quan trọng của mô hình khuếch tán trong thị giác máy tính bao gồm:

* Tạo hình ảnh chất lượng cao từ văn bản (Text-to-Image Generation): Đây là ứng dụng nổi bật nhất và đã tạo ra một làn sóng "Nghệ thuật AI" trên toàn thế giới. Các mô hình như DALL-E, Midjourney, và Stable Diffusion cho phép người dùng tạo ra các hình ảnh phức tạp và chân thực chỉ từ một vài từ mô tả. Điều này yêu cầu mô hình không chỉ hiểu ngữ nghĩa của văn bản mà còn cả biểu diễn trực quan của cảnh được mô tả.
* Tổng hợp hình ảnh và video (Image and video synthesis): Mô hình khuếch tán có khả năng tạo ra các hình ảnh và chuỗi video mới, chân thực.
* Khử nhiễu hình ảnh (Image denoising): Bằng cách học cách đảo ngược quá trình thêm nhiễu, mô hình khuếch tán có thể loại bỏ nhiễu khỏi hình ảnh, cải thiện chất lượng của chúng.
* Điền vào chỗ trống hình ảnh (Image inpainting): Khả năng lấp đầy các phần bị thiếu hoặc bị hỏng của hình ảnh một cách hợp lý và nhất quán.
* Siêu phân giải (Super-resolution): Nâng cao độ phân giải của hình ảnh từ chất lượng thấp lên chất lượng cao hơn, tạo ra các chi tiết sắc nét.
* Tạo dữ liệu tổng hợp (Synthetic Data Generation): Mô hình khuếch tán có thể tạo ra các tập dữ liệu đa dạng và chân thực, được sử dụng để tăng cường tập dữ liệu huấn luyện cho các mô hình học máy khác, đặc biệt hữu ích khi dữ liệu thực tế khan hiếm.
* Phát hiện bất thường (Anomaly Detection): Bằng cách mô hình hóa phân phối dữ liệu bình thường, mô hình khuếch tán có thể giúp xác định các điểm bất thường lệch khỏi phân phối này.
* Tạo tham số mạng nơ-ron: Một ứng dụng mới nổi là khả năng tạo ra các tham số mạng nơ-ron hiệu suất cao, mở ra hướng nghiên cứu mới trong thiết kế mô hình AI.
* Tạo hình ảnh có điều kiện hoặc không điều kiện: Mô hình có thể tạo ra hình ảnh mà không có bất kỳ hướng dẫn nào (vô điều kiện) hoặc theo các điều kiện cụ thể (có điều kiện), chẳng hạn như từ văn bản hoặc nhãn lớp.

## So sánh các mô hình AI tạo sinh

|  |  |  |  |
| --- | --- | --- | --- |
| Đặc điểm | GANs | VAE | Diffusion Models |
| Nguyên lý | Trò chơi đối nghịch giữa Generator và Discriminator | Mã hóa và giải mã xác suất, tối ưu ELBO | Quá trình khuếch tán và khử nhiễu nhiều bước |
| Chất lượng ảnh | Rất cao, đôi khi chân thực nhất | Thường mờ hơn, kém chân thực hơn GANs | Cao nhất hiện nay, siêu thực và chi tiết |
| Độ ổn định huấn luyện | Rất khó, dễ gặp mode collapse | Tương đối ổn định, dễ huấn luyện hơn | Tốt, ít gặp mode collapse |
| Đa dạng mẫu | Dễ bị mode collapse, kém đa dạng nếu không cẩn thận | Khá tốt, không gian tiềm ẩn liên tục | Rất tốt, đa dạng mẫu tạo sinh |
| Tốc độ tạo sinh | Nhanh (một lần chuyển tiếp) | Nhanh (một lần chuyển tiếp) | Chậm (nhiều bước lặp lại), cải thiện với LDM |
| Kiểm soát tạo sinh | Khó kiểm soát trực tiếp, cần cGAN | Tốt hơn (qua không gian tiềm ẩn), cVAE | Rất tốt (qua văn bản, điều kiện, và các bước) |
| Ưu điểm nổi bật | Ảnh rất sắc nét, chân thực | Không gian tiềm ẩn có ý nghĩa, dễ huấn luyện | Chất lượng ảnh vượt trội, ổn định hơn |
| Nhược điểm nổi bật | Khó huấn luyện, mode collapse | Ảnh mờ, chất lượng kém hơn | Tốc độ tạo sinh chậm, chi phí cao (trước LDM) |
| Ứng dụng chính | Tạo ảnh chân thực, chuyển đổi phong cách, data augmentation | Nén dữ liệu, phát hiện dị thường, khám phá không gian tiềm ẩn | Tạo ảnh từ văn bản, chỉnh sửa ảnh, super-resolution |

*Bảng 3.1 Bảng so sánh mô hình GANs, VAE và Diffusion Models*

# THỰC NGHIỆM

Chương này sẽ trình bày một số ví dụ thực nghiệm và demo mã nguồn (code demo) minh họa cho hoạt động của các mô hình AI tạo sinh đã được giới thiệu. Mục tiêu để đưa ra cái nhìn trực quan hơn về cách các mô hình này hoạt động và kết quả mà chúng có thể tạo ra.

## Dữ liệu thực nghiệm

Các thí nghiệm được thực hiện trên tập dữ liệu MNIST (Modified National Institute of Standards and Technology database), một bộ dữ liệu chuẩn trong nghiên cứu thị giác máy tính. MNIST bao gồm 60.000 hình ảnh chữ số viết tay cho mục đích huấn luyện và 10.000 hình ảnh cho mục đích kiểm tra. Mỗi hình ảnh có độ phân giải 28x28 pixel và thuộc một trong 10 lớp, đại diện cho các chữ số từ 0 đến 9. Dữ liệu được tải trực tiếp từ thư viện torchvision.datasets và được chuyển đổi thành định dạng tensor, sẵn sàng cho quá trình huấn luyện mô hình.

## Môi trường và công cụ thực nghiệm

Ngôn ngữ lập trình: Python

Thư viện sử dụng:

* Torch, torchvision: dùng để xây dựng và huấn luyện mạng nơ-ron nhân tạo, xử lý dữ liệu và tải bộ dữ liệu MNIST.
* Torchvision.utils: hỗ trợ lưu ảnh đầu ra từ mô hình.
* Tqdm: hiển thị tiến trình huấn luyện dạng thanh tiến độ.
* Os: xử lý thao tác thư mục và lưu file ảnh.

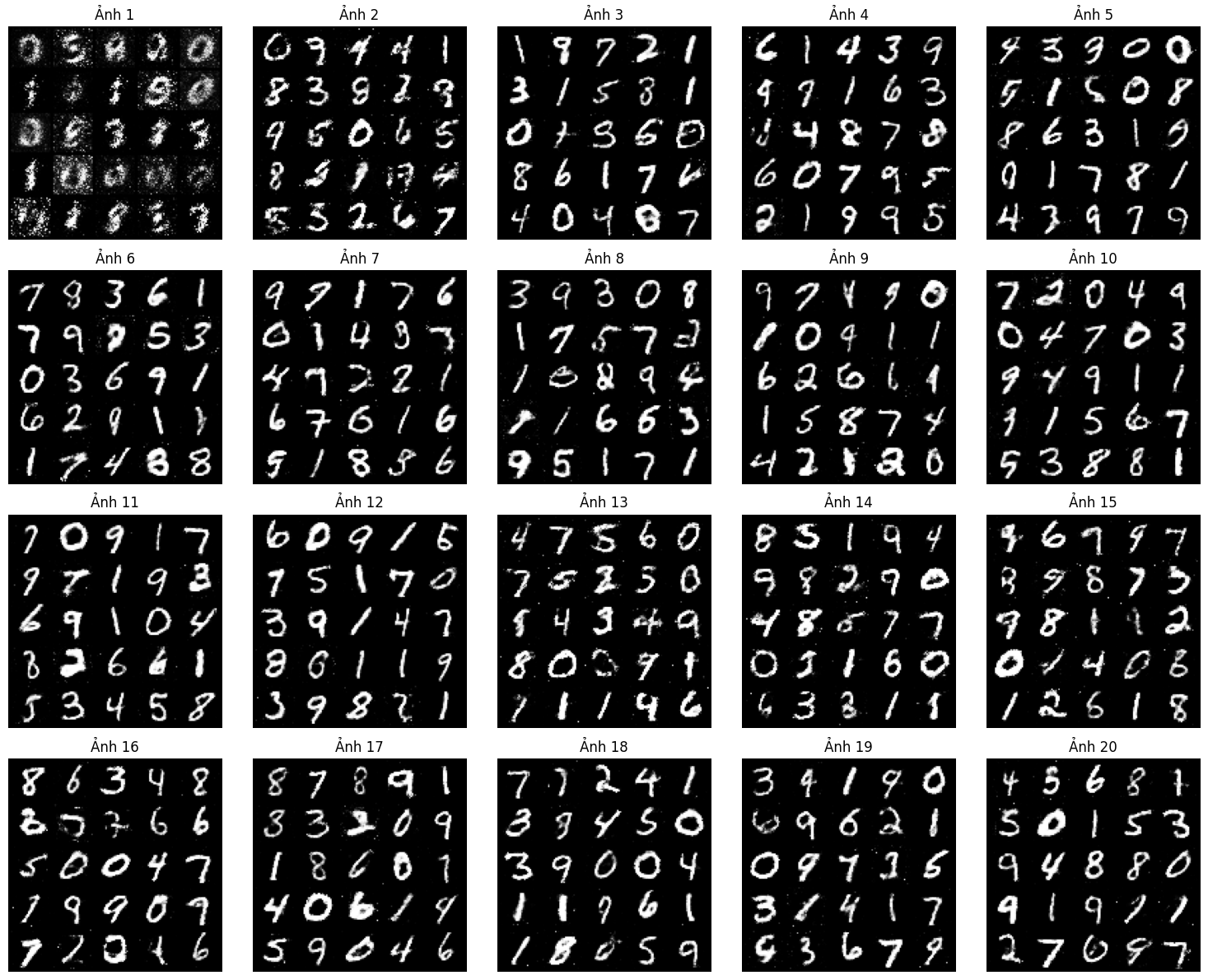
Thiết bị thực nghiệm: máy tính cá nhân với CPU (có hỗ trợ huấn luyện trên GPU nếu khả dụng)

## Demo 1: Tạo ảnh chữ số viết tay bằng GANs

Demo này sử dụng mạng GAN đơn giản (Vanilla GAN) được xây dựng từ đầu bằng thư viện PyTorch. Mô hình này được thiết kế để làm rõ cơ chế hoạt động cơ bản của GAN trong việc sinh ảnh chữ số viết tay trên bộ dữ liệu MNIST (ảnh grayscale kích thước 28×28).

Ý tưởng và cấu trúc mã nguồn:

* Khởi tạo mô hình:
* Tải dữ liệu huấn luyện MNIST và chuẩn hóa chúng.
* Xây dựng kiến trúc cho Generator (G) và Discriminator (D).
* Định nghĩa các hàm mất mát (loss functions) và bộ tối ưu hóa (optimizers) cho cả G và D.
* Quá trình huấn luyện:
* Lặp qua một số lượng epoch đã định (200 epoch).
* Trong mỗi epoch, huấn luyện bộ phân biệt bằng cách cung cấp cả ảnh thật và ảnh giả (do Generator tạo ra), sau đó cập nhật trọng số của nó dựa trên hàm mất mát.
* Tiếp theo, huấn luyện bộ tạo bằng cách cố gắng "đánh lừa" bộ phân biệt rằng ảnh giả là ảnh thật, sau đó cập nhật trọng số của nó.
* Ghi lại và hiển thị các giá trị D loss và G loss định kỳ để theo dõi sự hội tụ của mô hình.
* Tạo và lưu ảnh:
* Sau mỗi vài epoch, tạo ảnh chữ số viết tay từ vector nhiễu ngẫu nhiên bằng Generator.
* Lưu ảnh đầu ra để đánh giá chất lượng ảnh sinh được qua thời gian.



*Hình 4.1 Kết quả tạo ảnh chữ số viết tay bằng GANs*

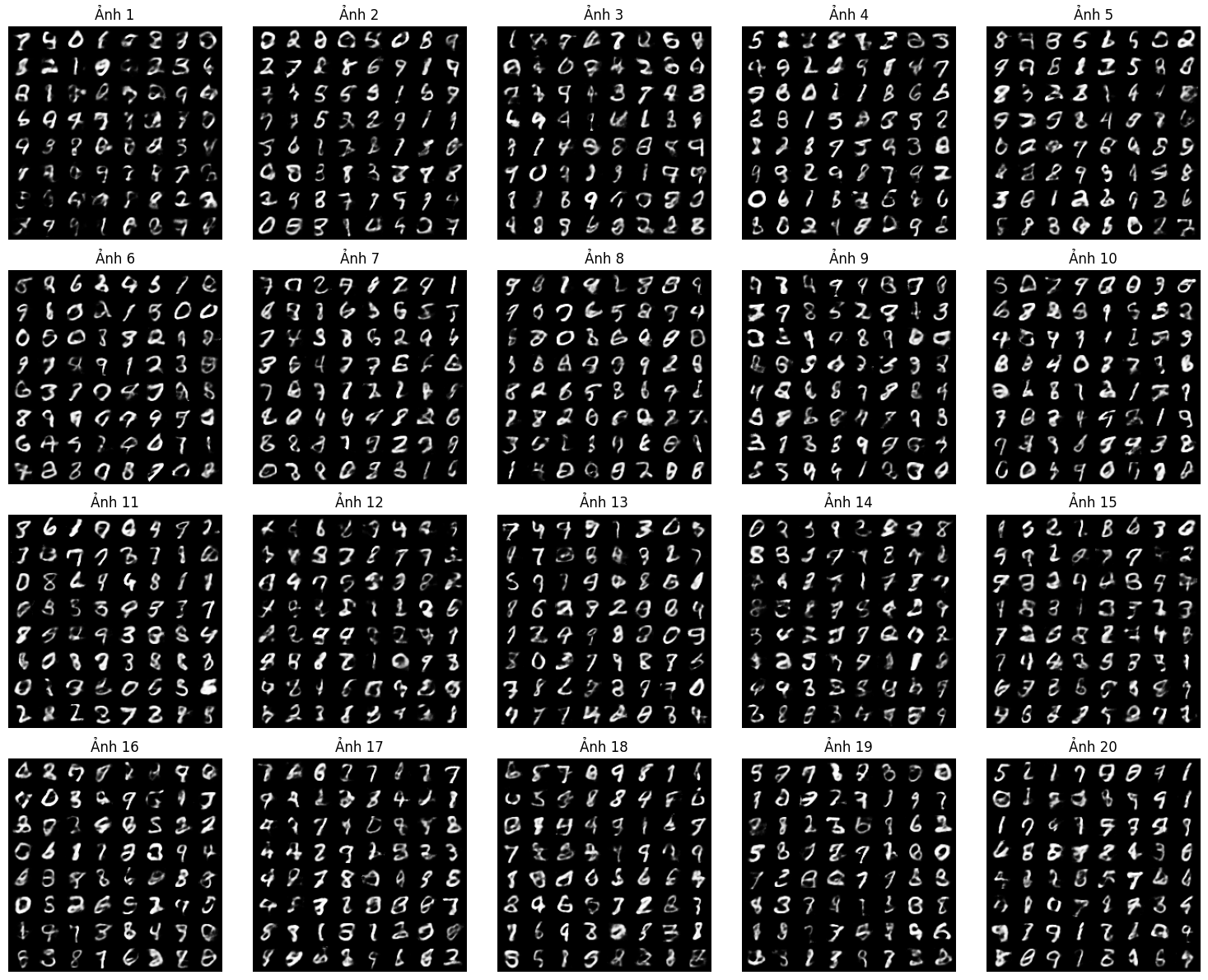
## Demo 2: Tạo ảnh chữ số viết tay bằng VAE

Demo này giới thiệu Bộ tự mã hóa biến phân (VAE), một kiến trúc mạng nơ-ron tạo sinh khác, cũng được triển khai bằng PyTorch. Mục tiêu của demo này là minh họa cách VAE học một biểu diễn tiềm ẩn của dữ liệu và sử dụng nó để tái tạo cũng như tạo ra các ảnh chữ số viết tay mới từ bộ dữ liệu MNIST.

Khác với GAN tập trung vào sự đối kháng giữa hai mạng, VAE học cách nén dữ liệu đầu vào thành một không gian tiềm ẩn phân phối Gauss và sau đó giải nén nó trở lại thành dữ liệu gốc, đồng thời đảm bảo không gian tiềm ẩn có cấu trúc mượt mà để sinh ảnh mới.

Ý tưởng và cấu trúc mã nguồn:

* Khởi tạo mô hình:
* Tải dữ liệu huấn luyện MNIST và chuẩn hóa chúng.
* Xây dựng kiến trúc cho Encoder (mã hóa đầu vào thành không gian tiềm ẩn) và Decoder (giải mã từ không gian tiềm ẩn thành dữ liệu đầu ra).
* Định nghĩa hàm mất mát của VAE, bao gồm mất mát tái tạo (Reconstruction Loss - BCE) và mất mát phân kỳ KL (KL Divergence).
* Định nghĩa bộ tối ưu hóa Adam cho mô hình VAE.
* Quá trình huấn luyện:
* Lặp qua các epoch huấn luyện.
* Trong mỗi epoch, dữ liệu đầu vào được truyền qua Encoder để tạo ra các tham số của phân phối tiềm ẩn (mean và log-variance).
* Một mẫu ngẫu nhiên được lấy từ phân phối này (kỹ thuật reparameterization trick).
* Mẫu này sau đó được truyền qua Decoder để tái tạo ảnh.
* Tính toán tổng mất mát (BCE Loss + KL Loss) và cập nhật trọng số của mô hình.
* Ghi lại và hiển thị giá trị mất mát trung bình (Average Loss) định kỳ để theo dõi sự hội tụ.
* Tạo và lưu ảnh:
* Sau khi huấn luyện, sử dụng mô hình Decoder đã huấn luyện.
* Tạo một vector nhiễu ngẫu nhiên từ phân phối Gauss (tức là lấy mẫu trực tiếp từ không gian tiềm ẩn).
* Decoder sẽ tạo ra một hình ảnh chữ số viết tay từ vector nhiễu đó.
* Lưu và hiển thị các hình ảnh tạo ra.



*Hình 4.2 Kết quả tạo ảnh chữ số viết tay bằng VAE*

## Demo 3: Tạo ảnh chữ số viết tay bằng Diffusion Model

Demo này giới thiệu Mô hình Khuếch tán (Diffusion Model), một kiến trúc tạo sinh mạnh mẽ, được triển khai hoàn toàn từ đầu bằng thư viện PyTorch. Mục tiêu của demo này là minh họa cách một Diffusion Model cơ bản học cách tạo ra ảnh chữ số viết tay từ bộ dữ liệu MNIST thông qua một quá trình khử nhiễu lặp đi lặp lại.

Ý tưởng và cấu trúc mã nguồn:

* Khởi tạo mô hình:
* Tải và chuẩn hóa bộ dữ liệu MNIST.
* Xây dựng kiến trúc U-Net cải tiến (ContextUnet) với nhúng thời gian và nhúng lớp.
* Tạo lịch trình nhiễu (ddpm\_schedules) theo công thức của DDPM.
* Khởi tạo mô hình DDPM, hàm mất mát MSE và tối ưu hóa bằng Adam.
* Quá trình huấn luyện:
* Thêm nhiễu vào ảnh theo bước thời gian ngẫu nhiên.
* Mô hình học dự đoán lại nhiễu dựa trên ảnh nhiễu, thời gian và lớp.
* Áp dụng context\_mask để mô hình học cả chế độ có điều kiện và không điều kiện.
* Cập nhật trọng số theo sai số giữa nhiễu dự đoán và nhiễu thật.
* Tạo và lưu ảnh:
* Sau mỗi epoch, sinh 20 ảnh từ nhiễu trắng qua 400 bước khử nhiễu.
* Áp dụng kỹ thuật classifier-free guidance để cải thiện ảnh sinh.
* Lưu ảnh đầu ra.



*Hình 4.3 Kết quả tạo ảnh chữ số viết tay bằng Diffusion Model*

# KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

## Kết luận

Trong bối cảnh AI tạo sinh đang ngày càng đóng vai trò trung tâm trong các ứng dụng thị giác máy tính, báo cáo này đã tập trung phân tích ba mô hình tiêu biểu gồm GANs, VAE và Diffusion Models – mỗi mô hình đại diện cho một hướng tiếp cận khác nhau trong quá trình sinh ảnh. Qua việc trình bày nguyên lý hoạt động, kiến trúc, ưu và nhược điểm cũng như các ứng dụng thực tiễn, từ đó có được cái nhìn toàn diện về cơ chế và tiềm năng của từng mô hình.

* GANs nổi bật với khả năng tạo ảnh chân thực, tốc độ suy luận nhanh, nhưng đi kèm là khó khăn trong huấn luyện và hiện tượng mode collapse.
* VAE mang lại không gian tiềm ẩn có cấu trúc, dễ diễn giải và ứng dụng tốt trong phát hiện dị thường, nén dữ liệu, tuy nhiên chất lượng ảnh thường bị mờ và kém chân thực.
* Diffusion Models được đánh giá là có chất lượng sinh ảnh cao nhất hiện nay, với độ ổn định huấn luyện tốt và khả năng kiểm soát mạnh mẽ, nhưng lại tốn thời gian suy luận và chi phí tính toán cao.

Thông qua các demo thực nghiệm với tập dữ liệu MNIST, báo cáo không chỉ minh họa trực quan khả năng tạo ảnh của các mô hình mà còn cung cấp căn cứ thực tiễn để so sánh hiệu quả giữa chúng. Từ đó, có thể khẳng định rằng không có mô hình nào vượt trội hoàn toàn, mà việc lựa chọn phụ thuộc vào bài toán cụ thể, yêu cầu về chất lượng, tốc độ và khả năng kiểm soát.

## Hướng phát triển

Trong tương lai, nghiên cứu và ứng dụng AI tạo sinh trong thị giác máy tính có thể tập trung vào các hướng sau:

* Tối ưu hóa mô hình Diffusion, giảm thời gian suy luận bằng các kỹ thuật như Latent Diffusion Model (LDM), kết hợp attention hiệu quả hơn, hoặc thiết kế pipeline lấy mẫu ít bước hơn mà vẫn giữ chất lượng ảnh.
* Tích hợp cấu trúc GANs với Diffusion hoặc VAE để tận dụng ưu thế về tốc độ suy luận và độ chân thực (như các mô hình GDM – GAN-enhanced Diffusion Models).
* Mở rộng khả năng sinh ảnh từ nhiều dạng dữ liệu kết hợp như văn bản, âm thanh, hay dữ liệu cảm biến – thúc đẩy các ứng dụng như AI trợ lý thiết kế, tổng hợp cảnh 3D từ ngữ cảnh, v.v.
* Ứng dụng trong y tế và công nghiệp để tạo ảnh y tế tổng hợp bảo vệ quyền riêng tư, mô phỏng lỗi trong kiểm tra chất lượng sản phẩm, hoặc tái tạo hình ảnh trong điều kiện thiếu dữ liệu thực tế.
* Phát triển các cơ chế kiểm soát trực quan và giải thích được quá trình tạo sinh, đặc biệt quan trọng với các lĩnh vực yêu cầu tính minh bạch như tài chính, y tế, luật pháp.

# TÀI LIỆU THAM KHẢO

1. Goodfellow, Ian & Pouget-Abadie, Jean & Mirza, Mehdi & Xu, Bing & Warde-Farley, David & Ozair, Sherjil & Courville, Aaron & Bengio, Y.. (2014). Generative Adversarial Networks. Advances in Neural Information Processing Systems. 3. 10.1145/3422622.
2. Mao, Xudong & Li, Qing & Xie, Haoran & Lau, Raymond & Zhen, Wang & Smolley, Stephen. (2017). Least Squares Generative Adversarial Networks. 2813-2821. 10.1109/ICCV.2017.304.
3. Arjovsky, Martin & Chintala, Soumith & Bottou, Léon. (2017). Wasserstein GAN. 10.48550/arXiv.1701.07875.
4. Gulrajani, Ishaan & Ahmed, Faruk & Arjovsky, Martin & Dumoulin, Vincent & Courville, Aaron. (2017). Improved Training of Wasserstein GANs. 10.48550/arXiv.1704.00028.
5. Lopez, Ryan & Atzberger, Paul. (2020). Variational Autoencoders for Learning Nonlinear Dynamics of Physical Systems. 10.48550/arXiv.2012.03448.
6. Rombach, Robin & Blattmann, Andreas & Lorenz, Dominik & Esser, Patrick & Ommer, Bjorn. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. 10674-10685. 10.1109/CVPR52688.2022.01042.
7. Croitoru, Florinel-Alin & Hondru, Vlad & Ionescu, Radu Tudor & Shah, Mubarak. (2023). Diffusion Models in Vision: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. PP. 1-20. 10.1109/TPAMI.2023.3261988.
8. Yang, L., Zhang, Z., Song, Y., Hong, S., & Yang, M.-H. (2023). A Survey of Diffusion Models in Generative Modeling. arXiv preprint arXiv:2303.04197. <https://arxiv.org/abs/2303.04197>
9. Iglesias, Guillermo & Talavera, Edgar & Díaz, Alberto. (2022). A survey on GANs for computer vision: Recent research, analysis and taxonomy. Computer Science Review. 48. 10.1016/j.cosrev.2023.100553.
10. Viblo. (2021). Generative Adversarial Networks (GAN) và ứng dụng của nó trong Deepfakes. Viblo. <https://viblo.asia/p/generative-adversarial-networksgan-va-ung-dung-cua-no-trong-deepfakes-jvElagwdKkw>
11. TeaPearce. (2022). *Conditional Diffusion for MNIST* [Source code]. GitHub. <https://github.com/TeaPearce/Conditional_Diffusion_MNIST>

# PHỤ LỤC

Nhằm minh họa trực quan cho nguyên lý hoạt động và khả năng sinh dữ liệu của các mô hình AI tạo sinh được trình bày trong báo cáo, ba ứng dụng demo đã được xây dựng và triển khai trên tập dữ liệu MNIST. Các mô hình đều được huấn luyện trên cùng một tập dữ liệu, với mục tiêu tạo ra các ảnh chữ số viết tay nhân tạo, từ đó hỗ trợ quá trình quan sát, đánh giá và so sánh hiệu quả sinh ảnh của từng kiến trúc.

Ba mô hình được triển khai bao gồm:

* Mạng đối kháng tạo sinh (Generative Adversarial Network – GAN)
* Bộ tự mã hóa biến phân (Variational Autoencoder – VAE)
* Mô hình khuếch tán (Diffusion Model)

Toàn bộ mã nguồn thực nghiệm được công khai và lưu trữ tại kho lưu trữ GitHub sau: [Link mã nguồn đầy đủ](https://github.com/Huynneh/GANs-VAE-Diffusion-MNIST.git)