**BỘ CÔNG THƯƠNG**



**TRƯỜNG ĐẠI HỌC CÔNG THƯƠNG TP HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

-----o0o-----



**ĐỒ ÁN CHUYÊN NGÀNH**

**ĐỀ TÀI: WEBSITE SÁNG TẠO VÀ CHIA SẺ HÌNH ẢNH**

**Giáo Viên Hướng Dẫn**: TS. Nguyễn Thị Bích Ngân

**Sinh Viên Thực Hiện**:

Trần Hoàng Anh Khôi

Nguyễn Bảo Long – 2001210745 – 12DHTH04

Nguyễn Văn Chí Thanh – 2001210055 – 12DHTH04

Thành Phố Hồ Chí Minh, tháng 12 năm 2024

**BỘ CÔNG THƯƠNG**



**TRƯỜNG ĐẠI HỌC CÔNG THƯƠNG TP HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

-----o0o-----



**ĐỒ ÁN CHUYÊN NGÀNH**

**ĐỀ TÀI: WEBSITE SÁNG TẠO VÀ CHIA SẺ HÌNH ẢNH**

**Giáo Viên Hướng Dẫn**: TS. Nguyễn Thị Bích Ngân

**Sinh Viên Thực Hiện**:

1. Trần Hoàng Anh Khôi
2. Nguyễn Bảo Long – 2001210745 – 12DHTH04
3. Nguyễn Văn Chí Thanh – 2001210055 – 12DHTH04

Thành Phố Hồ Chí Minh, tháng 12 năm 2024

# Lời Cam Đoan

Em xin cam đoan đây là công trình nghiên cứu do chính bản thân Em thực hiện. Các kết quả, số liệu và trích dẫn trong báo cáo này là hoàn toàn trung thực và chưa từng được sử dụng trong bất kỳ công trình nghiên cứu nào khác. Mọi sự trợ giúp, hỗ trợ từ các thầy, cô, bạn bè và những người khác đều được Em ghi nhận và cảm ơn.

Nếu có bất kỳ sai sót nào, Em xin chịu hoàn toàn trách nhiệm.

TP.HCM, ngày……tháng……năm 2024

Sinh viên thực hiện đồ án

(Ký và ghi rõ họ tên)

Sinh viên 1 Sinh viên 2 Sinh viên 3

# Lời Cảm Ơn

Trong quá trình thực hiện đồ án tốt nghiệp này, Em đã nhận được sự hỗ trợ, giúp đỡ quý báu từ nhiều người.

Trước hết, Em xin bày tỏ lời cảm ơn chân thành tới cô Nguyễn Thị Bích Ngân người đã tận tình hướng dẫn, góp ý chuyên môn và động viên em trong suốt quá trình thực hiện đồ án. Những chia sẻ và định hướng của cô đã giúp em hoàn thiện và nâng cao chất lượng của đồ án.

Nhóm của chúng Em cũng xin được gửi lời cảm ơn đến quý thầy, cô giáo trường Đại Học Công Thương TP.HCM, đặc biệt là các thầy, cô khoa Công Nghệ Thông Tin - những người đã truyền lửa và giảng dạy kiến thức cho em suốt thời gian qua.

Mặc dù đã có những đầu tư nhất định trong quá trình làm bài song cũng khó có thể tránh khỏi những sai sót, Em kính mong nhận được ý kiến đóng góp của quý thầy cô để đồ án này được hoàn thiện hơn

Nhóm của chúng Em xin chân thành cảm ơn!

Sinh viên thực hiện đồ án

Sinh viên 1 Sinh viên 2 Sinh viên 3

# Nhận Xét Của Giáo Viên Hướng Dẫn

Mục lục

[Lời Cam Đoan i](#_Toc183712245)

[Lời Cảm Ơn ii](#_Toc183712246)

[Nhận Xét Của Giáo Viên Hướng Dẫn iii](#_Toc183712247)

[Danh Mục Hình Ảnh vii](#_Toc183712248)

[Danh Mục Từ Viết Tắt viii](#_Toc183712249)

[LỜI Mở Đầu 1](#_Toc183712250)

[Chương 1. Tổng Quan Về Đề Tài 2](#_Toc183712251)

[1.1 Lý Do Chọn Đề Tài 2](#_Toc183712252)

[1.2 Mục Tiêu Đề Tài 3](#_Toc183712253)

[1.3 Phạm Vi Đề Tài 3](#_Toc183712254)

[1.4 Công Nghệ Sử Dụng 3](#_Toc183712255)

[1.5 Khó Khăn và Thách Thức 4](#_Toc183712256)

[1.5.1 Thách thức về công nghệ tạo sinh 4](#_Toc183712257)

[1.5.2 Cơ sở dữ liệu và bảo mật 5](#_Toc183712258)

[1.5.3 Quản lý nội dung và trải nghiệm người dùng 5](#_Toc183712259)

[1.5.4 Khó khăn về thiết bị và xử lý dữ liệu 5](#_Toc183712260)

[1.6 Ý Nghĩa Phát Triển và Thực Tiển của Đề Tài 6](#_Toc183712261)

[1.6.1 Ý nghĩa phát triển 6](#_Toc183712262)

[1.6.2 Ý nghĩa thực tiễn 6](#_Toc183712263)

[1.7 Cấu Trúc Đồ Án 7](#_Toc183712264)

[Chương 2. Cơ Sở Lý Thuyết 8](#_Toc183712265)

[2.1 Khái niệm cơ bản 8](#_Toc183712266)

[2.1.1 Laravel là gì? 8](#_Toc183712267)

[2.1.2 Laravel hoạt động như thế nào? 8](#_Toc183712268)

[2.1.3 R2 Object Storage 9](#_Toc183712269)

[2.1.4 Thông báo realtime với Pusher 11](#_Toc183712270)

[2.1.5 Stable Diffusion mô hình tạo sinh 11](#_Toc183712271)

[2.1.6 Stable Diffusion gồm những kiến trúc nào? 14](#_Toc183712272)

[2.1.7 Mô hình nổi tiếng của Stable Diffusion 16](#_Toc183712273)

[2.1.8 Các Extension nổi tiếng của Stable Diffusion 21](#_Toc183712274)

[2.1.9 Ứng dụng mô hình AI tạo sinh với Comfyui 23](#_Toc183712275)

[Chương 3. Phân tích yêu cầu và thiết kế hệ thống 27](#_Toc183712276)

[Chương 4. ??? 1](#_Toc183712277)

[Kết Luận 1](#_Toc183712278)

[Tài Liệu Tham Khảo ii](#_Toc183712279)

# Danh Mục Hình Ảnh

# Danh Mục Từ Viết Tắt

# LỜI Mở Đầu

Trong bối cảnh công nghệ tạo sinh phát triển mạnh mẽ và nhu cầu chia sẻ hình ảnh do AI tạo sinh tạo ra ngày càng tăng cao, việc xây dựng một nền tảng trực tuyến hiệu quả để kết nối những người yêu thích sáng tạo và chia sẻ hình ảnh là một xu thế tất yếu. Đồ án tốt nghiệp của em tập trung nghiên cứu và phát triển một website có ý tưởng và chức năng tương tự Pinterest, đồng thời ứng dụng các công nghệ tiên tiến về xử lý và tạo ảnh.

Mục tiêu chính của đồ án là xây dựng một nền tảng web có khả năng:

* Cho phép người dùng tải lên, chia sẻ và lưu trữ hình ảnh
* Sử dụng các mô hình trí tuệ nhân tạo tiên tiến để tạo và chỉnh sửa ảnh
* Cung cấp trải nghiệm người dùng trực quan, thuận tiện và an toàn
* Tích hợp các tính năng tương tác xã hội như bình luận, thích và theo dõi

Trong quá trình thực hiện, Em đã áp dụng các kiến thức chuyên sâu về phát triển web, thiết kế giao diện người dùng, quản lý cơ sở dữ liệu và ứng dụng các công nghệ AI tiên tiến. Đồ án không chỉ là minh chứng cho quá trình học tập mà còn thể hiện khả năng ứng dụng công nghệ của bản thân.

Báo cáo này sẽ trình bày chi tiết quá trình nghiên cứu, phân tích, thiết kế và triển khai website, từ việc lựa chọn công nghệ, phân tích yêu cầu cho đến kết quả cuối cùng và các hướng phát triển tiềm năng.

Bằng cách kết hợp giữa công nghệ web hiện đại và các mô hình trí tuệ nhân tạo tiên tiến, em mong muốn đóng góp một giải pháp mới trong lĩnh vực chia sẻ và tạo lập hình ảnh trực tuyến.

# Tổng Quan Về Đề Tài

## Lý Do Chọn Đề Tài

Việc chọn đề tài “Website sáng tạo và chia sẻ hình ảnh” bắt nguồn từ các phong trào sử dụng sản phẩm của các AI tạo sinh đang ngày càng phổ biến. Trong đó hình ảnh do GenAI làm ra hiện đang rất phổ biến và việc phát triển một website chia sẻ những hình ảnh sản phẩm do GenAI tạo ra là một việc tất yếu dựa vào các yếu tố sau:

**Xu Hướng Chia Sẻ Hình Ảnh Số**. Trong kỷ nguyên số hiện nay, hình ảnh đã trở thành phương tiện giao tiếp quan trọng nhất. Các nền tảng chia sẻ hình ảnh như Instagram, Pinterest đã chứng minh sức hấp dẫn to lớn của mình, với hàng trăm triệu người dùng trên toàn cầu. Việc phát triển một nền tảng mới với những cải tiến về công nghệ và trải nghiệm người dùng là một thách thức hấp dẫn và có ý nghĩa.

**Ứng Dụng Công Nghệ Trí Tuệ Nhân Tạo**. Sự phát triển nhanh chóng của các mô hình AI như DALL-E, Midjourney và Stable Diffusion đã mở ra những khả năng hoàn toàn mới trong việc tạo và chỉnh sửa hình ảnh. Đồ án của Em muốn khai thác và tích hợp những công nghệ tiên tiến này, tạo ra một nền tảng không chỉ là nơi chia sẻ mà còn là không gian sáng tạo cho người dùng.

**Nhu Cầu CỘNG ĐỒNG SÁNG TẠO**. Qua quá trình nghiên cứu, Em nhận thấy có một cộng đồng ngày càng lớn mạnh của các nhà thiết kế, nhiếp ảnh gia, sinh viên nghệ thuật và những người yêu thích sáng tạo (Stable Diffusion Việt Nam với hơn 100 nghìn thành viên trên mạng xã hội Facebook). Họ cần một môi trường an toàn, thuận tiện để chia sẻ, học hỏi và phát triển ý tưởng.

**Cơ Hội Học Tập và Phát Triển**. Đồ án không chỉ là sản phẩm cuối khóa mà còn là cơ hội để áp dụng toàn bộ kiến thức đã học tại chuyên ngành Công nghệ Phần mềm. Từ việc phân tích yêu cầu, thiết kế hệ thống đến triển khai, Em có thể vận dụng trọn vẹn những kỹ năng chuyên môn.

## Mục Tiêu Đề Tài

Tìm hiểu các kỹ thuật mới trong việc xây dựng website.

Học hỏi và vận dụng các chức năng của framework Laravel một cách thành thạo.

Tìm hiểu các mô hình đã được huấn luyện sẵn để sáng tạo hình ảnh.

Xây dựng ứng dụng website.

## Phạm Vi Đề Tài

Tập trung vào nghiệp vụ chia sẻ hình ảnh giữa các người dùng.

Xây dựng các chức năng sáng tạo hình ảnh từ mô hình đã huấn luyện sẵn.

Xây dựng các chức năng thể hiện các hình ảnh theo sở thích.

Thêm các tính năng tương tác giữa các người dùng.

Quản lý hình ảnh từ người dùng tải lên cộng đồng

## Công Nghệ Sử Dụng

* Front-End:

**Ngôn ngữ sử dụng**: HTML, CSS, JavaScript – Jquery

**Framework**: BootStrap, Tailwind CSS

* Back-End:

**Ngôn ngữ sử dụng**: PHP

**Framework**: Laravel 11

* Database:

**Cơ sở dữ liệu**: MySQL

* Lưu trữ cloud:

**Lưu trữ dữ liệu**: CloudFlare R2

* Công nghệ mới:

**Mô hình tạo sinh**: Các mô hình được huấn luyện bởi Stability AI và cộng đồng người dùng.

* Websocket:

**Thông báo realtime**: Pusher

* Khác:

**Xác thực không phải là bot**: Turnstile

**Quản lý source code**: Github

## Khó Khăn và Thách Thức

### Thách thức về công nghệ tạo sinh

Việc tích hợp các mô hình đã được huấn luyện thông qua các API. Xử lý API phức tạp, khó đảm bảo tốc độ phản hồi nhanh. Khó khăn trong việc xử lý hàng loạt các yêu cầu tạo ảnh với tốc độ cao, tối ưu lưu trữ hình ảnh của người dùng sau khi sử dụng các chức năng tạo sinh cũng bị hạn chế trong 10GB dữ liệu với 1 triệu lượt ghi và 10 triệu lượt lấy dữ liệu từ cloud R2.

Không có cơ chế để đảm bảo chất lượng hình ảnh và hạn chế các sản phẩm ảnh nhạy cảm từ việc người dùng dùng AI tạo sinh cũng như các vấn đề về bản quyền.

### Cơ sở dữ liệu và bảo mật

Sử dụng cơ sở dữ liệu quan hệ Mysql dẫn đến khó khăn trong việc phát triển mở rộng hơn sau này. Tốc độ lấy dữ liệu không nhanh khi có hơn 1 triệu bản ghi hình ảnh.

Cơ chế bảo mật chống lại SQL Injection (SQLi) cũng hạn chế chỉ giới hạn việc sử dụng hàm Hash và @csrf có sẵn trong Laravel.

Không có cơ chế bảo mật API nào được thực hiện.

### Quản lý nội dung và trải nghiệm người dùng

Chưa xây dựng được hệ thống nhận diện nội dung, hình ảnh độc hại, phản cảm. Việc phân loại xét duyệt nội dùng phụ thuộc toàn bộ vào người dùng.

Tải lên hình ảnh và phân loại ảnh chưa thể tự động, việc phân loại là tùy thuộc vào người dùng đăng ảnh.

Giao diện UI/UX còn gặp khó khăn khi phát triển đa nền tảng. Hiện chỉ có thể tập trung vào nền tảng website trên máy tính và điện thoại.

Khâu thanh toán hiện chưa thể thực hiện do liên quan tới các vấn đề pháp lý và chi phí phát sinh. Thanh toán hiện tại chỉ là tượng chưng khi người dùng chọn thanh toán.

### Khó khăn về thiết bị và xử lý dữ liệu

Việc tích hợp nhiều mô hình tạo sinh được huấn luyện từ trước cần một GPU tiên tiến. Đi kèm với đó GPU càng tiên tiến chi phí vận hành và mua sắm thiết bị càng cao. Nhóm đang sử dụng GPU 3070ti với 6000 nhân CUDA sử dụng tối đa sức mạnh sẽ tiêu hao 300w điện và thiết bị làm lạnh cần phải được sử dụng liên tục để tránh tình trạng quá nhiệt.

Xử lý dữ liệu lớn còn rất hạn chế. Hiện không thể xử lý được các yêu cầu khi xảy ra tình trạng bùng nổ dữ liệu phát sinh hơn 1 triệu bản ghi.

## Ý Nghĩa Phát Triển và Thực Tiển của Đề Tài

### Ý nghĩa phát triển

Sử dụng AI tạo sinh (Generative AI) để tạo hình ảnh không chỉ mang tính đổi mới mà còn thúc đẩy sự phát triển của ngành công nghệ sáng tạo nội dung việc biến trí tưởng tượng trong tâm trí thành hiện thực đã không còn xa vời.

Khả năng tạo ra hình ảnh từ mô tả hoặc từ khuôn mặt cá nhân là bước tiến lớn trong cá nhân hóa trải nghiệm người dùng, mở ra cơ hội ứng dụng trong nhiều lĩnh vực khác nhau như nghệ thuật, giải trí, và giáo dục. Chức năng như "tạo người mẫu ảo" hoặc "tạo hình ảnh vui nhộn" giúp người dùng dễ dàng tham gia sáng tạo mà không cần kỹ năng thiết kế chuyên nghiệp.Tạo ra không gian cộng tác mới giữa các cá nhân và doanh nghiệp sáng tạo.

Xây dựng một nền tảng tương tự Pinterest với chức năng tương tác hình ảnh và tích hợp AI sẽ mang lại trải nghiệm đa dạng hơn, từ việc khám phá, lưu trữ đến sáng tạo nội dung mới.

### Ý nghĩa thực tiễn

Đề tài không chỉ mang ý nghĩa học thuật cao khi tích hợp thành công các công nghệ mới vào nền tảng web mà còn đáp ứng nhu cầu thực tế của thị trường và người dùng.

Với sự phát triển không ngừng của công nghệ AI, đề tài có tiềm năng mở rộng và triển khai trong nhiều lĩnh vực kinh tế, xã hội.Đây là một bước đi đáng giá để ứng dụng GenAI vào thực tiễn, nâng cao trải nghiệm người dùng và tạo ra giá trị mới cho cộng đồng số.

## Cấu Trúc Đồ Án

Lời mở đầu

Chương 1: Tổng quan về đề tài

Chương 2: Cơ sở lí thuyết

Chương 3: Thiết kế phân tích ứng dụng

Chương 4: Triển khai ứng dụng

Kết luận

Tài liệu tham khảo

# Cơ Sở Lý Thuyết

## Khái niệm cơ bản

### Laravel là gì?

Laravel là một framework PHP mã nguồn mở, được thiết kế để phát triển các ứng dụng web với cấu trúc rõ ràng và hiện đại. Nó giúp lập trình viên phát triển nhanh hơn, dễ bảo trì hơn và giảm thiểu viết đi viết lại những đoạn code phổ biến. Laravel đi kèm với nhiều tính năng mạnh mẽ như route, middleware, Eloquent ORM, Blade template, và công cụ Artisan CLI.

### Laravel hoạt động như thế nào?

Laravel sử dụng kiến trúc MVC, chia ứng dụng thành 3 phần chính:

* **Model:**  
  Chịu trách nhiệm làm việc với cơ sở dữ liệu. Model định nghĩa dữ liệu và các thao tác liên quan, sử dụng Eloquent ORM để thao tác với bảng dữ liệu.  
  *Ví dụ: Bảng users có model là User.php.*
* **View:**  
  Là nơi hiển thị giao diện người dùng. Laravel sử dụng Blade template để thiết kế giao diện, dễ dàng sử dụng các cú pháp động như hiển thị dữ liệu từ model.  
  *Ví dụ: File welcome.blade.php.*
* **Controller:**  
  Điều phối dữ liệu giữa Model và View. Controller nhận yêu cầu từ người dùng, xử lý logic, và trả kết quả cho View để hiển thị.  
  *Ví dụ: Controller UserController.php xử lý hiển thị danh sách người dùng từ model User.*

Ngoài ra còn nhiều thành phụ trợ khác cho để xây dựng các sự kiện xảy ra trong ứng dụng như:

* + Routing xác định các luồng yêu cầu.
  + Middleware kiểm tra trung gian trước khi tới controller.
  + Blade Template là template engine của Laravel, cho phép viết giao diện dễ dàng và động.

Ví dụ:

<h1>Hello, {{ $name }}</h1>

@foreach ($users as $user)

    <p>{{ $user->name }}</p>

@endforeach

### R2 Object Storage

**Cloudflare R2** là một dịch vụ lưu trữ đối tượng (object storage) do Cloudflare cung cấp, tương tự như Amazon S3, Google Cloud Storage hay Azure Blob Storage. Tuy nhiên, R2 nổi bật với khả năng giảm đáng kể chi phí lưu trữ dữ liệu bằng cách không tính phí băng thông egress (data egress fees) – một điểm mà các dịch vụ khác thường thu phí cao.

Các đặc điểm mạnh của R2:

**Lưu trữ đối tượng**

* R2 được thiết kế để lưu trữ dữ liệu dưới dạng các đối tượng (object), phù hợp cho các loại tệp lớn như hình ảnh, video, tài liệu, hoặc dữ liệu máy học.
* Dữ liệu được truy cập thông qua các URL hoặc API.

**Không tính phí egress**

* Điểm khác biệt lớn nhất của R2 là không thu phí khi truyền dữ liệu ra ngoài (egress). Điều này giúp giảm chi phí đáng kể cho các ứng dụng có lưu lượng truy cập lớn, đặc biệt là các dịch vụ cần phân phối nội dung.

**Tương thích S3 API**

* R2 tương thích với S3 API của Amazon, nghĩa là có thể sử dụng các thư viện hoặc công cụ được viết cho S3 mà không cần thay đổi nhiều.

**Hiệu suất cao**

* R2 tích hợp sâu với Cloudflare Network, giúp truyền tải dữ liệu nhanh hơn và ổn định hơn nhờ mạng lưới rộng khắp toàn cầu của Cloudflare.

**Hỗ trợ tính năng phiên bản (Versioning)**

* Lưu trữ nhiều phiên bản của cùng một tệp, cho phép khôi phục hoặc quản lý lịch sử thay đổi dễ dàng.

**Khả năng tích hợp với các dịch vụ Cloudflare khác**

* Kết hợp R2 với các sản phẩm khác như Cloudflare Workers, Pages, KV Store, hoặc Durable Objects để xây dựng ứng dụng hiện đại và tối ưu hóa hiệu suất.

**Chính sách giá cạnh tranh**

* **Chi phí lưu trữ:** Tính phí dựa trên dung lượng lưu trữ thực tế (tính theo GB).
* **Không gian lưu trữ miễn phí:** Lên tới 10GB và 1 triệu lượt thêm mới dữ liệu và 10 triệu lượt lấy dữ liệu.
* **Không phí băng thông:** Loại bỏ chi phí truyền dữ liệu ra ngoài, vốn là gánh nặng tài chính lớn ở các nền tảng khác.

### Thông báo realtime với Pusher

Pusher là một dịch vụ cung cấp nền tảng giao tiếp thời gian thực (real-time communication) dành cho các ứng dụng web và di động. Nó giúp các nhà phát triển tích hợp các tính năng realtime như thông báo, chat, theo dõi hoạt động trực tiếp, hoặc cập nhật nội dung ngay lập tức mà không cần phải viết và quản lý hạ tầng phức tạp.

Pusher nổi bật với dịch vụ Pusher Channels, được sử dụng phổ biến nhất, và cũng cung cấp các sản phẩm khác như Pusher Beams (thông báo đẩy) và Pusher Relay (GraphQL Subscriptions).

Pusher Channels cho phép giao tiếp hai chiều giữa máy chủ và trình duyệt hoặc ứng dụng di động bằng cách sử dụng công nghệ WebSockets hoặc fallback HTTP polling. Dưới đây là tính năng nổi bật:

**Real-Time Events (Sự kiện thời gian thực)**

* Gửi thông báo hoặc dữ liệu từ server đến client ngay khi có sự kiện xảy ra, không cần phải refresh hoặc gửi yêu cầu lặp lại từ phía client.
* Ví dụ: Hiển thị thông báo "Người dùng X vừa gửi tin nhắn" trong ứng dụng chat.

### Stable Diffusion mô hình tạo sinh

Stable Diffusion là một mô hình [trí tuệ nhân tạo tạo sinh (AI tạo sinh)](https://aws.amazon.com/ai/generative-ai/) có khả năng tạo ra hình ảnh tả thực độc đáo từ lời nhắc văn bản và hình ảnh. Mô hình này ra mắt lần đầu vào năm 2022. Ngoài hình ảnh, cũng có thể sử dụng mô hình này để tạo ra video và hoạt ảnh. Mô hình này dựa trên công nghệ khuếch tán và sử dụng không gian ngầm. Điều này làm giảm đáng kể yêu cầu xử lý và có thể chạy mô hình trên máy tính để bàn hoặc máy tính xách tay được trang bị GPU. Stable Diffusion có thể được tinh chỉnh để đáp ứng nhu cầu riêng với chỉ năm hình ảnh thông qua quá trình học chuyển giao.



Hình 1: Hình ảnh do Stable Diffusion tạo ra

Là một mô hình khuếch tán, Stable Diffusion khác với nhiều mô hình tạo hình ảnh khác. Về nguyên tắc, các mô hình khuếch tán sử dụng nhiễu Gaussian để mã hóa một hình ảnh. Sau đó, chúng sử dụng công cụ dự đoán nhiễu cùng với quá trình khuếch tán đảo ngược để tái tạo hình ảnh.

Ngoài các điểm khác biệt về kỹ thuật của mô hình khuếch tán, Stable Diffusion còn độc đáo ở chỗ nó không sử dụng không gian pixel của hình ảnh. Thay vào đó, nó sử dụng một không gian ngầm giảm độ nét.

Lý do là hình ảnh màu với độ phân giải 512x512 có thể có 786.432 giá trị. Để so sánh, Stable Diffusion sử dụng hình ảnh nén nhỏ hơn 48 lần với 16.384 giá trị. Điều này làm giảm đáng kể yêu cầu xử lý. Và đó là lý do tại sao có thể sử dụng Stable Diffusion trên máy tính để bàn có GPU NVIDIA với 8 GB RAM. Không gian ngầm nhỏ hơn làm việc hiệu quả vì hình ảnh tự nhiên không phải là ngẫu nhiên. Stable Diffusion sử dụng các tệp mã hóa tự động biến đổi (VAE) trong bộ giải mã để vẽ các chi tiết sắc nét như mắt.

Stable Diffusion V1 được đào tạo bằng cách sử dụng ba tập dữ liệu do LAION thu thập thông qua Common Crawl, bao gồm tập dữ liệu LAION-Aesthetics v2.6 với các hình ảnh có xếp hạng thẩm mỹ từ 6 trở lên.



Hình 2: Cách thức hoạt động cơ bản của Stable Diffusion

A screenshot of a computer

Description automatically generated

Hình 3: Giao diện người dùng A1111

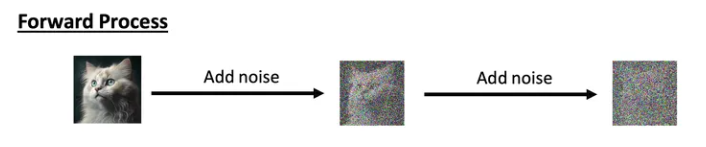
### Stable Diffusion gồm những kiến trúc nào?

* **Bộ mã hóa tự động biến đổi**

Bộ mã hóa tự động biến đổi bao gồm một bộ mã hóa và bộ giải mã riêng biệt. Bộ mã hóa nén hình ảnh 512x512 pixel thành mô hình 64x64 nhỏ hơn trong không gian ngầm dễ thao tác hơn. Bộ giải mã khôi phục mô hình từ không gian ngầm thành hình ảnh 512x512 pixel kích thước thực.

* **Khuếch tán thuận**

Khuếch tán thuận thêm dần nhiễu Gaussian vào hình ảnh cho đến khi chỉ còn lại nhiễu ngẫu nhiên. Không thể xác định hình ảnh ban đầu là gì từ hình ảnh nhiễu cuối cùng. Trong quá trình đào tạo, tất cả các hình ảnh đều đi qua quá trình này. Khuếch tán thuận không được sử dụng thêm trừ khi thực hiện chuyển đổi hình ảnh thành hình ảnh.



Hình 4: Khuếch tán thuận

* **Khuếch tán đảo ngược**

Quá trình này về cơ bản là một quá trình được tham số hóa hoàn tác lại nhiều lần quá trình khuếch tán thuận. Ví dụ: có thể huấn luyện mô hình chỉ với hai hình ảnh, như một con mèo và một con chó. Nếu làm vậy, quá trình đảo ngược sẽ lệch về phía con mèo hoặc con chó và không có hình ảnh nào ở giữa. Trong thực tế, đào tạo mô hình bao gồm hàng tỷ hình ảnh và sử dụng lời nhắc để tạo ra hình ảnh độc đáo.

A diagram of a cat's head

Description automatically generated

Hình 5: Khuếch tán đảo

* **Công cụ dự đoán nhiễu (U-Net)**

Công cụ dự đoán nhiễu là chìa khóa để khử nhiễu hình ảnh. Stable Diffusion sử dụng mô hình U-Net để thực hiện điều này. Mô hình U-Net là mạng nơ-ron tích chập ban đầu được phát triển để phân đoạn hình ảnh trong y sinh. Cụ thể, Stable Diffusion sử dụng mô hình Mạng nơ-ron dư (ResNet) được phát triển cho thị giác máy tính.

Công cụ dự đoán nhiễu ước tính lượng nhiễu trong không gian ngầm và loại bỏ lượng nhiễu này khỏi hình ảnh. Công cụ này lặp lại quá trình này theo một số lần định sẵn, giảm nhiễu theo các bước do người dùng chỉ định. Công cụ dự đoán nhiễu tiếp xúc với các lời nhắc biến đổi giúp xác định hình ảnh cuối cùng.

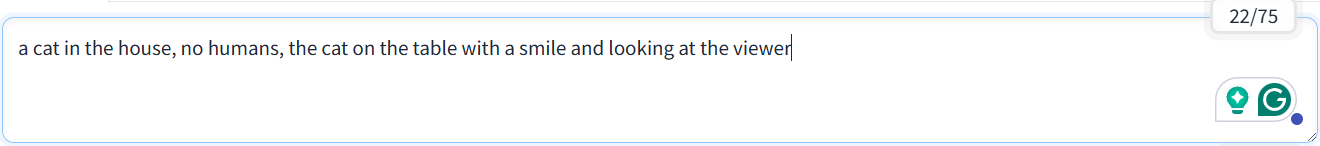
* **Biến đổi văn bản**

Hình thức biến đổi phổ biến nhất là lời nhắc văn bản (Prompt). Tác nhân token hóa CLIP phân tích từng từ trong lời nhắc văn bản và nhúng dữ liệu này vào vectơ 768 giá trị. Có thể sử dụng tối đa 75 token trong một lời nhắc. Stable Diffusion đưa các lời nhắc này từ bộ mã hóa văn bản đến công cụ dự đoán nhiễu U-Net bằng cách sử dụng bộ biến đổi văn bản. Bằng cách đặt hạt giống vào trình tạo số ngẫu nhiên, bạn có thể tạo ra các hình ảnh khác nhau trong không gian ngầm.

A diagram of a process

Description automatically generated with medium confidence

Hình 6: Cách Stable Diffusion biến đổi văn bản

****

Hình 7: Một đoạn Prompt trong Stable Diffusion

### Mô hình nổi tiếng của Stable Diffusion

* **Stable-Diffusion-v1-5**

Stable-Diffusion-v1-5 được khởi tạo với trọng số của điểm kiểm tra Stable-Diffusion-v1-2 và sau đó được tinh chỉnh trên 595 nghìn bước ở độ phân giải 512x512 trên "laion-aesthetics v2 5+" và giảm 10% điều hòa văn bản để cải thiện việc lấy mẫu hướng dẫn không cần phân loại.

Những hạn chế của mô hình này:

* + Mô hình không đạt được độ chân thực hoàn hảo.
  + Mô hình không thể hiển thị văn bản dễ đọc.
  + Mô hình này không thực hiện tốt các tác vụ khó hơn liên quan đến bố cục, chẳng hạn như hiển thị hình ảnh tương ứng với “Khối màu đỏ nằm trên hình cầu màu xanh”.
  + Khuôn mặt và con người nói chung có thể không được tạo đúng cách.
  + Mô hình được đào tạo chủ yếu với phụ đề tiếng Anh và sẽ không hoạt động tốt bằng các ngôn ngữ khác.
  + Phần tự động mã hóa của mô hình bị mất.
  + Mô hình này đã được đào tạo trên tập dữ liệu quy mô lớn LAION-5B chứa nội dung dành cho người lớn và không phù hợp để sử dụng sản phẩm nếu không có các cơ chế và cân nhắc an toàn bổ sung.



Hình 8: Hình ảnh do SD 1.5 tạo ra

* **Mô hình majicMIX realistic**

Là một mô hình do người dùng tạo ra. Mô hình này tập trung vào việc tạo ra các hình ảnh về phụ nữ Châu Á. Dựa trên mô hình mặc định SD 1.5.

Các hạn chế của mô hình này:

* + Dung lượng lớn khoảng 2GB.
  + Hình ảnh tạo ra ở độ phân giải 512x512 có thể bị hỏng khuôn mặt.
  + Hình ảnh tạo ra theo văn bản vẫn còn sai sót.
  + Có thể tạo ra hình ảnh không an toàn NSFW.
  + Phải dùng nhiều các Extension để hoàn thiện hình ảnh.
  + Không sửa được lỗi bàn tay.

A screenshot of a person

Description automatically generated

Hình 9: Hình ảnh do Model người dùng tạo ra

* **Stable Diffusion XL 1.0**

Stable Diffusion XL (SDXL) là mô hình hình ảnh AI mới nhất có thể tạo ra những con người chân thực, văn bản dễ đọc và nhiều phong cách nghệ thuật khác nhau với bố cục hình ảnh tuyệt vời. Đây là phiên bản lớn hơn và tốt hơn của mô hình Stable Diffusion v1.5 nổi tiếng , và do đó có tên là SDXL. Cải thiện mô hình khuếch tán tiềm ẩn để tổng hợp hình ảnh có độ phân giải cao ” của Podell và các cộng sự, Stable Diffusion XL tốt hơn so với mô hình v1.5 về mọi mặt.

Những cải tiến so với SD1.5

* + Hình ảnh chất lượng cao hơn
  + Thực hiện theo lời nhắc chặt chẽ hơn
  + Chi tiết tốt hơn
  + Kích thước hình ảnh lớn hơn
  + Khả năng tạo ra văn bản dễ đọc
  + Khả năng tạo ra hình ảnh tối hơn

A collage of a person

Description automatically generated

Hình 10: So sánh khả năng tạo hình ảnh của SD1.5 và SDXL

* **FLUX.1 Model**

Flux là mô hình tạo hình ảnh AI mới do Black Forest Labs phát triển, một nhóm tách ra từ Stability AI. Nó đại diện cho một bước tiến đáng kể trong nghệ thuật do AI tạo ra, sử dụng "kiến trúc lai" kết hợp các kỹ thuật biến áp và khuếch tán, được mở rộng lên tới 12 tỷ tham số.

Mô hình này cung cấp khả năng tạo hình ảnh hiệu suất tiên tiến với tính năng theo dõi nhắc nhở hàng đầu, chất lượng hình ảnh, chi tiết hình ảnh và tính đa dạng đầu ra. Nhưng đi kèm với hiệu xuất đáng kể trên thì Flux thường yêu cầu phần cứng rất cao, các GPU được ra mắt muộn sau 2023 dường như không thể chạy Flux một cách nhanh chóng.

Những cải tiến so với SDXL:

* + Chất lượng hình ảnh và khả năng bám dính nhanh chóng: Flux được biết đến với chất lượng hình ảnh vượt trội và khả năng bám dính nhanh chóng so với Stable Diffusion 3. Nó được coi là ngang bằng với Midjourney V6, trở thành đối thủ cạnh tranh mạnh mẽ trong lĩnh vực tạo hình ảnh AI.
  + Tốc độ và hiệu quả: Mô hình này cung cấp khả năng tạo hình ảnh nhanh chóng, với các phiên bản khác nhau được thiết kế riêng cho các nhu cầu khác nhau. Ví dụ, Flux Schnell tạo hình ảnh nhanh hơn khoảng 10 lần so với mô hình Pro, mặc dù chất lượng giảm.
  + Kích thước tham số: Với 12 tỷ tham số, Flux lớn hơn đáng kể so với các mô hình trước đây như Stable Diffusion XL, có khoảng 3,5 tỷ tham số. Sự gia tăng tham số này góp phần nâng cao hiệu suất và khả năng của nó nhưng đồng thời dung lượng cũng rất lớn

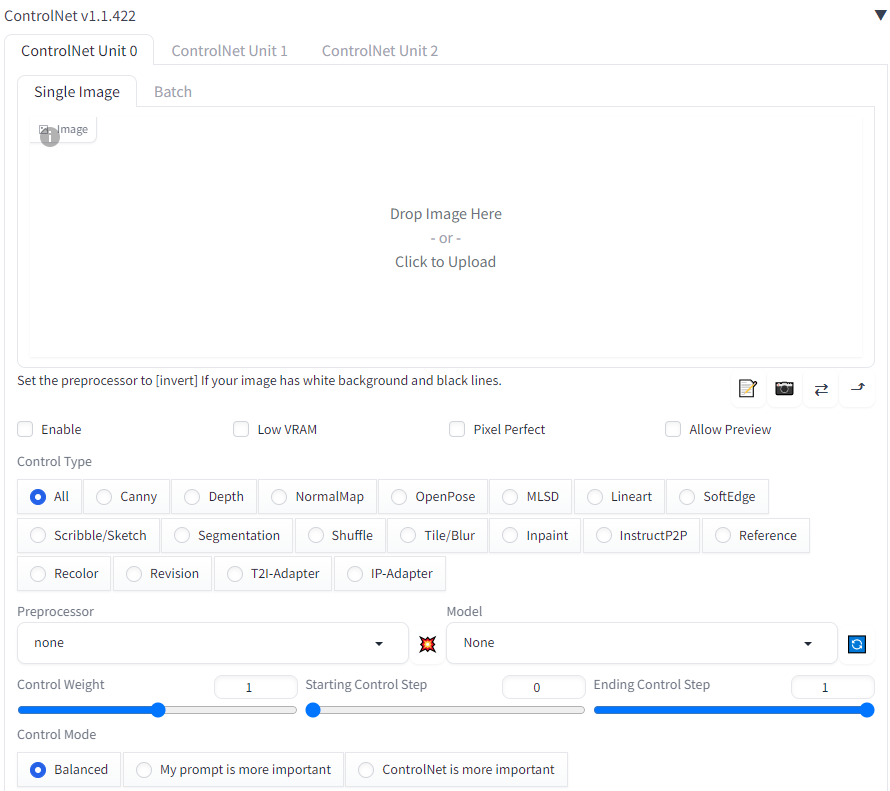


Hình 11: Hình ảnh do Flux Model tạo ra

### Các Extension nổi tiếng của Stable Diffusion

* **ControlNet**

Là công nghệ tiên tiến kết hợp sức mạnh của việc tạo hình ảnh AI với tính ổn định của ControlNet. Bằng cách tận dụng các kỹ thuật khuếch tán ổn định, mô hình này có thể tạo ra hình ảnh chân thực, chất lượng cao dựa trên thông tin đầu vào của người dùng. Cho dù đó là tạo ra nghệ thuật, nâng cao hình ảnh hay tạo ra các tư thế khác nhau, ControlNet Stable Diffusion sẽ mở ra một thế giới đầy khả năng. Hơn nữa, ImageFlash tích hợp liền mạch với mô hình Khuếch tán ổn định ControlNet, tận dụng tính ổn định và khả năng kiểm soát của nó để nâng cao quá trình tạo hình ảnh. Sự kết hợp của hai mô hình mạnh mẽ này đảm bảo rằng hình ảnh được tạo ra không chỉ hấp dẫn về mặt hình ảnh mà còn duy trì tính nhất quán và mạch lạc.



Hình 12: Giao diện của ControlNet

ControlNet tạo ra hình ảnh chính xác và chân thực mô tả tư thế của con người với độ chính xác vượt trội. Nó có khả năng hiển thị ngay cả những tư thế phức tạp nhất mà chân tay có thể bị che khuất hoặc bị cong, do đó đảm bảo mô tả trung thực tư thế đầu vào.

Công cụ NormalMap trong ControlNet sẽ làm cho hình ảnh thông thường tạo điều kiện cho người dùng tập trung vào tính nhất quán của đối tượng thay vì môi trường xung quanh và độ sâu của đối tượng, từ đó cho phép sửa đổi chính xác hơn cả nền và đối tượng. Tính năng này tăng cường khả năng kiểm soát của người dùng đối với kết quả chỉnh sửa đồng thời giảm thiểu các hiện vật không mong muốn trong hình ảnh.

A screenshot of a computer

Description automatically generated

Hình 13: Normal Map trong Controlnet

Quá trình đổi màu và cách điệu hóa hình ảnh trong ControlNet bao gồm một sự biến đổi đáng chú ý có thể nâng chúng lên ngang tầm với những bức tượng hoặc bức tranh hiện thực. Kỹ thuật phức tạp này đòi hỏi thao tác khéo léo cũng như con mắt tinh tế và thẩm mỹ để đạt được kết quả thực sự ấn tượng.

ControlNet sao chép chính xác phong thái và lập trường của một hình ảnh cụ thể lên một hình ảnh khác. Sao chép phong thái và lập trường của một hình ảnh cụ thể sang một hình ảnh khác có thể là một nhiệm vụ khó khăn

Khả năng đặc biệt của ControlNet trong việc giữ lại các sắc thái nghệ thuật đường nét phức tạp khiến nó trở thành một công cụ rất tốt để đưa tác phẩm nghệ thuật đơn sắc vào màu sắc. Mặc dù thực tế là tính năng này vẫn chưa được ra mắt nhưng việc phát hành sắp tới của nó hứa hẹn sẽ giải quyết được cả những vấn đề khó khăn về mặt kỹ thuật.

A screenshot of a person wearing a white dress

Description automatically generated

Hình 14: OpenPose trong ControlNet

### Ứng dụng mô hình AI tạo sinh với Comfyui

ComfyUI là một giao diện người dùng (UI) mã nguồn mở được thiết kế để quản lý, xây dựng và tương tác với các mô hình AI tạo sinh (generative AI). Giao diện này giúp người dùng dễ dàng tạo, chỉnh sửa, và kiểm soát các pipeline AI mà không cần nhiều kinh nghiệm về lập trình. ComfyUI được sử dụng phổ biến trong các lĩnh vực như tạo hình ảnh, xử lý ngôn ngữ tự nhiên (NLP), hoặc âm thanh, đặc biệt khi làm việc với các mô hình AI hiện đại như Stable Diffusion.

Tính năng nổi bật của ComfyUI:

**Giao diện dạng đồ họa (Node-based Workflow)**

* Node-based UI giúp người dùng xây dựng pipeline bằng cách kéo thả các thành phần hay còn được gọi là Workflow. Mỗi node đại diện cho một tác vụ như nạp dữ liệu, chạy mô hình, hoặc áp dụng xử lý hậu kỳ.
* Tính trực quan cao, giúp dễ dàng theo dõi và chỉnh sửa luồng xử lý.

A screenshot of a computer program

Description automatically generated

Hình 15: Một giao diện Workflow cơ bản trong Comfyui (Text2Image)

**Hỗ trợ đa dạng mô hình AI**

* Tích hợp và làm việc tốt với các mô hình tạo sinh phổ biến như Stable Diffusion, ControlNet, và các mô hình khác hỗ trợ tạo hình ảnh hoặc video.

**Kiểm soát chi tiết**

* Người dùng có thể điều chỉnh các tham số của mô hình như số bước lặp (steps), độ nhiễu (noise), tỷ lệ hướng dẫn (guidance scale) và các tùy chỉnh khác để tinh chỉnh kết quả đầu ra.

**Tích hợp xử lý nhiều giai đoạn**

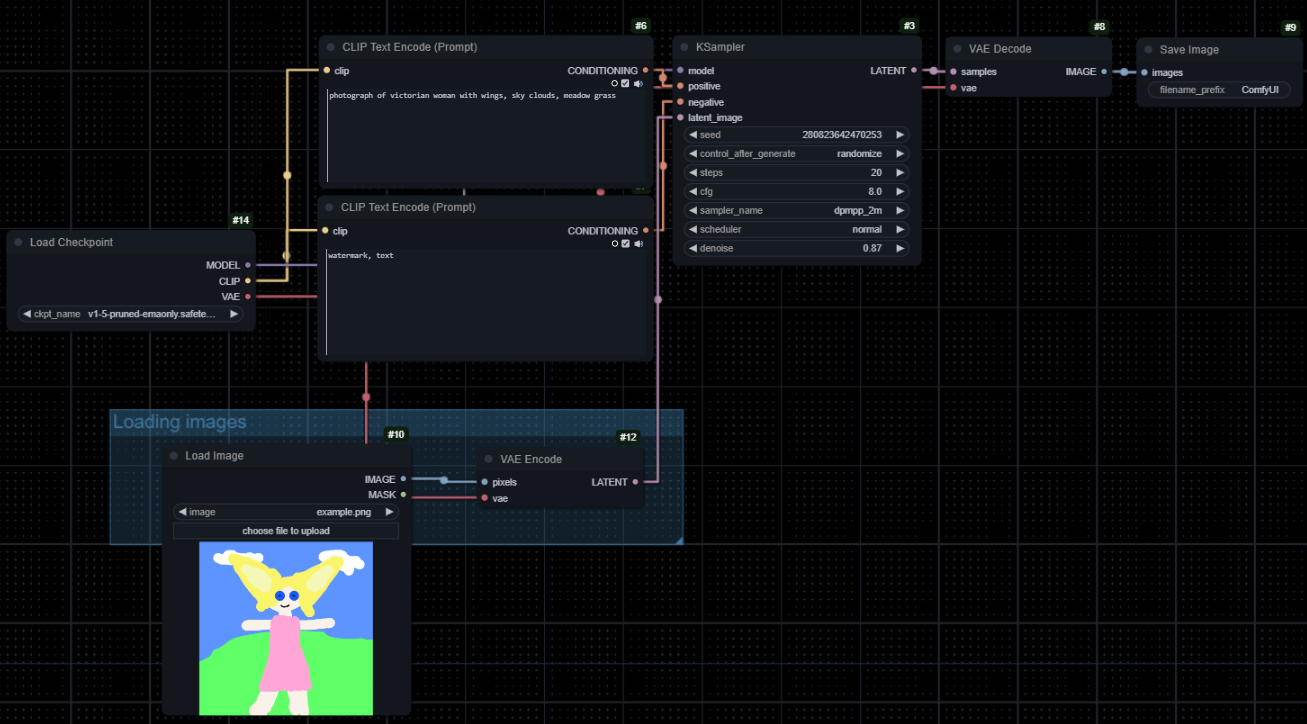
* Hỗ trợ các pipeline phức tạp như xử lý nhiều bước, tạo hình ảnh từ văn bản, điều chỉnh màu sắc, hoặc áp dụng hiệu ứng đặc biệt.

**Mở rộng và tùy chỉnh**

* Dễ dàng mở rộng với các node tùy chỉnh, plugin, hoặc tích hợp các công cụ AI khác.

**Mã nguồn mở và miễn phí**

* Là dự án mã nguồn mở, ComfyUI được xây dựng bởi cộng đồng, dễ dàng sửa đổi và tích hợp vào các dự án hiện có.



Hình 16: Một Workflow đơn giản trong Comfyui (Image2Image)

Ứng dụng của Comfyui

**Tạo hình ảnh từ mô tả văn bản (Text-to-Image)**

* Sử dụng các mô hình như Stable Diffusion để tạo hình ảnh chất lượng cao từ văn bản.

**Chỉnh sửa hình ảnh (Image-to-Image)**

* Điều chỉnh hoặc thêm hiệu ứng lên hình ảnh hiện có.

**Xử lý ảnh nghệ thuật**

* Tạo ra các bức tranh kỹ thuật số, phong cách nghệ thuật độc đáo.

**Tích hợp với các pipeline phức tạp**

* Kết hợp các mô hình AI khác nhau để tạo ra quy trình làm việc nâng cao, ví dụ: sử dụng ControlNet để kiểm soát bố cục hình ảnh.

**Phát triển dự án AI tùy chỉnh**

* Phù hợp cho các nhà nghiên cứu hoặc lập trình viên muốn xây dựng hệ thống AI dựa trên mô hình tạo sinh.

A computer screen shot of a computer

Description automatically generated

Hình 17: Một Workflow phức tạp trong Comfyui trong việc sao chép hình ảnh

# Phân tích yêu cầu và thiết kế hệ thống

## Phân tích yêu cầu người dùng

### Đối với người dùng cuối

Yêu cầu chức năng

* **Tìm kiếm:**

Tìm kiếm hình ảnh qua từ khóa, mô tả.

Hỗ trợ tìm kiếm ngược (bằng hình ảnh).

* **Tính năng AI tạo sinh:**

Tạo hình ảnh từ mô tả văn bản.

Tạo người mẫu ảo từ ảnh chụp người dùng.

Tạo ảnh sáng tạo (vui nhộn, phong cách nghệ thuật).

* **Quản lý nội dung:**

Lưu trữ, chia sẻ hình ảnh.

Xây dựng bảng (board) riêng để quản lý ý tưởng.

# ???

# Kết Luận

# Tài Liệu Tham Khảo