MỞ ĐẦU

1. Đặt vấn đề

Thị trường Game Di động Toàn cầu đã và đang phát triển mạnh mẽ, trong đó, thể loại Casual Puzzle Game (Game xếp hình, giải đố thông thường, như Match-3, Block-out,...) luôn giữ vững vị thế là một trong những phân khúc mang lại doanh thu cao nhất. Đặc trưng của thể loại này là lối chơi đơn giản, dễ tiếp cận nhưng đòi hỏi một lượng nội dung Khổng lồ để duy trì sự gắn kết và trải nghiệm mới mẻ cho người chơi. Đối với các game hàng đầu, số lượng màn chơi có thể lên đến hàng nghìn, thậm chí hàng chục nghìn.

Nhu cầu về nội dung là thách thức lớn nhất đối với đội ngũ phát triển game. Quy trình thiết kế màn chơi truyền thống, mặc dù đảm bảo tính sáng tạo và cân bằng, nhưng bộc lộ nhiều hạn chế nghiêm trọng khi cần mở rộng quy mô:

- Tốn kém về thời gian và nguồn lực: Mỗi màn chơi đòi hỏi nhà thiết kế sắp xếp thủ công các vật thể, chướng ngại vật theo luật chơi và đảm bảo độ khó tăng dần hợp lý. Quá trình này rất mất thời gian và phụ thuộc hoàn toàn vào kinh nghiệm cá nhân.
- Thiếu tính nhất quán và lặp lại: Khi số lượng màn chơi tăng lên, nhà thiết kế dễ bị cạn kiệt ý tưởng, dẫn đến sự lặp lại về bố cục và cấu trúc, làm giảm trải nghiệm của người chơi.
- Khó khăn trong việc tạo biến thể: Việc điều chỉnh các tham số nhỏ để tạo ra hàng trăm biến thể màn chơi mới theo một chủ đề cụ thể trở nên phi thực tế nếu thực hiện thủ công.

Trước những hạn chế trên, việc ứng dụng Trí tuệ nhân tạo (AI) vào Quy trình Tạo sinh Nội dung Thủ tục (PCG) đã trở thành giải pháp tất yếu. Tuy nhiên, các phương pháp PCG truyền thống chủ yếu dựa trên các thuật toán quy tắc hoặc mô hình học tăng cường, thường chỉ tập trung vào việc tạo ra logic và cấu trúc màn chơi dưới dạng dữ liệu thô (ví dụ: mảng 2D mô tả vị trí các vật phẩm), mà chưa giải quyết được khâu thị giác và thẩm mỹ của màn chơi.

Trong những năm gần đây, sự trỗi dậy của các Mô hình AI Tạo sinh (Generative AI) đã mở ra một hướng đi đột phá. Đặc biệt, Mô hình Khuếch tán (Diffusion Models) như Stable Diffusion, Midjourney đã chứng minh khả năng vượt trội trong việc tạo ra hình ảnh chất lượng cao, có độ chi tiết và tính thẩm mỹ ấn tượng dựa trên mô tả ngôn ngữ.

Tuy nhiên, trong bối cảnh tạo màn chơi game, Mô hình Khuếch tán đơn thuần không thể hiểu được luật chơi hay mục tiêu thiết kế cần thiết (ví dụ: màn chơi A cần một lối đi dài

với 5 khối chướng ngại vật loại B). Đây là lúc cần đến sự can thiệp của Mô hình Ngôn ngữ lớn (LLM).

- Vai trò của LLM: LLM có khả năng hiểu, lập luận và tạo ra logic thiết kế màn chơi dưới dạng văn bản đóng vai trò là bộ não kiến tạo cấu trúc.
- Vai trò của Diffusion Models: Diffusion Models sẽ đóng vai trò là nghệ sĩ, nhận đầu vào là logic thiết kế do LLM tạo ra để tạo ra hình ảnh bố cục màn chơi

Sự kết hợp giữa LLM (tạo Logic) và Diffusion Models (tạo Visual) là một giải pháp tiềm năng, giải quyết đồng thời hai vấn đề cốt lõi của PCG: Tính đúng đắn về mặt luật chơi và Tính thẩm mỹ về mặt thị giác.

Nắm bắt được tiềm năng này, Đồ án "NGHIÊN CỨU ỨNG DỤNG MÔ HÌNH TẠO SINH ẢNH HỖ TRỢ TẠO MÀN CHƠI CASUAL PUZZLE GAME" được đề xuất.

Mục tiêu của Đồ án này là nghiên cứu về cơ chế hoạt động, khả năng tùy biến của các mô hình Diffusion, đặc biệt thông qua kỹ thuật DreamBooth, để huấn luyện mô hình sinh ra các bố cục màn chơi có tính thẩm mỹ và tuân thủ phong cách nghệ thuật của một game casual. Đồ án sẽ tập trung xây dựng một quy trình làm việc tích hợp, nơi người thiết kế có thể sử dụng câu lệnh tự nhiên để mô tả màn chơi, và hệ thống sẽ phản hồi bằng hình ảnh bố cục màn chơi tức thì, rút ngắn đáng kể thời gian từ ý tưởng đến sản phẩm mẫu.

2. Đối tượng nghiên cứu

Đối tượng nghiên cứu của đồ án là kỹ thuật Mô hình Khuếch tán (Diffusion Models) và Mô hình Ngôn ngữ lớn (LLM) để hỗ trợ quá trình tạo màn chơi trong thể loại game casual.

3. Phạm vi nghiên cứu

Phạm vi nghiên cứu của đồ án bao gồm:

1. Nghiên cứu cơ sở lý thuyết liên quan đến Mô hình Ngôn ngữ lớn và Mô hình Khuếch tán.
2. Khảo sát nhu cầu và thực trạng thiết kế màn chơi trong thể loại game casual tại Việt Nam.
3. Thử nghiệm xây dựng công cụ sinh hình ảnh màn chơi bằng cách tích hợp LLM và Mô hình Khuếch tán

4. Phương pháp nghiên cứu

- **Phương pháp Thu thập và Kế thừa Tài liệu:** Nghiên cứu các công trình, bài báo khoa học trong và ngoài nước liên quan đến ứng dụng AI trong phát triển game và các mô hình tạo sinh hình ảnh.
- **Phương pháp Phân tích, Tổng hợp:** So sánh, đánh giá các mô hình (LLM, Diffusion) và phương pháp hiện có để lựa chọn hướng tiếp cận phù hợp nhất cho việc tạo sinh bố cục màn chơi.
- **Phương pháp Thực nghiệm:** Xây dựng, huấn luyện và thử nghiệm công cụ sinh màn chơi. Đánh giá kết quả dựa trên các tiêu chí về tính sáng tạo, tính khả thi của bố cục, và mức độ ứng dụng thực tế trong quy trình thiết kế.

5. Các nội dung nghiên cứu

Báo cáo đồ án gồm các nội dung chính:

- **Chương 1:** Nghiên cứu thực trạng và nhu cầu thiết kế màn chơi trong thể loại game casual**.** Tổng quan về thị trường Casual Puzzle, tầm quan trọng của thiết kế màn chơi, và phân tích thực trạng các studio game tại Việt Nam. Giới thiệu kiến trúc LLM và Diffusion Models như là giải pháp tiềm năng.
- **Chương 2:** Nghiên cứu, phát triển và ứng dụng công nghệ để sinh hình ảnh màn chơi**.** Phân tích yêu cầu kỹ thuật và thiết kế kiến trúc hệ thống tích hợp LLM (cho logic) và Diffusion (cho hình ảnh). So sánh, lựa chọn mô hình nền tảng và mô tả luồng hoạt động cơ bản của hệ thống.
- **Chương 3:** Triển khai và thử nghiệm công cụ**.** Giới thiệu game thử nghiệm. Mô tả quá trình triển khai, bao gồm chuẩn bị tập dữ liệu huấn luyện, các bước huấn luyện mô hình bằng kỹ thuật DreamBooth. Cuối cùng, trình bày việc thử nghiệm ứng dụng và đánh giá kết quả về chất lượng hình ảnh và tốc độ sinh ảnh của hệ thống.

6. Kết quả dự kiến

- Báo cáo kết quả nghiên cứu.
- Công cụ hoặc mô hình ứng dụng LLM với Mô hình Khuếch tán để sinh hình ảnh bố cục màn chơi game casual.

CHƯƠNG 1: CƠ SỞ LÝ THUYẾT VÀ TỔNG QUAN NGHIÊN CỨU

1.1. Tổng quan về thể loại Game Casual

Thể loại Game Casual nổi bật là một phân khúc có tốc độ phát triển ấn tượng trong lĩnh vực game di động. Đặc trưng của game Casual là lối chơi đơn giản, cơ chế điều khiển trực quan, mục tiêu rõ ràng và khả năng tiếp cận nhanh chóng cho người chơi thuộc mọi lứa tuổi và trình độ. Tính chất chơi nhanh và dừng nhanh (phù hợp với những phút nghỉ ngắn) khiến Casual Game chiếm vị thế then chốt, đặc biệt trên nền tảng điện thoại thông minh.

Về mặt tâm lý, Game Casual mang lại cảm giác giải trí tức thời, giảm căng thẳng và tạo cảm giác thành tựu nhờ vào việc thiết kế mục tiêu nhỏ gọn, rõ ràng. Sự thành công của thể loại này dựa trên mô hình free-to-play kết hợp quảng cáo hoặc giao dịch nhỏ.

Thách thức cốt lõi: Người chơi thường vượt qua các cấp độ trong thời gian ngắn và luôn mong chờ sự đổi mới liên tục. Điều này buộc các nhà phát triển phải tạo ra một số lượng màn chơi khổng lồ (hàng trăm đến hàng nghìn) trong thời gian ngắn, đồng thời phải đảm bảo tính logic, mức độ khó phù hợp, và tính khác biệt giữa các màn chơi. Điều này tạo ra áp lực rất lớn đối với đội ngũ thiết kế. Trong bối cảnh này, việc ứng dụng các giải pháp hỗ trợ bằng Trí tuệ Nhân tạo là một yêu cầu cấp thiết.

1.2. Quy trình thiết kế màn chơi trong Game Casual hiện nay

Thiết kế màn chơi là thành tố quan trọng nhằm tối ưu hóa các chỉ số vận hành (tỉ lệ giữ chân, tỉ lệ hoàn thành màn chơi). Trong thực tế sản xuất game casual, quy trình thiết kế vận hành theo một vòng lặp liên tục với các bước chính:

1. **Phân tích Mục tiêu và Xác định Cơ chế Chủ đạo:** Xác định mục tiêu chơi chính (giải đố, ghép cặp, vượt chướng ngại vật), hành vi người chơi mong muốn và phong cách trình bày. Thiết lập nền tảng để người chơi làm quen dễ dàng.
2. **Xây dựng Khung Màn chơi và Cân bằng Độ khó:** Triển khai ý tưởng thành bố cục thực tế, thiết lập không gian, chướng ngại vật, phần thưởng và giới hạn. Cân bằng độ khó là yếu tố trọng tâm: màn chơi phải tạo ra đường cong độ khó hợp lý, tăng dần qua từng màn nhưng vẫn đa dạng. Màn dễ quá gây nhàm chán, khó quá dẫn đến tỉ lệ bỏ cuộc cao.
3. **Tạo mẫu, Thử nghiệm và Thu thập Dữ liệu Hành vi:** Tạo phiên bản mẫu và thử nghiệm với người dùng thật. Ghi nhận thời gian hoàn thành, các điểm gây khó khăn và hành vi bất thường. Phương pháp thường áp dụng là phỏng vấn trực tiếp, phân tích dữ liệu hành vi trong game và đo lường sinh trắc học.

4. **Điều chỉnh, Tối ưu và Phân tích Chỉ số:** Dựa trên dữ liệu thử nghiệm, điều chỉnh các thông số (bố cục, số lượt đi, phần thưởng,...) để tối ưu hóa tỷ lệ hoàn thành, tăng tính hấp dẫn và cải thiện cảm giác tiến bộ. Màn chơi cần đáp ứng nhu cầu vận hành dài hạn thông qua việc duy trì động lực.
5. **Sản xuất Hàng loạt và Vận hành Liên tục:** Liên tục cung cấp các màn chơi mới để duy trì sự mới mẻ. Giai đoạn này đòi hỏi quy mô lớn, tốc độ cao và khả năng xử lý dữ liệu linh hoạt, điều này khó khăn với quy trình thủ công.

Nhận xét: Quy trình này đòi hỏi sự lặp lại, tốn thời gian và nhân lực lớn, dẫn đến nhu cầu ứng dụng công nghệ AI để hỗ trợ thiết kế màn chơi.

1.3. Thực trạng thiết kế màn chơi trong các Studio Game Việt Nam

Mặc dù thị trường game Việt Nam đã có sự bùng nổ, đặc biệt ở mảng mobile và casual game, khâu thiết kế màn chơi vẫn là một điểm nghẽn lớn tại nhiều studio:

- **Phụ thuộc vào Quy trình Thiết kế Truyền thống:** Đại đa số studio vẫn thao tác thủ công trực tiếp trên các game engine như Unity. Phương pháp này không thể mở rộng quy mô sản xuất hàng trăm, hàng nghìn màn chơi, gây tiêu tốn nguồn lực khổng lồ và dẫn đến thiếu đồng bộ về chất lượng giữa các màn chơi do sự tham gia của nhiều nhân sự.
- **Thiếu Công cụ Nội bộ Hỗ trợ:** Các công cụ quan trọng như trình sinh màn tự động (PCG), hệ thống đánh giá độ khó, công cụ mô phỏng người chơi, hoặc các editor tùy chỉnh thường không được phát triển do thiếu nguồn lực kỹ thuật chuyên sâu. Điều này buộc level designer phải làm việc thủ công trong hầu hết các khâu.
- **Hạn chế về Nguồn Nhân lực và Đào tạo:** Thị trường khan hiếm đội ngũ Level Design được đào tạo bài bản. Nhiều nhân sự là người chuyển ngạch, thiếu nền tảng lý thuyết vững chắc để xây dựng hệ thống màn chơi phức tạp và duy trì đường cong độ khó theo tiêu chuẩn quốc tế. Việc kiêm nhiệm nhiều đầu việc phụ càng làm giảm chất lượng chuyên môn.
- **Văn hóa Thiết kế Dựa trên Dữ liệu chưa Phổ biến:** Nhiều studio chưa sở hữu hệ thống thu thập hoặc khai thác dữ liệu hành vi người dùng đủ mạnh. Việc tinh chỉnh màn chơi thường dựa trên cảm nhận cá nhân hoặc phản hồi nhỏ lẻ, thay vì dựa trên dữ liệu thực nghiệm, dẫn đến hiệu quả tối ưu hóa thấp và kém ổn định.
- **Khả năng Ứng dụng AI còn Hạn chế:** Dù xu hướng AI (LLM, Diffusion Models) đã phổ biến trên thế giới, nhưng tại Việt Nam, việc ứng dụng vẫn ở giai

đoạn sơ khai do thiếu hiểu biết, chi phí đầu tư lớn và hạn chế về tài nguyên tính toán.

Kết luận: Nhu cầu chuyển đổi sang quy trình thiết kế màn chơi có hỗ trợ AI là yêu cầu thiết yếu đối với các studio Việt Nam để tăng năng suất, giảm chi phí và duy trì vị thế cạnh tranh trên thị trường toàn cầu.

1.4. Nhu cầu và xu hướng ứng dụng AI trong thiết kế màn chơi

Trí tuệ nhân tạo (AI), đặc biệt là các phương pháp Học sâu (Deep Learning), đã mở rộng ứng dụng trong ngành công nghiệp game, không chỉ giới hạn ở việc điều khiển NPC mà còn tác động mạnh mẽ đến quy trình phát triển nội dung.

Xu hướng ứng dụng AI

- **Procedural Content Generation (PCG):** AI hỗ trợ tự động hoặc bán tự động tạo ra các yếu tố như bố cục màn chơi, bản đồ, và thử thách gameplay, giúp giảm đáng kể khối lượng công việc thủ công.
- **AI Tạo sinh (Generative AI) – Hướng Tiếp cận Mới:**
    - **Mô hình Ngôn ngữ lớn (LLM):** Khả năng xử lý mô tả bằng ngôn ngữ tự nhiên và học các quy tắc gameplay, cho phép LLM đề xuất cấu trúc, logic màn chơi, hoặc sinh ra các biến thể bố cục dưới dạng dữ liệu cấu hình.
    - **Mô hình Khuếch tán (Diffusion Models):** Hiệu quả cao trong việc tạo sinh hình ảnh, tài nguyên đồ họa và **bố cục không gian trực quan** cho màn chơi, đảm bảo sự đồng nhất về phong cách mỹ thuật.
- **Thiết kế Hỗ trợ bằng AI (AI-assisted Level Design):** Sự kết hợp giữa LLM và Diffusion Models cho phép chuyển đổi mô tả trừu tượng của nhà thiết kế thành các biểu diễn trực quan cụ thể, rút ngắn đáng kể thời gian sản xuất và khối lượng công việc cho đội ngũ thiết kế.
- **Phân tích và Cân bằng Độ khó:** AI được huấn luyện dựa trên dữ liệu hành vi người chơi để dự đoán khả năng vượt màn, thời gian hoàn thành, hoặc mô phỏng người chơi ảo nhằm tự động kiểm tra màn chơi trước khi phát hành, đảm bảo đường cong tiến trình hợp lý.

Nhu cầu tại Việt Nam

Nhu cầu ứng dụng AI tại Việt Nam ngày càng rõ rệt do áp lực cạnh tranh quốc tế và hạn chế về nhân lực. AI giúp:

- **Tiết kiệm thời gian và giảm chi phí:** Tự động hóa nhiều công đoạn thủ công.
- **Nâng cao tính nhất quán:** Đảm bảo chất lượng đồng đều giữa các màn chơi.
- **Duy trì vị thế cạnh tranh:** Tăng cường năng suất sản xuất nội dung, vốn là yếu tố sống còn trong game casual.

1.5. Đánh giá tổng quan và Xác định Vấn đề Nghiên cứu

Thiết kế màn chơi là yếu tố quyết định khả năng giữ chân người chơi trong Game Casual. Tuy nhiên, thực trạng tại các studio Việt Nam đang tồn tại nhiều hạn chế: quy trình thủ công tốn kém, thiếu công cụ hỗ trợ và văn hóa thiết kế thiếu nền tảng dữ liệu.

Sự phát triển của **LLM và Mô hình Khuếch tán** mang đến cơ hội khả thi để tái định hình quy trình phát triển màn chơi. AI có thể giải quyết các hạn chế bằng cách:

1. **Tự động tạo ra các biến thể layout** với tốc độ nhanh.
2. **Phân tích/Dự đoán độ khó** dựa trên dữ liệu hoặc mô phỏng.
3. **Tăng tính đồng nhất** về chất lượng và phong cách giữa các màn.

Trọng tâm Nghiên cứu: Đồ án tập trung nghiên cứu, thử nghiệm và ứng dụng sự kết hợp giữa mô hình ngôn ngữ lớn và mô hình khuyếch tán nhằm xây dựng một **công cụ hỗ trợ sinh màn chơi theo hướng tự động**.

Vấn đề Nghiên cứu Xác định: Bằng cách nào có thể khai thác và kết hợp hiệu quả mô hình ngôn ngữ lớn cùng mô hình khuếch tán để xây dựng một hệ thống có khả năng tiếp nhận mô tả màn chơi, tạo sinh bố cục hình ảnh phù hợp và nâng cao tốc độ sản xuất các màn trong game casual?

Mục tiêu cuối cùng là đánh giá khả năng triển khai mô hình trong môi trường sản xuất thực tế và đề xuất các hướng cải tiến.

1.6. Kết luận Chương

Chương 1 đã thiết lập nền tảng lý thuyết, làm rõ thách thức về nhu cầu nội dung khổng lồ của Game Casual, sự kém hiệu quả của quy trình thiết kế truyền thống, và thực trạng hạn chế tại các studio Việt Nam. Chương này khẳng định AI Tạo sinh (LLM + Diffusion Models) là giải pháp tiềm năng để hiện đại hóa quy trình phát triển game.

Trên cơ sở này, Chương 2 sẽ đi sâu vào các công nghệ AI liên quan (nguyên lý hoạt động của LLM và Diffusion Models, kỹ thuật DreamBooth) làm tiền đề cho việc xây dựng và triển khai công cụ hỗ trợ sinh màn chơi trong Chương 3.

CHƯƠNG 2: PHÂN TÍCH YÊU CẦU VÀ THIẾT KẾ HỆ THỐNG HỖ TRỢ TẠO MÀN CHƠI

Chương này trình bày cơ sở lý thuyết của Mô hình Ngôn ngữ lớn (LLM) và Mô hình Khuếch tán (Diffusion Models), đồng thời đề xuất kiến trúc kết hợp hai mô hình này để giải quyết bài toán sinh màn chơi game casual, đảm bảo cả tính logic và tính thẩm mỹ.

2.1. Mô hình Ngôn ngữ lớn (LLM)

2.1.1. Tổng quan và Kiến trúc Transformer

Sự phát triển của LLM gắn liền với kiến trúc Transformer, đã thay thế các mô hình tuần tự bằng cơ chế Attention. Cơ chế này cho phép mô hình phân tích toàn bộ chuỗi dữ liệu đầu vào cùng lúc, nắm bắt được mối quan hệ ngữ nghĩa ở khoảng cách xa, giúp biểu diễn ngữ nghĩa phong phú và tổng quát hơn.

Kiến trúc Decoder-only (như GPT, Llama) được sử dụng trong các LLM hiện đại, phù hợp cho các bài toán sinh văn bản và duy trì logic nội tại, làm nền tảng lý tưởng cho việc tạo nội dung có cấu trúc.

2.1.2. Khả năng Tạo Nội dung có Cấu trúc và Suy luận Logic

Điểm mạnh của LLM là khả năng tạo nội dung có cấu trúc (ví dụ: file JSON, ma trận) và suy luận logic, điều này rất quan trọng đối với game puzzle:

- **Tạo cấu trúc màn chơi:** LLM có thể sinh ra các ma trận hai chiều thể hiện bố cục màn chơi, bao gồm vị trí item, hướng di chuyển, và tránh được các lỗi cấu trúc cơ bản.
- **Suy luận độ khó:** LLM có thể phân tích và giải thích màn chơi, ước tính số bước giải, chỉ ra điểm nghẽn hoặc nhận diện cấu trúc gây bế tắc.
- **Kiểm chứng:** Mô hình có thể mô phỏng quá trình chơi từng bước để kiểm chứng màn chơi có khả thi hay không, hỗ trợ kiểm tra lỗi tự động.

2.1.3. Lý do LLM phù hợp cho Nhiệm vụ Sinh Màn chơi Puzzle

Game puzzle dựa vào cấu trúc rõ ràng, quy tắc cố định và ràng buộc không gian, rất phù hợp với khả năng của LLM. LLM giúp:

- **Trực quan hóa thiết kế:** Nhà thiết kế chỉ cần mô tả bằng ngôn ngữ tự nhiên.
- **Biểu diễn không gian:** Xử lý dữ liệu dưới dạng ma trận hoặc tọa độ để hình dung mối quan hệ giữa các vật phẩm.

- **Tạo biến thể:** Dễ dàng tạo ra nhiều biến thể màn chơi từ một mẫu ban đầu, tăng tốc độ sản xuất nội dung quy mô lớn.

2.2. Mô hình Khuếch tán

2.2.1. Tổng quan và Nguyên lý Hoạt động

Mô hình khuếch tán đã trở thành phương pháp mạnh mẽ nhất để tạo ra hình ảnh chất lượng cao, vượt trội hơn các mô hình trước đây (VAE, GAN) về độ ổn định và khả năng kiểm soát.

Nguyên lý hoạt động: Diffusion Models hoạt động dựa trên hai quá trình Markovian đối lập:

1. **Thêm Nhiễu (Khuếch tán Thuận):** Quá trình cố định, không cần học. Hình ảnh ban đầu được thêm nhiễu Gaussian dần dần cho đến khi trở thành nhiễu hoàn toàn.
2. **Loại bỏ Nhiễu (Khuếch tán Ngược):** Quá trình phải học. Mô hình học cách khôi phục ảnh ít nhiễu hơn từ ảnh bị nhiễu. Bắt đầu từ nhiễu thuần, mô hình dần dần tái tạo lại hình ảnh ban đầu.

Ưu điểm lớn nhất là độ ổn định cao khi huấn luyện và khả năng sinh ra hình ảnh có độ chi tiết cao và độ đa dạng lớn, là nền tảng cho các hệ thống như Stable Diffusion.

2.2.2. So sánh Mô hình Khuếch tán và GAN

Tiêu chí	Diffusion Models	GAN (Generative Adversarial Network)
Cơ chế hoạt động	Thêm nhiễu rồi học khử nhiễu dần.	Hai mạng đối kháng.
Độ ổn định khi huấn luyện	Rất ổn định, dễ train.	Dễ mất ổn định.
Chất lượng hình ảnh	Cực cao, chi tiết sắc nét.	Rất tốt nhưng kém chi tiết hơn trong phân giải cao.
Tính đa dạng của mẫu sinh	Cao.	Dễ bị thiếu đa dạng.
Khả năng điều khiển	Xuất sắc.	Kém linh hoạt hơn.

Diffusion Models là lựa chọn tối ưu hơn cho bài toán game casual do yêu cầu sự đa dạng và độ ổn định cao về chất lượng.

2.2.3. Các Mô hình Khuếch tán Hiện nay

- **Stable Diffusion:** Phổ biến nhất, cân bằng giữa chất lượng và hiệu năng.

- **ControlNet:** Bước tiến lớn về khả năng kiểm soát cấu trúc. Cho phép ràng buộc quá trình sinh ảnh dựa trên các dữ liệu cấu trúc như bản phác layout, đường viền, bản đồ chiều sâu. Điều này cực kỳ quan trọng để biến layout logic của LLM thành hình ảnh màn chơi trực quan.

2.2.4. Ứng dụng trong Sinh Nội dung Game Casual

Diffusion Models giúp tự động hóa hoàn toàn phần hình ảnh:

- **Chuyển đổi Layout Logic thành Hình ảnh:** Diffusion Models (kết hợp với ControlNet) sử dụng layout logic (tạo bởi LLM) làm điều kiện đầu vào để sinh ra hình ảnh màn chơi, đảm bảo đúng bố cục và luật chơi.
- **Sinh Asset và Item:** Tạo ra hàng trăm biến thể item nhỏ (khối gỗ, viên đá) với phong cách đồng nhất, giảm thời gian thiết kế thủ công.
- **Tùy biến Theo Chủ đề:** Khả năng mở rộng nội dung nhanh chóng theo mùa hoặc sự kiện bằng cách điều chỉnh prompt, giúp duy trì sự mới mẻ cho trò chơi.

2.3. Kiến trúc Kết hợp LLM + Diffusion

Kiến trúc này được xây dựng dựa trên nguyên tắc phân tách rõ ràng giữa khả năng hiểu và diễn giải logic của LLM, và khả năng tái tạo hình ảnh của mô hình khuếch tán.

2.3.1. Vai trò của LLM (Tầng Tiền xử lý & Điều phối)

LLM không trực tiếp tạo ảnh mà đóng vai trò là cầu nối ngữ nghĩa:

- **Phân tích và Trích xuất Thuộc tính:** Phân tích mô tả ngôn ngữ tự nhiên (thường mơ hồ) của người dùng.
- **Chuẩn hóa và Cụ thể hóa Mô tả:** Chuyển đổi mô tả thành prompt tối ưu, rõ ràng hơn, giảm khoảng cách giữa ngôn ngữ tự nhiên và ngôn ngữ điều khiển của mô hình khuếch tán.
- **Tạo và Tối ưu Prompt:** Tổ chức lại thứ tự, gán trọng số ưu tiên, bổ sung chi tiết còn thiếu cho prompt.

2.3.2. Vai trò của Mô hình Khuếch tán

- **Tái tạo Hình ảnh:** Thực hiện quá trình sinh hình ảnh chất lượng cao dựa trên prompt đã được LLM chuẩn hóa.
- **Điều khiển Cấu trúc:** Kết hợp với ControlNet để ràng buộc hình ảnh sinh ra theo các thông tin hình học hoặc bố cục cụ thể.

2.3.3. Cách thức Phối hợp và Khả năng Mở rộng

Quy trình:

1. **Input:** Người dùng nhập mô tả bằng ngôn ngữ tự nhiên.
2. **LLM:** Phân tích, trích xuất cấu trúc và tạo prompt tối ưu.
3. **Diffusion Model:** Sinh hình ảnh bố cục màn chơi dựa trên prompt đã được chuẩn hóa và các ràng buộc cấu trúc (layout).
4. **Output:** Hình ảnh bố cục màn chơi.

Sự phân công rõ ràng này giúp hệ thống ổn định, dễ huấn luyện và kiểm soát chất lượng hơn.

2.4. Phân tích Lựa chọn Mô hình

So sánh Fine-tuning và Xây dựng từ Đầu

Đồ án chọn hướng Fine-tuning mô hình có sẵn vì lý do chi phí, nguồn lực và tính khả thi:

Tiêu chí	Fine-tuning mô hình có sẵn	Tự xây dựng mô hình từ đầu
Chi phí huấn luyện	Rất thấp (chỉ cần huấn luyện một phần nhỏ trọng số).	Rất cao (cần GPU mạnh, thời gian dài).
Khối lượng dữ liệu	Ít (vài trăm đến vài nghìn mẫu).	Rất lớn (hàng chục đến hàng trăm triệu mẫu).
Độ ổn định	Cao (kế thừa tri thức từ mô hình gốc).	Thấp (dễ bị lỗi ổn định, overfit).
Phù hợp với Đồ án	Rất phù hợp.	Không phù hợp (quá tốn kém, rủi ro cao).

Lựa chọn Mô hình Khuếch tán Nền tảng

Sau khi phân tích các mô hình phổ biến (SD 1.5, SDXL, SD 3.5, Mô hình Thương mại), đồ án lựa chọn Stable Diffusion 3.5 làm mô hình chính để triển khai quá trình fine-tuning:

Tên mô hình	Ưu điểm (Lý do Lựa chọn)	Nhược điểm
Stable Diffusion 3.5	• Chất lượng hình ảnh cao, chi tiết sắc nét.	• Yêu cầu phần cứng mạnh hơn SD 1.5 (≥12 GB VRAM để fine-tuning).

	• Khả năng kiểm soát bố cục. • tương thích với tín hiệu cấu trúc từ LLM. • Dễ fine-tuning	

Quyết định: Fine-tuning Stable Diffusion 3.5 là lựa chọn tối ưu, đáp ứng yêu cầu về chất lượng hình ảnh, khả năng kiểm soát cấu trúc và tính khả thi về mặt tài nguyên.

2.5. Thiết kế Hệ thống

Hệ thống được thiết kế để sinh ra hình ảnh bố cục của màn chơi game casual dựa trên mô tả đầu vào bằng ngôn ngữ tự nhiên.

Đầu vào: Mô tả màn chơi bằng ngôn ngữ tự nhiên.

- *Ví dụ:* "Tạo một màn chơi có kích thước là 5x5, có 5 item các item có một trong bốn màu xanh đỏ tím vàng."

Đầu ra: Một hình ảnh bố cục màn chơi được sinh tự động, có cấu trúc và phong cách tương tự như các màn chơi đã được huấn luyện.

Quy trình hoạt động:

1. Người dùng nhập mô tả màn chơi.
2. Hệ thống Phân tích và Sinh bố cục Logic LLM phân tích prompt, tạo ra các ràng buộc cấu trúc (ví dụ: ma trận 5x5, vị trí 5 item, màu sắc item) và prompt tối ưu cho Diffusion Model.
3. Hệ thống Sinh Hình ảnh: Mô hình SD 3.5 (đã được fine-tuning) nhận prompt và ràng buộc cấu trúc, tiến hành sinh hình ảnh bố cục màn chơi.
4. Hệ thống trả về hình ảnh bố cục màn chơi cho người dùng.

Quy trình này đảm bảo tính đơn giản cho người dùng nhưng vẫn đảm bảo tính phức tạp và chính xác về mặt kỹ thuật .

CHƯƠNG 3: XÂY DỰNG, TRIỂN KHAI VÀ ĐÁNH GIÁ HỆ THỐNG

3.1. Xây dựng Hệ thống

3.1.1. Giới thiệu Game Thử nghiệm: Color Block Jam

Color Block Jam là một trò chơi casual puzzle có lối chơi đơn giản và sở hữu phong cách hình ảnh đặc trưng, dễ nhận diện, phù hợp với xu hướng thiết kế tối giản hiện đại. Game này được chọn làm đối tượng nghiên cứu để mô hình có thể học và tái tạo phong cách đồ họa cụ thể khi sinh ra các màn chơi mới.

Đặc trưng Hình ảnh:

- **Cấu trúc:** Sử dụng bảng lưới có đường viền mảnh, các khối màu đặt trên lưới, và các cửa thoát
- **Block Màu:** Hình dạng vuông/chữ nhật, bo tròn nhẹ, màu sắc tươi sáng, có hiệu ứng đổ bóng tinh tế để tạo cảm giác nổi 3D nhẹ.
- **Lưới (Grid):** Tối giản, đường viền mảnh, độ tương phản thấp, kích thước ô đồng nhất, giúp mô hình dễ học dạng cấu trúc lặp đều.

Sự đồng nhất và cấu trúc rõ ràng của phong cách Color Block Jam mang lại lợi điểm lớn cho việc huấn luyện mô hình sinh ảnh, giúp DreamBooth dễ học style và mô hình tái tạo tốt.

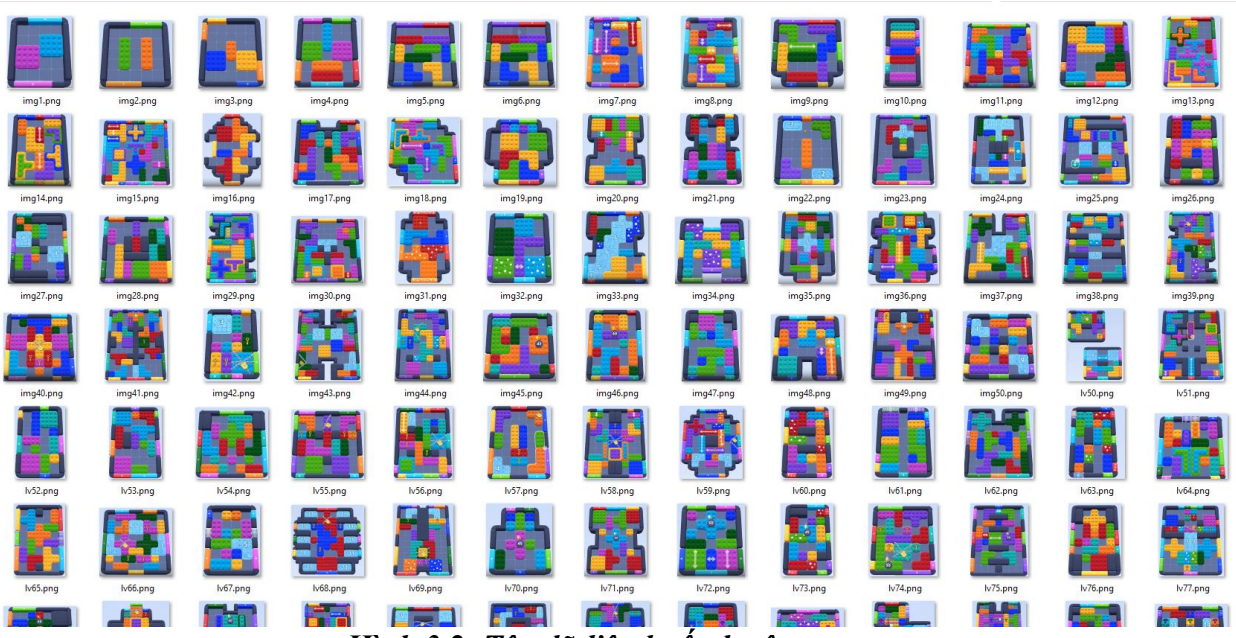


Hình ảnh game color block jam

3.1.2. Chuẩn bị Tập Dữ liệu Huấn luyện

Chất lượng dữ liệu đầu vào là yếu tố quyết định khả năng học và tái tạo phong cách của mô hình khuếch tán.

Quy trình Thu thập và Xử lý Dữ liệu:

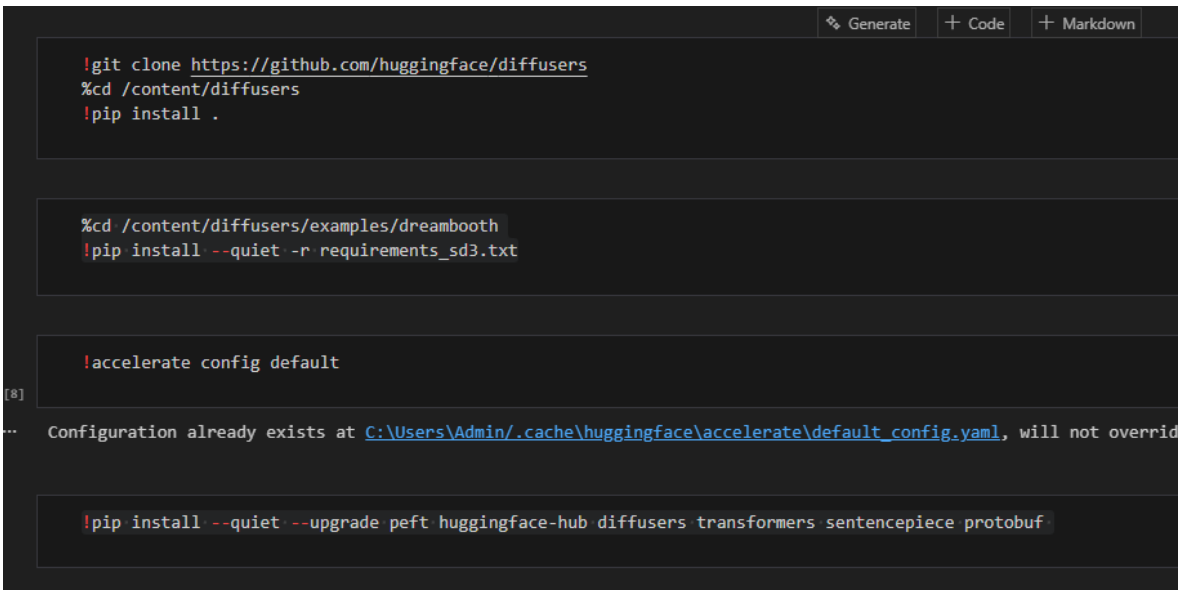
1. **Thu thập Dữ liệu:** Chụp màn hình các màn chơi từ game Color Block Jam ở nhiều cấp độ.
2. **Số lượng:** Thu thập khoảng 1000 ảnh chất lượng cao, đây là số lượng phù hợp cho phương pháp DreamBooth.
3. **Tiền xử lý:**
    - **Loại bỏ nhiễu:** Cắt bỏ các thành phần giao diện không cần thiết (nút bấm, thanh trạng thái).
    - **Chuẩn hóa kích thước:** Toàn bộ ảnh được thay đổi kích thước về độ phân giải huấn luyện 512x512.
4. **Thiết lập Nhãn Dữ liệu:** Áp dụng gán nhãn mô tả chi tiết thay vì nhãn ngắn gọn. Toàn bộ dữ liệu đều được gán chung một mô tả văn bản cụ thể và chi tiết về phong cách game. Việc này định hướng mô hình tập trung học vào toàn bộ bối cảnh và phong cách nghệ thuật.
5. 

Hình 3.2: Tập dữ liệu huấn luyện

3.1.3. Huấn luyện Mô hình

Môi trường Huấn luyện:

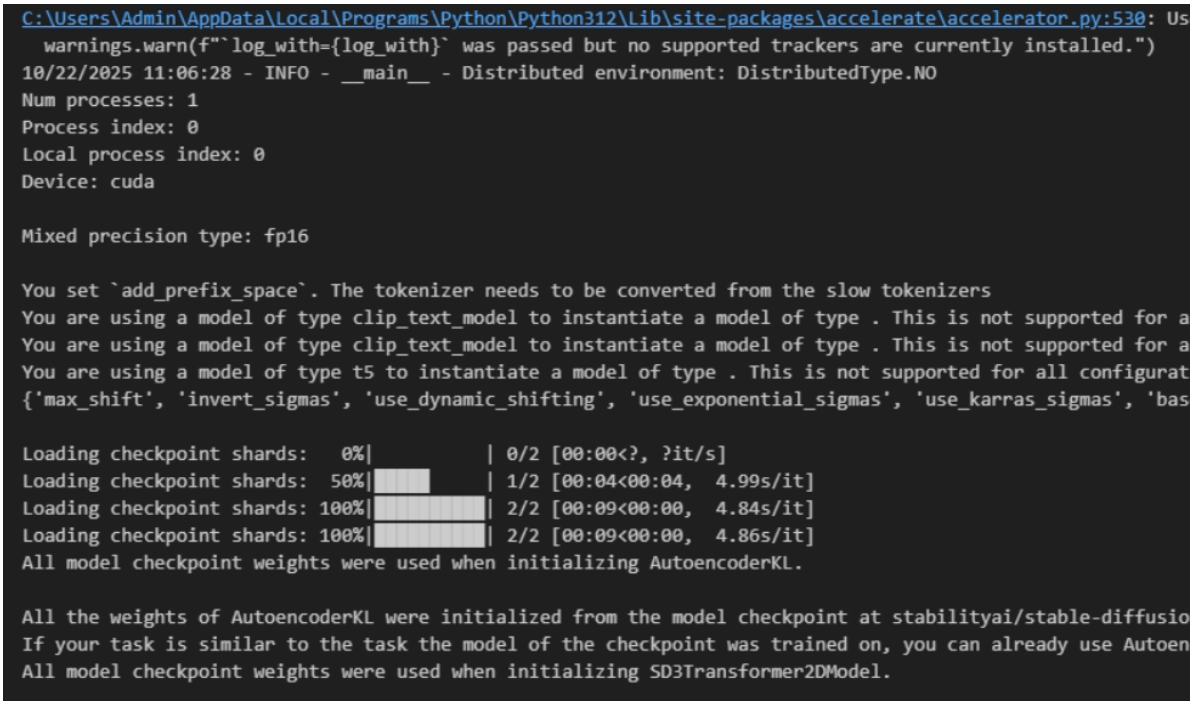
- **Phần cứng:** NVIDIA GeForce RTX 5090 (32 GB VRAM). Cấu hình VRAM lớn là then chốt để thực hiện DreamBooth Full Fine-tuning ở độ phân giải 512x512, cho phép tăng batch size và cải thiện tốc độ.
- **Phần mềm:** Python, PyTorch, thư viện HuggingFace Diffusers.



Hình 3.5: Ảnh cài đặt các thư viện

Bảng 3.1: Các tham số cần cấu hình trong khi huấn luyện

Tham số	Giá trị
Mô hình nền	stabilityai/stable-diffusion-3.5-medium
Độ phân giải	512
Batch size	1
Gradient Accumulation Steps	4
Learning Rate	5e-6
Max Train Steps	1000
Precision	Fp16
Instance Prompt	Mô tả chi tiết phong cách game Color Block Jam



Hình 3.8: Ảnh quá trình huấn luyện mô hình

Quy trình Huấn luyện: Sử dụng script train_dreambooth_sd3.py của thư viện diffusers. Quá trình lặp lại 1000 bước, trong đó trọng số của mô hình được cập nhật để học các đặc trưng thị giác (block, grid, màu sắc) theo *Instance Prompt*.

Kết quả Huấn luyện: Mô hình đã tinh chỉnh được lưu trữ dưới dạng các checkpoint (ví dụ: checkpoint-1000), bao gồm VAE, Transformer, tokenizer và text encoder, sẵn sàng cho quá trình sinh ảnh.

3.1.4. Đánh giá Mô hình Huấn luyện

Tiến hành sinh ảnh mẫu từ mô hình đã huấn luyện bằng một đoạn mã kiểm tra đơn giản. Mục tiêu là kiểm tra khả năng tái tạo phong cách nghệ thuật của Color Block Jam.

- **Đầu vào:** Sử dụng Instance Prompt chi tiết đã dùng trong huấn luyện.
- **Kết quả:** Các hình ảnh sinh ra phản ánh thành công phong cách đồ họa đặc trưng (khối màu nổi, lưới mảnh, ánh sáng nhẹ) của game. Việc này xác nhận mô hình DreamBooth đã học được các đặc trưng thị giác cốt lõi của tập dữ liệu.



Hình 3.11: Ảnh màn chơi được sinh ra từ mô hình



Hình 3.12: Ảnh màn chơi được sinh ra từ mô hình

3.2. Triển khai Hệ thống Sinh Hình ảnh Màn chơi

3.2.1. Định hướng Triển khai và Kiến trúc Tổng thể

Hệ thống được triển khai theo mô hình All-in-one, phù hợp cho mục tiêu nghiên cứu và demo, không phụ thuộc vào máy chủ hay kết nối mạng bên ngoài.

Kiến trúc: Ứng dụng được tổ chức thành ba khối chức năng chính:

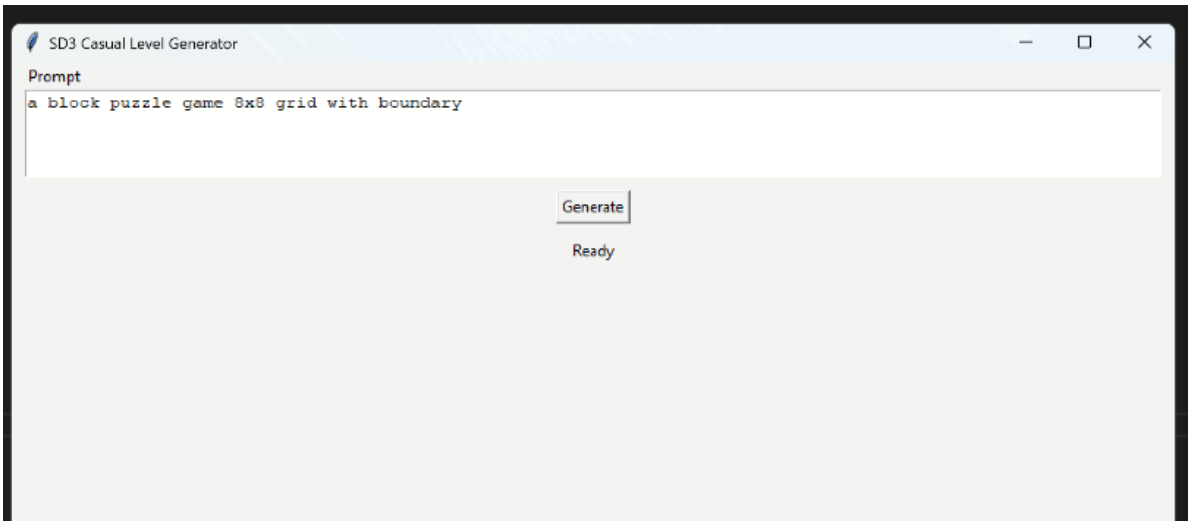
1. **Khối Giao diện Người dùng (GUI):** Cho phép người dùng nhập Prompt và quan sát kết quả.
2. **Khối Xử lý Prompt và Điều phối:** Tiếp nhận prompt, kiểm tra tính hợp lệ và điều phối luồng sinh ảnh.
3. **Khối Sinh ảnh bằng Mô hình Khuếch tán:** Sử dụng mô hình Stable Diffusion 3.5 đã tinh chỉnh để sinh ảnh.

3.2.2. Triển khai Ứng dụng và Cơ chế Tối ưu

Để tối ưu trải nghiệm người dùng, hệ thống áp dụng các cơ chế sau:

- **Lazy Loading:** Mô hình khuếch tán chỉ được load lần đầu tiên khi người dùng thực sự yêu cầu sinh ảnh, giúp giảm đáng kể thời gian khởi động ứng dụng và tránh chiếm dụng tài nguyên GPU khi không sử dụng.
- **Sinh Đa Hình ảnh:** Hệ thống sinh ra khoảng 5 hình ản**h** khác nhau cho cùng một prompt. Việc này cung cấp nhiều phương án thiết kế đa dạng cho nhà thiết kế.
- **Xử lý Song song (Threading):** Quá trình sinh ảnh được thực hiện trong luồng nền để tránh hiện tượng giao diện bị đóng băng (*freeze*), đảm bảo ứng dụng vẫn phản hồi và hiển thị trạng thái xử lý cho người dùng.

Sau khi hoàn thiện, ứng dụng được đóng gói thành một file cài đặt độc lập bằng công cụ đóng gói Python, giúp dễ dàng phân phối.



Hình 3.14: Ảnh giao diện ứng dụng

3.3. Thử nghiệm và Đánh giá

3.3.1. Mục tiêu và Phương pháp Thử nghiệm

Mục tiêu:

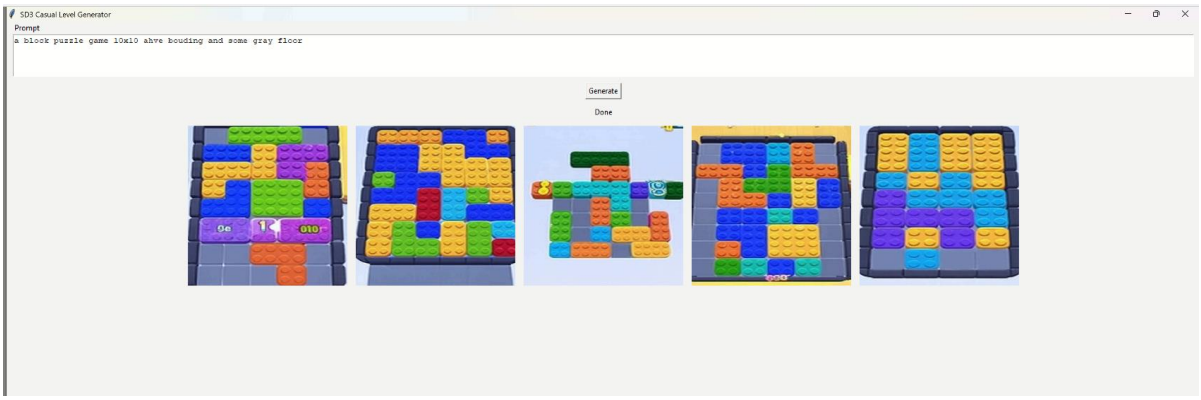
- Đánh giá Chất lượng hình ảnh (mức độ bám sát prompt, bố cục, phong cách đồ họa).
- Đo lường Tốc độ sinh ảnh (thời gian từ lúc gửi yêu cầu đến khi hiển thị kết quả).
- Đánh giá Tính ổn định và Khả năng sử dụng (hệ thống hoạt động liên tục).

Phương pháp: Thử nghiệm với hai nhóm prompt:

1. Prompt Cơ bản**:** Mô tả ngắn, đơn giản.
2. Prompt Chi tiết: Mô tả phức tạp hơn về bố cục, vật thể, màu sắc và phong cách.

Mỗi prompt sinh ra 5 hình ảnh để đánh giá khả năng tạo đa dạng thiết kế.

3.3.2. Kết quả Thử nghiệm



Hình 3.15: Ảnh hệ thống khi sinh màn chơi

- **Chất lượng Hình ảnh:**
    - Các hình ảnh sinh ra phản ánh được bố cục tổng quát (lưới, khối màu) và phong cách nghệ thuật đã học (sự đồng nhất, màu sắc tươi sáng, hiệu ứng 3D nhẹ).
    - **Kết quả mang tính minh họa và gợi ý ý tưởng** (Concept Art), có giá trị tham khảo đáng kể cho giai đoạn hình thành ý tưởng, nhưng chưa thể thay thế hoàn toàn bản thiết kế thủ công chi tiết. Mức độ chính xác phụ thuộc mạnh vào nội dung prompt.
    - *Ví dụ Prompt Chi tiết:* Hình ảnh sinh ra từ prompt chi tiết (*"top-down isometric view of a 2D puzzle board from the game Color Block, surrounded by raised outer walls with several colorful exit gates, inside the board are multiple glossy plastic blocks of various bright colors..."*) thể hiện mức độ bám sát cấu trúc tốt hơn hẳn.
- **Tốc độ Sinh ảnh:**
    - Hệ thống sinh ra 5 hình ảnh cho mỗi prompt trong khoảng thời gian vài chục giây đến khoảng một phút (tùy thuộc vào cấu hình GPU).
    - Thời gian xử lý này là mức chấp nhận được đối với một công cụ hỗ trợ thiết kế concept và nghiên cứu.
- **Tính Ổn định:**
    - Hệ thống hoạt động ổn định trong các thử nghiệm liên tục nhờ cơ chế Lazy Loading và tái sử dụng mô hình sau lần khởi tạo đầu tiên.

3.4. Kết luận Chương

Chương 3 đã trình bày thành công quá trình xây dựng, huấn luyện (DreamBooth Full Fine-tuning trên Stable Diffusion 3.5), và triển khai hệ thống sinh ảnh hỗ trợ thiết kế màn chơi game casual theo mô hình All-in-one.

Kết quả thực nghiệm cho thấy tính khả thi của việc ứng dụng mô hình khuếch tán để hỗ trợ thiết kế màn chơi, tạo ra các hình ảnh gợi ý ý tưởng ban đầu với phong cách đồ họa đồng nhất. Việc áp dụng các kỹ thuật tối ưu (Lazy Loading, xử lý song song) đã cải thiện trải nghiệm người dùng.

Mặc dù chất lượng hình ảnh sinh ra chủ yếu mang tính là gợi ý, kết quả này là cơ sở vững chắc để tiếp tục nghiên cứu, đặc biệt trong việc tích hợp LLM để cung cấp ràng buộc cấu trúc chính xác hơn nhằm nâng cao độ chính xác và tính hoàn thiện của sản phẩm.

KẾT QUẢ ĐẠT ĐƯỢC

Sau quá trình nghiên cứu, xây dựng và thử nghiệm, đồ án đã đạt được một số kết quả chính như sau:

- Đã xây dựng được một hệ thống sinh hình ảnh màn chơi dựa trên mô hình khuếch tán, cho phép người dùng nhập mô tả màn chơi bằng ngôn ngữ tự nhiên và sinh ra nhiều hình ảnh minh họa tương ứng cho cùng một yêu cầu.
- Hệ thống được triển khai theo mô hình all-in-one, tích hợp cả giao diện người dùng và mô hình xử lý trong cùng một ứng dụng, giúp quá trình cài đặt và sử dụng trở nên đơn giản.
- Kết quả thử nghiệm cho thấy các hình ảnh sinh ra có nội dung gần tương tự với mô tả đầu vào, thể hiện được bố cục tổng thể, không gian và các thành phần chính của màn chơi, qua đó hỗ trợ hiệu quả cho việc hình thành và phát triển ý tưởng thiết kế.
- Thời gian sinh ảnh và mức độ ổn định của hệ thống đáp ứng được yêu cầu trong môi trường thử nghiệm, mô hình được tải một lần và tái sử dụng cho nhiều lần sinh ảnh liên tiếp, giúp tối ưu hiệu năng.
- Đồ án đã bước đầu chứng minh tính khả thi của việc ứng dụng mô hình sinh ảnh vào lĩnh vực hỗ trợ thiết kế game, đặc biệt ở giai đoạn lên ý tưởng và phác thảo ban đầu.

KẾT LUẬN VÀ KHUYẾN NGHỊ

Thông qua đồ án, nghiên cứu và triển khai thành công một hệ thống sinh ảnh dựa trên trí tuệ nhân tạo nhằm hỗ trợ quá trình thiết kế màn chơi trong game. Hệ thống cho phép chuyển đổi mô tả bằng ngôn ngữ tự nhiên thành các hình ảnh minh họa, góp phần giảm bớt thời gian và công sức trong giai đoạn thiết kế ý tưởng.

Mặc dù chất lượng hình ảnh sinh ra chưa đạt mức hoàn chỉnh có thể từ đó để thay thế hoàn toàn các công cụ thiết kế truyền thống, kết quả thu được cho thấy hệ thống có giá trị tham khảo cao và phù hợp với mục tiêu nghiên cứu ban đầu của đồ án. Điều này khẳng định tiềm năng ứng dụng của các mô hình khuếch tán trong lĩnh vực phát triển game và thiết kế nội dung số.

Nhìn chung, đồ án đã hoàn thành các mục tiêu đặt ra, đồng thời tạo nền tảng cho các nghiên cứu và phát triển tiếp theo trong tương lai.

Bên cạnh các kết quả đạt được, hệ thống vẫn còn một số hạn chế và có thể tiếp tục được cải tiến trong tương lai. Một số hướng phát triển và khuyến nghị được đề xuất như sau:

- Nâng cao chất lượng hình ảnh sinh ra bằng cách tinh chỉnh prompt, huấn luyện bổ sung hoặc fine-tune mô hình với tập dữ liệu chuyên biệt cho thiết kế màn chơi.
- Cải tiến giao diện người dùng, bổ sung các tùy chọn điều chỉnh tham số sinh ảnh nhằm tăng khả năng kiểm soát kết quả đầu ra cho người sử dụng.
- Tối ưu hiệu năng và bộ nhớ để hệ thống có thể chạy ổn định trên các cấu hình phần cứng thấp hơn hoặc triển khai trên môi trường máy chủ.
- Mở rộng hệ thống theo hướng tích hợp với các công cụ thiết kế game hiện có, giúp quá trình chuyển đổi từ ý tưởng sang sản phẩm hoàn chỉnh trở nên thuận tiện hơn.
- Tiếp tục đánh giá hệ thống trên tập người dùng thực tế để thu thập phản hồi và cải thiện trải nghiệm sử dụng.