THÔNG TIN CHUNG CỦA BÁO CÁO

• Họ và Tên: Võ Hoàng Vũ
• MSSV: CH1902039
 
	• Lớp: CS2205.APR2023
	• Tự đánh giá (điểm tổng kết môn): 7.5/10
	• Số buổi vắng: 0
	• Số câu hỏi QT cá nhân: 1
	• Số câu hỏi QT của cả nhóm: 4
	• Link Github: https://github.com/vuvh87/CS2205.APR2023
	• Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:
	  ○ Tìm đề tài
	  ○ Viết để cương

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

NGHIÊN CỨU MÔ HÌNH KHUẾCH TÁN ẨN TRONG TỔNG HỢP ẢNH ĐỘ PHÂN GIẢI CAO

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

HIGH-RESOLUTION IMAGE SYNTHESIS WITH LATENT DIFFUSION MODELS

TÓM TẮT *(Tối đa 400 từ)*

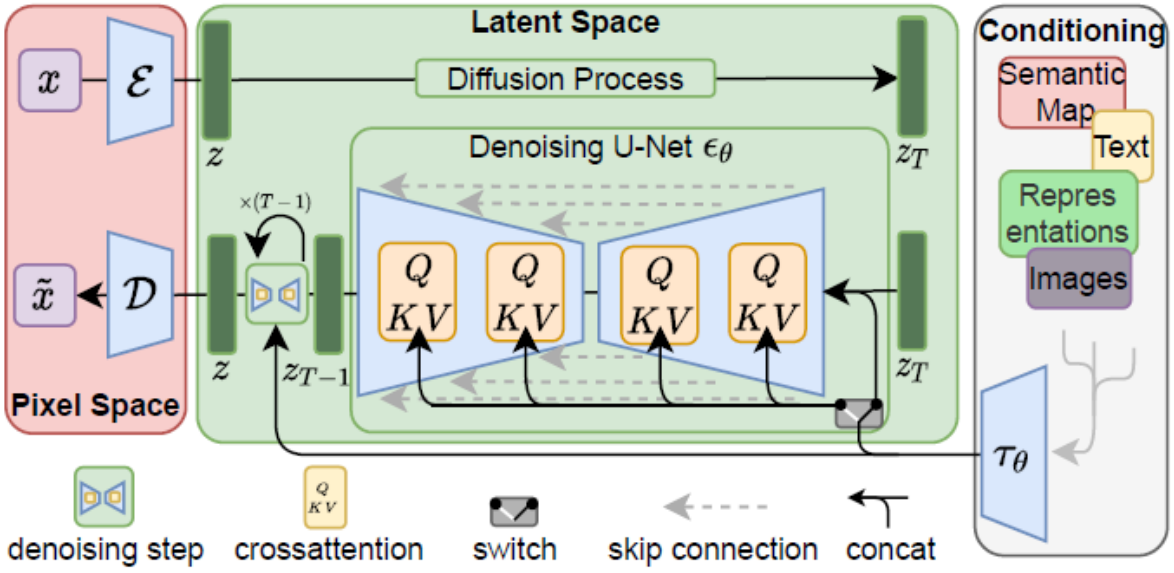
Với việc chia quá trình tạo ảnh thành các bước áp dụng tuần tự các bộ mã hóa tự động để khử nhiễu, các mô hình khuếch tán (Diffusion Model) đã đạt được các kết quả ưu việt trong tổng hợp dữ liệu hình ảnh. Hơn thế, nó còn có thể sử dụng làm cơ chế kiểm soát việc tạo hình ảnh mà không cần huấn luyện lại. Tuy nhiên, do các mô hình này thường thao tác tới từng điểm ảnh, nên đòi hỏi phải sử dụng nhiều tài nguyên GPU để tối ưu hóa các DM cũng như đánh giá theo tuần tự. Để có thể huấn luyện các DM trong điều kiện tài nguyên tính toán hạn chế mà vẫn duy trì được chất lượng và tính linh hoạt của chúng, chúng tôi thực hiện việc huấn luyện trong không gian ẩn của các bộ mã hóa tự động đã được huấn luyện trước. Ngược lại với các nghiên cứu trước, việc huấn luyện các mô hình khuếch tán theo phương pháp này đã đạt tới điểm gần tối ưu giữa việc giảm độ phức tạp và bảo toàn chi tiết, từ đó tăng đáng kể độ trung thực của hình ảnh. Với việc thêm các lớp cross-attention vào kiến trúc mô hình, chúng tôi biến các mô hình khuếch tán thành các trình tạo mạnh mẽ và linh hoạt dùng các điều kiện đầu vào như văn bản hoặc khung giới hạn để tổng hợp với độ phân giải cao bằng phương pháp tích chập. Mô hình khuếch tán ẩn (Latent Diffusion Model) của chúng tôi đã tạo ra một mốc ưu việt mới trong việc in ảnh và tổng hợp hình ảnh có điều kiện theo lớp, và hiệu suất cao trong các tác vụ khác nhau, bao gồm tạo hình ảnh không điều kiện, tổng hợp văn bản thành hình ảnh và siêu phân giải. Đồng thời, mô hình giảm đáng kể yêu cầu tính toán so với phương pháp trực tiếp dựa trên pixel.

GIỚI THIỆU *(Tối đa 1 trang A4)*

Những phát triển gần đây của lĩnh vực thị giác máy tính đã tạo nên những tiến bộ ấn tượng trong tổng hợp hình ảnh. Việc tổng hợp các cảnh tự nhiên, phức tạp với độ phân giải cao có thể thực hiện được được bằng cách nhân rộng các mô hình likelihood-based với hàng tỷ tham số trong quá trình biến đổi tự hồi quy (autoregressive transformer). Tuy nhiên, GAN[1], một mô hình phổ biến, đã bộc lộ những hạn chế do quy trình học tập đối nghịch (adversarial learning). Ngược lại, các mô hình khuếch tán (DM)[2], được xây dựng từ các bộ mã hóa tự động để khử nhiễu, đã cho thấy kết quả ấn tượng trong việc tổng hợp hình ảnh, tổng hợp hình ảnh theo lớp điều kiện và độ phân giải siêu cao.

Tuy nhiên, DM có khối lượng tính toán lớn, đòi hỏi nhiều tài nguyên để huấn luyện và suy luận[3]. Hạn chế này khiến số nhà nghiên cứu tham gia vào chỉ dừng lại ở một con số khiêm tốn và làm tăng mối lo ngại về môi trường do lượng khí thải carbon lớn từ việc huấn luyện các mô hình như vậy.

Để giải quyết những thách thức này, chúng tôi đề xuất một phương pháp để giảm độ phức tạp tính toán của DM mà không ảnh hưởng đến hiệu suất của chúng. Đó là Mô hình Khuếch tán ẩn (LDM), gồm việc huấn luyện các bộ mã hóa tự động để tìm ra một không gian tương đương nhưng khối lượng tính toán phù hợp hơn để huấn luyện các DM. Cách tiếp cận này cho phép tạo hình ảnh hiệu quả và giảm việc nén không gian quá mức.



Nhìn chung, cách tiếp cận này đơn giản hóa quá trình tổng hợp hình ảnh có độ phân giải cao, làm cho các mô hình mạnh mẽ như DM dễ tiếp cận hơn với các nhà nghiên cứu đồng thời giảm mức tiêu thụ tài nguyên đáng kể và tác động đến môi trường. Sự kết hợp giữa bộ mã hóa tự động và mô hình khuếch tán cho phép khám phá hiệu quả nhiều loại tác vụ và nâng cao trình độ tiên tiến nhất trong lĩnh vực tổng hợp hình ảnh và các lĩnh vực liên quan.

MỤC TIÊU

(Viết trong vòng 3 mục tiêu, lưu ý về tính khả thi và có thể đánh giá được)

Tìm hiểu mô hình tri giác nén (Perceptual compression model) dựa trên bộ mã hóa tự động, nhằm mục đích tái tạo lại hình ảnh một cách hiệu quả trong khi vẫn giữ được các chi tiết trong không gian tiềm ẩn đã học.

Tìm hiểu các DM để tổng hợp hình ảnh, giới thiệu các ưu điểm của các mô hình tri giác nén đã được huấn luyện để tạo mô hình tổng quát của các biểu diễn ẩn với số chiều thấp.

Cải tiến các mô hình khuếch tán (DM) để trở thành các trình tạo hình ảnh có điều kiện linh hoạt hơn bằng cách kết hợp cơ chế cross-attention, cho phép chúng lập mô hình phân phối có điều kiện dựa trên các đầu vào như văn bản, bản đồ ngữ nghĩa hoặc các tác vụ dịch từ hình ảnh sang hình ảnh khác, từ đó mở rộng khả năng tổng hợp của

chúng tới các lớp nhãn và các biến thể hình ảnh bị làm mờ.

NỘI DUNG VÀ PHƯƠNG PHÁP

(Viết nội dung và phương pháp thực hiện để đạt được các mục tiêu đã nêu)

Tiến hành tìm hiểu các công trình nghiên cứu liên quan đến mô hình tri giác nén và các bộ mã hóa tự động liên quan. Xây dựng mô hình tri giác nén dựa trên bộ mã hóa tự động, bao gồm việc xác định kiến trúc của bộ mã hóa và bộ giải mã. Áp dụng các phương pháp học máy và huấn luyện mô hình tri giác nén trên tập dữ liệu ảnh phù hợp để đạt được kết quả tối ưu. Đánh giá mô hình bằng cách so sánh các độ đo và thực hiện kiểm tra trên các tập dữ liệu kiểm tra khác nhau để kiểm tra hiệu suất của mô hình tri giác nén. Đề xuất các cải tiến và thử nghiệm các biến thể của mô hình để cải thiện hiệu suất và hiệu quả trong việc tái tạo hình ảnh.

Tiến hành tìm hiểu và phân tích các mô hình khuếch tán (DMs) đã được nghiên cứu. Đánh giá và so sánh các ưu điểm của các mô hình đã được huấn luyện trước, đặc biệt là khả năng tạo ra mô hình tổng quát của các biểu diễn ẩn với số chiều thấp. Xây dựng mô hình tổng quát sử dụng các mô hình tri giác nén tốt nhất được lựa chọn sau đánh giá. Áp dụng các phương pháp đánh giá và thử nghiệm để đảm bảo tính hiệu quả và hiệu suất của mô hình tổng quát này.

Nghiên cứu cải tiến các mô hình khuếch tán (Diffusion Models - DMs) để trở thành các trình tạo hình ảnh có điều kiện linh hoạt hơn bằng cách tích hợp cơ chế cross-attention. Điều này cho phép mô hình thực hiện việc tổng hợp hình ảnh dựa trên các đầu vào như văn bản, bản đồ ngữ nghĩa hoặc các tác vụ dịch từ hình ảnh sang hình ảnh khác. Qua đó, khả năng tổng hợp hình ảnh của các mô hình DM được mở rộng tới các lớp nhãn và các biến thể hình ảnh bị làm mờ.

KẾT QUẢ MONG ĐỢI

(Viết kết quả phù hợp với mục tiêu đặt ra, trên cơ sở nội dung nghiên cứu ở trên)

Cải tiến thành công mô hình khuếch tán ẩn, một mô hình đơn giản và hiệu quả để cải thiện đáng kể cả việc huấn luyện và lấy mẫu của các mô hình khuếch tán khử nhiễu mà không làm giảm chất lượng của chúng. Kết hợp với cơ chế điều khiển

cross-attention, các thí nghiệm của chúng tôi chứng minh kết quả khả quan hơn so với các phương pháp state-of-the-art trên nhiều tác vụ tổng hợp hình ảnh theo điều kiện mà không cần xây dựng từng kiến trúc riêng cho mỗi tác vụ.

TÀI LIỆU THAM KHẢO *(Định dạng DBLP)*

[1] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. In Int. Conf. Learn. Represent., 2019. 1, 2, 6, 7, 8, 19, 26

[2] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. CoRR, abs/1503.03585, 2015. 1, 3, 4, 15

[3] Prafulla Dhariwal and Alex Nichol. Diffusion models beat gans on image synthesis. CoRR, abs/2105.05233, 2021. 1, 2, 3, 4, 6, 7, 8, 15, 19, 23, 24, 26

[4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In NIPS, pages 5998–6008, 2017. 3, 4, 5, 6