

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

Võ Hoàng Vũ - CH1902039

Tóm tắt

- Lớp: CS2205.APR2023
- Link Github: <https://github.com/vuvh87/CS2205.APR2023>
- Link YouTube video:



Giới thiệu

- Những tiến bộ trong lĩnh vực tổng hợp hình ảnh:
 - Tổng hợp hình ảnh có độ phân giải cao sử dụng các mô hình likelihood-based có hàng tỷ tham số
 - Các mô hình khuếch tán (Diffusion Models) mang lại các kết quả đầy hứa hẹn trong việc tổng hợp hình ảnh phức tạp
- Thách thức:
 - Các mô hình DM yêu cầu cao về xử lý tính toán và bị hạn chế trong khả năng tiếp cận

Giới thiệu

- Mô hình cải tiến được đề xuất:
 - Latent Diffusion Model: mô hình khuếch tán ẩn
 - Ưu điểm: Dễ mở rộng, tối ưu hiệu suất, tái tạo hình ảnh trung thực, đa dạng phương pháp huấn luyện, tổng hợp hình ảnh có độ phân giải cao với mức tiêu thụ tài nguyên thấp

Mục tiêu

- Tìm hiểu mô hình tri giác nén (Perceptual Compression Model)
- Tổng hợp hình ảnh bằng LDM và ưu điểm của mô hình
 - Tạo mô hình tổng quát với biểu diễn ẩn với số chiều thấp
 - Cải tiến DM với cross-attention cho khả năng tạo hình ảnh có điều kiện linh hoạt
- Mở rộng khả năng tổng hợp với LDM
 - Sử dụng văn bản, bản đồ ngữ nghĩa và tác vụ dịch hình ảnh để tạo mô hình phân phối có điều kiện
 - Tổng hợp hình ảnh với các lớp nhãn và biến thể hình ảnh bị làm mờ

Nội dung và Phương pháp

- Xây Dựng và Đánh Giá Mô Hình Tri Giác Nén
 - Tìm hiểu các công trình nghiên cứu liên quan đến mô hình tri giác nén và bộ mã hóa tự động
 - Xây dựng mô hình tri giác nén dựa trên bộ mã hóa tự động với kiến trúc đã xác định
 - Áp dụng phương pháp học máy và huấn luyện mô hình trên tập dữ liệu ảnh phù hợp để đạt kết quả tối ưu
 - Đánh giá mô hình bằng so sánh các độ đo và kiểm tra trên các tập dữ liệu khác nhau
 - Đề xuất cải tiến và thử nghiệm các biến thể để cải thiện hiệu suất và hiệu quả tái tạo hình ảnh

Nội dung và Phương pháp

- Nghiên Cứu và Xây Dựng Mô Hình Tổng Quát LDM
 - Tìm hiểu và phân tích các mô hình khuếch tán (DMs) đã được nghiên cứu
 - Đánh giá và so sánh ưu điểm của các mô hình huấn luyện trước
 - Khả năng tạo mô hình tổng quát của biểu diễn ẩn với số chiều thấp
 - Xây dựng mô hình tổng quát sử dụng các DM tốt nhất sau đánh giá
 - Áp dụng phương pháp đánh giá và thử nghiệm để đảm bảo tính hiệu quả và hiệu suất

Nội dung và Phương pháp

- Cải Tiến Mô Hình Khuếch Tán (DMs) với Cross-Attention
 - Tạo mô hình tạo hình ảnh có điều kiện linh hoạt hơn
 - Cơ chế Cross-Attention: Tích hợp cơ chế này để tổng hợp hình ảnh dựa trên các đầu vào như văn bản, bản đồ ngữ nghĩa, và các tác vụ dịch hình ảnh
 - Mở rộng khả năng tổng hợp: Đạt được khả năng tổng hợp hình ảnh linh hoạt hơn, bao gồm các lớp nhãn và biến thể hình ảnh bị làm mờ

Kết quả dự kiến

- Mô hình khuếch tán cải tiến với không gian ẩn
 - Cải thiện việc huấn luyện và lấy mẫu của mô hình khuếch tán khử nhiễu
 - Mô hình đơn giản và hiệu quả, không giảm chất lượng
- Kết hợp cơ chế điều khiển Cross-Attention
 - Sử dụng một kiến trúc cho nhiều tác vụ tổng hợp hình ảnh theo điều kiện
 - Các thực nghiệm chứng minh kết quả khả quan hơn so với các phương pháp state-of-the-art

Tài liệu tham khảo

- [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