

TẠO HÌNH ẢNH KHUÔN MẶT TỪ MÔ TẢ VĂN BẢN SỬ DỤNG STYLEGAN2

Võ Ngô Văn Tiền - 230201033

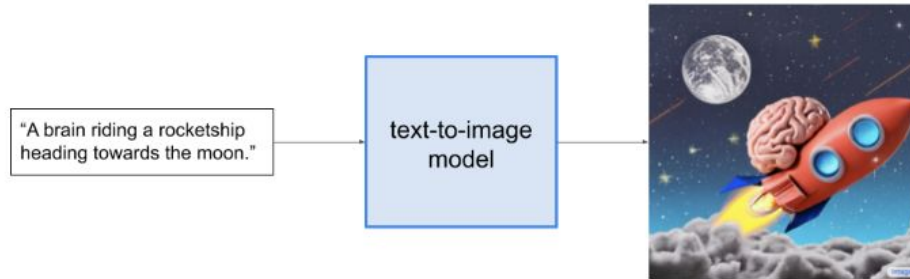
Tóm tắt

- Lớp: CS2205.MAR2024
- Link Github:
<https://github.com/vongovantien/CS2205.CH181>
- Link YouTube video:
https://youtu.be/_5qb8vxQ8-q
- Ảnh + Họ và Tên: Võ Ngô Văn Tiền



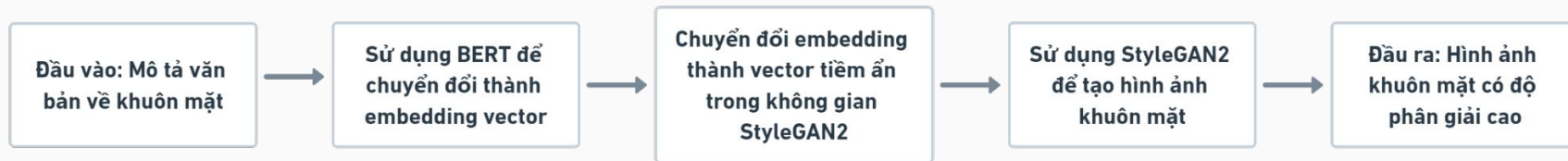
Giới thiệu

- Trong những năm gần đây, AI đã có những bước tiến vượt bậc trong việc tạo ra hình ảnh chất lượng cao.
- Mạng Đối Kháng Tạo Sinh (GANs) đã chứng minh khả năng tạo ra dữ liệu hình ảnh chân thực và kiểm soát nhiều đặc điểm khác nhau của hình ảnh.



Giới thiệu

- Text-to-face generation (T2F) là lĩnh vực đầy thách thức do sự đa dạng của các thuộc tính khuôn mặt nhưng có nhiều ứng dụng quan trọng trong an ninh và sinh trắc học. StyleGAN2 là một trong những công nghệ tiên tiến nhất trong việc tạo hình ảnh giúp cải thiện chất lượng và sự khớp ngữ nghĩa giữa mô tả văn bản và hình ảnh
- Đầu vào của hệ thống là các mô tả văn bản chi tiết về **các đặc điểm khuôn mặt**. Đầu ra là **hình ảnh khuôn mặt với các chi tiết khớp với văn bản**



Mục tiêu

- **Tạo hình ảnh khuôn mặt có độ phân giải cao từ mô tả văn bản:** Sử dụng StyleGAN2 để tạo ra hình ảnh khuôn mặt có độ phân giải cao (1024x1024 pixel) từ các mô tả văn bản chi tiết.
- **Đảm bảo sự khớp với mô tả văn bản:** Phát triển mô hình chuyển đổi từ văn bản sang không gian tiềm ẩn của StyleGAN2, đảm bảo rằng các đặc điểm khuôn mặt được mô tả bằng văn bản sẽ được thể hiện chính xác trong hình ảnh tạo ra.
- **Đánh giá và cải thiện tính khả thi:** Thử nghiệm và phân tích định lượng để xác định hiệu suất của mô hình, sử dụng các tập dữ liệu từ CelebA và Text2FaceGAN để huấn luyện và kiểm tra.

Nội dung và Phương pháp

Nội dung 1: Tạo hình ảnh khuôn mặt có độ phân giải cao từ mô tả văn bản

Sử dụng BERT để xử lý văn bản: Trích xuất các đặc điểm từ mô tả văn bản bằng mô hình ngôn ngữ BERT, mã hóa văn bản thành các vector ngữ nghĩa phong phú.

Ánh xạ vector ngữ nghĩa vào không gian tiềm ẩn của StyleGAN2: Sử dụng mạng thần kinh nhiều lớp (MLP) để chuyển đổi các vector ngữ nghĩa thành vector tiềm ẩn.

Nội dung và Phương pháp

Nội dung 2: Đảm bảo sự khớp với mô tả văn bản

Tạo hình ảnh từ vector tiềm ẩn: Sử dụng StyleGAN2 để tạo ra hình ảnh khuôn mặt từ các vector tiềm ẩn đã được chuyển đổi. Vector tiềm ẩn được đưa vào StyleGAN2, sau đó mô hình sẽ tổng hợp hình ảnh khuôn mặt tương ứng.

Kiểm tra và đánh giá: Hình ảnh tạo ra được kiểm tra và đánh giá dựa trên các chỉ số như độ phân giải, độ chân thực và sự phù hợp với mô tả văn bản. Các chỉ số như Face Semantic Distance (FSD) và Face Semantic Similarity (FSS) được sử dụng để đánh giá sự phù hợp ngữ nghĩa.

Nội dung và Phương pháp

Nội dung 3: Đánh giá và cải thiện tính khả thi

Huấn luyện và kiểm tra mô hình: Sử dụng các tập dữ liệu lớn như CelebA và Text2FaceGAN để huấn luyện và kiểm tra mô hình. Quá trình huấn luyện bao gồm việc sử dụng logistic regression và linear regression để tìm ra các hướng đặc điểm trong không gian tiềm ẩn.

Đánh giá chất lượng hình ảnh: Sử dụng Fréchet Inception Distance (FID) để đánh giá chất lượng hình ảnh tạo ra, đảm bảo rằng hình ảnh gần giống với hình ảnh thật. Các chỉ số FID thấp hơn chỉ ra rằng hình ảnh tạo ra có chất lượng cao hơn.

Kết quả dự kiến

Nghiên cứu dự kiến tạo ra hình ảnh khuôn mặt có độ phân giải cao (1024x1024 pixel) từ mô tả văn bản, đảm bảo phản ánh chính xác các đặc điểm như màu tóc, kiểu tóc, và hình dạng khuôn mặt.

Ứng dụng thực tế bao gồm điều tra tội phạm, nghệ thuật kỹ thuật số, và hệ thống nhận dạng khuôn mặt. Công việc tương lai sẽ tập trung giải quyết các hạn chế hiện tại và nghiên cứu thêm về tính đa dạng của các khuôn mặt tạo ra từ một mô tả văn bản duy nhất.

Tài liệu tham khảo

- [1]. Mohamed Shawky Sabae, Mohamed Ahmed Dardir, Remonda Talaat Eskarous, Mohamed Ramzy Ebbed. StyleT2F: Generating Human Faces from Textual Description Using StyleGAN2. arXiv preprint arXiv:2204.07924, 2022.
- [2]. D. M. A. Ayanthi, Sarasi Munasinghe. Text-to-Face Generation with StyleGAN2. In: David C. Wyld et al. (Eds): FCST, CMIT, SE, SIPM, SAIM, SNLP - 2022, CS & IT - CSCP 2022, pp. 49-64, 2022. DOI: 10.5121/csit.2022.120805
- [3]. X. Cao, H. Wei, P. Wang, C. Zhang, S. Huang, H. Li. High Quality Coal Foreign Object Image Generation Method Based on StyleGAN-DSAD. Sensors, vol. 23, no. 1, p. 374, Dec. 2022. DOI: 10.3390/s23010374
- [4]. A. H. Bermano, R. Gal, Y. Alaluf, R. Mokady, Y. Nitzan, O. Patashnik, D. Cohen-Or. State-of-the-Art in the Architecture, Methods and Applications of StyleGAN. Computer Graphics Forum, vol. 41, no. 2, pp. 591-611, May 2022. DOI: 10.1111/cgf.14503