


THÔNG TIN CHUNG CỦA BÁO CÁO

- Link YouTube video của báo cáo (tối đa 5 phút):
(ví dụ: <https://www.youtube.com/watch?v=AWq7uw-36Ng>)
- Link slides (dạng .pdf đặt trên Github):
(ví dụ: <https://github.com/mynameuit/CS2205.APR2023/TenDeTai.pdf>)
- Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới
- Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in

<ul style="list-style-type: none">● Họ và Tên: Võ Ngô Văn Tiền● MSSV: 230201033 	<ul style="list-style-type: none">● Lớp: CS2205.CH181● Tự đánh giá (điểm tổng kết môn): 8.5/10● Số buổi vắng: 0● Số câu hỏi QT cá nhân: 3● Link Github: https://github.com/vongovantien/CS2205.CH181/● Link Youtube: https://youtu.be/_5qb8vxQ8-g
---	---

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

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

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

GENERATING FACIAL IMAGES FROM TEXT DESCRIPTIONS USING
STYLEGAN2

TÓM TẮT (Tối đa 400 từ)

Trong những năm gần đây, việc tạo hình ảnh dựa trên AI đã có những bước tiến đáng kể. Đặc biệt, các mạng đối kháng tạo sinh (GANs) đã trở thành một trong những mô hình tạo sinh thành công nhất, không chỉ tạo ra dữ liệu chất lượng cao và chân thực mà còn có khả năng kiểm soát kết quả đầu ra. Mục tiêu của nghiên cứu này là ứng dụng GANs vào việc tạo khuôn mặt từ các mô tả văn bản chi tiết.

Nghiên cứu đề xuất phát triển một phương pháp sử dụng GANs để tự động tổng hợp hình ảnh khuôn mặt từ mô tả văn bản. Mục tiêu là đảm bảo rằng hình ảnh đầu ra khớp với mô tả chi tiết về các đặc điểm khuôn mặt như màu tóc, màu mắt, hình dáng khuôn mặt và các đặc điểm khác. Ví dụ, mô tả như “Một người có mái tóc xoăn, khuôn mặt trái xoan và ria mép” cần phải được chuyển đổi thành hình ảnh khuôn mặt cụ thể.

Để đạt được mục tiêu này, chúng tôi sẽ thực hiện các bước sau: Thứ nhất, sử dụng BERT để chuyển đổi các mô tả văn bản thành các embedding vector, cung cấp một biểu diễn ngữ nghĩa phong phú cho quá trình tạo hình ảnh. Thứ hai, phát triển một mô hình học sâu để chuyển đổi các embedding từ BERT thành các vector tiềm ẩn trong không gian StyleGAN2, cho phép điều khiển chính xác các đặc điểm khuôn mặt dựa trên mô tả văn bản. Thứ ba, sử dụng mạng tổng hợp của StyleGAN2 để tạo ra các hình ảnh khuôn mặt từ các vector tiềm ẩn đã được điều chỉnh, đảm bảo độ phân giải chi tiết rõ nét.

Chúng tôi kỳ vọng rằng phương pháp này sẽ tạo ra hình ảnh khuôn mặt phù hợp với mô tả văn bản, đảm bảo tính nhất quán giữa văn bản đầu vào và hình ảnh đầu ra. Phương pháp này có tiềm năng lớn, mở ra nhiều cơ hội ứng dụng trong các lĩnh vực như an ninh công cộng và điều tra tội phạm, giúp nhận dạng và xác định danh tính dựa trên mô tả văn bản.

GIỚI THIỆU *(Tối đa 1 trang A4)*

Việc tạo ra hình ảnh từ văn bản đã trở thành một lĩnh vực nghiên cứu hấp dẫn nhờ sự phát triển của Mạng Tạo sinh Đối kháng (GANs). GANs có khả năng tạo ra hình ảnh chất lượng cao và chân thực. Tuy nhiên, các phiên bản đầu tiên của GAN gặp khó khăn trong việc kiểm soát các đặc điểm cụ thể của hình ảnh đầu ra dựa trên mô tả đầu vào. Để giải quyết vấn đề này, các mô hình GAN điều kiện đã được phát triển, đặc biệt trong việc tạo hình ảnh từ văn bản (Text-to-Image generation - TTI).

Text-to-Face generation (TTF) là một nhánh cụ thể của TTI tập trung vào việc tạo hình ảnh khuôn mặt từ văn bản có nhiều ứng dụng, đặc biệt trong điều tra tội phạm, giúp xác định danh tính từ mô tả của nhân chứng. Mặc dù có tiềm năng lớn, lĩnh vực này chưa được khai thác đầy đủ do các thách thức kỹ thuật.

Đầu vào của hệ thống bao gồm các mô tả văn bản chi tiết về khuôn mặt như hình dáng, màu tóc, kiểu tóc và râu ria. Bộ dữ liệu huấn luyện gồm hình ảnh khuôn mặt và mô tả văn bản tương ứng, được mã hóa bằng mô hình BERT để biểu diễn trong không gian tiềm ẩn của StyleGAN2.

Đầu ra của hệ thống là các hình ảnh khuôn mặt độ phân giải cao (1024x1024) được tạo ra từ mô tả văn bản, với độ tương đồng cao với hình ảnh thực tế. Mô hình cụ thể như sau:



MỤC TIÊU

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

Mục tiêu đầu tiên của nghiên cứu là 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. StyleGAN2 được chọn vì khả năng tạo ra hình ảnh chân thực và có độ phân giải cao. Để cải thiện chất lượng hình ảnh và giảm thiểu hiện tượng tạo ảnh giả (artifacts), chúng tôi sẽ áp dụng các kỹ thuật như điều chỉnh trọng số (weight demodulation), chuẩn hóa đường dẫn (path length regularization) và loại bỏ quá trình tăng trưởng dần (progressive growing).

Một mục tiêu quan trọng khác là tạo ra các hình ảnh không chỉ có chất lượng cao mà còn phải phù hợp ngữ nghĩa với mô tả đầu vào. Điều này có nghĩa là các thuộc tính khuôn mặt được mô tả bằng văn bản cần được thể hiện chính xác trong hình ảnh tạo ra. Để đánh giá điều này, chúng tôi sẽ sử dụng các chỉ số như Face Semantic Distance (FSD) và Face Semantic Similarity (FSS).

Mục tiêu cuối cùng là đánh giá và cải thiện tính khả thi của mô hình thông qua các thử nghiệm và phân tích định lượng. Chúng tôi sẽ sử dụng các tập dữ liệu từ CelebA và Text2FaceGAN để huấn luyện và kiểm tra mô hình. Thử nghiệm sẽ xác định không gian tiềm ẩn phù hợp nhất và các lớp trích xuất đặc trưng tốt nhất trong mô hình tồn thất cảm nhận. Hiệu suất của mô hình sẽ được đánh giá bằng các chỉ số như Fréchet Inception Distance (FID), Face Semantic Distance (FSD) và Face Semantic Similarity (FSS).

NỘI DUNG VÀ PHƯƠNG PHÁP

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

Nội dung 1: Tạo ra hình ảnh khuôn mặt có độ phân giải cao từ mô tả văn bản bằng cách sử dụng StyleGAN2

Nghiên cứu và áp dụng StyleGAN2 để tạo ra hình ảnh khuôn mặt có độ phân giải cao từ mô tả văn bản chi tiết. Chúng tôi sẽ sử dụng bộ dữ liệu hình ảnh khuôn mặt từ CelebA để huấn luyện StyleGAN2, áp dụng các kỹ thuật tiên tiến như điều chỉnh trọng số, chuẩn hóa đường dẫn và loại bỏ quá trình tăng trưởng dần để cải thiện chất lượng hình ảnh và giảm thiểu hiện tượng tạo ảnh giả. Hình ảnh tạo ra sẽ được kiểm thử và đánh giá bằng các chỉ số như Fréchet Inception Distance (FID) để đảm bảo chất lượng.

Nội dung 2: Đảm bảo tính nhất quán ngữ nghĩa giữa mô tả văn bản và hình ảnh tạo ra

Sử dụng mô hình BERT để mã hóa các mô tả văn bản thành các vector ngữ nghĩa, giúp hiểu sâu về ngữ cảnh và ý nghĩa của các mô tả. Sau đó, phát triển mô hình chuyển đổi từ văn bản sang không gian tiềm ẩn của StyleGAN2 để ánh xạ các mô tả vào không gian đầu vào của StyleGAN2. Tính nhất quán ngữ nghĩa sẽ được đánh giá bằng các chỉ số như Khoảng cách ngữ nghĩa khuôn mặt (FSD) và Độ tương đồng ngữ nghĩa khuôn mặt (FSS).

Nội dung 3: Đánh giá và cải thiện tính khả thi của mô hình qua các thử nghiệm và phân tích định lượng

Đánh giá và cải thiện tính khả thi của mô hình thông qua các thử nghiệm và phân tích định lượng. Chúng tôi sẽ sử dụng các tập dữ liệu từ CelebA và Text2FaceGAN để huấn luyện và kiểm tra mô hình, xác định không gian tiềm ẩn phù hợp nhất và các lớp trích xuất đặc trưng tốt nhất. Hiệu suất của mô hình sẽ được đánh giá bằng các chỉ số như FID, FSD và FSS để đảm bảo hiệu quả tối ưu trong việc tạo hình ảnh từ mô tả văn bản.

KẾT QUẢ MONG ĐỢI

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

Chúng tôi kỳ vọng sẽ tạo ra các hình ảnh khuôn mặt có độ phân giải cao (1024x1024 pixel) từ mô tả văn bản chi tiết, bằng cách sử dụng StyleGAN2 và các kỹ thuật tiên tiến. Các hình ảnh này sẽ rõ ràng và chân thực, phản ánh chính xác các đặc điểm khuôn mặt được mô tả. Đặc biệt, hình ảnh tạo ra sẽ đảm bảo tính nhất quán ngữ nghĩa cao, thể hiện đúng các yếu tố như màu tóc, kiểu tóc và hình dạng khuôn mặt, duy trì đầy đủ các yếu tố ngữ nghĩa từ văn bản đầu vào.

Nghiên cứu này sẽ phát triển một đề xuất mới cho việc tạo ảnh từ văn bản, mở ra tiềm năng ứng dụng rộng rãi trong nhiều lĩnh vực như điều tra tội phạm, nghệ thuật kỹ thuật số và hệ thống nhận dạng khuôn mặt, đáp ứng nhu cầu sử dụng thực tế. Kết quả nghiên cứu cũng sẽ đóng góp quan trọng vào sự phát triển và ứng dụng của công nghệ AI, mang lại nhiều lợi ích thực tiễn trong tương lai.

TÀI LIỆU THAM KHẢO *(Định dạng DBLP)*

- [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

