

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

Võ Ngô Văn Tiền¹

¹ Trường ĐH Công Nghệ Thông Tin TP.HCM- ĐHQG TP.HCM

What ?

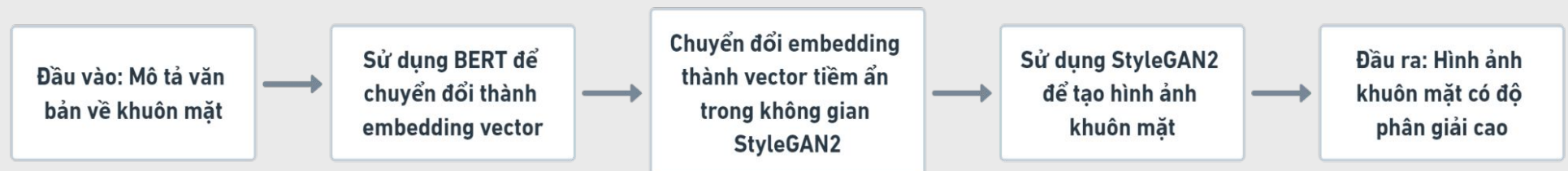
Nghiên cứu này khác biệt và cải thiện so với các nghiên cứu trước đây bằng cách sử dụng mô hình StyleGAN2 để tạo ra hình ảnh khuôn mặt có độ phân giải cao từ mô tả văn bản. Trước đây, các mô hình Text-to-Face (T2F) thường gặp khó khăn trong việc tạo ra hình ảnh có độ phân giải cao và phù hợp với mô tả. Bằng cách sử dụng StyleGAN2 và mã hóa BERT, nghiên cứu này giải quyết vấn đề chất lượng hình ảnh và sự phù hợp ngữ nghĩa, cải thiện đáng kể so với các phương pháp trước đó

Why ?

Nghiên cứu này giới thiệu một phương pháp mới để tạo hình ảnh khuôn mặt từ mô tả văn bản. Các nghiên cứu trước thường tạo ra hình ảnh có độ phân giải thấp và không giống với mô tả văn bản. Nghiên cứu này không chỉ cải thiện chất lượng hình ảnh mà còn giảm thiểu các vấn đề về độ chính xác ngữ nghĩa, với kết quả vượt trội cho thấy tiềm năng ứng dụng trong an ninh công cộng, như điều tra tội phạm và nghiên cứu sinh trắc học

Overview

Mạng đối kháng sinh (GANs) đã đạt được những tiến bộ vượt bậc trong việc tạo ra hình ảnh chất lượng cao, tái hiện chính xác dữ liệu thực tế. Text-to-Face Generation (T2F) là một lĩnh vực nghiên cứu đầy thách thức nhưng mang lại nhiều ứng dụng quan trọng trong an ninh công cộng và nghiên cứu sinh trắc học. Sử dụng các công nghệ tiên tiến như StyleGAN2 và BERT, nghiên cứu này tập trung vào việc tạo ra hình ảnh khuôn mặt có độ phân giải cao từ mô tả văn bản, đảm bảo tính chính xác ngữ nghĩa và chất lượng hình ảnh vượt trội so với các phương pháp trước đây.

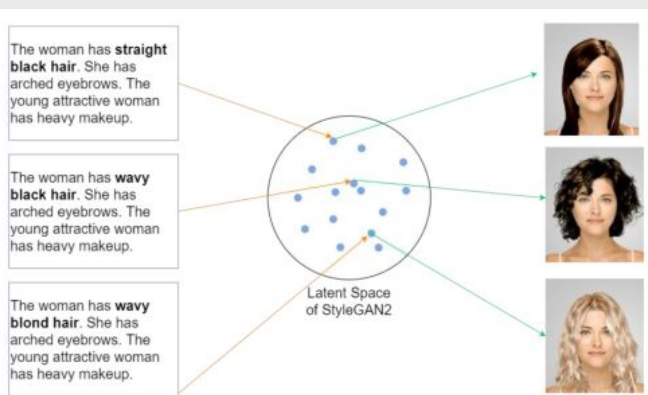


Description

1. Xử lý văn bản

Đầu tiên, các mô tả văn bản được mã hóa thành các vector ngữ nghĩa phong phú thông qua mô hình ngôn ngữ BERT. BERT, với khả năng đào tạo hai chiều, giúp hiểu rõ ngữ cảnh và các đặc điểm được mô tả từ cả hai phía của câu văn. Quá trình mã hóa này đảm bảo rằng các đặc điểm khuôn mặt được mô tả trong văn bản được chuyển đổi chính xác thành các vector ngữ nghĩa, đại diện cho các thuộc tính khuôn mặt cần thiết.

2. Tạo mã tiềm ẩn



Hình 1. Biểu diễn mô tả trong không gian tiềm ẩn của StyleGAN2 và tương ứng hình ảnh khi các thuộc tính trong mô tả được thay đổi.

Tiếp theo, các vector ngữ nghĩa này được chuyển đổi thành vector tiềm ẩn để tạo ra hình ảnh khuôn mặt. Quá trình này sử dụng mạng thần kinh nhiều lớp (MLP) để ánh xạ các vector ngữ nghĩa vào không gian tiềm ẩn của StyleGAN2.

Quá trình này bao gồm ba bước chính:

- Trích xuất hướng đặc trưng: Các hướng trong không gian tiềm ẩn được trích xuất để tách biệt và điều khiển các đặc điểm khuôn mặt cụ thể.
- Tạo vector khởi tạo: Bắt đầu từ một vector tiềm ẩn ngẫu nhiên, sau đó sử dụng các hướng đặc trưng để đạt được vector tiềm ẩn mục tiêu.
- Điều khiển vector tiềm ẩn: Điều khiển các hướng đặc trưng để đạt được giá trị đặc trưng mục tiêu, đảm bảo các đặc điểm khuôn mặt được mô tả chính xác.

3. Tạo hình ảnh và đánh giá

Các vector tiềm ẩn được chuyển đổi được đưa vào StyleGAN2 để tạo ra hình ảnh khuôn mặt, phản ánh chính xác các đặc điểm trong mô tả văn bản. Chất lượng hình ảnh được đánh giá bằng các chỉ số như Face Semantic Distance (FSD), Face Semantic Similarity (FSS) và Fréchet Inception Distance (FID), đảm bảo tính chân thực và phù hợp ngữ nghĩa.

Phương pháp này sử dụng StyleGAN2 và BERT, cải thiện chất lượng và độ chính xác của hình ảnh, mở ra nhiều ứng dụng tiềm năng trong an ninh, giải trí và nghiên cứu sinh trắc học.