

Xây dựng mô hình chuyển đổi văn bản thành hình ảnh sử dụng GAN

Sinh viên Nguyễn Tuấn Sinh Phạm Trung Sơn & Nguyễn Tiến Sĩ & Trần Hồ Phi Long
Giảng viên dạy PGS. TS Huỳnh Trung Hiếu & Lưu Giang Nam
Hội đồng phản biện PGS. TS Huỳnh Trung Hiếu & TS. Nguyễn Chí Kiên & TS. Bùi Thanh Hùng & Lưu Giang Nam

Tóm tắt nội dung

Mô hình tạo văn bản thành hình ảnh (text2image – T2I) nhằm mục đích tạo ra các hình ảnh chân thực về mặt ngữ nghĩa nhất quán từ các mô tả văn bản. Với những tiên bộ gần đây trong kĩ thuật Deep Learning đặc biệt là kiến trúc mạng generative adversarial networks (GAN) thì các mô hình T2I đã có những bước tiến vượt bậc. Tuy nhiên, việc kiểm tra kỹ các hình ảnh được tạo ra từ mô hình T2I có một số hạn chế như: Các phương pháp chuẩn hóa hàng loạt được áp dụng trên toàn bộ bản đồ đặc trưng hình ảnh như nhau, bỏ qua ngữ nghĩa cục bộ; Bộ mã hóa văn bản được cố định trong khi đào tạo, nên được đào tạo với trình tạo hình ảnh cùng nhau để tìm hiểu các biểu diễn văn bản tốt hơn để tạo hình ảnh. Để giải quyết những hạn chế này, chúng tôi đề xuất một mô hình GAN được đào tạo theo kiểu end-to-end để bộ mã hóa văn bản có thể khai thác thông tin văn bản tốt hơn. Cụ thể, mô hình sẽ học cách chuyển đổi ngữ nghĩa thích nghi dựa trên văn bản để kết hợp hiệu quả các tính năng văn bản và tính năng hình ảnh. Đồng thời, mô hình học bản đồ mặt nạ theo cách được giám sát vào quá trình hợp nhất hình ảnh văn bản hiện tại theo thứ tự để chuyển đổi theo không gian. Một số tập data được sử dụng trong bài toán này như (COCO, 102flowers, CUB2002011). Chúng tôi đã đạt được một số kết quả so với các phương pháp hiện đại gần đây, liên quan đến cả độ trung thực và sự liên kết với mô tả văn bản đầu vào.

Bố cục bài báo cáo

1. Giới thiệu
2. Nghiên cứu liên quan
3. Giới thiệu mô hình GAN
4. Thí nghiệm
5. Bộ dữ liệu và số liệu đánh giá
6. Kết luận

Tài liệu

- [1] Y. Chen, Y. -K. Lai and Y. -J. Liu, CartoonGAN: Generative Adversarial Networks for Photo Cartoonization 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 9465-9474, doi: 10.1109/CVPR.2018.00986.
- [2] F. Andersson, S. Arvidsson, Generative Adversarial Networks for photo to Hayao Miyazaki style cartoons arXiv:2005.07702, 2020, doi: 10.48550/arXiv.2005.07702.