

PixelBrush: Tạo nghệ thuật từ văn bản với GAN

Sinh viên Phạm Hà Văn Đông & Phạm Thế Hiển & Tăng Hoài Duy & Phạm Tài Đức Phú Đa & Lê Tân Đạt
 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

Trong những năm gần đây các mô hình GAN ngày càng phát triển mạnh mẽ và có tính ứng dụng cao trong thực tế, vì thế trong dự án này chúng tôi đề xuất áp dụng mô hình GAN trong lĩnh vực nghệ thuật tạo ra các tác phẩm nghệ thuật một cách dễ dàng từ mô tả bằng văn bản. Trong dự án này chúng tôi đã hợp nhất tập huấn luyện, val và tập kiểm tra lại với nhau và chọn khoảng 3000 hình ảnh từ Bò, Chó, Ngựa, Cừu để huấn luyện và khoảng 600 hình ảnh từ 4 danh mục đó để thử nghiệm từ bộ dữ liệu tranh Oxford bộ dữ liệu này chứa 8629 hình ảnh trong 10 danh mục: hàng không, chim, thuyền, ghế, bò, bàn, chó, ngựa, cừu, xe lửa để áp dụng vào thuật toán của chúng tôi bằng cách tiếp cận là đào tạo các mạng đối thủ tạo ra có điều kiện dựa trên các đặc điểm văn bản được mã hóa bằng cách nhúng văn bản vectơ suy nghĩ bỏ qua được mã hóa từ một mô tả văn bản. Cho trước dữ liệu huấn luyện x và văn bản y , cả mạng tạo và mạng phân biệt đều được huấn luyện cùng nhau để có thể học được phân phối của x . từ đó tạo ra được một tác phẩm nghệ thuật chân thật nhất.

Bổ cục bài báo cáo

1. Giới Thiệu: Trong dự án này, chúng tôi đề xuất một công cụ mới gọi là PixelBrush đưa ra mô tả về một bức tranh, tạo ra một hình ảnh nghệ thuật từ mô tả đó. Đầu vào cho công cụ của chúng tôi là một đoạn văn bản ngắn, chẳng hạn như một con chim đang bay trên bầu trời. Sau đó, chúng tôi sử dụng RNN để xử lý văn bản đầu vào thành một vectơ và sau đó sử dụng GANs để tạo ra các hình ảnh nghệ thuật.
2. Những công việc liên quan: Trong công việc của chúng tôi, chúng tôi đề xuất một cách tạo tác phẩm nghệ thuật từ các mô tả văn bản. Chúng tôi sẽ sử dụng ngôn ngữ tự nhiên làm đầu vào để mọi người có thể mô tả loại tác phẩm nghệ thuật mà họ muốn, sau đó công cụ PixelBrush của chúng tôi sẽ tạo ra một hình ảnh theo mô tả được cung cấp.
3. Background: Giới thiệu về GAN, cGAN, Text Embedding
4. Phương Pháp: Kiến trúc mạng, iến trúc của generator, discriminator
5. Dữ Liệu: Chúng tôi sử dụng bộ dữ liệu tranh Oxford để giúp chúng tôi phát triển thuật toán của mình. Bộ dữ liệu này chứa 8629 hình ảnh trong 10 danh mục: hàng không, chim, thuyền, ghế, bò, bàn, chó, ngựa, cừu, xe lửa. Nó được chia thành các tập training, validation và test.
6. Thí Nghiệm:
 - Phương pháp đánh giá: Sử dụng Inception Score để đánh giá chất lượng hình ảnh được tạo
 - So sánh với baselines: Sử dụng kiến trúc generator Simple cho cả cGAN và DCGAN rồi so sánh kết quả.
 - So sánh giữa các generator khác nhau: Sử dụng 3 kiến trúc generator bao gồm: Simple, Normal và Deep sau đó so sánh Inception Score sau mỗi 1000 Iterations.
 - Phân tích Entropy: So sánh giá trị Entropy của ảnh thực và ảnh được tạo ra trên cùng 1 số lượng Iteration nhất định.
7. Kết luận: Trong công việc này, chúng tôi đã đề xuất một công cụ PixelBrush mới giúp cGAN tạo ra các hình ảnh chân thực và nghệ thuật từ các mô tả văn bản. Chúng tôi cho thấy rằng bằng cách thêm mô tả văn bản làm điều kiện, hình ảnh được tạo trông chân thực hơn và có thể nhìn thấy các đối tượng dễ nhận dạng hơn trong hình ảnh được tạo. Chúng tôi cũng chỉ ra rằng độ sâu generator khác nhau ảnh hưởng như thế nào đến chất lượng hình ảnh được tạo bằng cách sử dụng Inception Score làm đánh giá. Đóng góp chính của công việc này là chỉ ra rằng cGAN có thể tạo ra hình ảnh chân thực và nghệ thuật trên bộ dữ liệu phức tạp hơn – bộ dữ liệu tranh Oxford mặc dù bộ dữ liệu CUB 200 thường được sử dụng.

8. Công việc tương lai: - Thử nghiệm thuật toán trên tập dữ liệu The Behance Artistic Media Dataset.

- Công việc của chúng tôi cho thấy rằng chúng tôi có thể tạo ra các hình ảnh thực tế từ các mô tả văn bản, nhưng để đưa điều này vào ứng dụng, chúng tôi cần tạo ra các hình ảnh lớn hơn với độ phân giải cao hơn. Hiện tại, độ phân giải hình ảnh đầu ra bị giới hạn ở 64×64 .

<http://cs231n.stanford.edu/reports/2017/pdfs/322.pdf> <https://www.robots.ox.ac.uk/vgg/data/paintings/> <https://github.com/ajai-nair/art-generation-gan>

Tài liệu

- [1] J. Zhi (2017). PixelBrush: Art Generation from text with GANs. cs231n.stanford.edu/reports/2017/pdfs/322.pdf
- [2] E. Crowley and A. Zisserman. (2014) The State of the Art: Object Retrieval in Paintings using Discriminative Regions. *In Proceedings British Machine Vision Conference 2014*. doi: 10.5244/C.28.38
- [3] E. Crowley and A. Zisserman. (2015) In Search of Art. *Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science*, vol 8925. Springer, Cham. doi: 10.1007/978-3-319-16178-54
- [4] A. Radford, L. Metz, S. Chintala (2015) Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv:1511.06434*. doi: 10.48550/arXiv.1511.06434