

Chinese Painting Generation Using Generative Adversarial Networks

1st Phạm Thế Hiển

Trường Đại Học Công Nghiệp Thành Phố Hồ Chí Minh
Thành Phố Hồ Chí Minh, Việt Nam
Hienchipham1212@gmail.com

3rd Lê Tấn Đạt

Trường Đại Học Công Nghiệp Thành Phố Hồ Chí Minh
Thành Phố Hồ Chí Minh, Việt Nam
Tandat9a4tiengiang@gmail.com

5th Phạm Hà Văn Đông

Trường Đại Học Công Nghiệp Thành Phố Hồ Chí Minh
Thành Phố Hồ Chí Minh, Việt Nam
vandong875@gmail.com

2nd Phạm Tài Đức Phú Đa

Trường Đại Học Công Nghiệp Thành Phố Hồ Chí Minh
Thành Phố Hồ Chí Minh, Việt Nam
phuda575@gmail.com

4th Tăng Hoài Duy

Trường Đại Học Công Nghiệp Thành Phố Hồ Chí Minh
Thành Phố Hồ Chí Minh, Việt Nam
tanghoaiduy360@gmail.com

TÓM TẮT

Dự án này triển khai mô hình DCGAN trên bộ dữ liệu Chinese Painting của chính chúng tôi để tạo ra các bức tranh giống như thật.

I. GIỚI THIỆU

Trong những năm gần đây, học sâu đã được chứng minh là một trong những công cụ mạnh mẽ nhất trong trí tuệ nhân tạo và đang bắt đầu thay đổi cuộc sống của chúng ta. Có rất nhiều ứng dụng thành công của deep learning. Ví dụ: Google DeepMind đã phát triển AlphaGo, một trí tuệ nhân tạo dạy mạng thần kinh tích chập sâu để chơi trò chơi cờ vây [8]. Gần đây, AlphaGo đã đánh bại Ke Jie, kỳ thủ hàng đầu thế giới trong ván đấu thứ ba liên tiếp. Trong bối cảnh thị giác máy tính, học sâu đã được sử dụng rộng rãi trong nhận dạng khuôn mặt, nhận dạng hình ảnh chế độ xem phổ, truy xuất hình ảnh, v.v.

Một cuộc thảo luận thú vị khác hiện nay trong lĩnh vực học sâu là cách nó có thể tác động và định hình quá trình sản xuất văn hóa và nghệ thuật trong tương lai của chúng ta. Tuy nhiên, tạo ra các tác phẩm nghệ thuật là một nhiệm vụ đầy thách thức, đặc biệt là đối với các bức tranh phong cảnh của Trung Quốc. Những khó khăn nằm ở các khía cạnh sau: (a) Bối cảnh của tranh phong cảnh Trung Quốc khó nhận ra hơn, các họa sĩ Trung Quốc thích sử dụng mây và sương mù để tạo ra bầu không khí mờ ảo; (b) Trong một bức tranh Trung Quốc thường có nhiều đối tượng như núi, sông, cầu và sương mù; (c) Hình dạng của những vật thể đó thường không đều đặn; (d) Hiện không có bộ dữ liệu nguồn mở nào về tranh Trung Quốc.

Mục tiêu của dự án của chúng tôi là tạo ra những bức tranh Trung Quốc bằng cách sử dụng deep neural network. Chúng tôi sử dụng Deep Convolutional GAN (DCGAN) để đào tạo hàng nghìn bức tranh phong cảnh Trung Quốc và sau đó tạo ra các bức tranh chân thực một cách tổng hợp. Đầu vào là bản đồ cạnh của một bức tranh và chúng tôi muốn tái tạo lại hình ảnh mục tiêu đó.

II. CÔNG VIỆC LIÊN QUAN

GAN lần đầu tiên được giới thiệu bởi Goodfellow et al. [10] năm 2014, trong bài báo này, các tác giả sử dụng GAN để tạo ảnh tự nhiên bằng bộ dữ liệu MNIST [17], TFD [23] và CIFAR-10 [16]. Hóa ra hình ảnh do GAN tạo ra sắc nét hơn đáng kể so với hình ảnh được đào tạo bằng các phương pháp khác dựa trên các mục tiêu đào tạo có maximum likelihood. Sau đó, GAN chủ yếu được áp dụng để mô hình hóa các hình ảnh tự nhiên và các biến thể khác nhau của GAN đã được giới thiệu. Năm 2014, Mirza và cộng sự. các mạng đối thủ tạo ra có điều kiện (cGAN) [20] được đề xuất, kiểm soát các chế độ của dữ liệu được tạo. Tuy nhiên, GAN được biết là không ổn định để huấn luyện và thường gặp sự cố sập chế độ. Vào năm 2015, Radford đã đề xuất một kiến trúc ổn định hơn được gọi là mạng đối nghịch thể hệ tích chập sâu (DCGAN) mở rộng quy mô GAN bằng cách sử dụng mạng thần kinh tích chập (CNN) để mô hình hóa hình ảnh. Gần đây, Arjovsky et al. [5] đã định nghĩa một dạng GAN mới được gọi là GAN Wasserstein (WGAN). WGAN cải thiện tính ổn định của việc học và có thể tạo hình ảnh ở các chế độ khác nhau. Có nhiều biến thể khác của GAN như LSGAN [19], Bayesian GAN [25], WGAN cải tiến [11], Cramer GAN [6].

Nhờ khả năng mạnh mẽ trong việc tạo ảnh chất lượng cao của GAN, có rất nhiều ứng dụng thú vị của GAN bao gồm

tạo khuôn mặt [9], tổng hợp văn bản thành ảnh [22], dịch ảnh sang ảnh [14], dịch ảnh không ghép nối thành ảnh [27], tô màu tự động [15], độ phân giải siêu cao cho một ảnh chân thực [18].

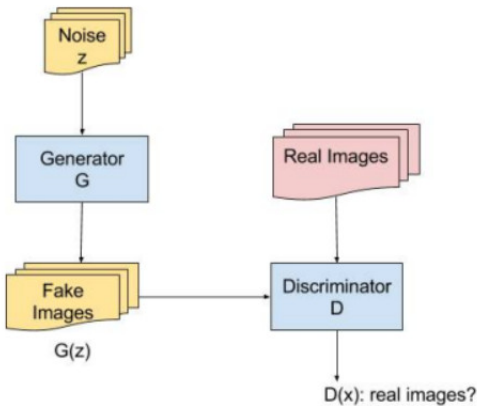
Tuy nhiên, đã có rất ít bài báo cố gắng tạo ra các tác phẩm nghệ thuật bằng cách sử dụng GAN, gần đây, ArtGAN [24] đã được giới thiệu để tạo ra các bức tranh phương Tây. ArtGAN sử dụng Condition GAN dựa trên thông tin nhân. Nhưng một điều đáng lo ngại là một số tranh phương Tây, đặc biệt là tranh Trừu tượng, miêu tả các vật thể được “trừu tượng hóa” khỏi thiên nhiên. Những loại tranh đó có thể khó hiểu hơn đối với những người không phải là chuyên gia so với tranh tượng trưng. Do đó, có thể không dễ dàng so sánh kết quả của ArtGAN với các biến thể khác của GAN một cách trực quan.

III. PHƯƠNG PHÁP

Trong phần này, trước tiên chúng tôi sẽ xem xét ngắn gọn về GAN và sau đó giới thiệu các thuật toán của một số biến thể của GAN như DCGAN, WGAN và cGAN.

3.1. Generative Adversarial Networks

Generative Adversarial Networks GAN, như tên gọi của nó, chứa hai adversarial networks, một generator và một discriminator. Nó có thể được xây dựng như một trò chơi minimax hai người chơi. Generator (G) tạo ảnh từ random noise, tức là, nó học cách ánh xạ từ vectơ nhiễu ngẫu nhiên z sang ảnh y , $G: z \rightarrow y$. Discriminator (D) đồng thời học một ánh xạ từ một hình ảnh x đến một số giá trị từ 0 đến 1, cho biết xác suất đầu vào đến từ phân phối dữ liệu thực. Mục tiêu của D là phân biệt các mẫu từ mô hình tổng quát và các mẫu từ dữ liệu huấn luyện, trong khi mục tiêu của G là tạo ra các hình ảnh không thể phân biệt được với các hình ảnh từ hình ảnh huấn luyện (nghĩa là đánh lừa D), các thủ tục là thể hiện trong Hình 1 [3]. Tại điểm cân bằng, G sẽ tạo ra hình ảnh thực và D sẽ tạo ra xác suất là 0,5.

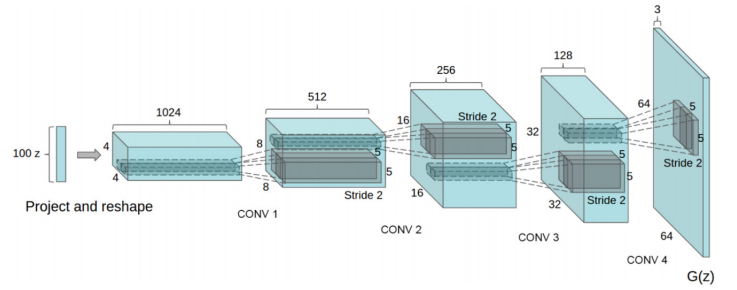


Hình 1. Kiến trúc của GAN

Cụ thể, thủ tục có thể được coi là minimax chức năng mục tiêu sau đây.

$$\min_G \max_D \mathcal{L}_{GAN}(D, G) = \min_G \max_D \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log(1 - D(G(z)))].$$

3.2. Deep Convolutional Generative Adversarial Networks



Hình 2. Kiến trúc của DCGAN

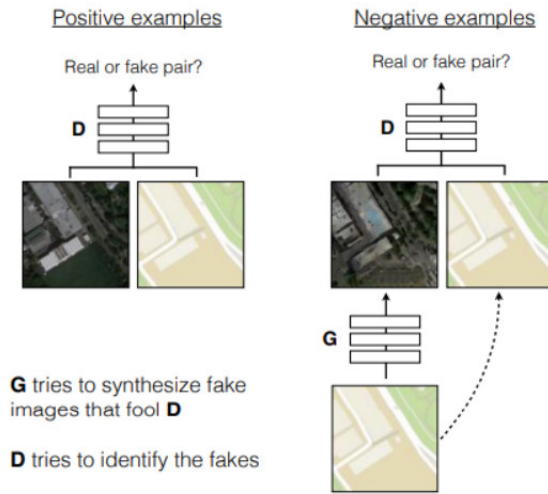
Deep Convolutional Generative Adversarial Networks (DCGAN) là một phiên bản cải tiến của GAN ban đầu, bổ sung thêm các lớp tích chập trong cả kiến trúc của generator và discriminator. Kiến trúc khác với GAN ban đầu ở các khía cạnh chính sau đây, như được đề xuất bởi [21]:

- Sử dụng các kết chập có bước cho discriminator và các kết chập chuyển vị cho generator.
- Sử dụng batch normalization trong cả generator và discriminator
- Loại bỏ các lớp ẩn FC để có kiến trúc sâu hơn.

DCGAN ổn định hơn GAN gốc và xuất ra hình ảnh sắc nét hơn, nhưng vẫn có một số dạng không ổn định. Cả GAN và DCGAN đều gặp phải sự cố sập chế độ và biến mất độ dốc.

3.3. Conditional Generative Adversarial Networks

Condition GAN bảo tồn hầu hết tất cả các cấu trúc của GAN, vẫn có G và D và chúng chơi một trò chơi minimax. Sự khác biệt là đầu vào của G và D . Giờ đây, G học cách ánh xạ từ ảnh được quan sát x và vectơ nhiễu ngẫu nhiên z , tới ảnh y , $G: (x, z) \rightarrow y$. Tương tự, D học một ánh xạ từ (x, y) đến $[0, 1]$, cố gắng phân biệt hình ảnh với dữ liệu huấn luyện và trình tạo. Quy trình đào tạo được sơ đồ hóa trong Hình 3 từ [14].



Hình 3. Kiến trúc của cGAN

Cụ thể, thủ tục có thể được coi là minimax chức năng mục tiêu mới

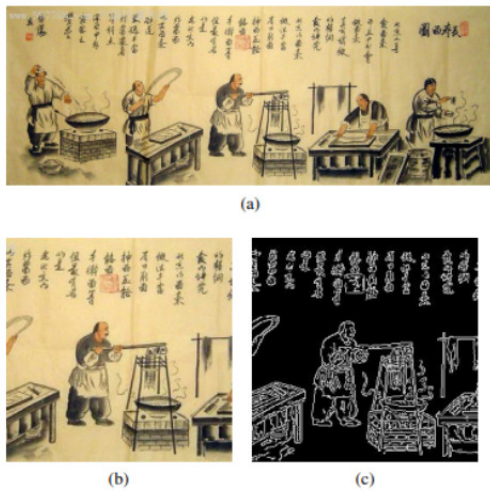
$$\min_G \max_D \mathcal{L}_{cGAN}(D, G)$$

Trong đó:

$$\mathcal{L}_{cGAN}(D, G) = \min_G \max_D \mathbb{E}_{x, y \sim p_{data}(x, y)} [\log D(x, y)] + \mathbb{E}_{z \sim p_z, y \sim p_y(y)} [\log(1 - D(y, G(y, z)))].$$

IV. DỮ LIỆU VÀ ĐẶC TRƯNG

Chúng tôi có tổng cộng 5798 bức tranh trong tập dữ liệu. Mỗi bức tranh đều có kích thước 256x256. Sau đó chúng tôi trích xuất các cạnh của bức tranh và dùng nó làm đầu vào.



Hình 4. Một ví dụ về điểm dữ liệu được xử lý trước. (a) Các hình gốc, (b) hình đã cắt, (c) cạnh của hình đã cắt nhân vật.

V. THÍ NGHIỆM VÀ KẾT QUẢ

5.1. Tạo tranh Trung Quốc sử dụng DCGAN

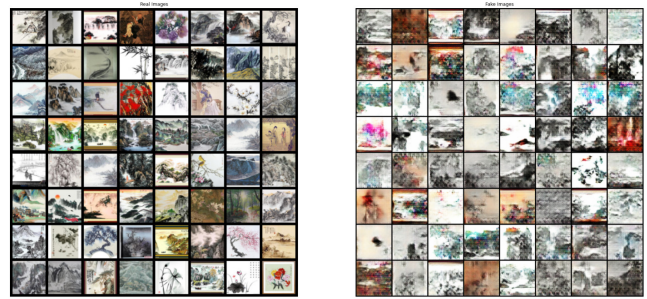
5.1.1. Cài đặt thử nghiệm

Chúng tôi sử dụng các kết chập có bước tiến cho D và các kết chập chuyển vị cho G . Batchnorm được sử dụng trong cả D và G . Đối với DCGAN, chúng tôi sử dụng trình tối ưu hóa Adam với learning rate 0,0002.

Đầu ra của bộ D là một xác suất cho DCGAN (áp dụng sigmoid trong D). Đầu vào là hình ảnh 3 kênh có kích thước 256×256 pixel. Có thể tìm thấy mã cho mô hình DCGAN cơ sở mà chúng tôi đã sử dụng tại đây [2].

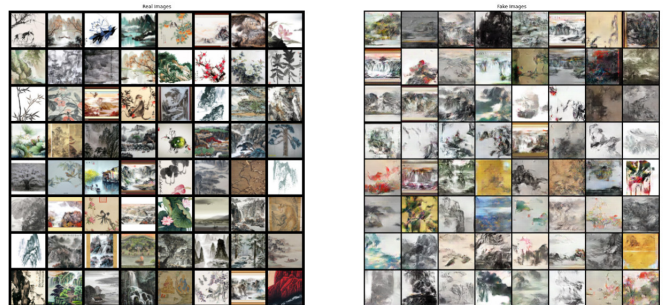
5.1.2. Kết quả thí nghiệm và so sánh

Thử nghiệm đầu tiên của chúng tôi đã đào tạo từng DCGAN trong 200 epoch, với batch size 64 và output size $64 \times 64 \times 3$. Kết quả cuối cùng được hiển thị trong Hình 5.



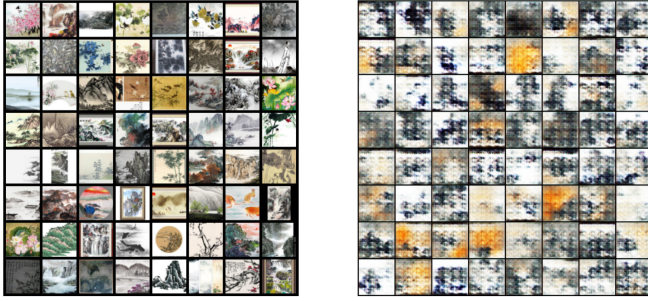
Hình 5. Đầu ra kết quả 64×64 sau 200 epoch và ảnh thật của DCGAN

Chúng tôi đã thực hiện thử nghiệm thứ hai, thuật toán DCGAN được đào tạo trong 500 epoch, với batch size 64 và output size $64 \times 64 \times 3$. Kết quả cuối cùng được hiển thị trong Hình 6.



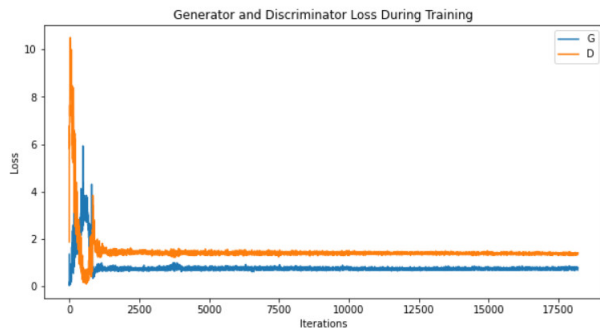
Hình 6. Đầu ra kết quả 64×64 sau 500 epoch và ảnh thật của DCGAN

Chúng tôi thực hiện thêm thử nghiệm thứ ba, thuật toán DCGAN được đào tạo trong 100 epoch, với batch size 16 và output size $64 \times 64 \times 3$. Kết quả cuối cùng được hiển thị trong Hình 7.

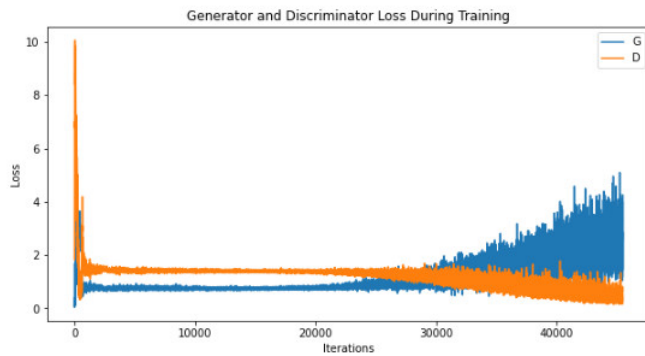


Hình 7. Đầu ra kết quả 64 x 64 sau 100 epoch và ảnh thật của DCGAN

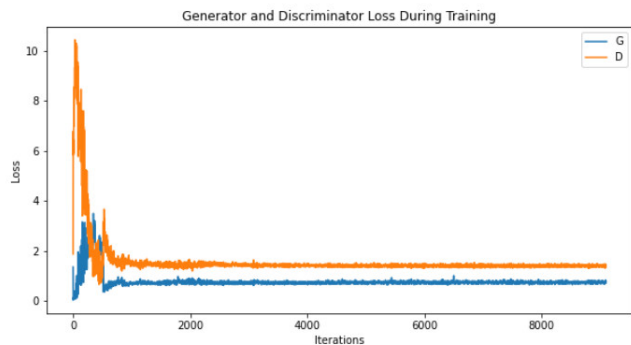
5.1.3. Kết quả định lượng



Hình 8. Loss của generator và discriminator sau 200 epoch



Hình 9. Loss của generator và discriminator sau 500 epoch



Hình 10. Loss của generator và discriminator sau 100 epoch

VI. KẾT LUẬN VÀ CÔNG VIỆC TƯƠNG LAI

Đối với việc tạo tác phẩm nghệ thuật, chúng tôi đã triển khai DCGAN theo nhiều số lượng epoch khác nhau và so sánh chúng.

Trong tương lai, có một số điều đáng để thử. Trước tiên, chúng tôi có thể triển khai WGAN, modified WGAN và so sánh với DCGAN. Triển khai DCGAN trên các bộ dữ liệu khác để xem liệu nó có còn mạnh mẽ trong việc cung cấp nhiều loại kết quả đầu ra hay không hoặc giải thích điều này về mặt lý thuyết. Cuối cùng nhưng không kém phần quan trọng, chúng ta có thể khám phá các loại khác cho GAN (có thể sử dụng tổng khoảng cách biến đổi, khoảng cách Hellinger hoặc khoảng cách Levy-Prokhorov) có thể dẫn đến kết quả thực nghiệm đầy hứa hẹn.

References

- [1] Gemini 2: The intelligent duplicate file finder. <https://macpaw.com/gemini>.
- [2] A tensorflow implementation of "deep convolutional generative adversarial networks". <https://github.com/carpedm20/DCGAN-tensorflow>.
- [3] Deep learning 13: Understanding generative adversarial network. <https://ireneli.eu/2016/11/17/>, 2016.
- [4] M. Akten. Source code and pretrained model for webcam pix2pix. <https://github.com/memo/webcam-pix2pix-tensorflow>, 2017.
- [5] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein gan. arXiv preprint arXiv:1701.07875, 2017.
- [6] M. G. Bellemare, I. Danihelka, W. Dabney, S. Mohamed, B. Lakshminarayanan, S. Hoyer, and R. Munos. The cramer distance as a solution to biased wasserstein gradients. arXiv preprint arXiv:1705.10743, 2017.
- [7] J. Canny. A computational approach to edge detection. IEEE Transactions on pattern analysis and machine intelligence, (6):679–698, 1986.
- [8] C. Clark and A. Storkey. Teaching deep convolutional neural networks to play go. arXiv preprint arXiv:1412.3409, 2014.
- [9] J. Gauthier. Conditional generative adversarial nets for convolutional face generation. Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester, 2014(5):2, 2014.
- [10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Advances in neural information processing systems, pages 2672–2680, 2014.
- [11] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. Improved training of wasserstein gans. arXiv preprint arXiv:1704.00028, 2017.
- [12] Z. He. Learning gan from fundamental principles to implementing a demo (in chinese). <https://zhuanlan.zhihu.com/p/24767059>, 2017.
- [13] H.-z. L. Hong Bao, Ye Liang and D. Xu. A dataset for research of traditional chinese painting and calligraphy images.

In International Proceedings of Computer Science and Information Technology, volume 42, pages 136–143. International Association of Computer Science and Information Technology Press, 2010.

[14] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Imagenet-to-image translation with conditional adversarial networks. arXiv preprint arXiv:1611.07004, 2016.

[15] S. Koo. Automatic colorization with deep convolutional generative adversarial networks.

[16] A. Krizhevsky and G. Hinton. Learning multiple layers of features from tiny images. 2009.

[17] Y. Lecun and C. Cortes. The MNIST database of handwritten digits.

[18] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. arXiv preprint arXiv:1609.04802, 2016.

[19] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley. Least squares generative adversarial networks. arXiv preprint arXiv:1611.04076, 2016.

[20] M. Mirza and S. Osindero. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014.

[21] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434, 2015.

[22] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee. Generative adversarial text to image synthesis. In Proceedings of The 33rd International Conference on Machine Learning, volume 3, 2016.

[23] J. M. Susskind, A. K. Anderson, and G. E. Hinton. The toronto face database. Department of Computer Science, University of Toronto, Toronto, ON, Canada, Tech. Rep, 3, 2010.

[24] W. R. Tan, C. S. Chan, H. Aguirre, and K. Tanaka. Artgan: Artwork synthesis with conditional categorical gans. arXiv preprint arXiv:1702.03410, 2017.

[25] D. Tran, R. Ranganath, and D. M. Blei. Deep and hierarchical implicit models. arXiv preprint arXiv:1702.08896, 2017.

[26] K. Ullah. How to download all full sized images from google image search without any software - 2 easy steps only. <https://32hertz.blogspot.com/2015/03/download-all-images-from-google-search.html>, 2015.

[27] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired imagenet-to-image translation using cycle-consistent adversarial networks. arXiv preprint arXiv:1703.10593, 2017.