Location Conditional Image Generation using Generative Adversarial Networks

Mayur Saxena
Department of Computer Science
University of Oxford
mayurajaysaxena@gmail.com

Aidan Gomez
Department of Computer Science
University of Oxford
aidan.gomez@sjc.ox.ac.uk

Yarin Gal
Department of Computer Science
University of Oxford
yarin@cs.ox.ac.uk

Abstract

Can an AI-artist instil the emotion of sense of place in its audience? Motivated by this thought, this paper presents our endeavours to make a GANs model learn the visual characteristics of locations to achieve creativity. The project's novelty lies in addressing the problem of the hardness of GANs training for an extremely diverse dataset in a contextual setting. The project explores GANs as an impressionist artist who adds its perspective to the artwork without hampering photo realism.

1 Problem Definition

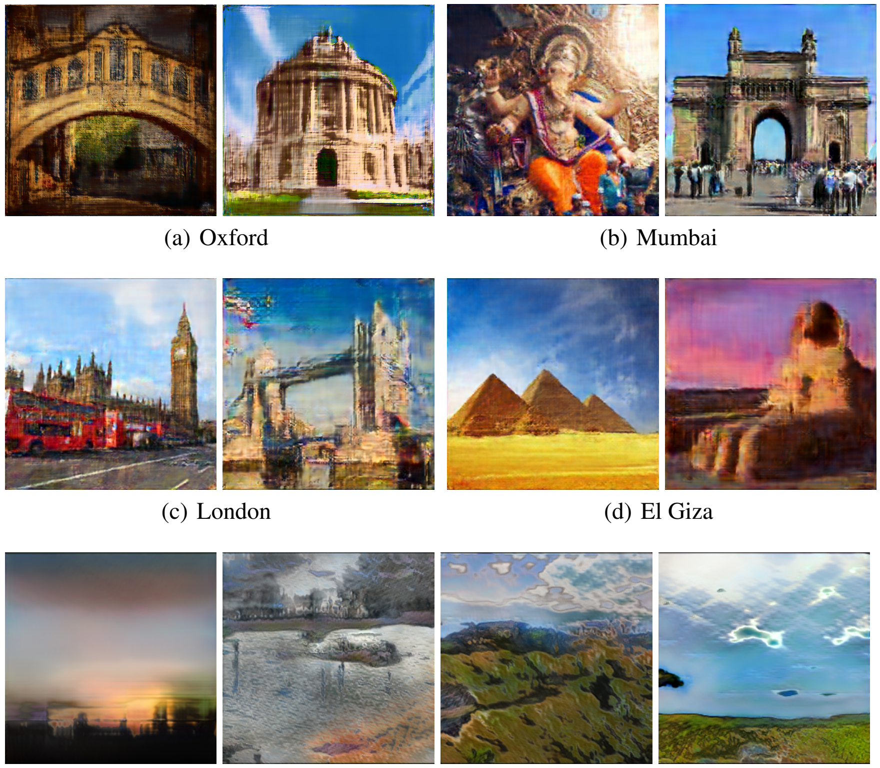
Formally, this project aimed at creating a GANs model that takes as input a city's geo-location coordinates and generate an image which represents the city, inspired by what people capture across social media. Since the conception of Generative Adversarial Networks[1], there have been substantial publications around its usage for image generation. But most of these only cater to datasets having some intrinsic repetitive pattern or style, such as MNIST[2], CelebA[3] or CIFAR-10[4] datasets. Since the features of the training data lie in the pixel space, a repetitive pattern is easier to learn as it can be imagined as a curve in this space. If we even consider images of one landmark posted over social media, there is huge variability in its images: camera angle, lighting condition, filters, etc. Now imagine the variability present in the images from cities around the world.

2 Experiments

We first curated a dataset of ∼8 million geo-location tagged images covering 13000 cities spread globally using Flickr's search APIs[5] and World Cities Database[6]. Next, we developed a plug and play framework which amalgamates the three GANs architectures: Progressive Growing GANs[7], Wasserstein GANs[8] and Conditional GANs[9]. Due to our doubts about the ease of training GANs for complex data distributions, we opted for an incremental approach to our explorations: starting with just one city, then 25 cities and finally for 13000 cities. Through our experiments, we captured the reaction of GANs towards different heuristics and learning techniques and used them to achieve stable learning for the complex data distribution. The resultant enhancement involves: Modelling GANs as a two-player game and defining its state of convergence as its pure/mixed Nash equilibrium(for model selection)[10], employing Wasserstein GANs, removing outliers, context normalization,& introducing Gaussian Noise only in the context.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

3 Results and Evaluation



(a) Oxford (b) Mumbai

(c) London (d) El Giza

(e) New York (f) Moscow (g) Zurich (h) Cancun

Figure 1: Generated Samples. It is worth noting that none of the latent vectors here were handpicked.

Fig. 1(a) presents images generated for the single city: Oxford. Fig. 1(b),1(c) & 1(d) showcase the samples created during 25 cities explorations. The uncanny resemblance of the generated samples to the actual landmarks makes it very interesting. What makes these experiments further alluring is the aspect of innovation or perspective added by the Generative Adversarial Networks in the generated images (Refer to Appendix A). Fig. 1(e), 1(f), 1(g) & 1(h) presents results achieved for data of 13000 cities. It is worth noting that almost all the generated images have a natural theme and no man-made structures, unlike the other two analysis. This is attributed to GANs' nature wherein it tries to gain information from a city's surroundings too. To validate the relevance of the generated samples, we employed Google Image Search[11] with location information. The top results featured realistic images which were visually similar to the generated sample. To further validate the photorealism of all generated images, we queried these images without any context. The interesting aspect about the results was that for some images realistic images were retrieved (proving photorealism) while for others, paintings (validating model's artistic power). To cater to subjectivity associated with result evaluation, we organized an exhibition featuring some of the samples[12]. Attended by 200+ visitors, the work received a very positive response from both technical and fine arts community.

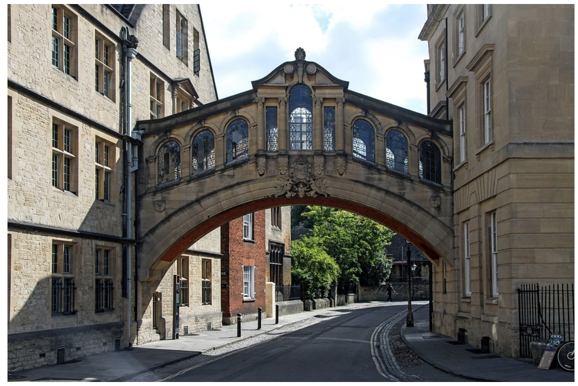
4 Conclusion

This project wasn't just another use-case of GANs-created art but also an attempt towards stable GANs training. The resultant Machine Learning model not only generated photo-realistic high definition images but was also successful in capturing the emotion of sense of place by understanding the location-based context. The generated images were not a mere replication or distortion of the original image/landmark. Instead the model also instilled its perspective into the samples and hence achieving creativity.

Appendix

A GANs' Perspective in Generated Images

(a) Generated Image



(b) Real Image

Figure 2: Bridge of Sighs, Oxford

Here we present an example on how the GANs model was adding its own perspective to the generated samples. Fig. 2(a) presents an image created during single city exploration on the city of Oxford. The image has an uncanny resemblance to the actual landmark named *Bridge of Sighs* (Refer to Fig. 2(b)). On comparison of the images of original Bridge of Sighs and the generated version, the first thing that pops out is the difference in the colour of the building. While the original has a very pale tone, the generated sample has a strong mustard-coloured tone. Also, the presence of a dark sky and sunlit tree in the generated image is nowhere close to reality. Hence the model is adding its perspective into the image to make it look and feel more like Oxford.

References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[2] Yann LeCun. The mnist database of handwritten digits. *http://yann. lecun. com/exdb/mnist/*, 1998.

[3] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Large-scale celebfaces attributes (celeba) dataset. *Retrieved August*, 15:2018, 2018.

[4] Cifar-10 and cifar-100 datasets. `https://www.cs.toronto.edu/~kriz/cifar.html`. (Accessed on 09/01/2019).

[5] Flickr services: Flickr api: flickr.photos.geo.photosforlocation. `https://www.flickr.com/services/api/flickr.photos.geo.photosForLocation.html`. (Accessed on 08/10/2019).

[6] World cities database | simplemaps.com. `https://simplemaps.com/data/world-cities`. (Accessed on 07/28/2019).

[7] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

[8] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017.

[9] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[10] John F Nash. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49, 1950.

[11] Google images. `https://www.google.com/imghp?hl=en`. (Accessed on 08/26/2019).

[12] Oxford: An ai-art view - dataseries - medium. `https://medium.com/dataseries/oxford-an-ai-art-view-6a49e34de387`. (Accessed on 09/01/2019).