A critical analysis on Generative Adversarial Nets

220432052
t.d.joloko@se22.qmul.ac.uk

***Abstract*— Generative adversarial networks (GANs) introduce a novel approach to generative modelling by introducing the concept of an adversarial component where two models are trained simultaneously, a generator network and a discriminator network. This method has gained high popularity over the years, with various versions developed to address different problems. This report critically analyses some key ideas, general trends, and challenges around the GAN framework.**

I. Introduction

Models that can learn a data distribution to generate new samples to match the learned distribution are called generative models. A Generative adversarial network (GAN) is one of such models introduced by Goodfellow et al. in 2014 that approximates a generative model by introducing an adversary [1]. In this framework, two models are simultaneously trained to compete against one another, a generative model, and a discriminative model (the adversary) [1]. The discriminator is trained to distinguish between the real data from the training set and fake data from the generator. In contrast, the generator is trained to generate new samples to fool the discriminator. This process has been likened to a two-player minimax game where each player tries to maximise their reward while minimising the reward of the other player [1].

The generator and discriminator are multilayer perceptrons [1], where the input to the generator is a noise vector while the input to the discriminator is either the real or fake data. Both networks are iteratively trained until a global optimum is achieved [1]. At this point, the discriminator can no longer differentiate between real or fake data and outputs a classification probability of 0.5. The training is done by backpropagating the gradients of the loss function to both networks, with the gradients going through the discriminator first before flowing to the generator [1]. Once the training is done, new samples can be generated by a simple feed-forward mechanism [1].

The following sections of this report a) analyse some of the key works and trends that have emerged following the introduction of the GAN approach for generative models and b) discuss some of the associated problems, improvements, and evaluation methods.

II. Keyworks and trends

Various extensions to the original paper [1] attempt to solve some of the known problems of GANs, such as instability [2], [3], [4] and expand the functionality to a broader scope [2], [5], [6], [7], [8]. Some models suggest improvements in the architecture's general structure [2], [5], [6], [7], [8] or specific model components, like the loss function [3] or gradient process [4]. The section will focus on the methods related to architectural changes, while the subsequent section will focus on the latter.

One of the improvements suggested by the original GAN paper was adding a conditioning factor to the generative model to control the data generation process [1]. Mirza et al. actualised this with Conditional Generative Adversarial Nets (CGAN) [5], which introduce a conditioning variable fed to the generator and discriminator input layer alongside the original inputs. This conditioning variable is highly dependent on the particular use case and can range from being the class labels to some other modality describing the data samples [5]. The LAPGAN (Laplacian Generative Adversarial Networks) [6] is one of such conditional models which merges a Laplacian Pyramid with a conditional GAN. In this case, there is a generator-discriminator network pair at each level of the pyramid, where the current level's conditioning variable is the output of the generator from the previous level. While this multi-scale approach to GANs produces higher-quality images than the original GAN and incorporates continuous refinement at each level [6], there is also much noise in the resulting output [2].

An innovative approach that combines the power of CNNs (Convolutional Neural Networks) with GANs is the DCGAN (Deep Convolutional Generative Adversarial Network) [2]. This model uses a modified CNN architecture to introduce a more stable way to train GANs and generate a higher-resolution image. The CNN used in DCGAN uses strided convolutions instead of max-pooling, ReLU, LeakyReLU and Tanh as the activation functions and batch normalisation, which helps the generator avoid the mode-collapse problem [2]. Using CNNs allows an even deeper understanding of the spatial representation of the images, thereby resulting in the generation of more meaningful samples. Some notable variations utilising CNNs are InfoGAN (Information Maximizing Generative Adversarial Networks) [7] and StyleGAN [8]. Both these methods force the generator to learn a disentangled representation that characterises different attributes of the data distribution. Examples are a) the width or thickness of hand-written digits and b) pose, identity or hairstyle for facial images. The InfoGAN achieves this by separating the original input noise into an incompressible noise and a latent code representing the attributes [7]. StyleGAN instead injects a latent code into each layer of the convolution [8].

III. Discussion

The GAN framework for generative modelling has experienced widespread success in the world of deep learning, with applications in areas such as image super-resolution [9], high-quality image generation [8], Information Retrieval [10] and Image-to-Image Translation [11]. However, one of the major problems encountered during the training process of GANs is the mode-collapse problem where the generator outputs samples that are too similar. Arjovsky et al. [3] proposed a variation to mitigate against this problem called WGAN. WGAN utilizes a new loss function based on the Earth-Mover (EM) Distance. In contrast to the Jensen-Shannon divergence with vanishing gradients used in the original literature [1], the EM distance is continuously differentiable and can be trained till the generator is able to learn an optimal distribution [3]. Kodali et al. [4] theorizes that this mode-collapse problem occurs due to the gradients crashing to an unwanted local equilibrium, and introduces a method called DRAGAN that uses the loss function from the original work but applies a penalty to regularize the gradient of the discriminator. Both WGAN and DRAGAN have been identified to improve the stability of training the networks.

Apart from the problem associated with training GANs, there is also the issue of evaluating them. Various methods for evaluating GANs have emerged over the years as there is no single unified or best method agreed upon, unlike in supervised learning where the accuracy of a discriminative model can easily be evaluated based on the correct predictions. For instance, earlier GANs like [1], [5], [6] use the Parzen window estimation which is a log-likelihood estimation, while some use the Inception Score [4], [12] which has been noted to be similar to perceptual evaluations. While these methods are good and provide an objective way to evaluate methods, manual evaluation may still be involved during some evaluation of the generated samples as some sort of cross-validation to the results of the quantitative measures. This is due to the intrinsic nature of generative models where we would like to visually analyse if the generated samples indeed match our perceived expectations. StyleGAN goes a step further to introduce a quantifying measure for latent space disentanglement called perceptual path length and linear separability [8]. Because the StyleGAN is based on the CNN architecture, these two measures might only be useful for GANs related to image, video, or other visual-related generations.

IV. Conclusion

This report has been able to provide a critical analysis of GANs by looking at some of the trends, key ideas, problems, and evaluation measures. There is still work to be done in a) stabilizing GANs as this is not perfected yet, and b) exploring the latent space that describes the learned factors of the generator. Understanding the latent space will provide deeper insight that allows for interpretable models and can lead to better developed models for generation at all scale and modalities, as well as understanding if there is any bias in the process of generation.

References

[1] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative Adversarial Nets. Advances in Neural Information Processing Systems. 2014;27.

[2] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434. 2015 Nov 19.

[3] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN. arXiv preprint arXiv arXiv:1701.07875. 2017 Dec 6.

[4] Kodali N, Abernethy J, Hays J, Kira Z. On convergence and stability of gans. arXiv preprint arXiv:1705.07215. 2017 May 19.

[5] Mirza M, Osindero S. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784. 2014 Nov 6.

[6] Denton EL, Chintala S, Fergus R. Deep generative image models using a laplacian pyramid of adversarial networks. Advances in neural information processing systems. 2015;28.

[7] Chen X, Duan Y, Houthooft R, Schulman J, Sutskever I, Abbeel P. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. Advances in neural information processing systems. 2016;29.

[8] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. InProceedings of the IEEE/CVF conference on computer vision and pattern recognition 2019 (pp. 4401-4410).

[9] Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Qiao Y, Change Loy C. Esrgan: Enhanced super-resolution generative adversarial networks. InProceedings of the European conference on computer vision (ECCV) workshops 2018 (pp. 0-0).

[10] Zhang W. Generative adversarial nets for information retrieval: Fundamentals and advances. InThe 41st International ACM SIGIR Conference on Research & Development in Information Retrieval 2018 Jun 27 (pp. 1375-1378).

[11] Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. InProceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 1125-1134).

[12] Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved techniques for training gans. Advances in neural information processing systems. 2016;29.