© 2018 IJSRSET | Volume 4 | Issue 2 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
National Conference on Advanced Research Trends in Information and
g Technologies (NCARTICT-2018), Department of IT, L. D. College of Engineering,
Ahmedabad, Gujarat, India In association with
International Journal of Scientific Research in Science, Engineering and Technology

Generative Adversarial Networks(GANs): Applications and Future Scope

Ketakee Nimavat, Tushar Champaneria
[1]UG student, Computer Engineering Department, L. D. C. E, Ahmedabad, Gujarat, India
ketakeenimavat91@gmail.com[1]
[2] Assistant Professor, Computer Engineering Department, L. D. C. E, Ahmedabad, Gujarat, India
tushar.1985@gmail.com[2]

ABSTRACT

Generative Adversarial Networks, GANs are the most recent and exciting development in the field of generative models. Inspired by game theory, the system consists of two models, one is a generator that generates the information and whose goal is to fool the discriminator who has been trained to differentiate between real and fake samples. In this paper, the architecture of the models is explored along with its ideology. Research on application of GANs have shown improvement in various possible areas over the traditional models. Especially areas pertaining to images such as image filling and text to image have shown promising results. However, because of GANs being relatively new, there remains a lot to be done. In this paper, we explore the existing models and the drawbacks or possible areas of improvement of these models. We finally discuss the open-ended research areas such as repetition, system equilibrium, finding appropriate metrics for evaluation and various others that would help make GANs more adaptable and would help achieve the potential they have for effectiveness in real-world scenarios such as realistic speech generation.

Keywords: GANs, Image Processing, Generative models, game theory

I. INTRODUCTION

Over the past few years, innumerable strides have been made in machine intelligence right from ELIZA to conquering jeopardy to beating the world's best player at chess and the most recent ones such as winning a game of GO and winning at DOTA. While we have excelled at various tasks of narrow domain, true human intelligence in terms of human characteristics such as empathy, creativity, spontaneity, context awareness and other such traits remain to be displayed yet. One of these characteristics, creativity, consists of generation of information, be it pictures or text or music. What makes us human is not the ability to calculate in our heads but the ability to understand the context and respond. Generative algorithms have often been seen as the measure of 'understanding' that a machine has of the training data. Since they generate after learning from a set of information, if they generate correctly, they must've understood the information correctly as well. Hence, generation is seen as a measure of understanding of the system. They are used in generating texts, images, speech and to generate system behaviours or system states. One such generative algorithm set is Generative Adversarial Networks or GANS.

Generative algorithms span a broad spectrum, from the Naïve Bayes to deep belief networks. The most recent addition to the existing toolset of Gaussian Mixture model, Hidden Markov models, Latent Dirichlet Allocation and Restricted Boltzmann machines are Generative Adversarial Networks or GANs. In 2014, Ian Goodfellow et al proposed a new model called Generative Adversarial Networks (GANs) in the paper presented at NIPS.[1] Since then, they have been hailed as the next step in machine learning. Yoshua Bengio has called GANs to be the most revolutionary technology in machine learning in the past decade.

GANs have often been favoured for unlike previous networks they fit to complex real-life data very well. For example, the quality of the images they produce and the closeness of the image to the real image is superior to that achieved using existing systems. What makes GANs different is that the system is based on an approach which is a combination of game theory and supervised learning. On the other hand, Deep Belief network, a popular neural network-based method for generation, is a complex one and is found to be difficult to scale due to its reliance on energy functions that involve large amounts of computations. GANs hence give us a new perspective to approach machine learning with. GANs also are a new concept because of their architecture which enables them to grasp the major characteristics of the system. The architecture is based on Game theory and is discussed in the next section.

In this paper, the working of GANs is discussed in section 2. Section 3 explores various research work wherein GANs are applied in different tasks and in Section 4, the areas of GANs that require more work are discussed.

II. THE WORKING OF GANS

The architecture consists of two networks: one generator network and one discriminator network. The generator generates data that the discriminator guesses as either real or fake. The goal of the generator is to fool the discriminator into thinking the item it produced is the real one. The goal of the discriminator is to not be fooled by fake generations. An analogy of a forger and an original painting can be used here, the forger first uses random colors and tries to create a piece similar to the painting. Then it keeps on editing the painting until it looks like the original painting but it isn't exactly the same. The goal of the forger is to fool the buyer into thinking that it is indeed an original painting. The buyer's job is to reject the painting if it doesn't look like the original one, the buyer is a trained expert and knows fake paintings from real ones. The generator is the forger and the buyer is the discriminator network.

Both the generator and the discriminator can be multilayer perceptron or neural networks. Both the networks have back propagation enabled to learn better. Since the discriminator is trained to differentiate between real and fake. The discriminator maps from input space to [0,1] and is a binary classifier.

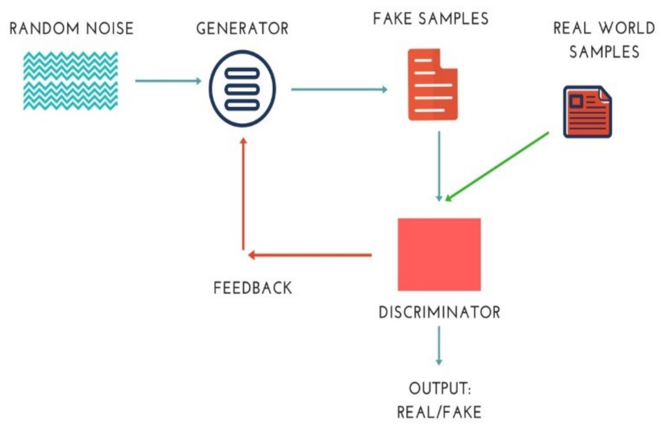

Figure 1 : GAN components and working

The generator generates from the random set with a pre-decided seed and from a predefined latent space. The key here is that the generator is trained after generating a sample not before. This way, the generator gets feedback on its generated input and can learn from it without wasting any chances of learning. The generated items are sent to a discriminator along with real samples and the discriminator either labels each image as real or fake. The feedback is then sent back to the generator network which learns via backpropagation.

This particular nature of two competing networks have made GANs be powerful systems, various research approaches have portrayed their efficiency. In the coming section, we discuss various possible uses of GANs and the research papers that have portrayed their use.

III. APPLICATION AREAS OF GANS

Since they were first proposed in 2014, they have been experimented with and used for various real-world scenarios. Variations of GANs involve using different types of networks and different combinations of networks. These variations are discussed below:

Image to image translation[2] (Conditional GAN(cGAN))

This approach uses a conditional GAN for solving image translation problems such as translation of an image from a Black and white space to an RGB space. CGANs prove to be a successful approach as compared to other existing approaches and also successfully maintain the details in the picture.

Another approach using a similar cGAN finds that cGANs are able to generate faces without overfitting the training dataset. It also proves that providing a conditional matrix to GAN does indeed improve performance.

Image generation [3] (Recurrent GAN)

This approach uses a recurrent GAN to show that an iterative convolutional neural net when trained to update features on a single canvas can generate images in general and not just images based on training images.

Enhancing resolution of images [4] (SRGAN (GAN for image Super Resolution))

When producing high-resolution images from low-resolution ones, GANs have seen to surpass traditional methods. In particular, they have been successful at maintaining the details even at a scale which other methods would blur out.

Text to image synthesis [5], [6][7]: (DCGAN, STACKGAN)

One approach uses a DCGAN which is conditioned on word embedding to generate images but finds that the images are recognizable and possess parts of the description when GANs are used with interpolation.

In another approach called StackGAN, two stacked GANs have been used to generate images. The first layer consists of the generation of basic features and components based on the description and the second layer of GAN then edits the image for resolution and details to make it more accurate. The approach has seen to generate more photo realistic details at a higher resolution with more accuracy and more diversity than at lower resolutions.

Image filling [8]

Adobe created a recent application using GANs wherein the user draws a shape and the system fills in the possible textures and colors. Such applications would help introduce more functions and enhance current functions in image editing software.

The applications of GANs are impressive but not without drawbacks. The drawbacks, however, present opportunities for making GANs more flexible and adaptable. In the next section, they are discussed with an aim to provide researchers with a direction to work in and for implementers to be aware of the shortcomings of the system.

IV. OPEN-ENDED RESEARCH AREAS IN GANS

In the end, however, it mustn't be forgotten that these are mathematical approaches to a generation that hope to arrive at a solution. Owing to this nature, there are some fundamental flaws such as drawing four eyes instead of two, lack of perspective and lack of 3D understanding.

This happens mainly because the system doesn't know the boundaries within which it can experiment. The system attempts to find which change is best suitable and closest to the original but in doing so it has no constraints other than the measures of similarity. It doesn't know the properties of the very object it is changing. This can lead to discrepancies such as a cat with two faces or a cow with two bodies. In image related applications, as the dimensionality of the picture increases so does the scope of error. For such a scenario, perhaps finding a way to store information about boundaries of experimentation will help train the system faster. What this tells us is that the system does indeed learn the general representation of an object but fails to capture the characteristics of the object.

Mode Collapse: An open-ended research area is the tendency of the system to collapse, also called mode collapse. The system is a combination of two interacting entities. Feedback and input from each are pivotal to the survival of the other. In this situation, they might go on forever or failure of one might cause overall failure.

Equilibrium: From the system's perspective, the system is designed to minimize the cost function of each the generator and the discriminator but it isn't designed to reach Nash Equilibrium which causes the system to keep going without reaching anywhere. Hence, the convergence of the system is not guaranteed. Hence another open-ended research area is algorithms that help a system reach equilibrium.

Repetition: GANs often fall into a trap wherein when one image is accepted by the discriminator i.e. manages to fool the discriminator, they either create the same image again and again or give minor variations of the image as output. This limits the diversity in the output and introduces inflexibility in the system. One proposed solution to this issue by training the data in mini-batches. This has shown to reduce redundancy. [9]

Evaluation metrics: As of now, there is a lack of proper evaluation metrics for GANs. The ones being used are nearest neighbor algorithm for measuring the distance between the produced image/sample and the training set.

Another proposed approach is battling two GANs to see which would win. [3]

Hence, GANs are in their nascent stage and hold a lot of potential after implementation. Various improvements such as training in mini batches have been proposed in [9]. Solving the above-mentioned problems will help make GANs more accessible for various other fields of implementation such as speech generation and real-time image generation applications such as image editors and video editors.

V. CONCLUSION

GANs produce quality outputs and provide a novel way to create generations similar to real-world scenarios. At their current stage, there is a lot that can be done in GANs in areas such as ensuring the variety of output and finding suitable ways to measure and rate the output. Research is needed to make GANs more controllable and mouldable. Apart from that, various problems such as stability and variety in the output are areas where work can be done to provide a more usable result. GANs will be useful in image applications, spontaneous and realistic speech generation, cross-field applications such as an image to text, image description, image captioning etc. Image editing softwares could benefit largely from GANs. It is hoped that future work will bring a step closer in our attempt to embed human qualities such as creativity in the intelligence we build.

VI. REFERENCES

[1] I. Goodfellow et al., "Generative Adversarial Nets," Adv. Neural Inf. Process. Syst. 27, pp. 2672–2680, 2014.

[2] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 2016.

[3] D. J. Im, C. D. Kim, H. Jiang, and R. Memisevic, "Generating images with recurrent adversarial networks," 2016.

[4] C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," 2016.

[5] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative Adversarial Text to Image Synthesis," Icml, pp. 1060–1069, 2016.

[6] H. Zhang et al., "StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks," 2016.

[7] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," pp. 1–16, 2015.

[8] J. Y. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros, "Generative visual manipulation on the natural image manifold," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9909 LNCS, pp. 597–613, 2016.

[9] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved Techniques for Training GANs," pp. 1–10, 2016.