International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 11 | Nov 2020 www.irjet.net p-ISSN: 2395-0072

Applications of GAN

Aarya Vakharia[1], Achal Shah[2], Shrey Mehta[3], Pavan Doshi[4]

[1-4]*Student, Computer Science and Engineering, Indus University, Gujarat, India*

---***---

Abstract- *In the past six years since GAN's were first introduced in 2014 by Ian Goodfellow, there have been radical improvements and changes in the efficiency and applications of GAN's. GAN's have time and again proved their capabilities in various fields including computer vision and natural language processing. In the following paper, we discuss the various applications of GAN's and how they can be employed in computer vision as well as other tasks. We talk about how they can be applied in different conditions and debate about the advantages of the same.*

Keywords- **Generative Adversarial Networks, Computer Vision, Domain Translation, Applications of GAN's**

1. INTRODUCTION

A GAN is a neural network that is used to generate different outputs from a given dataset. It works on the basis of opposition or conflict (i.e Adversarial) and tries to better itself by a feedback loop. A GAN consists of two separate neural networks; namely a generator (G) and a discriminator (D). These two neural networks try to "Outsmart" each other and thus better themselves. [1]

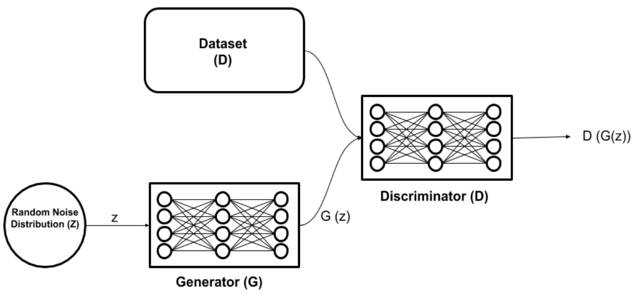

Fig. 1 Structure of a Generative Adversarial Networks

A. Generator

The generator is used to create an output from a random set of inputs. The inputs are taken in the form of an array known as a **random distribution (Z).** This (i.e random distribution), when fed into the generator, produces an output. [1]

B. Discriminator

The discriminators' task is to distinguish the real data from the fake (generated) data. The discriminator is fed data from the generator as well as the **dataset(D).** The discriminator outputs a 1 (real) or a 0 (fake) and checks to see if the result is correct or not. It then proceeds to give feedback to itself and the generator and the cycle continues. [1]

Both methods can be applied using a GAN. GANS can further be categorized into 2 sections:

C. Supervised Translation

Here the input consists of ground-truth image pairs on which the output is based upon. The disadvantage of this technique is that the ground truth images pair datasets are difficult to create. [14]

D. Unsupervised Translation

Instead of having pairs of corresponding images in the dataset, we have two datasets without any correspondence between them. We make certain assumptions that can make correspondences between different images in different domains. [14]

Generative Adversarial Networks can be implemented in multiple leading languages such as Java and Python. GAN's can be implemented in java using the Deeplearning4java (DL4J) library. Deeplearning4java includes functions such as ActivationLayer() which applies the activation function to the neural network. Python uses libraries such as TensorFlow and PyTorch.

2. APPLICATIONS

GAN's have a wide range of applications. They are used prominently to generate images. There are mainly 2 types of image synthesis applications which are text-to-image and image-to-image translation. Below listed are the major applications of GAN's.

A. Text to Image Generation [2]

The main purpose of Text-to-Image Generation is to generate images from text and also allow users to make changes in an image using natural language descriptions in one framework. It creates a high-quality image effortlessly by writing its description in natural languages. It works with different visual attributes and allows parts of the synthetic image to be manipulated accurately. This can be achieved using ControlGAN.

Uses:

- Using Text-to-Image Generation can edit videos as well.

- It is also used in making Games.
- Using it we can also change the color of objects according to our requirements.

Fig. 2 Output for text the bird has a yellow back and rump, gray outer rectrices, and a light gray breast.

B. Neural Style Transfer [3]

The basic principle behind the network is to transfer the style of the *style image* to the *content image*. Prior to neural nets, there have been similar attempts to transfer a style of an image to another image. But most face a similar problem i.e a single algorithm can only be used to transfer a particular style. Neural nets can be used in correspondence with Convolutional Neural Networks (CNN) to overcome the problem faced by prior algorithms.

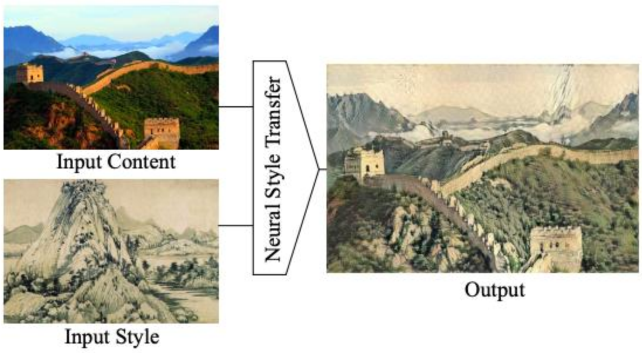

Fig. 3 Example of NST Algorithm to transfer the style of a Chinese painting onto a given photograph

Neural Nets can be categorized mainly into 2 sub-genres:

Image-optimization-based Online Neural Methods (IOB-NST) - The following network's basic idea is to extract the style and content information from the Style input image and recombine it with the content image. It then iteratively reconstructs the image until it matches the target. A major drawback of this method is that it is computationally expensive.

Model-optimization-based Offline Neural Methods (MOB-NST) - In this network the model is pre-trained onto a large data set. This provides an advantage over the prior network i.e it can be trained over multiple data sets involving multiple artistic styles.

C. Face Editing GAN with User's Sketch and Color [4]

It consists of training data and network architecture. Network architecture has a generator and a discriminator. The image is computed with all the necessary equations and mathematical formulae. The below example demonstrates face image editing and restoration.

Image editing system based on end to end trainable GAN with novel GAN loss is presented. The system is trained on high-resolution imagery based on celebA-HQ dataset and shows a variety of successful and realistic image editing in many cases

Fig. 4 Face image editing and restoration corresponding to user input

D. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs [5]

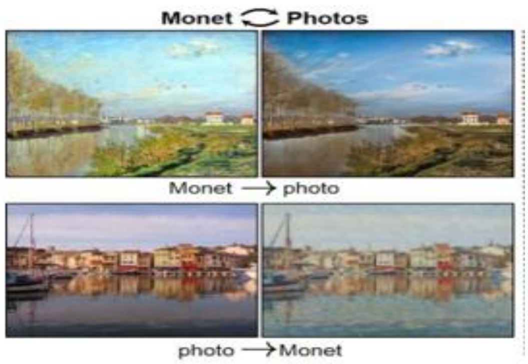

Fig. 5 Example of semantic manipulation from a monet to a photo and vice versa.

Using this feature, we can generate high-resolution photorealistic images from semantic label maps using conditional generative adversarial networks (conditional GANs). Using semantic segmentation methods, we can transform images into a semantic label domain, edit the objects in the label domain, and then transform them back to the image domain. This method also gives us new tools for higher-level image editing.

1) Image-to-Image translation framework: In this, the goal is to learn the mapping between an input image and an output image.

2) pix2pix method: In this, there are two main pieces, the Generator, and the Discriminator. The Generator transforms the input image to get the output image.

Uses:

- Using this we can do object manipulations such as removing/adding objects and changing the object category.
- We can allow users to edit the object as per requirements.
- We can use it to create synthetic training data for training visual recognition algorithms.
- We can change color, background, texture, etc in the existing image.

E. GANimation [6]

The paper talks about a novel application of GAN aptly named GANimation. It is the process of animating the facial expression from a single input image. The GAN is built on the EmotioNet dataset which provides Action Units (AU) for facial expressions. These can then be manipulated to form different facial expressions.The algorithm works by first creating the desired expression from a single image input, and then rendered-back to the original pose of the input image.

Advantages over other networks: It works much more efficiently in low light conditions and with a busy background. Unlike other networks the GANimation network learns the anatomy of the face and thus can manipulate it to make accurate facial expressions in multiple different conditions.

The model reaches its limits when dealing with extreme facial expressions, non-human inputs and when trying to use it on animal expressions. This mostly is due to insufficient training data.

F. Faces: from photos to emojis [7]

The problem of transferring a sample in one domain to an analog sample in another domain is discussed. The method is applied to visual domains including digits and face images and demonstrates its ability to generate convincing novel images of previously unseen entities while preserving their identity. As the main application challenge, the problem of emoji generation for a given facial image is tackled.

For face images, sets of one million random images without identity information are used. The set consists of assorted facial avatars (emoji) created by an online service (bitmoji.com). The emoji images were processed by a fully automatic process that localizes, based on a set of heuristics, the center of the irides, and the tip of the nose.

Fig.6 Unsupervised translation from photos to emojis.

G. 3D Generative-Adversarial Modeling [8]

The task here is to learn the latent space of objects and then recreate them in 3D space. This method uses unsupervised learning to create high-resolution 3D models, unlike supervised learning methods.The generator establishes a mapping from a low-dimensional space to a 3D space, this alleviates the need for a reference image or a CAD model. A powerful 3D shape descriptor is provided by adversarial discriminator which, when learned without supervision, has wide applications in 3D object recognition. Deep learning is also possible from 3d data.To further build on to the 3D-GAN model, the new proposed 3D-VAE-GAN adds a new encoder which can be used to create a latent vector *z* from a 2D image. This vector can then be used to render a 3D object.On further evaluation the 3D-GAN model was able to produce higher resolution images compared to previous state-of-the-art models.

The 3D-GAN model was trained on ShapeNet which contained shapes such as chairs, sofas, tables, boats, airplanes, rifles, and cars. The model was later tested on the ModelNet dataset to produce the results.

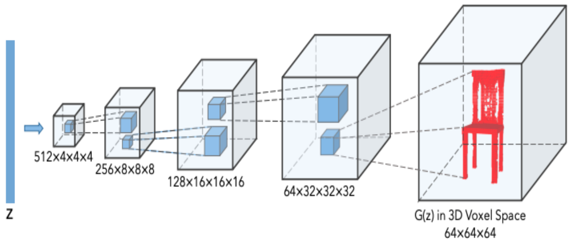

Fig.7 Architecture of 3D-GAN

3. MERITS AND DEMERITS

GAN's have an upper hand compared to many similar supervised and unsupervised networks. The benefits of using GAN's are as follows.

A. Merits

1) GAN's are a good way to classify data: The trained discriminator can be used to classify objects according to the training. [9]

2) GAN's can be trained on minimal data: unlike other networks such as CNN's which are data-hungry [9]

3) GAN's are a good way to classify using semi-supervised learning methods: It consists of a dataset that contains a minute amount of labeled data and a large amount of unlabeled data. [10]

4) GAN's understand the internal representation of the data: and therefore can be used with unlabeled datasets.[11]

5) GAN produces data indistinguishable from real data: which can be used in multiple real-world applications.[11]

B. Demerits

1) The generator suffers from a situation known as model collapse wherein the generator lacks the ability to produce different varieties of samples.[12]

2) If the discriminator becomes too good at its job it hinders the generator's ability to learn: Thus the GAN cannot produce the correct output. [12]

3) GAN's are less efficient at classifying data compared to CNN when datasets are large. [9]

4) GAN's can be affected by non-convergence. This may be due to parameter oscillation and destabilization. [12][13]

4. CONCLUSION

In the following review paper, we discuss the different applications of Generative Adversarial Networks (GAN). We discuss the way a GAN works, by using a Generator and Discriminator to produce realistic outputs. The paper also mentions the multiple implementation techniques and ways GAN's can be used in real-world applications. These include Text-to-Image translation, Image-to-Image translation, Image Synthesis and Semantic Manipulation etc. GAN's can be implemented in various languages such as Java and Python.

Each application builds on to the originally proposed model by Ian Goodfellow, and uses the basic functions used to build the GAN. The paper also mentions the advantages and limitations faced by GAN's.

ACKNOWLEDGEMENT

We would like to thank Prof. Pruthvi Patel for her expertise and assistance throughout all aspects of our study.

REFERENCES

1) https://www.youtube.com/user/kumarahlad/about
2) Controllable Text-to-Image Generation Bowen Li, Xiaojuan Qi, Thomas Lukasiewicz, Philip H. S. Torr University of Oxford
3) Neural Style Transfer: A Review Yongcheng Jing, Yezhou Yang, Member, IEEE, Zunlei Feng, Jingwen Ye, Yizhou Yu, Senior Member, IEEE, and Mingli Song, Senior Member, IEEE
4) SC-FEGAN: Face Editing Generative Adversarial Network with User's Sketch and Color, Youngjoo Jo Jongyoul Park ETRI, South Korea
5) High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs Ting-Chun Wang1 Ming-Yu Liu1 Jun-Yan Zhu2 Andrew Tao1 Jan Kautz1 Bryan Catanzaro1 1NVIDIA Corporation 2UC Berkeley Pumarola and A. Agudo and A.M. Martinez and A. Sanfeliu and F. Moreno-Noguer, GANimation: One-Shot Anatomically Consistent Facial Animation, International Journal of Computer Vision (IJCV), 2019
6) UNSUPERVISED CROSS-DOMAIN IMAGE GENERATION, Yaniv Taigman, Adam Polyak & Lior Wolf, Facebook AI Research, Tel-Aviv, Israel
7) Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling, Jiajun Wu* ,Chengkai Zhang*, Tianfan Xue, William T. Freeman, Joshua B. Tenenbaum
8) https://www.researchgate.net/post/Can_anyone_inform_me_please_about_the_advantages_and_limitations_of_Generative_Adversarial_Networks_GANs
9) Improved Techniques for Training GANs Tim Salimans Ian Goodfellow Wojciech Zaremba Vicki Cheung Alec Radford Xi Chen [10]

10) https://subscription.packtpub.com/book/big_data_and_business_intelligence/9781789136678/1/ch01lvl1sec14/advantages-of-gans
11) https://medium.com/@jonathan_hui/gan-why-it-is-so-hard-to-train-generative-advisory-networks-819a86b3750b
12) https://towardsdatascience.com/gan-ways-to-improve-gan-performance-acf37f9f59b
13) An Introduction to Image Synthesis with Generative Adversarial Nets, He Huang, Philip S. Yu and Changhu Wang