WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Survey on Synthesizing Images with Generative Adversial Networks

Jiby Mariya Jose, Shibu Kumar KB

Department of Computer Science and Engineering, Rajiv Gandhi Institute of Technology, Kerala 686501, INDIA.

A B S T R A C T

Over the past few years, there has been a extreme growth of research in Generative Adversarial Nets (GANs).Proposed in 2014 by Ian Good Fellow, GAN has been employed to a wide variety of applications such as computer vision and natural language processing, in which it attained a fascinating performance. Among the many applications of GAN, image synthesis is the most well-studied one, and research in this area has already showed the huge potential of implementing GAN in image synthesis. In this paper, we provide are view of some GANs employed in image synthesis.

Keywords: Computer Vision, Deep Learning, Generative Adversial Networks, Image Synthesis

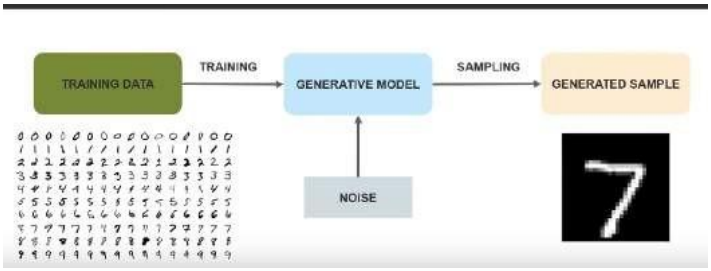
1. Introduction

Generative Adversial Networks [1], or GANs for short is a machine learning frameworks developed by Ian. Good fellow and his colleagues in 2014.GANs basically comprises a system of two neural networks which competes to analyze the differences within the dataset. For this purpose, GAN uses two neural network one is 'Generator' and the other one is the 'Discriminator'. Consider a scenario, In which there are Police (Let it be the Discriminator) and Counterfeiters (Let it be the Generatorhere), The mainaim of the Counterfeiter is to generate fake note sand circulate it in such a way that the police should not find it. On the other hand, The Police tries to find this fake notes by comparing between the real and the fake one to find the fake one out. This is the way how the Discriminator and the Generator works in a GAN. This system of Neural Network s have been a hot topic for research till now due to its wide applicability in industries such as cyber security, computer gaming, photography, computer vision and many more. This paper will mainly focus on the image synthesis application of GAN which comes under as a sub part of computer vision. The main goal of this paper to provide an overview of variety of technique s used in Image synthesis using the GAN and point out the pros and cons of each method. The GANs are classified into the into three main sub categories [2] for the ease of study, based o the techniques involved i.e. direct methods, hierarchical methods and iterative methods. And for each method there are GANs either performing Image Synthesis in text to image method or image to image method which will be discussed in the further section.The rest of this paper is organized as follows. In Section 2 we discuss some basic concepts of GAN, as well as some variants and training issues. Then in Section 3 we introduce three general approaches for image synthesis . In Section 4 we discussed some of the popular GANs which are highly used for image synthesis, while in Section 5 the conclusions are given.

* *Corresponding author.*
E-mail address: *jibyjose14@gmail.com*

2. Generative Adversial Networks

In this section, we review some core concepts of Generative Adversarial Nets (GANs) Before moving to in depth of the topic it is necessary to discuss



what the letter 'G', 'A' and 'N' exactly denote. The letter 'G' here stands for 'Generative'.

Fig 1 : Block of Generative model

A Generative model is a Unsupervised Learning approach, which has the ability to generate new samples from the dataset as shown in figure1.Now 'A' represents the setting at which the model is trained, here it is adversial. Hence 'A' is 'Adversial'. And 'N' indicates the 'Network' ie, The model is trained using the Neural Network. Hence, A GAN can be defined as a generative model trained under adversial conditions using Neural Network. A GAN uses two systems of neural network ie, Discriminator network and the Generator networks. The generator network takes the sample as the input and generates an output which is then fed into the discriminator network, The Discriminator network on the other hand is a probabilistic one. Based on the probability score attained, it decides whether the the sample is from the generator or the real sample. If the Discriminator fails to identify the generator generated image then the Generator wins and the iteration is stopped, else the iteration of this generator generating the image and discriminator continues as shown in the figure 2 below.The GAN has become a topic with wide applicability.They are widely used in generating images that look realistic, for generating variations in the facial expressions and for implementing many more editing techniques. Even more, the GANs are also used to tackle security concerns and also have application in generating data in those areas in which the data is least available It can also be used in autonomous driving, text generation and even in medical field such as in drug discovery ,medical imaging and many more. In short the application of GAN is a never ending list, it is the next step in deep learning evolution [12].Thebaseideaofthismodelisfoundontheindirecttrainingthroughthediscriminator, which is being varied dynamically. Unlike the discriminator, the generator is trained only to fool the discriminator. This enables the model to be an example of unsupervised manner. The generator is generally a de-convolutional neural network and the discriminator is implemented as a convolutional neural network.

3. Generative Adversial Networks

In this section, we discuss the three main approaches used in generating images, i.e. direct methods, iterative methods and hierarchical methods respectively, used for image synthesis.

A. Direct methods

In this approach GANs system of neural network uses only one generator and only one discriminator respectively. Many of the earliest models come in this category, like DCGAN [3], Improved GAN [4], InfoGAN [11] of which DCGAN is the most classic one

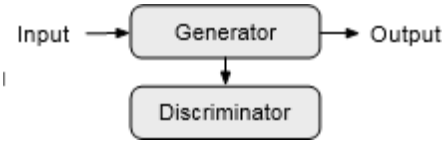


Fig 3 : Block of Direct method[2]

This is a simple design based on the complexity and implementation cost as compared to other two methods. Some of the most common combine all your researched in formation in form of a journal or research paper.In this researcher can take the reference of already accomplished work as a starting building block of its paper. Some of the GANs under this category are the following: DCGAN[3], This idea is derived from the paper titled as "Unsupervised representation learning with Deep Convolutional Generative Adversial Networks", authored by "Alec Rutherford et.al " in 2016.Here the author proposed

the idea of training the GAN store present the images.This idea emerged when the GAN using CNN to model images got failed. Hence here the author implemented this model with some small changes such as replacing the max-pooling layers with the strided convolutional layers, secondly, eliminating the fully connected layers and finally implementing the generator and discriminator using the Batch Normalization to stabilize the input . Here the author have used three datasets : Large- scale Scene Understanding (LSUN) (Yu et al., 2015),Imagenet-1k and a newly assembled Faces dataset respectively used for training And for testing and validation of the model CIFAR-10 is used.The results have shown that this model is much stable one for training the GAN to learn good representation of images. This model is not completely perfect, when trained for longer time period the model got collapsed and also this model is implemented only for the images in this paper which can we extended to video representation. Second GAN is Improved GAN[4],This paper "Improved Techniques for Training GANs" This paper, introduced several techniques intended to encourage convergence of the GANs game .These methods are adopted from the heuristic approaches of the non-convergence problem Existing GAN training have used gradient descent on each player's cost leading to problems in convergence. And hence this paper used Feature matching, one sided label smoothing, virtual batch normalization etc., which lead to heuristically motivate to encourage convergence. This paper discusses more stable training methods and also tries to figure out a evaluation metric i.e, the Inception score which gives a basis for comparing the quality of different GAN. Third GAN is InfoGAN[5],this paper "InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets" discusses the idea of implementing an 'information-theoretic extension to the Generative Adversarial Network that is able to learn disentangled representations in a completely unsupervised manner'[5],Here this author discusses about the changes that they made in the existing GAN to make it learn interpretable and meaningful representations by maximizing the information between the noise variables and the observations. The experimental results have shown that the InfoGAN successfully extracted writing styles from the number shapes from the MNIST dataset. It also discovered facial features for example–Presence or absence of glasses, facial emotions from the celeb A datasets. Experimental results have shown that the InfoGAN is highly successful in interpreting representations in images as compared to its other competitive models. All the above methods mainly focuses on Image-to- Image synthesis.

B. HierarchicalMethods

Unlike other methods, the Hierarchical methods[2] uses two generators and two discriminators respectively. The two generators uses two part sof the same image such as-foreground & background,styles & structures etc.,The relation between the two generators are either parallel or sequential as shown in figure 5 . Some of the major GANs under this category are: SS-GAN[6],This paper "Generative Image Modelling using Style and Structure Adversarial Networks" focuses on a model having two generators, One for generating a surface map from random noise and the second one is the Style-GAN that accepts two inputs namely- the surface map and noise. And the discriminator has also a Style-Discriminator which- accepts a single input formed by combining the surface map of aim age with its real image. A major drawback identified for SS-GAN is that it needs to use Kinect to fetch the ground truth for surface normal maps[6].The second GANisLR-GAN[7] introduced in the paper" Lr-gan: Layered recursive generative adversarial networks for image generation" discusses generation of the background and fore ground contents, which uses different generators but only one discriminator to evaluate the images. Experiments with LR-GAN demonstrated that it is possible to segregate the creation of background and foreground content and generate more sharp pictures.

C. Iterative Methods

This method is unique in itself from the previous methods in two ways. Firstly, rather than using two different generators that are capable to perform different functionalities, the model in this category use different generators performing same function as shown in the figure 4,and eventually due to which they generate fine images. Secondly, the structure is identical in this different GANs. This method can share weights among the generators which is not possible with other methods. Different GANs which can under this category are: LAPGAN [8],introduced in the paper" Deep generative image models using a laplacian pyramid of adversarial networks" focusses on the GAN that initially used this method to generate fine images from the coarse one using Laplacian pyramid [9].Different generators present in LAPGAN perform similar task ie., accepts an image from preceding generator and a noise vector as the input, and outputs the features(a residual image) that could create the picture more sharp when combined to the input . The dissimilarity present in those generators is the dimensions of input or output, while an anomaly is that the generator at the minimal level only takes a noise vector as the input and outputs an picture. LAPGAN surpass original GAN model and illustrates that iterative method can create more sharp pictures than the initial direct method. The second GAN under this category is StackGAN [10], introduced in paper "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks" is an iterative method, which uses only dual layers of generators. First generator get hold of an input and then outputs a fuzzy image that depicts a rough figure along with the fuzzy details of the objects present in it, while the other generator takes the input from the previous generator and same input which is fed to the first one and then produces an enlarged figure with more realistic features. This GAN involves the Text-To-Image synthesis. Synthesizing quality pictures from text is a challenging task in computer vision and many other applications. Images generated by the existing text to image approaches rarely depicts the meaning of the given description and also fails to define the necessary details in the image. Another example of GAN under this category is SGAN [11] described in the paper "Stacked generative adversarial networks," in which mound generators accepts minimal level details as the input and produces maximal level details,while the base generator accepts noise vector as the input whereas the peak generator outputs a picture. The need of utilizing distinct generators for non-identical levels is that SGAN is composed of an encoder and a decoder. Unlike other models, here the training set consists of images along with its associated labels. Here initially the image is encoded and then fed into the generator which is different from the models accepting the input. Experimental results have shown that this model much powerful on and able to create much realistic images from text.

4. Some other Miscellaneousgans

Under this category we introduce some most popular GANs based on the functions, Which are specially intended to image synthesis rather than the above GANs which have many applicability other than image synthesis. Some of the popular GANs are- PPGN [13] referred from the paper" Plug & play generative networks: Conditional iterative generation of images in latent space," generates attractive pictures in varied applications, such as class-conditioned image synthesis [14], text-to- image synthesis [15] , image in painting [16], and many more. This idea evolved when the author found that the DGN-AM[19],the previously existed model lacked diversity in the generated samples. Here in this paper the DGN-AM is combined with an denoising auto encoder (DAE) [17] which resulted in improving the sample quality and thereby the diversity with higher resolution than the previous models. This paper mainly focusses upon the synthesizing text to image.The main drawbacks identified in this paper is that it fails to produce high-quality pictures for certain data which are not present in the training sets. And also this model focusses more on learning the overall details of the image rather than the concepts associated with each of the objects in it. This idea here in this paper focused only in the image generation domain which can be generalized to many other types of data. Another example under this category is DualGAN[18] referred in the paper "Dualgan: Unsupervised dual learning for image- to-image translation" introduced a GAN that uses dual learning concept with cyclic mapping [20].In this paper this GAN is an example from unsupervised learning method which synthesizes image to image. The gap identified in this paper is that this method consumes a large amount of time for training as compared to other models. Experimental results have shown that this GAN model can highly improves the output as compared to other GAN which performs image to image translation tasks. Another example in this category is DistanceGAN[21] referred from the paper" Image-to-Image Translation with Distance Adversarial Generative Networks" discusses the idea of a GAN which is combined with an encoder.It is an unsupervised method which synthesizes image from image. The discriminator here calculates the difference by finding the distance between the encoded image and the generator generated image. Experimental results have shown that the model attained a good result than state-of-the-art methods across benchmark datasets: synthetic, MNIST, MNIST-1K,CelebA,STL-10datasets[22].

5. Conclusions

In this paper,we review some Generative Adversarial Nets(GAN)[1],that are mainly focused on the image synthesis. And also classified the image synthesis based on the general approaches into three types ie., direct, heirarchial and iterative, and also illustrated it based on the popularity of usage. Fortext-to-image synthesis, existing mechanism work well on datasets where each distinct pictures consist object such as CUB [57] and Oxford-102 [55], but the performa on complex datasets such as MSCOCO [61] is much worse'[2].This drawbacks probably came from the GANs incapacity to acquire knowledge of varying notion of things. For image-to-image translation, we review some of the general methods. This translation is definitely an fascinating implementation of GAN, which has substantial possibilities to be subsumedin to other software products, mostly mobile applications. Inspite of the fact that research in unsupervised method simply to be more famous, supervised methods may be more practicable as they still generate better synthetic images than the unsupervised one.

REFERENCES

[1] Goodfellow,J.Pouget-Abadie,M.Mirza,B.Xu,D.Warde-Farley,S.Ozair,A.Courville,andY.Bengio,"Generativeadversarialnets,"in *Neural Information Processing Systems*, pp. 2672– 2680, 2014.

[2] HeHuang,PhilipS.Yu,andChanguWang,"AnIntroductiontoImageSynthesiswithGenerativeAdversialNets" inarXiv:1803.04469v2 [cs.CV] ,pp2-3,2018

[3] A. Radford, L. Metz, and S. Chintala, "Unsupervised represen- tation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.B. Smith, "An approach to graphs of linear forms (Unpublished work style),"unpublished.

[4] T.Salimans, I.Goodfellow,W.Zaremba,V.Cheung, A.Radford,andX.Chen,"Improvedtechniquesfortraininggans,"in*Ad-vancesinNeural Information Processing Systems*, 2016, pp. 2226–2234.

[5] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever,andP. Abbeel, "Infogan: Interpretable representation learning by in- formation maximizing generative adversarial nets," in *Advances In Neural Information Processing Systems*, 2016, pp. 2172–2180.

[6] . X. Wang and A. Gupta, "Generative image modeling us- ing style and structure adversarial networks," *arXiv preprint arXiv:1603.05631*, 2016

[7] J. Yang, A. Kannan, D. Batra, and D. Parikh, "Lr-gan: Layered recursive generative adversarial networks for image generation," *arXiv preprint arXiv:1703.01560*,2017

[8] E.L.Denton,S.Chintala,a.szlam,andR.Fergus,"Deepgen-erativeimagemodelsusingalaplacianpyramidofadversarialnetworks,"in Advances in Neural Information Processing Systems28. Curran Associates, Inc., 2015, pp. 1486–1494

[9] P.J.BurtandE.H.Adelson,"Thelaplacianpyramidasacompactimagecode,"in *ReadingsinComputerVision*.Elsevier,1987,pp.671–679.

[10] H.Zhang,T.Xu,H.Li,S.Zhang,X.Huang,X.Wang,andD.Metaxas,"Stackgan:Texttophoto-realisticimagesynthesiswithstackedgenerative adversarial networks," *arXiv preprint arXiv:1612.03242*, 2016

[11] X. Huang, Y. Li, O. Poursaeed, J. Hopcroft, and S. Be- longie, "Stacked generative adversarial networks," *arXiv preprint arXiv:1612.04357*, 2016.

[12] Nicole Hemsoth.().A reality check on the future of GANs retrived from https://www.nextplatform.com/2019/01/30/a-reality-check-on-the- future-of-gans/amp/

[13] A. Nguyen, J. Yosinski, Y. Bengio, A. Dosovitskiy, and J. Clune, "Plug & play generative networks: Conditional iterative generation ofimages in latent space," *arXiv preprint arXiv:1612.00005*, 2016

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

[15] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," arXiv preprint arXiv:1605.05396, 2016

[16] R. A. Yeh, C. Chen, T. Lim, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with perceptual and contextual losses," CoRR, vol. abs/1607.07539, 2016. [Online]. Available:http://arxiv.org/abs/1607.07539

[17] P.Vincent,H.Larochelle,Y.Bengio,andP.-A.Manzagol,"Ex-tractingandcomposingrobustfeatureswithdenoisingautoencoders,"in *Proceedings of the 25th international conference on Ma- chine learning*. ACM, 2008, pp. 1096–1103.

[18] Z. Yi, H. Zhang, P. T. Gong *et al.*, "Dualgan: Unsupervised dual learning for image-to-image translation," *arXiv preprint arXiv:1704.02510*, 2017

[19] A Nguyen, D Alexey,Y Jason,"Syntheisizing the preferref inputs for neurons in neural networks via deep generator networks"arXiv:1605.09304v5 [cs.NE],2016

[20] G Arash,L Maria"A deep machine learning method for classifying cycic times series of biological signals using time growing neuralnetwork"

in IEEE Transactions on neural networks and learning system,2017

[21] S. Benaim and L. Wolf, "One-sided unsupervised domain map- ping," *arXiv preprint arXiv:1706.00826*,2017

[22] TNgoc,TAnh,"Dist-GAN:animprovedGANusingdistanceconstraints",inNTTran,ATBuietal.arXiv:1803.08887v3[cs.CV]15Dec2018