IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

IJARSCT
Impact Factor: 6.252

Volume 2, Issue 8, May 2022

Image Synthesis Using Generative Adversarial Network

Shrikant A. Shinde, Mahesh R. Kate, Kshitija T. Nalawade, Akhil A. Shinde, Mangesh B. Narwade,
Department of Computer Engineering
Sinhgad Intitute of Technology and Science, Pune, Maharashtra, India

Abstract: *Generative Adversarial Networks (GANs) are a deep learning based generative model. GANs are a model for training a generative model and it is common to use deep learning models. Generative Adversarial Network(GANs) are a powerful class of neural networks that are used for unsupervised learning. GANs are basically made up of two competing neural network models which compete with each other and are able to analyze, capture and copy the variations within dataset. GANs achieve high level realism by pairing a generator which learns to produce a target output with a discriminator which learns to distinguish true data from the output of the generator. GANs used for Image Synthesis generates high resolution images. Text to face generation is a sub domain of text to image synthesis, and it has a huge impact along with the wide range of applications on public safety domain. Our proposed fully trained GAN outperformed by generating the good quality images with accordance to the input sentence. StackGAN aims at generating high resolution photorealistic images. The Stage-I GAN sketches the primitive shape and colors of a scene based on a given text description, yielding low-resolution images. The Stage-II GAN takes Stage-I results and the text description as inputs, and generates high-resolution images with photo-realistic details.*

Keywords: Generative Adversarial Network (GAN), Stage-I GAN, Stage-II GAN

I. INTRODUCTION

Deep Learning is a branch of machine learning which is completely based on artificial neural network. Deep learning is a particular kind of machine learning that achieves great power and flexibility by learning to represent the world as a nested hierarchy of concepts, with each concept defined in relation to simpler concepts, and more abstract representations computed in terms of less abstract ones. Generative Adversarial Networks (GANs) were proposed by Goodfellow et al.. In the original setting, GANs are composed of a generator and a discriminator that are trained with competing goals. The generator is trained to generate samples towards the true data distribution while the discriminator is optimized to distinguish between real samples from the true data distribution and fake samples produced by the generator. GANs is be trained to generate high resolution image by breaking the difficult generative task into sub problems with the progressive goals. StackGAN (Stack Generative Adversarial Network) is trained to produce high resolution realistic images i.e. Stage-I GAN and Stage-II GAN. The text description will be given to Stage-I GAN which will produce image based of text description. Then the output from Stage-I GAN will be given to Stage-II GAN where the realistic and the image with more resolution the Stage-I will be generated. More generators will be trained for generating high resolution images and realistic image. Low-resolution images are first generated by our Stage-I GAN. On top of the Stage-I GAN, we stack Stage-II GAN to generate highresolution (e.g., 256_256) images. By conditioning on the Stage-I result and the text again, Stage-II GAN learns to capture the text information that is omitted by Stage-I GAN and draws more details.

II. RELATED WORK

Recently, Generative Adversarial Networks (GANs) have shown promising performance for generating sharper images. Generative image modeling is a fundamental problem in computer vision. There has been remarkable progress in this direction with the emergence of deep learning techniques. Variational Autoencoders (VAEs) formulate the problem with probabilistic graphical models with the goal of maximizing the lower bound of data likelihood. Autoregressive models (e.g., PixelRNN) that utilize neural networks to model the conditional distribution of the pixel space have also generated appealing synthetic images.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, May 2022
Impact Factor: 6.252

Built upon these generative models, conditional image generation has also been studied. Most methods utilize simple conditioning variables such as attributes or class labels . There are also works conditioned on images to generate images, including photo editing, domain transfer and super-resolution. Recently, several methods have been developed to generate images from unstructured text. [1]Mansimov et al. built an AlignDRAW model by learning to estimate alignment between text and the generating canvas. [2]Reed et al. used conditional PixelCNN to generate images using text descriptions and object location constraints.

[3] Nguyen et al. used an approximate Langevin sampling approach to generate images conditioned on text. Kususm Lata used Conditional GAN to translate image based on the given condition.[4] Yitang Men[5] used Attribute decomposed GAN for synthesizing controllable person image. Rahul Nehra[6], built Cycle Generative Adversarial Netwrok for synthesizing Radiology Image. Grigory Antipov built GAN model with Ranking-CNN for face aging.[7] Muhammad Zeeshan Khan used fully trained GAN in which encoder and decoder are trained to generate high quality images.

[8] Ji Lin1built Anycost GANs for Interactive Image Synthesis and Editing. Dong Nie[9], built Supervised Deep Convolutional Adversarial Network Residual Learning for the Generator for medical Image synthesis. Supervised deep convolutional model is proposed for estimating a target image from a source image via adversarial learning, even for the cases where the target and the source images belong to different modalities.

Ting Chung Wang built[10], Deep Visual Manipulation, Multi Scale Discriminator, Improved Adversarial loss for High-Resolution Image Synthesis and Semantic Manipulation. Yuusuke Kataoka[11], the attention mechanism is implemented by recurrent neural networks. Additionally, the Generative Adversarial Networks (GANs) approach has been proposed to generate more realistic images. For generator “U-Net”-based architecture and for discriminator convolutional “PatchGAN” classifier is used by Phillip Isola[12] for Image-Image to translation.

III. PROPOSED METHODOLOGY

3.1 Overview:

Generative Adversarial Networks (GANs) are composed of two models that are alternatively trained to compete with each other. The generator G is optimized to reproduce the true data distribution data by generating images that are difficult for the discriminator D to differentiate from real images. Fig. 1 shows the system architecture of the proposed system.

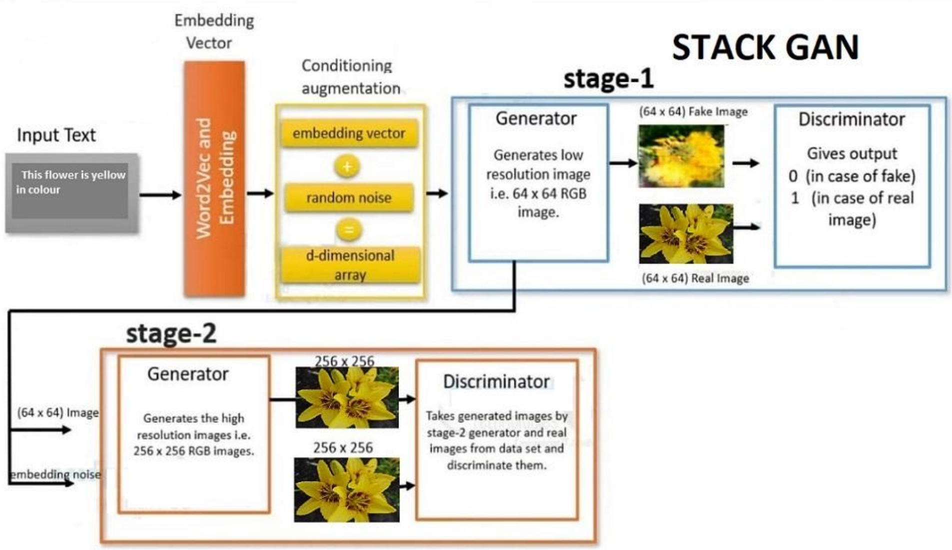

Figure 1: Proposed System Architecture

As shown in fig.1, the architecture is classified in two parts Stage-I GAN and Stage-II GAN each having Generator and Discriminator. The Generator is responsible for generation of images from the text input. The given Input sentence is splited with the help of conditioning augmentation and sent to generator for image generation.

3.2 Proposed System

The proposed system contains generator and discriminator where it generates the image from the image from the text input. Then it checks whether the generated image is true or false with the help of discriminator.

Algorithm1: Proposed Algorithm

1. Take input text from user.
2. Process the input with the help of embeddings for noise removal and pass to conditioning augmentation.
3. Pass the text to generator for image generation at Stage-I.
4. Pass the image generated from Stage-I to Stage-II for further image synthesis.
5. Check the generated image with the help of discriminator.
6. Display the generated high resolution image .

Algorithm 1 gives steps for proposed algorithm. The text input after processing is fed to Stage-I generator where the for image generation with basic shape and colour. Then this generated image from Stage-I is fed to Stage-II generator for further synthesis of image and generating high resolution image.

3.3 Mathematical Model

A mathematical model of proposed system is given below :

1. The training procedure is a min-max two player game, with following objective function:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}}[\log D(x)] + \mathbb{E}_{z\sim p_z}[\log(1 - D(G(z)))],$$

where x is a real image from the true data distribution pdata, and z is a noise vector.

2. Stage-I GAN trains the discriminator D0 and the generator G0 by alternatively maximizing LD0 and minimizing LG0 as given below

$$\mathcal{L}_{D_0} = \mathbb{E}_{(I_0,t)\sim p_{data}}[\log D_0(I_0, \varphi_t)] + \mathbb{E}_{z\sim p_z, t\sim p_{data}}[\log(1 - D_0(G_0(z, \hat{c}_0), \varphi_t))],$$

$$\mathcal{L}_{G_0} = \mathbb{E}_{z\sim p_z, t\sim p_{data}}[-\log D_0(G_0(z, \hat{c}_0), \varphi_t)] + \lambda D_{KL}(\mathcal{N}(\mu_0(\varphi_t), \Sigma_0(\varphi_t)) \,||\, \mathcal{N}(0, I)),$$

where the real image I0 and the text description t are from the true data distribution pdata. z is a noise vector.

3. Stage-II GAN are trained by alternatively maximizing LD and minimizing LG.

$$\mathcal{L}_D = \mathbb{E}_{(I,t)\sim p_{data}}[\log D(I, \varphi_t)] + \mathbb{E}_{s_0\sim p_{G_0}, t\sim p_{data}}[\log(1 - D(G(s_0, \hat{c}), \varphi_t))],$$

$$\mathcal{L}_G = \mathbb{E}_{s_0\sim p_{G_0}, t\sim p_{data}}[-\log D(G(s_0, \hat{c}), \varphi_t)] + \lambda D_{KL}(\mathcal{N}(\mu(\varphi_t), \Sigma(\varphi_t)) \,||\, \mathcal{N}(0, I)),$$

IV. EXPERIMENTAL SETUP

The result and and development is nothing but and Image synthesis product. To build an system we need an computer and Python IDE. Several packages are used which helps produce image from text input. The Stage-I Image generated will be stored and used for further synthesis at Stage-II Generator.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Volume 2, Issue 8, May 2022

Impact Factor: 6.252

V. RESULTS AND DISCUSSION

In the proposed method, the is generated from the given text input which is high resolution of 256*256 pixels. The Stage-I generator will generate image with basic shape and colour and will fed to Stage-II for further synthesis where high resolution image is generated.

Fig.2 shows the main login or sign up page for the users. It has attributes like name and password for already registered users. And also for new user sign up option has been added. It will be mandatory for the user to register themselves.

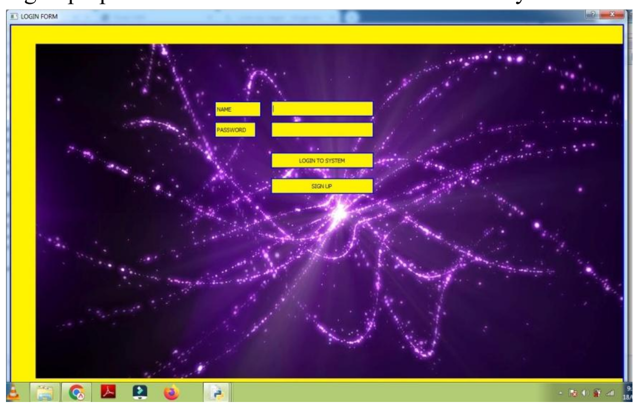

Fig. 2. Login/SignUp page

Following figures shows the testing, training and performance analysis of the model. The input text and the generated training images and the final image generated are show below.

```
'this flower is white and blue in color ,  with petals that are oval shaped . ',
'this flower has petals that are green with stringy purple stamen',
'this flower is blue and green in color ,  with petals that are oval shaped . ',
```

Fig. 3 Input text description

Fig. 4 Trained Images

Fig.5 Final Generated Images

DOI: 10.48175/IJARSCT-4542

VI. CONCLUSION AND FUTURE SCOPE

Stack Generative Adversarial Netwrok decomposes the problem of generating high resolution images into more manageable form. The Stage-I GAN with conditioning augmentation is proposed for text-to-image synthesis with basic colour colour and shape. Which further is fed to Stage-II GAN for generating high resolution image. In future work more dataset can be generated for image generation i.e. human face dataset which will generating sketches more efficiently and faster than traditional method of hand drawing .

VII. ACKNOWLEDGEMENT

The authors would like to thank the publishers, researchers for making their resources available and teachers for their guidance. We also thank the college authority for providing the required infrastructure and support. Finaly we would like to extend a heartfelt gratitude to friends and family members.

REFERENCES

[1]. E. Mansimov, E. Parisotto, L. J. Ba, and R. Salakhutdinov. Generating images from captions with attention. In ICLR, 2016.

[2]. S. Reed, A. van den Oord, N. Kalchbrenner, V. Bapst, M. Botvinick, and N. de Freitas. Generating interpretable images with controllable structure. Technical report, 2016.

[3]. Samaneh Azadi, Matthew Fisher, Vladimir Kim, Zhaowen Wang, Eli Shechtman, and Trevor Darrell. Multi-content GAN for few-shot font style transfer. In IEEE Conference on Computer Vision and Pattern Recognition, CVPR '18, pages 7564–7573, 2018.

[4]. Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, Honglak Lee. Generative Adversarial Text to Image Synthesis. 2016.

[5]. Yun Fu, Guodong Guo, and Thomas S Huang, "Age synthesis and estimation via faces: A survey," IEEE . 32, no. 11, pp. 1955–1976, 2010.

[6]. Rahul Nehra, Abhisikta Pal and B Baranidharan, "Radiological Image Synthesis Using Cycle-Consistent Generative Adversarial Netwrok". EasyChair Reprint No.5338.

[7]. Nilsback, Maria-Elena, and Andrew Zisserman. "Automated flower classifi- cation over a large number of classes." Computer Vision, Graphics & Image Processing, 2008. ICVGIP'08. Sixth Indian Conference on. IEEE, 2008.

[8]. Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems. 2014.

[9]. Zhang, Han, et al. "Stackgan++: Realistic image synthesis with stacked generative adversarial networks." arXiv preprint arXiv:1710.10916 (2017).

[10]. J.-Y. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros, "Generative visual manipulation on the natural image manifold," in European Conference on Computer Vision. pp. 597–613 Springer, 2016,

[11]. K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, "Unsupervised pixel-level domain adaptation with generative adversarial networks," in IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[12]. Brandon Amos, Bartosz Ludwiczuk, and Mahadev Satyanarayanan. OpenFace: A general-purpose face recognition library with mobile applications. Technical report, CMUCS- 16-118, CMU School of Computer Science, 2016.