First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Survey on Text to Image Synthesis

Chaitanya Ghadling[1], Firosh Vasudevan[1], Ruchin Dhama[1], Shreya Lad[1], Dr. Sunil Rathod[2]
[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India
[2]Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

One of the most difficult things for current Artificial Intelligence and Machine Learning systems to replicate is human creativity and imagination. Humans have the ability to create mental images of objects by just visualizing and having the general looks description of that particular object. In recent years with the evolution of GANs (Generative Adversarial Network) and its gaining popularity for being able to somewhat, replicate human creativity and imagination, research on generating high quality images from text description is boosted tremendously.

Through this research paper, we are trying to explore various GANs architectures to develop a model to generate plausible images of birds from detailed text descriptions with visual realism and semantic accuracy.

Keywords: GAN, AI, ML

I. INTRODUCTION

GAN (Generative Adversarial network) :

GANs consists of two components- Generator and Discriminator which are constantly in touch with each other working in tandem. The generator generates images and the discriminator then assess those images and provide feedback to generator about the correctness of the generated image in comparison with real images of the same object. The two neural networks constantly compete with each other to become more accurate in their predictions. The generator creates new images based on the feedback provided by the discriminator and the discriminator is trained by providing real images. The generator improves to fool the discriminator and the discriminator trains itself to not get fooled by the generator. The basic structure of GAN is shown in Fig. 1.

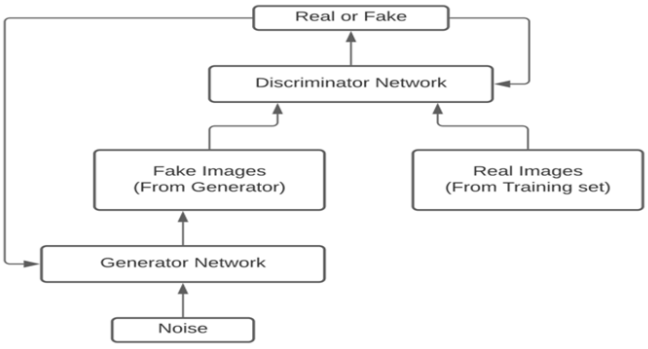

Fig. 1. Basic Structure of GAN

II. LITERATURE SURVEY

In 2014, Ian Goodfellow and his colleagues designed Generative Adversarial Network with the idea of

broadening scope of neural networks from just prediction and classification to allowing them to generate their own images. Though originally proposed as a form of generative model for unsupervised learning, GANs have also proven useful for semi-supervised learning, fully supervised learning and reinforcement learning. After various architectures developed to generate images by providing text description the quality of images along with semantic accuracy can be discussed from the Table-1.

Sr. No.	Paper Name	Advantages	Limitations
1.	Generative Adversarial Text to Image synthesis	1st major model for text to image synthesis	Lacks image quality. Does not work properly with different variety of datasets
2.	StackGAN++: Realistic Image synthesis with Stacked Generative Adversarial Networks	Improves the quality of image substantially	Difficult to train. Highly unstable and sensitive to hyper parameters.
3.	MirrorGAN: Learning Text to Image Generation by Redescription	Semantic consistency of image is highly improved.	Modules are not jointly optimized with complete end-to-end training.
4.	Learn, Imagine and Create: Text to Image Generation from prior knowledge.	Both visual realism and semantic accuracy is highly improved over baseline models.	Modules are not jointly optimized with complete end-to-end training.

Table-1 : Literature Survey

III. TAXONOMY CHART

Attributes Model	Image Quality	Semantic Accuracy	Inception Score (COCO dataset)	Inception Score (CUB dataset)
DC GAN	LOW	LOW	8.20	3.6
STACK GAN	MEDIUM	LOW	8.45	3.7
STACK GAN++	HIGH	MEDIUM	8.30	3.82
MIRROR GAN	MEDIUM	HIGH	26.47	4.56
LEICA GAN	MEDIUM	MEDIUM	20.42	4.62

Table-2: Taxonomy Chart

IV. CONCLUSION

Considering the results of the existing system, we will address the limitations on the quality of image with semantic accuracy and visual realism, by proposing a system that can improve both the parameter's and improve the overall inception score considerably. Also maintaining the consistency of output quality on other different non-standardized datasets will be one of the goals while developing a module.

V. ACKNOWLEDGEMENT

It gives us a great pleasure in presenting the paper on "A SURVEY ON TEXT TO IMAGE SYNTHESIS". We are really grateful to Dr Sunil Rathod for giving an opportunity to work with R&D cell of our department and providing us with necessary guidance with our project. We would like to take this opportunity to thank Dr. Pankaj Agarkar, Head of Computer Engineering Department, DYPSOE, Pune for providing us with an opportunity to present this paper. Our special thanks to Dr. Ashok Kasnale, Principal DYPSOE who motivated us and created a healthy environment for us to learn in the best possible way. We also thank all the staff members of our college for their support and guidance.

VI. REFERENCES

[1]. AttnGAN: Fine - grained Text to Image Generation with Attentional Generative Adversarial Networks.
[2]. StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks.
[3]. Generative Adversarial Text to Image Synthesis.
[4]. https://en.wikipedia.org/wiki/Generative_adversarial_network
[5]. https://paperswithcode.com/task/text-to-image-generation