Text to Image using Deep Learning

Article *in* International Journal of Engineering and Technical Research · March 2022

CITATION
1

READS
5,533

4 authors, including:

Sainath Patil
Vidyavardhini's College of Engineering And Technology
7 PUBLICATIONS **3** CITATIONS

SEE PROFILE

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 04, April-2021

Text to Image using Deep Learning

Akanksha Singh
Information Technology Vidyavardhini's
College of Engineering and Technology
Vasai, India

Ritika Shenoy
Information Technology Vidyavardhini's
College of Engineering and Technology
Vasai, India

Sonam Anekar
Information Technology Vidyavardhini's
College of Engineering and Technology
Vasai, India

Prof. Sainath Patil
Information Technology Vidyavardhini's
College of Engineering and Technology
Vasai, India

***Abstract*—Text to image synthesis refers to the method of generating images from the input text automatically. Deciphering data between picture and text is a major issue in artificial intelligence. Automatic image synthesis is highly beneficial in many ways. Generation of the image is one of the applications of conditional generative models. For generating images, GAN(Generative Adversarial Models) are used. Recent progress has been made using Generative Adversarial Networks (GAN). The conversion of the text to image is an extremely appropriate example of deep learning.**

Keywords—Generative Adversarial Networks (GANs), text-to-image synthesis, Generator, Discriminator.

I. INTRODUCTION

The generation of images from the regular language has numerous potential applications later on once the innovation is prepared for business applications. Generative Adversarial Networks have a place with the arrangement of generative models. It implies that they can create new substances. Text is translated into picture pixels. For eg: "Flower with pink petals." GAN comprise of an arrangement of two contending neural organization models that compete with one another and observe, catch and duplicate the varieties inside a dataset. Text to image synthesis is all about converting text descriptions into appropriate images. Nowadays, GAN models are widely used for better results. Also, there is one problem with deep learning is that there are many probable configurations for single text descriptions but it can be overcome by training the model.

A. Problem Statement

As it is difficult to understand the text by reading it and visualizing can become an issue. Also in some cases, there are words which can be wrongly interpreted. If text is represented in the image format it becomes a lot easier to acknowledge. Images are more attractive compared to text.Visual aids can deliver information more directly.Visual content grabs the attention and keeps people engaged. Key activities such as presentation, learning, and all involve visual Communication to some degree. If designed well, it offers numerous benefits.

B. Understanding Deep Learning

Deep learning is a subset of Artificial Intelligence which processes data for converting languages, recognizing objects by imitating the working of human brain. Deep learning has been evolved over the years and has brought huge amount of data which is easily accessible nowadays and mostly all the data are unstructured so it takes large amount of time for humans to extract relevant information but deep learning has resolved this issue so that it is easy to understand and process. Deep learning uses artificial neural networks which is meant to simulate the functioning of a human brain[4]. The hierarchical architecture of neural networks helps in processing the data across a series of layers. There are various neural network architecture such as Convolutional Neural Networks, Recurrent Neural Networks that are used widely. Hence, Artificial Intelligence is helping in transforming many scenarios[2].

II. METHOD

The deep learning technique that we have used is Generative Adversarial Network (GAN) which includes a generator and a discriminator. We have also used tensorflow, numpy, nltk, tensorlayer for generation of text to image. So basically tensorflow is a library of machine learning. It has a faster compilation time than other deep learning libraries. It also supports both CPU and GPU computing devices[5]. For text division means to convert bigger text into smaller parts like words we have used NLTK (Natural Language Toolkit) tokenizer. It helps the computer to analyze, pre process and understand the written text which is taken as an input from the user. Tensor layer which is a library built on top of tensor flow is used in training of the model to generate various layers like input layer, convolutional 2D layer, dense layer, etc. in the network for generator and discriminator. Python Pickle module is used for serializing the data means it converts the objects into byte so that it is easy to store in a file or to transport the data.

A. *Generative Adversarial Networks*

Generative Adversarial Network(GAN) is an approach for unsupervised learning which involves producing new examples. Generative Adversarial Networks uses neural networks in order to generate new instances of data. It can be used for image and voice generation. Generative adversarial networks refers to learn a generative model and training that model using neural networks[6]. GAN has two sub-parts (generator and discriminator):

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 04, April-2021

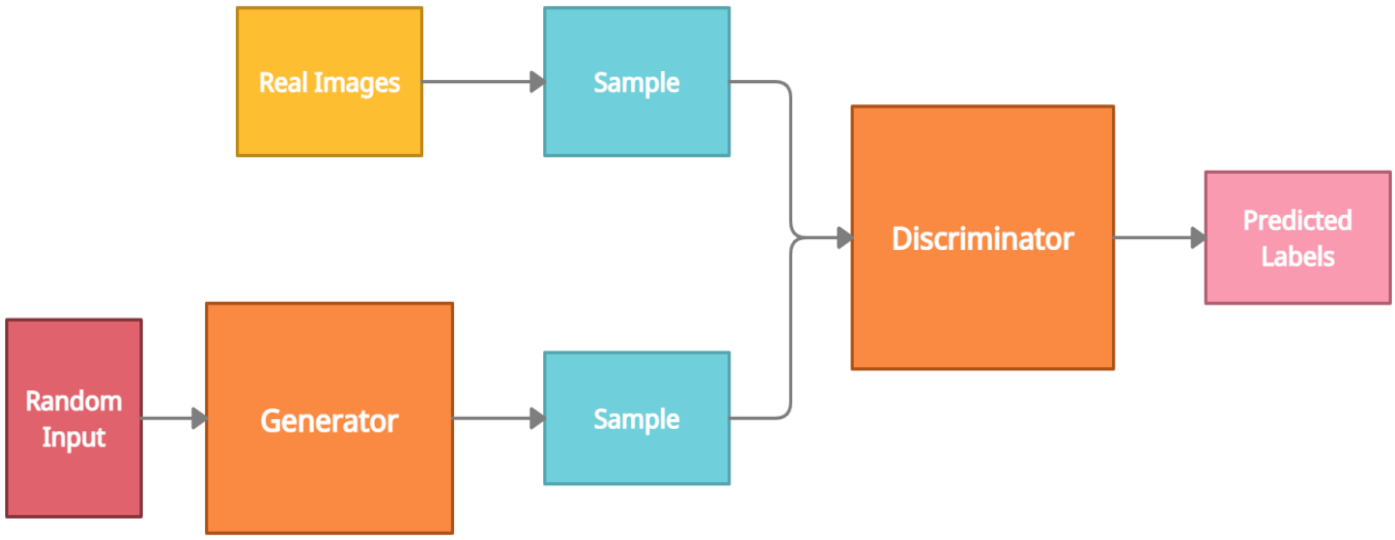

Fig. 1: Generative Adversarial Network Architecture

Generator:
It generates new instances of data which are mostly fake samples and passes it to Discriminator and tries to confuse the Discriminator.
Discriminator:
It distinguishes between real samples and fake samples generated from the Generator.

Generator and Discriminator are deep neural networks. The goal of Generator is to fool the Discriminator whereas the goal of Discriminator is to identify correct data. Generator and Discriminator both compete with each other. Generator makes all the attempts to convince the Discriminator that the generated fake instances are the real samples of data and also increases the probability of mistakes whereas the Discriminator figures out the real ones. Hence, these steps are repeated many times and both the sub-models get trained much better. First,Discriminator is trained on the real data samples to verify if it can identify those samples as real[1]. Again, the Discriminator is trained on generated fake data to see if is able to discriminate between actual and fake image. Generator is also trained depending upon the results of Discriminator so it can improves itself. Deep Convolutional GAN is extension of Generative Adversarial Network which is also widely used[3]. Here, Generator has to generate a vector for generating new data where vectors are made up of latent variables. GAN model takes large amount of time to train itself.

B. Algorithm

We have used GAN CLS algorithm for training the discriminator and generator.
GAN-CLS Training Algorithm:
1. Input - minibatch images, matching text
2. Encode matching text description
3. Encode mismatching text description
4. Draw random noise sample
5. Generator will pass it to Discriminator
6. The pairs will be:
{actual image, correct text}
{actual image, incorrect text}
{fake image, correct text}
7. Update discriminator
8. Update generator

According to the algorithm, as a generator will generate fake samples and pass it to Discriminator, there are three pairs of inputs will be provided to Discriminator[7]. Correct text with actual image, incorrect text with actual image and fake image with correct text out of which the pair of correct text and actual image is the most accurate output. These inputs are used to train the Discriminator more accurately.

C. Dataset

The dataset used is Oxford-102 flower dataset. Oxford-102 flowers include total 8,192 images of flowers of all different species. We have considered 8000 images for training the model and 189 flowers images for testing. Also, 10 captions are taken into consideration per image so that it is easy to train and is beneficial to produce accurate outputs.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 04, April-2021

Fig. 2: Dataset images

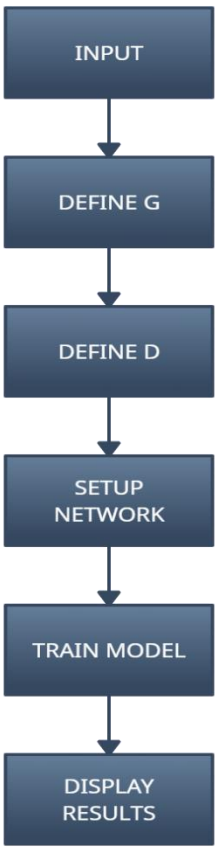

Fig. 3: Flowchart

This is the flowchart of the project which describes the actual flow wherein the first input is given to the generator and discriminator which is in the form of text for that generator and discriminator and various parameters such as image resolution are defined to set up a network. Based on this network, training of the model is done using the algorithm and the results are displayed.

D. Graphical User Interface

For Graphical User Interface(GUI), we have used PySimpleGUI, which is a python package. The theme is specially designed for designing beautiful windows to showcase the user creativity on the GUI with colors. It is easier to understand and implement. Adding a GUI, makes the project more interesting and approachable. Good GUI makes more interacting for the user to understand the process.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 04, April-2021

III. RESULTS

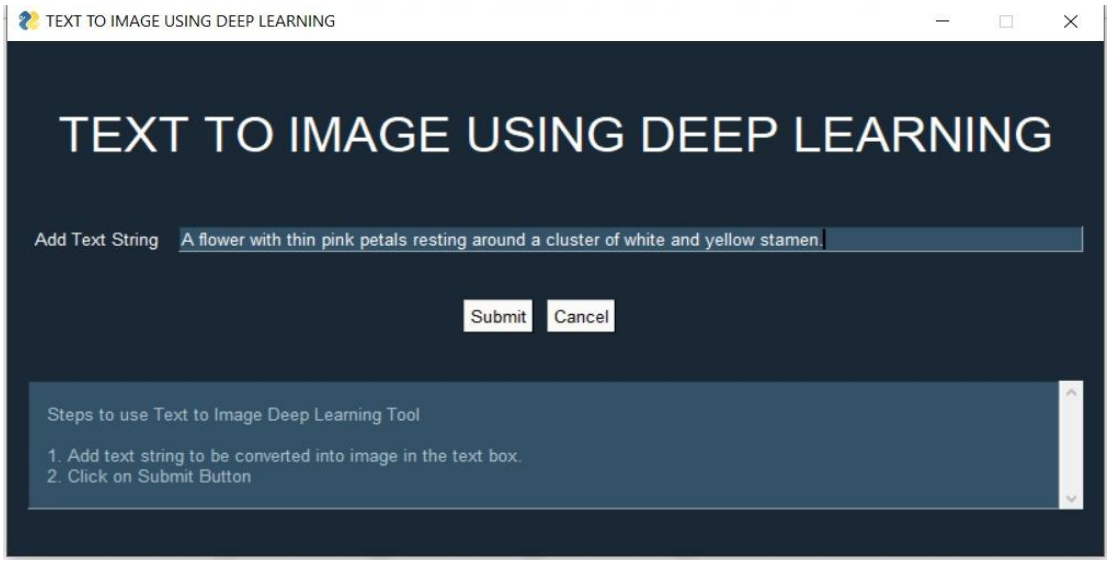

Fig. 4: GUI for taking input text

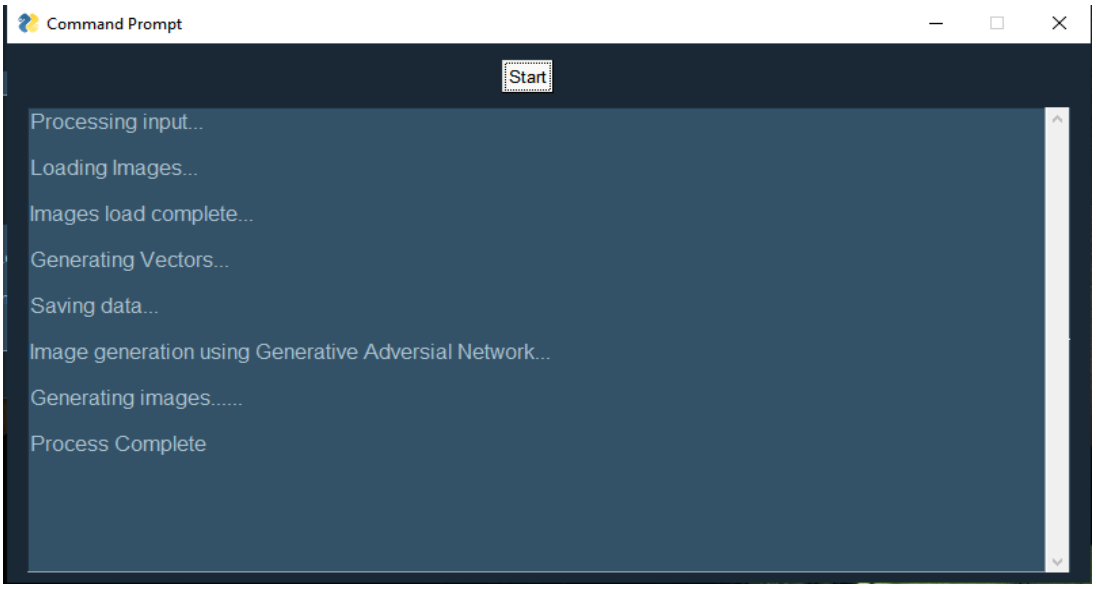

Fig .5: Steps of image generation

Fig. 6: Output

We used the GAN architeture and GAN-CLS algorithm with image text matching discriminator results on the Oxford-102 Flowers dataset. The basic GAN tends to possess the foremost variety in flower morphology. GUI takes the input text from the user and then processes that text for displaying the correct image. As a result, the image is displayed which matches the text description.

IV. CONCLUSION

After conducting a combined study of the papers and planning the project implementation, we developed an easy and efficient model for image generation. In future, we would like to improve the model so as to get pictures having a high resolution and will use this model on other dataset as well.

ACKNOWLEDGEMENT

We would like to acknowledge our guide Prof. Sainath Patil for his valued guidance and helpful discussions. We are also thankful to the Department of Information Technology (Vidyavardhini's College of Engineering and Technology) for their suggestions and support throughout the project.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 04, April-2021

REFERENCES

[1] Ankit Yadav1, Dinesh Kumar Vishwakarma2, Recent Developments in Generative Adversarial Networks: A Review (Workshop Paper),2020.

[2] Gregor, K., Danihelka, I., Graves, A., Rezende, D., and Wierstra, D. Draw: A recurrent neural network for image generation. In ICML, 2015.

[3] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, “StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks" in Rutgers University and Lehigh University August 2017.

[4] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, Honglak Lee , “Generative Adversarial Text to Image Synthesis" in University of Michigan and Max Planc Institute for Informatics June 2016.

[5] Stian Bodnar, Jon Shapiro, “Text to Image Synthesis Using Generative Adversarial Networks" in The University of Manchester May 2018.

[6] Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He. Attngan: Fine-grained text to image generation with attentional generative adversarial networks. CoRR, abs/1711.10485, 2017.

[7] Mehdi Mirza, Simon Osindero, Conditional Generative Adversarial Nets, 2014.