International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)

| ISSN: 2395-7852 | www.ijarasem.com | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 6, November 2020 |

Data Exploration with Adversarial Autoencoders using GAN

Aftab Kazi, Alisha Khatani, Mahira Majgaonkar

Department of Information Technology, Finolex Academy of Management and Technology, Ratnagiri, Maharashtra, India

ABSTRACT: The Data Exploration with Adversarial Autoencoders - Clustering with AAE text to image generation using generative adversarial networks (GAN) sis a deep learning model which can generate images from their corresponding text description. The image retrieval from input descriptions has been one of the most important applications with increasing numbers of videos on the internet. Although the text to image generation is a challenging work since semantic tasks between features of texts and videos exist. Conversion of information between text and image is a problem in artificial intelligence i.e. AI that connects natural language processing. So, we propose a system, that is a new retrieval framework based on the generative adversarial network (GAN) which will improve the performance by simple procedures. In the addition There are lot studies which are basically focused on generation of high-quality image. which semantically aligns with the given text description. Thorough experiments on two public benchmark datasets demonstrate the superiority of GAN over other representative state-of-the-art methods. Addition to the latent vector, the encoder will also generate a *categorical vector* or so-called *one-hot encoded* vector, so a vector with one of its values being 1 and all other values being 0. This vector will also be fed into the decoder for reconstruction so that it will be likely to also carry relevant information about the input data, basically some sort of additional *hint* for the reconstruction. The idea is that this hint will be the same for *similar* input data, and these similarities are what we want the encoder to find and encode into categories.

KEYWORDS: Adversarial Autoencoder (AAE), Generative Adversarial Network (GAN), Convolutional Neural Network (CNN), Recurrent Neural Networks (RNN).

I. INTRODUCTION

In today's artificial intelligence world, image generation is evaluated progressively. For generating the images of 3d objects convolution neural networks (CNN) were used. Also, algorithms have been designed that can extract the information from images upon being queried, this visual question answering the machine was a state of art research in the field of image analysis. Conditional image generation could be achieved using pixel CNN architecture.
Changing over common language content depictions into pictures is a stunning show of Deep Learning. Content order errands, for example, opinion investigation have been fruitful with Deep Recurrent Neural Networks that can take in discriminative vector portrayals from content. In another space, Deep Convolution GANs can blend pictures, for example, insides of rooms from an arbitrary commotion vector examined from an ordinary appropriation. The focal point is to associate advances in Deep RNN, enlivened by the possibility of Conditional-GANs. Contingent GANs work by contributing a one-hot class mark vector as contribution to the generator and discriminator notwithstanding the arbitrarily inspected commotion vector. These outcomes in higher preparing strength, all the more outwardly engaging outcomes, just as controllable generator yields. The contrast between conventional Conditional-GANs and the Text-to-Image model introduced is in the molding input. Rather than attempting to develop a meager visual ascribe descriptor to condition GANs, the GANs are molded on a book implanting learned with a Deep Neural Network. Notwithstanding building great content embeddings, making an interpretation from content to pictures is exceptionally multi-modular. The term 'multi-modular' is a significant one Revised Manuscript Received on April 25, 2020. * Correspondence Author Kambhampati.Monica, CSE, V R Siddhartha Engineering College, Vijayawada, India. Email: monica.kambhampati@gmail.com Dr. Duvvada Rajeswara Rao, CSE, V R Siddhartha Engineering College, Vijayawada, India. Email: rajeshpitam@gmail.com to get comfortable with in Deep Learning research. This alludes to the way that there are a wide range of pictures of winged creatures which relate to the content depiction "feathered creature". Another model in discourse is that there are a wide range of accents and so forth that would bring about various sounds compared to the content "flying creature". Multi-modular learning is additionally present in picture subtitling, (picture to-content). In any case, this is enormously encouraged because of the successive structure of content to such an extent that the model can anticipate the following word adapted on the picture just as the newly anticipated words. Multi-modular learning is generally troublesome, yet is made a lot simpler with the headway of GANs (Generative Adversarial Networks), this system makes a versatile misfortune work which is appropriate for multi-modular undertakings, for example, content-to-picture.

| ISSN: 2395-7852 | www.ijarasem.com | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 6, November 2020 |

A. Aim of the project

The main aim of this project is to generate the image from text using generative adversarial network(GAN), which will help in many image processing sectors. The generation of images from natural language has many possible applications in the future once the technology is ready for commercial applications. that the autoencoder is trained, we can use it to generate class labels for all our data.

B. Objectives of the Project

The objectives of the systems development are:

- Attention-driven, multi stage refinement for fine-grained text to image generation using GAN.
- It will provide the deep attentional multi model and a complete semantic layout.
- The synthesization of different subregions of the image by paying attention to the relevant words in the natural language description.
- Data exploration after GAN encoder and clustering with different classes.

C. Scope of the Project

Generative Adversarial Networks abbreviated as GANs are evolutionary technology in the field of Deep Learning which can generate images from random noise.
Now, a day's people are very much interested in generation through machines like Image Generation, audio or video generation, etc. And these all are possible in this era of Deep Learning. Because of the advancement of Deep Learning technologies, we can generate anything we want. One good example is that: machines generate faces which do not exist yet this thing helps a lot in today's world. The scope of the proposed system is that it will generate a lot of images which also do not exist.

- image by paying attention to the relevant words in the natural language description.
- Data exploration after GAN encoder and clustering with different classes.

II. LITERATURE SURVEY

Title	Authors	Advantage	Disadvantage	Scope
MirrorGAN: Learning Text-to-image Generation by Redescription	Tingting Qiao, Jing Zhang, Dacheng Tao	Mirror GAN provide high-quality and visually realistic images	Guaranteeing semantic alignment of the generated image with the input text remains challenging	Able to generate fine scaled image upto a optimized computational resource
Object-driven Text-to-Image Synthesis via Adversarial Training	Wenbo Li, Pengchuan Zhang, Lei Zhang.	Obj-GANs that allow object-centered text-to-image synthesis for complex scenes, provide a two-step generation process.	While providing a semantic layout it takes the input as a sentence.	Able to refine the synthesized image with the help of object wise discriminator
Semantics Disentangling for Text-to-Image Generation	Guojun Yin, Bin Liu, Lu Sheng, Nenghai Yu	It provides a more natural and photo realistic image with high semantic layout	The model which is used here is very complex and poses a challenge at some stage while creating the realistic image.	Here the generated images can keep generation consistency under expression variants

StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks	Han Zhang1, Tao Xu, Hongsheng Li.	Provide a 256x256 high resolution image using stage 1and stage 2 GAN, and also provide a high level semantic layout.	The major problem is that it works on the whole sentence. and process only the specific words as a global vectors means as a sentence	Able to generate high resolution images(256 x 256) with more photo realistic details
Photographic Text-to-Image Synthesiswith a Hierarchically-nested Adversarial Network	Z. Zhang, Y. Xie, and L.Yang	A novel end-to-end method that can directly model high-resolution image statistics and generate photographic images.	In most cases,the image features lack in matching with text features	Regularize mid-level representations and assist generator training to capture the complex image statistics.

III. PROBLEM DEFINITION

A. Problem Statement

Text-to-image generation, i.e. generating an image given a text description, is a very challenging task. Translating an image from a given text description is an amazing thing in deep learning. Common text—conversion methods employ Generative Adversarial Networks to generate high-text-images, but the images produced do not always represent the meaning of the phrase provided to the model as input. Using a captioning network to caption generated images, we tackle this problem and exploit the gap between ground truth captions and generated captions to further enhance the network. Text-to-Image synthesis is a difficult problem with plenty of space for progress despite the current state-of - the-art results. Synthesized images from current methods give the described image a rough sketch but do not capture the true essence of what the text describes.

B. Existing System

In the existing system, first generate an initial image with rough shape and color, and then refine the initial image to a high-resolution one. The main problem in the existing system is, these methods heavily depend on the quality of initial images. If the initial image is not well initialized, the following processes can hardly refine the image to a satisfactory quality.

C. Disadvantages of Existing System

1. Depends on the quality of initial images.
2. Each word contributes a different level of importance.

D. Proposed System

A GAN is a generative model that is trained using two neural network models. One model is called the “generator” or “generative network” model that learns to generate new plausible samples. The other model is called the “discriminator” or “discriminative network” and learns to differentiate generated examples from real examples.

In our proposed system, we propose Generative Adversarial Network to generate high-quality images. The proposed method introduces a module to refine the image contents, when the initial images are not well generated. In this system we are using a GAN(generative adversarial network) that can synthesize fine-grained details at different parts of the image by paying attention to the relevant words in the natural language description. The main purpose of this project is to translate text to image using GAN.which produces better results than existing systems.

E. Advantages of Proposed System

1. Using this method we produce high quality images.
2. Helpful for photo editing.
3. Also we produce images for emojis.
4. This system is very helpful for different types of applications.

| ISSN: 2395-7852 | www.ijarasem.com | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 6, November 2020 |

F. Methodology

This project we will develop using python. For this system we are using Tensorflow which is an open-source library designed for high performance numerical and scientific computations. It is one of the most widely used libraries for machine learning research. TensorFlow offers both low level and high-level APIs which make development flexible and allow fast iteration. In our project first we will first collect the dataset for the developing system. After that we will train the dataset and generate a module for proposed output. We will give input as an image with text and showing the result image is real or fake.

IV. HARDWARE & SOFTWARE REQUIREMENTS

A. Hardware

1. Processor: Intel Core i5 or more.
2. RAM: 8GB or more.
3. Hard disk: 250 GB or more.

B. Software

1. Operating System : Windows 10, 7, 8.
2. Python.
3. Anaconda.
4. Spyder, Jupyter notebook
5. MYSQL

C. Technologies Used:

1. Python

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

Often, programmers fall in love with Python because of the increased productivity it provides. Since there is no compilation step, the edit-test-debug cycle is incredibly fast. Debugging Python programs is easy: a bug or bad input will never cause a segmentation fault. Instead, when the interpreter discovers an error, it raises an exception. When the program doesn't catch the exception, the interpreter prints a stack trace. A source level debugger allows inspection of local and global variables, evaluation of arbitrary expressions, setting breakpoints, stepping through the code a line at a time, and so on. The debugger is written in Python itself, testifying to Python's introspective power. On the other hand, often the quickest way to debug a program is to add a few print statements to the source: the fast edit-test-debug cycle makes this simple approach very effective.

2. MySQL

MySQL is well known as the world's most widely used open-source database (back-end). It is the most supportive database for PHP as PHP-MySQL is the most frequently used open-source scripting database pair. The user-interface which WAMP, LAMP and XAMPP servers provide for MySQL is easiest and reduces our work to a large extent.

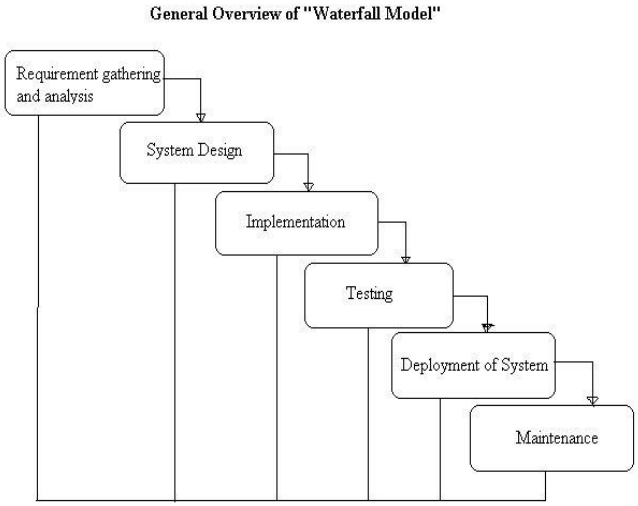
V. PLANNING & ESTIMATION

A. Software development Life Cycle

The entire project spanned for a duration of 6 months. In order to effectively design and develop a cost-effective model the Waterfall model was practiced.

| ISSN: 2395-7852 | www.ijarasem.com | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 6, November 2020 |



B. Requirement gathering and Analysis phase

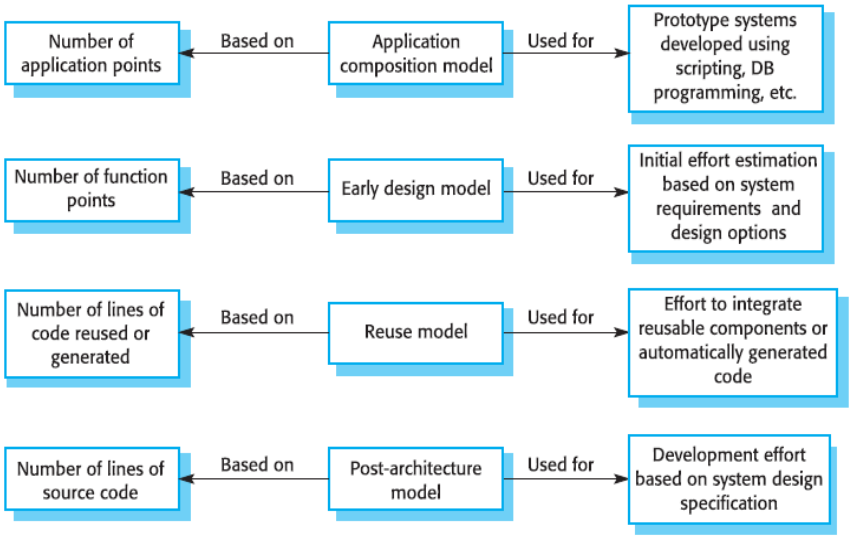
This phase started at the beginning of our project, we had formed groups and modularized the project. Important points of consideration were

1. Define and visualize all the objectives clearly.
2.Gather requirements and evaluate them

Consider the technical requirements needed and then collect technical specifications of various peripheral components (Hardware) required.

3. Analyze the coding languages needed for the project.
4. Define coding strategies.
5. Analyze future risks / problems.
6. Define strategies to avoid these risks else define alternate solutions to these risks.
7. Check financial feasibility.
8. Define Gantt charts and assign time span for each phase.

C. Cost Estimation



Now, a day's people are very much interested in generation through machines like Image Generation, audio or video generation, etc. And these all are possible in this era of Deep Learning. Because of the advancement of Deep Learning technologies, we can generate anything we want. One good example is that: machines generate faces which do not exist yet this thing helps a lot in today's world. The scope of the proposed system is that it will generate a lot of images which also do not exist.

| ISSN: 2395-7852 | www.ijarasem.com | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 6, November 2020 |

VI. FUTURE MODIFICATIONS

In future it is intended to improve the system developed by increasing the dataset to make it more accurate. update more cleaned dataset and check the accuracy of each algorithm, and use higher accuracy algorithms to improve the accuracy. In future we also make multiple applications like image to image translation, image to emojis, photo editing etc.

VII. CONCLUSION

In this project we are making the model of text to image translation which will help in multiple applications like photo editing, image to emoji translation. Our Proposed System helps to get fine grained text to image synthesis and generate high quality image through a multi-stage process. Additionally, we also compute the text-image matching loss for increasing the accuracy of the generator. Compared to existing text to image models, our model generates higher resolution images with more photo-realistic details.

REFERENCES

[1] Dong H, Zhang J, McIlwraith D, Guo Y. I2t2i: Learning text to image synthesis with textual data augmentation. In2017 IEEE International Conference on Image Processing (ICIP) 2017 Sep 17 (pp. 2015- 2019). IEEE.

[2] Muhammad A, Ahmad F, Martinez-Enriquez AM, Naseer M, Muhammad A, Ashraf M. Image to Multilingual Text Conversion for Literacy Education. In2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) 2018 Dec 17 (pp. 1328-1332). IEEE.

[3] Zhang C, Peng Y. Stacking vae and gan for context-aware text-toimage generation. In2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM) 2018 Sep 13 (pp. 1-5). IEEE.

[4] Rampurkar VV, Shah SK, Chhajed GJ, Biswas SK. An approach towards text detection from complex images using morphological techniques. In2018 2nd International. Conference on Inventive Systems and Control (ICISC) 2018 Jan 19 (pp. 969-973). IEEE.

[5] Yanagi R, Togo R, Ogawa T, Haseyama M. Query is GAN: Scene Retrieval With Attentional Text-to-Image Generative Adversarial Network. IEEE Access. 2019 Oct 14;7:153183-93.

[6] Xu T, Zhang P, Huang Q, Zhang H, Gan Z, Huang X, He X. Attngan: Fine-grained text to image generation with attentional generative adversarial networks. InProceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 1316-1324).

[7] Wang X, Gupta A. Generative image modeling using style and structure adversarial networks. InEuropean Conference on Computer Vision 2016 Oct 8 (pp. 318-335). Springer, Cham.

[8] Reed S, Akata Z, Yan X, Logeswaran L, Schiele B, Lee H. Generative adversarial text to image synthesis. arXiv preprint arXiv:1605.05396. 2016 May 17.

[9] Cha M, Gwon Y, Kung HT. Adversarial nets with perceptual losses for text-to-image synthesis. In2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP) 2017 Sep 25 (pp. 1-6). IEEE.

[10] Dai B, Fidler S, Urtasun R, Lin D. Towards diverse and natural image descriptions via a conditional gan. InProceedings of the IEEE International Conference on Computer Vision 2017 (pp. 2970-2979).