International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Deep Learning Based Text to Image Generation

G. Ajay[*1], Ch. Sai Teja[2], P. Baswaraj[3], V. Vasanth[4], Dr. G. Sreenivasulu[5]
[*1-4]B.Tech. Student, [5]Professor
CSE Department, JB Institute of Engineering and Technology, Hyderabad, India

ARTICLEINFO

Article History:
Accepted: 05 April 2023
Published: 23 April 2023

Publication Issue
Volume 10, Issue 2
March-April-2023

Page Number
623-628

ABSTRACT

Text-to-image generation is a method used for generating images related to given textual descriptions. It has a significant influence on many research areas as well as a diverse set of applications (e.g., photo-searching, photo-editing, art generation, computer-aided design, image re-construction, captioning, and portrait drawing). The most challenging task is to consistently produce realistic images according to given conditions. Existing algorithms for text-to-image generation create pictures that do not properly match the text. We considered this issue in our study and built a deep learning-based architecture for semantically consistent image generation: recurrent convolutional generative adversarial network (RC-GAN). RC-GAN successfully bridges the advancements in text and picture modelling, converting visual notions from words to pixels. The proposed model was trained on the Oxford-102 flowers dataset, and its performance was evaluated using an inception score and PSNR. The experimental results demonstrate that our model is capable of generating more realistic photos of flowers from given captions, with an inception score of 4.15 and a PSNR value of 30.12 dB, respectively. Generating images from natural language is one of the primary applications of conditional generative models. This project uses Generative Adversarial Networks (GANs) to generate an image given a text description. GANs are Deep Neural Networks that are generative models of data. Given a group of coaching data, GANs can learn to estimate the underlying probability distribution of the info. In this project, the model is trained on the Caltech birds dataset. Recent progress has been made using GANs.

Keywords: PSNR, GAN, Caltech birds dataset, NLP, CNN, RNN, CNN

I. INTRODUCTION

When people listen to or read a narrative, they quickly create pictures in their mind to visualize the content. Many cognitive functions, such as memorization, reasoning ability, and thinking, rely on visual mental imaging or "seeing with the mind's eye". Developing a technology that recognizes the connection between vision and words and can produce pictures that represent the meaning of written descriptions is a big step toward user intellectual ability.Image- processing techniques and applications of computer vision (CV)

have grown immensely in recent years from advances made possible by artificial intelligence and deep learning's success. One of these growing fields is text-to image generation. The term text-to- image (T2I) is the generation of visually realistic pictures from text inputs. T2I generation is the reverse process of image captioning, also known as image-to-text (I2T) generation,which is the generation of textual description from an input image. In T2I generation, themodel takes an input in the form of human written description and produces a RGB image that matches the description. T2I generation has been an important field of study due to its tremendous capability in multiple areas. Photo-searching, photo-editing, art generation, captioning, portrait drawing, industrial design, and image manipulation are some common applications of creating photo-realistic images from text. It consists of two neural networks: one for generating data and the other for classifying real/fake data. GANs are based on game theory for learning generative models. Its major purpose is to train a generator (G) to generate samples and a discriminator (D)to discern between true and false data. For generating better-quality realistic image, we performed text encoding using recurrent neural networks (RNN), and convolutional layers were used for image decoding. We developed recurrent convolution GAN (RC- GAN), a simple an effective framework for appealing to image synthesis from human written textual descriptions. The model was trained on the Oxford-102 Flowers Dataset and ensures the identity of the synthesized pictures.

II. LITERATURE SURVEY

An alternative to guided graphical models with hidden variables are indirect graphical models that have hidden variables admire the physicist machine, the Deep Boltzmann machine and their variants. The interaction all told states of random variables is noted within the model because the production of extraordinary potential functions normalized by world aggregates / integrations. This form (division function) and its gradient are all ennobling however terribly trivial examples, though they'll be calculable exploitation the Markov chain Monte Carlo (MCMC) methods. mix algorithms are a vital a part of learning a way to trust MCMC.Deep trust networks (DBNs) are hybrid models that consist of single single directed layers and serverly direct layers. Despite the rapid approximation layer- by level training standard, DBNs face computational problems associated with indirect and guiding models.Alternative criteria that don't predict or preclude log chance also are proposed, cherish score matching and noisedistinction estimation. The chance density discovered for those 2 should be nominative analytically as much as the standardisation constant. Note that during numerous interstring generative fashions with a couple of layers of latent variables (together with DBN and DBM), its now no longer even ability to get a traceable lognormal chance density. Some models, such as the Dinois machine Encoder and also the Contract machine Encoder, have learning rules similar to score matching that apply to TBM.

III. EXISTING SYSTEM

Deep learning uses neural networks to learn useful representations of features directly from data. For example, you can use a pretrained neural network to identify and remove artifacts like noise from images.

Image classification with deep learning most often involves convolutional neural networks, or CNNs. In CNNs, the nodes in the hidden layers don't always share their output with every node in the next layer (known as convolutional layers). Deep learning allows machines to identify and extract features from images.Text-to image generation works by feeding natural language processing (NLP). The two main deep learning architectures for text classification are Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). Convolutional Neural

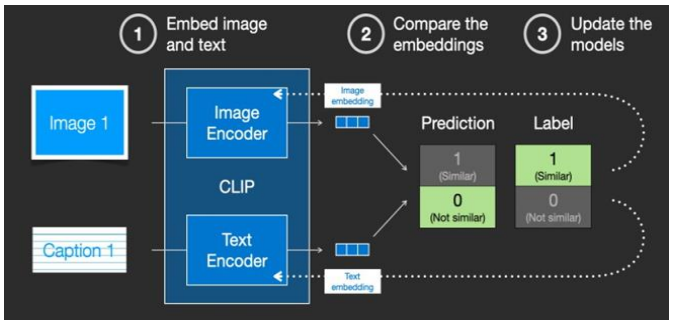
Networks (CNN) is a type of artificial neural network, which is widely used for image/object recognition and classification. Recurrent neural network (RNN) is a type of artificial neural network which uses sequential data or time series data.

IV. PROPOSED SYSTEM

The goal of this application is to develop a system which generates text to image using deep learning. Conditional GANs were used with recurrent neural networks (RNNs) and convolutional neural networks (CNNs) for generating meaningful images from a textual description. The dataset used consisted of images of materials and their relevant textual descriptions. Purpose of Proposed System:

1. Developing a image from textual descriptions.
2. Generating meaningful images accurately from input text.
3. Providing corrective and creative measures for the process.

BLOCK DIAGRAM



V. IMPLEMENTATION

Step 1 : Define project objectives and gather the resources.

(1) What are the services that we are going to provide using this project? (2) What are the conditions and restrictions used in this project? (3) What are the client needs and demands? (4) How many number of support staff do we need after developing the system

Step 2: Design the Deep learning based text to image generation System Architecture. Once the website objectives have been confirmed and communicated, there are multiple key factors that needed to be considered when designing the system architecture.

User Interface (UI) : Interface should be clear and easy to understand. Webpage must be eye catchy to attract users. With Django framework in addition we can built it with ease.

Backend: Python and javascript is the coding language we use in backend which connects to the server which helps to send the results obtained from the algorithm to frontend.

Database: Database stores information about diseases and users and profiles including their search and feedback.

Step 3 : Implementation – Developing code After we have clear idea on the architecture of the Deep learning based text to image generation system, we will start developing the code. Here, we have 3 stages to develop the code:

Stage 1 : Frontend -HTML and ReactJS Stage 2 : Backend - Python and Django Stage 3 : Database - MySQL

Step 4 : Deployment and Testing After completing the development of code we need to deploy the project in the compatible system which satisfies all the hardware and software requirement specifications. After deploying successfully, we need to check that the software meets the requirements and expectations of the end-users. Code should undergo testing processes like alpha, beta testing, unit testing, integration testing and functional testing to identify mistakes in code developed.

Step 5 : Maintain

The final step is risk analysis i.e. after testing process every unsuccessful step has some risk which disturbs

users accessibility. Our aim is maintain the website with support staff and to make servers available all the time and to analyze risks and update the code to avoid those risks. We need to fix some bugs and features in future as per user feedback. Website should be updated with new features in the future using renewing software models which increases project efficiency and productivity.

VI. SOFTWARE TOOLS

Operating system: windows 11 Coding language: Python3 Python IDLE, Vscode.
Modules: Tensorflow, Keras, Django.
The software for the development has been selected based on several factors such as Support

- ➢ Cost Effectiveness
- ➢ Development Speed
- ➢ Stability
- ➢ Accuracy

VII. FUNCTIONAL REQUIREMENTS

User shall upload the texts.
The system shall allow the user to upload texts.

The system shall generate the image and validate it.
The system shall provide the correct output
i.e. related to texts.
System shall allow user to see the desired images.

NON-FUNCTIONAL REQUIREMENTS

Scalability: System should be able to handle a large number of users. The system is capable enough to work properly.
Speed: The application should be fast. It should not slow down with the increase of number of users. Search functionality should be fast to enable better end-user experience. The system should be quick enough to be able to respond to user actions with a short period of time.
Usability: User interface should be simple and clear to break to understand to any user. At every step of this project user seems to be familiar with the interfaces as they are easy to use.
Availability: The system should be available at every moment to the user. It should be ensured that there should be minimum or no downtime to ensure better user experience for students.
Reliability: The system should be reliable and yield correct results if a user performs any actions. Also, if the user uploads a text, the system should ensure that the correct text is delivered to the correct destination without any loss of content.
Testability: The application is tested for validation, uploading texts, message structures and works fine.

VIII. ADVANTAGES

High-Quality Image Generation: Stable Diffusion can generate high-quality images that are visually consistent with the input text. The diffusion process encourages exploration of the entire image space, which allows for a diverse range of high-quality images.
Diverse Image Generation: Stable Diffusion can generate diverse images with different variations of the same textual description. This is because the diffusion process allows for exploration of different parts of the image space, leading to a diverse range of images.
Robust to Noise: Stable Diffusion is robust to Gaussian noise, which is added to the image during the diffusion process. This robustness helps to prevent the model from overfitting to the training data and can improve the generalization performance.

IX. DISADVANTAGES

Computationally Intensive: Stable Diffusion can be computationally intensive, especially for larger images

and complex datasets. The iterative nature of the diffusion process requires a large number of computations, which can increase the training time and resource requirements.

Lack of Interpretability: Stable Diffusion is a black-box model, which means that it is difficult to interpret how the model generates the images. This lack of interpretability can be a disadvantage in some applications where interpretability is critical.

X. APPLICATIONS

Product design: Text-to-image generation can be used to design and create product prototypes, allowing designers to visualize their ideas before investing in physical prototypes.

Interior decorating: Text-to-image generation can be used to generate images of interior spaces, allowing designers and clients to visualize different design options.

Fashion design: Text-to-image generation can be used to generate designs for clothing and accessories, allowing designers to quickly and easily create new designs.

XI. CONCLUSION

In conclusion, text-to-image generation is a challenging task that requires a deep understanding of both text and image data. Generative Adversarial Networks (GANs) have shown great promise for this task, and Stable Diffusion is a deep learning technique that has shown to be effective for generating high-quality and diverse images. Stable Diffusion involves applying a diffusion process to the image noise, gradually transforming the noise into an image. This process encourages exploration of the entire image space, leading to high-quality and diverse image generation. However, Stable Diffusion is computationally intensive and requires careful hyperparameter tuning. Despite these challenges, Stable Diffusion is a promising technique for text-to-image generation, offering high-quality and diverse image generation. Future research in this area can further optimize the technique for specific applications and datasets, and explore novel applications of Stable Diffusion in other domains such as natural language processing and speech synthesis.

XII.FUTURE ASPECTS

Future aspects of text to image generation project using deep learning include the integration of machine learning and artificial intelligence algorithms to improve efficiency and accuracy. The development of low-cost, portable, and user-friendly tools will reach the technology that involves human behaviour and human minds.

XIII. REFERENCES

[1]. T. Han, X. Zhang, Y. Xu, Y. Xu, and X. Tao, “Text-to-Image Generation via Stable Diffusion,” in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[2]. A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” arXiv preprint arXiv:1511.06434, 2015.

[3]. A. Brock, J. Donahue, and K. Simonyan, “Large scale GAN training for high fidelity natural image synthesis,” in International Conference on Learning Representations (ICLR), 2019.

[4]. M. Mirza and S. Osindero, “Conditional generative adversarial nets,” arXiv preprint arXiv:1411.1784, 2014.

[5]. D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in International Conference on Learning Representations (ICLR), 2015.

Cite this article as :

G. Ajay, Ch. Sai Teja, P. Baswaraj, V. Vasanth, Dr. G. Sreenivasulu, B.Tech. Student, Professor , "Deep Learning Based Text to Image Generation", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 10 Issue 2, pp. 623-628, March-April 2023.
Journal URL : https://ijsrset.com/IJSRSET23102105