iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VIII **Month of publication:** August 2021

DOI: https://doi.org/10.22214/ijraset.2021.37812

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

Image Synthesis Based On Feature Description

Rohan Bolusani
Student of Information Technology, Chaitanya Bharati Institute of Technology, India

Abstract: ***Generating realistic images from text is innovative and interesting, but modern-day machine learning models are still far from this goal. With research and development in the field of natural language processing, neural network architectures have been developed to learn discriminative text feature representations. Meanwhile, in the field of machine learning, generative adversarial networks (GANs) have begun to generate extremely accurate images of especially in categories, such as faces, album covers, and room interiors. In this work, the main goal is to develop a neural network to bridge these advances in text and image modelling, by essentially translating characters to pixels the project will demonstrate the capability of generative models by taking detailed text descriptions and generate plausible images.***
Keywords: Deep Learning, Computer Vision, NLP, Generative Adversarial Networks

I. INTRODUCTION

Generating images according to human provided natural language descriptions is a fundamental problem in many applications, such as art generation and computer-aided design. There are many applications to developing such generative models like generating cartoon/Anime Characters, generation of Realistic Photographs, generation of a human face using AI to catch criminals based on characteristics during investigations and generation of image datasets without having to take photos, to name a few. The main motivation of this research is to devise a generative neural network that maps words to images by using natural language processing and machine learning techniques.

II. DATASET AND IMPLEMENTATION

A. Dataset

The dataset consists of 102 flower categories. The flowers chosen while creating the dataset are the most commonly occurring flowers in the United Kingdom. Each category consists of between 40 and 258 images. The images have variations in large scale, pose and light. Additionally, there are categories that have large variations within the category and several very similar categories.

B. Implementation

The implementation can be seen in fig 4 and has the following steps:

1) *Sentence Encoding:* Sentence Encoding is done through Skip-thoughts. It is simple neural networks model that contains an encoder and decoder network and is frequently used for learning fixed length representations of sentences in any Natural Language without any labelled data or supervised learning, it uses the ordering of sentences from a natural language corpus that is provided as training data. The main reason behind using a skip thoughts model is, this model uses the order of sentences to self-supervise itself by using a recurrent neural network that remembers the previous sentence and this leads to a better reconstruction of the neighbouring sentence. The decoders are trained to minimise the reconstruction error of the previous and the next sentences given the embedding of these sentences, the errors are then back propagated to the encoder. The product of this exercise is a trained encoder that can be used to generate skip through vectors.
2) *Training:* The training starts by providing the encoded sentence and the images from the dataset as input to the generative adversarial network. The generative adversarial network essentially is made up of two neural networks, generator and a discriminator, the discriminator is trained to distinguish the real data instance from fake generated instance. Real data instances, such as real pictures of flowers are positive examples during training. Fake data instances created by the generator are negative examples during training. While the discriminator trains the generator is left untouched i.e., the weights do not change while it produces examples for the discriminator to train on. The generator training requires tighter integration between the two neural networks, once the generator creates a fake sample and sends as an input sample to the discriminator, based on the output from the discriminator, the generator is penalised and this generator loss backpropagated to improve the generator.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

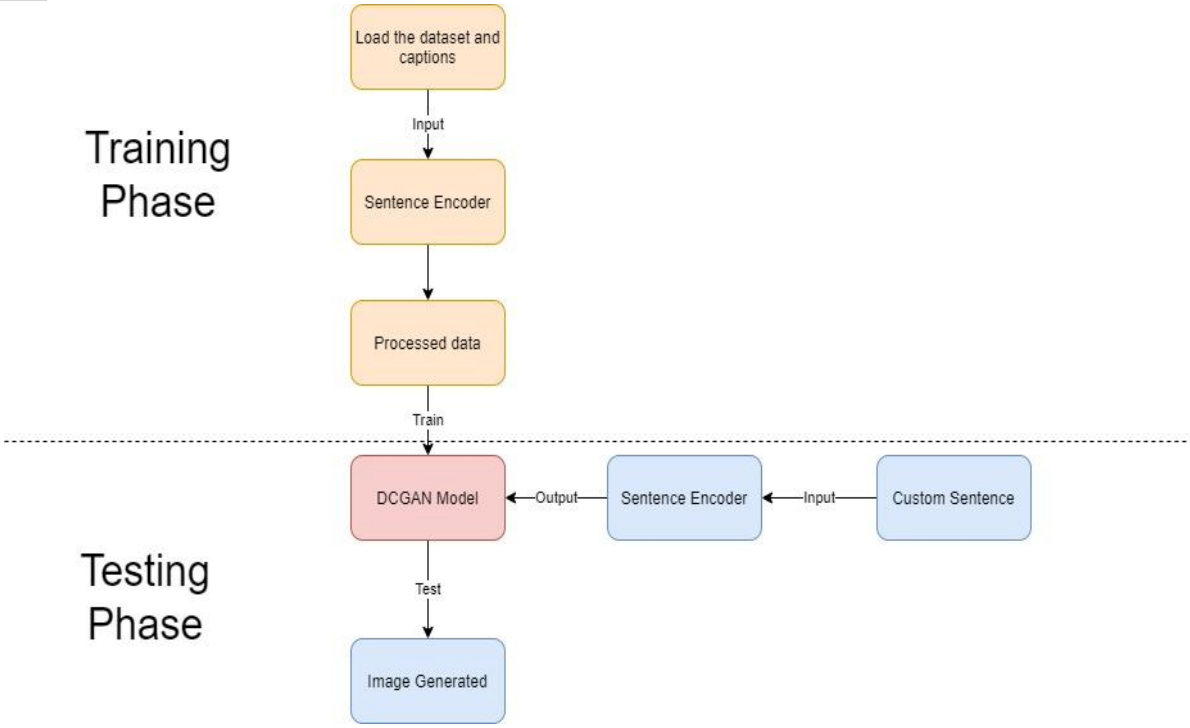


Fig. 1. Machine Learning pipeline for training and testing

3) Testing: Once I have a trained generative adversarial network and a sentence encoder, we can pass in the user sentence that contains description of the flower to be generated as seen in Fig 1. The trained model produces the result as seen in Fig 2,3.

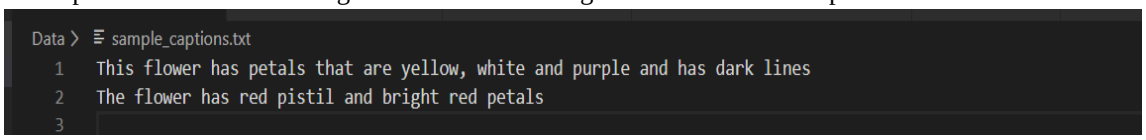


Fig. 2. Image description



Fig. 3. Generated images of first sentence



Fig. 4. Generated images of second sentence

III.LIMITATIONS

A. Time Consuming

The time taken to train the model and to generate these images is substantial as the model must learn the features and generate them based on the caption.

B. Expensive

The resolution of the images is not of the highest quality because it is a very highly graphic intensive task and requires a capable GPU which can be expensive

©IJRASET: All Rights are Reserved

IV. CONCLUSION

To conclude the paper, the model that I have devised is able to generate plausible images that are quite similar to the user provided captions and this shows the huge potential GANs research and development holds, by being able to mimic any distribution of data generative adversarial networks can create worlds extremely similar to our own in any domain like: images, music. The applications can range from generating new music without copy rights to solving difficult criminal cases by using well developed generative models to generate images based on the criminal's description.

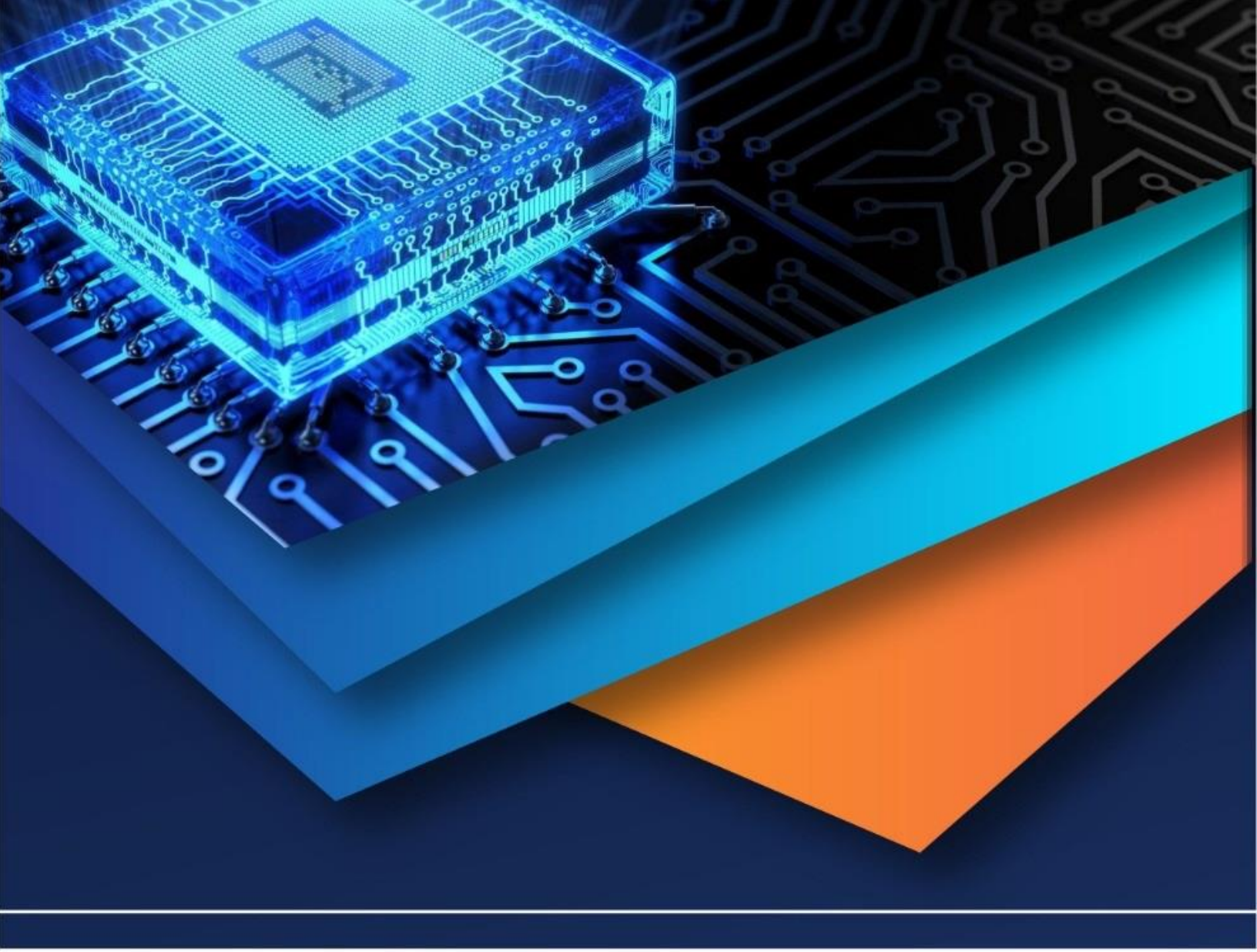
V. FUTURE SCOPE

Dealing with the limitations of the proposed system for Generating high quality images based on the dataset is the most immediate feature that can be implemented on to the model. The datasets that can be successfully and efficiently used to train the model with different scenarios are limited. That scope can be explored to bring in diverse datasets to train the GAN model. Training with human faces is the most challenging aspect of the model considering the hardware and time required to make use of the trained model is limited with the resources as of now. Thus, the future scope of the project will concentrate more on fixing the limitations and bringing in various image samples and applying in various fields like animation sector where in an artist can create a rough idea of this imagination based on a simple text caption.

REFERENCES

[1] Zhang, Chenrui, and Yuxin Peng. "Stacking vae and gan for context-aware text-to image generation." 2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM). IEEE, 2018.
[2] Xu, Tao, et al. "Attngan: Fine-grained text to image generation with attentional generative adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
[3] Reed, Scott, et al. "Generative adversarial text to image synthesis." arXiv preprint arXiv:1605.05396 (2016).
[4] Porter, T., & Duff, T. "Compositing digital images". In ACM Siggraph Computer Graphics (Vol. 18, No. 3, pp. 253-259).
[5] Radford, A., Metz, L., & Chintala, S. (2015). "Unsupervised representation learning with deep convolutional generative adversarial networks". arXiv preprint arXiv:1511.06434.
[6] Zhang, H., Xu, T., Li, H., Zhang, S., Huang, X., Wang, X., & Metaxas, D. "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks". In ICCV, 2017.
[7] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. "Improved techniques for training gans". In NIPS, 2016.
[8] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. "Generative adversarial nets". In NIPS, 2014.
[9] Mirza, M., & Osindero, S. (2014). "Conditional generative adversarial nets". arXiv preprint arXiv:1411.1784.
[10] Arjovsky, M., Chintala, S., & Bottou, L. (2017). "Wasserstein gan". arXiv preprint arXiv:1701.07875. 39
[11] H. Dong, J. Zhang, D. McIlwraith, and Y. Guo. "I2t2i: Learning text to image synthesis with textual data augmentation". In ICIP, 2017.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)