International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958 (Online), Volume-9 Issue-4, April 2020

Text to Image Translation using Cycle GAN

Kambhampati. Monica, Duvvada Rajeswara Rao

Abstract: ***In the recent past, text-to-image translation was an active field of research. The ability of a network to know a sentence's context and to create a specific picture that represents the sentence demonstrates the model's ability to think more like humans. Common text--translation 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. We present detailed similarities between our system and the methods already in place. 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. The re-penny achievement of Generative Adversarial Networks (GANs) demonstrates that they are a decent contender for the decision of design to move toward this issue***.

Keywords: Generative Adversarial Networks, Image, Synthesis, Text, Translation.

I. INTRODUCTION

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 of Hao et al. [1] 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 of from content to pictures is exceptionally multi-modular. The term 'multi-modular' is a significant one to get comfortable with in Deep Learning research. This alludes to the way that there are a wide range of pictures of winged creatures with 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 comparing 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 recently 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.

II. RELATED WORK

Hao Dong et.al., [1] propose another preparation strategy called Image-Text-Image which coordinates content to-picture and picture to-content (picture inscribing) amalgamation to improve the exhibition of content to-picture union. They show that I2T2I can create better multi-classifications pictures utilizing MSCOCO than the best in class. We additionally exhibit that I2T2I can accomplish move learning by utilizing a pre-prepared picture inscribing module to create human pictures on the MPII Human Pose dataset without utilizing sentence comment. Muhammad Ajmal et.al., built up an application. Right now, an application is created to make an interpretation of the content coordinated to pictures for visual education. Further, a few methodologies for picture to-content multilingual interpreter is looked into in detail. By defeating the holes, which are distinguished by intensive audit of the writing, an improved procedure is proposed [2]. Accordingly, the improvement of utilization experiences four significant stages including: catching, extraction, acknowledgment and interpretation. In addition, Optical Character Recognition calculation is especially utilized for character extraction and acknowledgment with high precision under various ecological conditions. Chenrui Zhang et.al., propose a setting mindful way to deal with perform content to-picture age, which isolates foundation and closer view for creating top notch pictures,

Revised Manuscript Received on April 16, 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

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

just as uses correlatively between Variation Auto encoder (VAE) and Generative Adversarial Network (GAN) for powerful content to-picture age [3]. To begin with, setting mindful restrictive VAE is proposed to catch pictures' essential format and shading dependent on content, which gives diverse consideration on the foundation and frontal area of pictures for powerful content picture arrangement.

At that point, restrictive GAN is received for refining the age of VAE, which recoups lost subtleties and adjusts the imperfections for sensible picture age. Vyankatesh V. Rampurkar et.al.,[4] proposed procedures can discover content strings by utilizing structure-based parcel and gathering techniques utilizing morphological activities. Proposed framework endeavors toward Morphological approachs that guide programmed identification, division and acknowledgment of visual content substances in complex a few pictures and hence bringing about ideal execution as contrasted and existing methods. Two distinct strategies utilizing morphological tasks are proposed to discover content strings from any characteristic scene pictures. Proposed procedures depend on kin's strategy for example nearby character gathering technique and content line gathering strategy. Content line gathering technique can find content strings arranged at subjective directions.

Rintaro yanagi et.al.,[5] attempt to take care of this issue by using a book to-picture Generative Adversarial Network (GAN), which has gotten one of the most appealing exploration themes as of late. The content to-picture GAN is a profound learning model that can create pictures from their relating portrayals. We propose another recovery system, "Question is GAN", and dependent on the content to picture GAN that definitely improves scene recovery execution by straightforward methods. Our clever thought utilizes pictures created by the content to-picture GAN as inquiries for the scene recovery task. Also, dissimilar to numerous investigations on content to-picture GANs that for the most part centered on the age of top notch pictures, we uncover that the created pictures have sensible visual highlights reasonable for the questions despite the fact that they are not outwardly charming. We show the viability of the proposed system through test assessment wherein scene recovery is performed from genuine video datasets.

Tao Xu et.al., proposed an Attentional Generative Adversarial Network that permits consideration driven, multi-arrange refinement for fine-grained content to-picture age. With a novel consideration generative system, the AttnGAN can incorporate fine-grained subtleties at various sub locales of the picture by paying considerations to the applicable words in the common language portrayal [6]. The proposed AttnGAN fundamentally outflanks the past cutting edge, boosting the best announced commencement score by 14.14% on the CUB dataset and 170.25% on the all the more testing COCO dataset. A nitty gritty investigation is additionally performed by imagining the consideration layers of the AttnGAN. It just because shows that the layered consideration GAN can consequently choose the condition at the word level for creating various pieces of the picture.

Xiaolong Wang et.al., factorize the picture age process and propose Style and Structure Generative Adversarial Network. Their S2-GAN has two parts: the Structure GAN creates a surface ordinary guide; the Style-GAN accepts the surface typical guide as info and produces the 2D picture [7]. Aside from a genuine versus created misfortune work, we utilize an extra misfortune with figured surface normals from produced pictures. The two GANs are first prepared freely, and afterward combined through joint learning. Scott Reed et.al., build up a novel profound engineering and GAN detailing to adequately connect these advances in content and picture demonstrating, deciphering visual ideas from characters to pixels. We show the capacity of our model to create conceivable pictures of flying creatures and blossoms from point by point content depictions [8]. Miriam Cha et.al., mean to expand cutting edge for GAN-based content to-picture amalgamation by improving perceptual nature of produced pictures. Separated from past work, our engineered picture generator improves on perceptual misfortune works that measure pixel; include actuation, and surface contrasts against a characteristic picture [9]. They present outwardly all the more convincing manufactured pictures of feathered creatures and blossoms produced from content portrayals in contrast with the absolute most noticeable existing work.

Bo Dai et.al., investigate an elective methodology, with the plan to improve the expectation and assorted variety – two fundamental properties of human articulation. In particular, we propose another system dependent on Conditional Generative Adversarial Networks (CGAN), which mutually learns a generator to create depictions molded on pictures and an evaluator to survey how well a portrayal fits the visual substance [10]. It is important that preparation an arrangement generator is nontrivial. They conquer the trouble by Policy Gradient, a system coming from Reinforcement Learning, which permits the generator to get early input en route.

III.MTHODOLOGY AND IMPLEMENTATION

In this phase, we describe the structure of our methodology and we discuss the methods, which we used in the architecture.

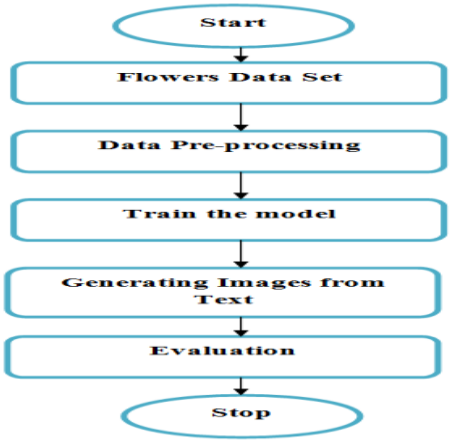

Fig 1. Proposed Methodology

Our methodology consists of following phases, they are;

- Data set Collection
- Data Pre-processing

© Copyright: All rights reserved.

- Train the Model
- Generate Images from the Text
- Evaluation

A.Data Set Collection

In this, we have made a 102 classification dataset, comprising of 102 blossom classes. The blossoms picked to be bloom ordinarily occurring in the United Kingdom. Each class comprises of somewhere in the range of 40 and 258 images. The pictures have huge scope, posture and light varieties. Moreover, there are classes that include huge varieties inside the classification and a few fundamentally the same as classifications. The dataset is envisioned utilizing isomap with shape and shading highlights.

Fig 2. Flowers Data Set

B.Data Pre-processing

In this section, Extract the skip-thought features for the captions, and prepare the training dataset by running the python script that is in our methodology implementation. This script will generate a series of pickled files that will be used in the directory during preparation.

C.Model Training

In this segment, we'll clarify how the GAN model performs training. We then clarify how the loss function can be easily expanded, so that the model can exploit some other potentially useful type of information. Let L_{DS} denote the loss of training related to the input source (real, fake or incorrect). L_{DS} is then calculated as a sum of the binary cross entropy denoted by H, between the discriminator's output and the desired value for each of the images.

Cyclic GAN

Our proposed model, the Cyclic GAN, creates pictures of size 128 × 128 that go along to the substance of the information content. To prepare our model, we utilize the Oxford-102 blossoms dataset, which has, for each picture, a class name and in any event five content depictions. For actualizing TAC-GAN we utilize the Tensor stream [1] execution of a Deep Convolution Generative Adversarial Network (DCGAN), in which G is displayed as a Deconvolutional Neural Network, and D is demonstrated as DCGAN-tensor stream a Convolution Neural Network (CNN).

In our approach we implement a version of Cyclic GAN, which is named to synthesize text images. Unlike AC-GANs, we condition the images produced from GANs on embedded text and not on class labels

The Generator Network is fundamentally the same as that of the ACGAN. Be that as it may, rather than taking care of the class name to which the incorporated picture should relate, we input the clamor vector, containing data identified with the literary portrayal of the picture. In our model, G is a neural system comprising of an arrangement of transposed convolution layers. It yields an upscale picture if of shape 128 × 128 × 3.

Our Generator network is composed by three transposed convolution layers with 256, 128 and 64 filter maps, respectively. The output of each layer has a size twice as big as that of the images fed to them as input. The output of the last layer is the produced If used as input to the Discriminator

D.Generate Images from the Text

To generate images from any text, do the following

- We Add Text Descriptions:
- Extract Skip-Thought Vectors:
- Generate Images:

By using above three steps, we generate images from text respectively. The process of generating images from text was implemented in python programming language. Generating images are visualized in results and discussion section.

E.Evaluation

We have used two metrics for evaluating Cyclic-GAN,

- Inception-Scope
- MS-SSIM score

Figure 2 displays some of the input images, along with the ground truth picture from the Oxford-102 dataset, for a given text summary. Our method can be seen to produce results whose content is compatible with the input text. The findings of our approach are contrasted with those of the Stack GAN model.

IV.RESULTS AND DISCUSSION

Furthermore, we show that our model confirms findings in other methods, i.e. it learns separate representations about the style and content of the images produced. The images in Figure 3 are generated by interpolating between two separate noise vectors while holding the same text input. Although the image's content remains largely unchanged, its composition seamlessly moves from one vector to the next. Since we use vector embedding for text explanations, it is also possible to interpolate between the two embedding. In Figure 3, we fix the same z for all images generated and interpolate between the embedding of the vector resulting from applying some to two text descriptions. The resulting photos can be seen retaining a similar style when switching seamlessly from one material to another. In existing literature most of the studies deals with only machine learning models for this problem. But, in our proposed methodology, we deal with deep learning models for getting better performance better than existing studies.

© Copyright: All rights reserved.

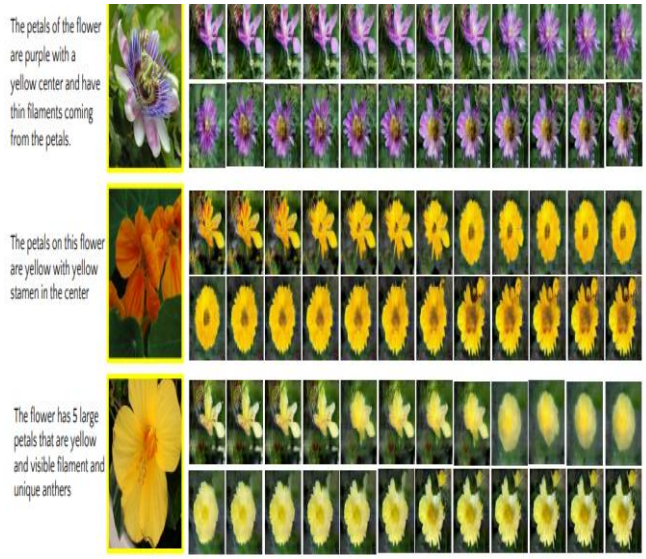

Fig 3. Generating images from the text

V. CONCLUSION

In this section, we presented the Cyclic GAN, a model fit for creating pictures dependent on printed portrayals. The outcomes created by our methodology are marginally better to those of other best in class draws near. The model is effectively extensible: it is conceivable to condition the systems on content, yet in some other kind of possibly valuable data. It stays to be inspected what impact the use of different kinds of data may have in the security of preparing, and the amount they help, rather than block, the limit of the model in creating better quality, higher goals pictures. Numerous methodologies have utilized a multi-organized design, where pictures created in the first stage are iteratively refined in quite a while. We accept that the consequences of our model can profit by such a pipeline, and have the option to additionally improve the outcomes revealed right now.

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-to-image 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, Biswash 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).
11. The dataset was available at http://www.robots.ox.ac.uk/~vgg/data/flowers/102/

AUTHORS PROFILE

Kambhampati.Monica, Studying Master of Technology, Department of Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada. Obtained B.Tech Degree in Computer Science and Engineering in NRI Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, India, in 2018

Duvvada Rajeswara Rao, Head of Department of Computer Science and Engineering (CSE), HOD, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada. He is qualified in Ph.D. in Computer Science and Engineering. He has 25 years of teaching experience and He published more than 62 journal papers and 5 international conference papers.

© Copyright: All rights reserved.