iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** III **Month of publication:** March 2023

DOI: https://doi.org/10.22214/ijraset.2023.49812

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: 7.538
Volume 11 Issue III Mar 2023- Available at www.ijraset.com

Text Guided Image Using Machine Learning

Dr.Krishna.K.Tripatis[1], Mohan.T.Mishra[2], Kajal.R.Dhende[3], Survesh Karande[4]
[1]*Professor,* [2, 3, 4]*Student, Department of Computer Engineering Shivajirao S Jondhale College of Engineering, Dombivli (E), Dist. Thane Maharashtra, India*

Abstract: ***Given a corrupted image, image inpainting aims to complete the image and outputs a plausible result. When we complete the missing region, we always borrow the information from a known area, which is aimless and causes unsatisfactory results. In our daily life, some other information is often used for corrupted images, such as text descriptions. Therefore, we introduce the use of text information to guide image inpainting. To fullfill this task, We introduce an inpainting model named TG Net (Text-Guided Inpainting Network). We provide a text-image gated feature fusion module to fuse text feature and image feature deeply. A mask attention module is proposed to enhance the consistency of known areas and the repaired area. Extensive quantitative and qualitative experiments on three public datasets with captions demonstrate the effectiveness of our method. The goal of our paper is to semantically edit parts of an image matching a given text that describes desired attributes (e.g., texture, colour, and background), while preserving other contents that are irrelevant to the text. To achieve this, we propose a novel generative adversarial network (Mani GAN), which contains two key components: text-image affine combination module (ACM) and detail correction module (DCM). The ACM selects image regions relevant to the given text and then correlates the regions with corresponding semantic words for effective manipulation. Meanwhile, it encodes original image features to help reconstruct text-irrelevant contents. The DCM rectifies mismatched attributes and completes missing contents of the synthetic image. Finally, we suggest a new metric for evaluating image manipulation results, in terms of both the generation of new attributes and the reconstruction of text-irrelevant contents. Extensive experiments on the CUB and COCO datasets demonstrate the superior performance of the proposed method. Code is available at this https URL.***

Keywords: Introduction, Existing System, Proposed System, Requirements

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 machine learning’s success. One of these growing fields is text guided image.

II. 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.

III. LITERATURE SURVEY

In “optimal text to image generator synthesis model for generating portrait image using generative adversarial network techniques” Mr. Mohammad berrahal proposed a system using Deep learning. It is the most approach Dl based of text to image synthesis to compare their efficiency in term of quantities assessment of generated image.[1]

In “ManuGan: Text guided image Manipulation” Mr. Bowel LiXiaojuam Qij2 Thomas Lukasiewicz Philip H.S Torr 11 university of Oxford2 university of Hong Kong.[2]

In “Text guided image using Deep learning” Rikita Shenoy Information Technology Vidyavardhin is college of Engineering and Technoloogy Vasai,india.[3]

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue III Mar 2023- Available at www.ijraset.com

IV. SYSTEM ARCHITECTURE

A text-to-image model is a machine learning model which takes as input a natural language description and produces an image matching that description. Such models began to be developed in the mid-2010s, as a result of advances in deep neural networks.

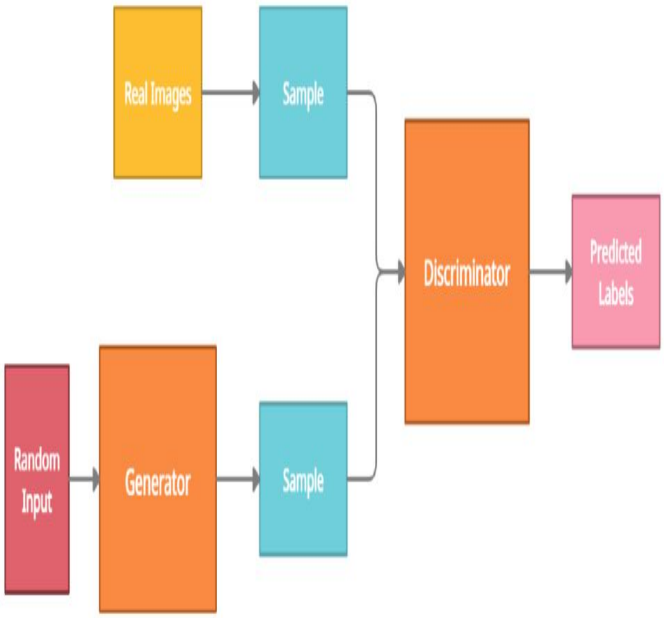


Fig. 1: Generative Adversarial Network Architecture

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 process.

V. REQUIREMENT ANALYSIS

1) Processor i3 (min)
2) Ram 4 gb ,8 gb for best performance.
3) Graphic 2 gb.
4) Hard disk 500 gb.

Programming languages used
Python

©IJRASET: All Rights are Reserved

VI. RESULT

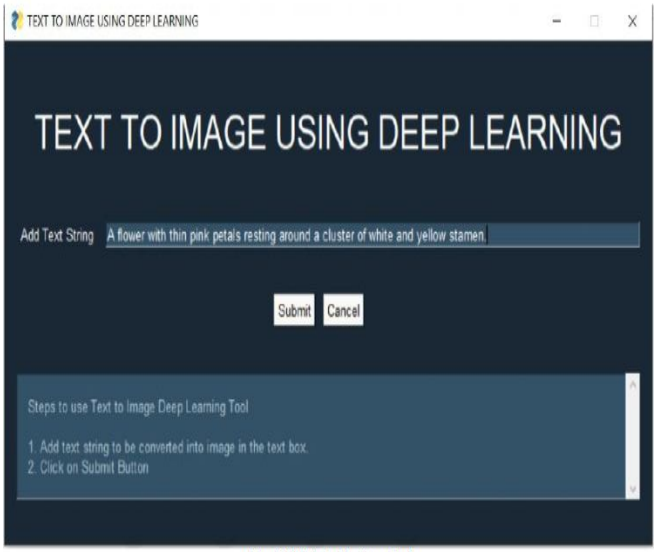


Fig. 4: GUI for taking input text

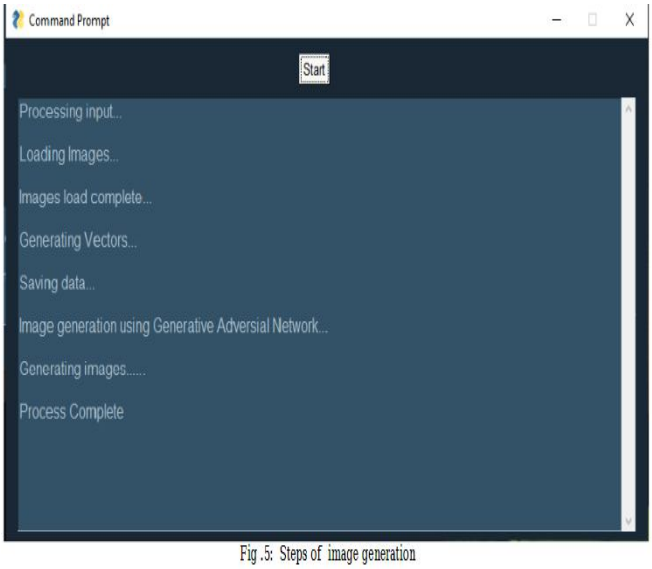


Fig .5: Steps of image generation



Fig. 6: Output

VII. FUTURE SCOPE

The departmental website will be updated as technology advances [9]. In order to maintain a smooth user experience, we will undoubtedly continue to implement updates on the website, including security updates [10]. Skills in development are becoming increasingly in demand as we move toward an even more technologically driven future. Machine learning, AI, the Internet of Things (IoT), quantum computing, and other similar technologies are transforming technology.

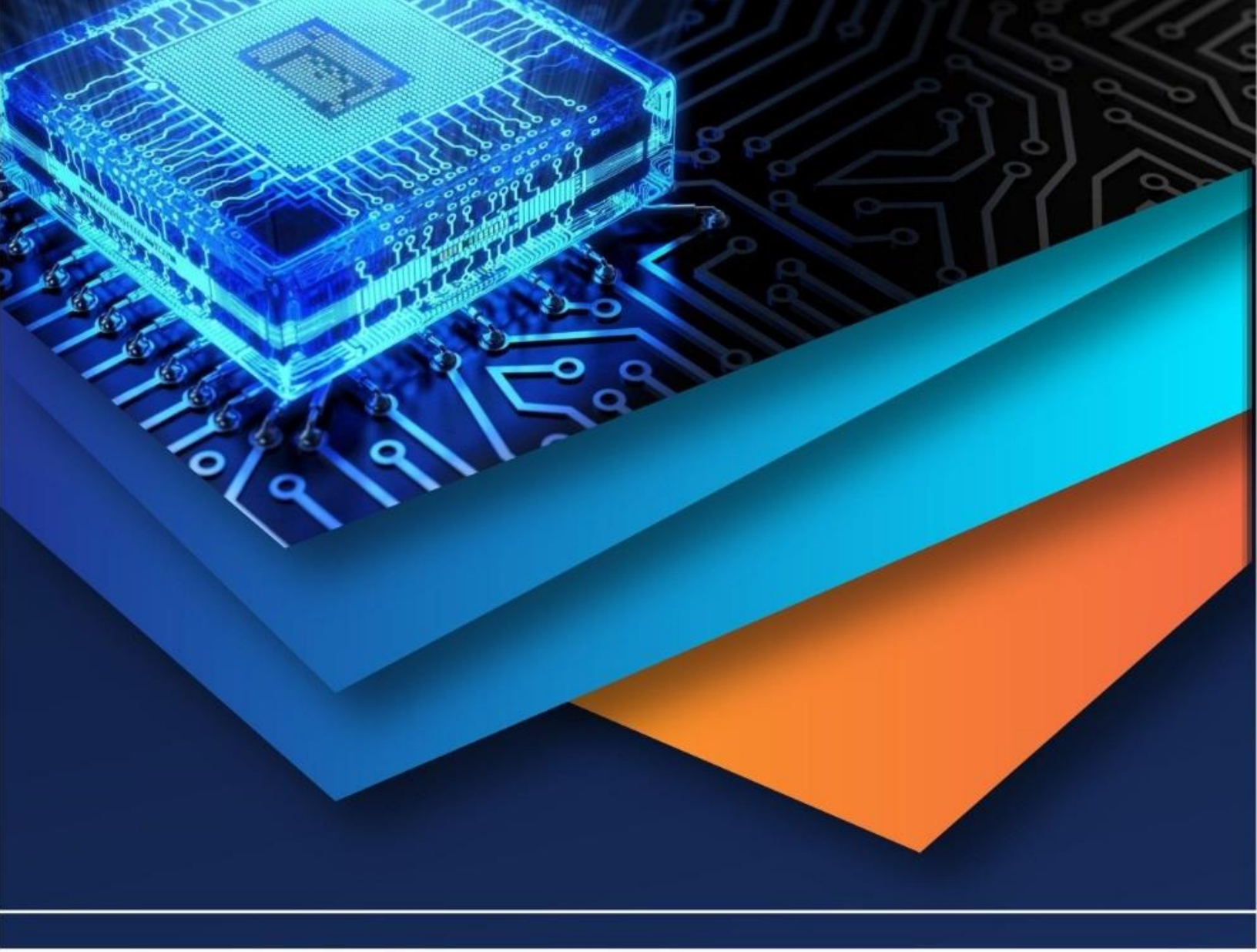
VIII. CONCLUSION

In the future, the efficiency and accuracy of the prototype can be enhanced to a certain level, and also enhance the user.

REFERENCES

[1] Kosslyn, S.M.; Ganis, G.; Thompson, W.L. Neural foundations of imagery. Nat. Rev. Neurosci. 2001, 2, 635–642. [CrossRef] [PubMed]
[2] Karpathy, A.; Fei-Fei, L. Deep visual-semantic alignments for generating image descriptions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3128–3137. ♣ 2. Karpathy, A.; Fei-Fei, L. Deep visual-semantic alignments for generating image descriptions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3128–3137.
[3] Vinyals, O.; Toshev, A.; Bengio, S.; Erhan, D. Show and tell: A neural image caption generator. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3156–3164.
[4] Youtube.com
[5] Google.com

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



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)