Report On

Text to image generation using TAC GAN

Submitted in partial fulfillment of the requirements of the Course project in Semester VIII of fourth year Artificial Intelligence and Data Science

by

Devashree Pawar (Roll No. 20)
Prasad Shah (Roll No.24)
Reena Vaidya (Roll No. 30)

Supervisor
Dr. Tatwadarshi P. N.

University of Mumbai
Vidyavardhini's College of Engineering & Technology
Department of Artificial Intelligence and Data Science

(2023-24)

Vidyavardhini's College of Engineering & Technology
Department of Artificial Intelligence and Data Science

CERTIFICATE

This is to certify that the project entitled “Text to image generation using TAC GAN ” is a bonafide work of “Devashree Pawar (Roll No. 20), Prasad Shah (Roll No.24), Reena Vaidya (Roll No. 30)” submitted to the University of Mumbai in partial fulfillment of the requirement for the Course project in Semester VIII of fourth year Artificial Intelligence and Data Science engineering.

Supervisor

Dr. Tatwadarshi P. N.

Dr. Tatwadarshi P. N.
Head of Department

Abstract

Text-to-Image production has attracted a lot of attention lately because of its many potential uses in virtual worlds, e-commerce, and content development. In this field, a promising paradigm that uses the power of Generative Adversarial Networks (GANs) to convert textual descriptions into realistic visuals is the Text-Conditioned Generative Adversarial Network (TAC-GAN). The goal of this theoretical framework is to clarify the underlying ideas and mechanics of the text-to-image creation process that TAC-GAN facilitates. Through an examination of the complex interactions between text and image modalities in the context of adversarial learning, this work seeks to shed light on the theoretical underpinnings that have guided the development of TAC-GAN-based image synthesis methods. This abstract attempts to contribute to a deeper understanding of the transformative potential of text-guided image generation in shaping the future of interactive digital experiences and visual content generation through an in-depth analysis of TAC-GAN's architecture, training methodology, and performance characteristics.

Table of Contents

Chapter No.		Title	Page No.
1		**Introduction**	
	1.1	Introduction	1
	1.2	Problem Statement	2
	1.3	Objective	2
2		**Proposed System**	
	2.1	Introduction	3
	2.2	Architecture/Framework	3
	2.3	Details of Hardware and Software	3
	2.4	Experiments and Results	4
	2.5	Conclusion	4
		References	5

Chapter 1: Introduction

1.1 Introduction

The fascinating field of text-to-image generation in artificial intelligence has attracted a lot of attention lately because of its many applications in various fields. Transforming written descriptions into vibrant visual representations has enormous potential for everything from content development to improving e-commerce and virtual environments.

In the field of text-to-image synthesis, the Text-Conditioned Generative Adversarial Network (TAC-GAN) is at the forefront of innovation. The powerful powers of Generative Adversarial Networks (GANs), a class of deep learning architectures well-known for producing lifelike data samples, are tapped into by TAC-GAN. Through the incorporation of textual descriptions as conditional inputs, TAC-GAN goes beyond the limits of traditional picture-generating methods, providing a means of smoothly converting language cues into visually captivating outputs.

A fundamental contribution to our understanding of text-to-image generating mechanisms is the conceptual foundation of TAC-GAN. Using an intricate competition between a discriminator and a generator network during adversarial training, TAC-GAN produces visuals that accurately match the accompanying textual descriptions. This adversarial paradigm creates a dynamic equilibrium that pushes the discriminator to reliably discern between real and fake images while forcing the generator to make ever more realistic images.

The key to TAC-GAN's effectiveness is its capacity to extract complex semantic details from text descriptions and make them appear in the images that are produced. TAC-GAN bridges the semantic gap between text and image modalities by use of feature alignment and contextual embedding, hence enabling the logical conversion of abstract notions into visually coherent representations. TAC-GAN's sophisticated comprehension of textual semantics enables it to generate images that capture not only the surface characteristics mentioned in the text but also the meaning and background expressed by the language input.

Moreover, TAC-GAN is a versatile framework that can be applied to a wide range of datasets and applications, which makes it easy to adopt and implement in a variety of contexts. When it comes to producing images of products based on textual product descriptions in e-commerce platforms or building immersive virtual environments using textual narratives as a guide, TAC-GAN demonstrates exceptional adaptability and scalability, highlighting its significance in modern AI-driven projects.

TAC-GAN's theoretical framework, which illuminates the underlying principles and mechanisms that support this transformative process, essentially marks a turning point in the evolution of text-to-image creation. TAC-GAN ushers in a new era of creativity and invention by seamlessly fusing textual semantics with visual fidelity. This blurring of the lines between language and vision creates a rich tapestry of synthetic experiences that are only constrained by the imagination.

1.2 Problem Statement and Objective

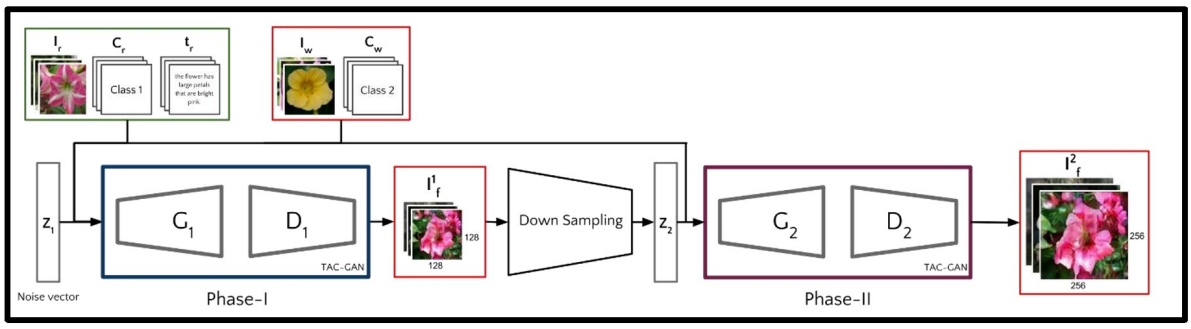
The synthesis of realistic and diverse images from textual descriptions poses a challenging problem with wide-ranging applications in various domains, including content creation, e-commerce, and virtual environments. Existing approaches, such as the Text-Conditioned Generative Adversarial Network (TAC-GAN), aim to address this challenge by leveraging Generative Adversarial Networks (GANs) and attention mechanisms. However, despite advancements, there remain several unresolved issues, including optimizing image quality and diversity while maintaining coherence with input text. Therefore, the primary objective of this project is to investigate and enhance the effectiveness of TAC-GAN for text-to-image generation, with a focus on improving image fidelity, diversity, and alignment with textual descriptions.

Chapter 2: Proposed System

2.1 Introduction

Text-to-Image generation has emerged as a significant area of research with applications spanning content creation, e-commerce, and virtual environments. This project explores the innovative Text-Conditioned Generative Adversarial Network (TAC-GAN) framework for generating realistic images from textual descriptions. At its core, TAC-GAN leverages Generative Adversarial Networks (GANs), employing a generator to synthesize images guided by text embeddings and a discriminator to discern between real and generated images. Incorporating attention mechanisms, TAC-GAN allocates resources to salient image regions, enhancing coherence and fidelity. The project aims to produce high-quality, diverse images through adversarial training and diversity-promoting objectives. This abstract encapsulates the theoretical framework and objectives driving the implementation of TAC-GAN for text-to-image generation.

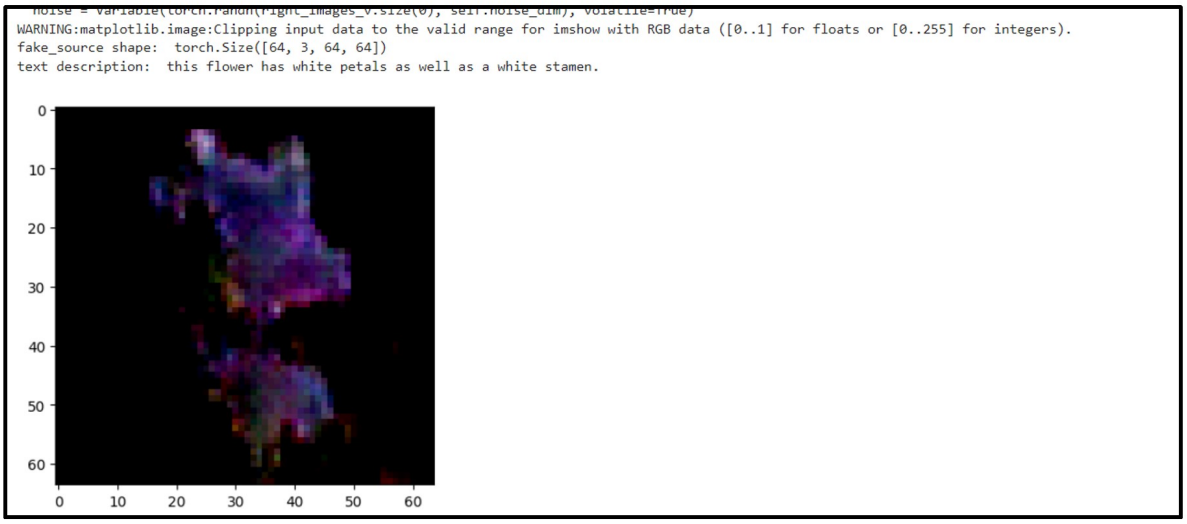
2.2 Architecture/ Framework/Block diagram

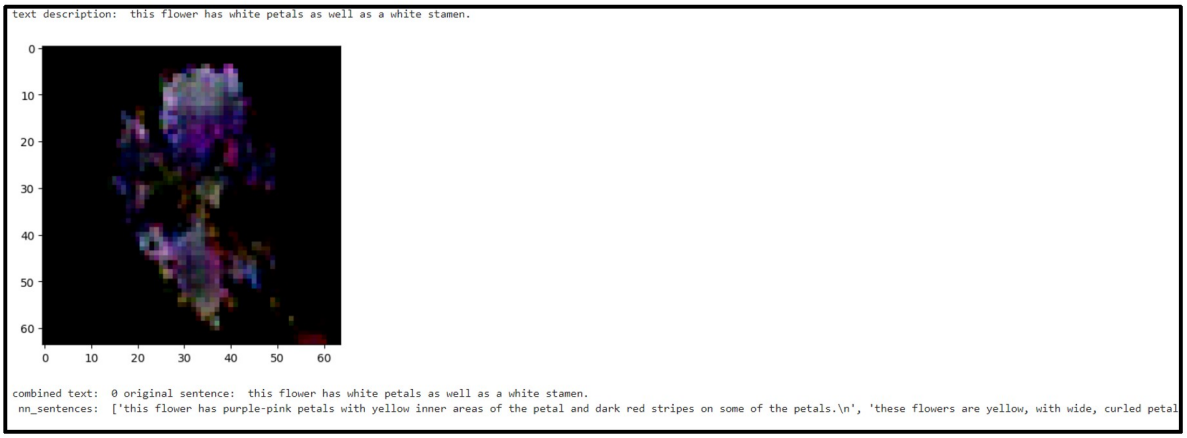


2.3 Details of Hardware and Software

- Python 3.8
- Keras
- Python Image Library
- PyTorch
- Tensorflow

2.4 Experiments and Results





2.5 Conclusion

In this project, we were successful in automatically coloring the grayscale images using Generative adversarial networks, to a visual degree which is acceptable. The images of CIPHER 10 with synthetic colors by GAN looked reasonably well and similar to the original images. There were some incidents where the model misunderstood the sea water for grass during the training process, but with further training, it was successful in coloring green color for grasses. We observed that the model faced an unusual problem with the color red, which it learnt after many epochs as compared to other colors.

References

[1]. Tiantian Wang, Ali Borji, Lihe Zhang, Pingping Zhang, Huchuan Lu, " A Stagewise Refinement Model for Detecting Salient Objects in Images", in *IEEE International Conference on Computer Vision*, DOI 10.1109/ICCV.2017.433, pp. 4039-4048, 2017.

[2]. Siyuan Qiao, Chenxi Liu, Wei Shen, Alan Yuille, "Few-Shot Image Recognition by Predicting Parameters from Activations", in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, DOI 10.1109/CVPR.2018.00755, pp. 7229-7238, 2018.

[3]. Maoke Yang, Kun Yu, Chi Zhang, Zhiwei Li, Kuiyuan Yang,"DenseASPP for Semantic Segmentation in Street Scenes", in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, DOI: 10.1109/CVPR.2018.00388, pp. 3684-3692, 2018.

[4]. Ran Yi, Yong-Jin Liu, Yu-Kun Lai, " Content-Sensitive Supervoxels via Uniform Tessellations on Video Manifolds", in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, DOI 10.1109/CVPR.2018.00074, pp. 646-855, 2018.

[5]. Luan Tran, Xiaoming Liu, "Nonlinear 3D Face Morphable Model", *in IEEE/CVF Conference on Computer Vision and Pattern Recognition*, DOI: 10.1109/CVPR.2018.00767, 2018.