MAJOR - 2 PROJECT

SOFTWARE REQUIREMENT SPECIFICATION (SRS)

For

Generative Adversarial Network Text To Image Synthesis

Submitted By

Specialization	SAP ID	Name
Artificial Intelligence & Machine Learning	500075183	Piyush Malviya
Artificial Intelligence & Machine Learning	500076400	Mudit Dagar
Artificial Intelligence & Machine Learning	500076579	Prateek Sehrawat
Artificial Intelligence & Machine Learning	500075154	Rahul Dhanola

Department of Informatics

School Of Computer Science

UNIVERSITY OF PETROLEUM & ENERGY STUDIES,

DEHRADUN- 248007. Uttarakhand

Dr. Virender Kadyan
Project Guide

Dr. Thipendra P Singh
Cluster Head

School of Computer Science
University of Petroleum & Energy Studies, Dehradun

Software Requirement Specification (SRS) Report

Generative Adversarial Network Text To Image Synthesis

Abstract

Generating distinct and distinguishable images from textual descriptions is a challenging task that has been tackled from various angles. One of the obstacles that need to be overcome is creating images that are both realistic and capture the meaning of the text. Generative Adversarial Networks (GANs) have demonstrated promising outcomes in image synthesis. The GANs use an adversarial training technique that employs the minimax algorithm. This involves training a generative model (G) and a discriminative model (D) simultaneously with opposing goals. G is trained to mimic the data distribution while D is trained to differentiate between real and generated data. To generate an image, an input noise vector (z) is passed through G. Since its inception, the framework has attracted a lot of attention. This project presents a GAN that can produce images based on a given textual description. The input vector of the Generative network is made up of a noise vector (z) and an embedded representation of the text description. Additionally, the Discriminator can be modified to receive the text information as input before performing its classification.

Table of Contents

Software Requirements Specification (SRS)

1. Introduction

Text to image synthesis is a process of automatically generating images from input text. Deciphering between data presented in pictures and text is a significant issue in the field of artificial intelligence. The capability of automatic image synthesis has several advantages and is a common application of conditional generative models. GANs are widely employed for image generation, and there have been significant improvements in this field recently. Converting text to images is an ideal example of deep learning. This innovation has numerous potential applications in the future once it is ready for commercial use. GANs are a type of generative model that can create new content, and they involve two neural network models competing with each other to capture variations in a dataset. Text to image synthesis aims to convert text descriptions into suitable images, and GAN models are now widely used for this purpose. One issue with deep learning is that a single text description can have many possible configurations, but this can be addressed through model training.

This project presents a GAN model that can generate images based on the provided text description. The Generative network's input comprises two vectors - one is a noise vector (z), and the other contains an embedded representation of the textual description. Additionally, the Discriminator can be augmented to receive the text information as input before performing its classification. The model's performance is evaluated using the Oxford-102 flower dataset, and text embeddings are created from image captions using Skip-Thought vectors.The images generated using GAN are diverse and highly discriminable.

Understanding text can be difficult, and visualization can sometimes be challenging. There are cases where words can be misinterpreted, but representing text in the image format makes it easier to comprehend. Images are more appealing than text and can deliver information more directly. Visual content can capture people's attention and keep them engaged, making it an essential component of activities such as learning and presentations. When designed effectively, visual communication offers numerous benefits.

2. Literature Review

- In recent years, there has been a rise in Deep Learning-based approaches for synthesizing images with varying degrees of success. Variational Autoencoders (VAE) have been used as generative models that produce samples from a distribution that approximates that of the training set.

- On the other hand, Generative Adversarial Networks (GAN) have garnered much interest and have been used in several tasks such as single-image super resolution, image-to-image translation, semantic image inpainting, and unsupervised learning.

- Several methods have been proposed to address the limitations of the basic GAN framework. One approach involves iteratively refining the generated images, such as in the Style and Structure GAN (S2-GAN) model, which uses a first GAN to produce the image structure, and a second GAN to generate the image style.

- The Laplacian GANs (LAPGAN) use a conditional form of GAN model integrated into a Laplacian pyramid with an indefinite number of stages.Other methods aim to make the network more aware of the data distribution that the GAN is supposed to model.

- For example, the Conditional GAN (CGAN) produces higher resolution images by conditioning the input on specific class labels. The Auxiliary Classifier GAN (ACGAN) can synthesize structurally coherent 128 × 128 images by training the discriminator to classify its input.

3. Problem Statement

The task of creating photo-realistic images from text is significant and has vast practical uses such as photo-editing, computer-aided design, and more, which can be accomplished through the use of GANs. The potential of GANs encouraged us to develop a system capable of generating fabricated images from the given input text.

4. Objective

To develop a Text to Image Synthesis system using the Generative Adversarial Networks (GAN) and its calculating the Generator/Discriminator loss over time.

Therefore, the following objectives need to be achieved to satisfy the development of the project.

- To study Generative Adversarial Network (GAN) and develop a system that is able to generate images
- To detect, extract and recognize image characterstics using Convolutional Neural Network (CNN) and,
- To study about Skip-Thought vectors to generate an embedding vector for each of our dataset captions and their corresponding images.

5. Methodology

4.1 Proposed Method

The proposed system when subjected to a scenario of a set of text description of images (mainly flowers related), the characters in the text description are converted to skip-thought vectors which have the features of the image to be generated and then the vector is passed to the GAN which generates the images based on the features identified by the text. Generative Adversarial Network is using the stages like preprocessing, feature extraction and recognition using neural network.

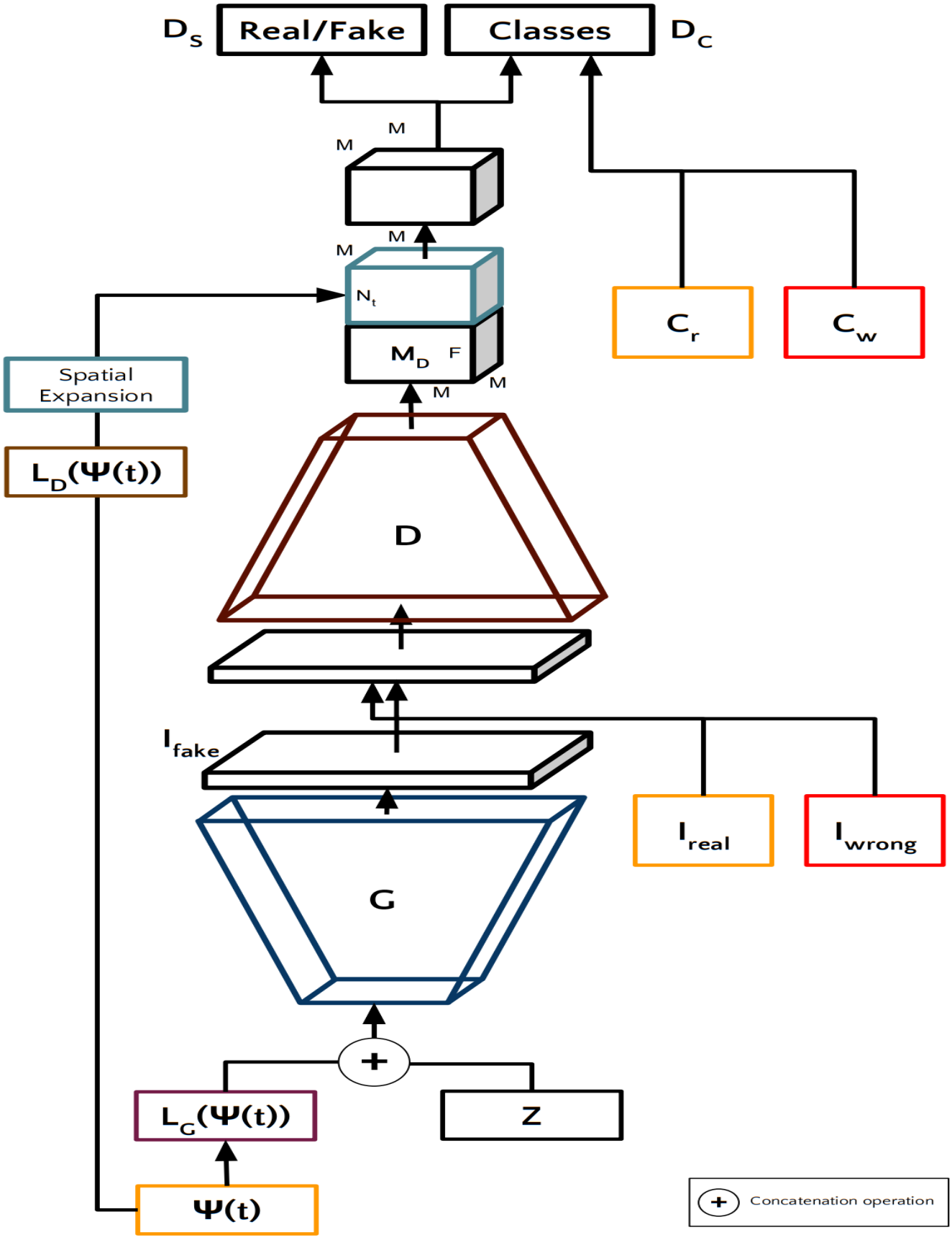
4.2 System Architecture

A text to image generation system receives an input in the form of text which contains some information about the image to generate. The output of this system is the image generated similar to as described in the text description. There are two Networks:

(A) Generator Network (G)

(B) Discriminator Network (D)

Each module is further described in detail as bellow:


D_S
Real/Fake
Classes
D_C
M
M
M
M
N_t
M_D F
M
M
C_r
C_w
Spatial Expansion
$L_D(\Psi(t))$
D
I_{fake}
I_{real}
I_{wrong}
G
+
$L_G(\Psi(t))$
Z
$\Psi(t)$
+ Concatenation operation

(A) The Generator Network (G) : The Generator Network (G) is similar to ACGAN but instead of taking in the class label, it takes in a noise vector zˆc which contains information related to the image description. The generator, G, is composed of transposed convolutional layers that produce an enlarged fake image, If, with dimensions of 128 × 128 × 3.

(B) Discriminator Network (D) : the Discriminator Network (D) comprises a series of convolutional layers that take in an image I from set A. The image is down sampled into MD of size M × M × F using convolutional layers. The spatially replicated lr vector of shape M × M × Nl is concatenated with MD in the F dimension. This concatenated vector is then passed through another convolutional layer of spatial dimension M × M. Finally, two fully connected layers F C1 and F C2 are used with 1 and Nc neurons, respectively, along with a sigmoid activation function. FC1 generates a probability distribution DS for the sources (real/fake), and FC2 produces a probability distribution DC for the class labels. Figure 1 provides further details on the architecture.

6. PERT CHART

Our month wise plan to complete the project is as follows –

Date	Expected Work
February 2023	• Research paper & feasibility understanding of the different methods used in the project. • Gather all the requirements like hardware and software tools and modules for the implementation. • Also collecting & understanding the required datasets.
April 2023	• Starting the Implementation part with various pre-processing steps. • Arranging the data according to the model requirements. • Training & testing of the model.
May 2023	• Optimizing & tuning the model parameters to increase the model accuracy. • Visualization of various approaches for better understanding. • Completion of the Project. • Preparation of Project Report.

References

1. M.-E. Nilsback and A. Zisserman. Automated flower classification over a large number of classes. In Computer Vision, Graphics & Image Processing, 2008. ICVGIP'08. Sixth Indian Conference on, pages 722–729. IEEE, 2008.
2. C. Doersch. Tutorial on variational autoencoders. arXiv preprint arXiv:1606.05908, 2016
3. D. P. Kingma and M. Welling. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114, 2013.
4. Nisha Sharma et al, "Recognition for handwritten English letters: A Re- view "International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 7, January 2013. Access Date:09/07/2015
5. Salvador España-Boquera, Maria J. C. B., Jorge G. M. and Francisco Z. M., "Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN Models", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 4, April 2011