Project Synopsis

Project Title:
Image Synthesis from text description using GAN.

Sponsored / Inhouse:
Inhouse

Team Members Names:
Saurabh Badgujar (BE402039)
Rushikesh Bhosle (BE402046)
Jagannath Gaidhani (BE402057)

Project Domain :
Machine Learning

Project Description :
Generating images from text descriptions is a challenging problem in computer vision and has many practical applications like computer aided creation. In this project we are going to generate image from text description using GAN(generative adversarial network).It consists of two stages the first stage deals with extracting features from text description and in the second stage the image is generated from the features extracted.

Literature Survey:

Sr no	Paper Title, Authors, year of publication	Concepts described in the paper	Gaps in the paper
1	Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran Bernt Schiele, Honglak Lee ,"Generative Adversarial Text to Image Synthesis" ,2016	Simple and effective model for generating images based on detailed visual descriptions.	Does not incorporate hierarchical structure into the image synthesis model in order to better handle complex multi-object scenes.
2	Han Zhang1, Tao Xu2, Hongsheng Li3 ,"StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks", 2017	Two Step StackGAN to generate photo-realistic images conditioned on text descriptions.	Low resolution image synthesis and limited set of images synthesized(dataset restricted)
3	Saifuddin Hitawala ,"Comparative Study on Generative Adversarial Networks", 2018	Comparison between different GAN networks on the basis of their methodology, architecture and performance.	Latest GANs like Boundary-Seeking Generative Adversarial Networks (Hjelm et al., 2017), Wasserstein Generative Adversarial Networks (WGAN),etc not compared
4	He Huang, Phillip S. Yu and Changhu Wang ,"An Introduction to Image Synthesis with Generative Adversarial Nets", 2018	Methods used in image synthesis, review different models for text-to-image synthesis	Image synthesis of complex objects not possible

Literature Gap:
While all literature surveys help us to understand how GAN(generative adversarial network) but , all the papers implemented only on the bird and flower datasets . we can improve the system by implementing on different datasets like car, cat, dog etc. And the images generated are low resolution and we can try to improve the image resolution.

Feasibility study (comparisons of existing systems):
Scott read and his team had implemented model with flower and bird dataset and there model was limited only for text description of birds and flower.

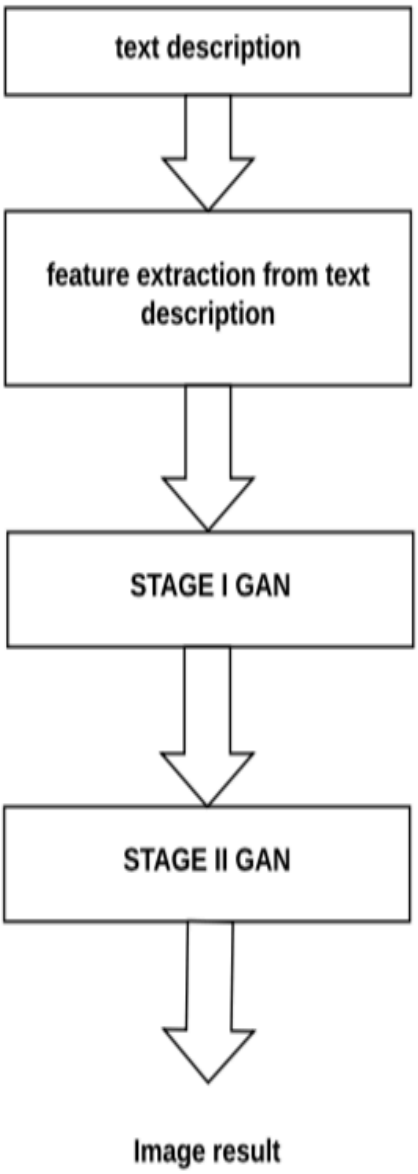
In another paper StackGAN they have proposed Stacked version of GAN(generative adversarial network) which has improved the image quality but still it was implemented on bird and flower dataset.

While these models were tested on standard dataset our system will be implemented on different dataset.

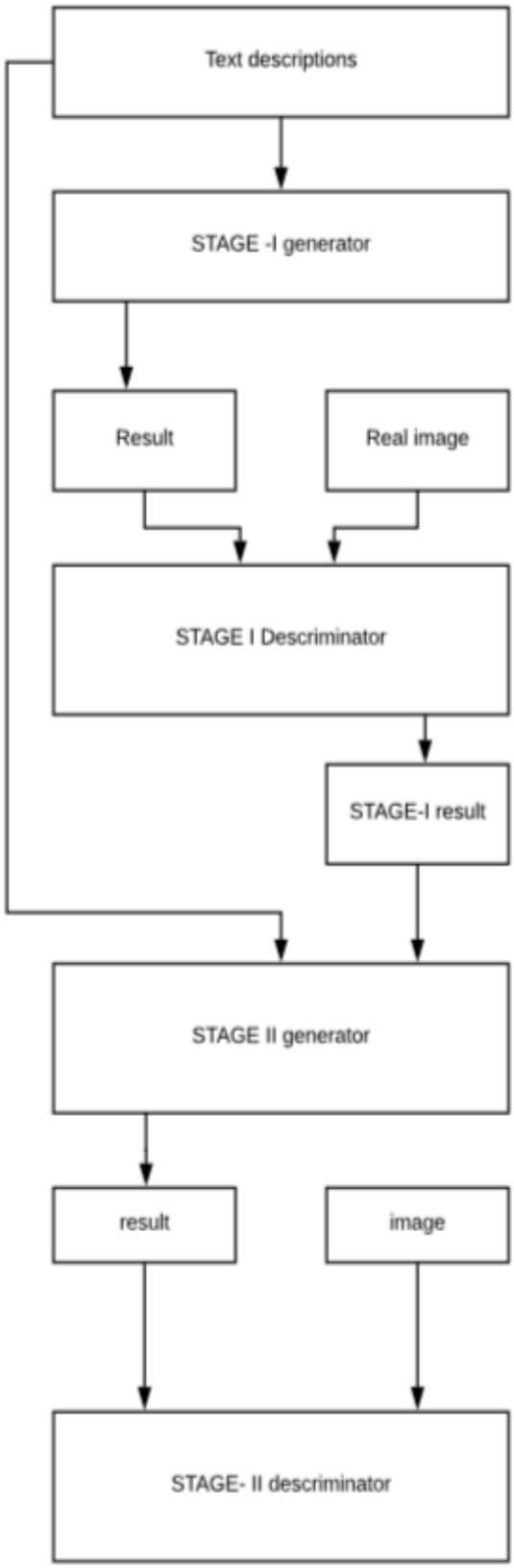
Project Scope(funtionalities to be implemented):
Following functionalities will be implemented

1. Generating images from text description
2. Taking voice commands and convert them to text
3. GUI will be provided so that user can select between voice mode and text mode

High Level Design:



Low Level Design :



Hardware Requirements:

1. 64-bit intel processor with 4GB RAM(min)
2. External GPU(in case of large datasets)

Software Requirements:

1. TensorFlow/keras
2. Python
3. tKinter library (python)

Technologies used:

PYTHON programming,

Testing and Development Strategies:
In this model the aim is to directly derive pixels from the features extracted from the text description . the datasets publically available for this application are the Oxford-102 flowers dataset and the Caltech CUB-200 birds dataset . the model is trained on these datasets and for generating new images the generator will add a noise to the existing image and give it to the discriminator, which will detect whether the image is real or fake. Real image is the image from the training set and the fake image is the image generated by the generator.

Future Enhancements:
The current model works on limited images like images of birds and flowers, in future we can train the model on different image sets to make the model work on different inputs . Another way to enhance the current model is to generate high resolution images because the images generated from the current model are low resolution .