**BE-COMP-A-PROJECT**

DAY-3,4

***Synopsis Preparation***

# Background :-

* Synthesizing photo-realistic images from text descriptions is a challenging problem. Previous studies have shown remarkable progresses on visual quality of the generated images. Where they have consider semantics from the input text descriptions in helping render photo-realistic images. However, diverse linguistic expressions pose challenges in extracting consistent semantics even they depict the same thing. To be specific, they designed a Siamese mechanism in the discriminator to learn consistent high-level semantics, and a visual-semantic embedding strategy by semantic-conditioned batch normalization to find diverse low-level semantics. Extensive experiments and ablation studies on CUB and MS-COCO datasets demonstrate the superiority of the proposed method in comparison to state-of-the-art methods. The input of each branch is a sentence of natural language description. The text encoder E aims at learning the feature representations from the natural language descriptions and they adopt a bi-directional Long Short-Term Memory (LSTM) that extracts semantic vectors from the text description. Generally, in the bi-directional LSTM, the hidden states are utilized to represent the semantic meaning of a word in the sentence while the last hidden states are adopted as the global sentence vector. They compared the proposed SD-GAN with the previous methods, i.e., StackGAN and AttnGAN . When users are asked to rank images based on their relevance to input text, they choose the generated images by SD-GAN as the best mostly, wining about 70% of the presented texts, much higher than others. This is consistent with the improvements of inception score. Furthermore,For each example, we compare the generation results from the descriptions of the same ground-truth image. Due to the lacking of the word-level details, StackGAN fails to predict the important semantic structure of object and scene. Although AttnGAN adopts the attention mechanism to extract details from the text, it is difficult to generate the corresponding visual concepts under linguistic expression variants. Comparing to them, the proposed SD-GAN generates more recognizable and semantically meaningful images based on the input texts.

# Literature Review :-

We are able to illustrate following synopsis after preparing Literature Survey.

* To overcome the problem of low resolution images by using stage 1 and stage 2 GAN to generate a much higher resolution version of the image by completing all the details.
* To handle the level of accuracy to reduce the gaps between the generated and original image.
* To work on data to resolve issues and be useful for criminal cases.
* To focus on text irrelevant contents in the original image and global coherent structure
* To increase the visual quality and resolution of the generated image.
* To select a better algorithm to suit and improve the selection of images by eliminating the wrong image and side image which should not be included in the training set.

# Problem definition :-

To model a system for generation of realistic face image that matches the text description using LSTM and progressive GAN. The input required is text and the desired output is photo realistic face image.

# IDENTIFICATION OF TOOLS AND TECHNOLOGIES :-

1.We can use GPU(Graphics Processing Unit) which takes 12-16 sec to processes one image when compared with CPU(Central Processing Unit) which takes 2-5 min to processes a single image.

2. But using GPU(hardware) is too costly.So another way is to use Google Colab which is a free Jupyter notebook environment provided by Google where in we can use free GPUs(Graphics Processing Unit) and TPUs(Tensor Processing Unit) which can solve all the issues and it is open source as well.

# List of expected outcomes :-

We are able to illustrate synopsis for phase wise outcomes to be generated

* In phase1 we will apply NLP (Natural Language processing) on our text data for feature representation.
* In phase2 we will do Exploratory Data Analysis and Data Visualization part by utilizing the above dataset and if required then we will apply feature engineering as well.
* In phase3 we will try to apply Convolutional Neural Networks and various GAN models
* In phase4 we will check the inception score for various dataset on our models.
* In phase5 we will try to get the desired output (which is generating realistic photo from text descriptions).