Progress Presentation-I

e-Yantra Summer Internship-2018

Text-to-Image/Video Synthesis using GANs

Aishwarya Kalloli
Deval Srivastava

Mentors:
Aditya Panwar, Kalind Karia

IIT Bombay

June 6, 2018

Overview of Project

- Project Name: Text-to-Image / Video Synthesis using GANs
- Objective: To generate image or video from given caption
- Deliverables:
  1. To create a model that can generate new images by getting trained on a given dataset
  2. Creation of video from these new set of images
  3. Prepare proper documentation and tutorial of the solution

Overview of Task

Project Task List		
Task	Task	Deadline (days)
1	Understanding the idea and create report on how it can be tackled using Machine Learning: a basic report of 2-5 pages highlighting various algorithms suitable for the task	2
2	Installing the required software	1
3	Perform a basic experiment to understand GANs (MNIST)	2
4	Gather the required data-set to train the model	1-2
5	Design the model, test its feasibility	2
6	Train the model and calculate the accuracy of the model	6
7	Generate new data-set of images/scenes from the text	4
8	Create a video/scene from the set of generated images	6
9	Develop proper tutorial and documentation (with video demo) on the implementation	4
10	Text to audio/music generation (optional)	3

Task Accomplished: Report

- Understanding the idea and create report on how it can be tackled using Machine Learning.

We went through several papers and researched about different GANs, based on that we decided to use DCGAN that will be conditioned on text embeddings generated by a Character RNN to Generate Images from text

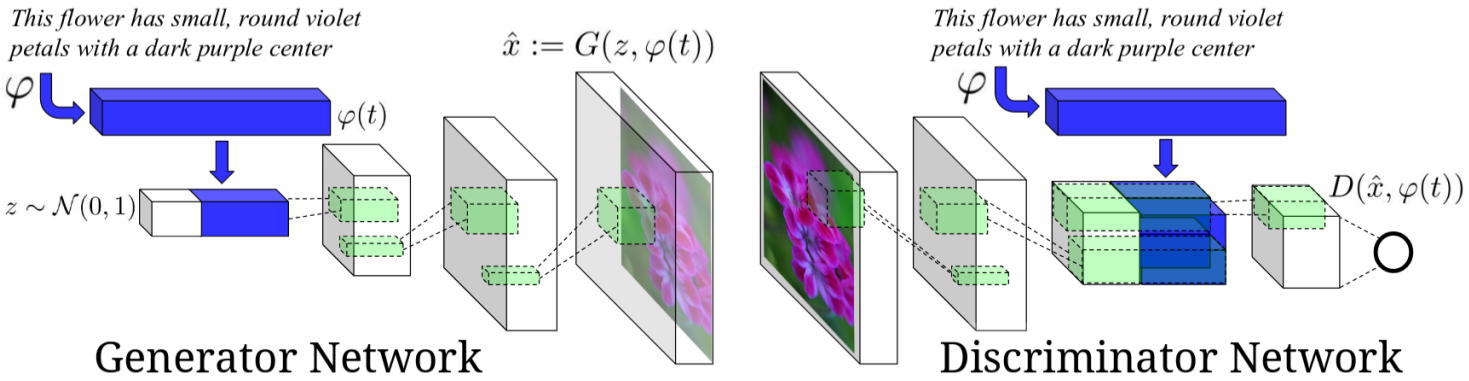


Figure 1: Architecture[1]

[1] Generative Adversarial Text to Image Synthesis by Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, Honglak Lee

Task Accomplished: Software Installation

- Installing the required software

1. Python, PyTorch and Torchvision were successfully installed
2. A Nvidia GTX 1080Ti was employed for training
3. CudaNN and Nvidia drivers were installed to allow training models on the GPU

Task Accomplished: DCGAN

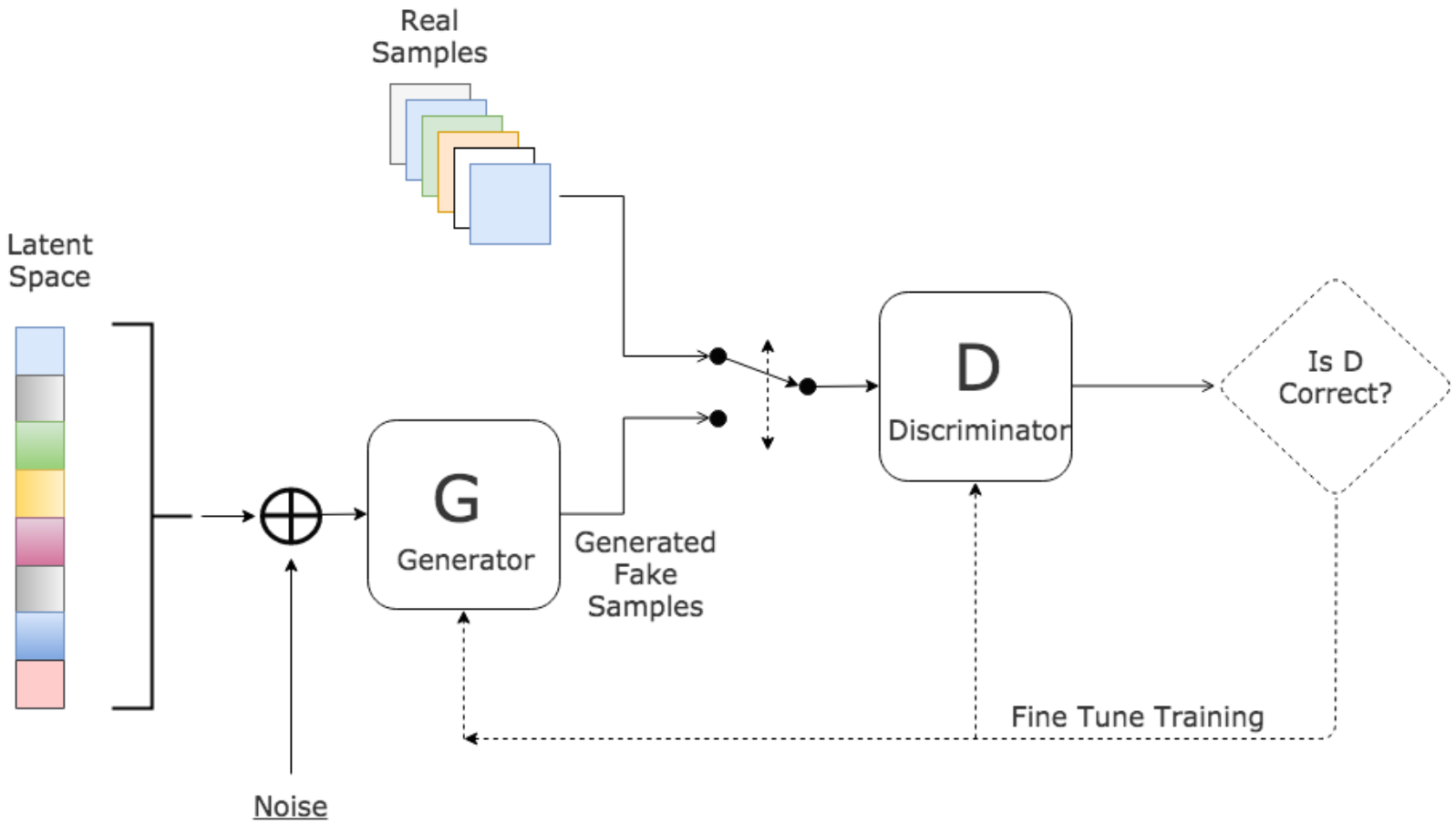


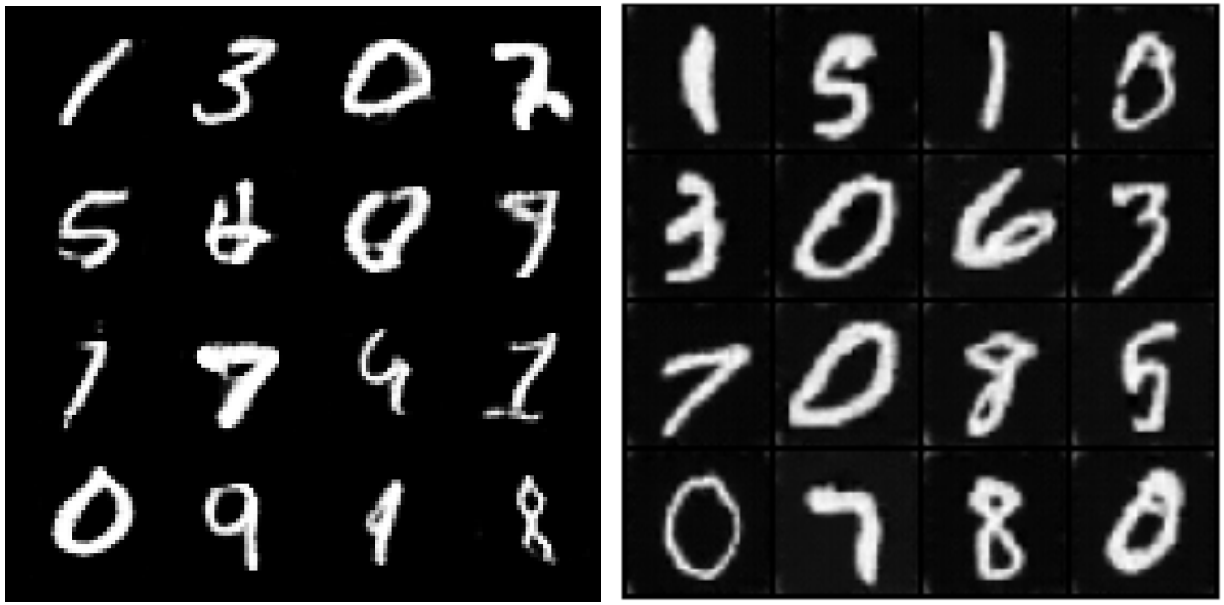
Figure 2: DCGAN block diagram[2]

[2] GANs from Scratch - Part 1 on Medium

Task Accomplished: MNIST example

- Perform a basic experiment to understand GANs (MNIST)

We used DCGAN to implement the task of generating new images from original MNIST dataset



(a) original MNIST (b) generated MNIST

Figure 3: Comparison of generated MNIST images

Task Accomplished: Gather Dataset

- Gather the required data-set to train the model for the final solution

We are planning to use COCO image dataset to train out text to image model. COCO is a large-scale object detection, segmentation, and captioning dataset. The dataset has been downloaded and prepared successfully.

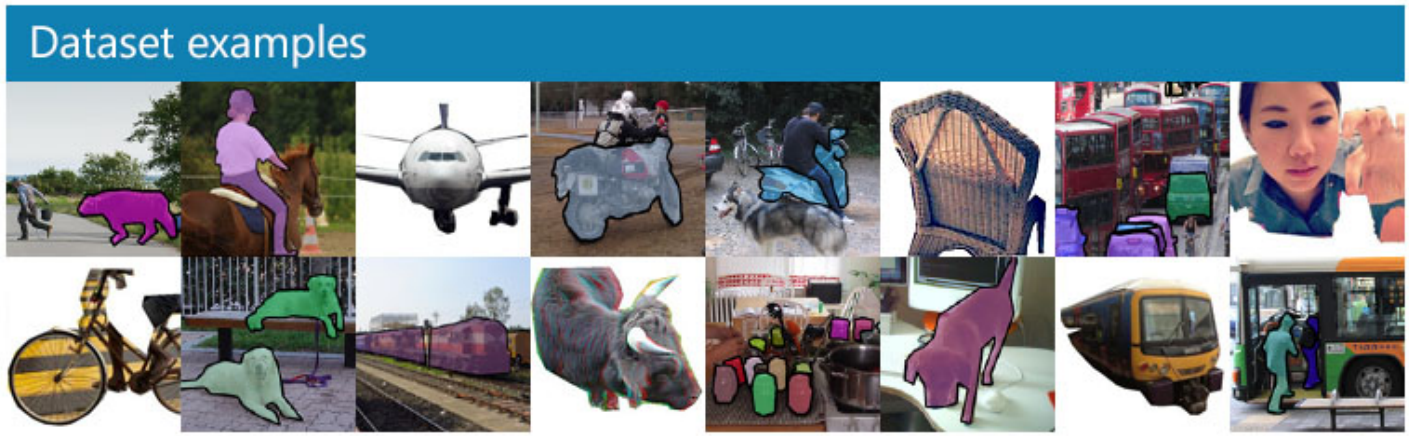


Figure 4: COCO Examples.[3]

[3] mscoco.org/dataset

Task Accomplished: Model Design

- Design the model and test its feasibilty

  Designing of model for generating images from text has has been completed but we are yet to test its feasibility and effectiveness. This current model incorporates the standard DCGAN architecture with conditioning on character data.

Challenges Faced

- Lack of knowledge about PyTorch before starting the internship
- Choosing hyper-parameters for GAN leading to efficient convergence
- Since GAN is fairly new and is still being actively researched, it took time to find the right algorithm for the task
- Finding a right Dataset for the task with captions and creating a dataloader for the same

Future Plans

- Test feasibility and effectiveness of our current model
- Train the model and calculate the accuracy of the model
- Generate new data-set of images/scenes from the text
- Next we plan to use these recent images to generate videos by either using a LSTM network to predict the next frame or a Temporal GAN

Thank You

THANK YOU !!!