Facial Image Generation from GANs

Arvind Ram
2020A7PS1210P
9th April 2023

Introduction:

The implementation of a Face generation algorithm using GANs is discussed in this report.
A facial image dataset was constructed from scratch by collecting images of people's faces and new images were constructed by the algorithm on the basis of text prompts.

Description of the dataset:

The dataset for this project was constructed by students who are a part of the Deep learning course. A set of 40 facial images were clicked by each student using their smartphone camera.

The images are in jpeg format and have a resolution of at least 500 × 500 pixels. The subjects in the image are facing the front with a neutral expression. The images are all solo images with no other objects or special lighting. A large part of the image is the face.

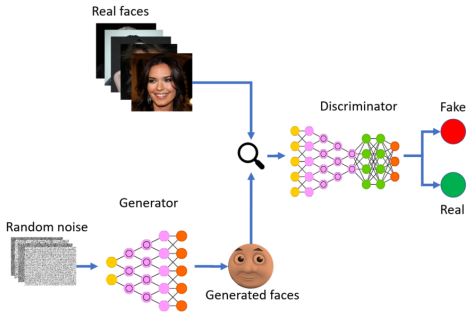
The following details are also collected for each image:

Type of Lip	Colour of Lip	Forehead	Cheeks with dimple
Ear	Eye Size	Eye Defect	Eye Color
Eyebrow style	Jaw Line	Skin Color	Hair Colour
Hair Style	Beard/Mustache	Age	Gender
Identification mark			

This combined Dataset is used to build our Face generation GAN.

Model Details:

GANs, or Generative Adversarial Networks, are a type of neural network architecture consisting of two neural networks, a generator and a discriminator. The generator is tasked with generating synthetic data that mimics the distribution of a given dataset, while the discriminator tries to distinguish between the real data and the generated data.
The two neural networks essentially compete with each other to try and become more accurate in their predictions.



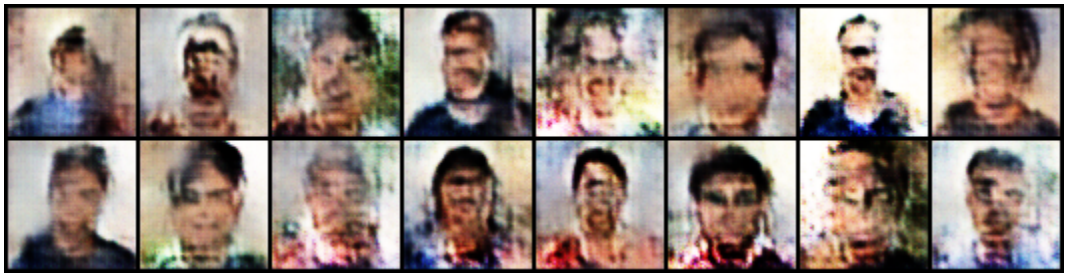
The training process of GANs involves an adversarial game between these two networks. The generator creates fake samples which are fed to the discriminator and real samples from the original dataset. The discriminator then tries to classify the samples as either real or fake, while the generator tries to generate samples that fool the discriminator into thinking they are real.

The discriminator and generator have a 5 layer architecture. The hidden layers of the discriminator use a leakyRelu activation function and the output layer uses a sigmoid activation function. The hidden layers of the generator use a Relu activation function and the output layer uses a tanh function.

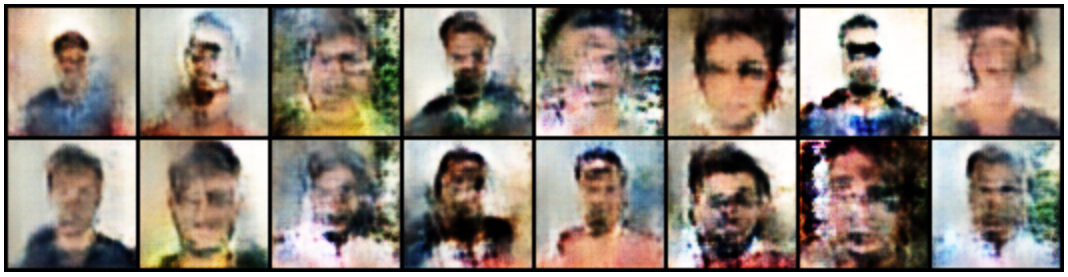
Results and Analysis:

The following sets of images were generated by the GANs in the 10th, 20th, 30th and 40th epoch.

10th Epoch



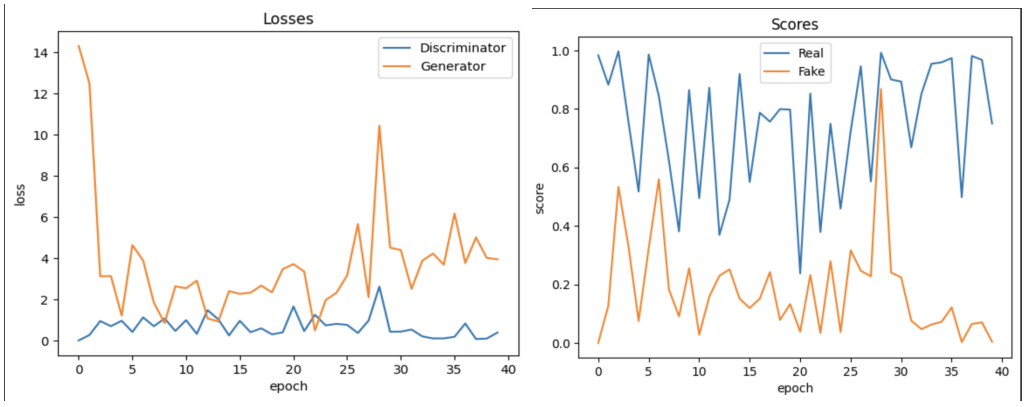
20th Epoch



30th Epoch

40th Epoch

Epoch	Generator Loss	Discriminator Loss	Real Score	Fake Score
10	3.9590	0.3992	0.7500	0.0036
20	4.5243	0.4415	0.9007	0.2403
30	3.4821	0.4104	0.7976	0.1325
40	2.6479	0.4776	0.8646	0.2552



The GAN Model built can simply generate new faces hence, further work is required in order to provide image details as input and generate new images based on text prompts.

References:

- Sun, Jianxin, et al. “Multi-caption text-to-face synthesis: Dataset and algorithm.” Proceedings of the 29th ACM International Conference on Multimedia. 2021.
- Xia, Weihao, et al. “Tedigan: Text-guided diverse face image generation and manipulation.” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.
- Sun, Jianxin, et al. “Anyface: Free-style text-to-face synthesis and manipulation.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.