Anime Character Generation Using Generative Learning

Name: Shreenidhi L Nayak
Registration number: 200968006
Section A (Batch 1)

Literature Review

For a better understanding of a way to approach the problem statement using generative learning, it is a good idea to have a look at the existing literature on this topic. Here are the papers reviewed to better understand the subject of generative deep learning using GANs and previous attempts at generating anime faces:

1) Generative Adversarial Networks - Goodfellow et al. 2014
   Link: https://dl.acm.org/doi/abs/10.1145/3422622
   Generative deep learning was an existent idea before this paper came out. Studies were done on Markov models and related areas. This paper argued that the generation can be done purely using the established techniques in deep learning. The idea was to use two deep learning models known as the generator and discriminator which were simple neural networks by themselves and pitch these two networks in a battle against each other. These could be trained purely using backpropagation as was already known. Thus was born the idea of Generative Adversarial Networks. The networks used were simple Fully connected and convolutional networks.
2) Towards the Automatic Anime Characters Creation with Generative Networks - Jin et al. (2017)
   Link: https://arxiv.org/abs/1708.05509
   This paper attempted the exact problem statement as this project. DRAGAN (Deep Regret Analytic GAN). Jin et al. used Illustration2Vec, a CNN-based tool to estimate tags for the anime character faces before training the network using these labels along with random sampled noise. The idea behind the labels was to have a human control over the factors of variation rather than just randomly sampled noise. This paper was able to achieve good results in generating anime character faces which according to the paper wasn't satisfactorily achieved by the time.
3) Towards Diverse Anime Face Generation: Active Label Completion and Style Feature Network - Li at al. (2019)

Link:
https://diglib.eg.org/bitstream/handle/10.2312/egs20191016/065-068.pdf?sequence=1&isAllowed=y&ref=https://codemonkey.link
This paper diverges a little from the area of this project. Li et al. introduce SGA-GAN (style gender based anime gan) which has gender as an input label of control along with random noise. The main takeaway from this paper is the SFN (Style Feature Network) training that was employed for active label completion. This network was trained on the training dataset for the gan to extract features.

4) A Style-Based Generator Architecture for Generative Adversarial Networks: Karras et al. (2019)
Link:
https://openaccess.thecvf.com/content_CVPR_2019/html/Karras_A_Style-Based_Generator_Architecture_for_Generative_Adversarial_Networks_CVPR_2019_paper.html
This paper by Nvidia presents an interesting new architecture for the generator part of the network. Firstly, the input layer as seen in most networks is eliminated. Instead, a learned constant is fed as input. The z vector sampled is first normalized and fed through a fully connected network to get an output transformed vector w. This vector is used like a style vector and is fed to the AdaIN (adaptive instance normalization) layer of the generator. Random noise is fed just before the AdaIn layers for stochastic variation. This paper goes in hand with the Progressive GAN paper, using that model as a baseline. The model is evaluated using FID (Frechet Inception Distance) score. The results of this paper are slightly better than pure Progressive GAN as measured by the FID scores (The lower the better).

Model Shortlist

The models that are planned to be attempted in this project are:
1) FCGAN (Fully Connected GAN)
2) DCGAN (Deep Convolutional GAN)
3) WGAN (Wasserstein GAN)
4) ProGAN (Progressive GAN)

Model Evaluation

1) FCGAN: The pro of this gan model is that it is the simplest gan that can be achieved and was the one used in the original Goodfellow paper along with DCGAN. This network is easy to implement given the existence of only FC Neural Networks. However this network has a lot of parameters to be optimized and might take computational time. This type of neural networks are also ill suited to handle images.
2) DCGAN: This type of gan uses Transpose Convolutions also known as deconv to upsample from the input vector and generate an image of desired dimensions. CNN is used in the discriminator network. This network is slightly harder to implement as the generator must be designed to output the same dimensional image as present in the dataset. The advantage here is that there are far less weights in this type of architecture and is also better for handling images.
3) WGAN: Wasserstein GAN misimises an approximation of the Earth-Mover's (EM) distance. This type of network is known to have a much lower chance of mode collapse wherein the generator generated a small subset of outputs despite different sampled z vector inputs.
4) ProGAN: Progressive GAN by Nvidia is a type of gan wherein the output images of the generator are small in dimension in the initial training iterations and progresses on to grow to the full required size by the end of the training process. This method has proven to have incredible stability during training as well as have a faster training time.

Baseline Models

The models used in the original GAN paper by Goodfellow et al., i.e. FCGAN and DCGAN will be taken as baseline models. The anime dataset will first be trained and slightly tuned to fit these architectures. For the purpose of scoring models teh FID score will be used. As the name might suggest, an Inception net is used for the scoring process. Once this is established, better models will be tried with an aim to improve the FID scores.