# StyleSwap Assignment2

Akshat Sharma

June 5, 2024

## 1 Introduction

This task involved defining and training a GAN (Generative Adverserial Network) on a dataset of Pokemon images, and displaying the images generated after training.
The architecture employed is that of a Deep Convolutional GAN (DCGAN), comprising two fully convolutional networks: generator and discriminator. The generator is built using transposed convolutional layers, whilst the discriminator uses convolutional layers.

## 2 Network Architecture

Hereafter we describe the architecture of the generator and the discriminator networks.

### 2.1 Generator

The generator takes as input, random latent vectors with 128 channels, having dimensions ($batchsize$, $latentsize$, 1, 1) with $batchsize = latentsize = 128$.
This vector is passed through four transposed convolutional layers, each followed by batch normalization and ReLU activation.
Finally, a transposed convolution is applied along with *tanh* (hyperbolic tangent) activation, to generate an output image of 3 channels, with pixel values between -1 and 1 (thanks to *tanh*).

### 2.2 Discriminator

The discriminator has an input shape of ($batchsize$, 3, 64, 64). The input images are processed using four convolutional layers each followed by Batch Norm and activation (Leaky ReLU), and finally a convolution, flattening and sigmoid activation, resulting in an output between 0 to 1, indicating the probability of the input image being real (*i.e.* belonging to the image dataset, and not generated).

## 3 Training and other details

The loss function and optimizer used in training both, the generator and discriminator, are *binary cross entropy* loss and *Adam* respectively. Training the generator involved adjusting its parameters such that it is able to produce more real-looking images, so that the discriminator gives an output closer to 1 even when the image is generated and not real. In other words, the weights of the generator are so optimized as to maximize the discriminator loss, whilst the discriminator weights are adjusted so as to minimize the same, *i.e.*, to make the discriminator better at differentiating between real and fake images.
The training dataset contains 40597 images of varying shape, each with 3 colour channels. The images are resized to a size of (64,64) and normalized to fall within the range [-1,1], so as to align them with the generated image range.