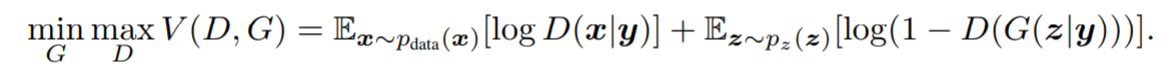
# Face Aging Using Conditional GAN

cGANs are a type of GAN that are conditioned on some extra information. We feed the extra information y to the generator as an additional input layer. In vanilla GANs, there is no control over the category of the generated images. When we add a condition y to the generator, we can generate images of a specific category, using y, which might be any kind of data, such as a class label or integer data. Vanilla GANs can learn only one category and it is extremely difficult to architect GANs for multiple categories. A cGAN, however, can be used to generate multi-modal models with different conditions for different categories.

The architecture of a cGAN is shown in the following diagram:

\*\* INSERT IMAGE OF ARCHITECTURE \*\*

The training objective function for cGANs can be expressed as follows:



Here, G is the generator network and D is the discriminator network. The loss for the discriminator is and the loss for the generator is . We can say the G(z|y) is modeling the distribution of our data given z and y. Here, z is a prior noise distribution

The architecture of a cGAN for face aging is slightly more complicated. The Age-cGan consists of four networks: an encoder, the FaceNet, a generator network, and a discriminator network. With the encoder, we learn the inverse mapping of input face images and the age condition with the latent vector http://localhost:5001/assets/577645cb-a80b-4a10-98e8-5cf21f336b77.png. FaceNet is a face recognition network that learns the difference between an input image x and a reconstructed image http://localhost:5001/assets/a77a78e0-3ea2-4ce3-8f96-deca552c9b9f.png. We have a generator network, which takes a hidden representation consisting of a face image and a condition vector and generates an image. The discriminator network is to discriminate between the real images and the fake images.

The problem with cGANs is that they can't learn the task of inverse mapping an input image x with attributes y to a latent vector z. The solution to this problem is to use an encoder network. We can train an encoder network to approximate the inverse mapping of the input images x.

# The encoder network

The primary goal of the encoder network is to generate a latent vector of the provided images. Basically, it takes an image of a dimension of (64, 64, 3) and converts it into a 100-dimensional vector. The encoder network is a deep convolutional neural network. The network contains four convolutional blocks and two dense layers. Each convolutional block contains a convolutional layer, a batch normalization layer, and an activation function. In each convolutional block, each convolutional layer is followed by a batch normalization layer, except the first convolutional layer

# Face recognition network

The primary goal of the face recognition network is to recognize a person's identity in a given image. For our task, we will be using the pre-trained or ResNet-50 model without fully connected layers. The pre-trained or ResNet-50 network, once provided with an image, returns the corresponding embedding. The extracted embeddings for the real image and the reconstructed image can be calculated by calculating the Euclidean distance of the embeddings

# The generator network

The primary goal of the generator is to generate an image of a dimension of (64, 64, 3). It takes a 100-dimensional latent vector and some extra information, y, andtries to generate realistic images. The generator network is a deep convolutional neural network too. It is made up of dense, upsampling, and convolutional layers. It takes two input values: a noise vector and a conditioning value. The conditioning value is the additional information provided to the network. For the Age-cGAN, this will be the age.

## The discriminator network

The primary goal of the discriminator network is to identify whether the provided image is fake or real. It does this by passing the image through a series of downsampling layers and some classification layers. In other words, it predicts whether the image is real or fake. Like the other networks, the discriminator network is another deep convolutional network. It contains several convolutional blocks. Each convolutional block contains a convolutional layer, a batch normalization layer, and an activation function, apart from the first convolutional block, which doesn't have the batch normalization layer.

## Aging-CGAN TRAINING

The training of the Age-cGAN is made up of three stages:

1. **Conditional GAN training:** In this stage, we train the generator network and the discriminator network.
2. **Initial latent vector approximation:** In this stage, we train the encoder network.
3. **Latent vector optimization:** In this stage, we optimize both the encoder and the generator network.

\*\*\* CREATE YOUR OWN DIAGRAM OF AGING-CGAN TRAINING AND PLACE HERE \*\*\*

## Conditional GAN training

In this stage, we train the generator network and the discriminator network. Once trained, the generator network can generate blurred images of a face. This stage is similar to training a vanilla GAN, in which we train both networks simultaneously.

## The training objective function

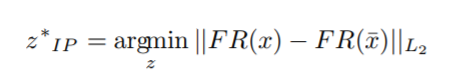
\*\* HANDWRITTEN IMAGE OF THE EQUATION AND EXPLAIN IT \*\*

## Initial latent vector approximation

Initial latent vector approximation is a method to approximate a latent vector to optimize the reconstruction of face images. To approximate a latent vector, we have an encoder network. We train the encoder network on the generated images and real images. Once trained, the encoder network will start generating latent vectors from the learned distribution. The training objective function for training the encoder network is the Euclidean distance loss.

Latent vector optimization

During latent vector optimization, we optimize the encoder network and the generator network simultaneously. The equation we use for latent vector optimization is as follows:



FR is the face recognition network. This equation indicates that the Euclidean distance between the real image and the reconstructed images should be minimal. In this stage, we try to minimize the distance to maximize identity preservation.