Report project 4

Daniel Toribio, Alejo González and Andrés Navarro

Introduction

In this project we have implemented a VAE with images from the dataset CelebA, and in the optional part with a 3-dimensional Gaussian Mixture Model with 10 components. Variational Autoencoders (VAEs) are a class of generative models particularly effective for tasks like image generation.

A VAE consists of an encoder and a decoder network. The encoder takes input data and maps it to a latent space representation, while the decoder takes samples from the latent space and reconstructs the original input data. The key innovation of VAEs lies in their ability to learn a probabilistic distribution over latent variables, allowing for the generation of diverse and realistic samples.

Some images from the CelebA dataset

Encoder

The encoder is used to obtain both $\mu\eta(x)$, $\sigma\eta(x)$, which determine the moments of the approximate posterior distribution $q(z|x)$. This neural network consist of 2 dimensions convolutional layers and a final linear layer with dimz*2 outputs, where the first k elements represent the mean, and we compute the variance from the last k elements using a soft-plus.

Decoder

The decoder is used to obtain the $\mu\theta(z)$, with transpose convolution layers. To evaluate the log-likelihood of an independent Gaussian distribution given the mean and the diagonal covariance matrix we had implemented a function.

Variational Autoencoder

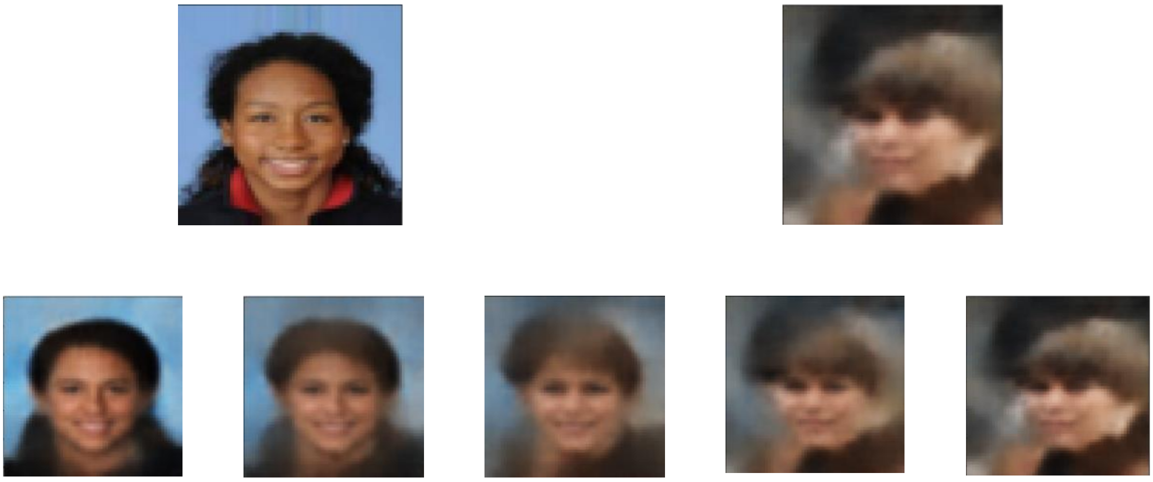
Once we have the encoder and decoder networks, we can create the VAE. The loss function is the ELBO lower bound. Before training the model, we have used pre-trained parameters with dim_z=50 trained for 200 epochs with var_x=0.1, lr=1e-3. The following image shows 20 generated images:

20 generated images from the CelebA dataset

We can see that the images have realistic details, and the model tries to vary the features of the images such as the smile, hair or background. We can also see the reconstruction of a real image:

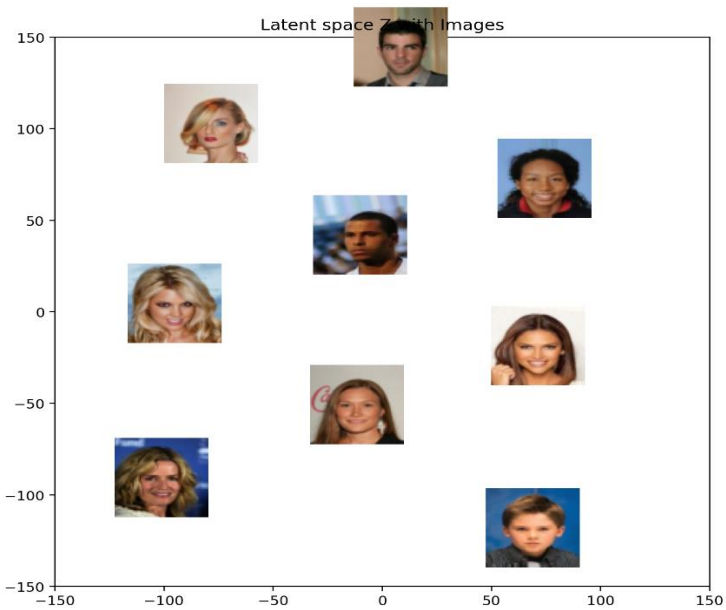
Image reconstruction by the VAE

We can see that the VAE reconstructs the image with some distortion but the main features such as the face and the hair are preserved. We can also interpolate between two images using their latent representations:



Interpolation of two images

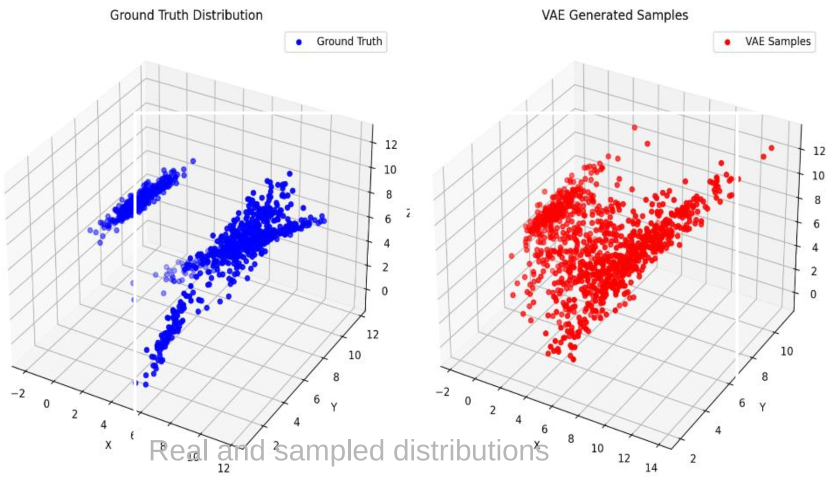
Here we can observe the representation in the latent space of the images, where the blonde people is placed on the left part.



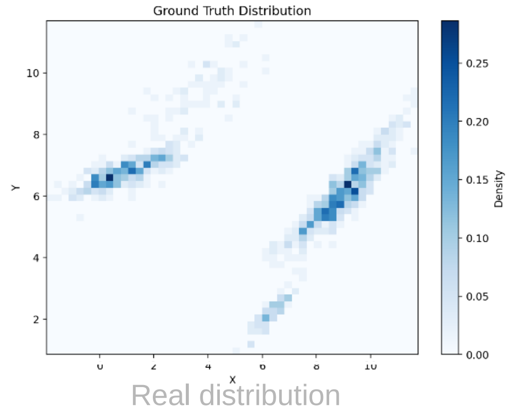
Latent representation of images

Optional part

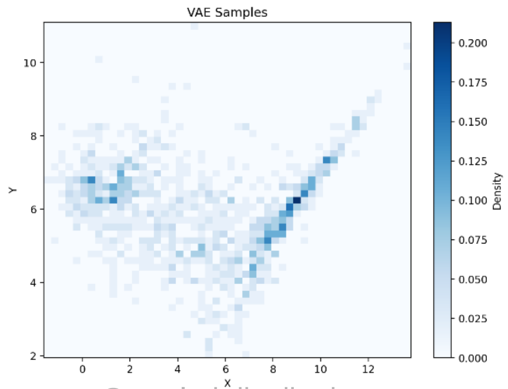
In this part we have used a VAE to sample from a 3-dimensional GMM. The results are displayed in the following images:



Real and sampled distributions

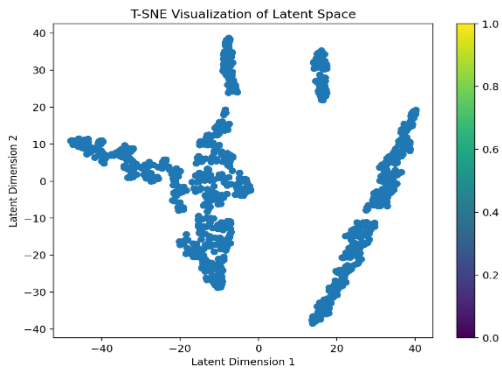


Real distribution



Sampled distribution

We can see that two components are effectively captured. Now we have used T-SNE to visualize the GMM samples in the latent space, where we can guess that we have between 4 to 5 clusters.



GMM samples in the latent space

Conclusions

In this project, a Variational Autoencoder (VAE) was implemented using images from the *CelebA* dataset, with an optional exploration of a 3-dimensional Gaussian Mixture Model (GMM). VAEs, comprising encoder and decoder networks, enable the learning of probabilistic distributions over latent variables, facilitating diverse and realistic image generation. The encoder extracts mean and variance values for the approximate posterior distribution, while the decoder reconstructs input data from samples in the latent space. The VAE successfully generated realistic images and preserved key features during reconstruction and interpolation tasks. Additionally, the exploration of a GMM demonstrated effective capture of components and clustering in the latent space, showing the versatility of VAEs for generative modelling tasks.