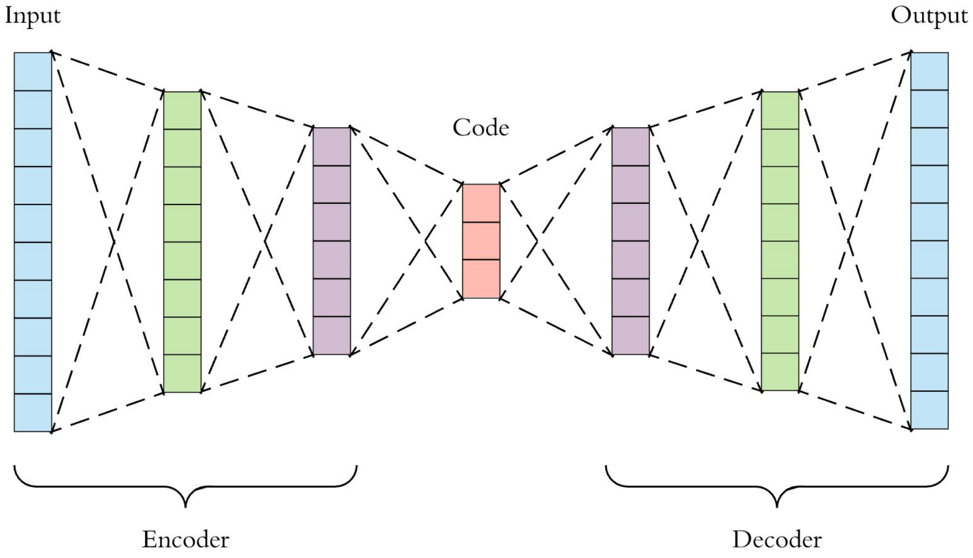
Autoencoders

Master InterUniversitario de Data Science
Santander, Spain
March 2019

Ignacio Heredia
iheredia@ifca.unican.es
Instituto de Física de Cantabria
(CSIC-UC)

Classical Autoencoders

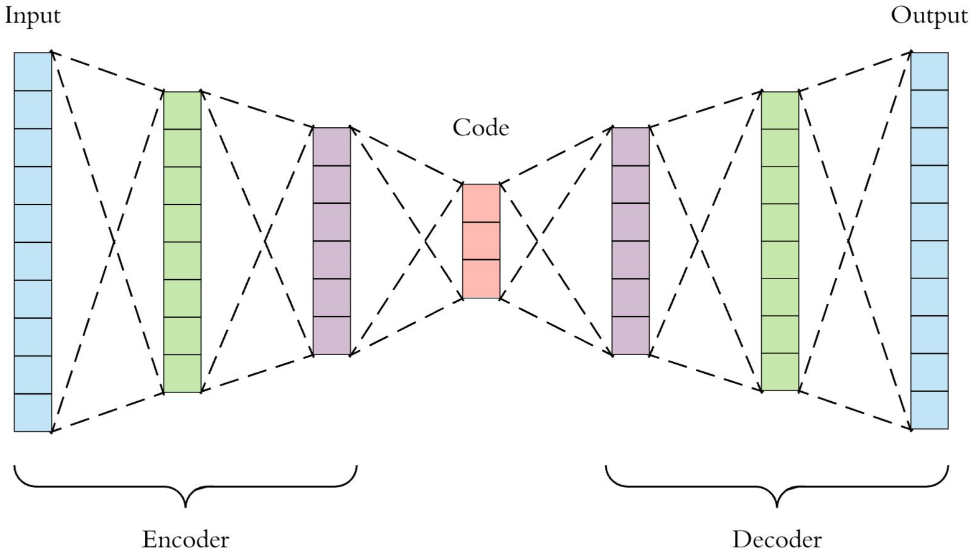
- Autoencoders (AE) are usually described as an unsupervised algorithm (no labels are needed for the training data) although they are more accurately a self-supervised algorithm (labels are automatically generated from inputs).
- The task during training is to reconstruct the input after having compressed it.



source

Classical Autoencoders

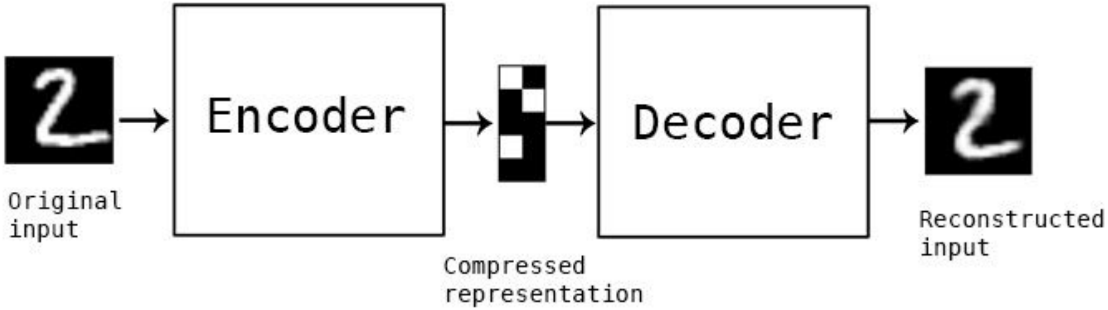
- An additional objective is to learn a compressed representation of your data.
- To build an autoencoder we need to define: an encoder function, a decoder function and a loss function.



source

Application to images

- We learn to compress an 3D RGB image to an 1D vector using a convolutional AE.

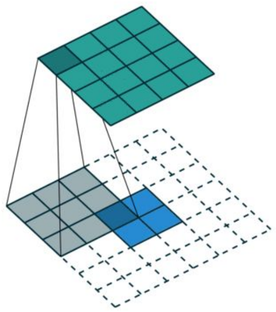


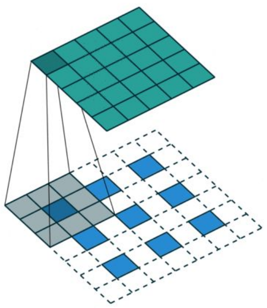
- **Encoder:** convolutional layers, pooling layers
- **Decoder:** transposed convolutional (or deconvolutional) layer, unpooling layers.

source

Application to images - Decoder layers

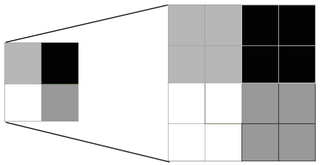
Possible transposed convolutions



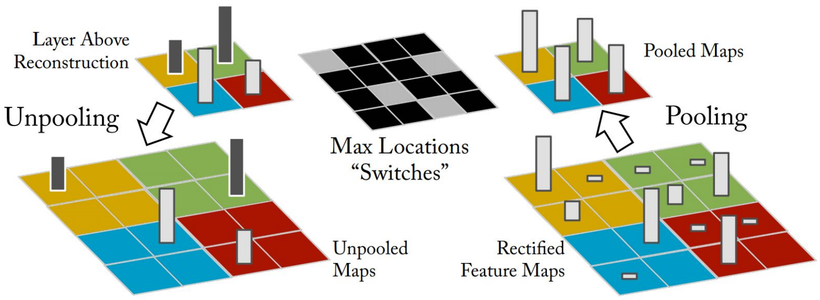


Animated GIF

Possible unpoolings



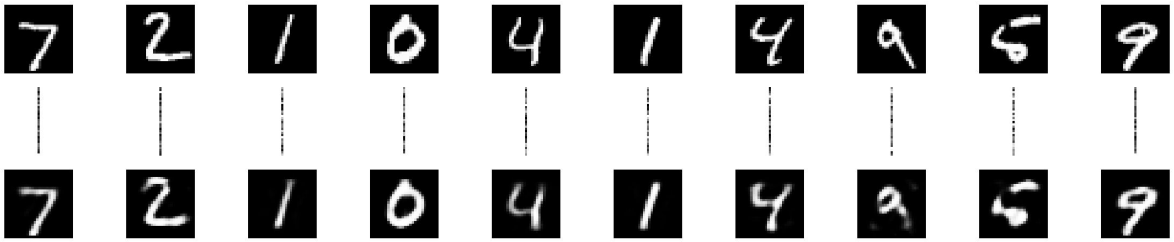
source



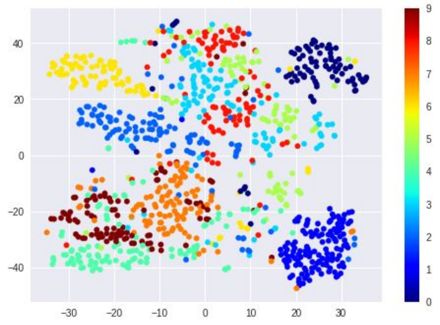
Paper: Visualizing and Understanding Convolutional Networks

Exercice 1 - Encoding MNIST

- Create a shallow autoencoder to encode MNIST data, using Dense layers (reshape input image to vector). *Optional: add sparsity*.
- Visualize the results of encoding and decoding of test data.

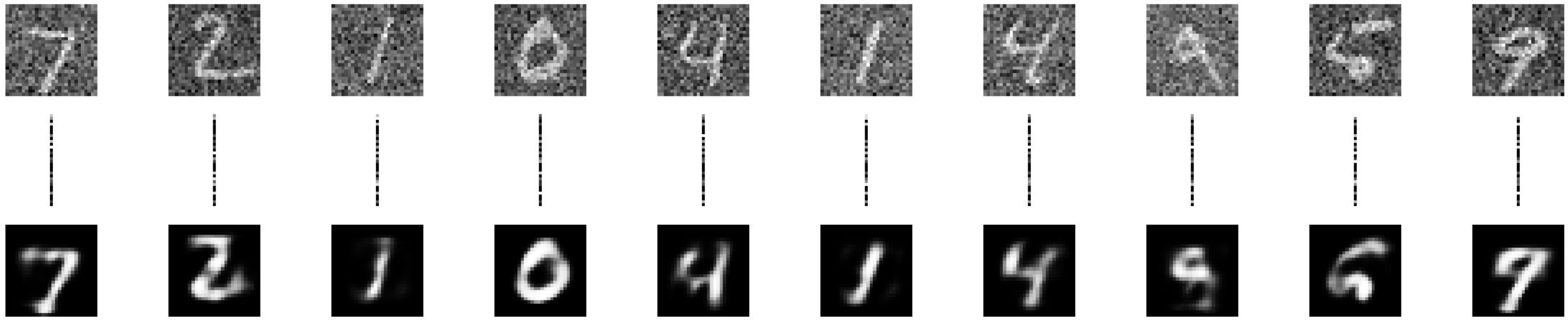


- Visualize the results of encoding of test data using the t-SNE algorithm.



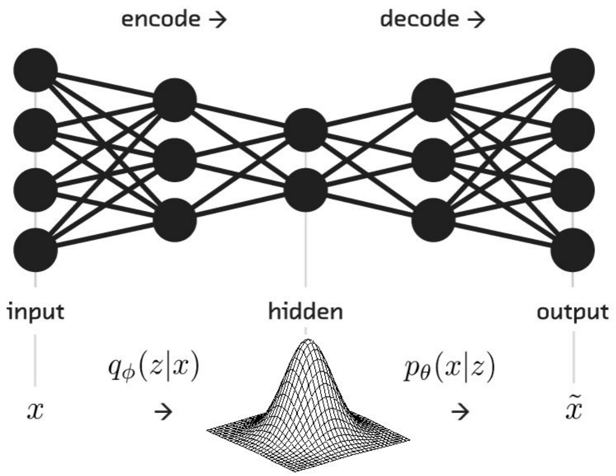
Exercice 2 - Denoising

- Use the previous model to create a denoising application.



Variational Autoencoders

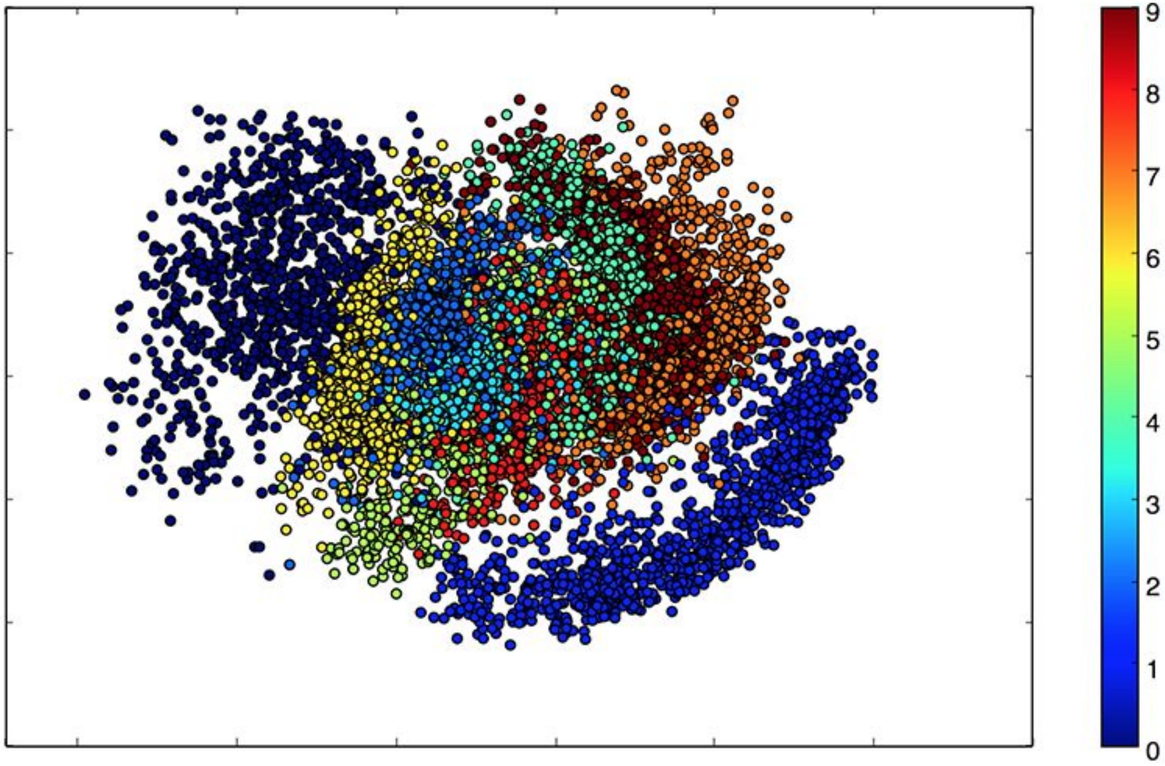
- Variational Autoencoders (VAE) are like AE but with added constraints.
- So instead of learning an arbitrary function to encode the input, you are learning the parameters of a probability distribution modeling your data.
- If you sample points from this distribution, you can generate new input data samples: a VAE is a "generative model".



source

Variational Autoencoders - Latent space

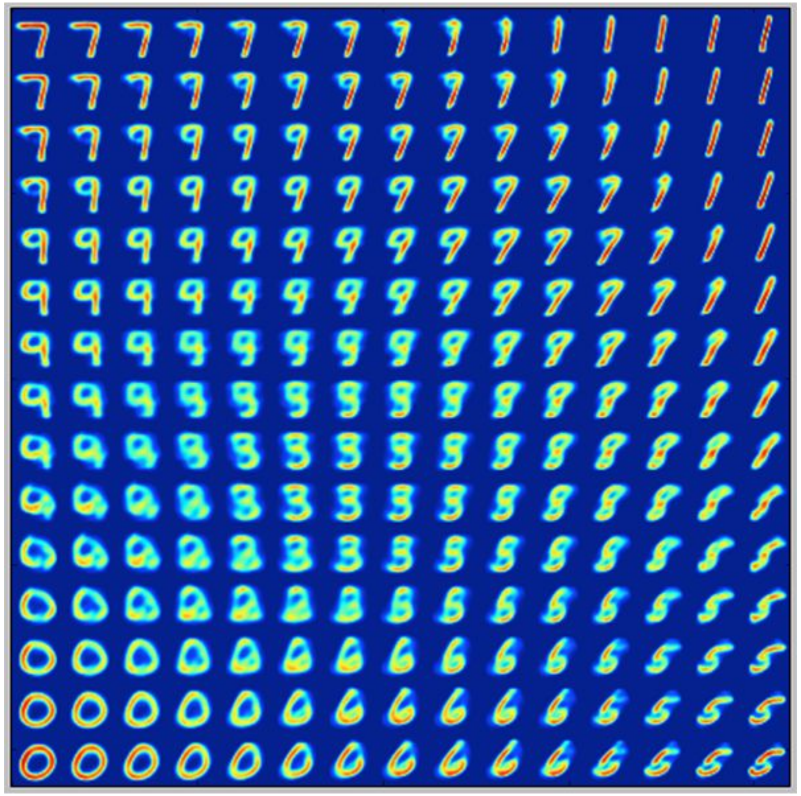
- 2D latent space (**not t-SNE**) Animated GIF



source

Variational Autoencoders - Latent space

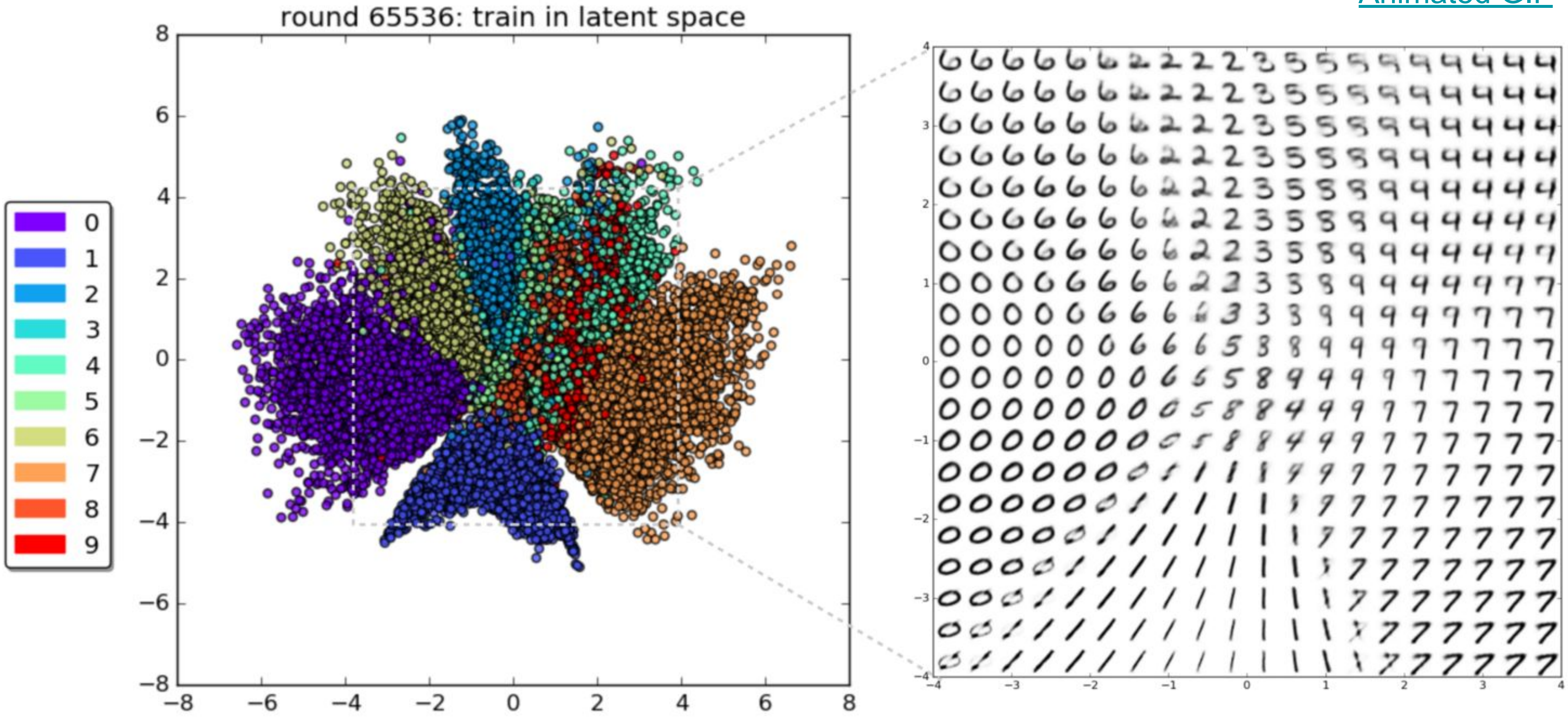
- Reconstruction from latent coordinates into the original data space.



Animated GIF

Variational Autoencoders - Latent space

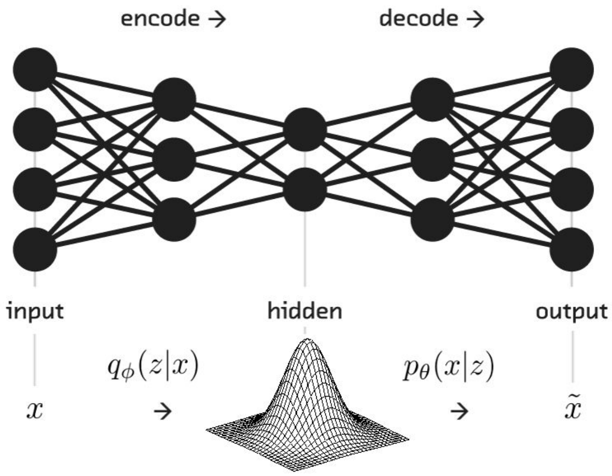
[Animated GIF]

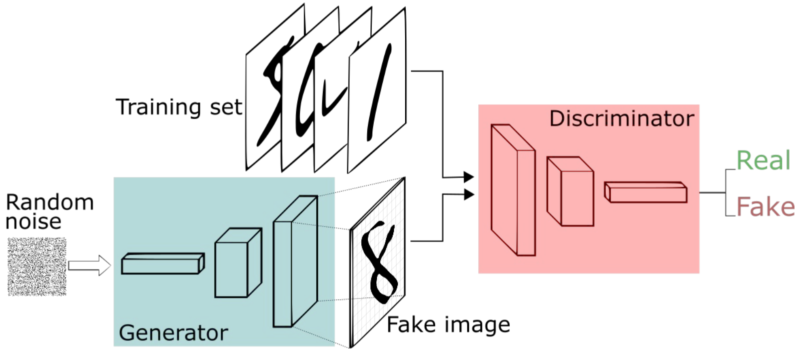


VAEs vs **GANs**

- Variational Autoencoders (VAE) are good at learning representations (interpretable dimensions, possibility to set complex priors).

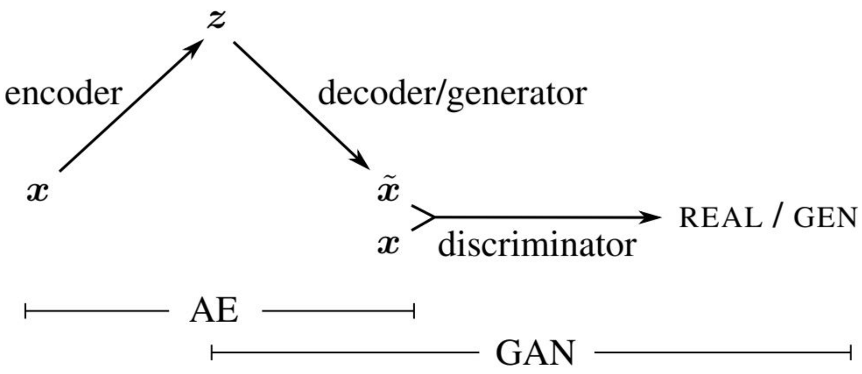
- Generative Adversarial Networks (GANs) are good at generating new samples (clever loss). Trickier to train.

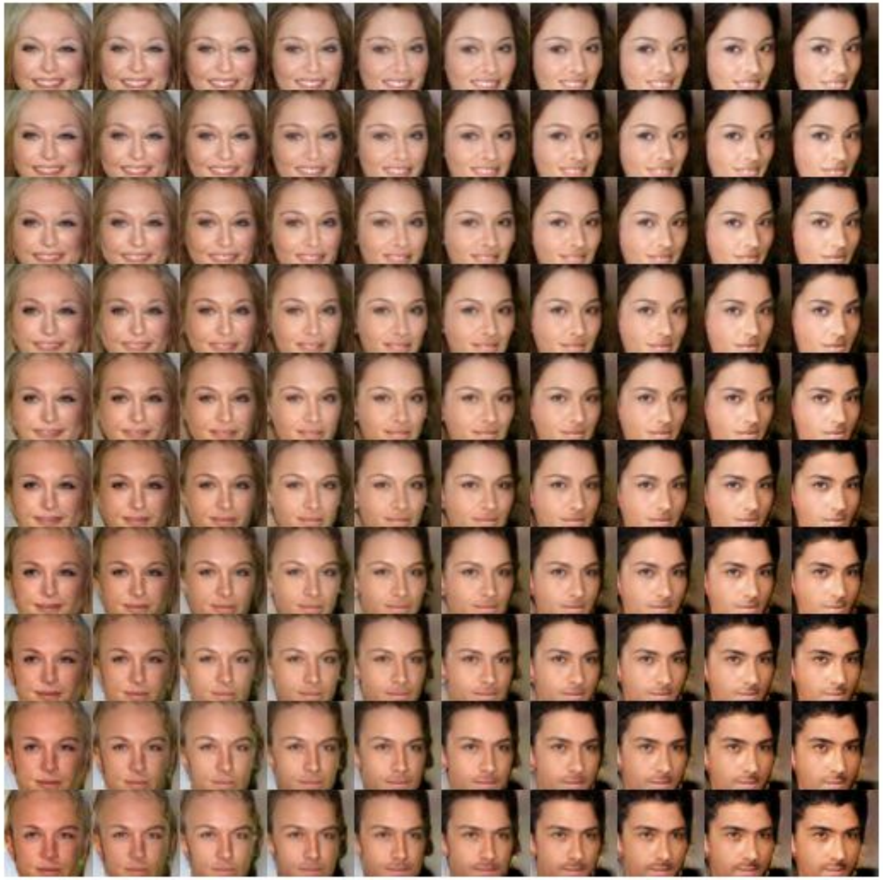




VAEs vs **GANs**

- They can be combined to get the better of both worlds.





source

Summary of applications

- Dimensionality reduction (and clustering if we apply k-means for example after the reduction).
- Data denoising
- Data generation (VAE)

More info

- Building Autoencoders in Keras
- Variational Autoencoders Part 1 + Part 2