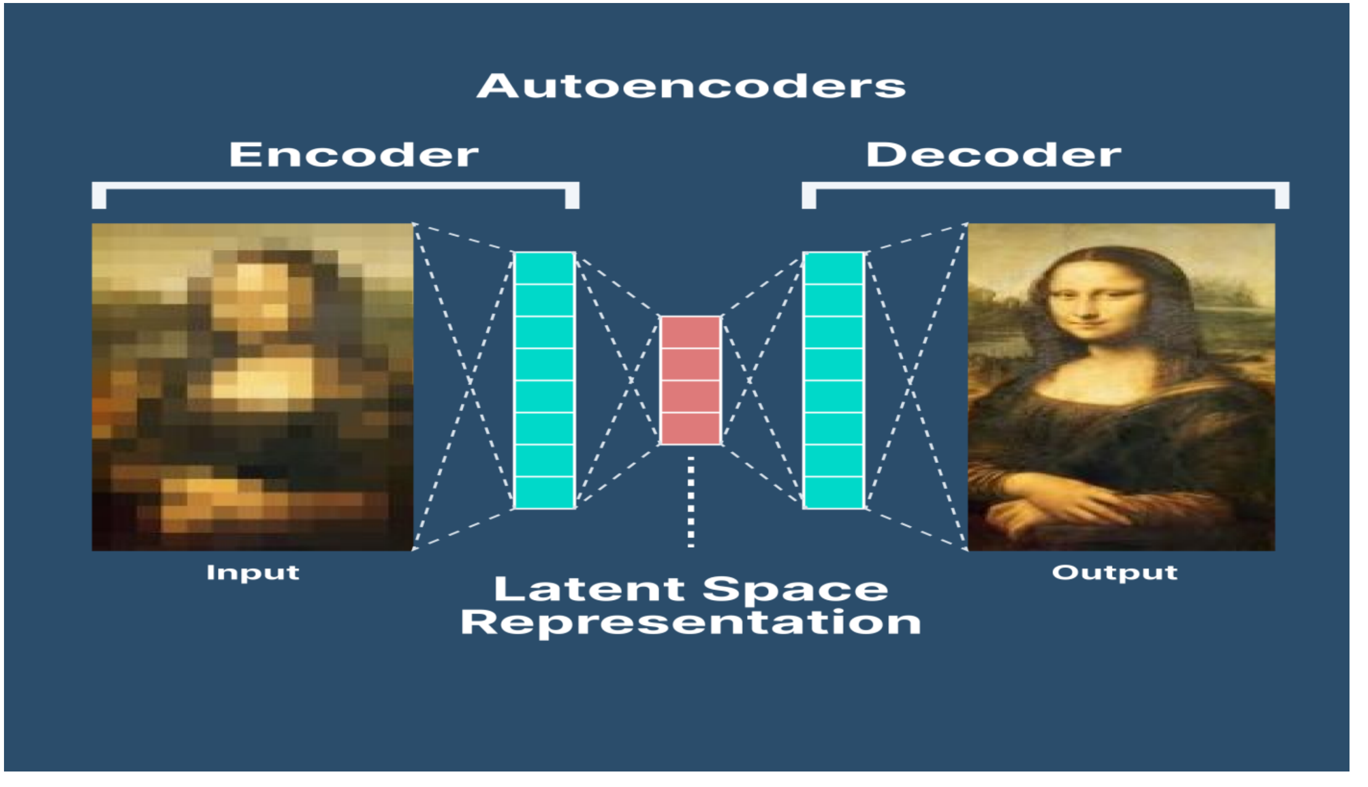
Autoencoder

By

Prof(Dr.)Premanand P. Ghadekar



https://www.v7labs.com/blog/autoencoders-guide

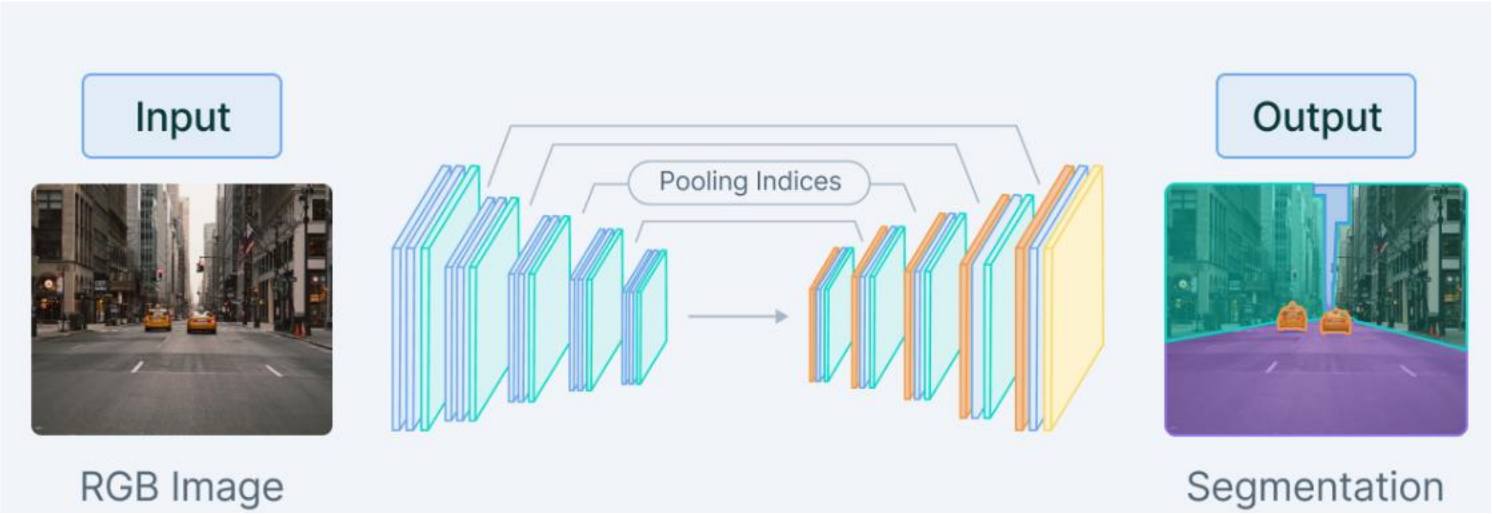
Outline

- Introduction
- Autoencoder
- Working
- Applications etc

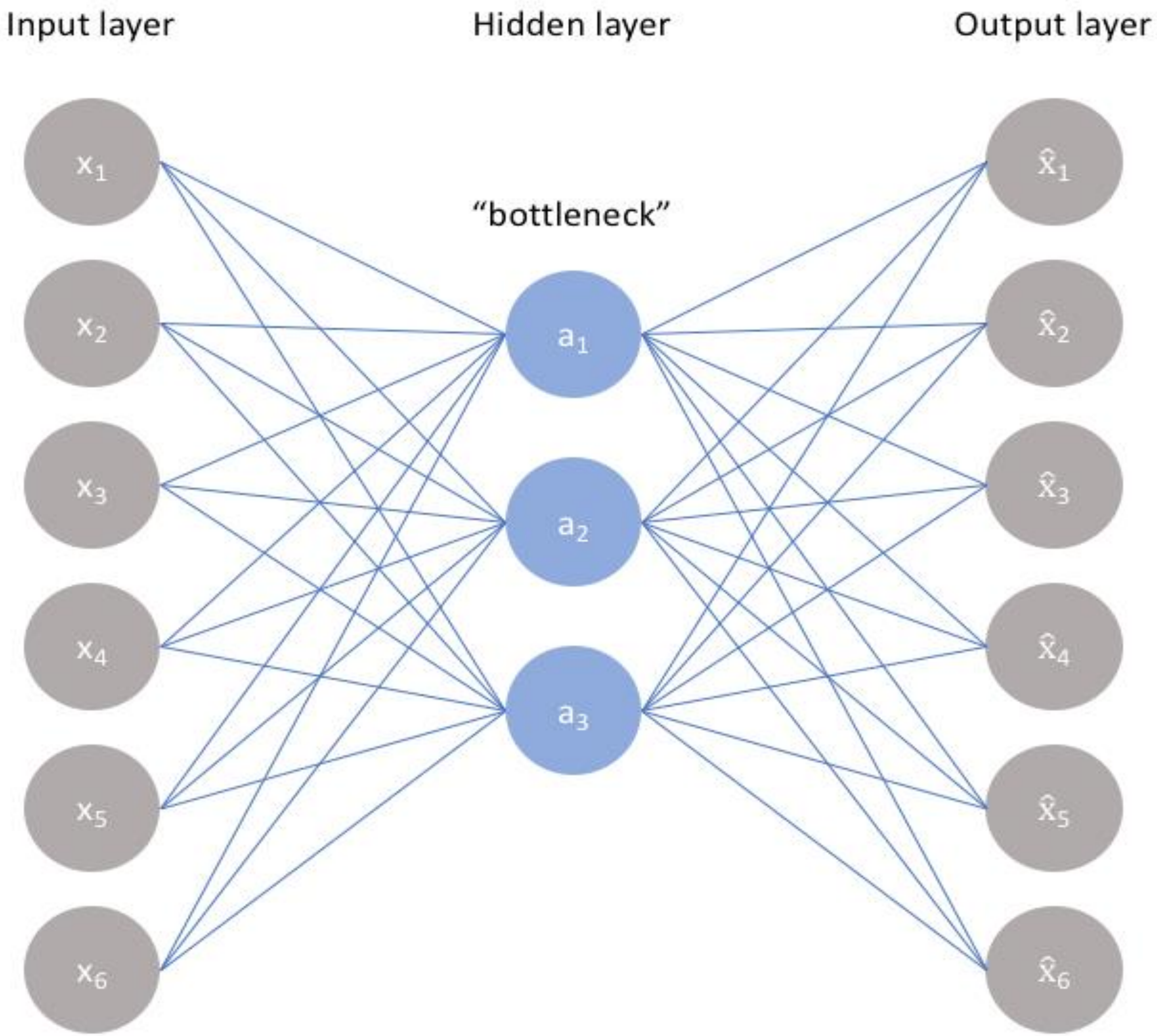
What is an Autoencoder?

❖ An autoencoder is a type of artificial neural network used **to learn data encodings in an unsupervised manner.**

❖ The aim of an autoencoder is to learn a lower-dimensional representation (encoding) for a higher-dimensional data, typically for dimensionality reduction, by training the network to capture the most important parts of the input image. **Convolutional Encoder Decoder**



The Architecture of Autoencoders



The relationship between the Encoder, Bottleneck, and Decoder

Encoder

The encoder is a set of convolutional blocks followed by pooling modules that compress the input to the model into a compact section called the bottleneck.

Bottleneck

The most important part of the neural network, and ironically the smallest one, is the bottleneck. The bottleneck exists to restrict the flow of information to the decoder from the encoder, thus, **allowing only the most vital information to pass through.**

Decoder

Finally, the decoder is a **set of upsampling and convolutional blocks that reconstructs the bottleneck's output.**

Since the input to the decoder is a compressed knowledge representation, the decoder serves as a “decompressor” and builds back the image from its latent attributes.

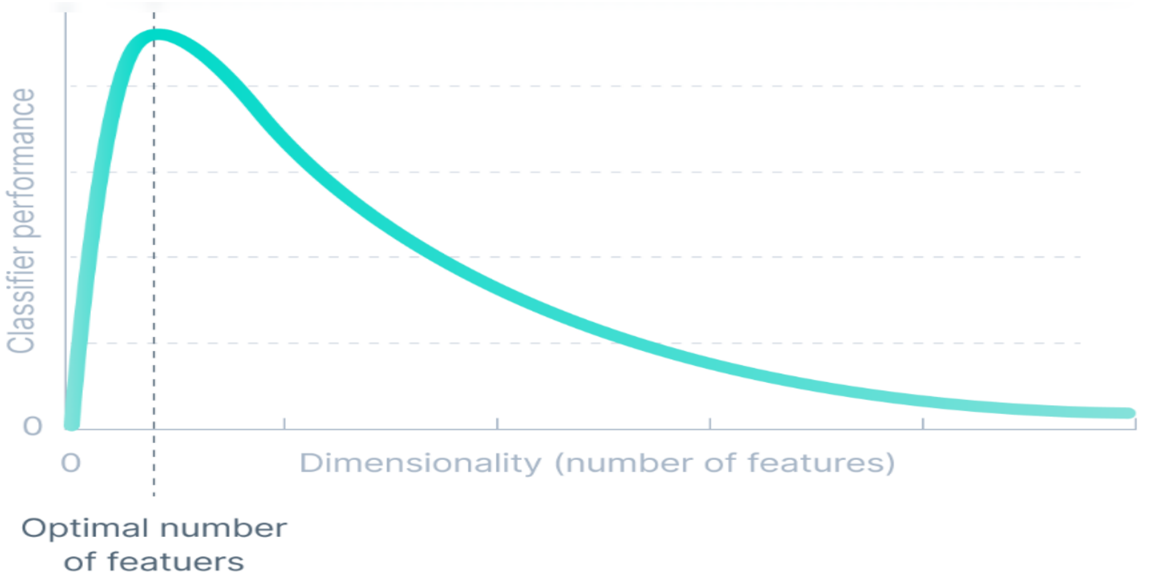
Types of Autoencoders

Five Popular Autoencoders are:

1. Undercomplete autoencoders
2. Sparse autoencoders
3. Contractive autoencoders
4. Denoising autoencoders
5. Variational Autoencoders (for generative modelling)

Undercomplete Autoencoders

- One of the simplest types of Autoencoders.
- Undercomplete autoencoder takes in an image and tries to predict the same image as output, thus reconstructing the image from the compressed bottleneck region.
- Undercomplete autoencoders are truly unsupervised as they do not take any form of label, the target being the same as the input.
- The primary use of autoencoders like such is the generation of the latent space or the bottleneck, which forms a compressed substitute of the input data and can be easily decompressed back with the help of the network when needed.
- This form of compression in the data can be modeled as a form of dimensionality reduction.



Undercomplete Autoencoders

- PCA can only build linear relationships. As a result, it is put at a disadvantage compared with methods like undercomplete autoencoders that can learn non-linear relationships and, therefore, perform better in dimensionality reduction.
- This form of nonlinear dimensionality reduction where the autoencoder learns a non-linear manifold is also termed as **manifold learning.**
- Effectively, if we remove all non-linear activations from an undercomplete autoencoder and use only linear layers, we reduce the undercomplete autoencoder into something that works at an equal footing with PCA.

Undercomplete Autoencoders

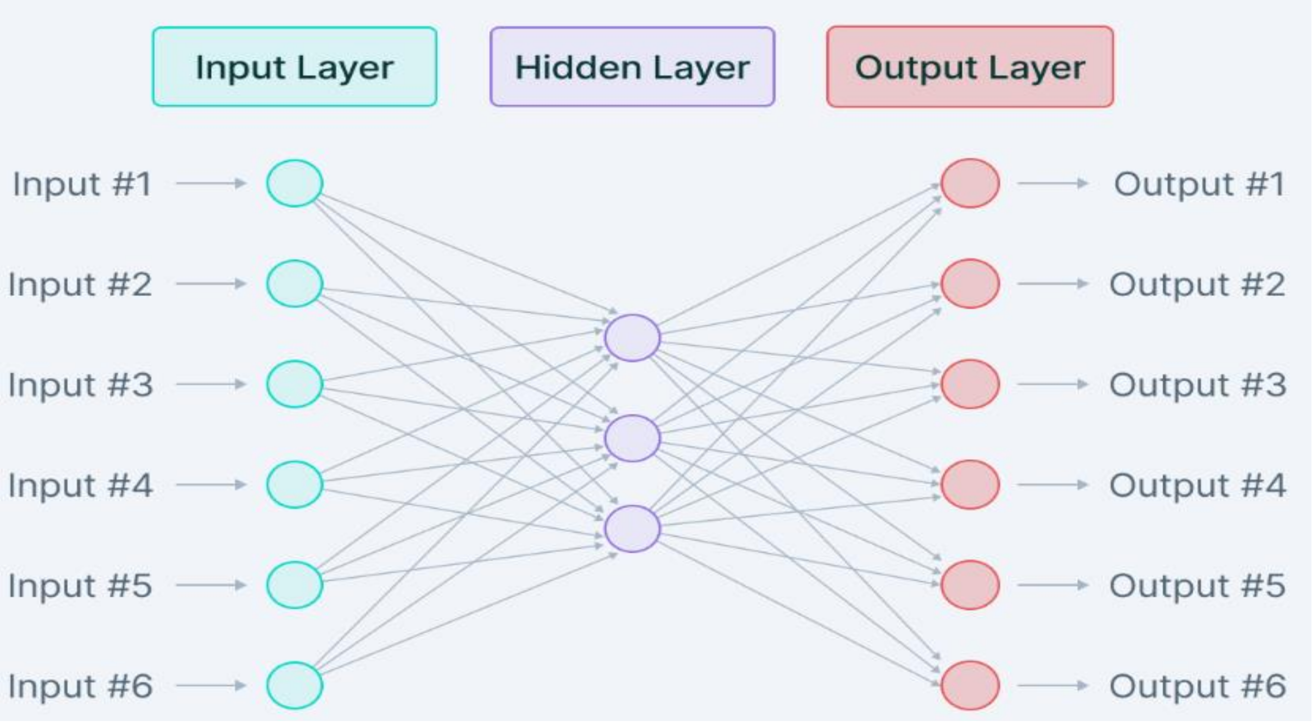
- The loss function used to train an undercomplete autoencoder is called **reconstruction loss**, as it is a check of how well the image has been reconstructed from the input.
- Although the reconstruction loss can be anything depending on the input and output, we will use an L1 loss to depict the term (also called the norm loss) represented by:

$$L = |x - \hat{x}|$$

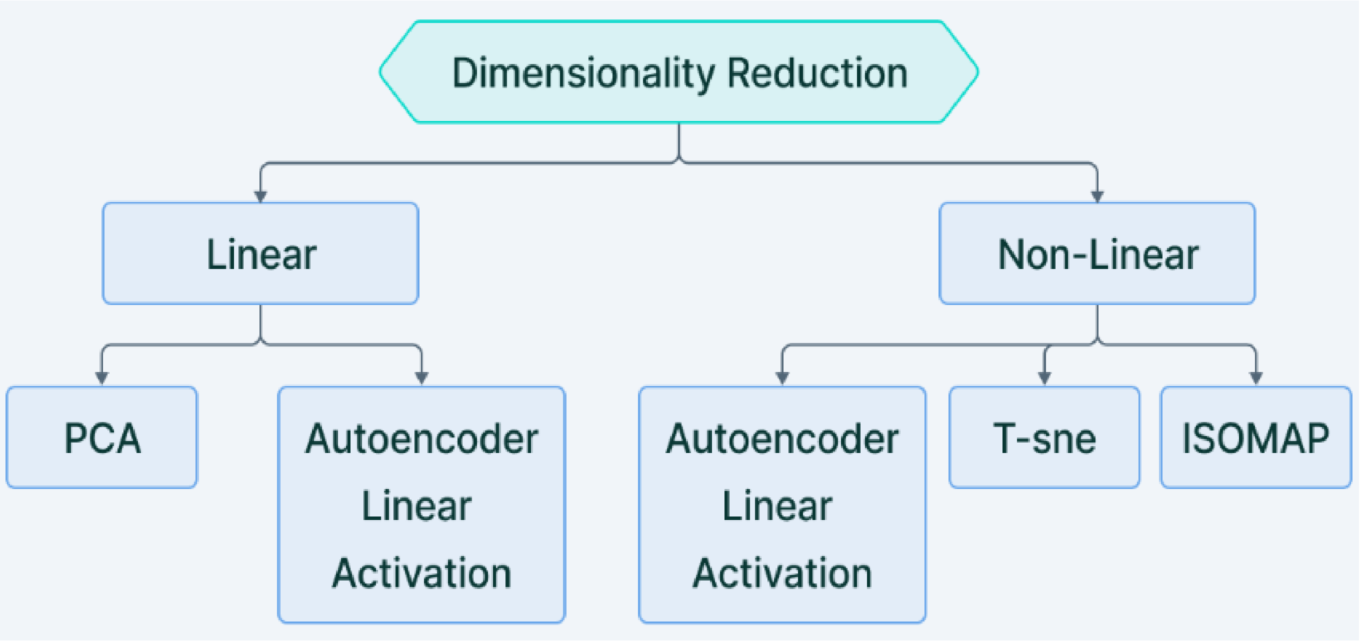
Where $\hat{x}$ represents the predicted output and x represents the ground truth.

Sparse Autoencoders

- Sparse autoencoders are similar to the undercomplete autoencoders in that they use the same image as input and ground truth. However—
- The means via which encoding of information is regulated is significantly different.



Sparse Autoencoders

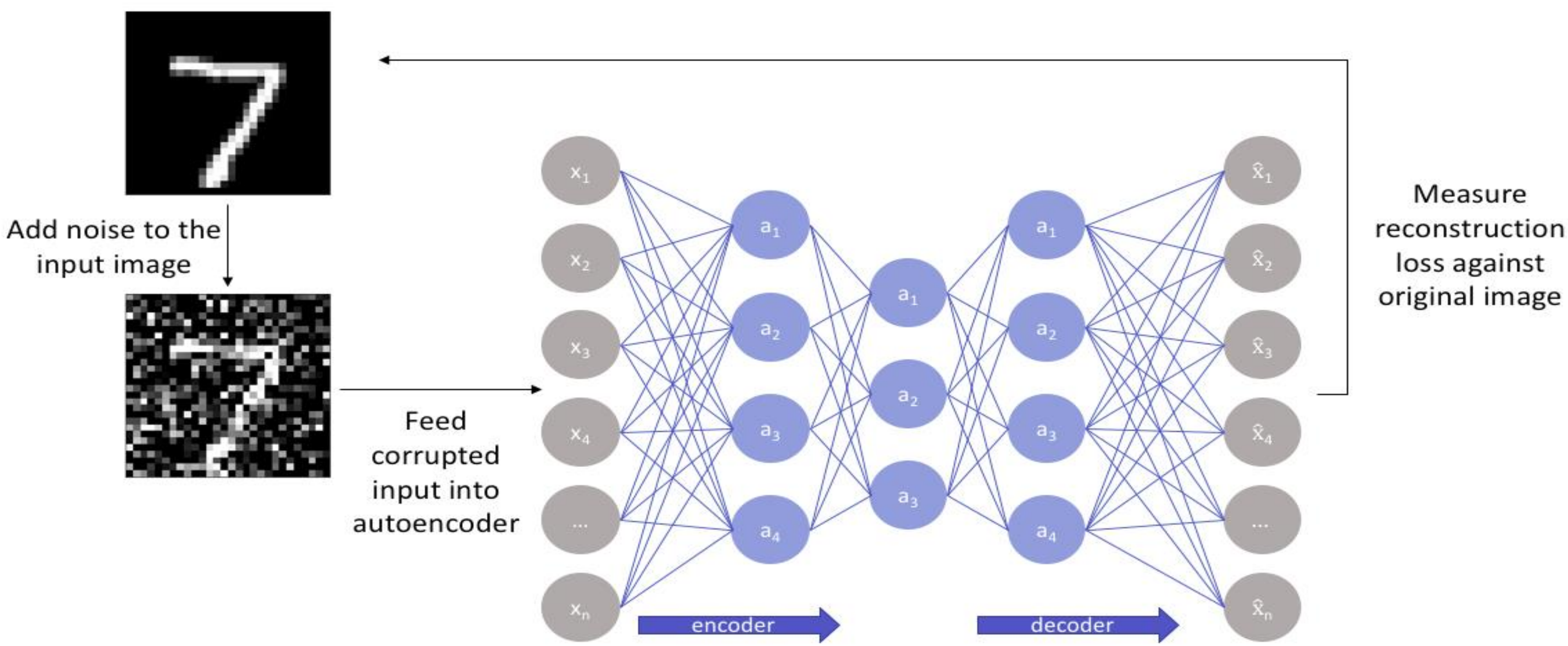


Denoising Autoencoders

Denoising autoencoders, as the name suggests, are autoencoders that remove noise from an image.

As opposed to autoencoders we’ve already covered, this is the first of its kind that does not have the input image as its ground truth.

In denoising autoencoders, we feed a noisy version of the image, where noise has been added via digital alterations. The noisy image is fed to the encoder-decoder architecture, and the output is compared with the ground truth image.



Denoising Autoencoders

The denoising autoencoder gets rid of noise by learning a representation of the input where the noise can be filtered out easily.

While removing noise directly from the image seems difficult, the autoencoder performs this by mapping the input data into a lower-dimensional manifold (like in undercomplete autoencoders), where filtering of noise becomes much easier.

Essentially, denoising autoencoders work with the help of non-linear dimensionality reduction.

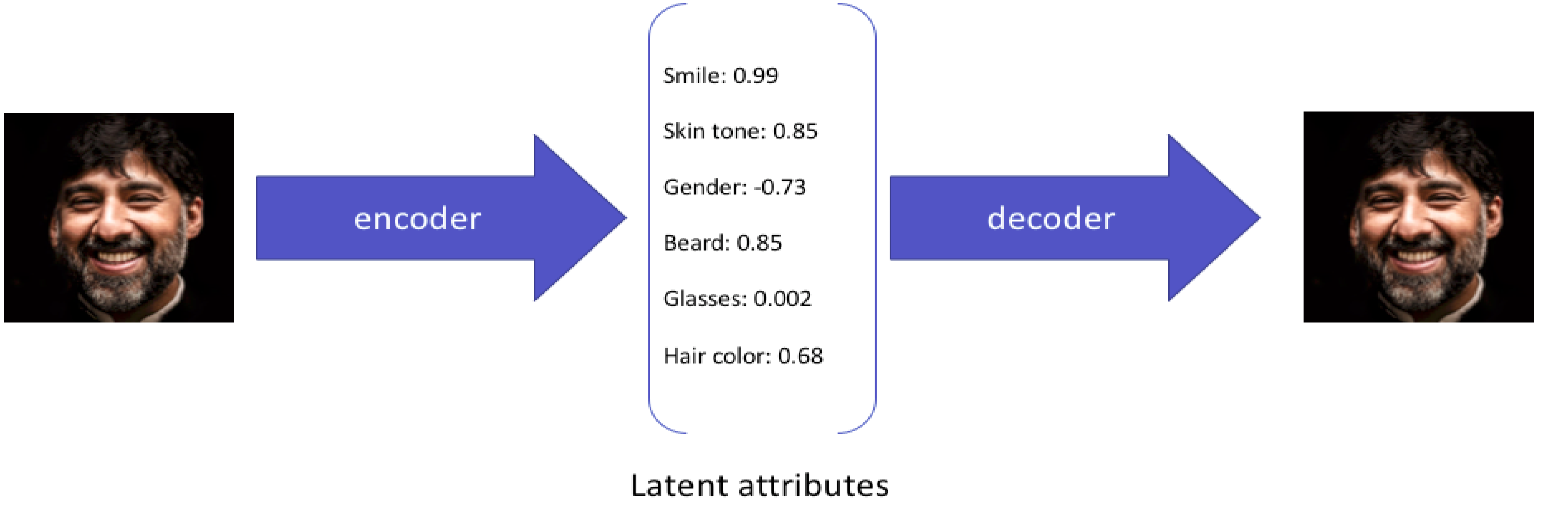
Variational Autoencoders

Standard and variational autoencoders learn to represent the input just in a compressed form called the latent space or the bottleneck.

Therefore, the latent space formed after training the model is not necessarily continuous and, in effect, might not be easy to interpolate.

For example—

This is what a variational autoencoder would learn from the input:

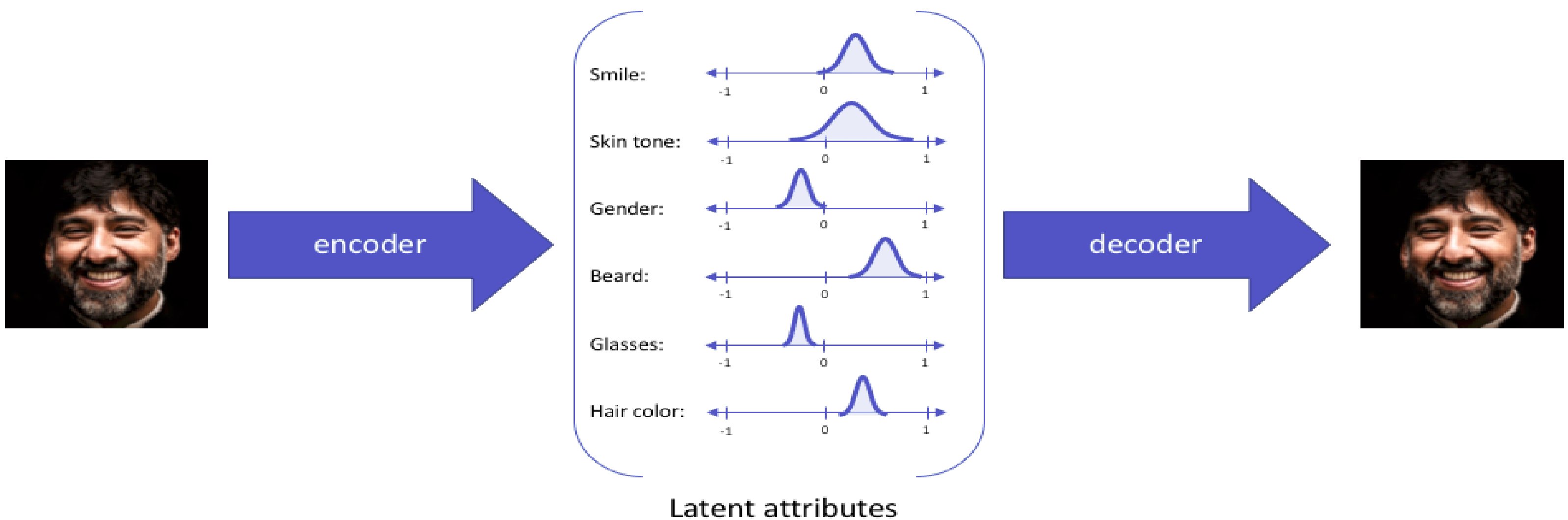


Variational Autoencoders

While these attributes explain the image and can be used in reconstructing the image from the compressed latent space, they do not allow the latent attributes to be expressed in a probabilistic fashion.

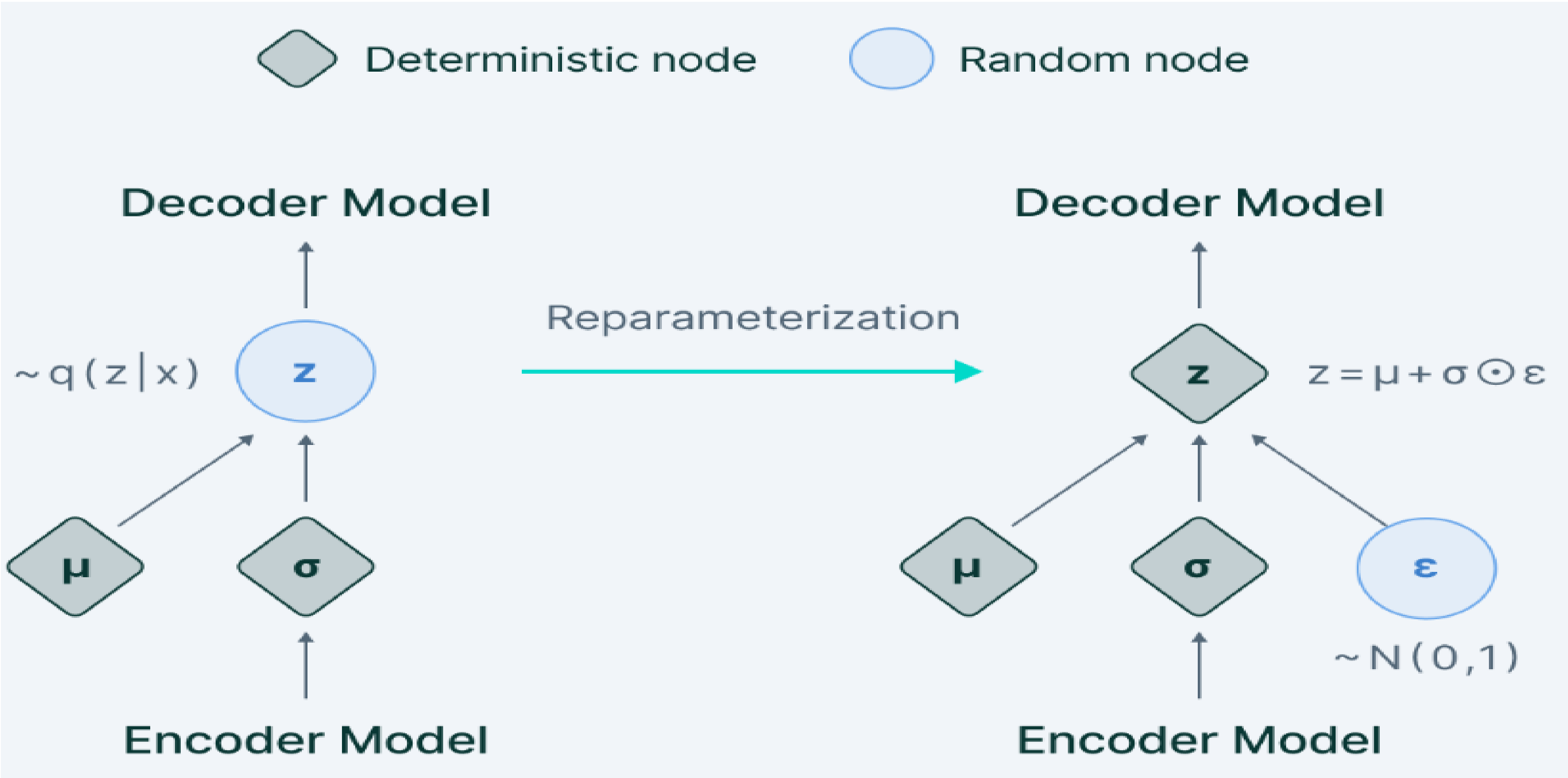
Variational autoencoders deal with this specific topic and express their latent attributes as a probability distribution, leading to the formation of a continuous latent space that can be easily sampled and interpolated.

When fed the same input, a variational autoencoder would construct latent attributes in the following manner:



Variational Autoencoders

A diagrammatic view of what we attain can be expressed as:



Variational Autoencoders

The summarised loss function can be expressed as:

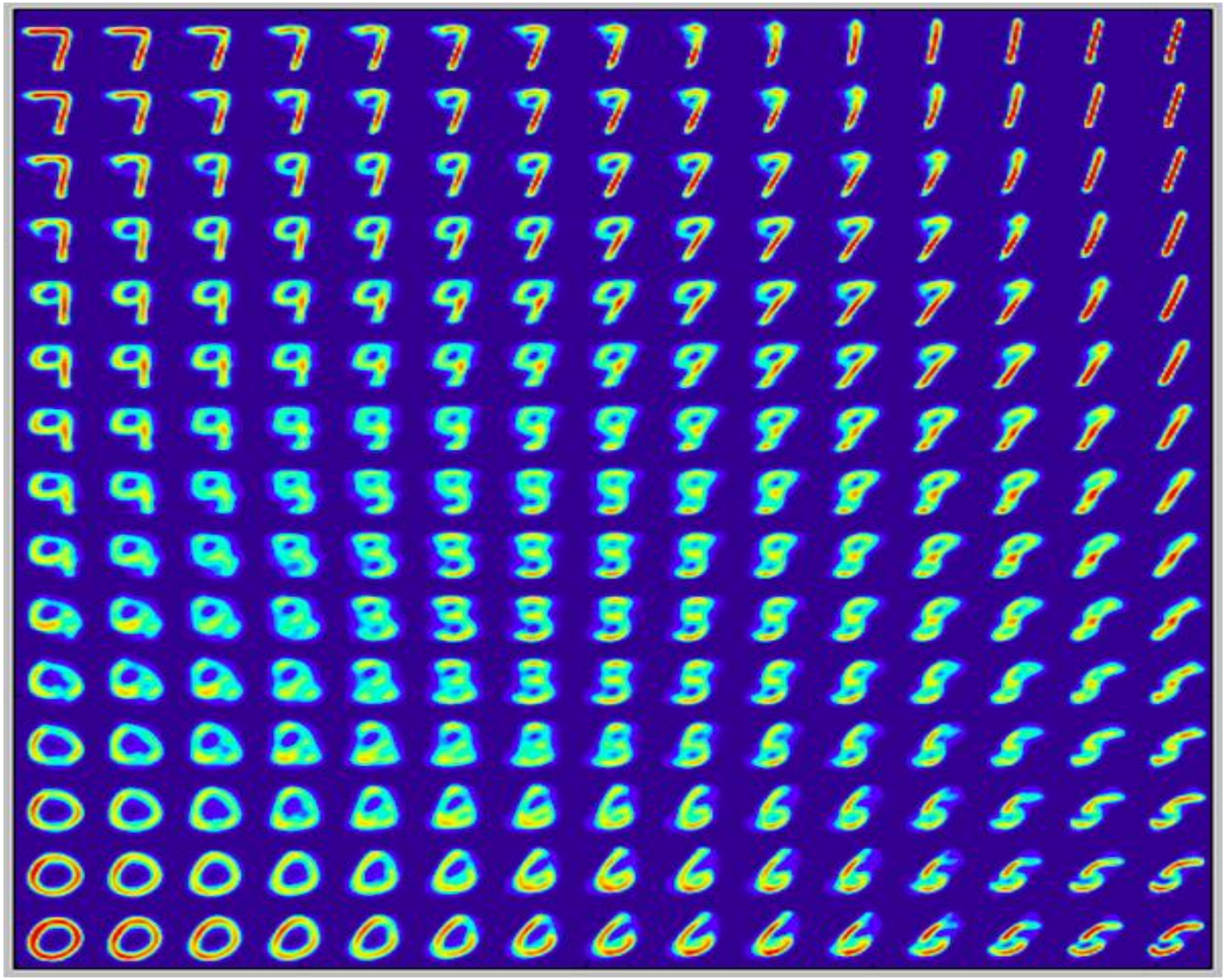
$$L = |x - \hat{x}| + \beta \sum_{i} KL(q_j(z|x) || \mathbb{N}(0, 1))$$

Where $\mathbb{N}$ denotes the unit normal distribution and β denotes a weighting factor.

The primary use of variational autoencoders can be seen in generative modeling. Sampling from the latent distribution trained and feeding the result to the decoder can lead to data being generated in the autoencoder.

Variational Autoencoders

A sample of MNIST digits generated by training a variational autoencoder is shown below:

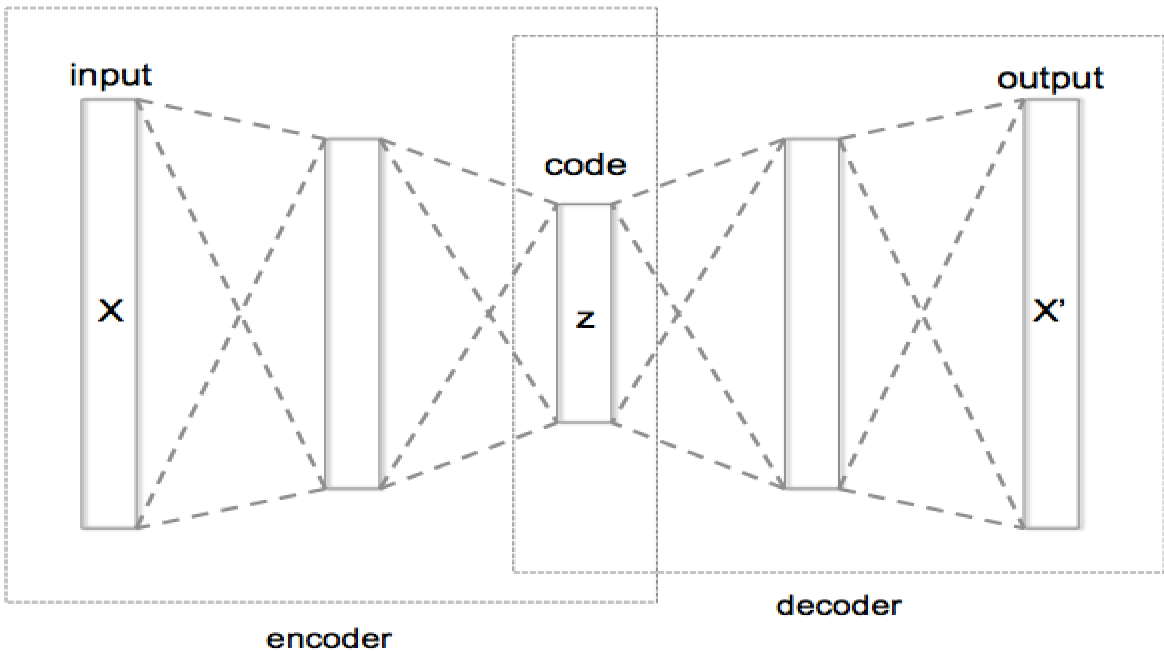


Applications of autoencoders

1. Dimensionality Reduction
2. Image denoising
3. Image Search
4. Anomaly Detection

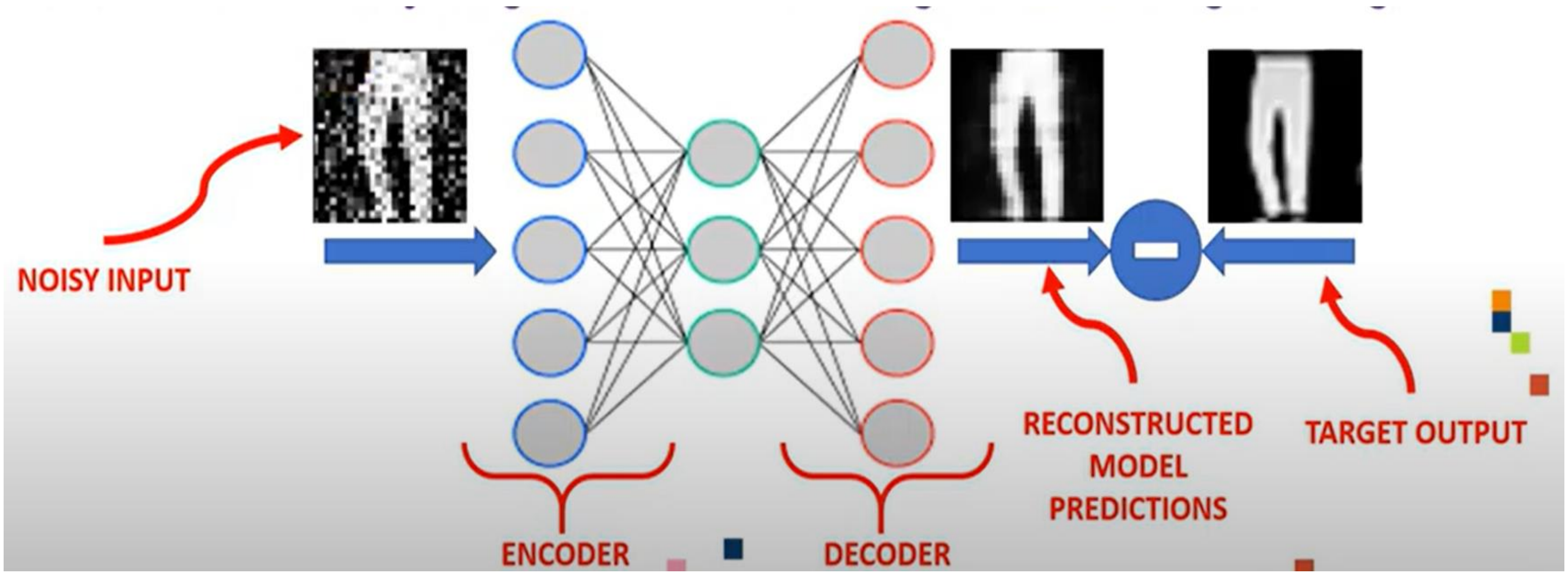
Dimensionality Reduction

1. AutoEncoder is an unsupervised Artificial Neural Network that attempts to encode the data by compressing it into the lower dimensions (bottleneck layer or code) and then decoding the data to reconstruct the original input. The bottleneck layer (or code) holds the compressed representation of the input data.



Denoising Autoencoder

1. One important application of Auto encoders is to perform denoising Operation.
2. Noisy image is used as an Input and Out put will be Enhanced Image.



Denoising Autoencoder

MNIST Fashion Dataset

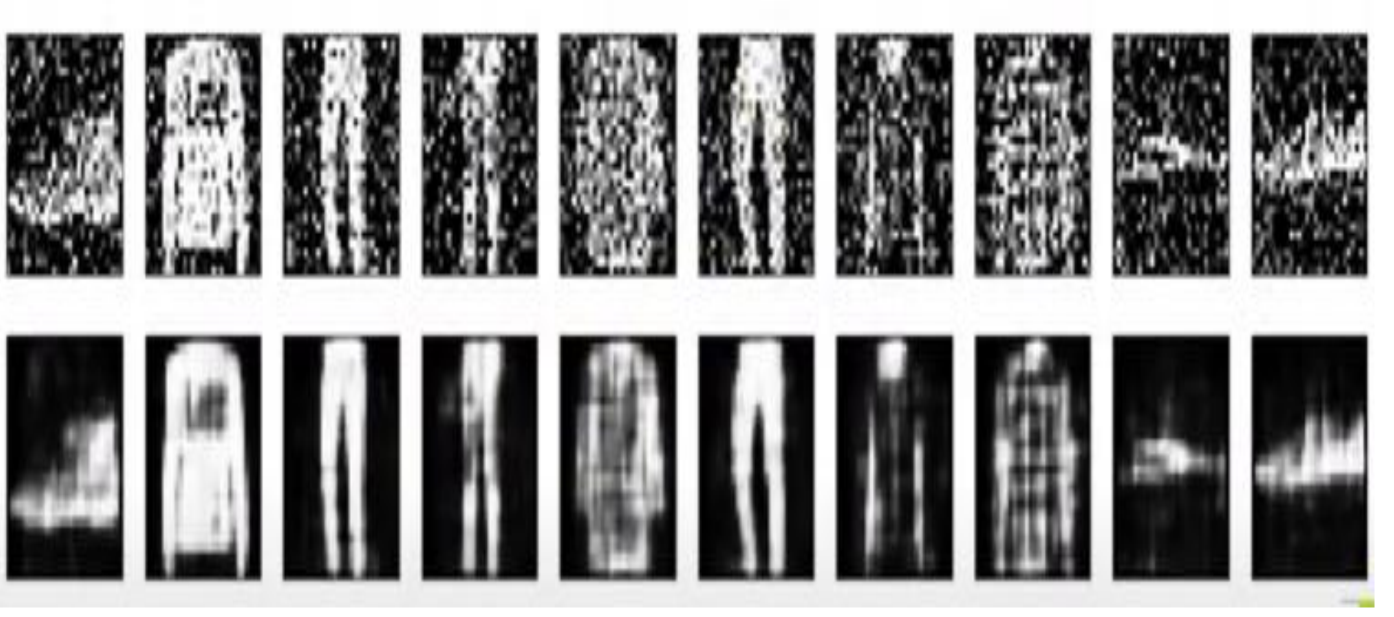


Image Compression

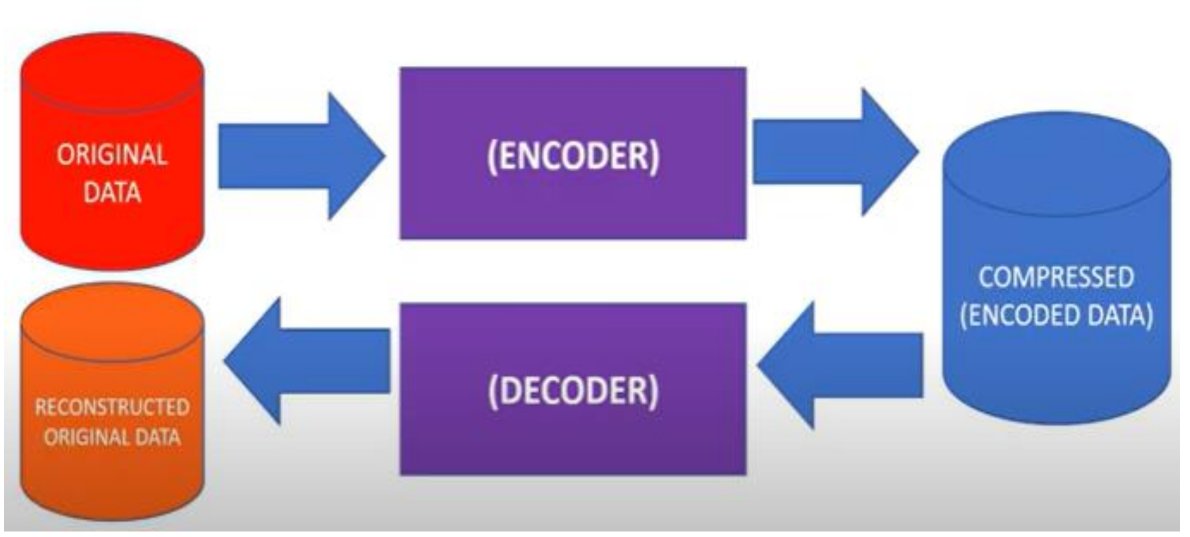
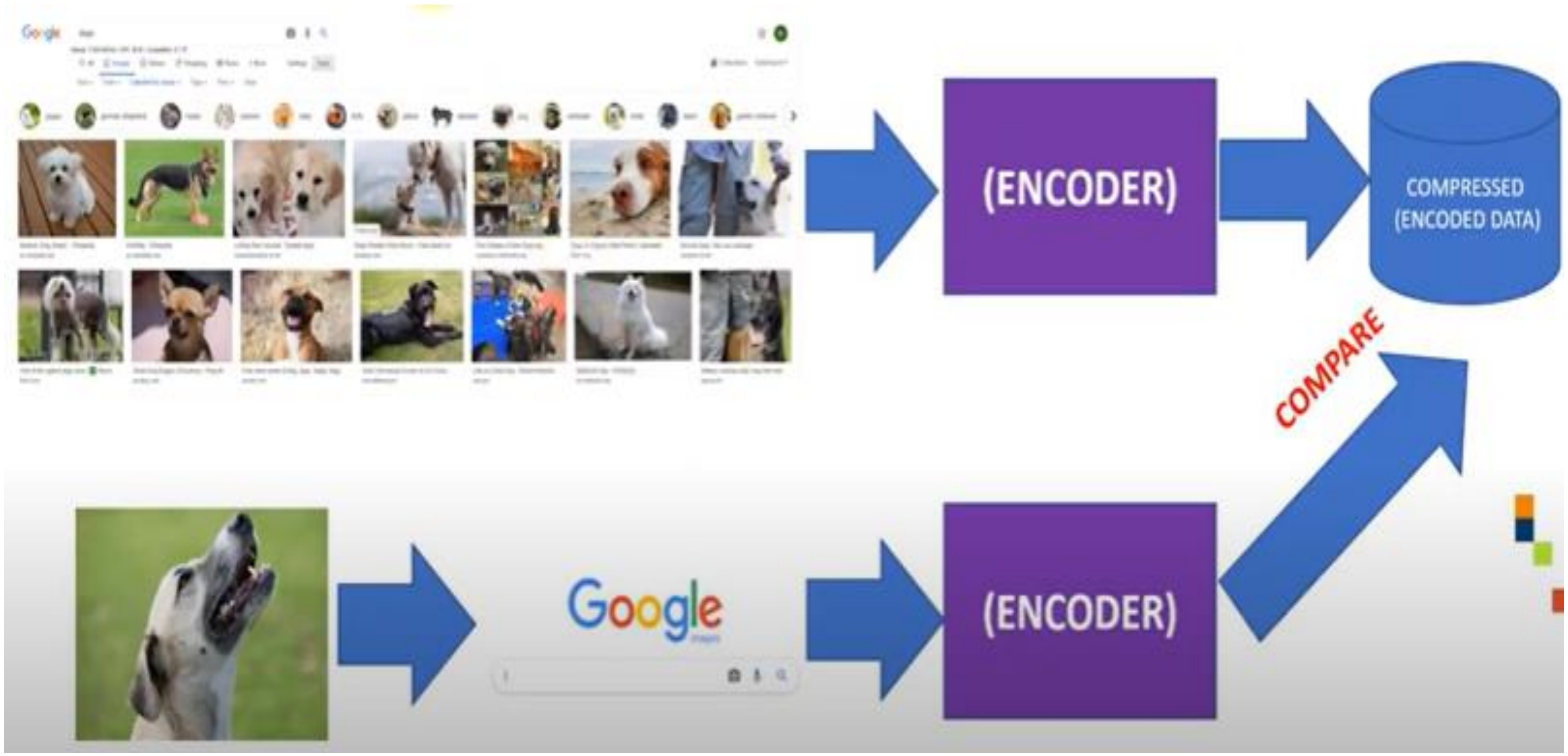


Image Search



Anomaly Detection

- Autoencoders are used anomaly detection such as Credit Card Fraud detection.
- Train Autoencoders on proper (Non-Fradulent) transactions only.
- Network becomes capable of reconstructing the input with good reconstruction loss.
- If you need in a fraudulent transaction (Anomaly), the reconstruction loss will be large.
- We can set threshold to perform anomaly detection.

Autoencoders in a nutshell: Key Takeaways

1. Well, that was a lot to take in. Let's do a quick recap of everything you've learned in this guide:
2. An autoencoder is an unsupervised learning technique for neural networks that learns efficient data representations (encoding) by training the network to ignore signal "noise."
3. Autoencoders can be used for image denoising, image compression, and, in some cases, even generation of image data.
4. While autoencoders might seem easy at the first glance (as they have a very simple theoretical background), making them learn a representation of the input that is meaningful is quite difficult.
5. Autoencoders like the undercomplete autoencoder and the sparse autoencoder do not have large scale applications in computer vision compared to VAEs and DAEs which are still used in works since being proposed in 2013 (by Kingmaet al).


Thank You
Merci
Salamat
Kop
Tack
Hvala
Diolch
Kiitos
Sheun
umesc
Kasih
Todah
Mamnoon
Shnorhakalutiun
Shokriya
Dziękuję
Shokrun
Spaas
Mul
Ači
Xie
Gamsahapnida
Teşekkür
Dekuju/Dekujeme
Ngiyabonga
Cam
Gra
or
al
Dhanyavaad
Dank
Takk
Dakujem
Waad
Daw
Dhanyavadaalu
Dankie
Go
Grazie
Faleminderit
Dhanyavad
Khopjai
Kruthagnathalu
kun
Arigatou
krap
Shukriya
Or
Dhonnobaad
ederim
Asante
Grazzi
raibh
Gracias
Nandree
Blagodariya
Gomapsupnida
Fyrir
Terima
Enkosi
Danke
dank
Euxaristo
Kun
Hain
Dhan
daa