Deep Autoencoders

V. Dimaridou, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 2.5

Artificial Intelligence &
Information Analysis Lab

Content

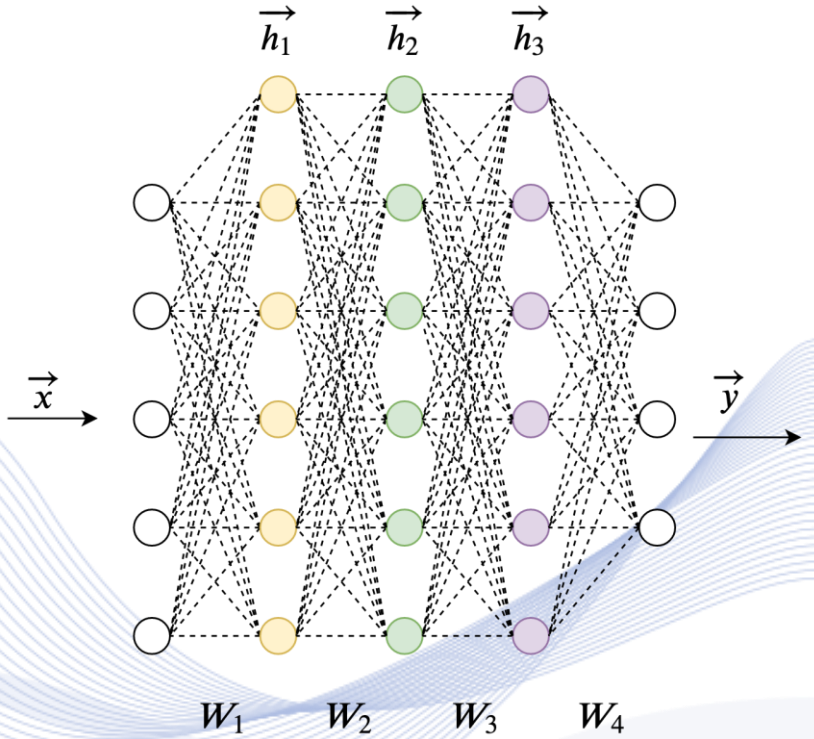
Deep learning *…is a **subset** of machine learning*

Can be split into 3 categories:

- Supervised → *uses labeled data only*
- Semi-supervised → *uses labeled & unlabeled data*
- Unsupervised → *uses only unlabeled data*

Connection to human brain :

- Deep neural networks draw inspiration from the human brain
- Human brain contains millions of neurons with thousands of connections
- Multilayered neural networks progressively extract higher level features from the raw input



Autoencoder Functionality

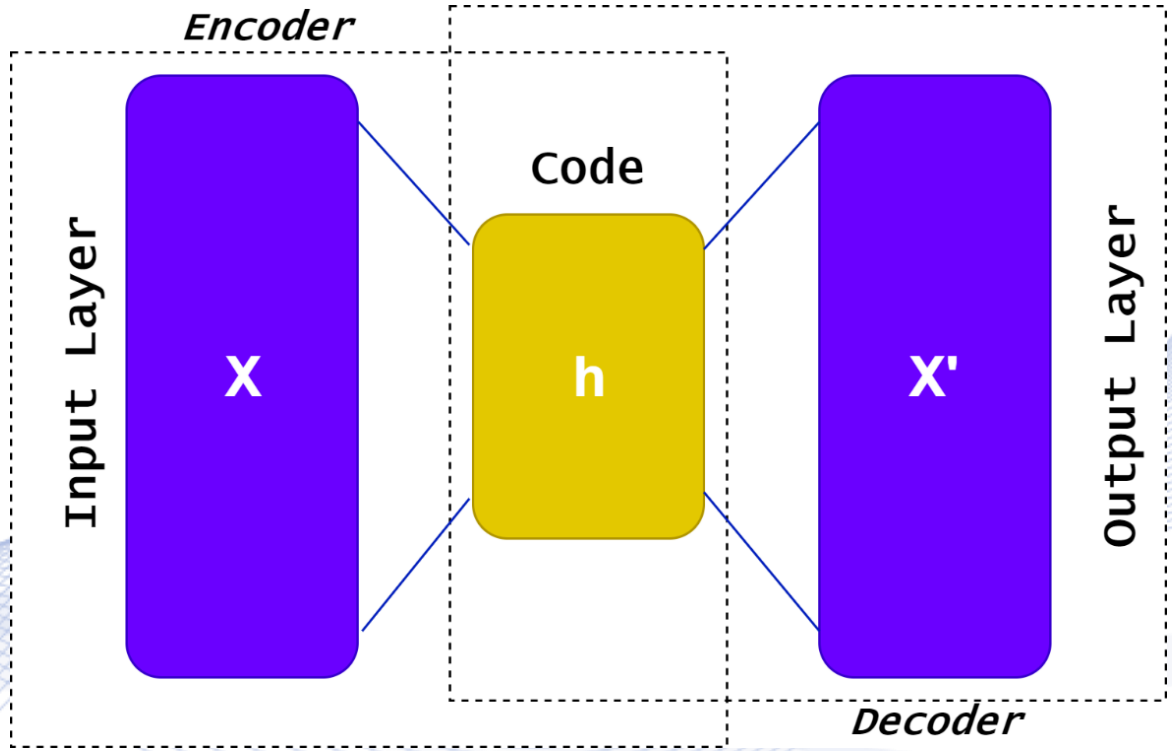
Autoencoders are unsupervised learners

- a family of neural networks for which the input is the same as the output.
- tight bottleneck of a few neurons in the middle, forcing them to effectively create representations that compress the input

Autoencoder Functionality

A classic autoencoder consists of:

- *Encoder layers*
- *Latent View Representation (code)*
- *Decoder layers*



Math time!

1

In general…

$$\boldsymbol{\varphi} : \mathbf{x} \rightarrow \mathbf{y}$$
$$\boldsymbol{\psi} : \mathbf{y} \rightarrow \mathbf{x}$$
$$\boldsymbol{\varphi}, \boldsymbol{\psi} = \ argmin_{\boldsymbol{\varphi},\boldsymbol{\psi}} \|\mathbf{x} - (\boldsymbol{\psi} \circ \boldsymbol{\varphi})\mathbf{x}\|$$

Where:

- $\mathbf{x}$ is the input vector
- $\mathbf{y}$ is the latent vector
- $\boldsymbol{\varphi}$ is the ***encoding*** function
- $\boldsymbol{\psi}$ is the ***decoding*** function
- $\boldsymbol{\psi} \circ \boldsymbol{\varphi}$: function synthesis

2

In the simplest case…

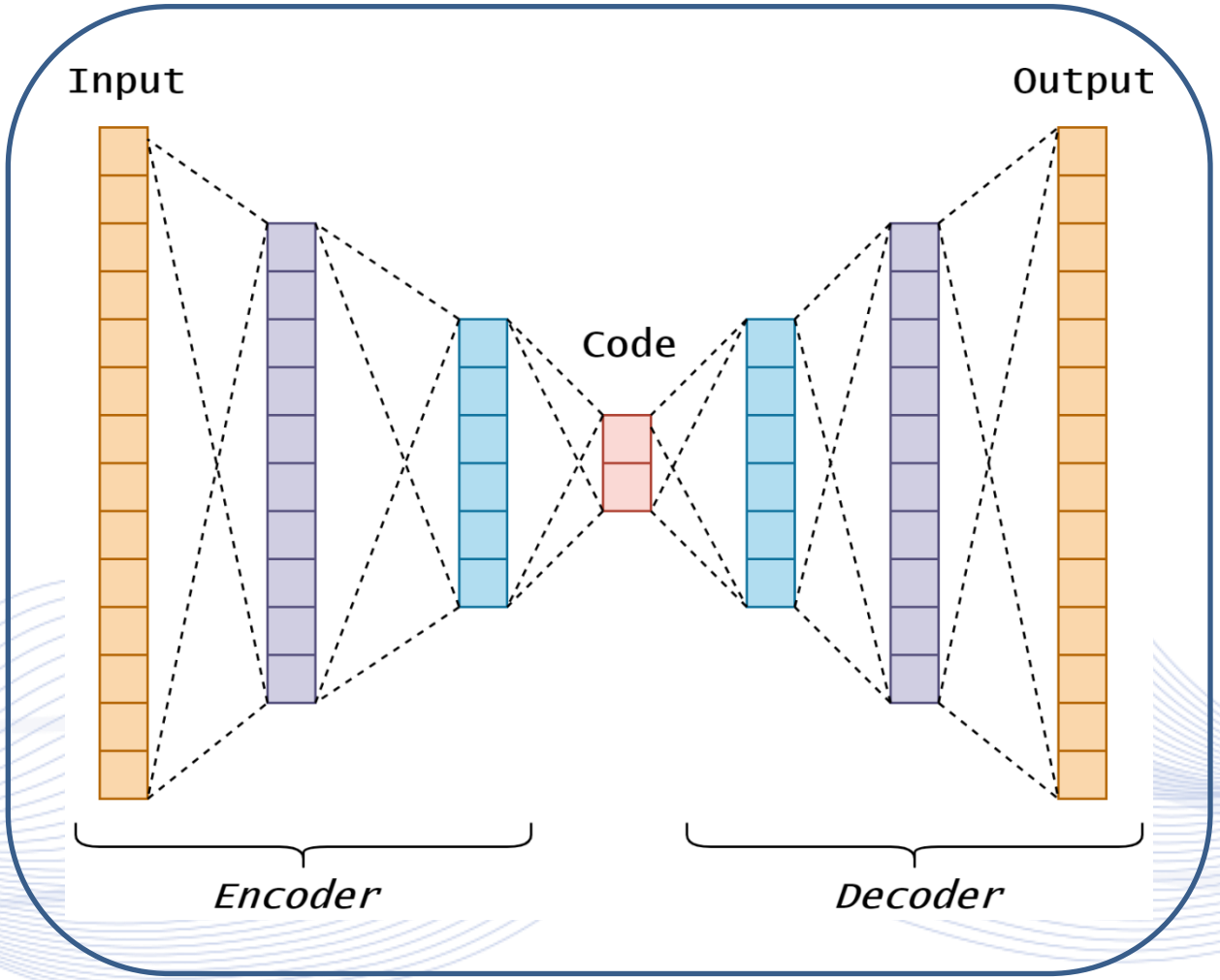
$$\mathbf{y} = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b})$$
$$\mathbf{x}' = \sigma'(\mathbf{W}\mathbf{x} + \mathbf{b})$$
$$L(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\|^2 = \|\mathbf{x} - \sigma'\mathbf{W}'(\sigma(\mathbf{W}\mathbf{x} + \mathbf{b}) + \mathbf{b}')\|^2$$

Where:

- $\mathbf{x}'$ is the reconstructed input
- $\mathbf{W}$ & $\mathbf{W}'$ are the weight matrixes
- σ & σ' are the activation functions
- $\mathbf{b}$ & $\mathbf{b}'$ are bias factors

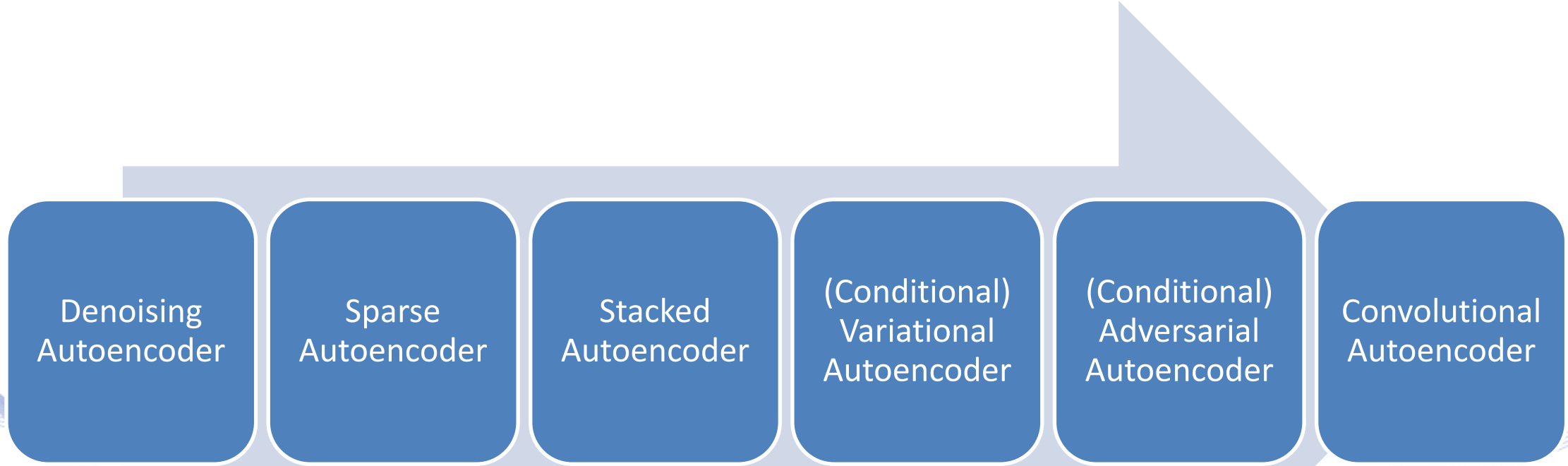
Deep Autoencoders

More complex datasets require more complex architectures



A deep autoencoder consists of two, symmetrical deep-belief networks

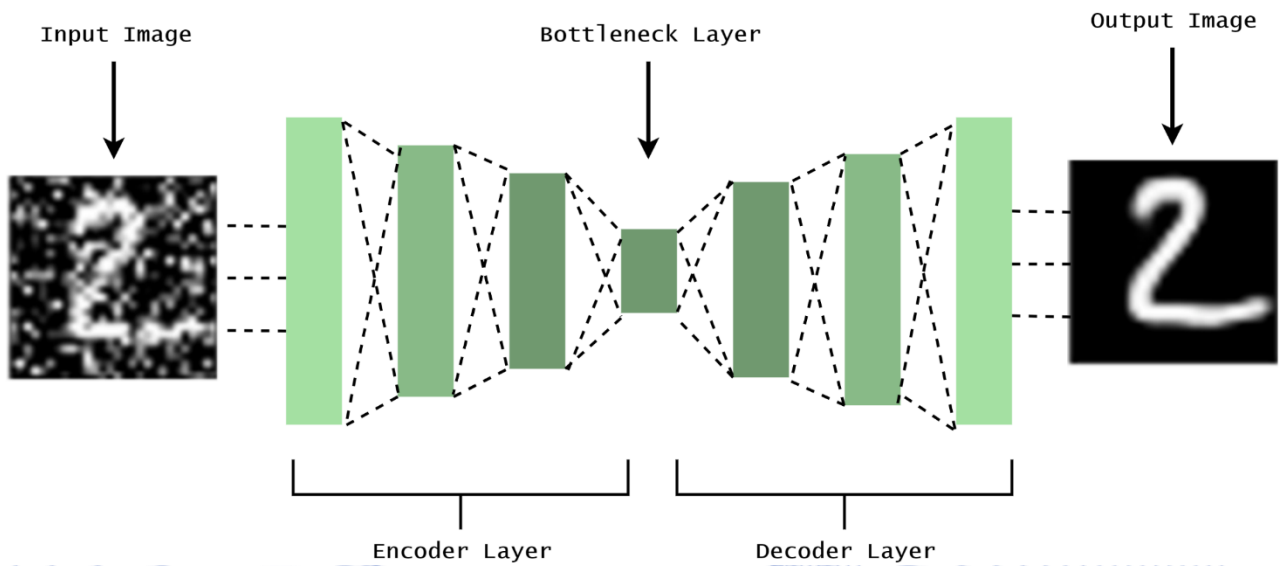
Autoencoder Types



1. Denoising Autoencoders

Tries to:

1. Encode the input from a corrupted version of it
2. Undo the effect of the corruption process

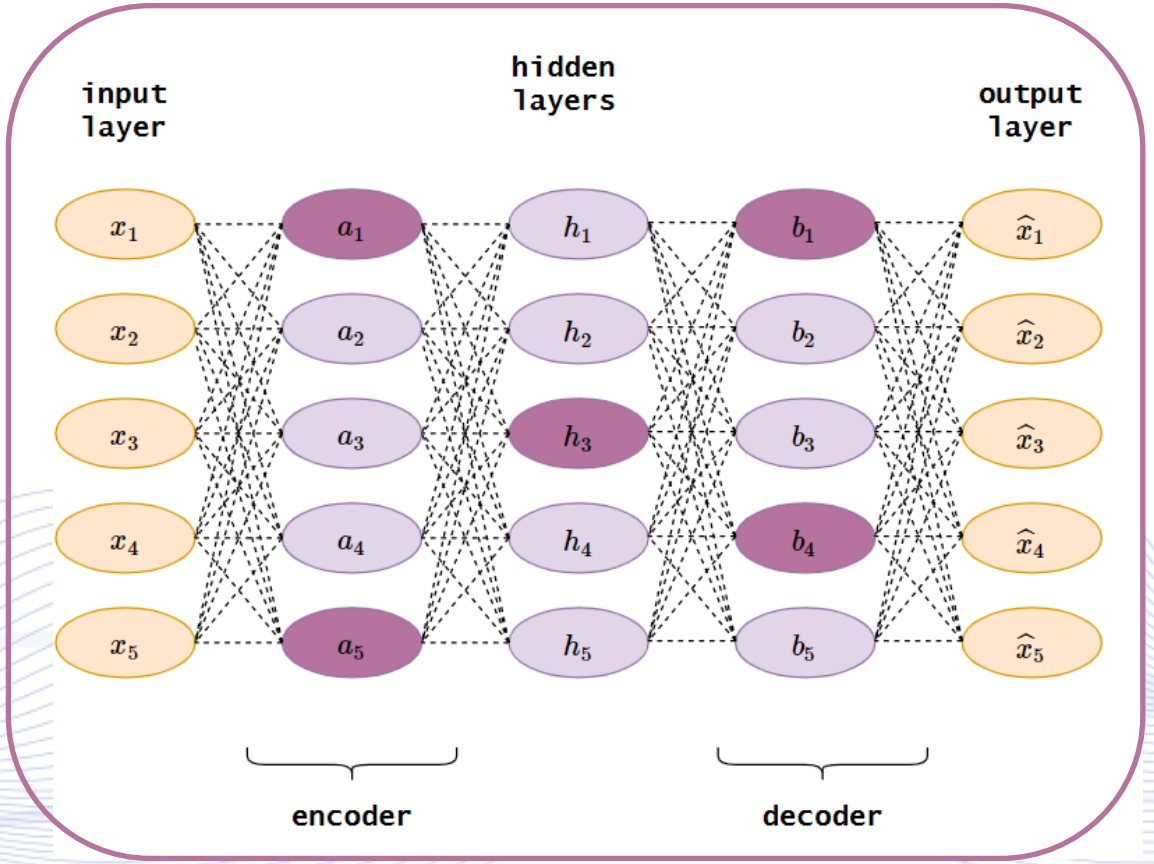


Data corruption typically in 30-50% of the pixels

In the loss function the output values are compared with the original input & not the corrupted output!

2. Sparse Autoencoders

- only a small number of the hidden units **can be active at once**.
- this sparsity constraint forces the model to respond to the **unique** statistical features of the input data used for training.



loss function penalizes activations (output value 0):

$$Loss = L(x, x') + \Omega(\mathbf{h})$$

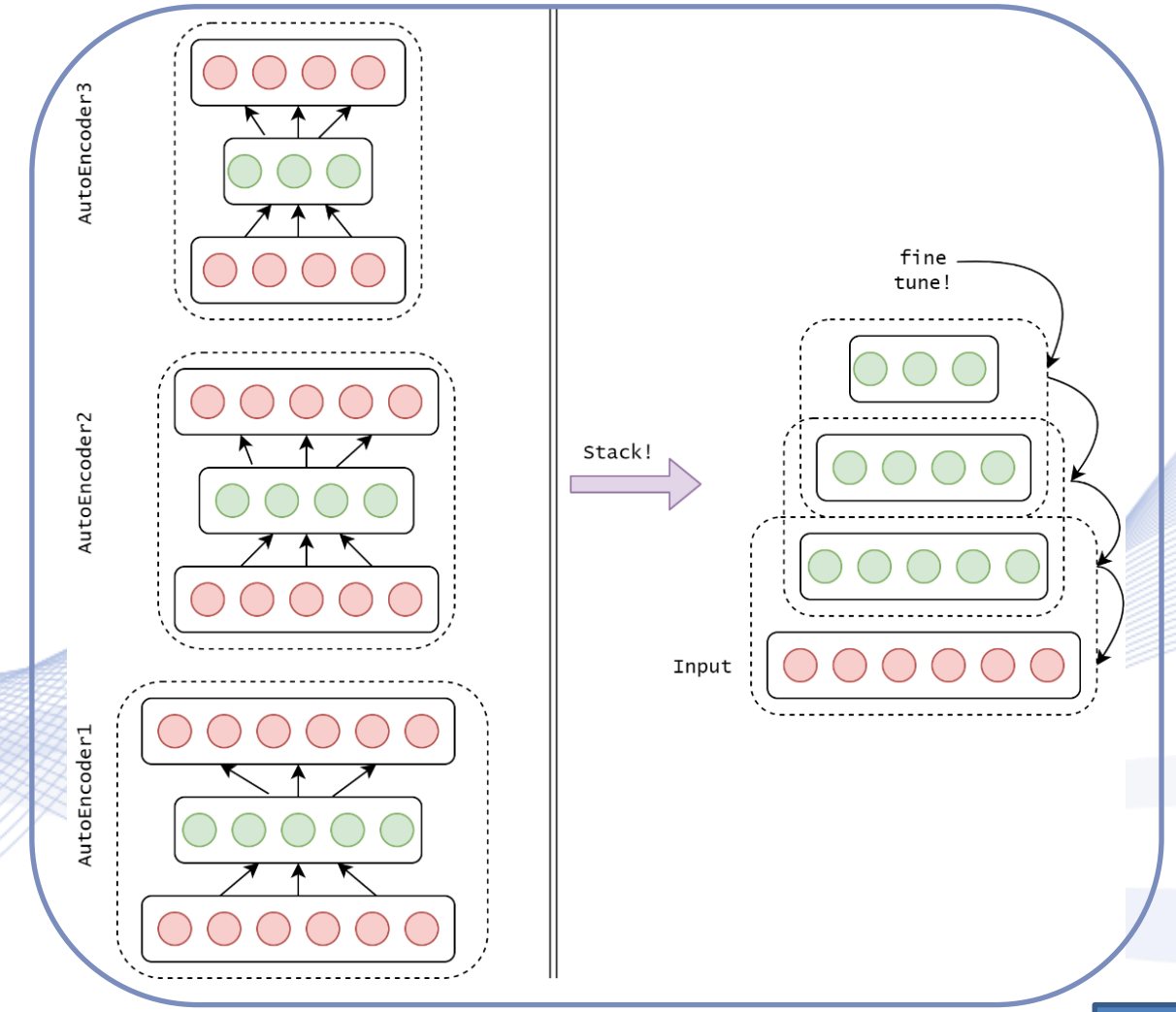
where Ω(h) is the sparsity penalty

3. Stacked Autoencoders

Consist of several layers of Autoencoders

Training procedure:

1. Train the 1st autoencoder by input data & obtain the learned feature vector.
2. The feature vector of the former layer is used as the input for the next layer, and this procedure is repeated until the training completes.
3. After all the hidden layers are trained, backpropagation algorithm (BP) is used to minimize the cost function and update the weights using the training set to achieve fine-tuning.



5. (Conditional) Variational Autoencoder

Classic Autoencoders learn to replicate input data. Latent space is discontinuous.

In data generation process the goal is to randomly sample from the latent space, resulting to an unseen image.

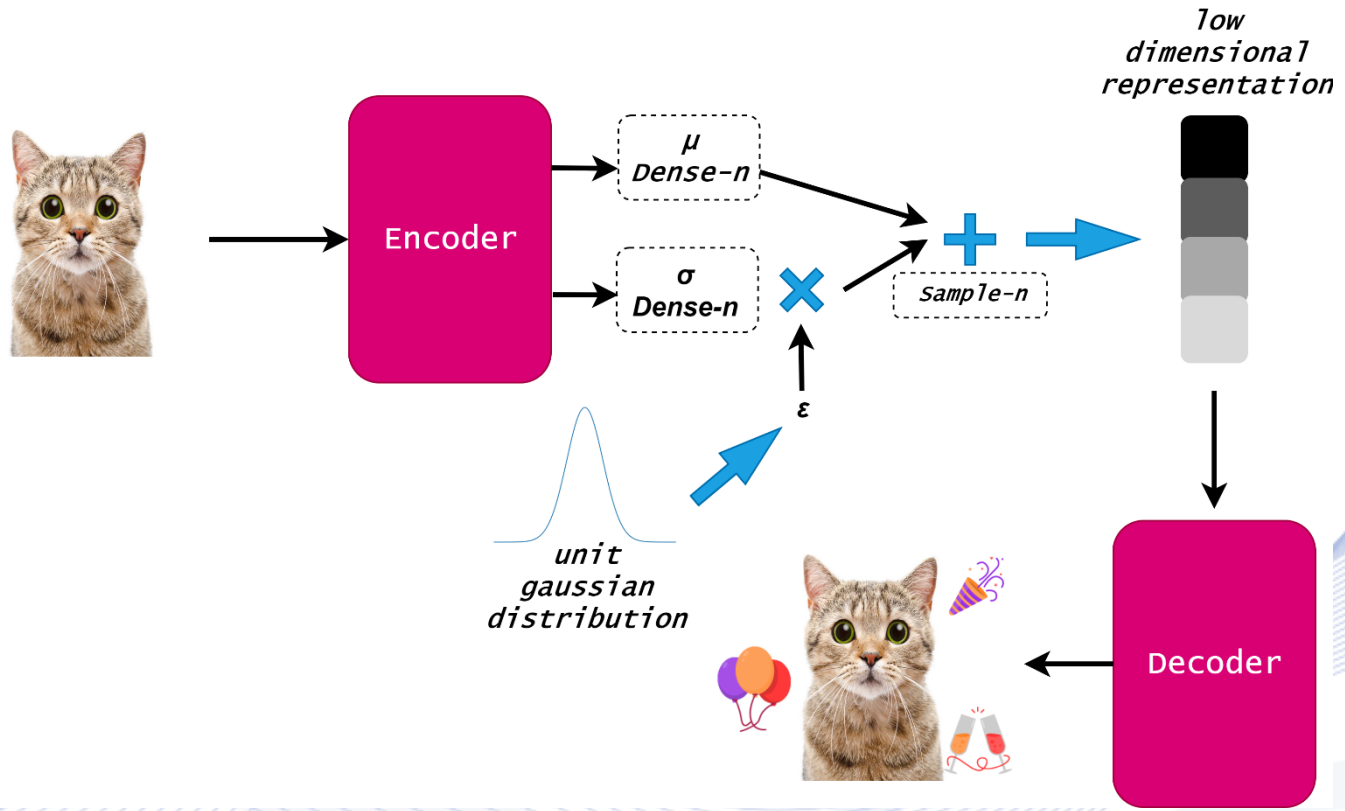
Generating from discontinuous space results to unrealistic output, because the decoder has *no idea* how to deal with that region of the latent space. During training, it *never saw* encoded vectors coming from that region of latent space.

5. (Conditional) Variational Autoencoder

- Variational Autoencoders latent spaces are, by design, continuous, allowing easy random sampling and interpolation.
- encoder outputs:

1. a vector of means, μ,
2. a vector of standard deviations, σ

- Formulation of parameters of a vector of n random variables.
- Sampling from n variables.



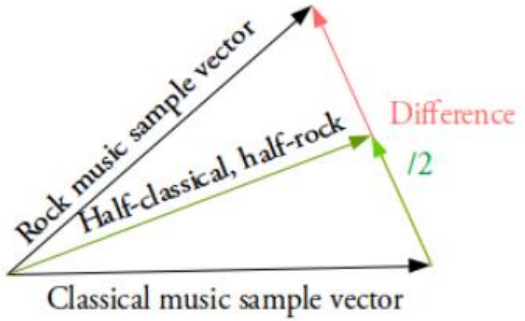
Due to stochastic generation the encoding will vary even for the same input

5. (Conditional) Variational Autoencoder

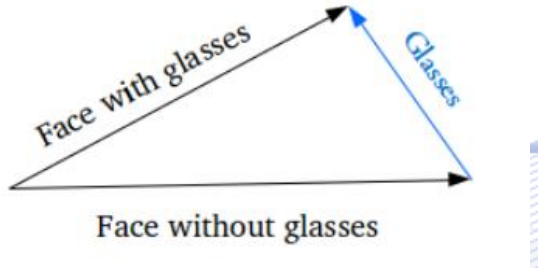
- Variational Autoencoders latent spaces are, by design, continuous, allowing easy random sampling and interpolation.
- encoder outputs:

1. a vector of means, μ,
2. a vector of standard deviations, σ

- Formulation of parameters of a vector of n random variables.
- Sampling from n variables.



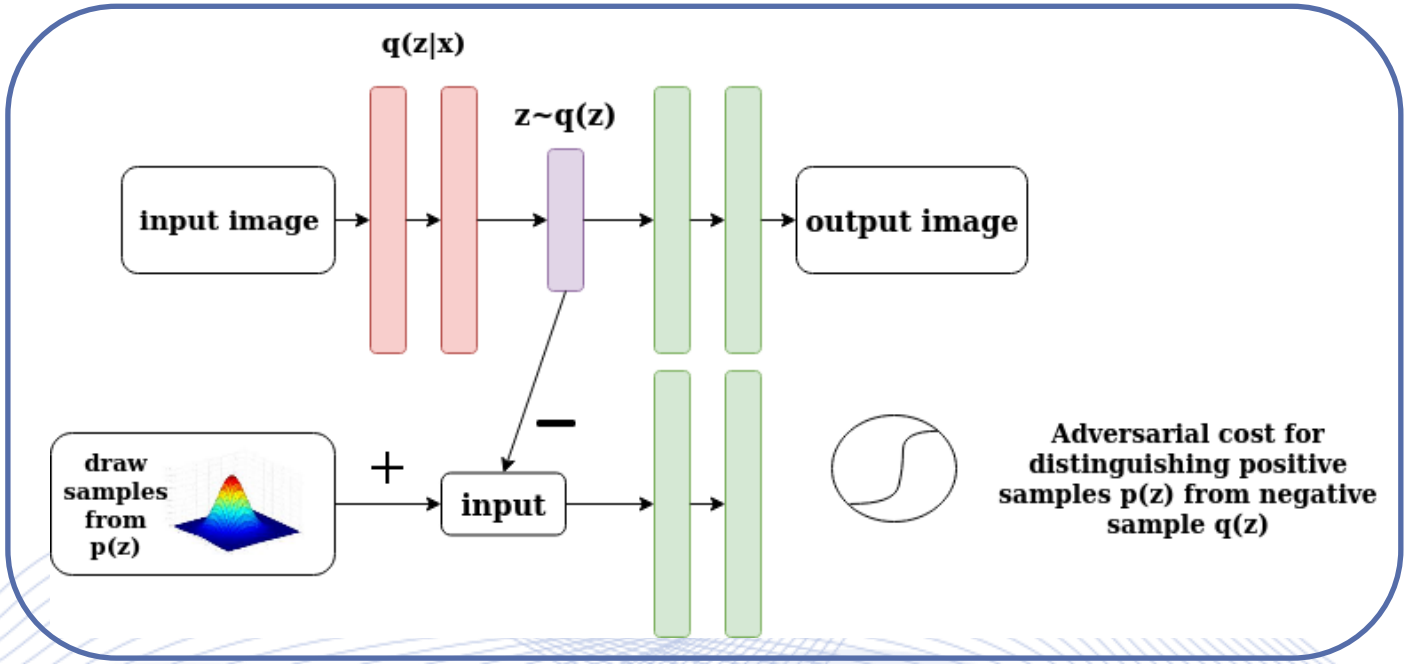
Interpolating between samples



Due to stochastic generation the encoding will vary even for the same input

6. (Conditional) Adversarial Autoencoder

- *AAEs aim to continuous latent space like VAEs.*
- Encoded vector is still composed of the mean value and standard deviation, but prior distribution $p(\mathbf{z})$ is used to model it.



2-phase training:

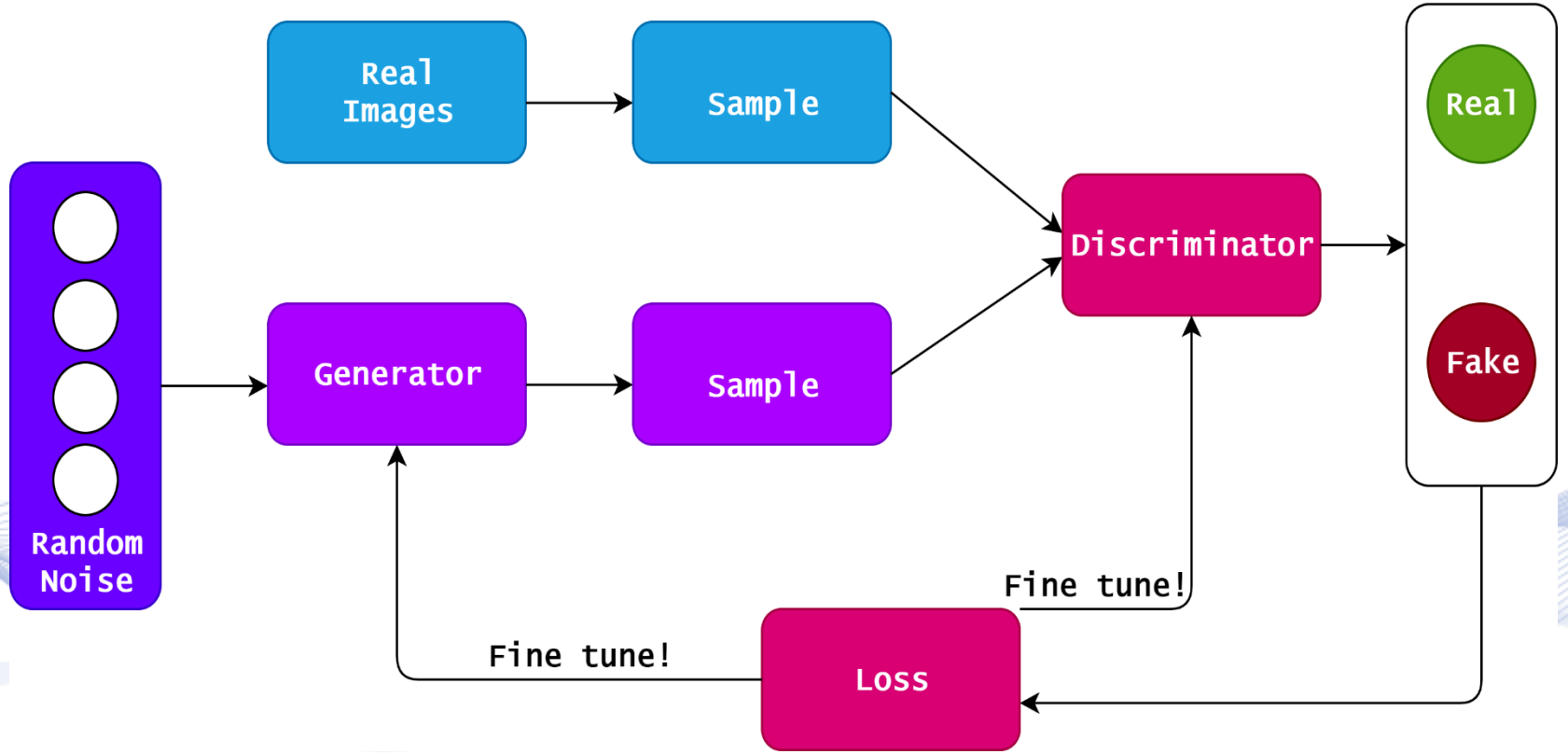
1. Reconstruction phase (encoder & decoder min reconstruction error)
2. Regularization phase (using adversarial loss)

The encoder ensures the autoencoder (generator) can fool the discriminative adversarial network into thinking that the hidden code $q(z)$ comes from the true prior distribution $p(z)$

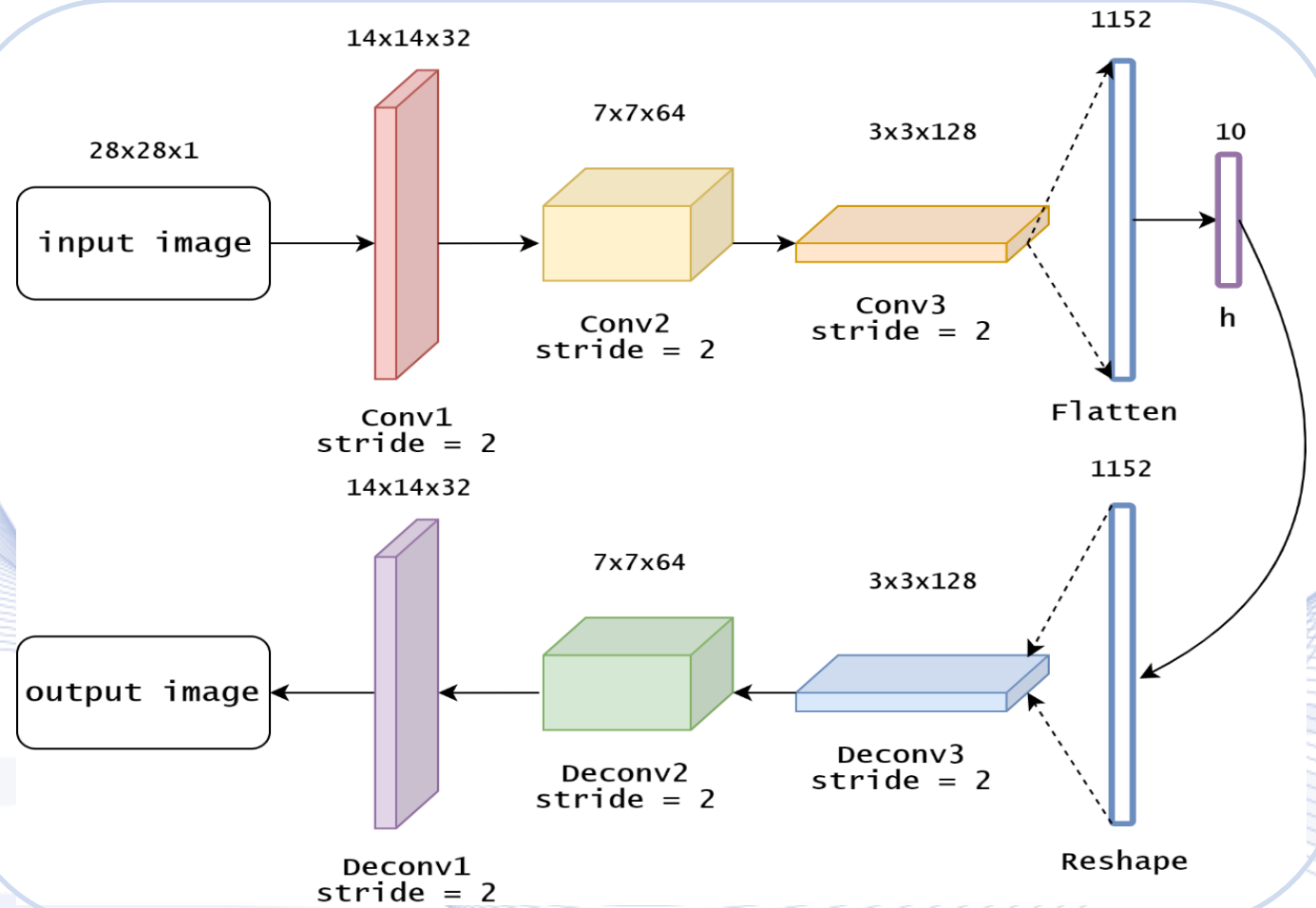
What is a GAN?

Generative **A**dversarial **N**etwork

- Generator & Discriminator *play a min-max game*!



6. Convolutional Autoencoder



- They encode the input in a set of simple signals and then reconstruct the input from them, modify the geometry or the reflectance of the image.
- They are the state-of-art tools for unsupervised learning of convolutional filters.
- There filters can be applied to any input in order to extract features (e.g. for classification).

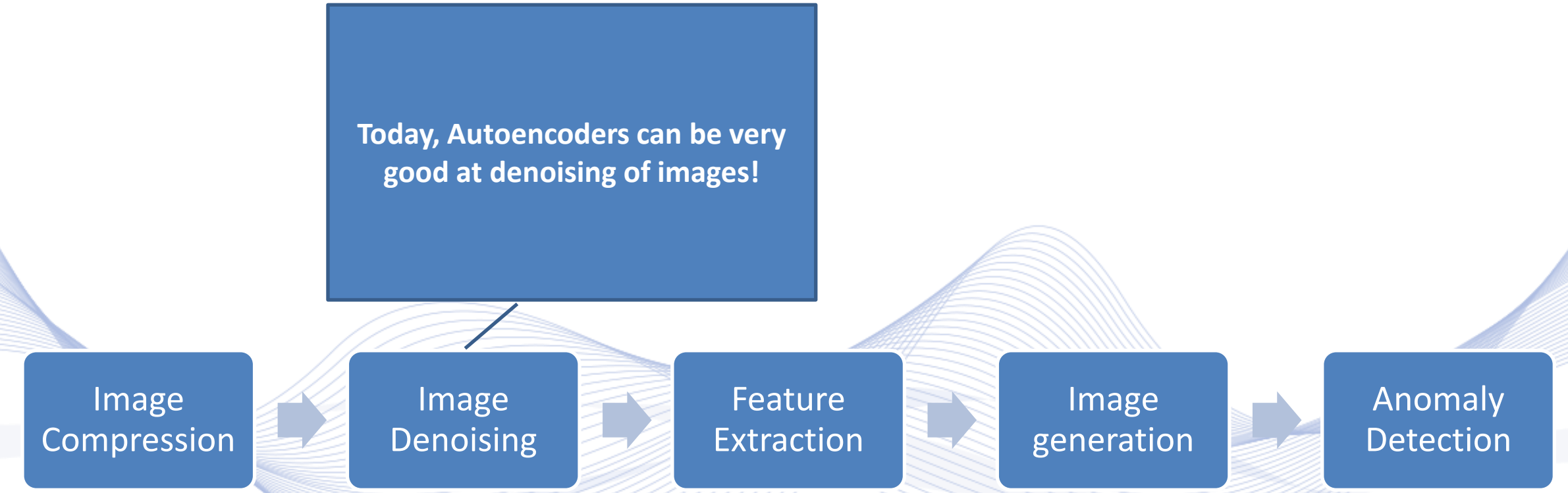
Applications in Computer Vision

- General autoencoder applicability
- Pose estimation autoencoder example
- Image denoising autoencoder example
- Image classification using autoencoder example
- Image generation example

General autoencoder applicability

However, **for image compression, it is pretty difficult for an autoencoder to do better than basic algorithms, like JPEG**

General autoencoder applicability



General autoencoder applicability

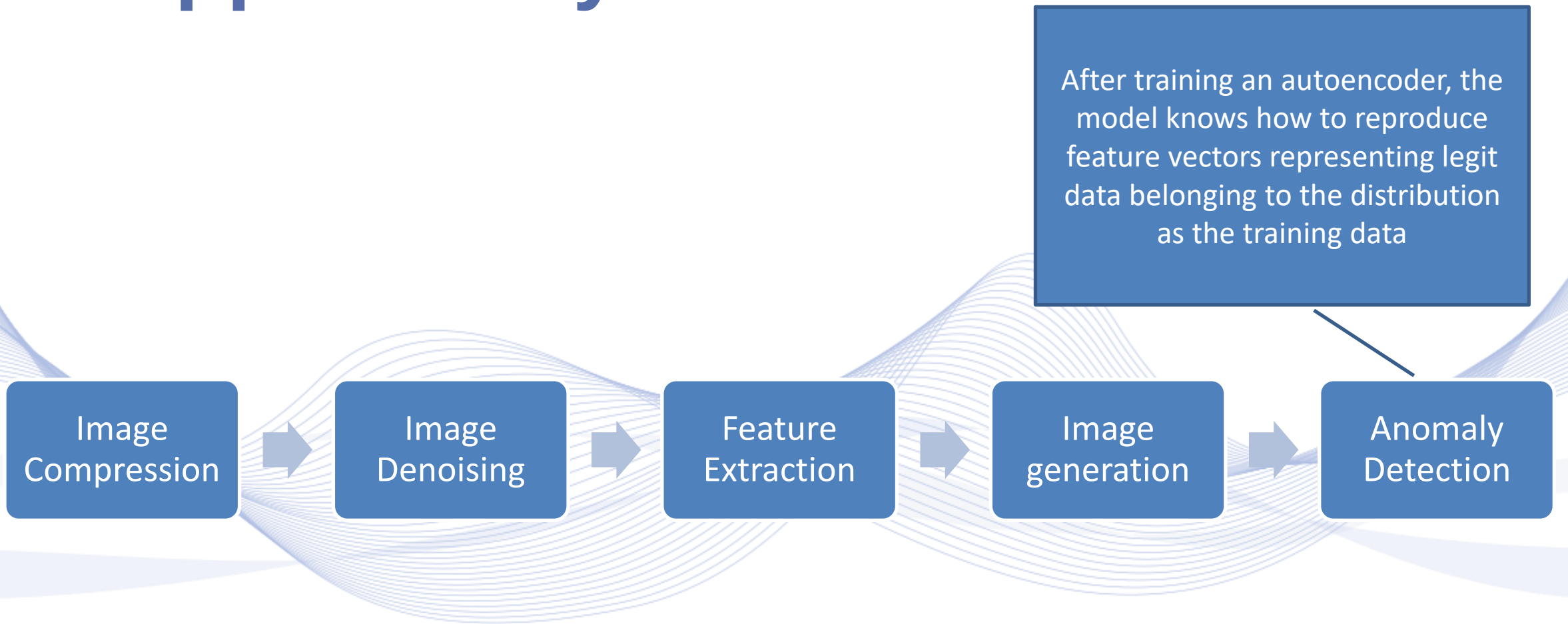
Encoding part of Autoencoders helps to learn important hidden features present in the input data, in the process to reduce the reconstruction error. During encoding, a new set of combination of original features is generated.

General autoencoder applicability

Using adversarial or variational (conditional or not) autoencoders

Image Compression → Image Denoising → Feature Extraction → Image generation → Anomaly Detection

General autoencoder applicability



Pose estimation autoencoder

"Multimodal deep autoencoder for human pose recovery"

Objective:

- Retrieve 3D pose representations from 2D images.

Methodology:

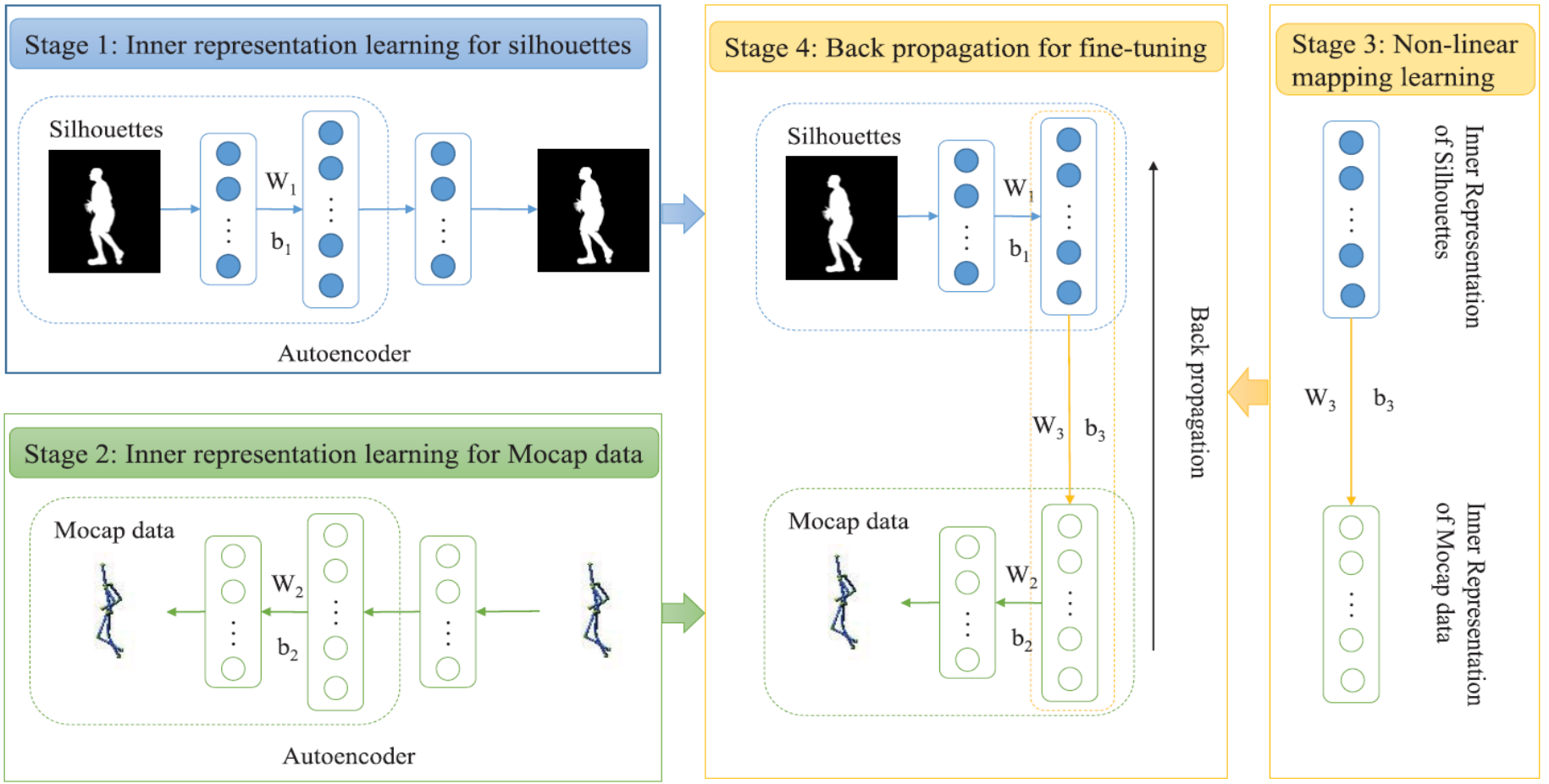
- Feature extraction of 2D & 3D representations using deep autoencoders
- Back-propagation deep learning for 2D – 3D mapping

Source [1]

Pose estimation autoencoder

"Multimodal deep autoencoder for human pose recovery"

Model Architecture:



- ***Step 1*** obtains the inner representations of the 2D images using features learned from AA.
- ***Step 2*** maps the 2D inner representations to the corresponding 3D inner representations.
- ***Step 3*** reconstructs the 3D poses based on the corresponding inner representations using BP NN.
- ***Step 4*** mocap data can be recovered with this mapping and back propagation is utilized to refine the mapping.

Source [1]

Image denoising autoencoder

"Medical image denoising using convolutional denoising autoencoders"

Objective:

- Denoise medical images as a preprocessing step in medical image analysis

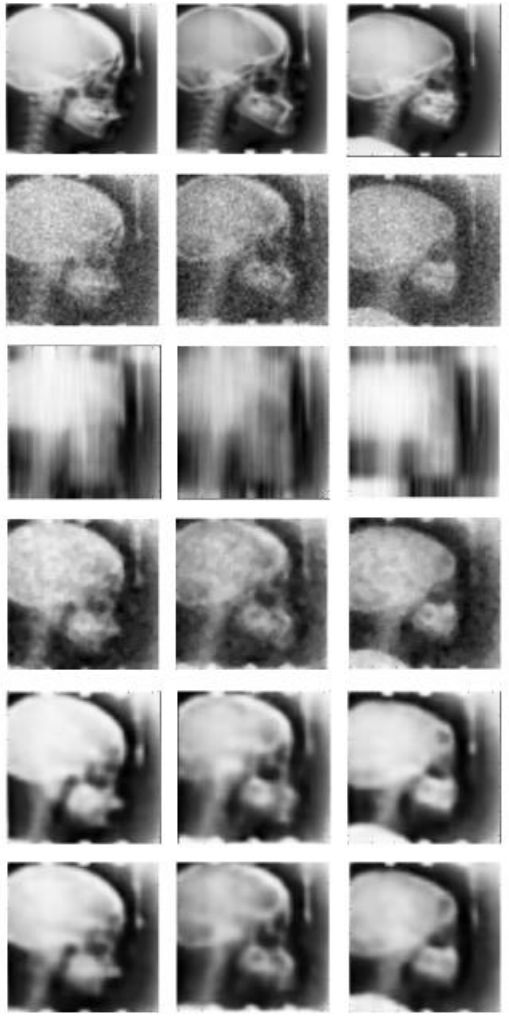
Methodology:

- Combination of convolutional, denoising & stacked autoencoder
- 2 datasets used, consisting of 722 high resolution images
- Gaussian & Poisson distribution introduced, with various noise proportion

Image denoising autoencoder

"Medical image denoising using convolutional denoising autoencoders"

Results:



Real Images

Noiser version with minimal noise

Denoising result of NL (Non-local mean filtering) means

Results of median filter

CNN DAE using smaller dataset (300 training samples)

CNN DAE using larger combined dataset

Source [2]

Image classification using autoencoder

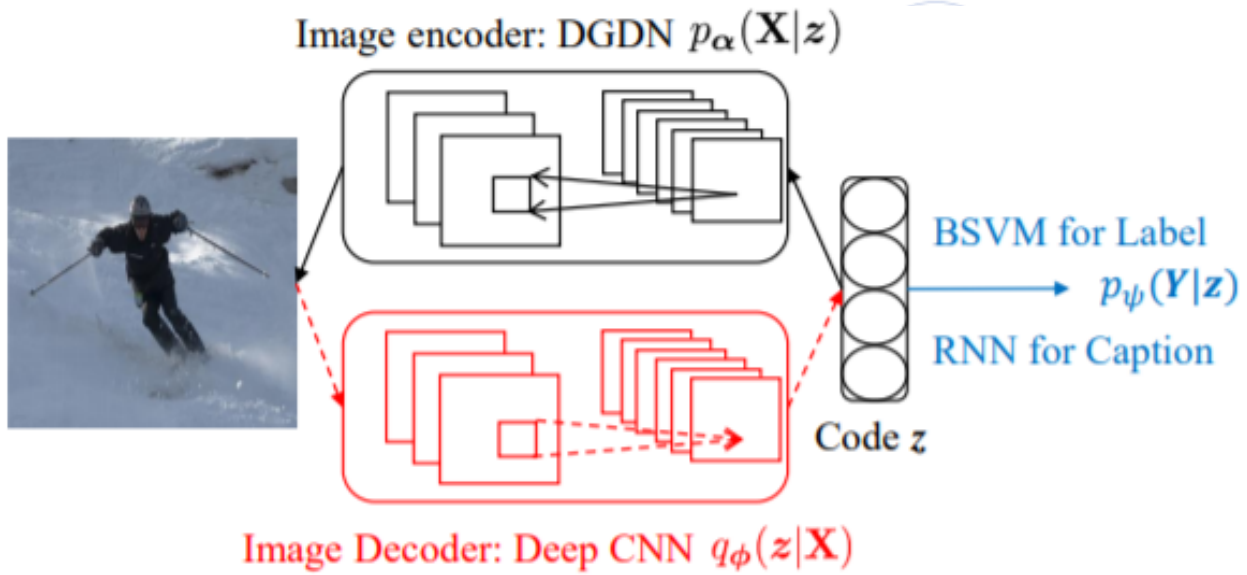
"Variational Autoencoder for Deep Learning of Images, Labels and Captions"

Objective:

- Generate image captions in unseen images

Methodology:

- Deep Generative Deconvolutional Network (DGDN) is used as a decoder of the latent image features
- Deep Convolutional Neural Network (CNN) is used as an image encoder
- Latent code is also linked to generative models for labels (Bayesian support vector machine) or captions (recurrent neural network)



Source [3]

Image classification using autoencoder

"Variational Autoencoder for Deep Learning of Images, Labels and Captions"

Objective:

- Generate image captions in unseen images

Methodology:

- Deep Generative Deconvolutional Network (DGDN) is used as a decoder of the latent image features
- Deep Convolutional Neural Network (CNN) is used as an image encoder
- Latent code is also linked to generative models for labels (Bayesian support vector machine) or captions (recurrent neural network)

Results:


a man with a snowboard next to a man with glasses


a big black dog standing on the grass


a player is holding a hockey stick


a desk with a keyboard


a man is standing next to a brown horse


a box full of apples and oranges

Source [3]

Image Generation

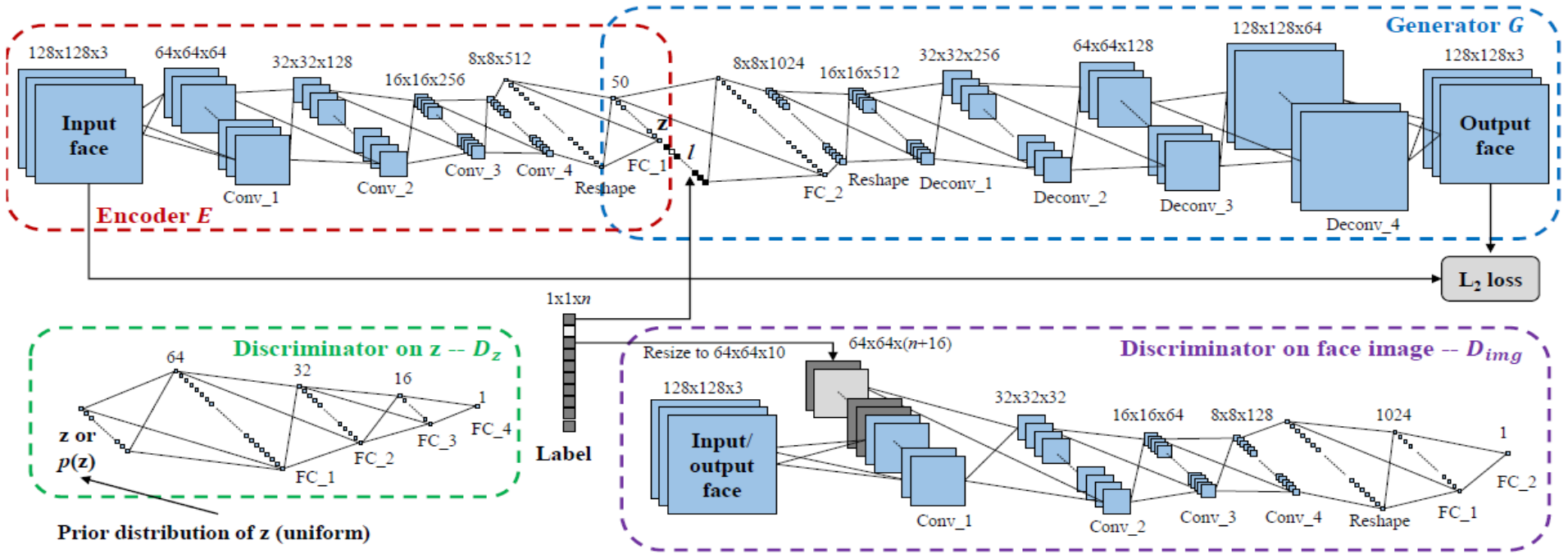
"Age progression/regression by conditional adversarial autoencoder."

Objective:

- Given an picture of a person, output an image of the same person in different ages.

Methodology:

- Conditional Adversarial Autoencoder proposed
- Encoder E maps the input face to a vector z (personality)
- Concatenate label l (age) to z, the new latent vector $[z; l]$ that is fed to the generator G
- The discriminator D_z imposes the uniform distribution on z
- The discriminator D_{img} forces the output face to be photo-realistic and plausible for a given age label l

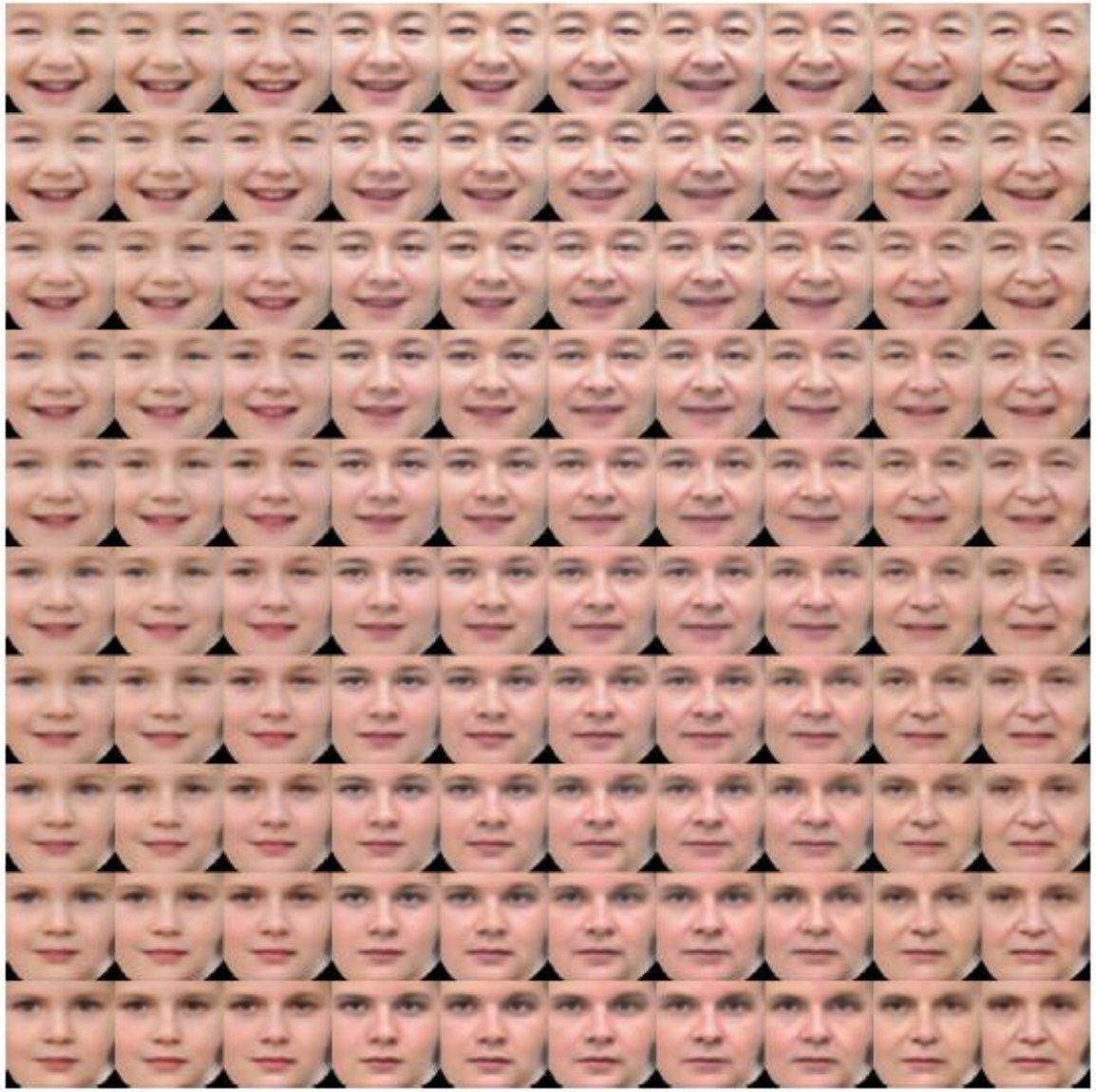


Source [4]

Image Generation

"Age progression/regression by conditional adversarial autoencoder."

Results:



Source [4]

References

[CHA2015] Chaoqun Hong et al. “Multimodal deep autoencoder for human pose recovery". In: IEEE Transactions on Image Processing 24.12 (2015), pp. 5659-5670.

[MIN2017] Min Chen et al. “Deep features learning for medical image analysis with convolutional autoencoder neural network". In: IEEE Transactions on Big Data (2017).

[YUN2016] Yunchen Pu et al. “Variational autoencoder for deep learning of images, labels and captions". In: Advances in neural information processing systems. 2016, pp. 352-2360.

[ZI2017] Zhifei Zhang, Yang Song, and Hairong Qi. “Age progression/regression by conditional adversarial autoencoder". In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017, pp. 5810-5818.

Q & A

Thank you very much for your attention!

More material in
http://icarus.csd.auth.gr/cvml-web-lecture-series/

Contact: Prof. I. Pitas
pitas@csd.auth.gr