Ketil Oppedal, Adjunct associate professor, Dept. electrical engineering and computer science

University of Stavanger

ELE680 - Deep Neural Networks

Introduction to autoencoders and GANs

Learning outcome

At the end of this lecture the student should have a basic understanding of:

- What an autoencoder is and how it works.
- Understand some example applications.
- What a variational autoencoder (VAE) is.
- What a general adversarial network (GAN) is and how it works.
- An example of how VAE and GAN together can be used in medical imaging applications to detect disease.

University of Stavanger

Lecture layout

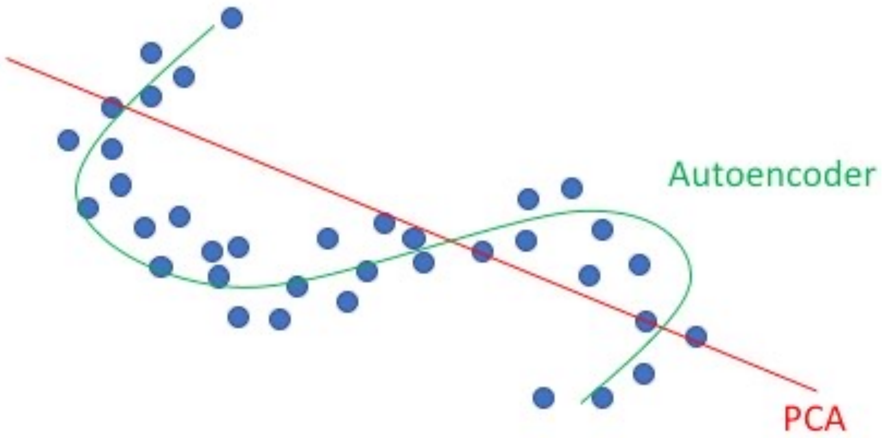
- Basic structure and building blocks of an autoencoder.
- Example applications.
- Variational autoencoders.
- General adversarial networks (GANs).
- Example from a medical imaging application.

Autoencoder

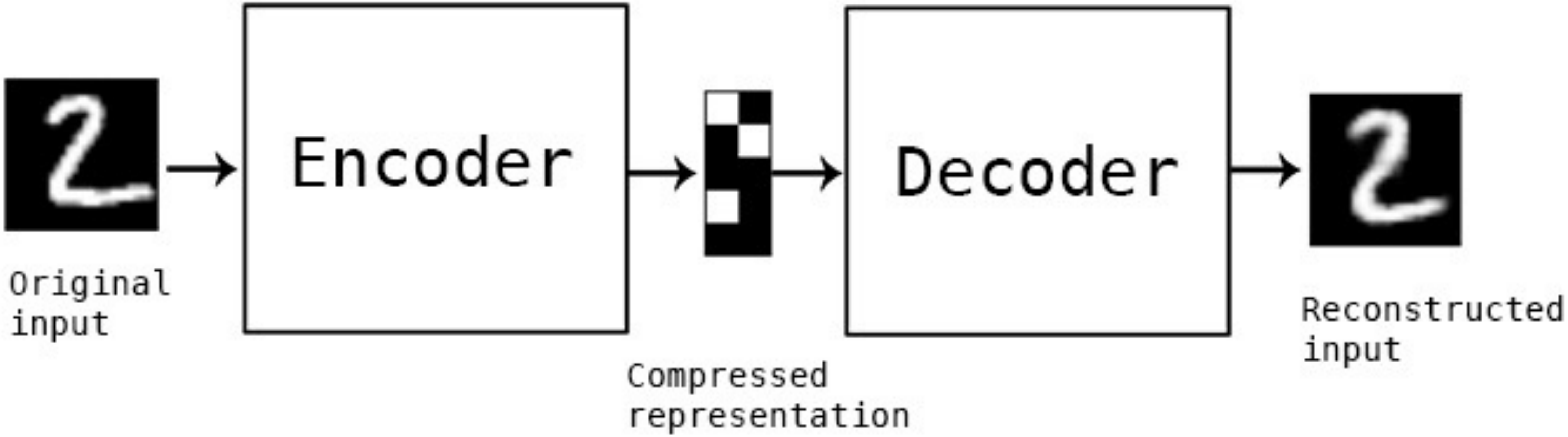
- Unsupervised learning.
- Representation learning.
- Bottleneck with lower-dimensional representation.
- Data should have dependencies across dimensions.
  - If data is independent, it is impossible to find a compressed representation.

University of Stavanger

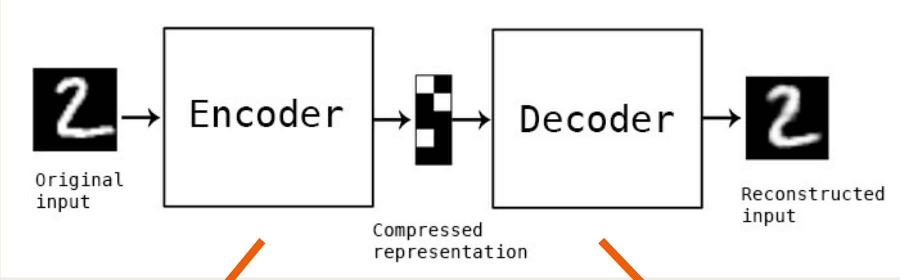
Autoencoder vs. PCA

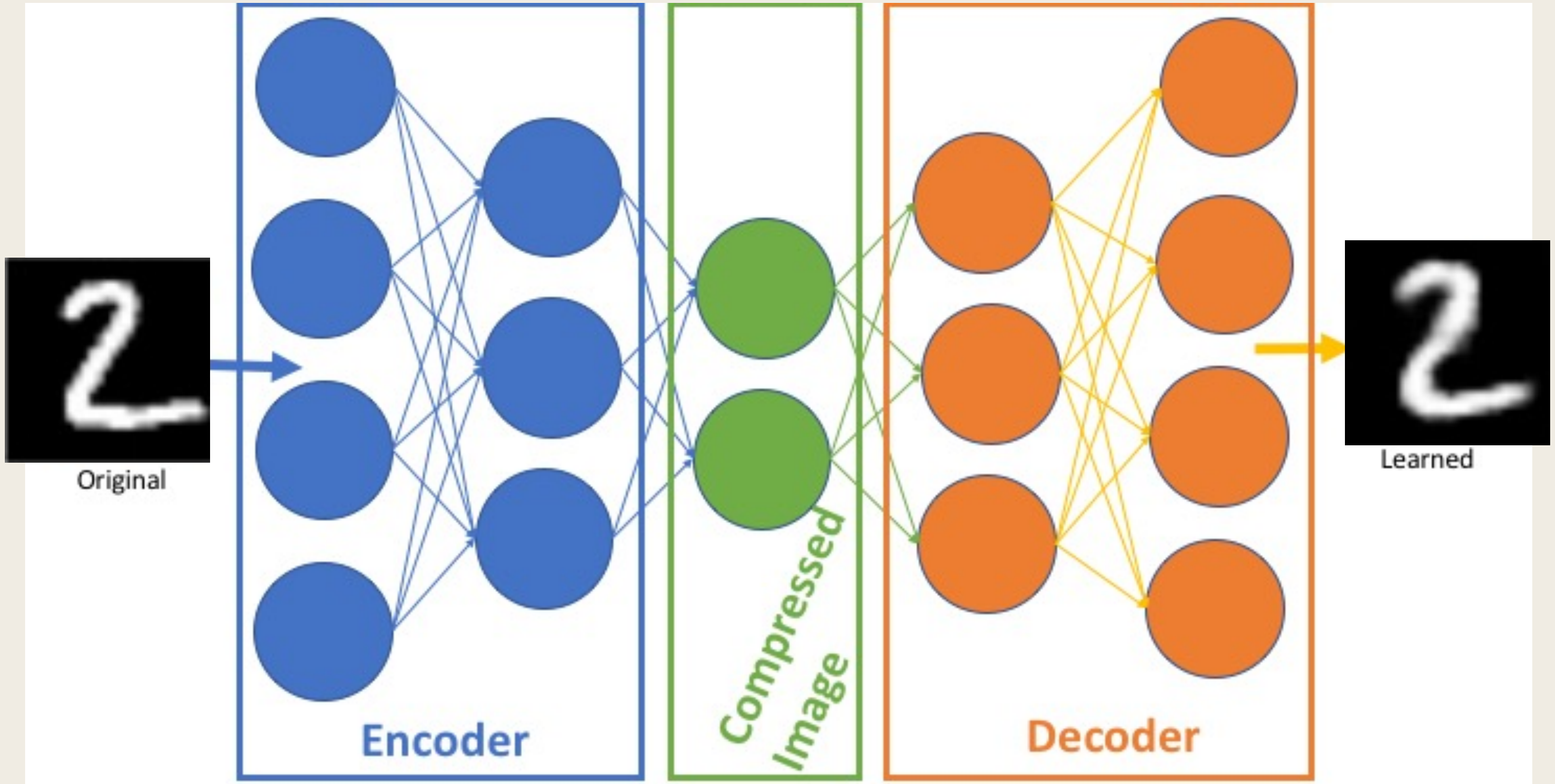


University of Stavanger

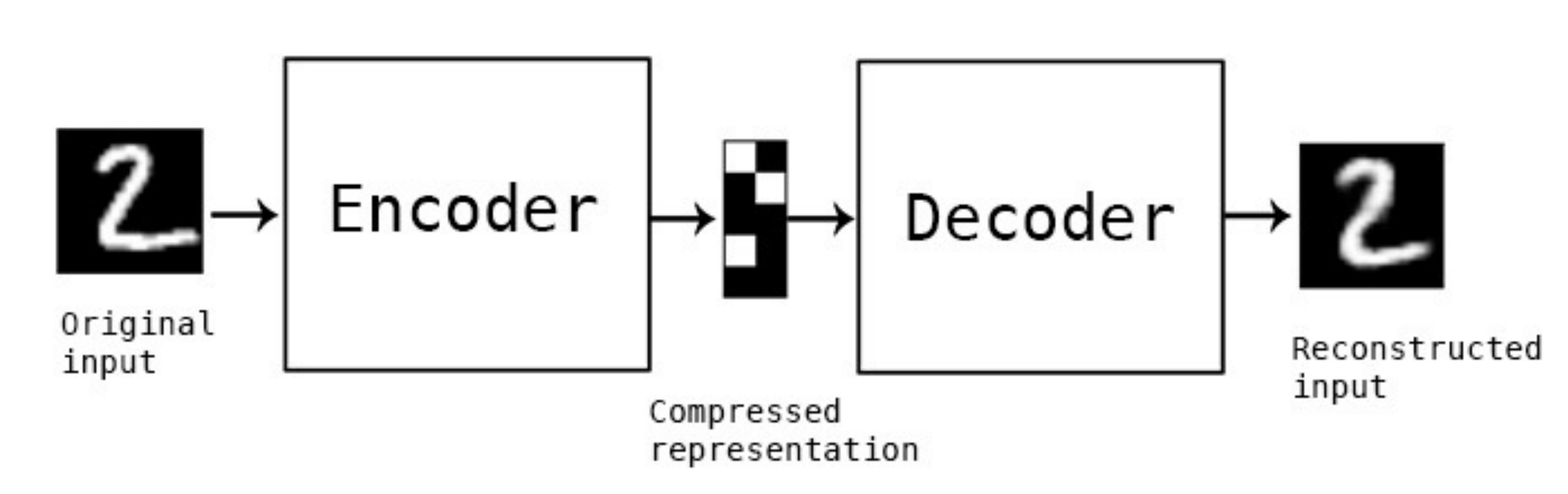


University of Stavanger





University of Stavanger



$x \quad f \quad f(x) \quad g \quad g(f(x)) = x'$

we want $(x' - x)$ as small as possible

University of Stavanger

The latent space keeps the most important attributes of the input data.

We can leverage the latent space to perform several intersting tasks.

Autoencoder applications

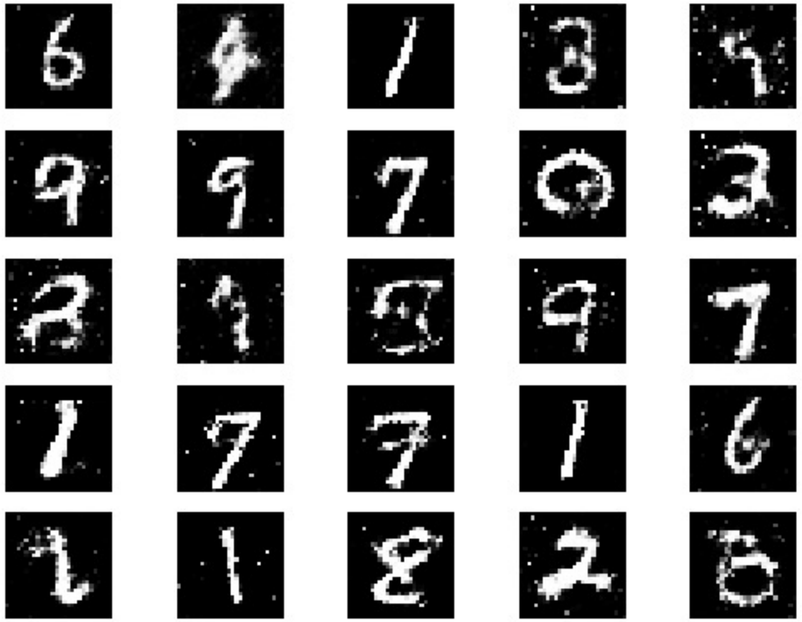
1. Use the compressed representation e.g. in data transfer.
2. Anomaly detection.
3. Denoising and watermark removal.
4. Pretraining for semi-supervised classification
5. Data generation.
6. other…

University of Stavanger

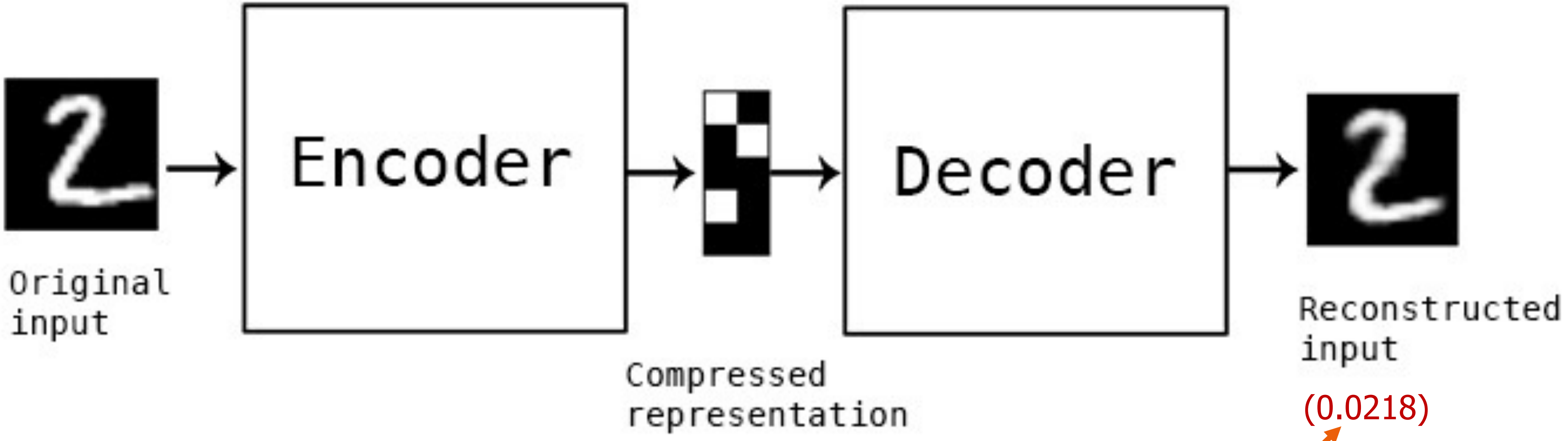
Input

Output

Poor reconstructions



Anomaly detection

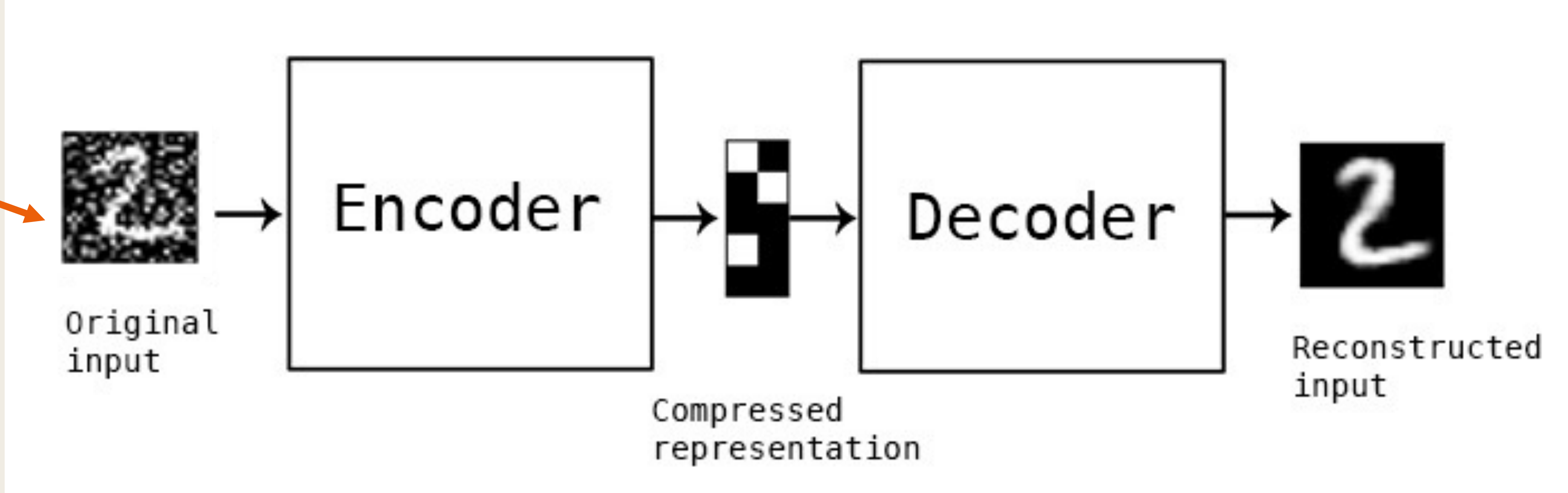


1. Max reconstruction loss during training is e.g. 0.04.
2. Anomaly if reconstructed image has loss > 0.04.

Reconstruction loss

Denoising

With noise

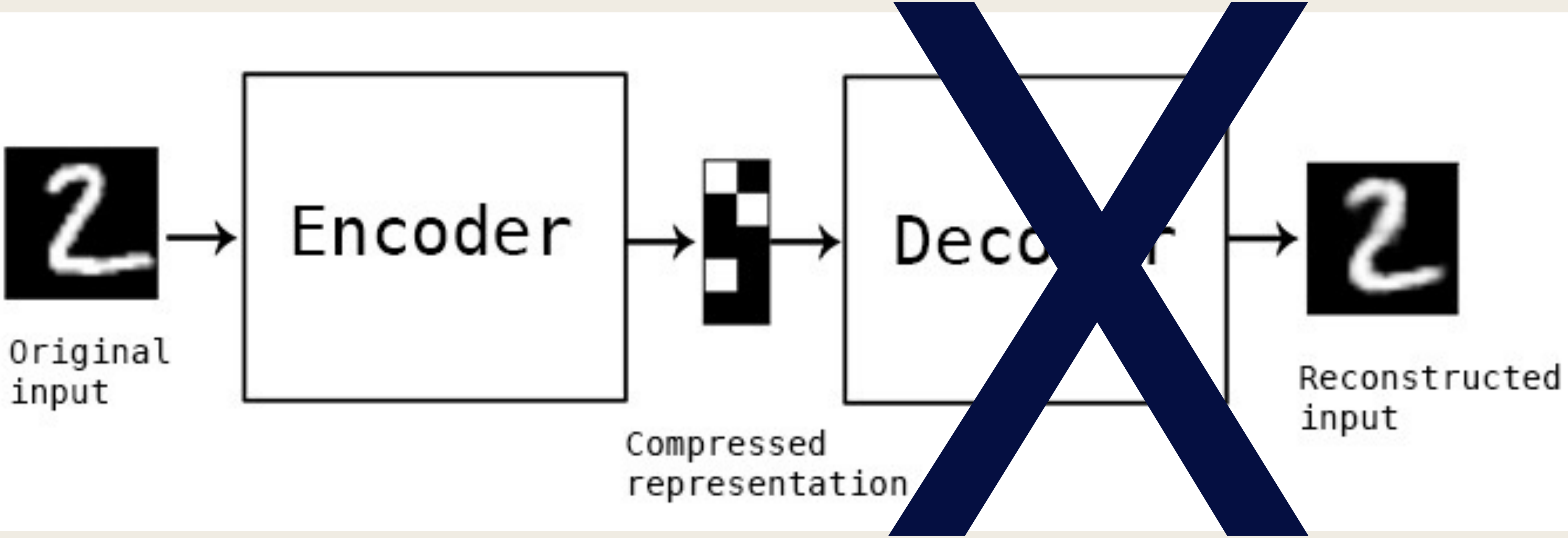


Input:

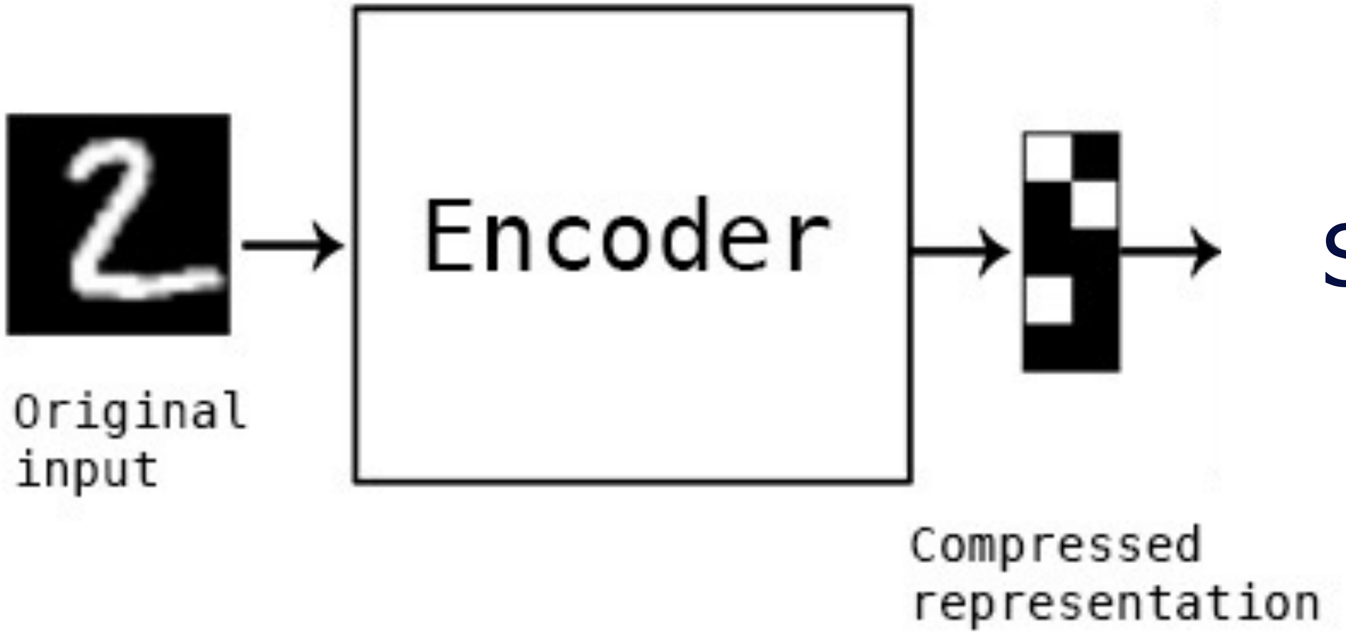
Output:

University of Stavanger

Pretraining



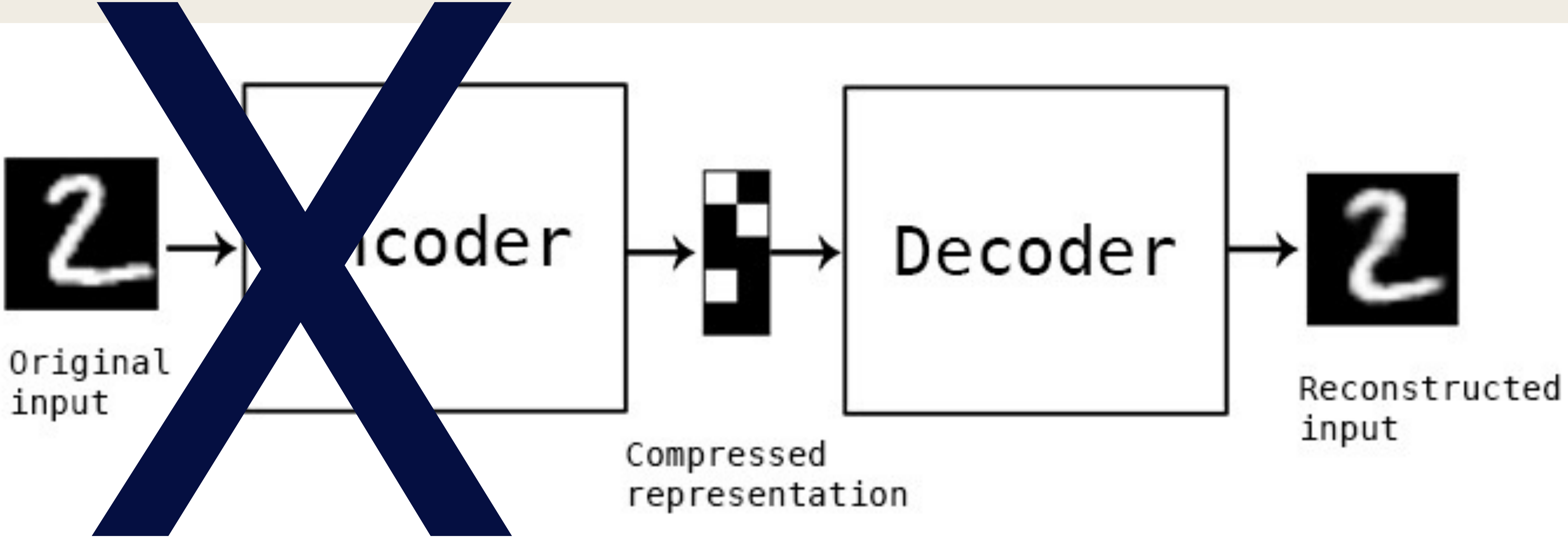
Pretraining



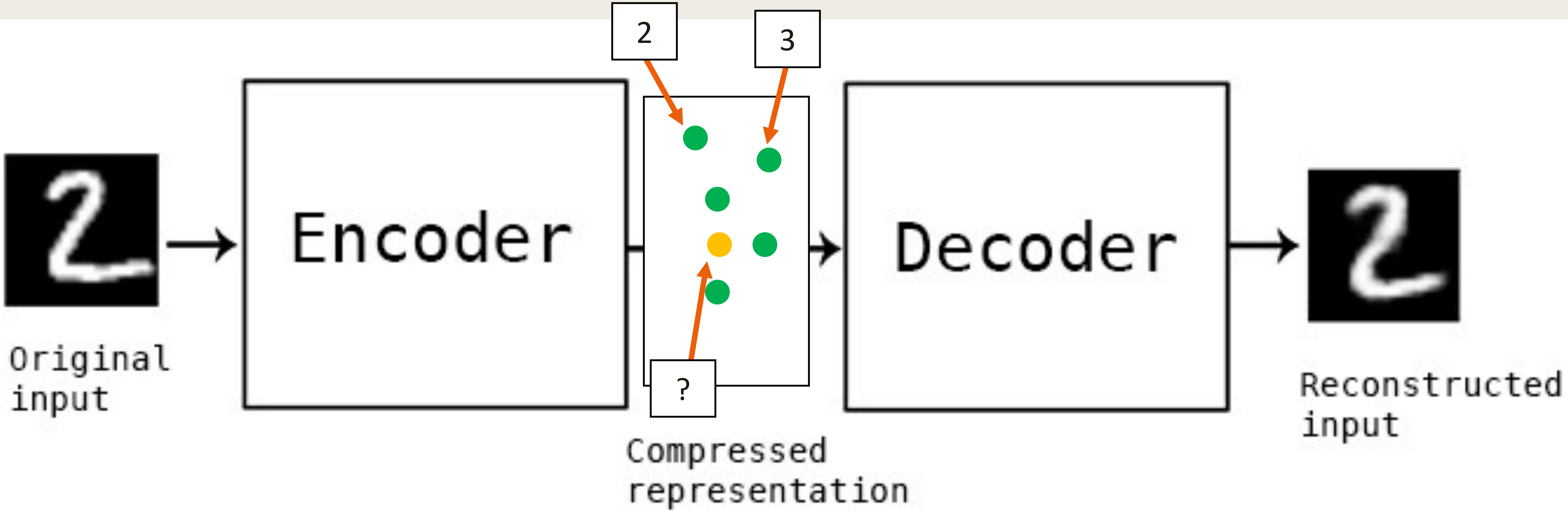
Softmax (10)

Semi-supervised training = pretraining + transfer training.

Data generation

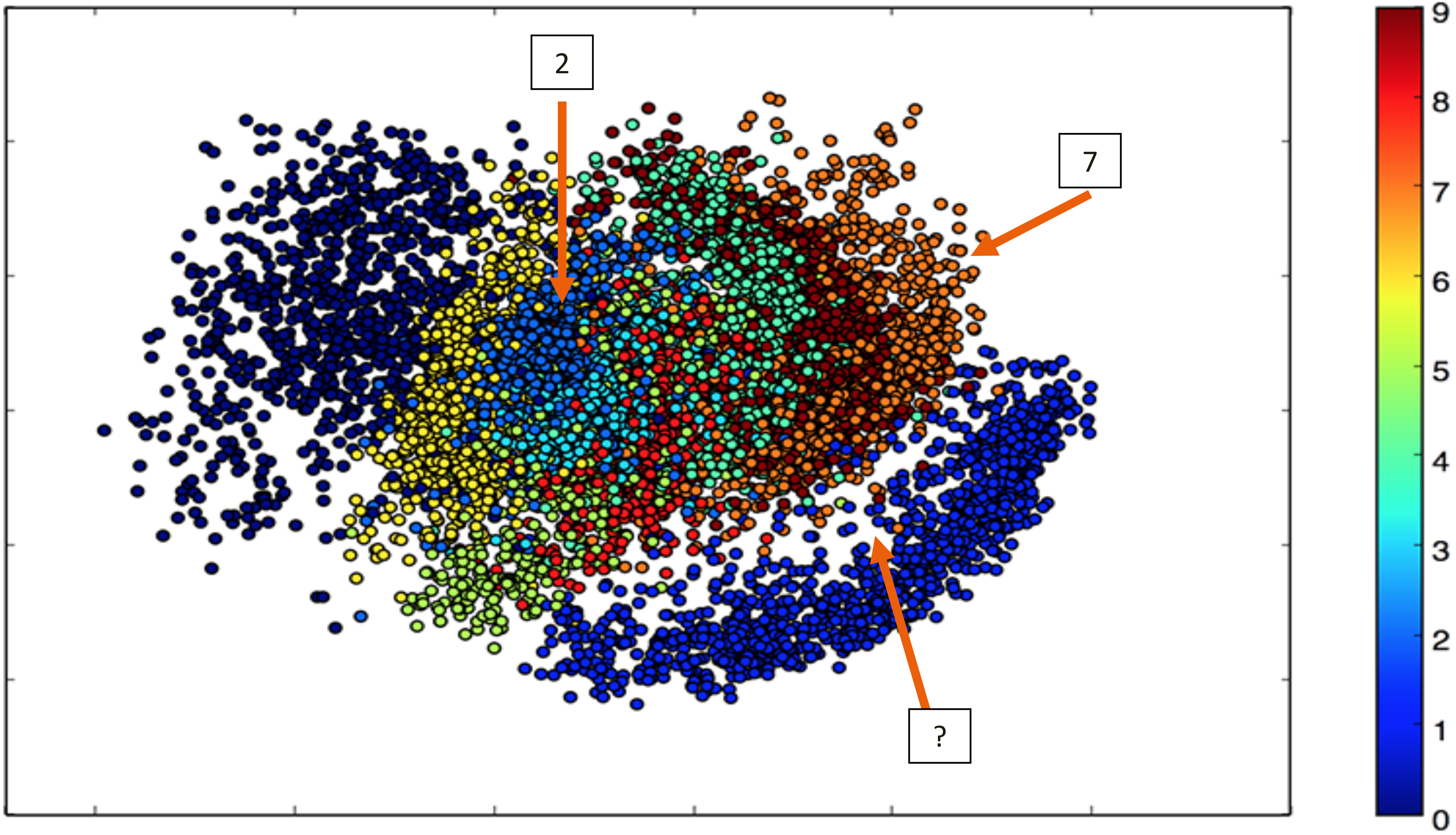


Data generation

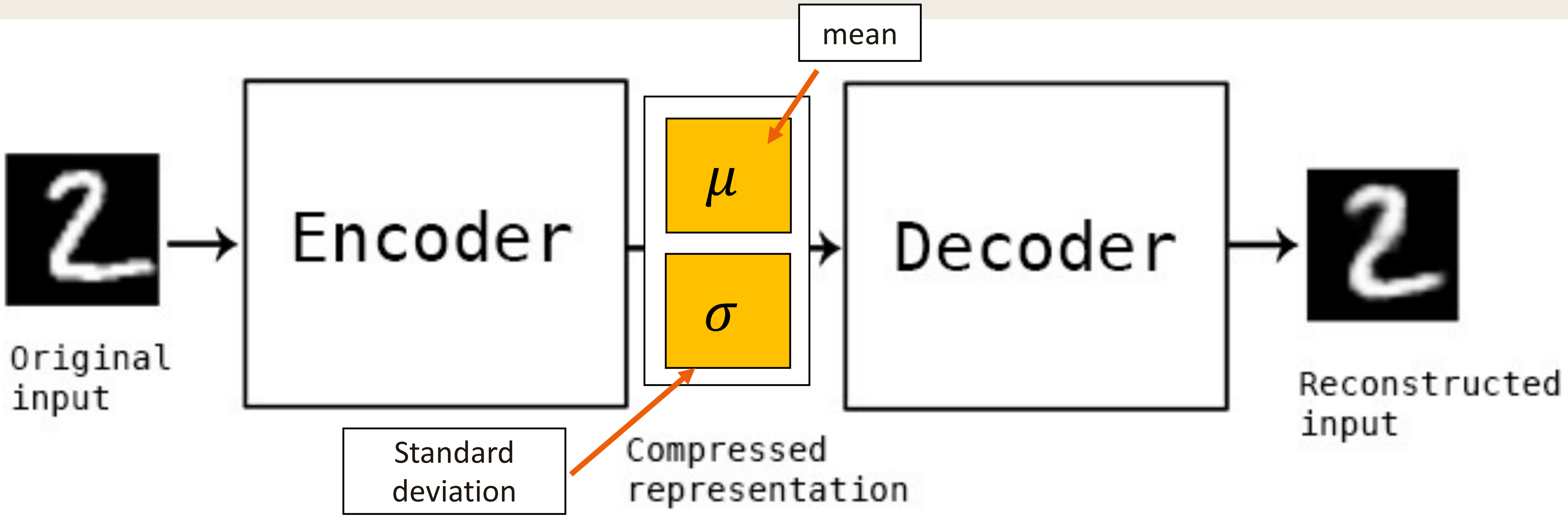


Semi-supervised training = pretraining + transfer training.

University of Stavanger


2
7
?
9
8
7
6
5
4
3
2
1
0

Variational autoencoder (VAE)



Loss function = reconstruction loss + Kullback-Leibler (KL) divergence loss

The reparameterization trick in VAE

Learned parameters (during backpropagation)

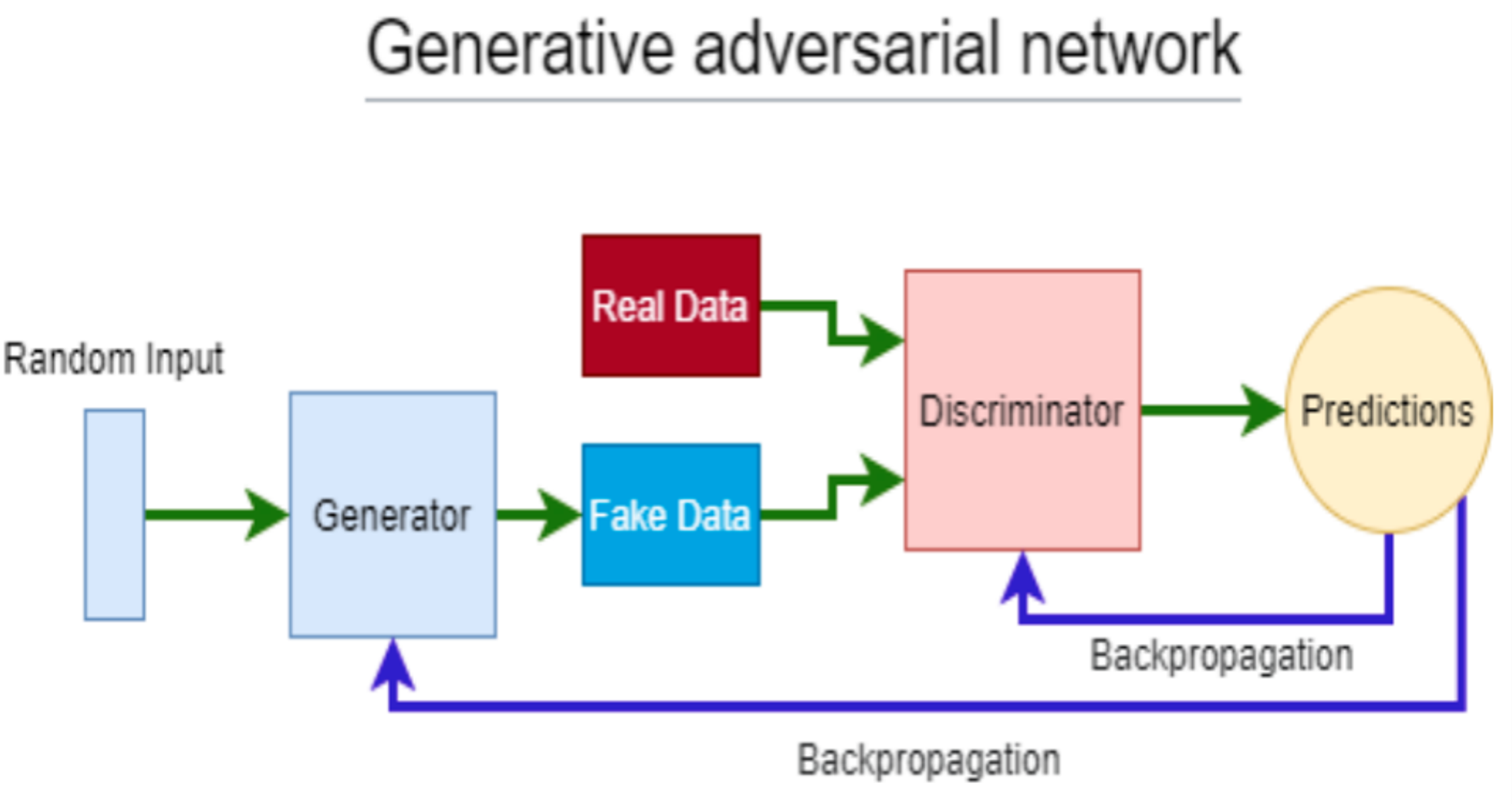
$$Z = \mu + \sigma \odot \varepsilon$$

- Standard normal distribution
- Fixed space where we randomly sample
- Not learned during backpropagation

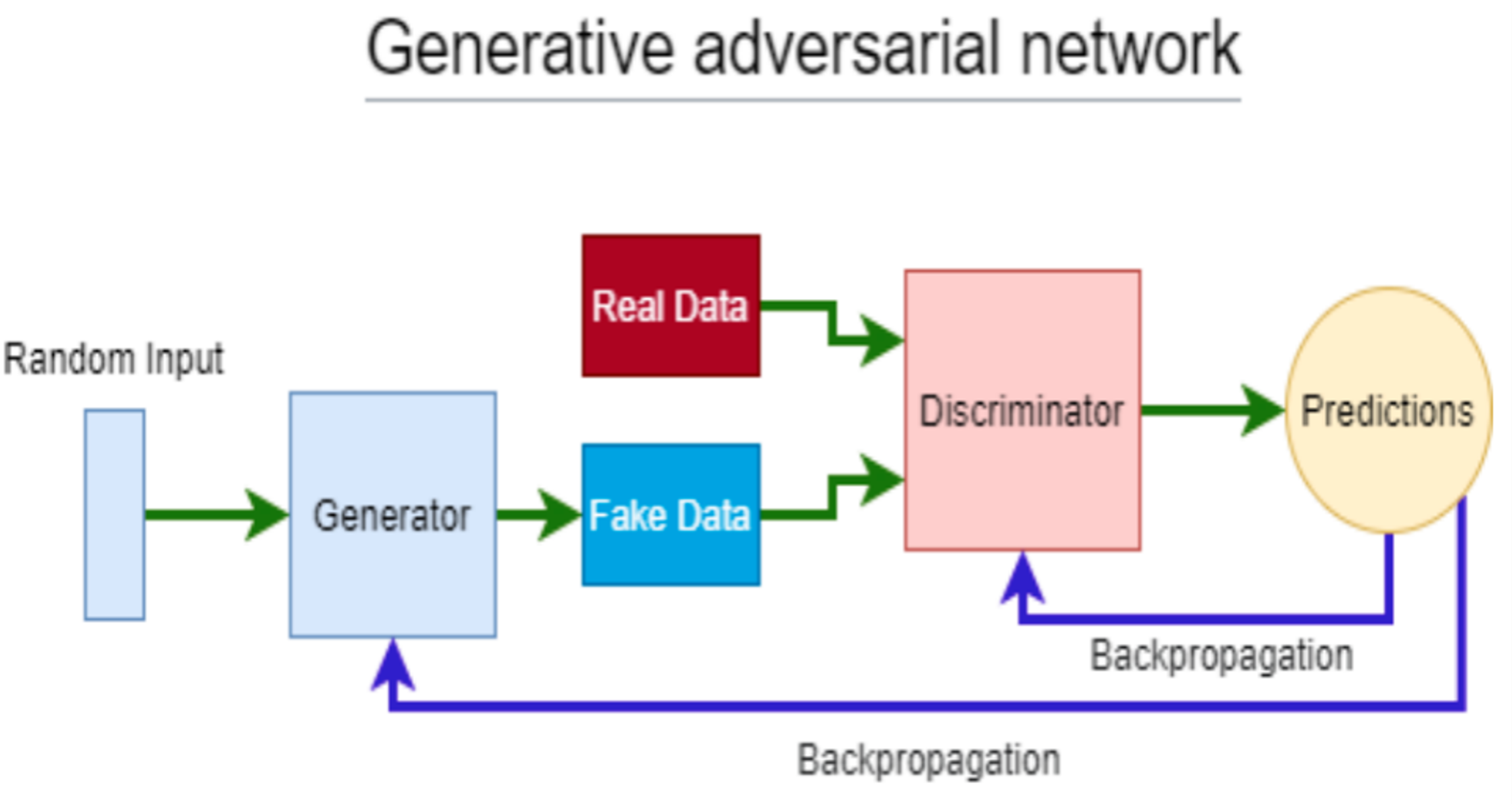
$$\varepsilon \sim \mathcal{N}(0, I)$$

University of Stavanger

GAN - Generative Adversarial Network



GAN - Generative Adversarial Network

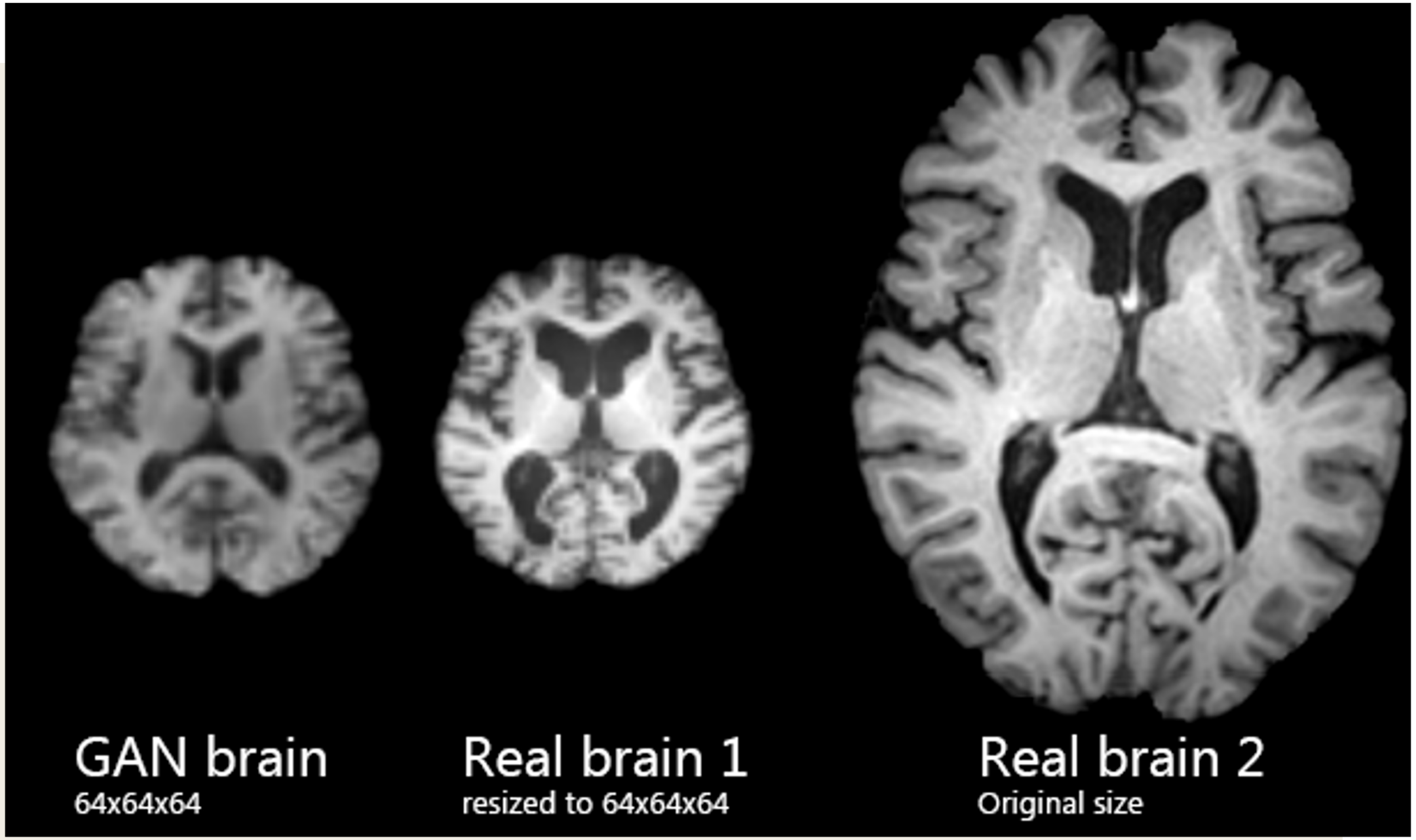


Weaknesses

- Non-convergence.
- Mode collapse.
- Diminished gradient.
- Unbalance between the training speed of the generator and discriminator.

Use cases

MSc. student Petter Minde



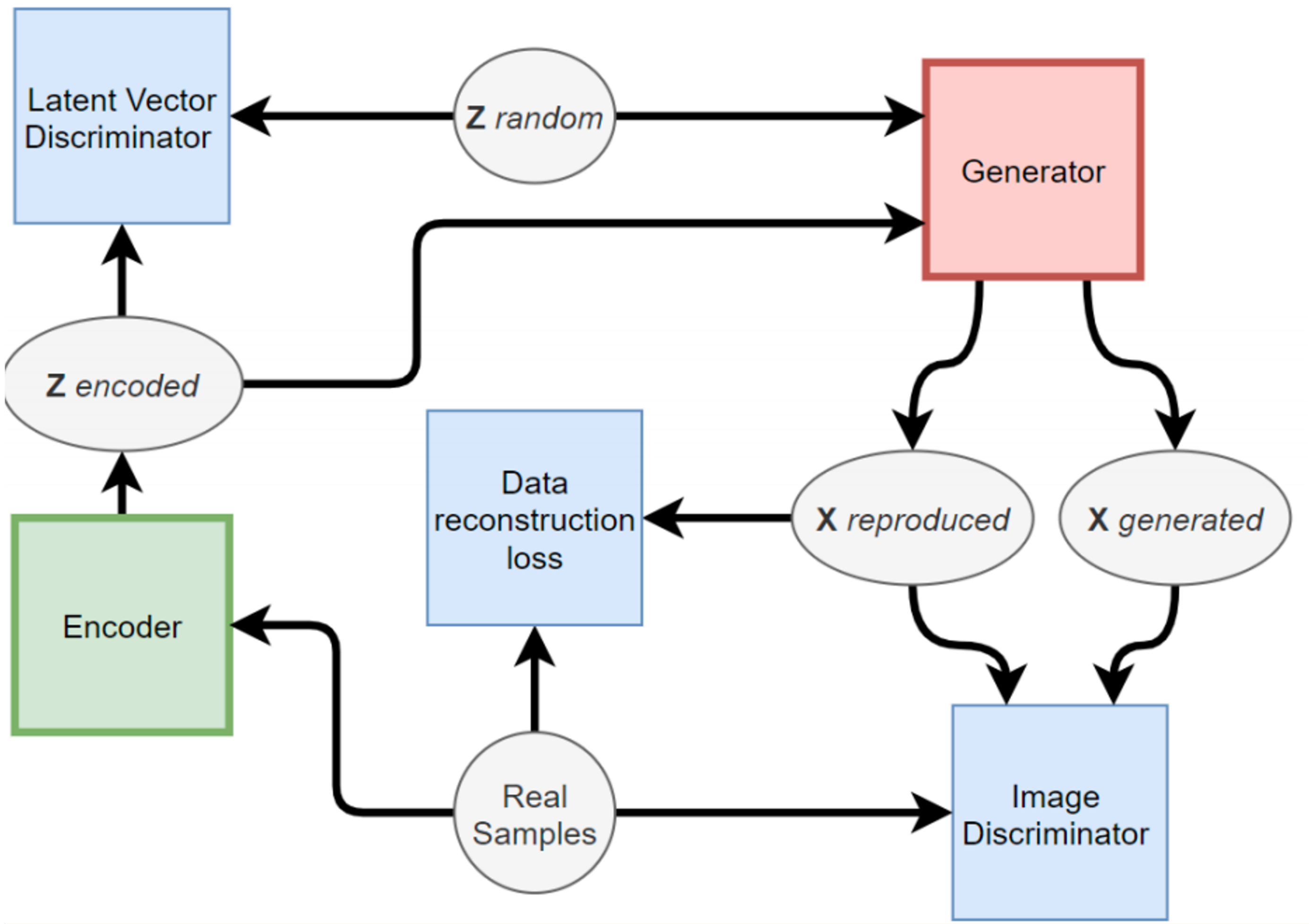
Most common uses are:

- Generating New images
- Upscaling images and Video
- Photo Inpainting

Our use:

- Generate new 3D images with different dementia diagnosis and supplement the real dataset with new “fake” data
- The brains had to be downscaled to 64x64x64 because it’s very computer intensive to generate 3D images

α-GAN



- 3D images struggles with the mode collapse because of the increased complexity
- VAE is free from mode collapse problem
  - Struggles with blurry images
- α-GAN combines GAN with a VAE to overcome mode collapse problem

GAN - Our results

6-fold CV Average results, with Standard Deviation(SD)				
	Validation Acc	SD	Test Acc	SD
GAN 33%	78,06	1,66	67,06	3,23
GAN 50%	82,32	3,11	67,15	4,21
GAN 50% w/Augmentation	82,32	1,58	67,15	3,19
Standard Dataset	66,71	4,21	63,55	2,02
Standard Dataset w/Augmentation	67,71	3,05	60,04	2,36

w/augmention = Roll 1-4 pixels in each direction, 95% chance of translation and 66% change for each direction, Rotate +-6 degree XY 90%.

Summary

In this lecture you have been introduced to:

- Autoencoders.
- Variational autoencoders (VAE).
- General adeversarial networks (GANs).
- How VAEs and GANs can be used in medical imaging appliations.

Taken from: https://www.centurysoft.com/blog/know-more-on-artificial-intelligence-ai.html