# (DEEP) GENERATIVE MODELS

## Variational Auto-Encoders and Generative Adversarial Networks

Mirko Polato, PhD
13.01.2021

# TABLE OF CONTENTS

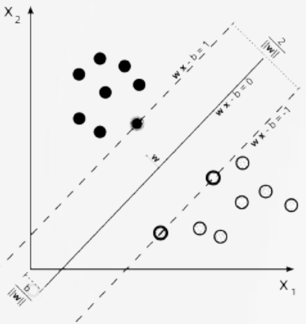
# 01

# GENERATIVE vs. DISCRIMINATIVE

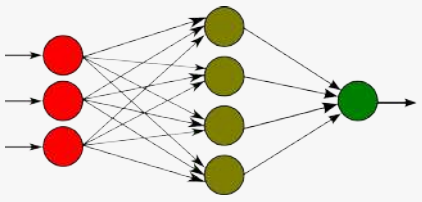
Learning from different perspectives

# DISCRIMINATIVE MODELS

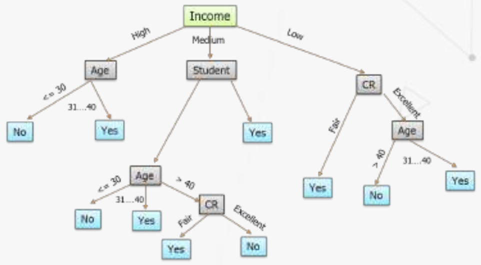
- Learn the **boundary** between classes
- Directly learn the conditional predictive distribution, $P(y|\boldsymbol{x})$



**SUPPORT VECTOR MACHINE**



**MULTI-LAYER PERCEPTRON**




**DECISION TREES**

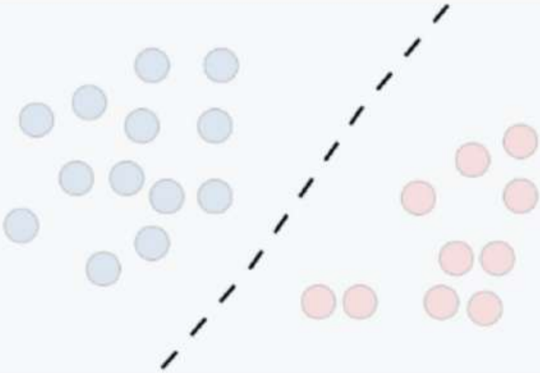
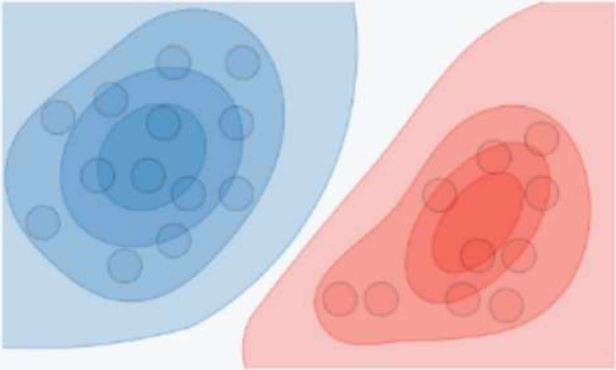
# GENERATIVE MODELS

- Can generate new data instances
- Capture the joint probability $P(\boldsymbol{x}, y)$ or just $P(\boldsymbol{x})$ if there are no labels
- Generative classifiers make the prediction by using Bayes rules

$$P(y|\boldsymbol{x}) = \frac{P(\boldsymbol{x},y)}{P(\boldsymbol{x})} = \frac{P(\boldsymbol{x}|y)P(y)}{P(\boldsymbol{x})}$$

- Naïve Bayes is an example of generative model:
  - Given y, you can draw a new example by sampling from $P(x_i|y)$

# DISCRIMINATIVE vs. GENERATIVE

| | Discriminative model | Generative model |
|---|---|---|
| **Goal** | Directly estimate $P(y\|x)$ | Estimate $P(x\|y)$ to then deduce $P(y\|x)$ |
| **What's learned** | Decision boundary | Probability distributions of the data |
| **Illustration** |  |  |

# DISCRIMINATIVE vs. GENERATIVE

| Feature | Discriminative | Generative |
|---|---|---|
| Support Unlabeled data | No | Yes |
| Can generate data | No | Yes |
| Can perform classification | Yes | Yes |
| Classification performance | Best | Very good |
| Computational complexity | Medium/High | High |
| Assumptions | Some | Many |
| Outlier detection | No | Yes |

# 02

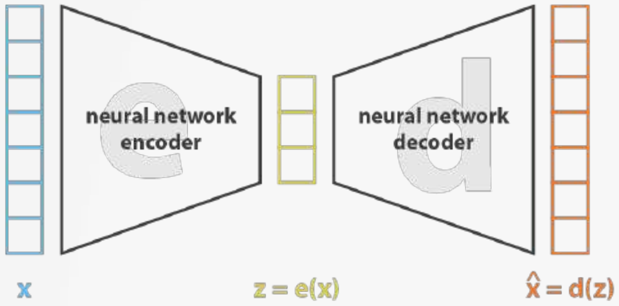
# VARIATIONAL AUTOENCODERS

From Auto-Encoders to Variational Auto-Encoders

# AUTO-ENCODERS (AE)

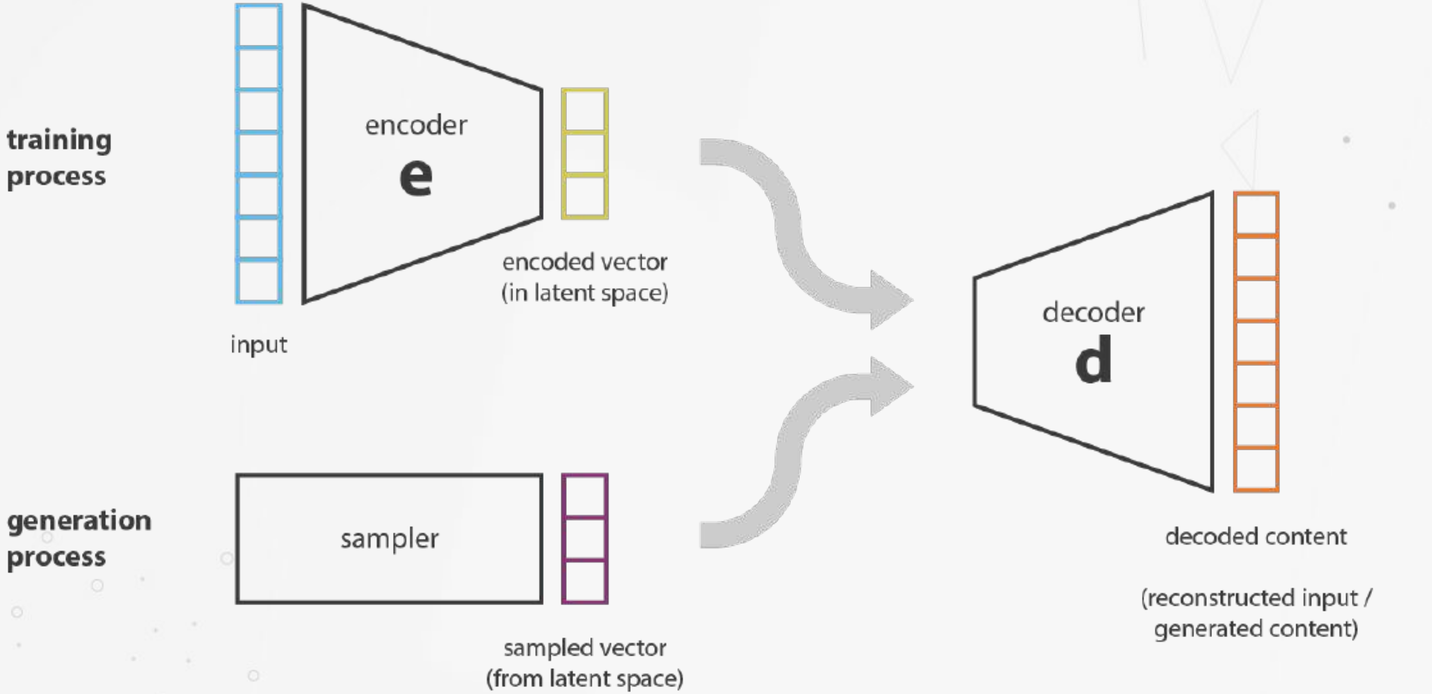
- Vanilla auto-encoders learn to represent (i.e., **encode**) the input in a lower dimensional space, while keeping the ability to reconstruct it (e.g., **decode**) as accurately as possible

- The **code** is said to be the latent representation of the input

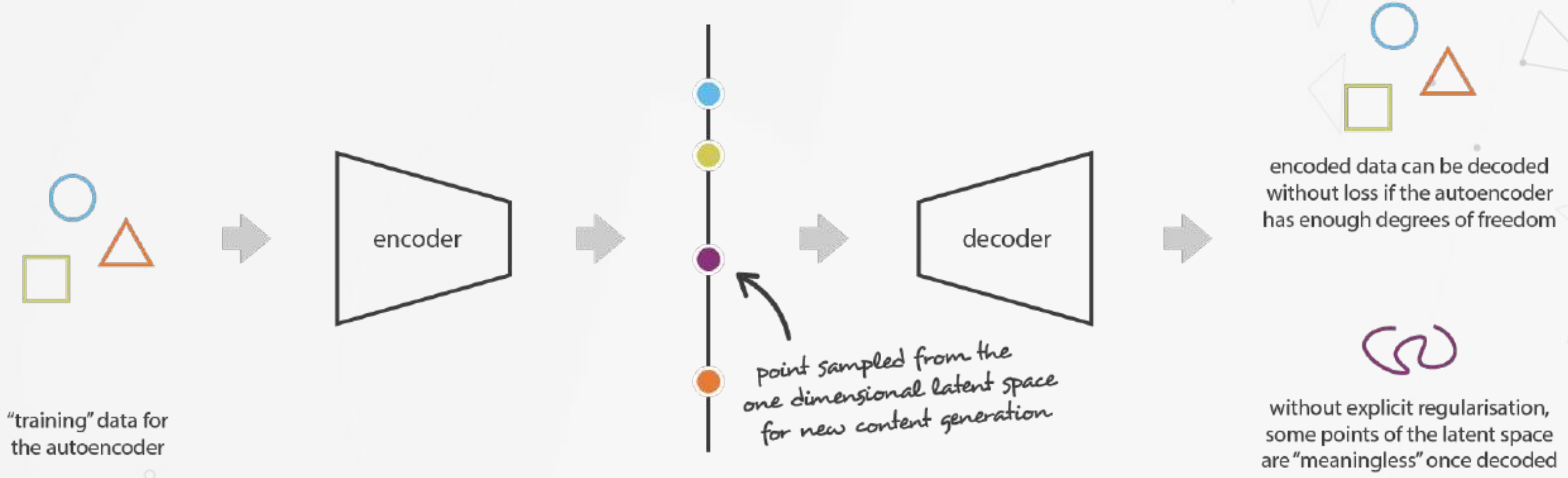



$$\text{loss} = \|x - \hat{x}\|^2 = \|x - d(z)\|^2 = \|x - d(e(x))\|^2$$

# AEs AS GENERATIVE MODELS



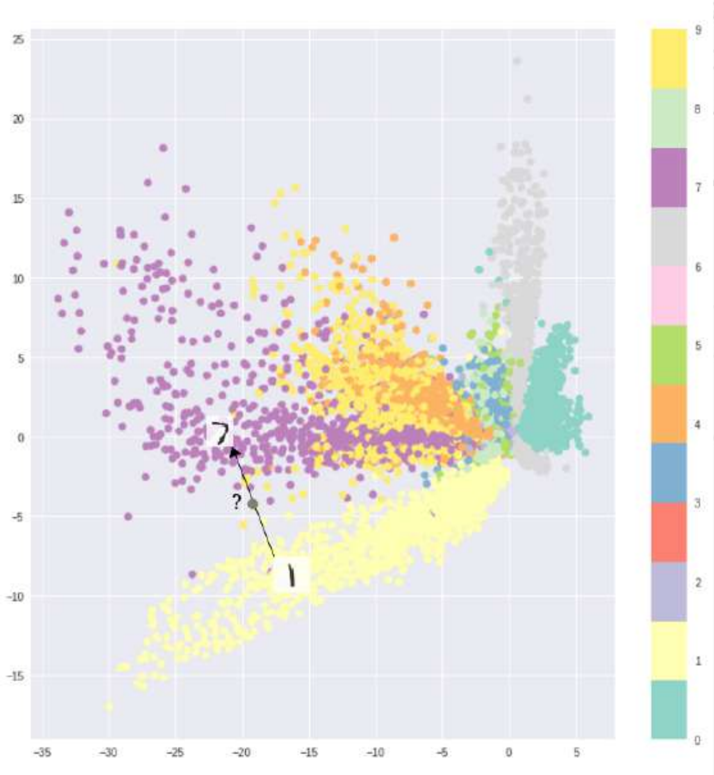
# AEs AS GENERATIVE MODELS, WHY NOT?




**Irregular latent spaces** make autoencoders not ideal for new content generation!

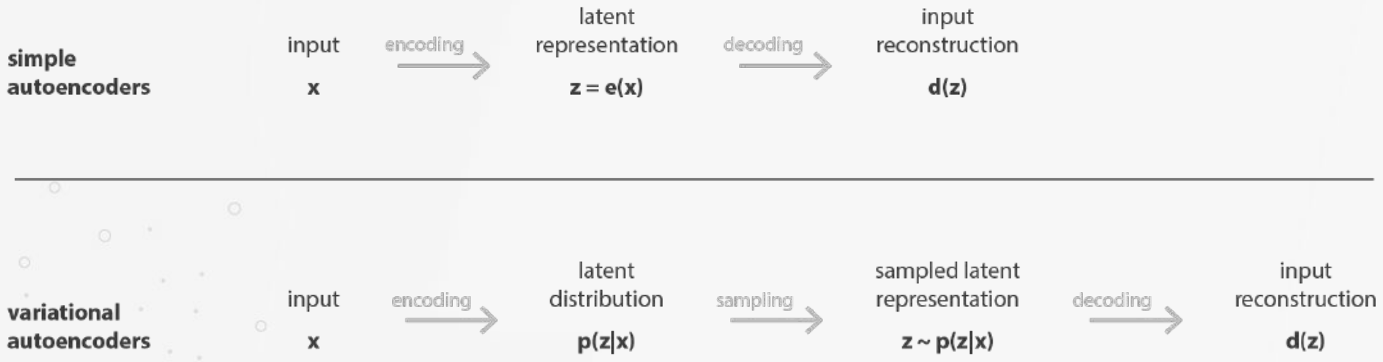
# AE's LATENT SPACE ON MNIST

- MNIST: dataset of handwritten digits
- GOOD: different digits are mapped on 'different' latent space regions
- BAD: latent space is not continuous
- What if we pick a latent representation outside the known regions?
  → The deconding would fail!!



# VARIATIONAL AUTO-ENCODERS (VAE)

- AE regularized to avoid overfitting
- AE regularized to ensure good properties of the latent space
  ➔ enables the generative process
- Encode inputs as a distribution over the latent space



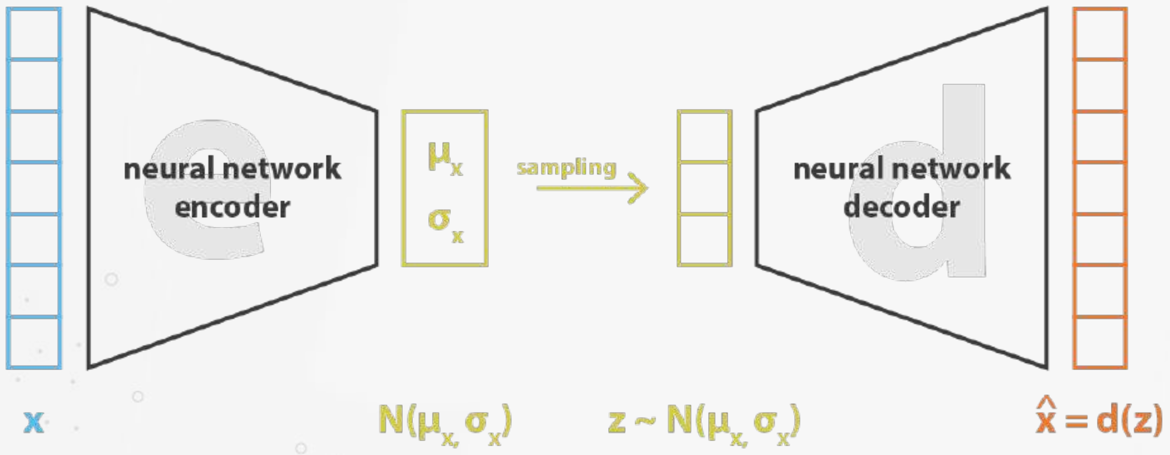
# GOOD LATENT SPACE PROPERTIES

- **Continuity**: close points in the latent space should remain (sufficiently) close also in the input space when decoded

- **Completeness**: for a given distribution, a point sampled from the latent space should give “meaningful” content once decoded

# VAE ARCHITECTURE

- The encoded distributions are (tipically) chosen to be standard gaussian
- The encoder outputs the **distribution means** and **variances**, separately
- The decoding is performed on a code (z) **sampled** from the latent distribution

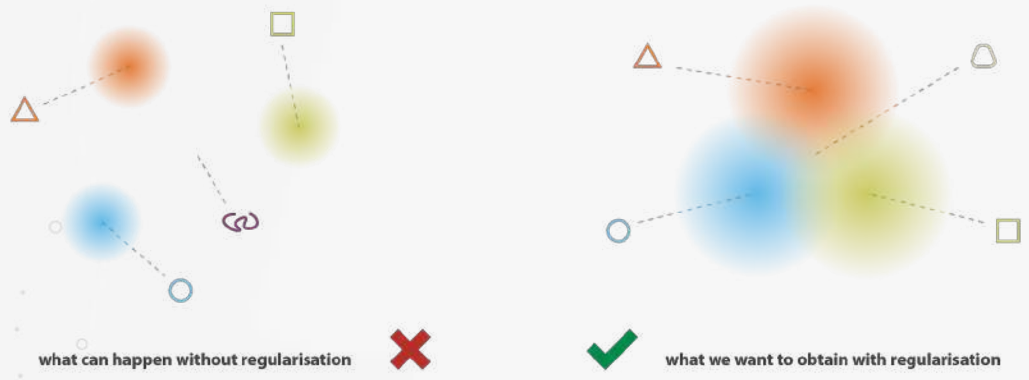


# VAE w/o REGULARIZATION

Encoding to distributions instead of points is not enough!

In terms of pure reconstruction error a ***not regularized*** VAE can:

- Return distributions with tiny variances → no completness
- Return distributions with very distant means → no continuity



# VAE's REGULARIZED LOSS

$L_{VAE}$ = **reconstruction loss** + **KL loss**

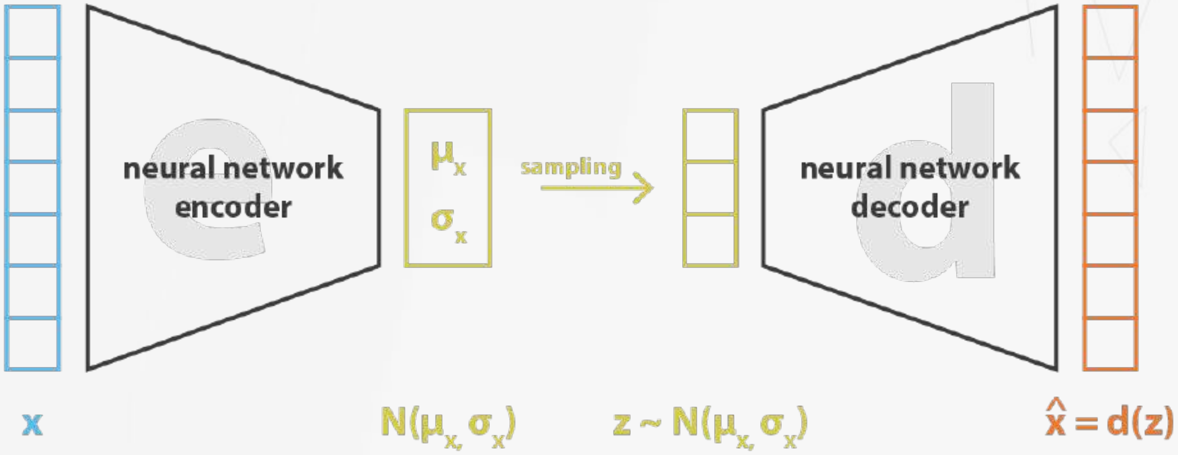
Standard AE loss: computes how far is the output w.r.t. the input

**Kulback-Leibler divergence**: Measures the difference between the latent factors' distribution and the target ones (usually standard Gaussian)

Force the covariance matrix being identity-like preventing punctual distribution

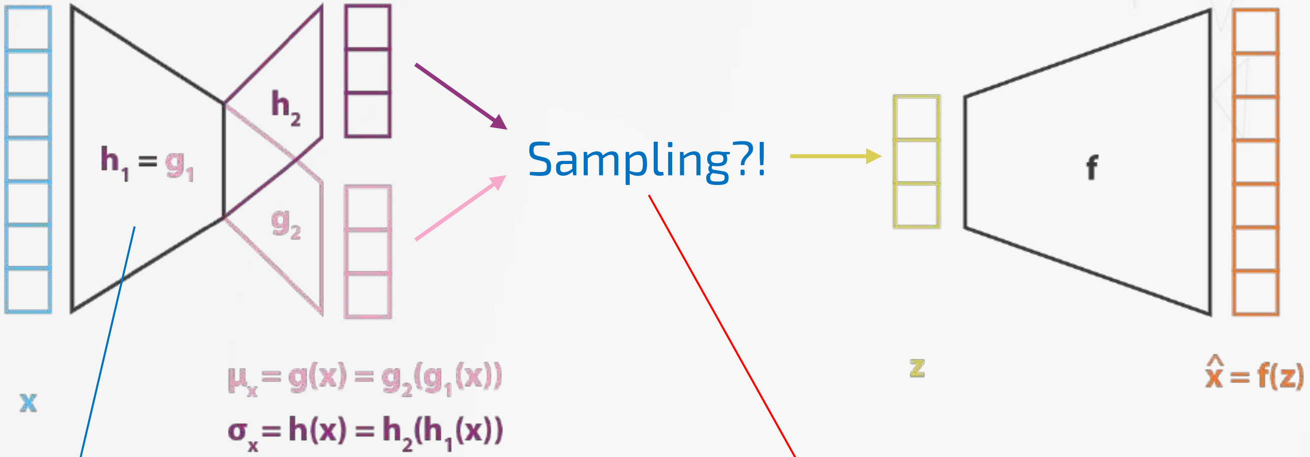
Force the means to be close to 0 preventing of having far apart distributions

# VAE ARCHITECTURE




$$\text{loss} = \| x - \hat{x} \|^2 + \mathrm{KL}[\, N(\mu_x, \sigma_x), N(0, I) \,] = \| x - d(z) \|^2 + \mathrm{KL}[\, N(\mu_x, \sigma_x), N(0, I) \,]$$

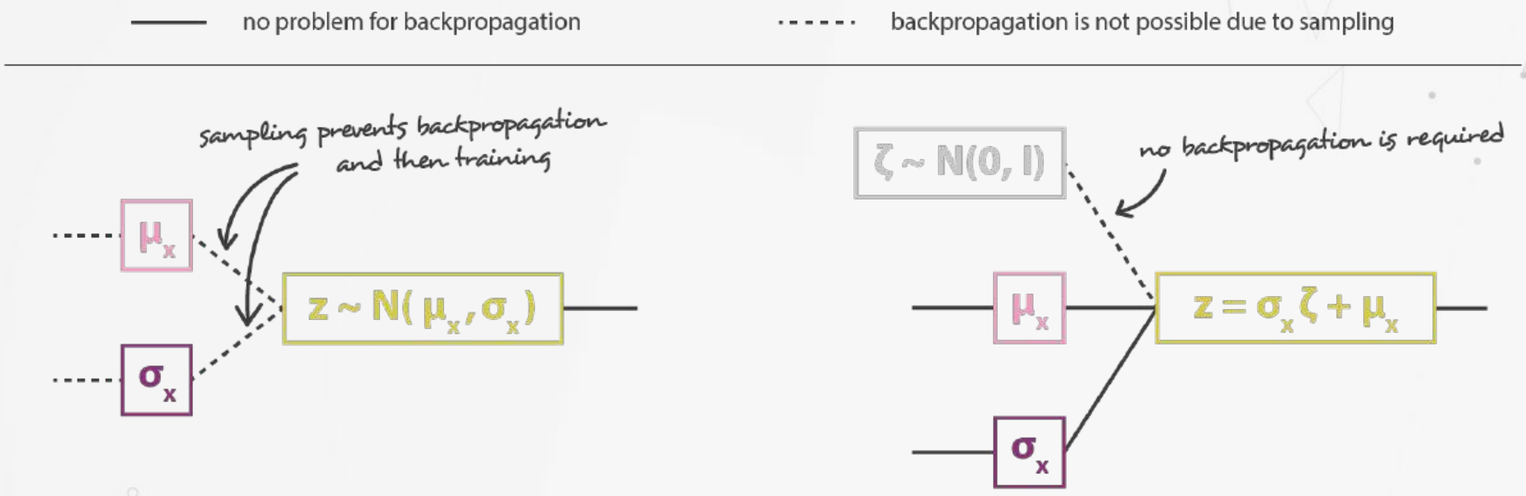
# VAE AS A NEURAL NETWORK




Parameters' sharing

The sampling operation does not allow the error to be backpropagate!

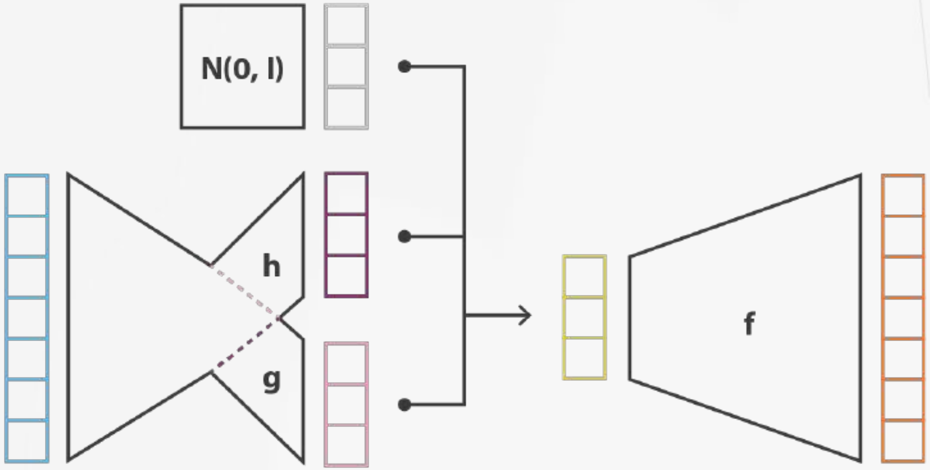
# “REPARAMETRIZATION” TRICK




**sampling without reparametrisation trick**

**sampling with reparametrisation trick**

# VAE FULL NN ARCHITECTURE




$x$

$\mu_x = g(x)$
$\sigma_x = h(x)$
$\zeta \sim N(0, I)$

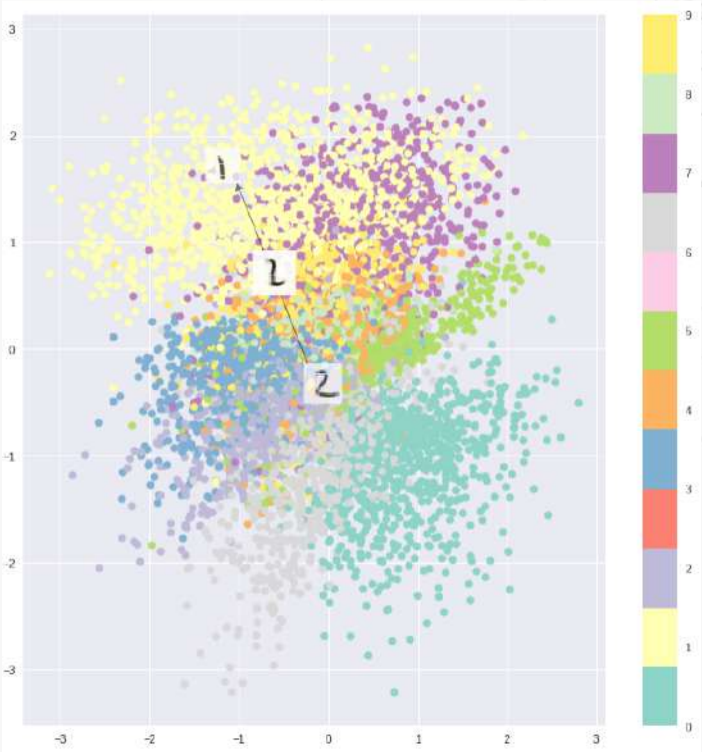
$z = \sigma_x \zeta + \mu_x$

$\hat{x} = f(z)$

$$\text{loss} = C\,\|x - \hat{x}\|^2 + KL[\,N(\mu_x, \sigma_x), N(0, I)\,] = C\,\|x - f(z)\|^2 + KL[\,N(g(x), h(x)), N(0, I)\,]$$

# VAE's LATENT SPACE ON MNIST

- **MNIST**: dataset of handwritten digits
- GOOD: different digits are mapped on 'different' (but somewhat overlapping) latent space regions
- GOOD: latent space is much more continuous and complete w.r.t. AE's one
- Points in "middle earth" regions are decoded to blended representations!



# 03

# GENERATIVE ADVERSARIAL NETWORKS

Game theory meets generative learning

*“GAN is the most interesting idea in the last 10 years in Machine Learning”*
*Yann LeCun*

# GENERATIVE PROCESS

## What is the idea behind a generic generative process?

EXAMPLE: "Complex" random variable generation

(pseudo) random variable

**Q**: Is this magic??

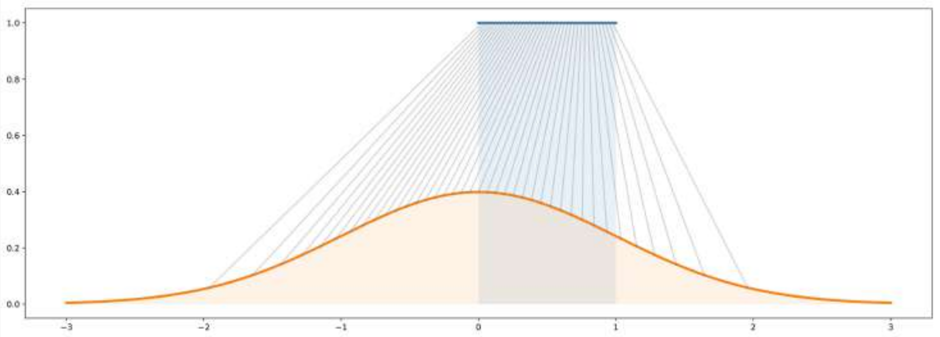
**A**: The machine generates (through a function) random variables as the result of a function applied to uniform random variables.

# GENERATIVE PROCESS

**Q**: What if we want to sample a random number from the standard Gaussian distribution?

**A: Inverse transform method**

$$f\colon x \sim U(0,1) \;\rightarrow f(x) \sim N(0,1)$$



# GENERATE REALLY COMPLEX RANDOM VARIABLES

**Q**: What if we want to generate dog pictures?

Rephrase

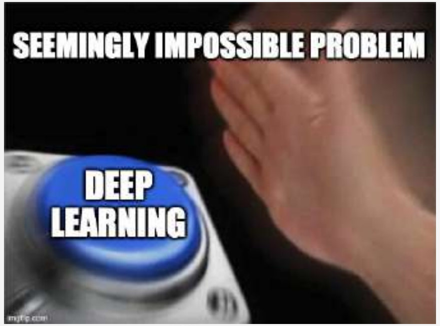
Generate new vectors following the "dog probability distribution" over the N dimensional vector space.

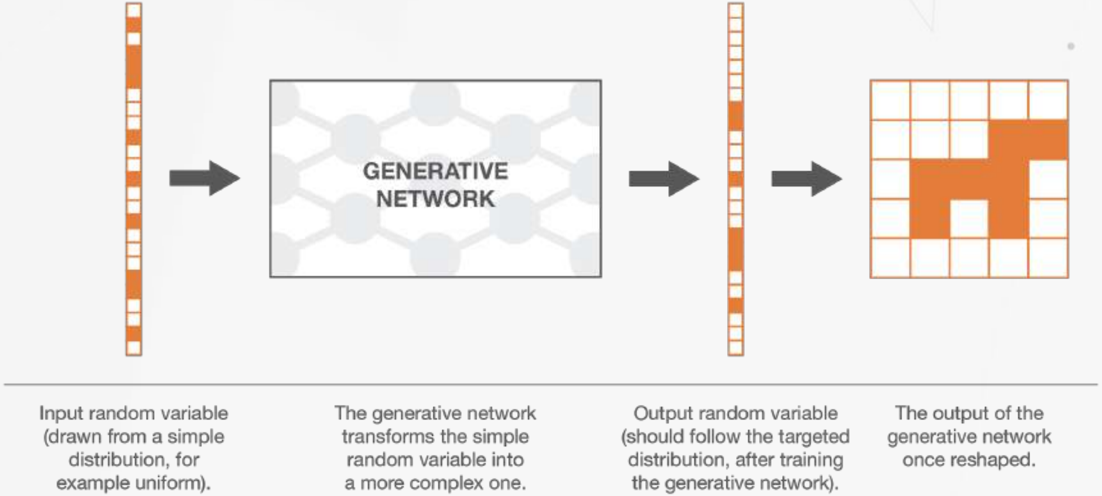
Very complex distribution


Even if it exists - We don't know how to express it explicitly

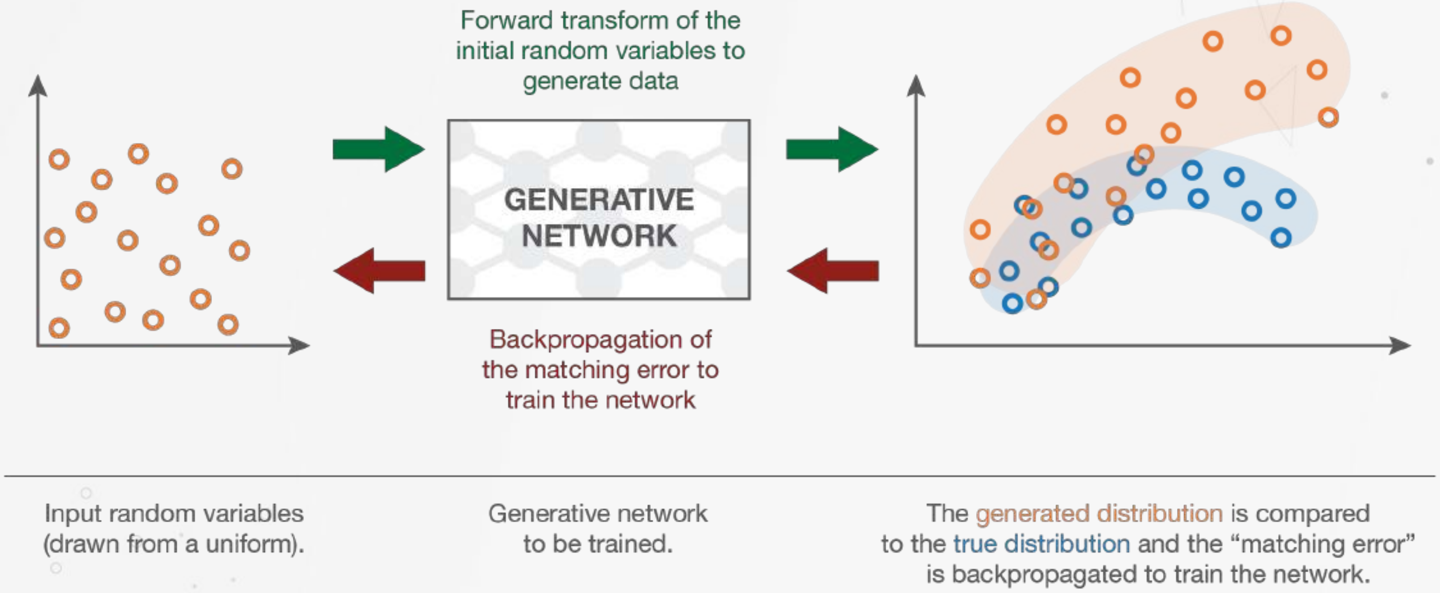
# GENERATIVE NEURAL NETWORK

In general, the functional f is hard to define → Lets use a **neural network**!!

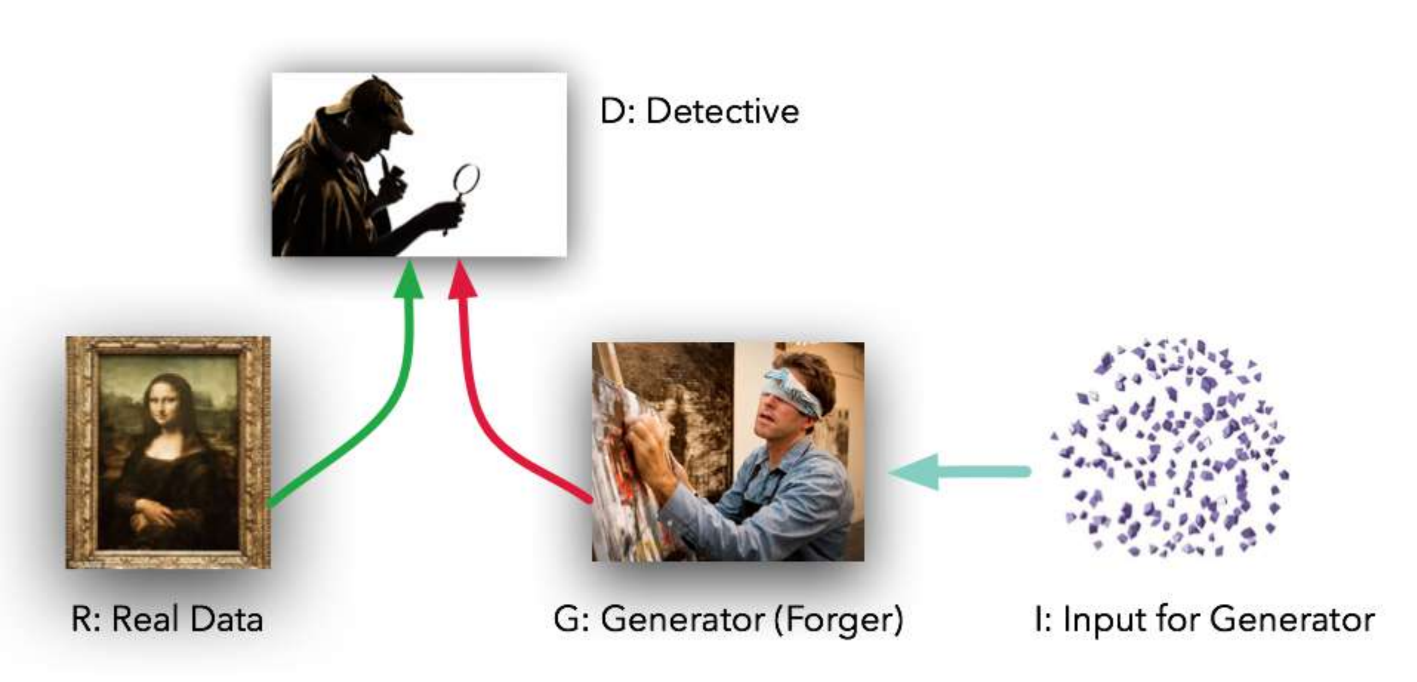


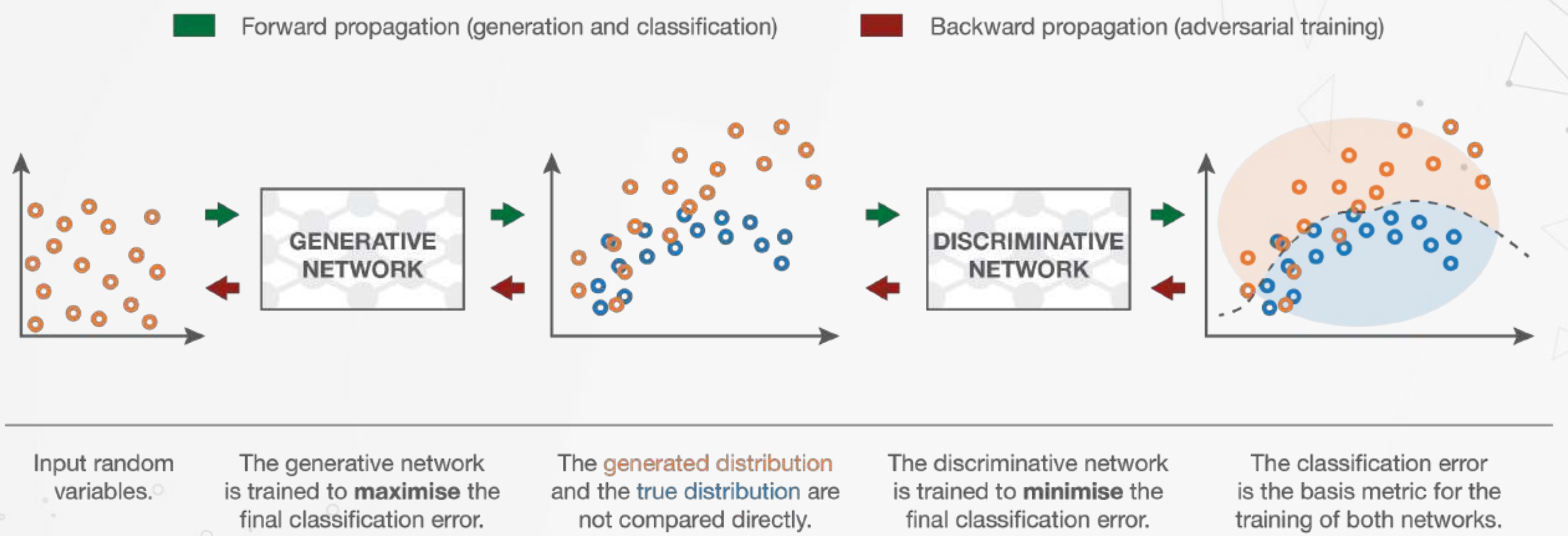
# GENERATIVE MATCHING NETWORK (GMN)



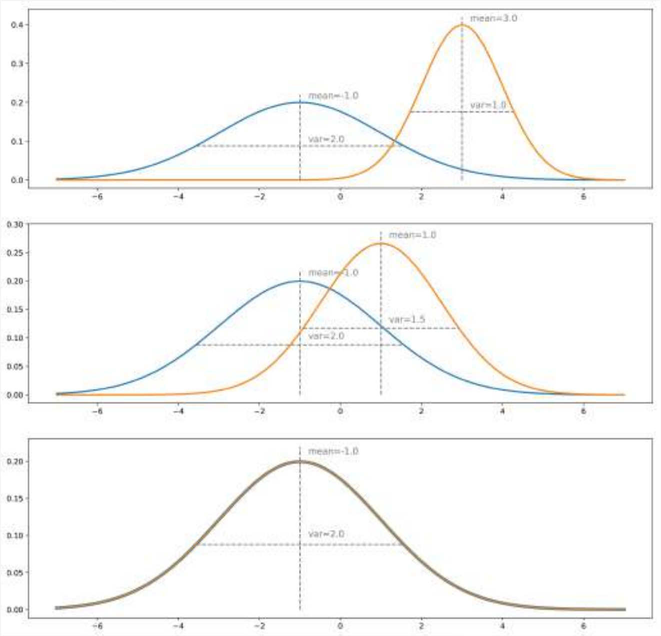
# GAN: GENERATIVE ADVERSARIAL NETWORK



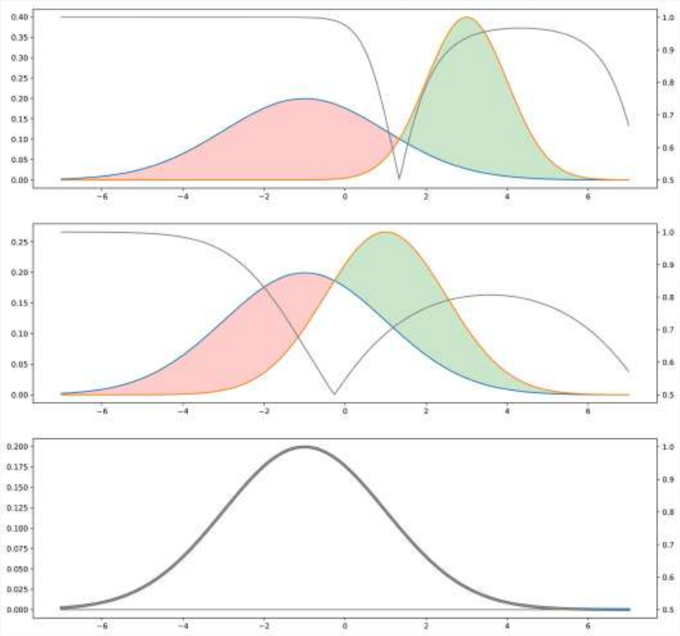
# GAN: TRAINING OVERVIEW



# GMN vs GAM: IDEAL LEARNING

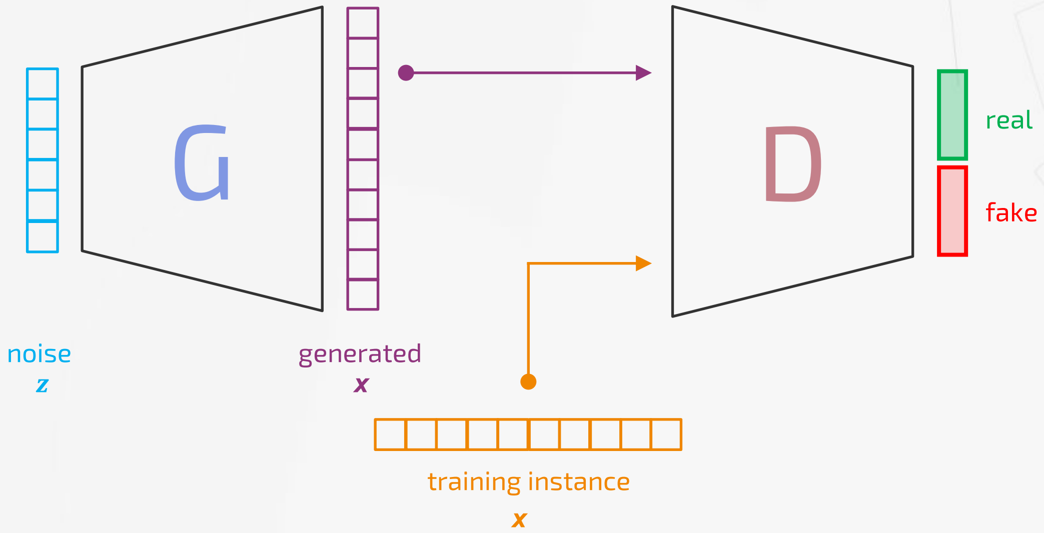



**GMN**



**GAN**

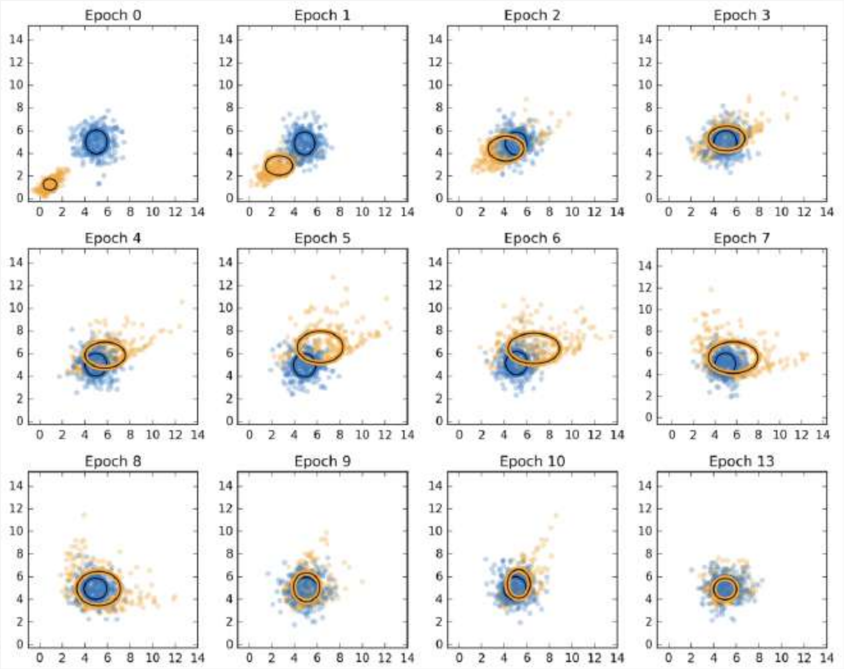
# GAN ARCHITECTURE



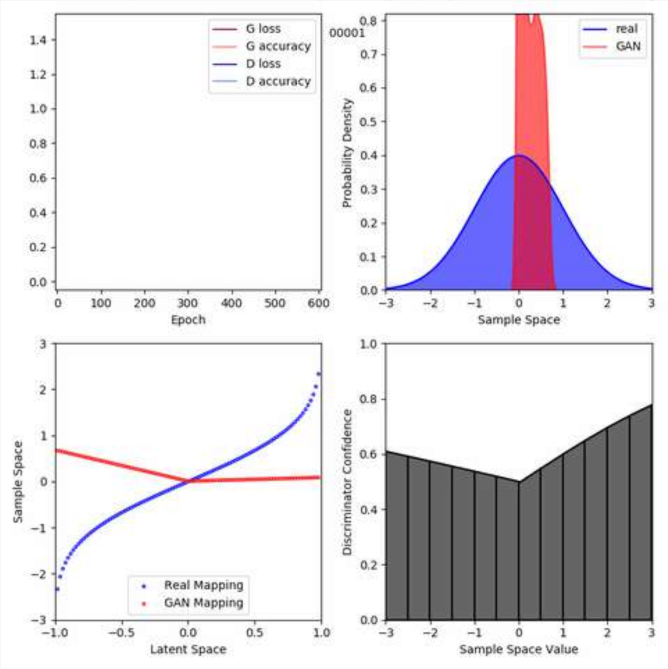

$$\min_G \mathrm{E}_{\mathbf{z}}[\log(1 - D(G(\mathbf{z}))] \qquad \max_D E_{\boldsymbol{x}}[\log(D(\boldsymbol{x}))]$$

$$\min_G \max_D E_{\boldsymbol{x}}[\log(D(\boldsymbol{x}))] + \mathrm{E}_{\mathbf{z}}[\log(1 - D(G(\mathbf{z}))]$$

# GAN: TRAINING EXAMPLES



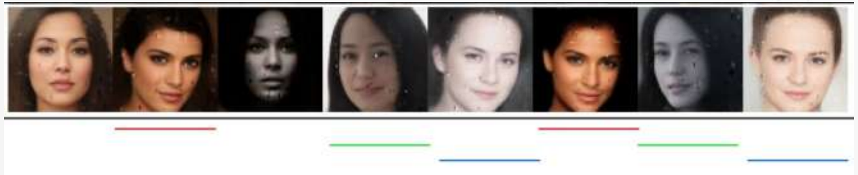

GAN's epochs on a 2D generative task




GAN to draw samples from the standard normal distribution N(0, 1)

# WHY DO GANs ARE HARD TO TRAIN?

- Very **delicate** **balance** **between discriminator and generator**
  - Mode collapse



  - Non-convergence / unstable gradient

Overspecialized generator

- **Hyper-parameters tuning**

# WHY GANs ARE (GENERALLY) PREFERRED TO VAEs FOR IMAGE GENERATION?

Input

VAE reconstruction

# 04

# APPLICATIONS

GANs and VAEs in action!

# GAN: DATA AUGMENTATION (GENERATION)

**Quiz time!**

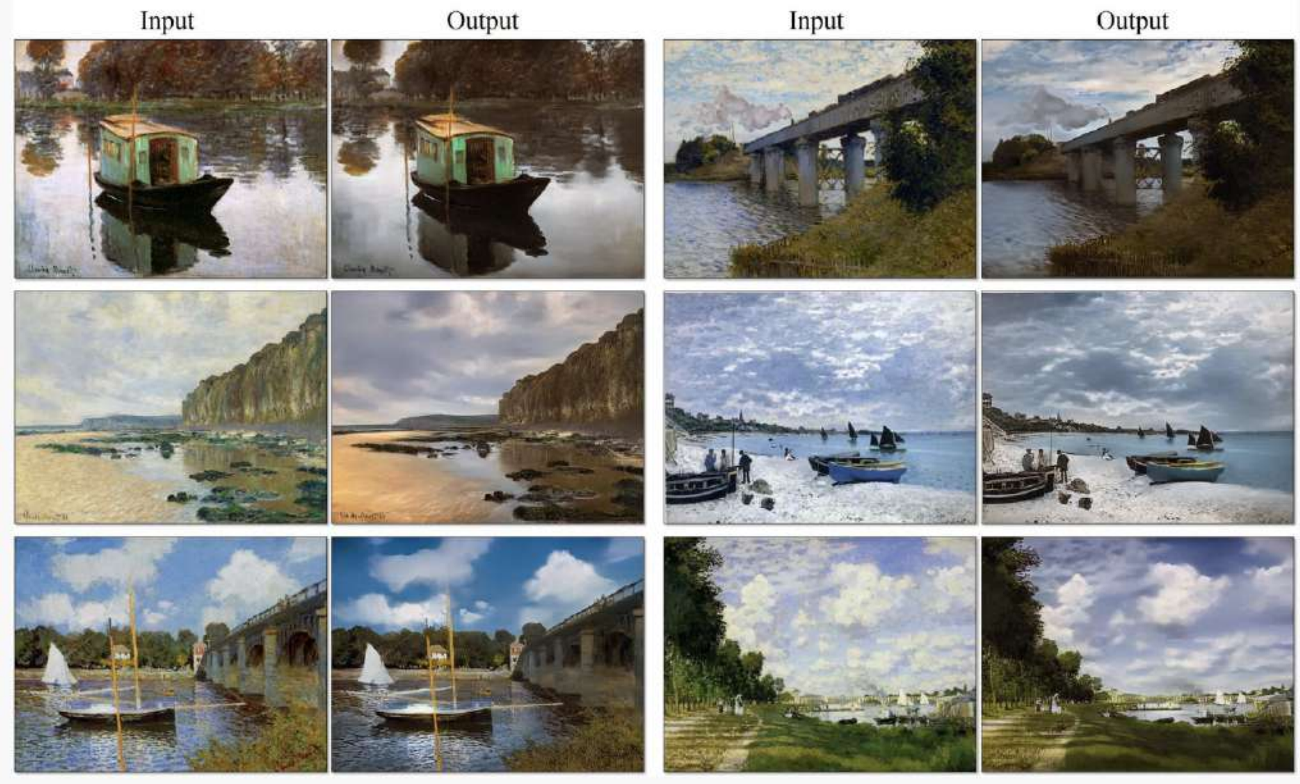
A

B

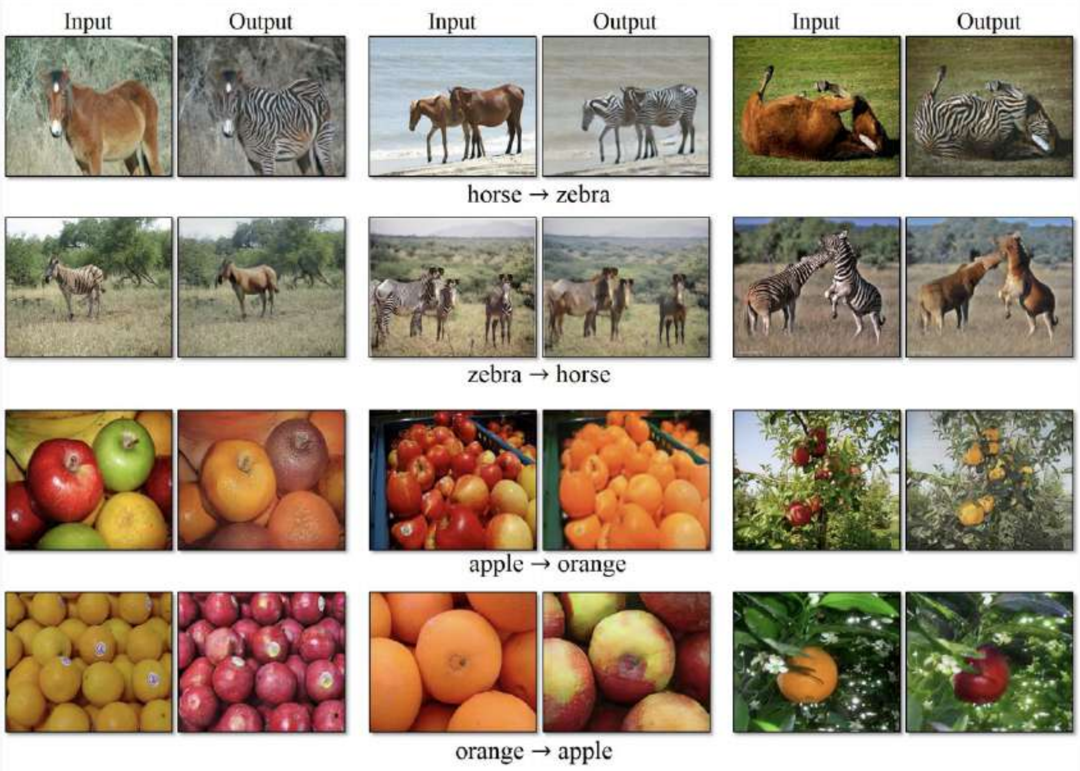
C

D

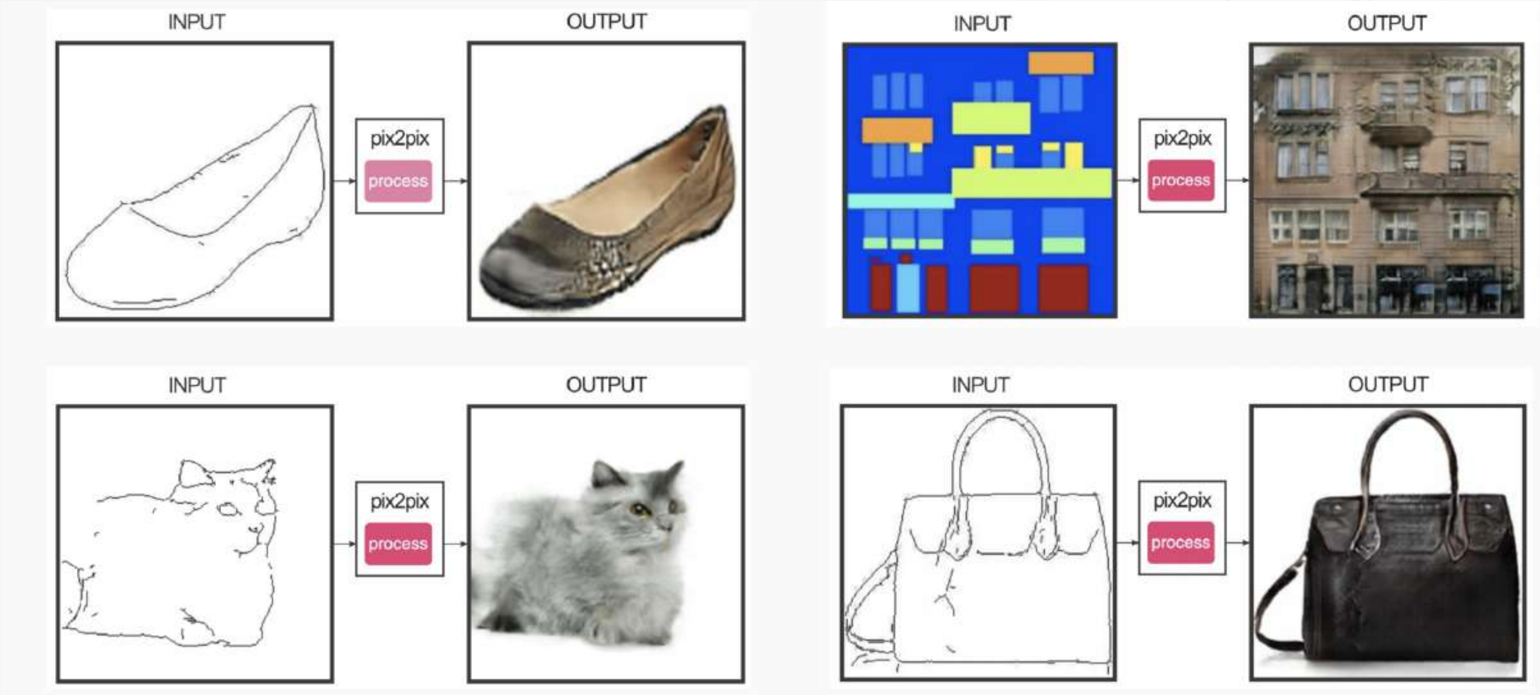
# GAN: STYLE TRANSFER



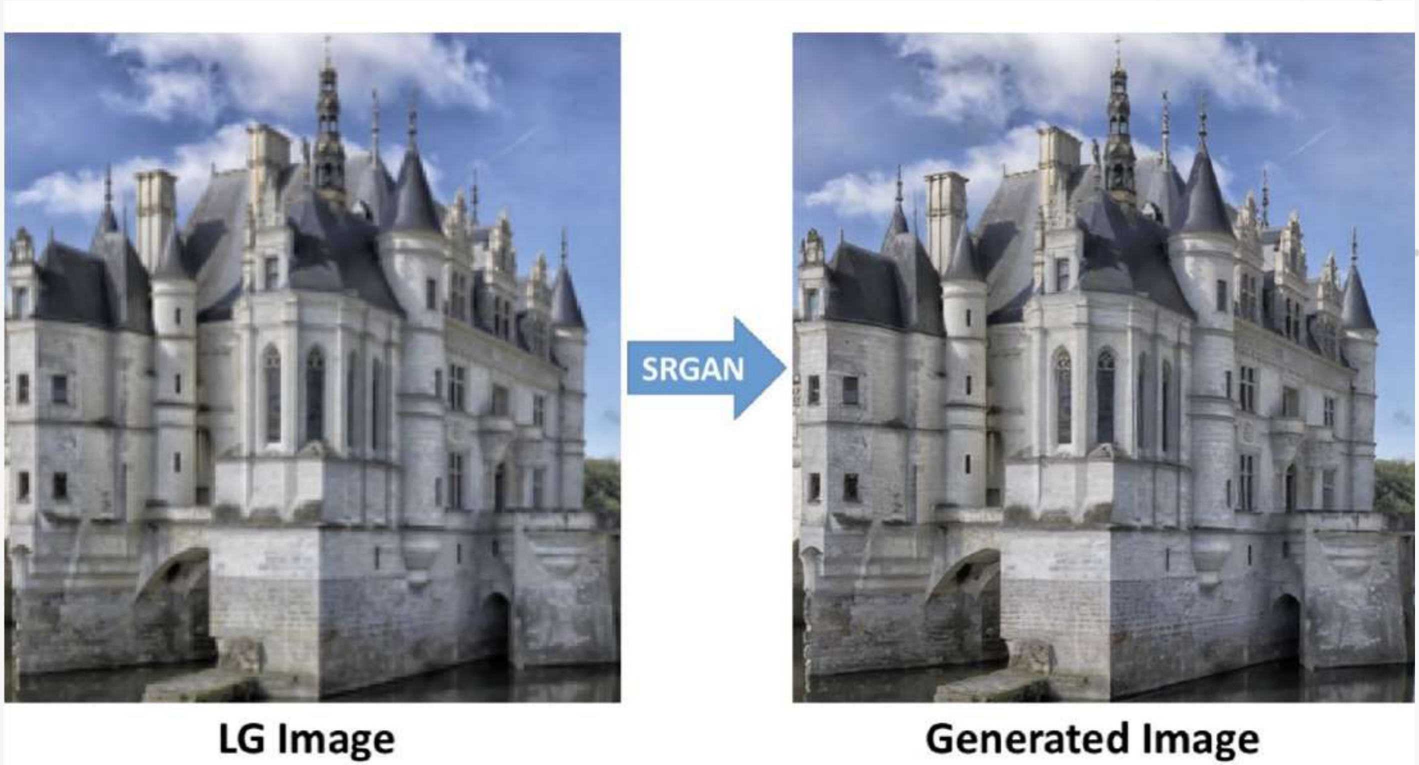
# GAN: STYLE TRANSFER



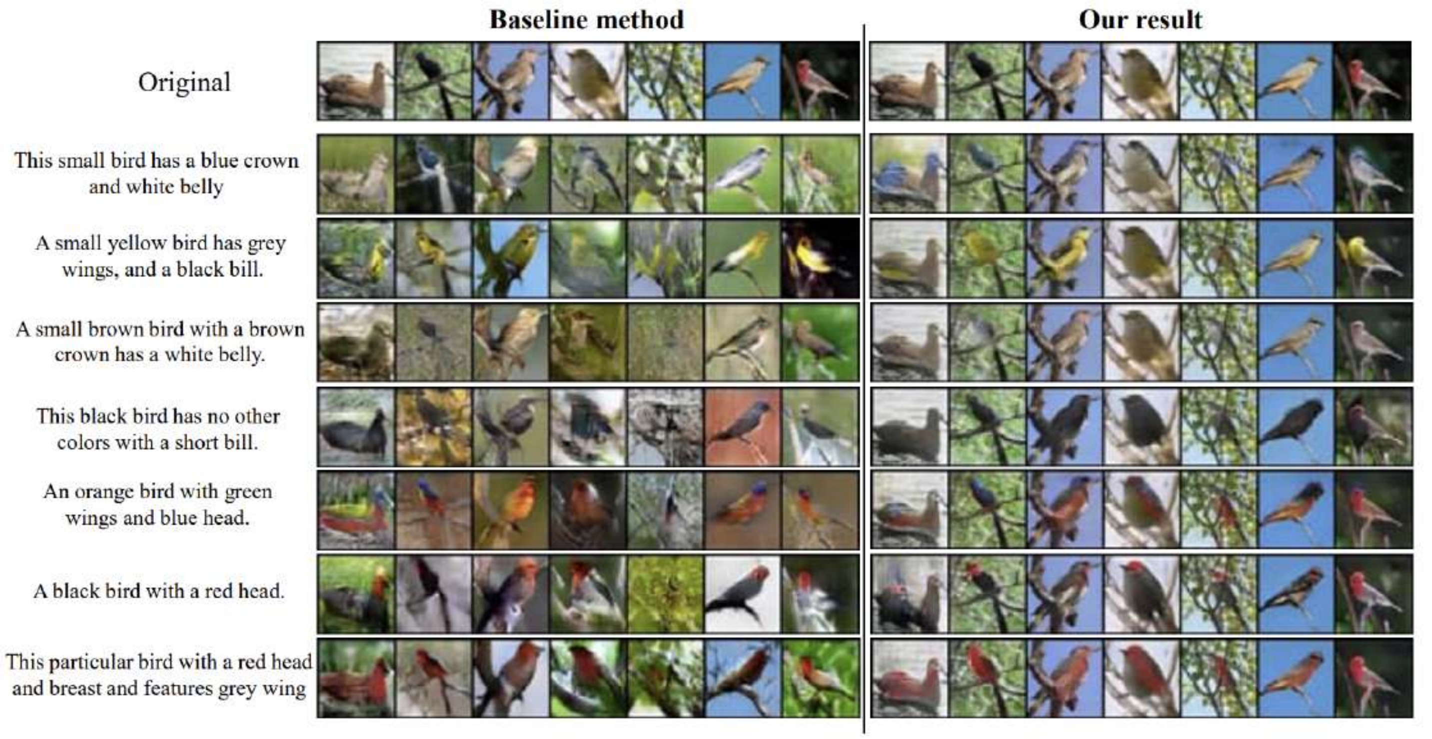
# GAN: STYLE TRANSFER



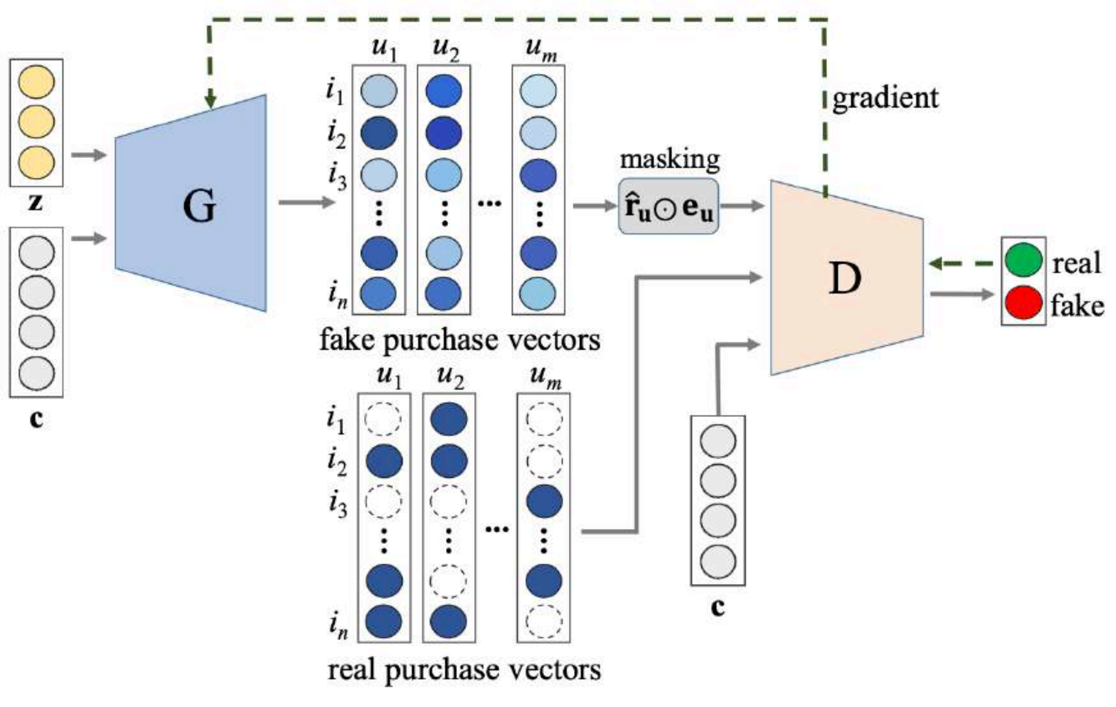
# GAN: SUPER RESOLUTION



# GAN: CONDITIONAL SYNTHESIS



# CF-GAN: COLLABORATIVE FILTERING

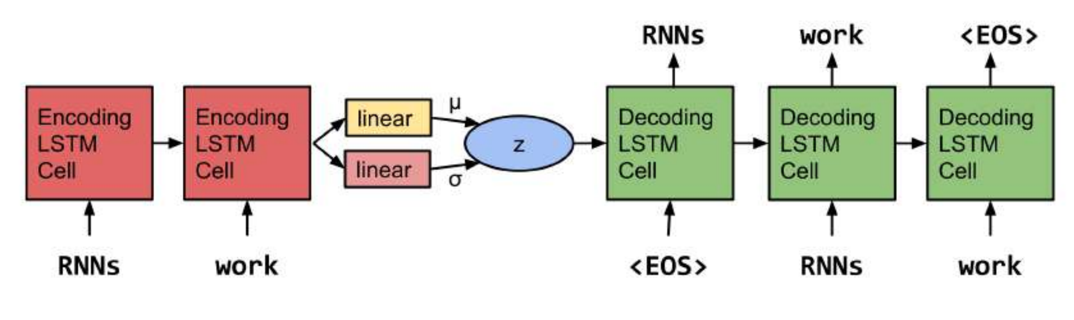


# VAE: SENTENCE INTERPOLATION

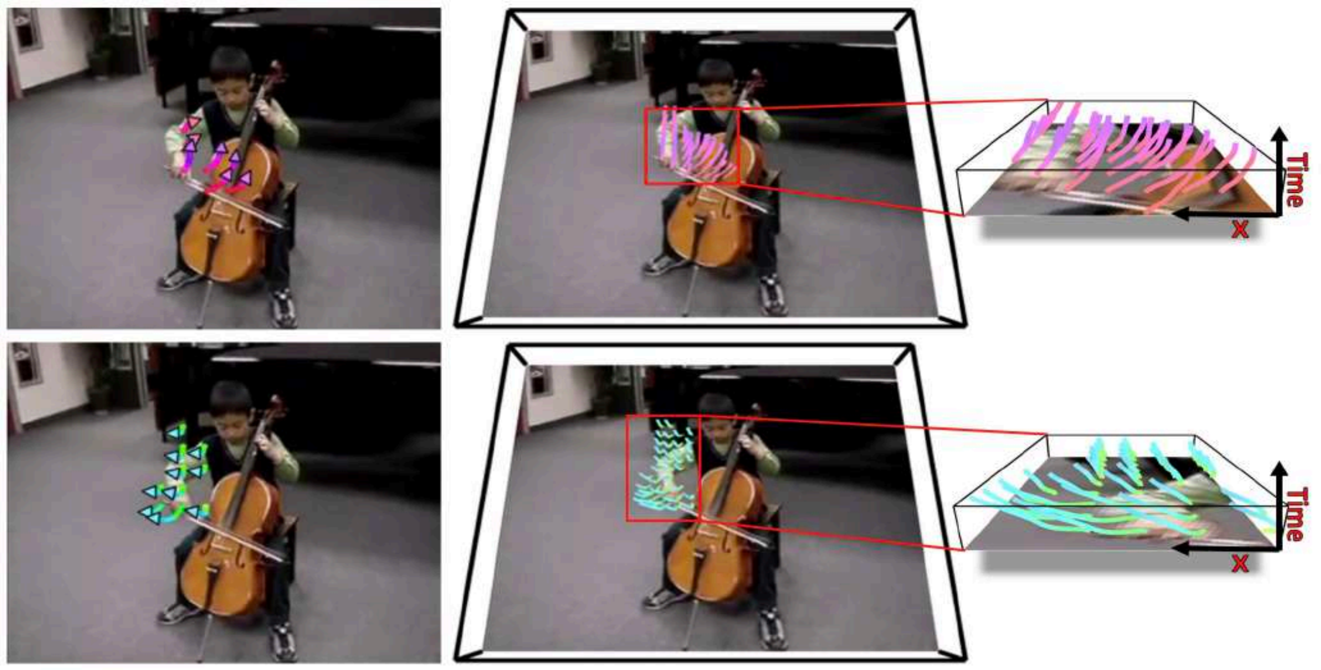
**" i want to talk to you . "**
*"i want to be with you . "*
*"i do n't want to be with you . "*
*i do n't want to be with you .*
**she did n't want to be with him .**

**he was silent for a long moment .**
*he was silent for a moment .*
*it was quiet for a moment .*
*it was dark and cold .*
*there was a pause .*
**it was my turn .**

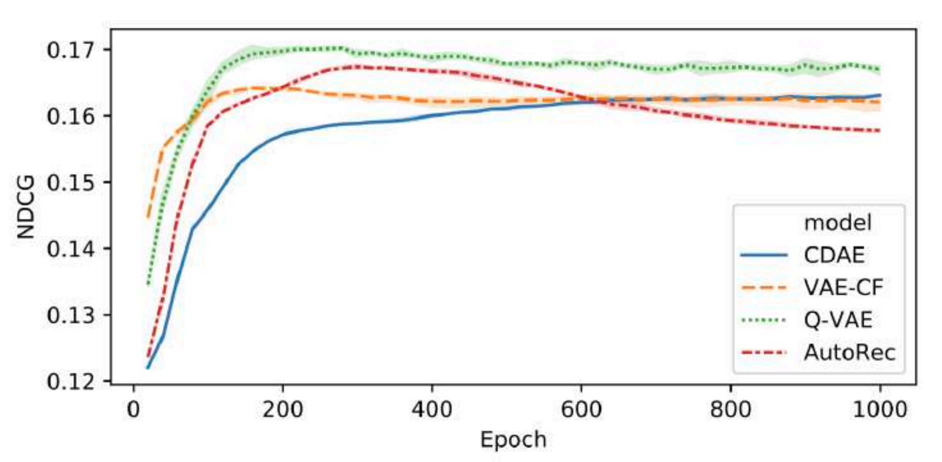
**there is no one else in the world .**
*there is no one else in sight .*
*they were the only ones who mattered .*
*they were the only ones left .*
*he had to be with me .*
*she had to be with him .*
*i had to do this .*
*i wanted to kill him .*
*i started to cry .*
**i turned to him .**



# VAE: MOTION FORECASTING IN STATIC IMAGES



# VAE: COLLABORATIVE FILTERING



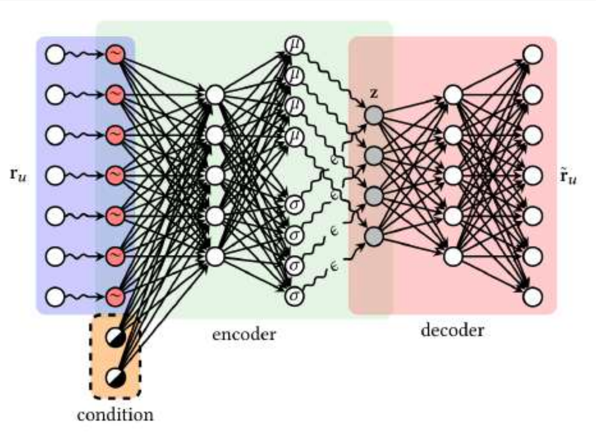

**(a) ML-20M**

| | Recall@20 | Recall@50 | NDCG@100 |
|---|---|---|---|
| Mult-VAE$^{PR}$ | **0.395** | **0.537** | **0.426** |
| Mult-DAE | 0.387 | 0.524 | 0.419 |
| WMF | 0.360 | 0.498 | 0.386 |
| SLIM | 0.370 | 0.495 | 0.401 |
| CDAE | 0.391 | 0.523 | 0.418 |

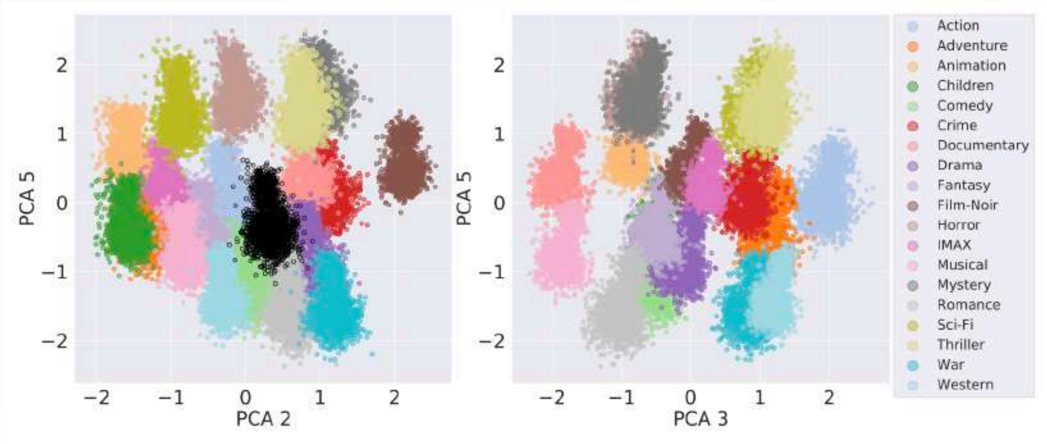
**(b) Netflix**

| | Recall@20 | Recall@50 | NDCG@100 |
|---|---|---|---|
| Mult-VAE$^{PR}$ | **0.351** | **0.444** | **0.386** |
| Mult-DAE | 0.344 | 0.438 | 0.380 |
| WMF | 0.316 | 0.404 | 0.351 |
| SLIM | 0.347 | 0.428 | 0.379 |
| CDAE | 0.343 | 0.428 | 0.376 |

# CONDITIONED VAE




C-VAE Architecture




Latent space visualization through PCA

# References


- PAPER GAN: https://arxiv.org/pdf/1406.2661.pdf
- PAPER VAE: https://arxiv.org/pdf/1312.6114.pdf
- PAPER VAE/GAN: https://arxiv.org/pdf/1512.09300.pdf
- PAPER C-VAE: https://dl.acm.org/doi/abs/10.1145/3386392.3399305

- Tutorial on GAN:
  https://towardsdatascience.com/comprehensive-introduction-to-turing-learning-and-gans-part-1-81f6d02e644d
  https://towardsdatascience.com/comprehensive-introduction-to-turing-learning-and-gans-part-2-fd8e4a70775
  https://towardsdatascience.com/gans-vs-autoencoders-comparison-of-deep-generative-models-985cf15936ea

- Tutorial on VAE: https://towardsdatascience.com/generating-images-with-autoencoders-77fd3a8dd368

- Interesting series of articles about VAE:
  https://towardsdatascience.com/the-variational-autoencoder-as-a-two-player-game-part-i-4c3737f0987b
  https://towardsdatascience.com/the-variational-autoencoder-as-a-two-player-game-part-ii-b80d48512f46
  https://towardsdatascience.com/the-variational-autoencoder-as-a-two-player-game-part-iii-d8d56c301600

- Why it is so hard to train GANs? https://medium.com/@jonathan_hui/gan-why-it-is-so-hard-to-train-generative-advisory-networks-819a86b3750b

# THANKS

Does anyone have any questions?

Mirko Polato, PhD
mpolato@math.unipd.it