Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

RESTART

SHAPE THE NOISE: A TUTORIAL ON GENERATIVE AI AND ITS TRANSFORMATIVE ROLE IN ICT

Danilo Comminiello
Sapienza University of Rome

SAPIENZA UNIVERSITÀ DI ROMA

Intelligent Signal Processing and MultiMedia lab

GTTI 2023 – Riunione Annuale del Gruppo Telecomunicazioni e Tecnologie dell'Informazione
Rome, Italy – September 13, 2023

Playing with Generative AI

Midjourney

DALL·E 2

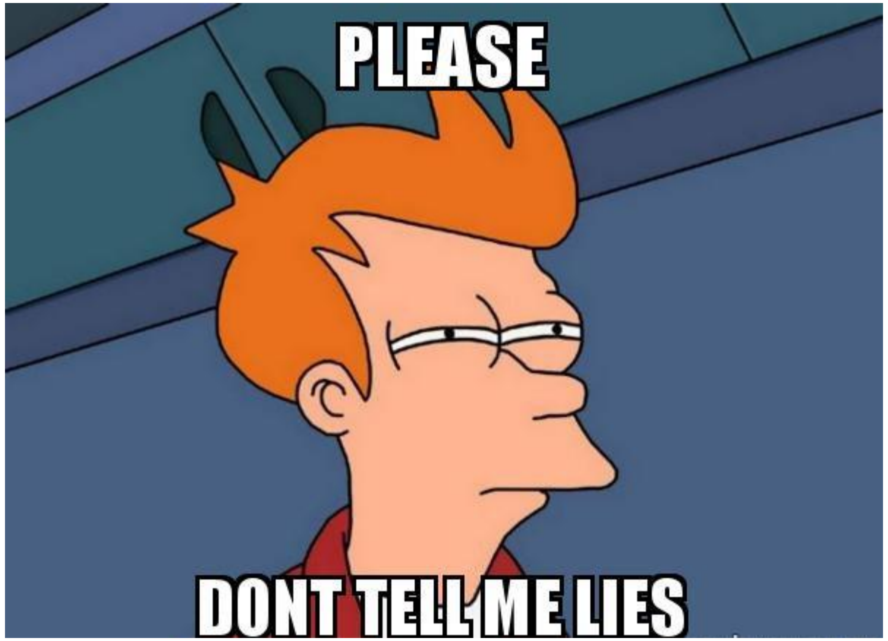
Bing

Stable Diffusion

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
BESTART

Playing with Generative AI

ChatGPT



Shape the noise

Creating noise from data is easy; creating data from noise is generative modeling.

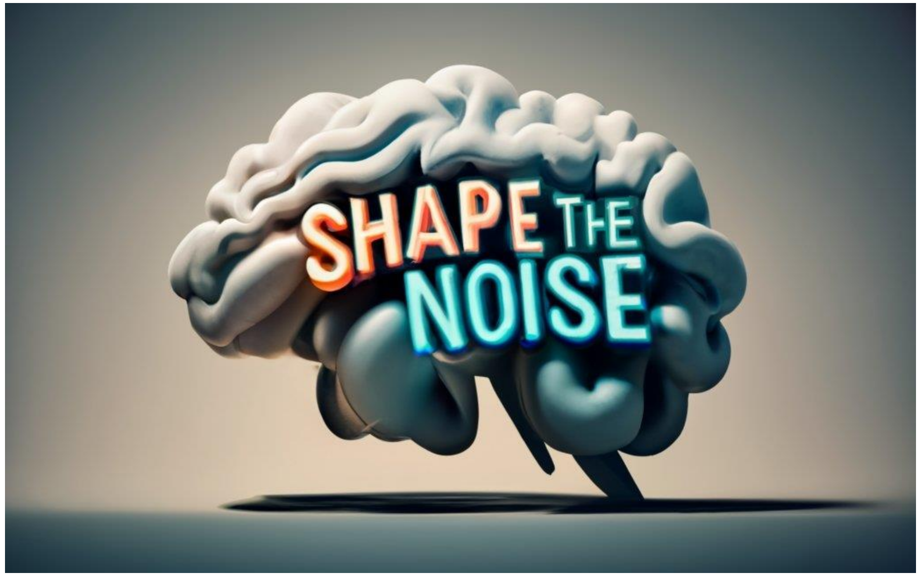
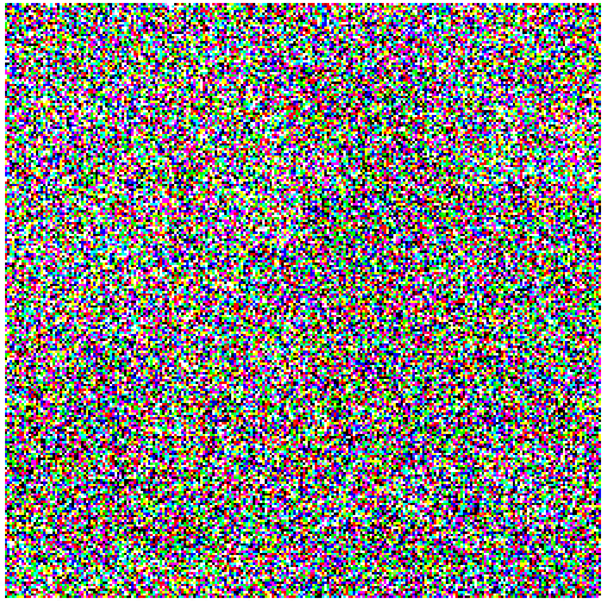


Image generated by ideogram.



Source: Yang Song blog.

Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, B. Poole, Score-Based Generative Modeling through Stochastic Differential Equations, *Int. Conf. On Learning Representations* (ICLR), 2021.

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
RESTART

Outline

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
RESTART

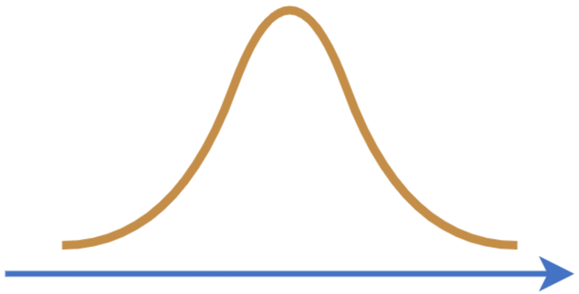
1 | Introduction to Generative Modeling

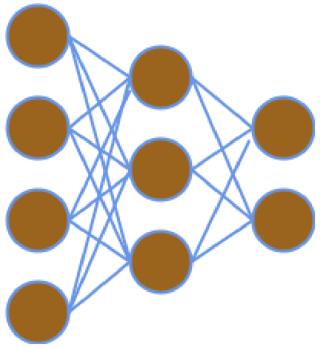
Where Generative AI comes from

Generative AI

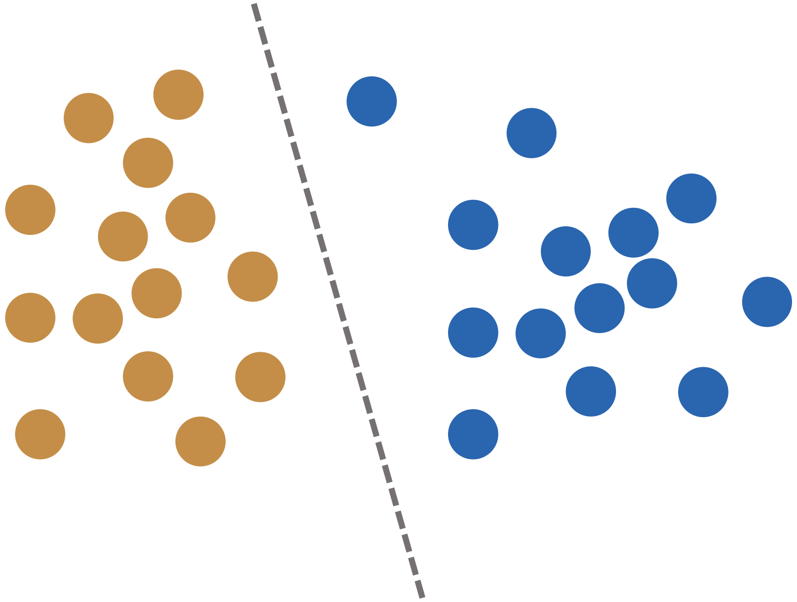
Generative Modeling

Deep Learning

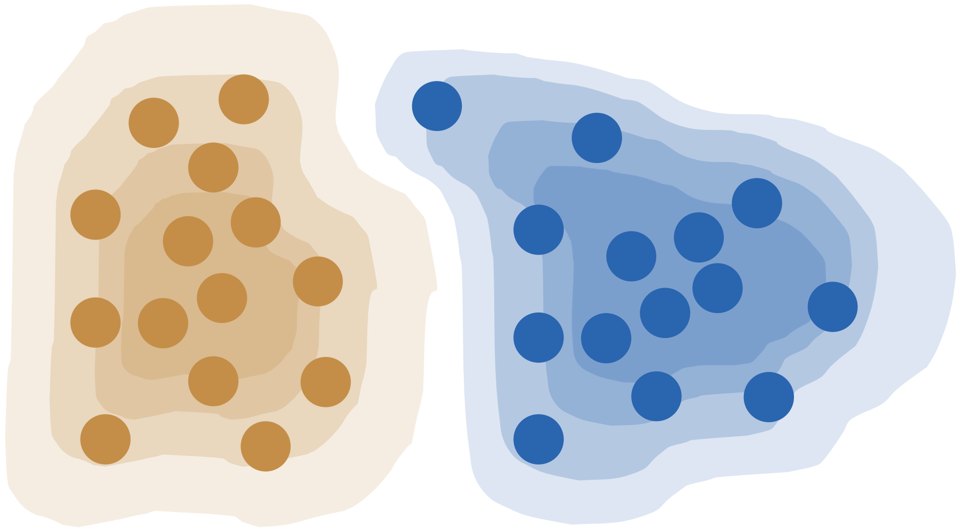




Discriminative vs Generative



Discriminative Modeling

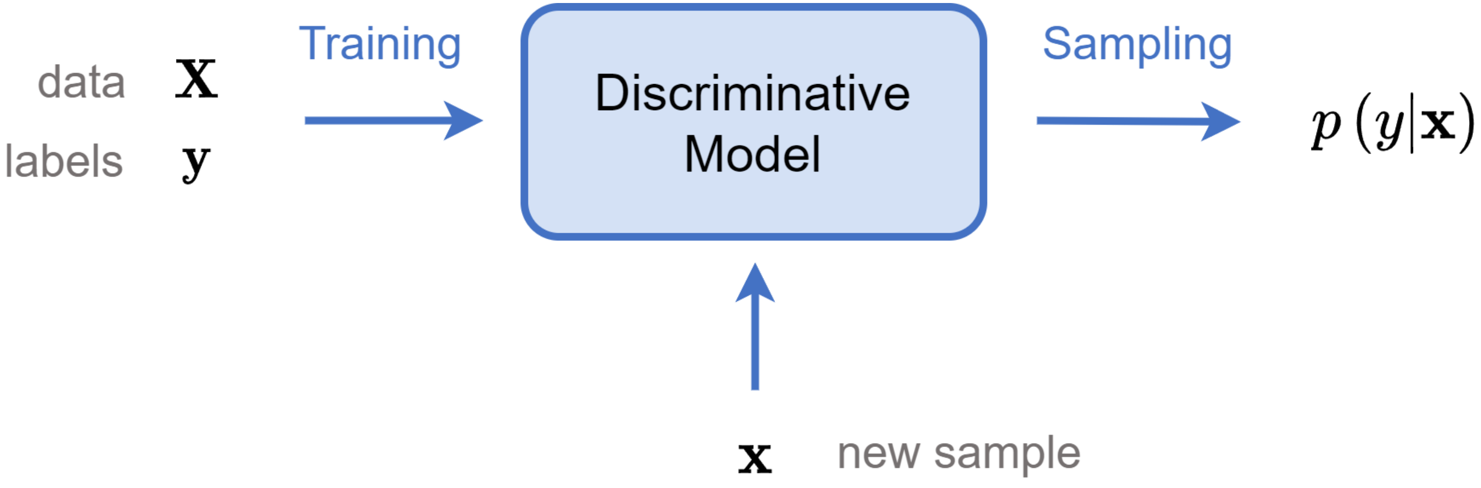


Generative Modeling

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RESTART

Discriminative modeling

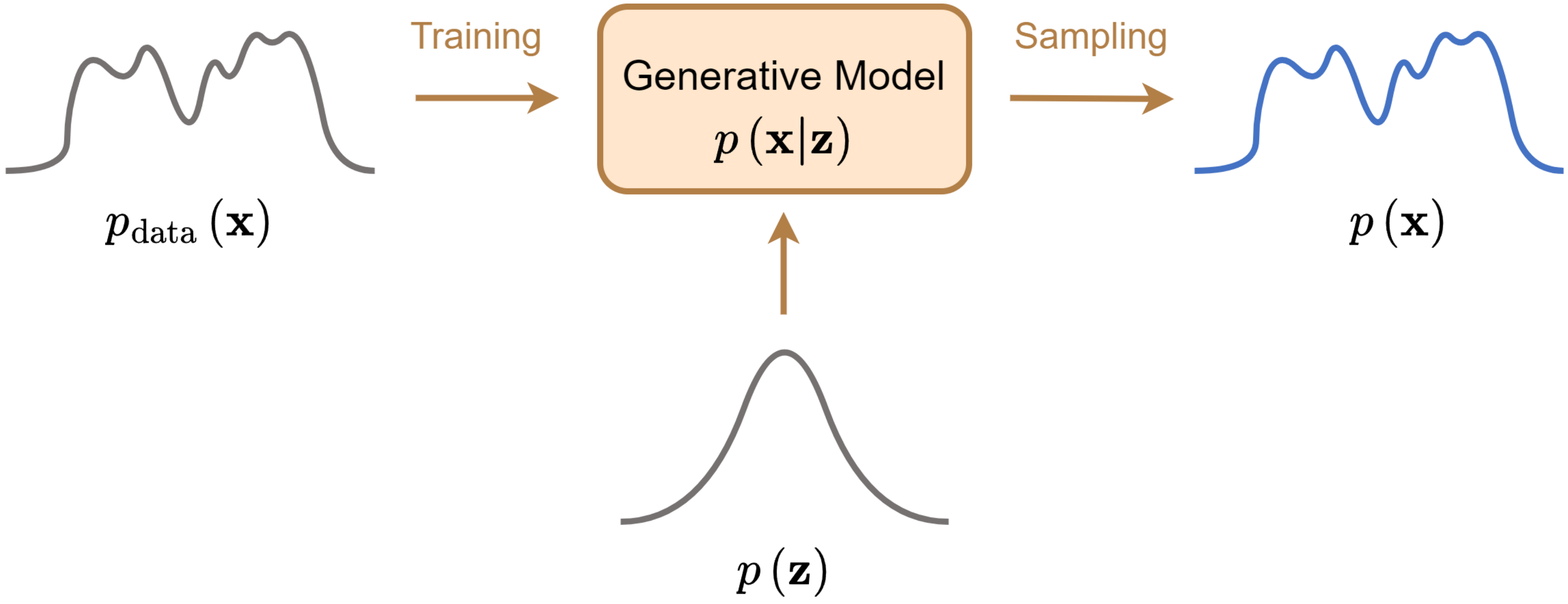
Discriminative modeling usually refers to supervised learning, or learning a function that maps an input to an output using a labeled dataset.



D. Foster, Generative Deep Learning – Teaching Machines to Paint, Write, Compose and Play, 2nd ed. O'Reilly Media, Inc., May 2023.

Generative modeling

A generative model describes how a dataset is generated, in terms of a probabilistic model. By sampling from this model, we can generate new data.



Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA | RESTART

The generative modeling framework

Let us consider a complex and large-scale dataset $\mathcal{X}$ of observations $\mathbf{x}$.

We assume $\mathbf{x}$ drawn from some unknown probability distribution p_{data}.

The goal of a generative model is to provide an estimate of the data distribution $p(\mathbf{x})$ from which drawing new observations that appear to have been drawn from p_{data}.

The fundamental laws of a generative model

We can say that a generative model is impressive if it satisfies the following fundamental rules:

1. It must generate new samples that appear to have been drawn from the original distribution of the data p_{data}.

2. It must generate new samples that did not exist before in the original dataset $\mathcal{X}$.

Example of a generative modeling

Let us consider a dataset of observations $\mathbf{x}$ whose elements are cities (blue points) of a map, which represents the original data distribution p_{data}.

We want to generate new points.

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
BESTART

Example of a generative modeling

The new generated points (orange) do not represent a reliable generative model, as they fall outside the data distribution, thus not satisfying the first fundamental rule of generative modeling.

Example of a generative modeling

The new generated points (yellow) do not represent a reliable generative model, as they significantly overlap with existing points, thus not satisfying the second fundamental rule of generative modeling.

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
BESTART

Example of a generative modeling

The new generated points (green) denote a good generative model, as they appear to have been drawn from the original data distribution and did not exist before, thus satisfying the fundamental rules of generative modeling.

This slide does not exist

Why generative models are attractive

- Generative processes are able to express physical laws, while considering meaningless details as *noise*.
- Generative models are usually highly intuitive and interpretable.
- Generative processes express causal relations, thus being able to generalize much better to *new situations* than mere correlations.

2 The Role of Deep Learning in Generative Frameworks

Dealing with real-world, complex, and large-scale data

Modern ICT applications cannot ignore the complexity of real data, which may exhibit:

- High-dimensional, complicated probability distributions;
- Damaged or missing information;
- Partially labeled or completely unlabeled data;
- High resolutions and QoS constraints;
- Multimodality.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RESTART

Generative modeling challenges

Generative modeling theory has been widely applied to classic machine learning algorithms (e.g., naïve Bayes classifiers).

However, 1) in presence of a big amount of data, it is not easy for a model cope with the high degree of conditional dependence between features.

Moreover, 2) the larger the input data space, the more difficult it is to produce an output that satisfies the generation constraints.

Deep learning is a solution

Deep learning is the key to solving both of these challenges due to its ability to form its own features in a lower-dimensional space.

The real power of deep learning, especially with regard to generative modeling, comes from its ability to work with complex and unstructured data.

Deep learning methods rely on multiple stacked layers of processing units to learn high-level representations from data.

Learning high-level representation

The way human brain learns is inherently hierarchical, thus being able to provide a *deeper* representation.

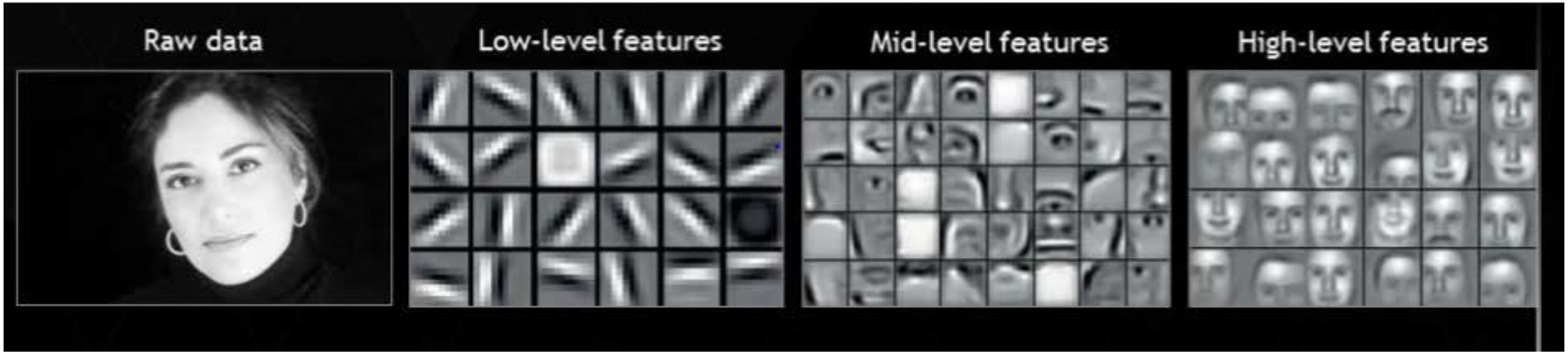
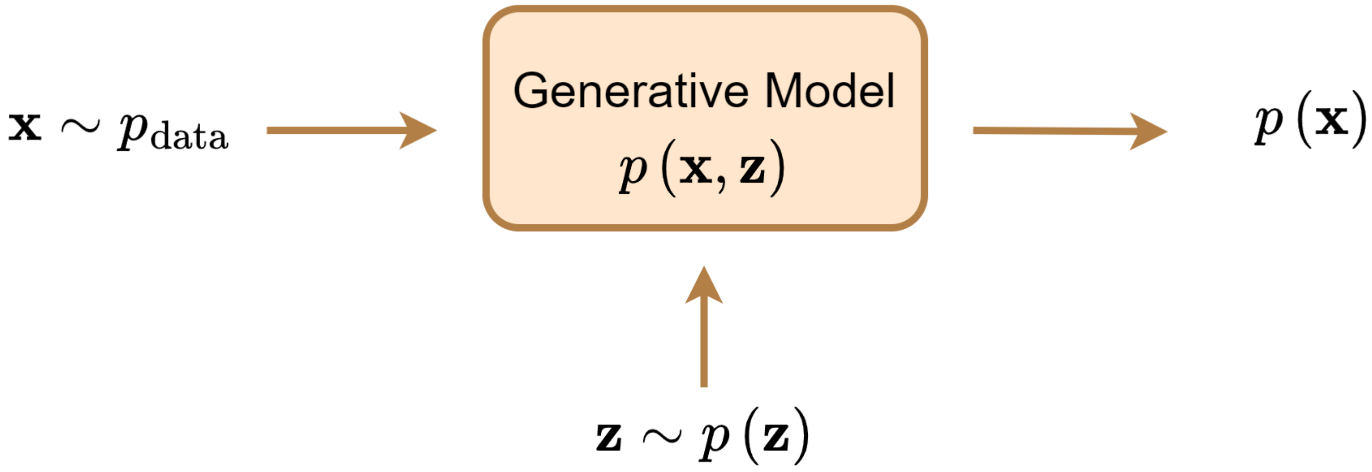


Image source: Analytics Vidhya.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RESTART

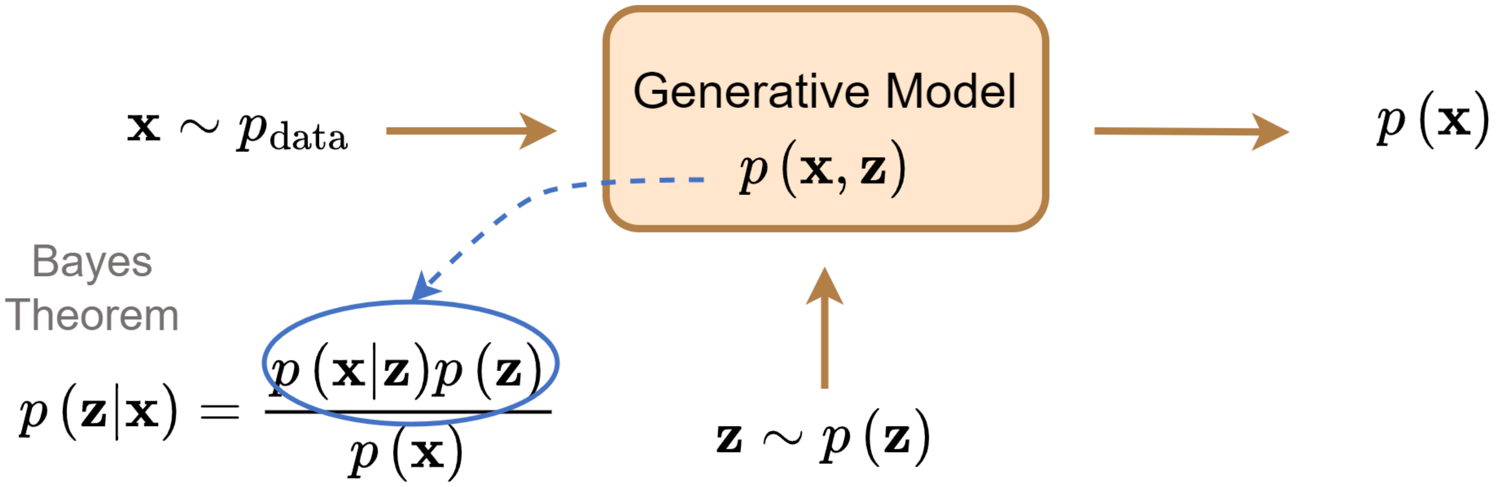
Neural networks parameterize generative models

Neural networks are flexible and powerful and, therefore, they are widely used to parameterize generative models.



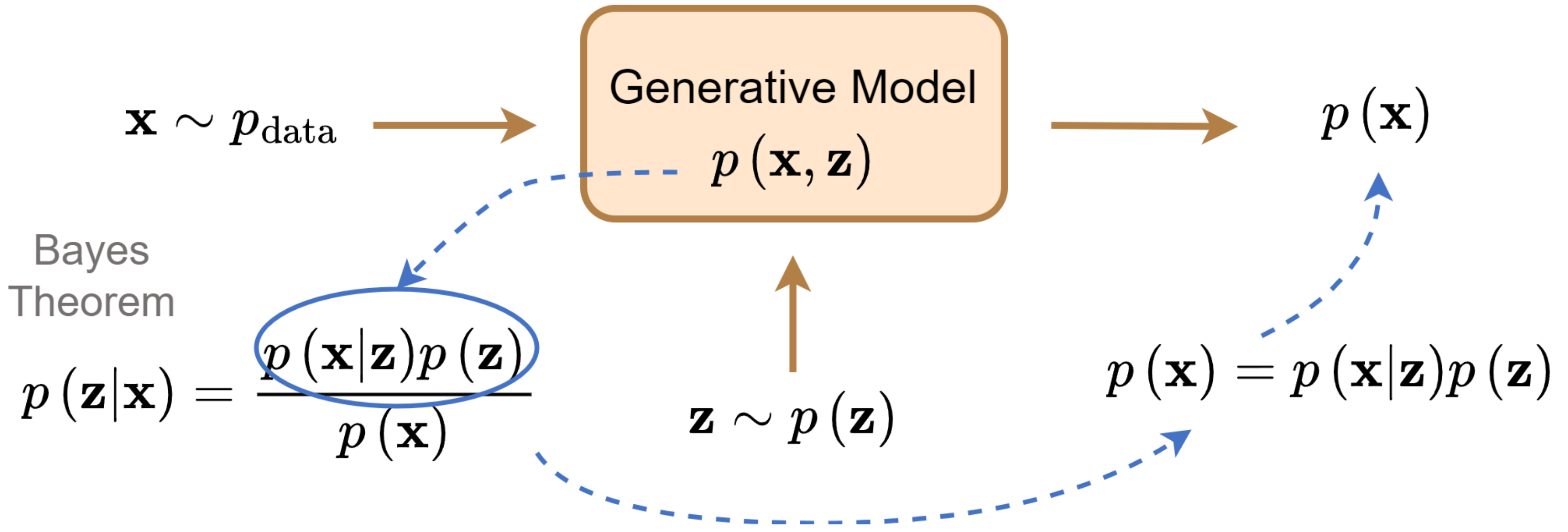
Neural networks parameterize generative models

Neural networks are flexible and powerful and, therefore, they are widely used to parameterize generative models.



Neural networks parameterize generative models

Neural networks are flexible and powerful and, therefore, they are widely used to parameterize generative models.



Neural networks parameterize generative models

Neural networks are flexible and powerful and, therefore, they are widely used to parameterize generative models.

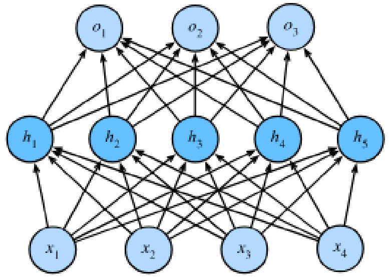
$$\mathbf{x} \sim p_{\text{data}} \longrightarrow \boxed{\begin{array}{c}\text{Generative Model}\\ p\left(\mathbf{x}, \mathbf{z}\right)\end{array}} \longrightarrow p\left(\mathbf{x}\right)$$

$$\mathbf{z} \sim p\left(\mathbf{z}\right) \uparrow$$

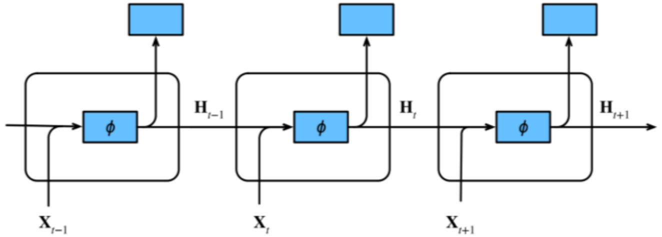
Neural network parameterization

$$p\left(\mathbf{x}\right) = p\left(\mathbf{x}|\mathbf{z}\right) p\left(\mathbf{z}\right)$$

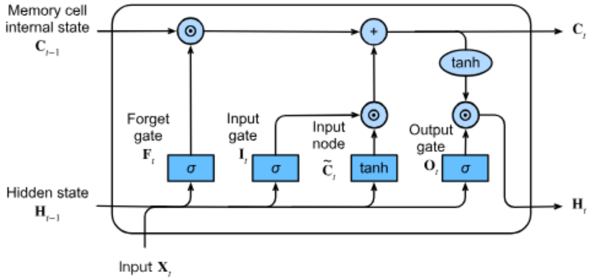
Generative models can benefit from different neural architectures



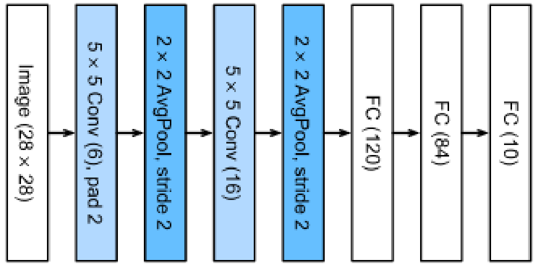
Multi-layer perceptron (MLP)



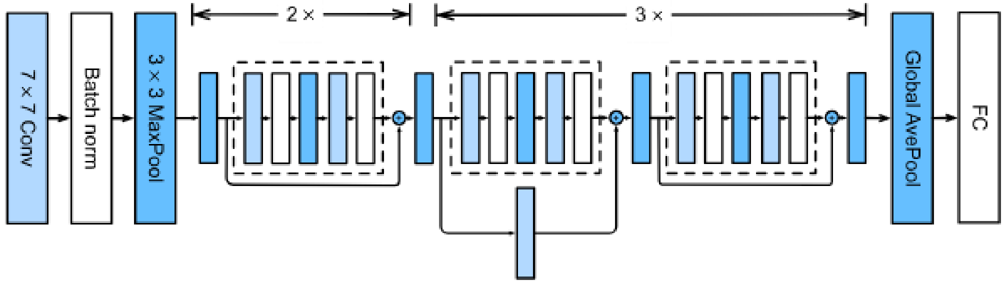
Recurrent neural network (RNN)



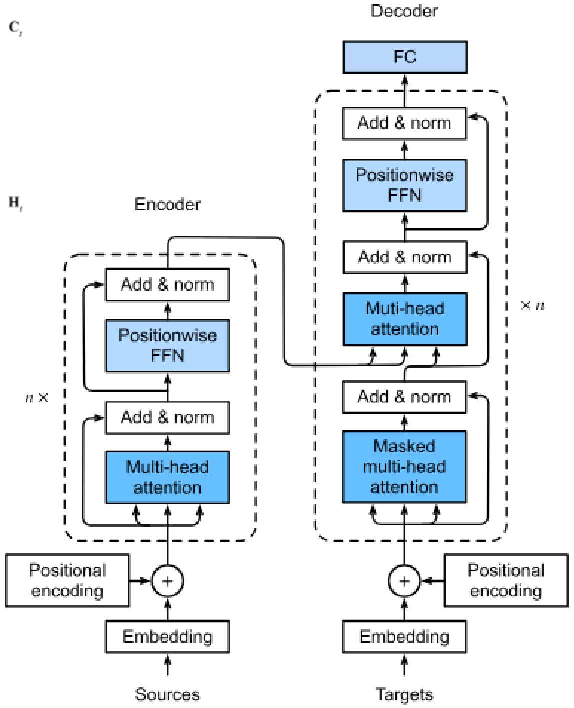
Long short-term memory (LSTM) network



Convolutional neural network (CNN)



ResNet-18 architecure



Vision Transformer

A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, Dive Into Deep Learning, 2020.

Steps for generating new data

The process to generate new data by using deep learning usually follows the next steps:

1. Sample from a known distribution
2. Optimize a likelihood-based loss (or any implicit variant).
3. Train the model.
4. Decode a representation to draw a new sample.

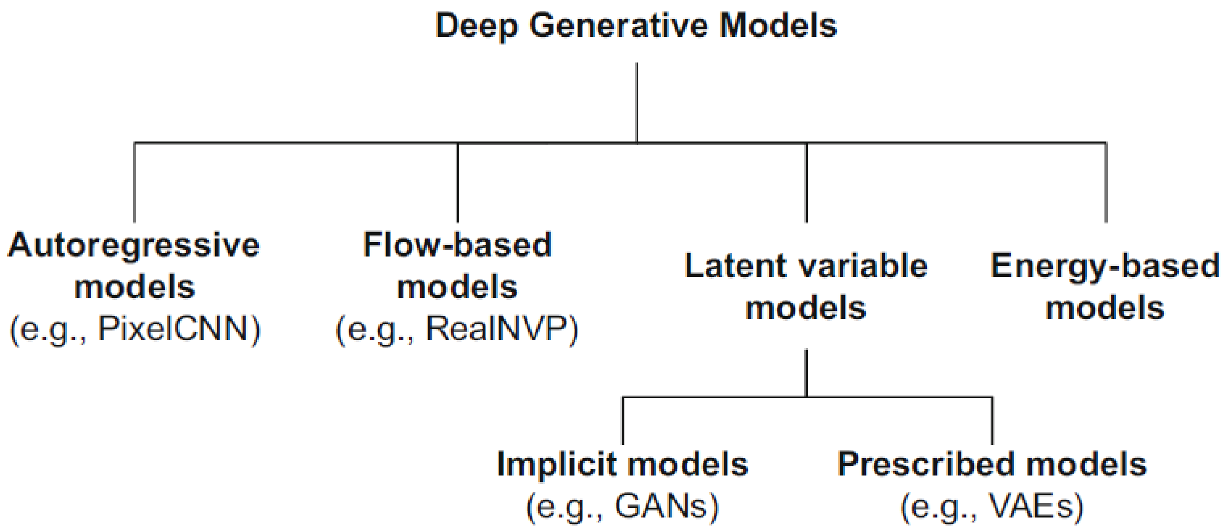
The new sample must belong to the underlying distribution of the data points and it must never have been seen before.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RESTART

3 A TAXONOMY OF DEEP GENERATIVE MODELS

Possible taxonomy of deep generative models

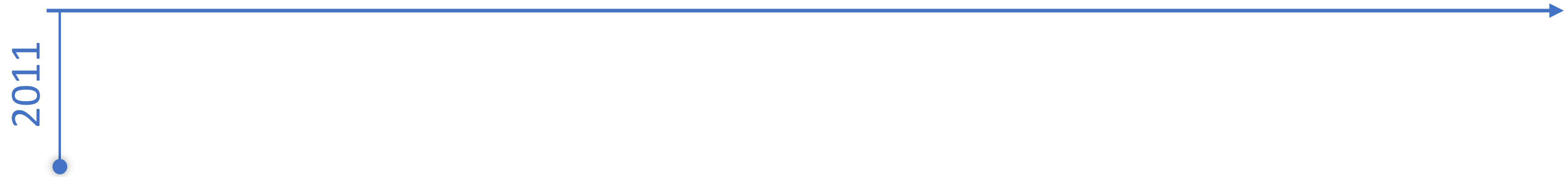
Deep generative models differentiate in how they express $p(\mathbf{x})$.



Characteristics of deep generative models

Generative models	Training	Likelihood	Sampling	Compression	Representation
Autoregressive models	Stable	Exact	Slow	Lossless	No
Flow-based models	Stable	Exact	Fast/slow	Lossless	Yes
Implicit models	Unstable	No	Fast	No	No
Prescribed models	Stable	Approximate	Fast	Lossy	Yes
Energy-based models	Stable	Unnormalized	Slow	Rather not	Yes

Timeline

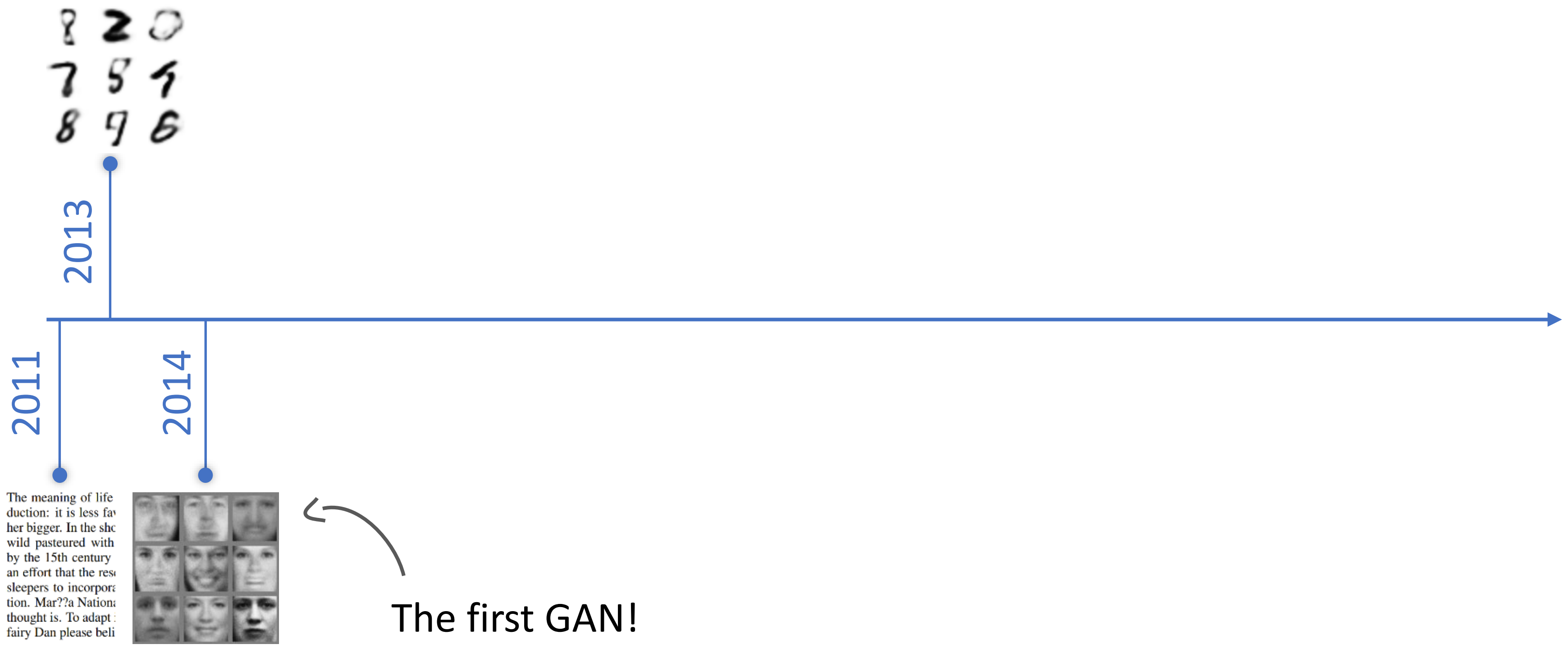


The meaning of life
duction: it is less fav
her bigger. In the sho
wild pasteured with
by the 15th century
an effort that the res
sleepers to incorpora
tion. Mar??a Nationa
thought is. To adapt
fairy Dan please beli

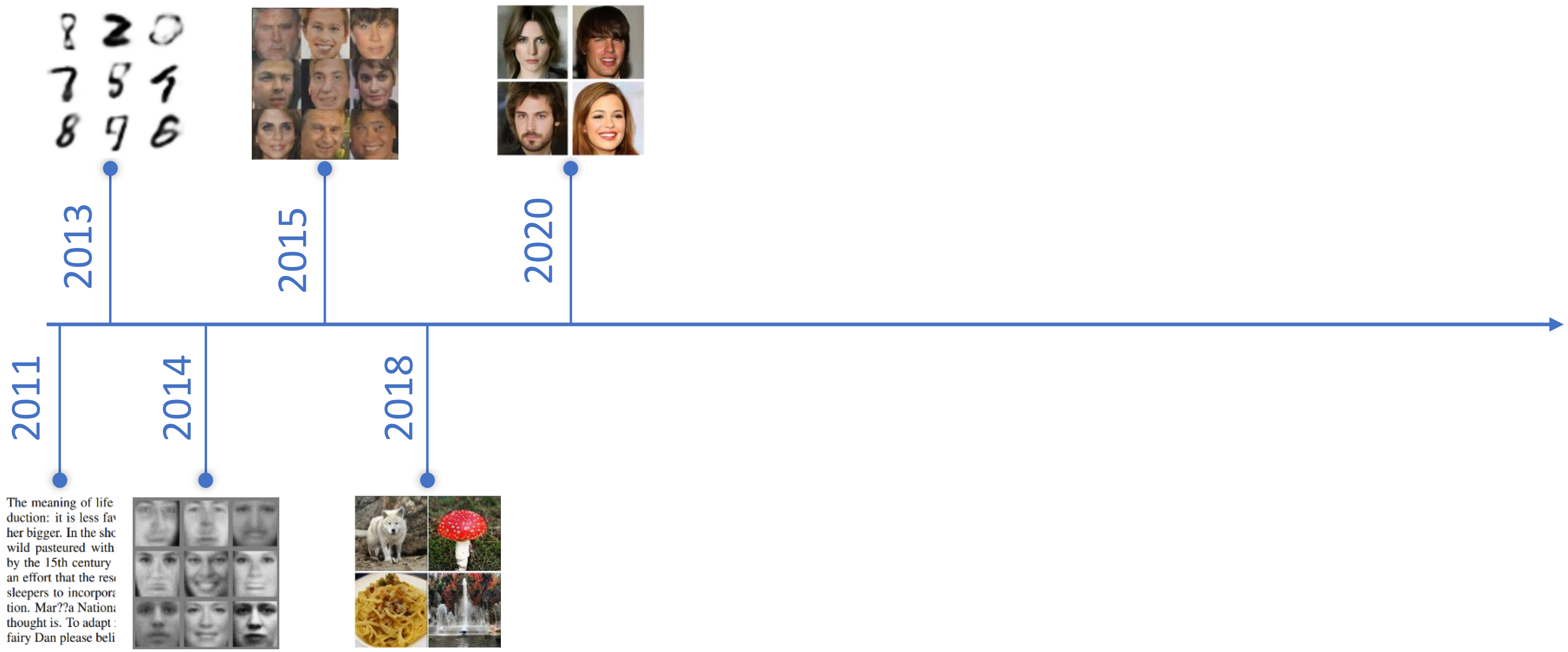
Timeline


The first VAE!
2013
2011
The meaning of life
duction: it is less fa
her bigger. In the sho
wild pasteured with
by the 15th century
an effort that the res
sleepers to incorpor
tion. Mar??a Nation
thought is. To adapt
fairy Dan please beli

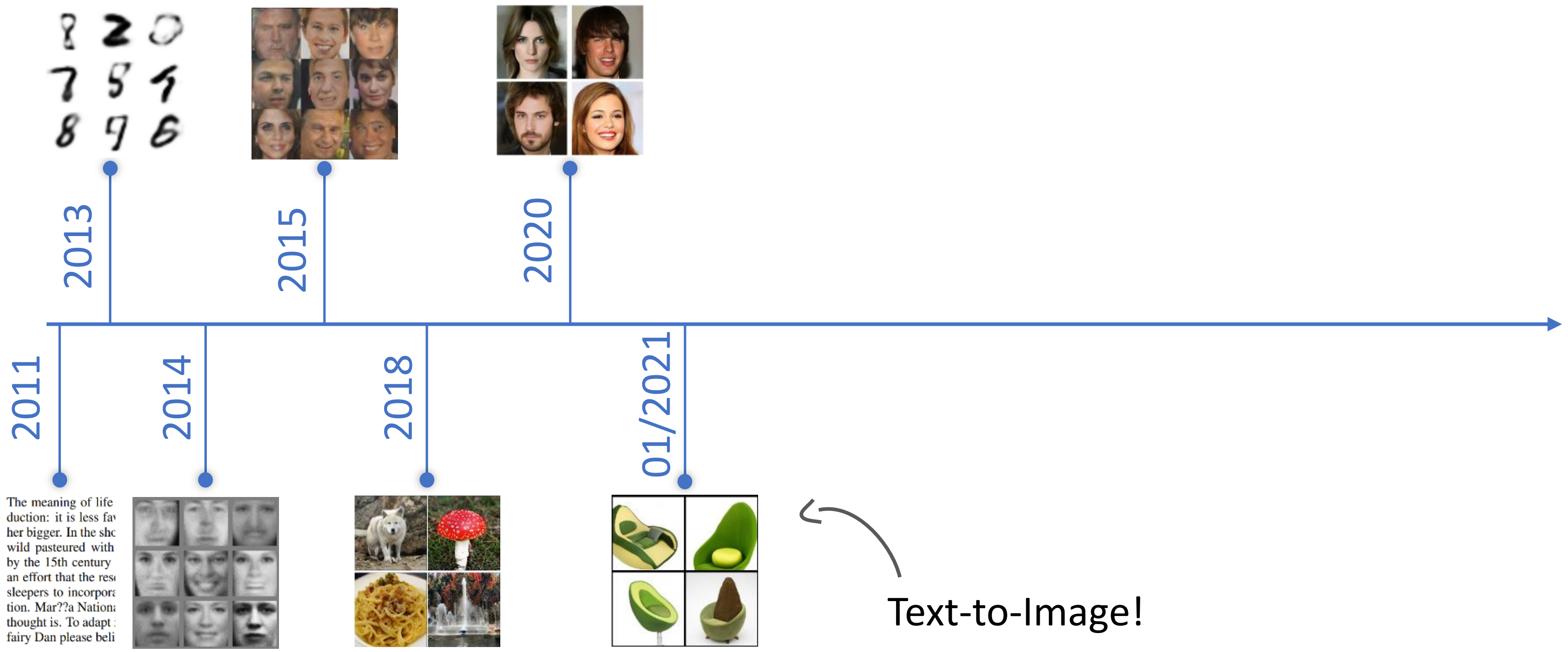
Timeline


2013
2011
2014
The meaning of life
duction: it is less fa
her bigger. In the sho
wild pasteured with
by the 15th century
an effort that the res
sleepers to incorpor
tion. Mar??a Nation
thought is. To adapt
fairy Dan please beli
The first GAN!

Timeline


2013
2015
2020
2011
2014
2018
The meaning of life
duction: it is less fa
her bigger. In the sho
wild pasteured with
by the 15th century
an effort that the res
sleepers to incorpor
tion. Mar??a Nation
thought is. To adapt
fairy Dan please beli

Timeline



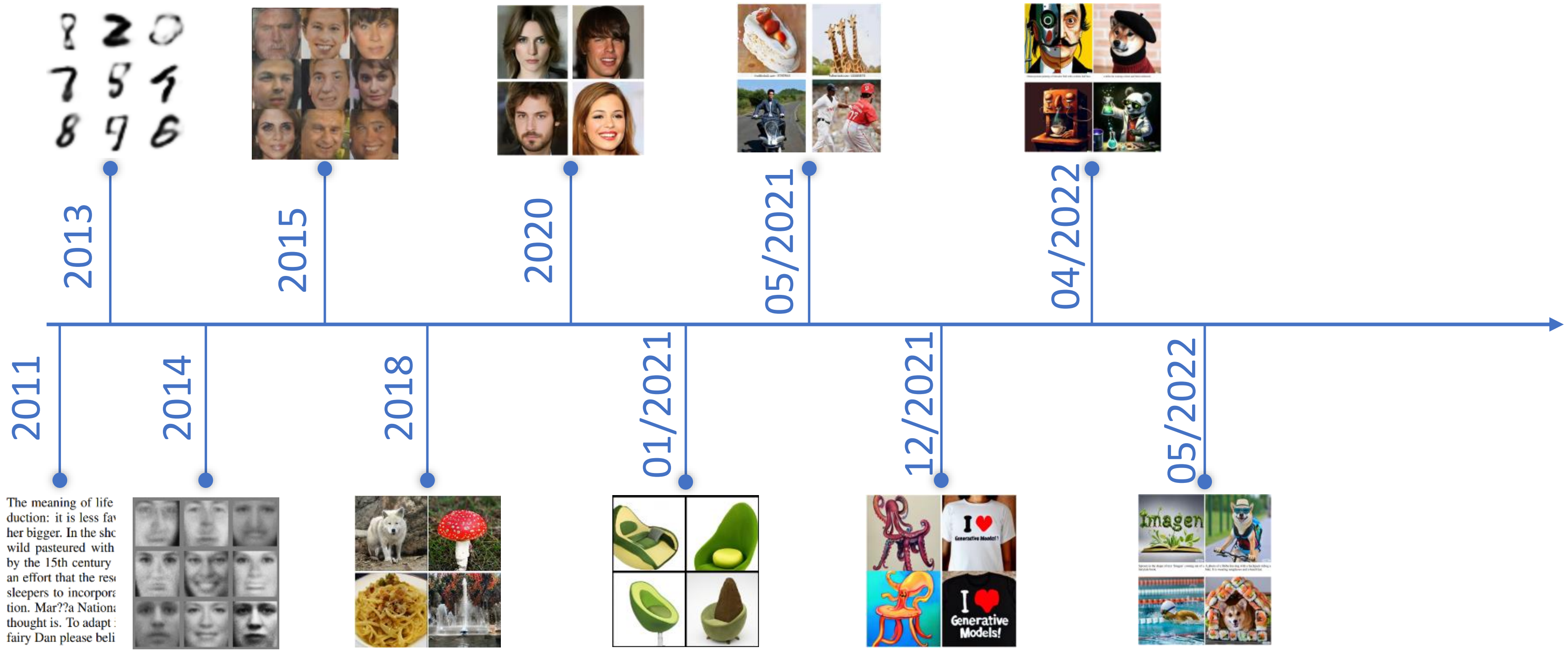
Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

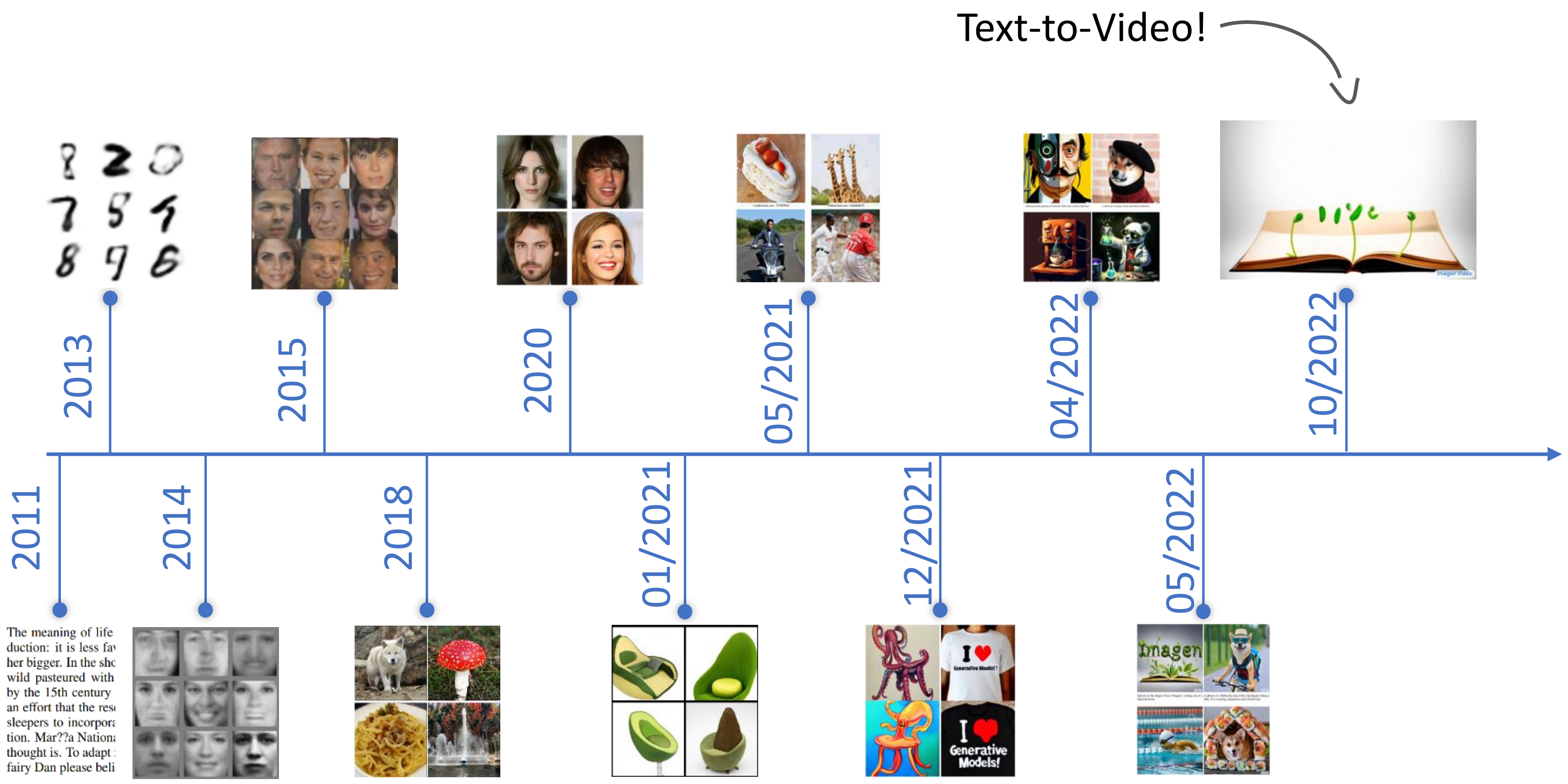
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

BESTART

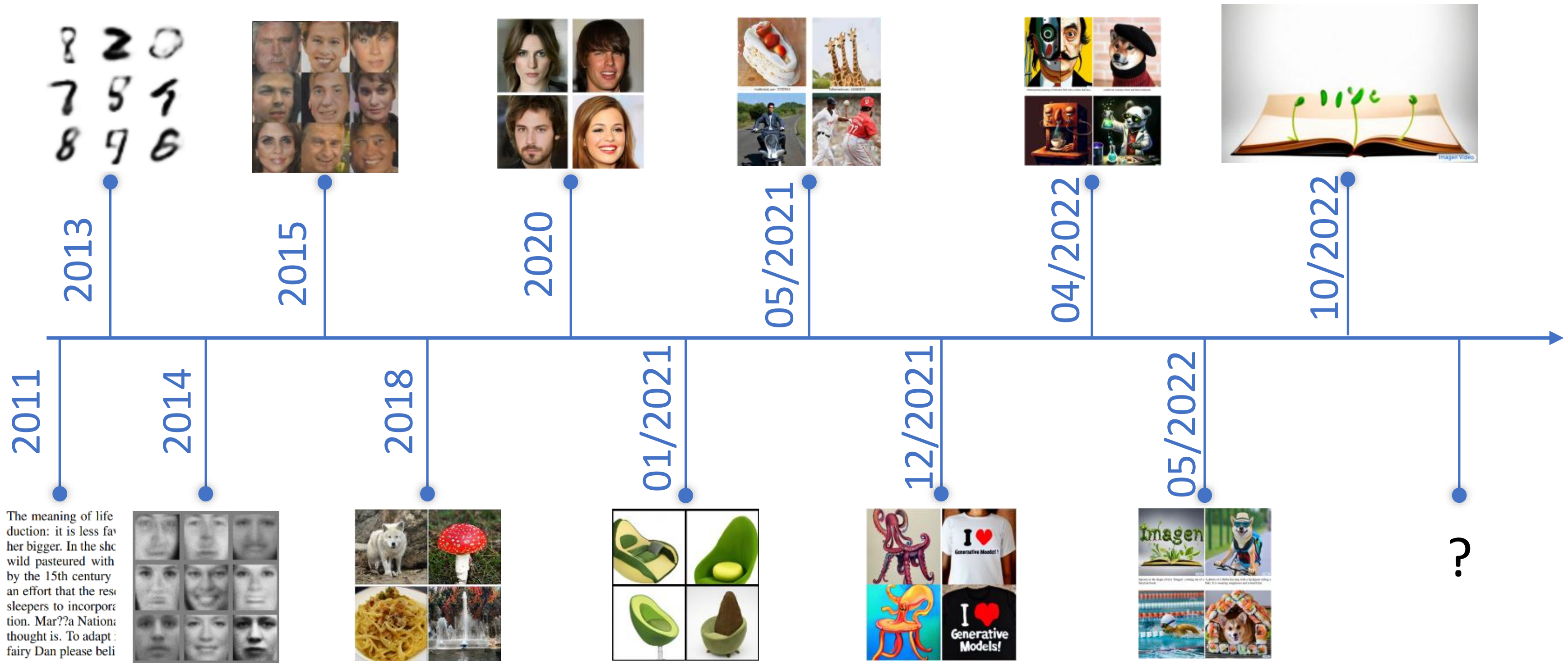
Timeline



Timeline


Text-to-Video!
2013
2015
2020
05/2021
04/2022
10/2022
2011
2014
2018
01/2021
12/2021
05/2022
The meaning of life
duction: it is less fa
her bigger. In the sh
wild pastured with
by the 15th century
an effort that the res
sleepers to incorpor
tion. Mar??a Nation
thought is. To adapt
fairy Dan please bel
I ♥ Generative Models!
Imagen

Timeline

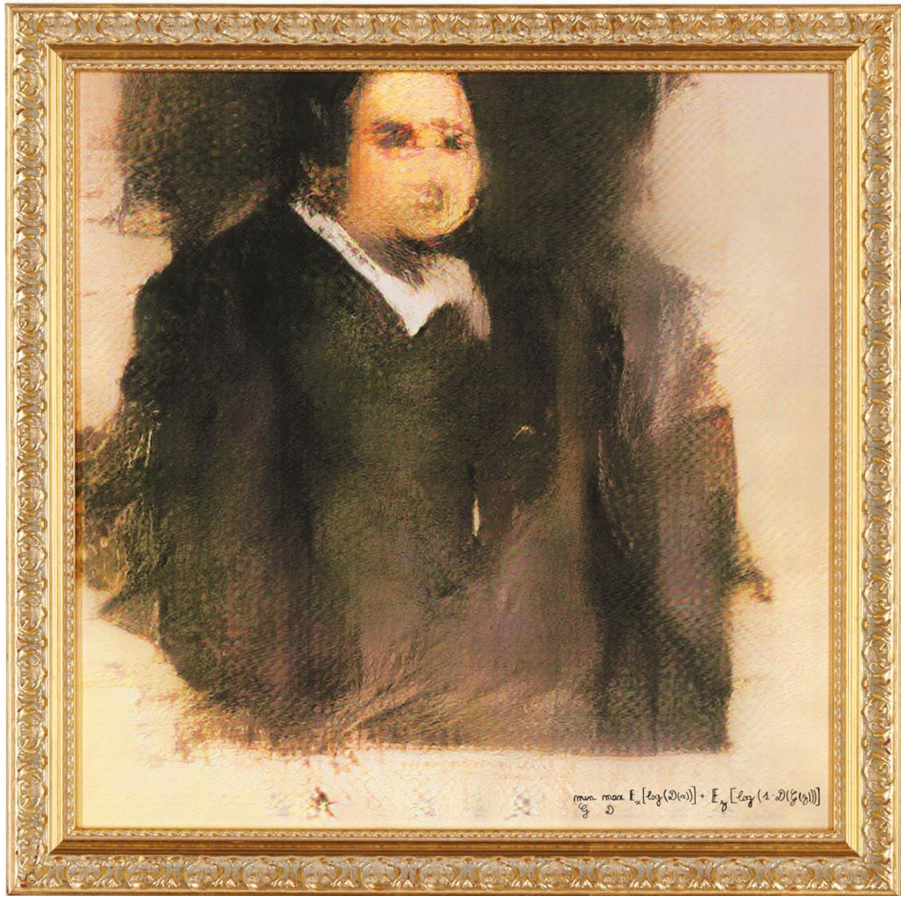

2013
2015
2020
05/2021
04/2022
10/2022
2011
2014
2018
01/2021
12/2021
05/2022
?
The meaning of life
duction: it is less fa
her bigger. In the sho
wild pasteured with
by the 15th century
an effort that the res
sleepers to incorpor
tion. Mar??a Nation
thought is. To adapt
fairy Dan please beli
I ♥ Generative Models!
Imagen

4 Generative Adversarial Networks

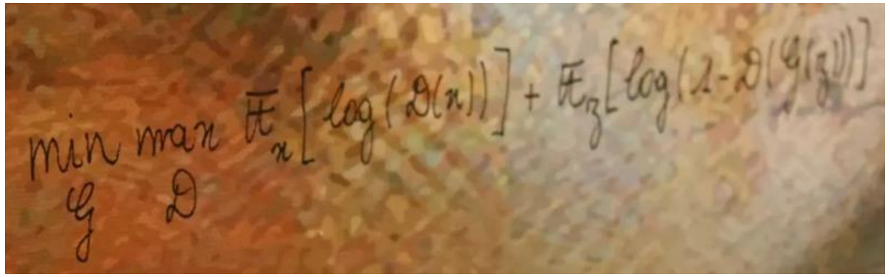
It was a dark and stormy night in Montréal..

One night in 2014, Ian Goodfellow, the "GANfather", came up with the generative adversarial network (GAN) while drinking beer in a pub with friends.

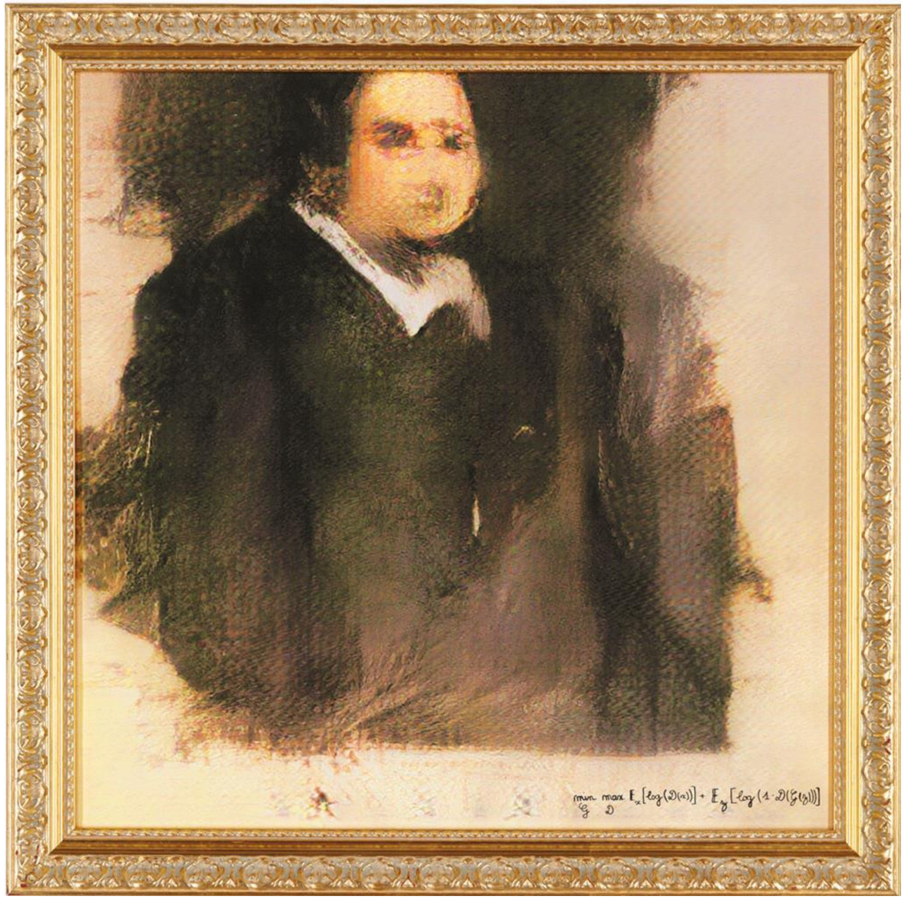
Generative art



Edmond de Belamy is a GAN portrait painting of 2018 by Paris-based arts-collective *Obvious*. It was selled as the first artwork created using articial intelligence in a Christie's auction for 432,500$.



Generative art



The New York Times

AI Art at Christie's Sells for $432,500

Give this article

"Edmond de Belamy, from La F...
Obvious, was sold on Thursday...

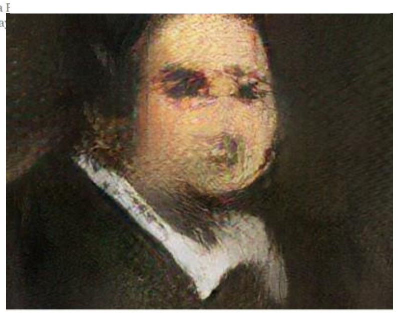
A portrait created by AI just sold for $432,000. But is it really art?

An image of Edmond de Belamy, created by a computer, has just been sold at Christie's. But no algorithm can capture our complex human consciousness

The Guardian

Portrait of Edmond Bellamy at Christie's in New York. Photograph: Timothy A Clary/AFP/Getty Images

From a distance, Portrait of Edmond de Belamy, which has just sold at Christie's in New York for $432,000 (£337,000), looks almost plausible. Up close, however, the paintwork becomes a grid of mechanical-looking dots, the man's face a golden blur with black



CHRISTIE'S

Is artificial intelligence set to become art's next medium?

AI artwork sells for $432,500 — nearly 45 times its high estimate — as Christie's becomes the first auction house to offer a work of art created by an algorithm

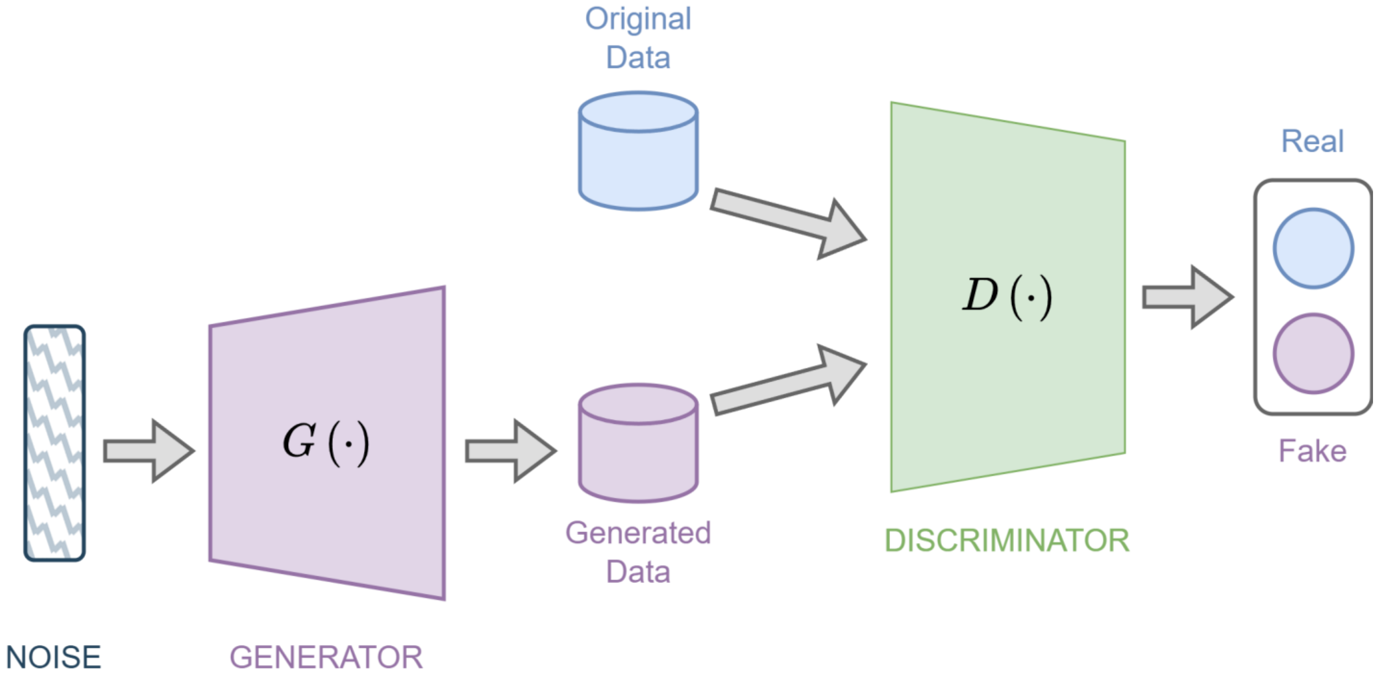
The Washington Post

Democracy Dies in Darkness

A 19-year-old developed the code for the AI portrait that sold for $432,000 at Christie's

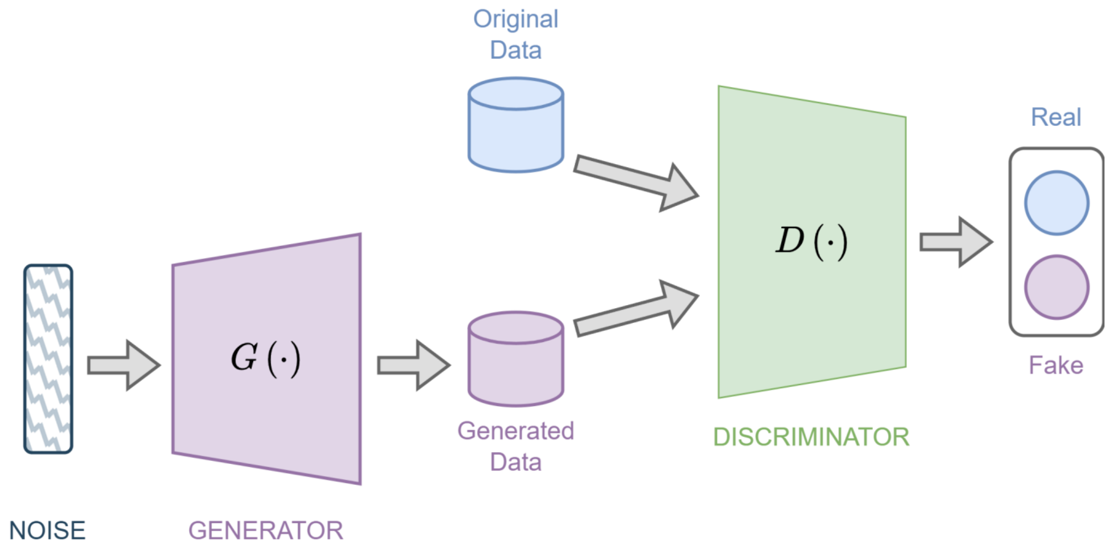
GAN framework

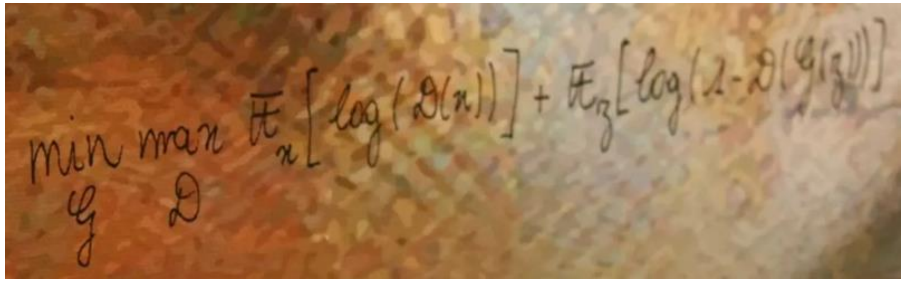
The generative adversarial network is composed of two models competing with each other, known as generator $G(\cdot)$ and discriminator $D(\cdot)$.



I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial networks, *arXiv preprint: arXiv:1406.2661v1*, 2014.

Adversarial training of a GAN

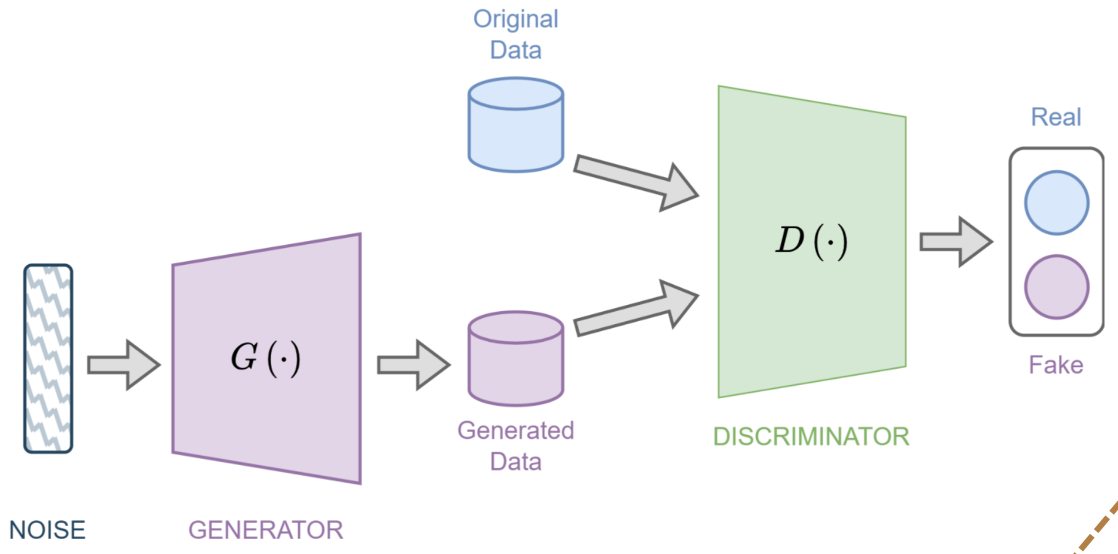




$$\min_G \max_D \mathrm{E}_{\mathbf{x}\sim p_{\text{data}}} \{\log D(\mathbf{x})\} + \mathrm{E}_{\mathbf{z}\sim p(\mathbf{z})} \{\log (1 - D(G(\mathbf{z})))\}$$

I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial networks, *arXiv preprint: arXiv:1406.2661v1*, 2014.

Goal of the generator



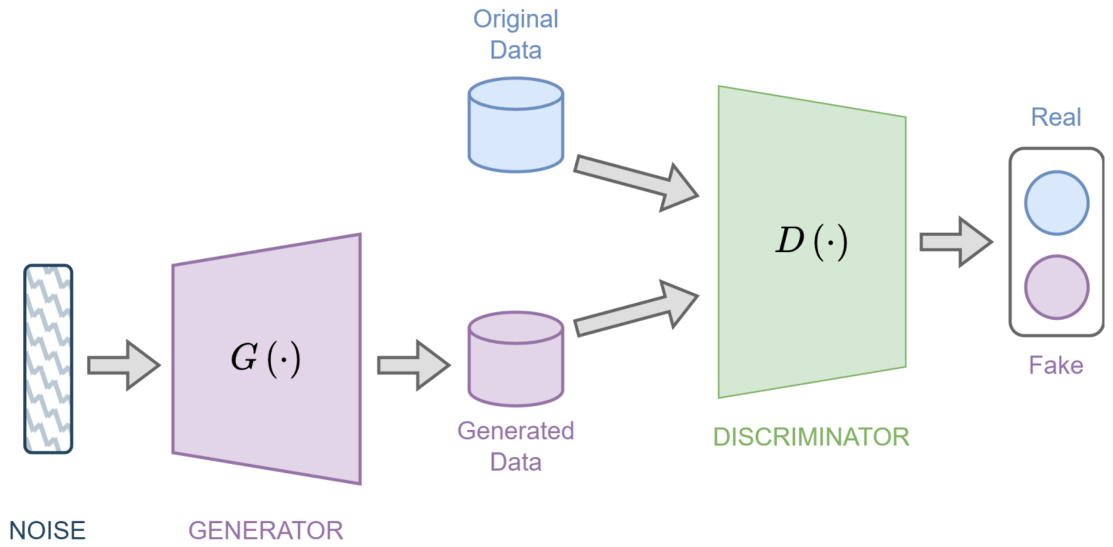
$D(\cdot)$ aims at maximizing the log-likelihood for the binary classification problem:

- original data: real (1)
- generated data: fake (0)

$$\min_G \max_D \mathrm{E}_{\mathbf{x}\sim p_{\text{data}}}\{\log D(\mathbf{x})\} + \mathrm{E}_{\mathbf{z}\sim p(\mathbf{z})}\{\log(1 - D(G(\mathbf{z})))\}$$

I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial networks, *arXiv preprint: arXiv:1406.2661v1*, 2014.

Goal of the discriminator



$D(\cdot)$ aims at maximizing the log-likelihood for the binary classification problem:

- original data: real (1)
- generated data: fake (0)

$G(\cdot)$ aims at minimizing the log-probability of its samples being classified as “fake” by $D(\cdot)$.

$$\min_G \max_D \mathrm{E}_{\mathbf{x}\sim p_{\text{data}}} \{\log D(\mathbf{x})\} + \mathrm{E}_{\mathbf{z}\sim p(\mathbf{z})} \{\log(1 - D(G(\mathbf{z})))\}$$

I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial networks, *arXiv preprint: arXiv:1406.2661v1*, 2014.

GAN pros and cons

- ⊕ High expressiveness and adaptability to many application cases.
- ⊕ Fast sampling, even for the generation of multiple samples at the same time.
- ⊕ Direct optimization for what you care about by using perceptual samples.
- ⊖ Unstable during training.
- ⊖ Poor control, as GANs do not work directly with data distribution.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RESTART

5 Generative Latent Variable Models

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
BESTART

Latent representation

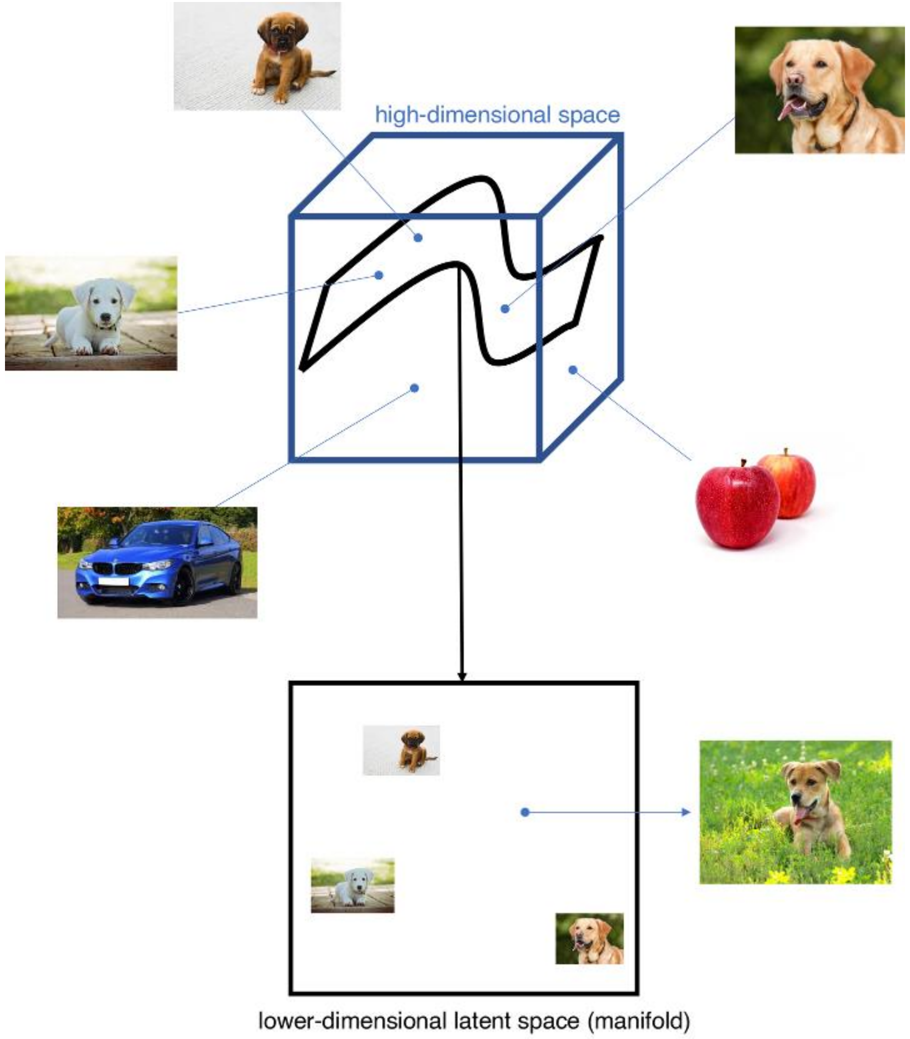
In the Plato's Allegory of the Cave, a group of people are chained inside a cave their entire life and can only see the two-dimensional shadows projected onto a wall in front of them, which are generated by unseen three-dimensional objects passed before a fire. To such people, everything they observe is actually determined by *higher-dimensional abstract concepts* that they can never be hold.

C. Luo, Understanding diffusion models: A unified perspective, *arXiv preprint: arXiv:2208.11970v1*, pp. 1–23, Aug. 2022.

Low-dimensional representation learning

Representation learning refers to the ability of neural networks to describe each observation in the training set using some *low-dimensional latent space*.

Each point in the latent space, known as a *latent variable*, is the most significant *representation* of some high-dimensional observation.



D. Foster, Generative Deep Learning – Teaching Machines to Paint, Write, Compose and Play, 2nd ed. O'Reilly Media, Inc., May 2023.

Generating by latent variables

Let us suppose that we have a collection of images with horses and we want to learn $p(\mathbf{x})$ to generate new images.

We can ask ourselves *how* we should generate a horse.

There are some *factors* in data (e.g., a silhouette, a color, a background) that are crucial for generating an object (here, a horse). Once decided, we can generate them by adding details.

D. Foster, Generative Deep Learning – Teaching Machines to Paint, Write, Compose and Play, 2nd ed. O'Reilly Media, Inc., May 2023.

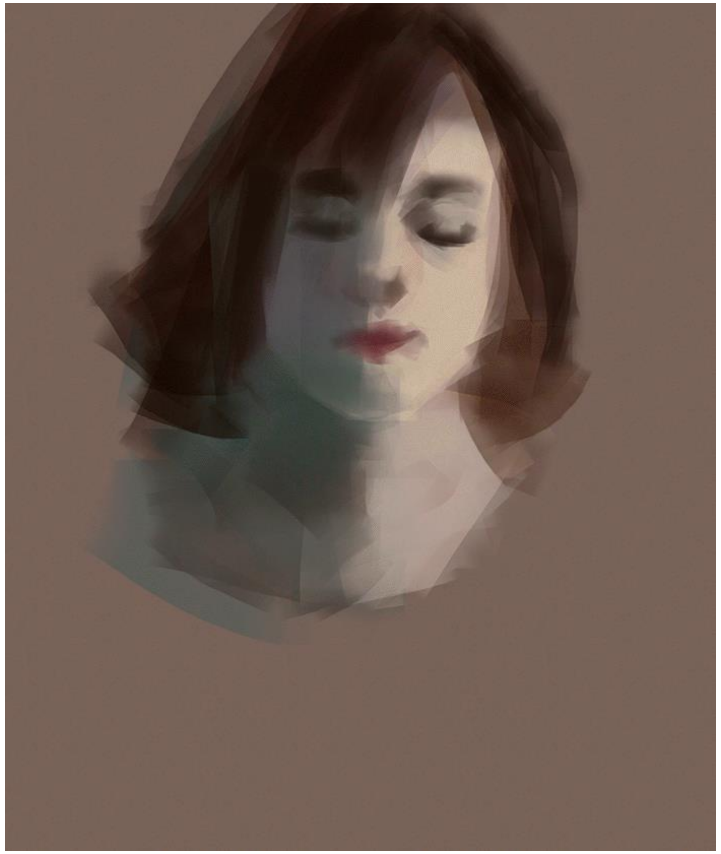
Generating by latent variables

Let us suppose that we have a collection of images with horses and we want to learn $p(\mathbf{x})$ to generate new images.

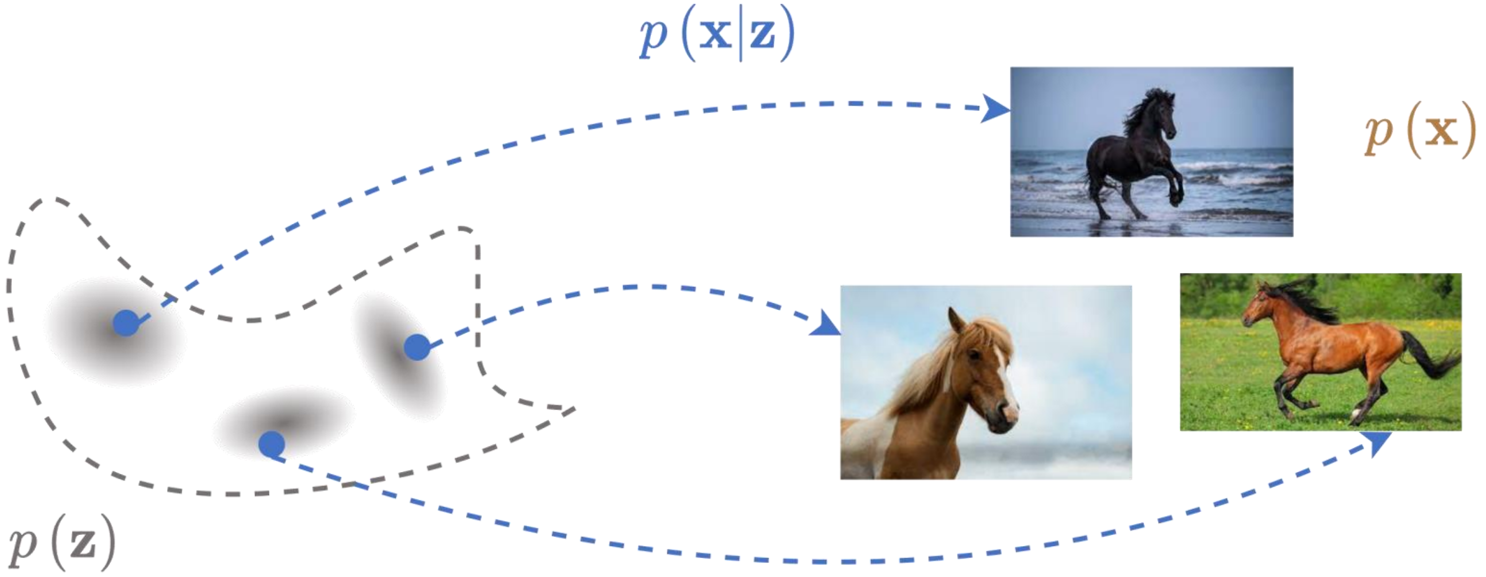
We can ask ourselves *how* we should generate a horse.

There are some *factors* in data (e.g., a silhouette, a color, a background) that are crucial for generating an object (here, a horse). Once decided, we can generate them by adding details.

Like painting.



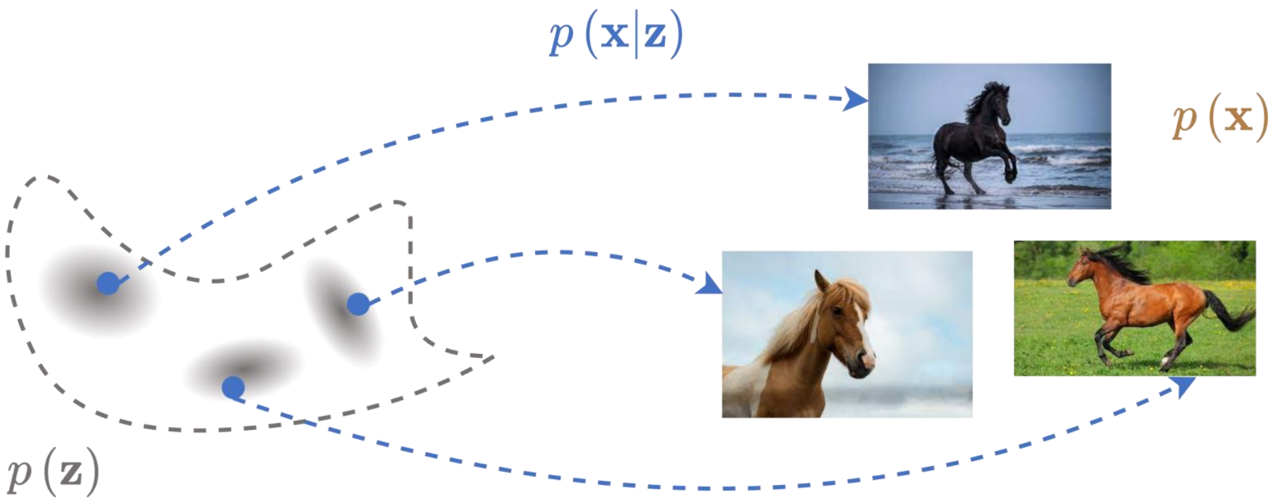
Generation process from a low-dimensional latent space



$$p(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z}$$

We first sample $\mathbf{z}$ (e.g., size, shape, and color of a horse) and then create an image with all necessary details, i.e., we sample $\mathbf{x}$ from the conditional distribution $p(\mathbf{x}|\mathbf{z})$.

Variational inference

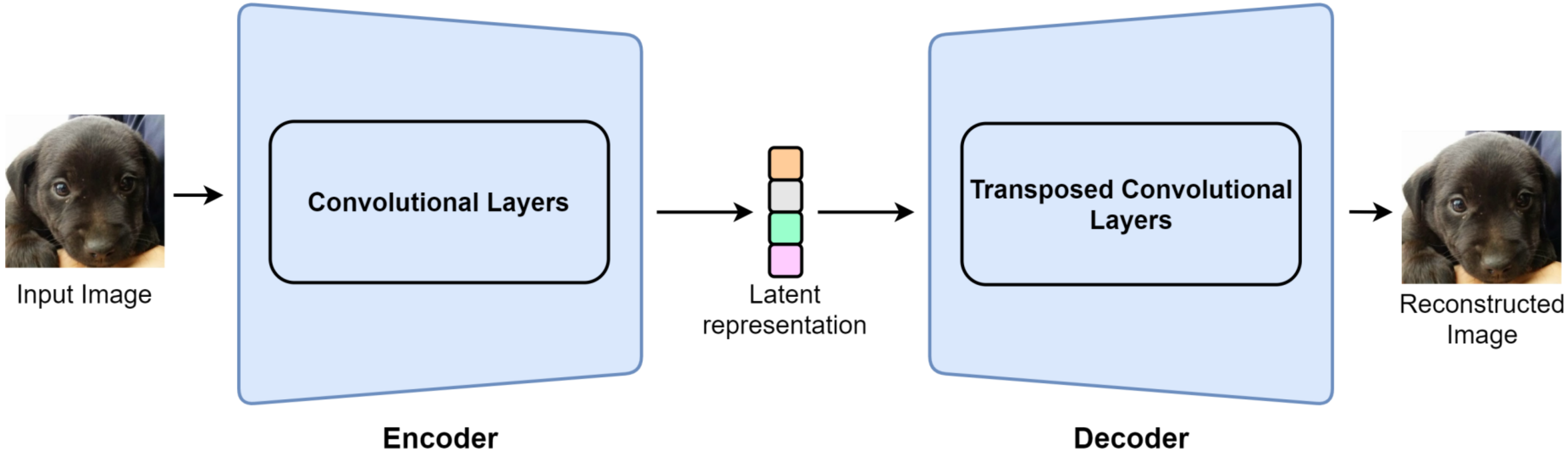


$$p(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z}$$

Intractable → approximate (or *variational*) inference

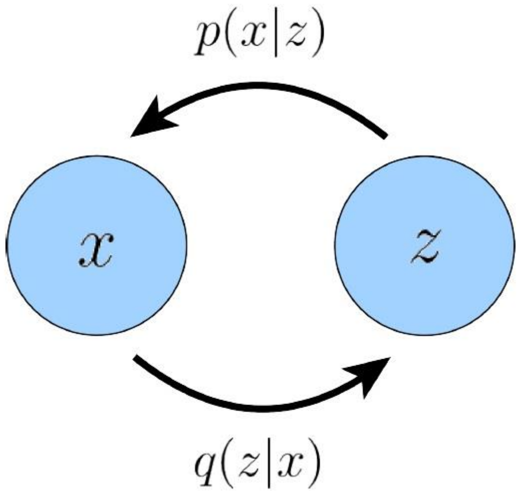
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
BESTART

Autoencoder



An autoencoder is a neural network architecture that is classically suited for denoising and reconstruction of an input, but it can be used also to reduce the input dimensionality and extract a latent representation.

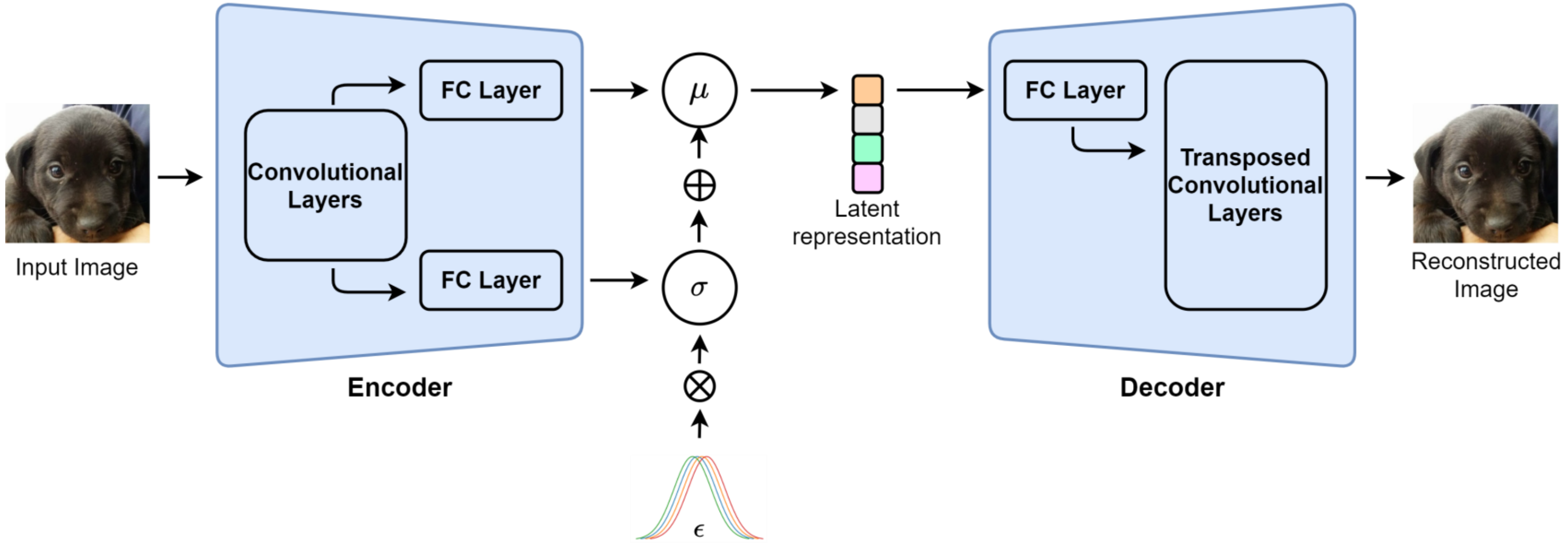
Variational autoencoder approach



In the Variational Autoencoder, the encoder $q(\mathbf{z}|\mathbf{x})$, or *recognition model*, defines a distribution over latent variables $\mathbf{z}$ for a set of observations $\mathbf{x}$, while the *generation model* $p(\mathbf{x}|\mathbf{z})$ decodes latent variables into observations.

A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, Dive Into Deep Learning, 2020.

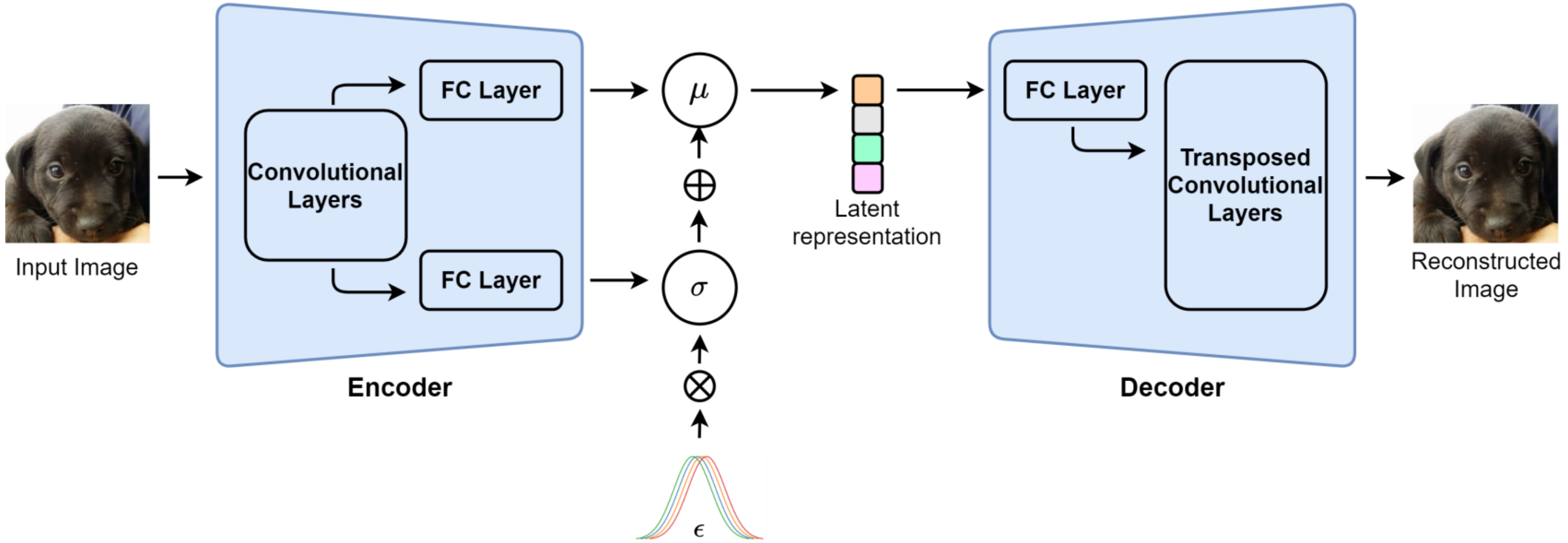
Variational autoencoder framework



A variational autoencoder introduces a control over the latent distribution.

D. P. Kingma and M. Welling, Auto-Encoding Variational Bayes, *arXiv preprint: arXiv:1312.6114v10*, May 2014.

Variational autoencoders: training

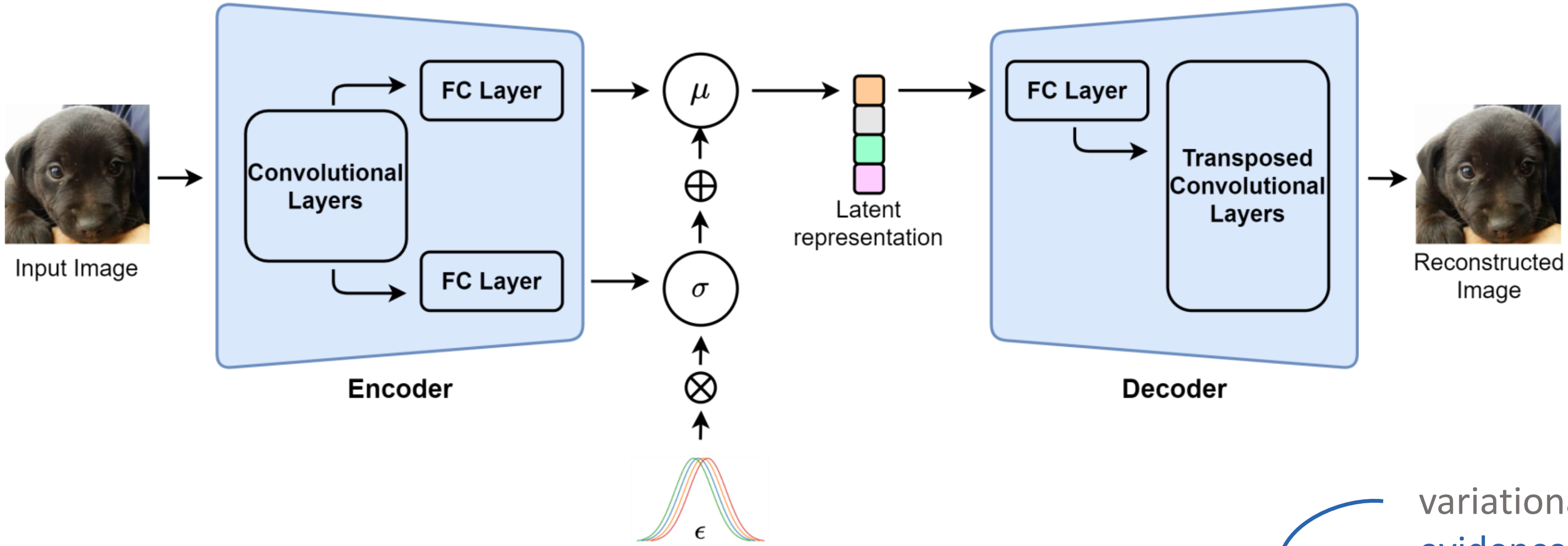


$$-\log p(x) \leq \underbrace{\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[-\log p_\theta(\mathbf{x}|\mathbf{z})\right]} + \underbrace{\mathrm{KL}\left(q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z})\right)} = \mathcal{L}_{\theta,\phi}(\mathbf{x})$$

focuses on the reconstruction

forces the latent distribution to be Gaussian

Variational autoencoders: training

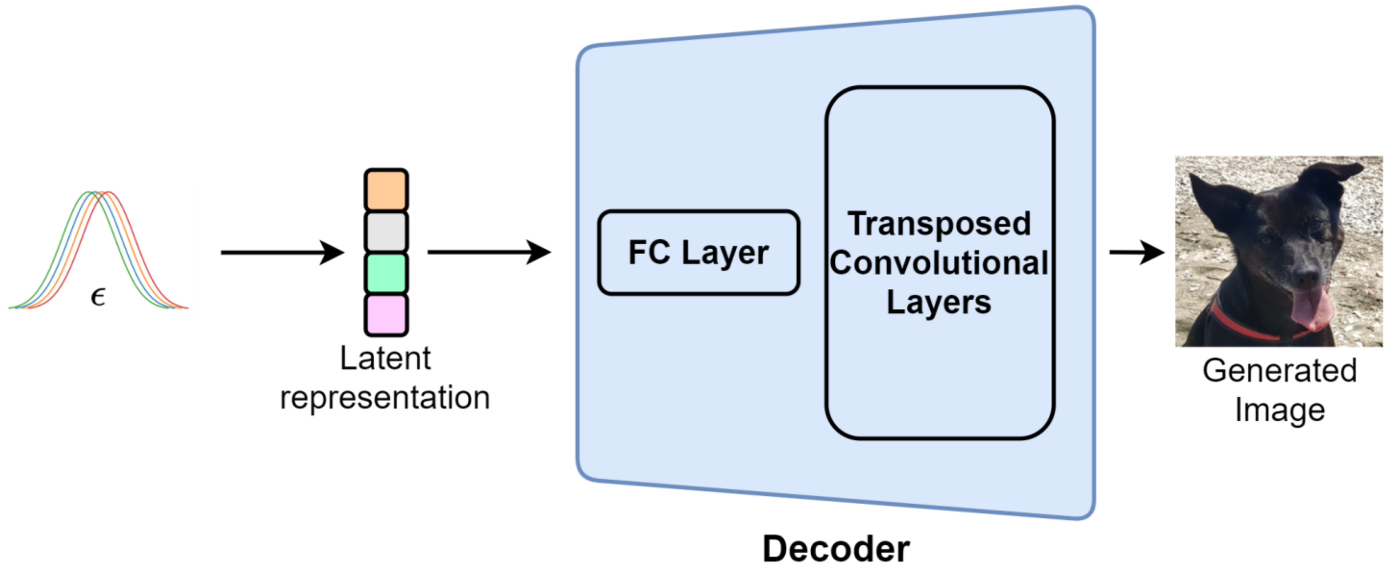


variational bound, or evidence lower bound (ELBO)

$$-\log p(x) \leq \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[-\log p_\theta(\mathbf{x}|\mathbf{z})\right] + \mathrm{KL}\left(q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z})\right) = \mathcal{L}_{\theta,\phi}(\mathbf{x})$$

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
BESTART

Variational autoencoders: generation



When a semantically meaningful latent space is learned, latent vectors can be edited before being passed to the decoder to more precisely control the data generated.

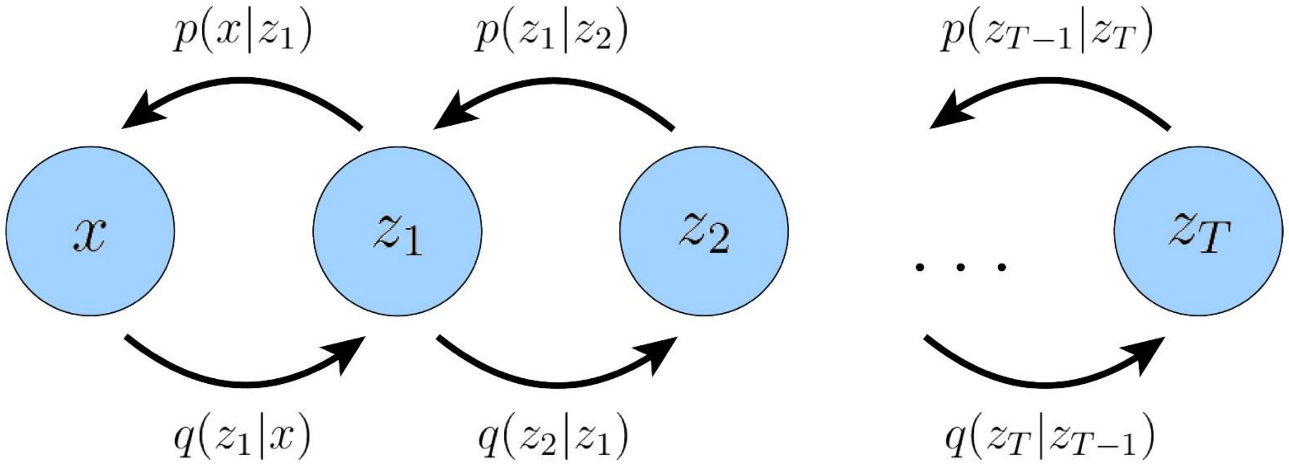
At inference, a sample is drawn from $\mathcal{N}(0, \mathbf{I})$ and passed to the decoder which generates new data.

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
RESTART

VAE pros and cons

- ⊕ The latent space can be very small therefore useful for many applications.
- ⊕ Fast sampling, even for the generation of multiple samples at the same time.
- ⊖ VAEs approximate the distribution of the data, so the samples can be imprecise.
- ⊖ VAEs are not easy to train due to the two terms in the function to be optimized.

Markovian hierarchical variational autoencoder



A Hierarchical Variational Autoencoder (HVAE) is a generalization of a VAE that extends to multiple hierarchies over latent variables.

A special case is the Markovian HVAE (MHVAE), in which the generative process is a Markov chain.

C. Luo, Understanding diffusion models: A unified perspective, *arXiv preprint: arXiv:2208.11970*, pp. 1–23, Aug. 2022.

What is a diffusion process?

The diffusion process aims to capture the dynamics of *how* latent variables disperse and diffuse in the data.

Image generated by ideogram.

Variational diffusion approach

A Variational Diffusion Model (VDM) can be seen as an MHVAE with three key assumptions:

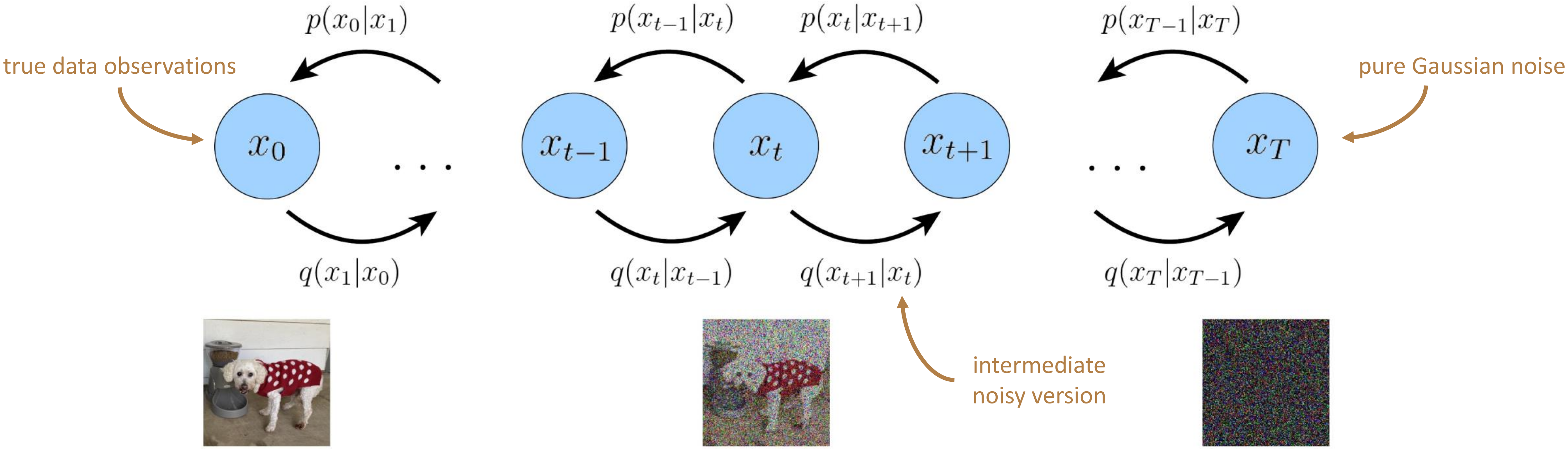
1. The latent dimension is exactly equal to the data dimension.
2. The structure of the latent encoder at each timestep is *not learned* but pre-defined as a linear Gaussian model (i.e., a Gaussian distribution centered around the output of the previous timestep).
3. The Gaussian parameters of the latent encoders vary over time in such a way that the distribution of the latent at final timestep T is a standard Gaussian.

Furthermore, the Markov property between hierarchical transitions is explicitly maintained from a standard MHVAE.

C. Luo, Understanding diffusion models: A unified perspective, *arXiv preprint: arXiv:2208.11970*, pp. 1–23, Aug. 2022.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
RESTART

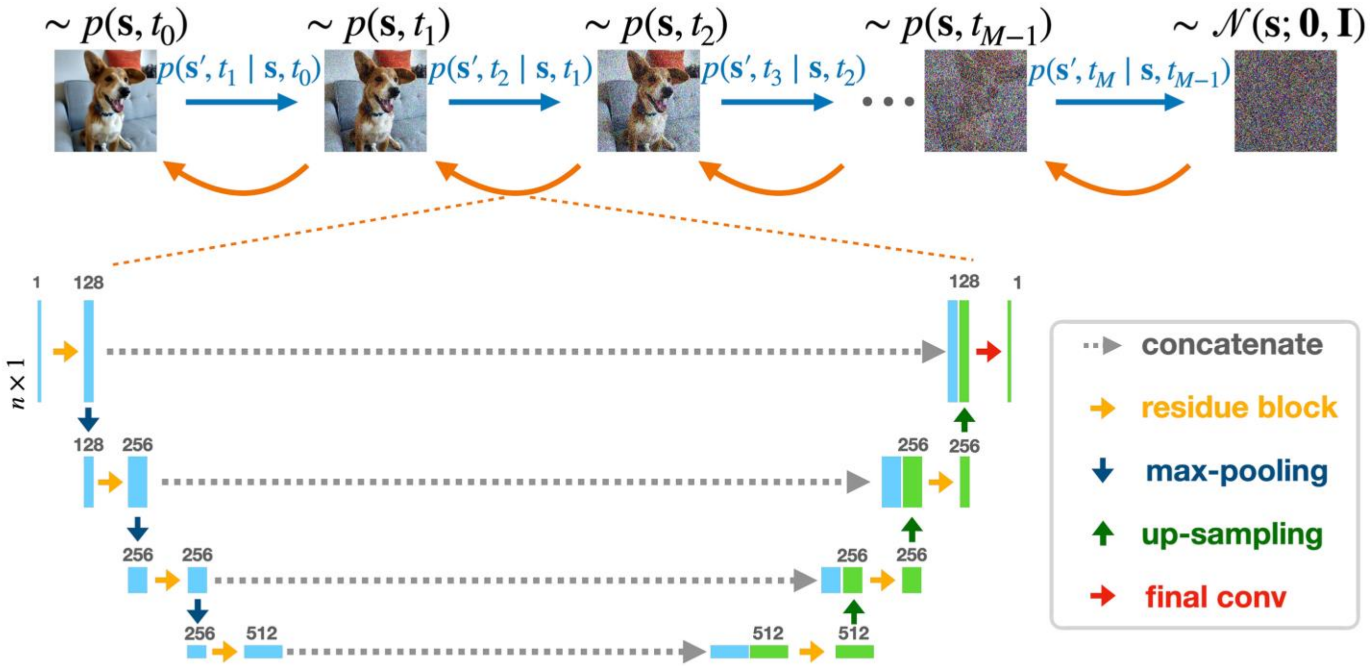
The variational diffusion process



Each encoder transition $q(x_t|x_{t-1})$ is modeled as a Gaussian distribution that uses the output of the previous state as its mean.

C. Luo, Understanding diffusion models: A unified perspective, *arXiv preprint: arXiv:2208.11970*, pp. 1–23, Aug. 2022.

Denoising diffusion probabilistic model



In the DDPM architecture a U-Net model can be used to learn predicting the source noise in diffusion models.

J. Ho, A. Jain, and P. Abbeel, Denoising Diffusion Probabilistic Models, in *Conference on Neural Information Processing Systems* (NeurIPS), vol. 33, 2020, pp. 6840–6851.

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
RESTART

Diffusion models pros and cons

- ⊕ High expressiveness, and even more controllable than GANs and VAEs.
- ⊕ Easier convergence with respect to GANs and VAEs.
- ⊖ Slower sampling with respect to GANs and VAEs.
- ⊖ Latent space dimension is large and requires huge computational resources.

6 Generative AI in ICT Applications

Transformative role in ICT

- Personalized experiences
- Autonomous networks
- Streamlines operations
- Data security and reliability
- Human in the loop

Google Cloud

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
BESTART

Solving Inverse problems

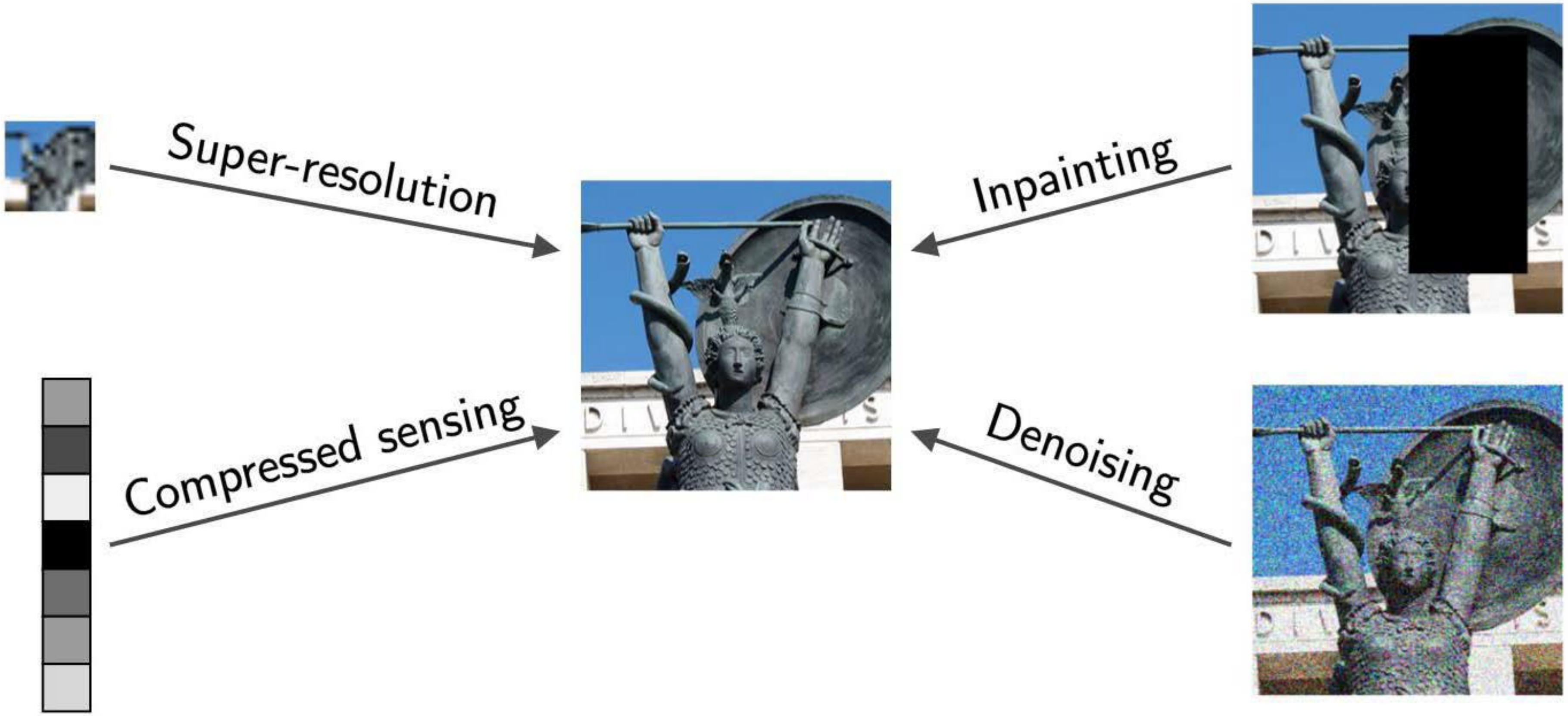
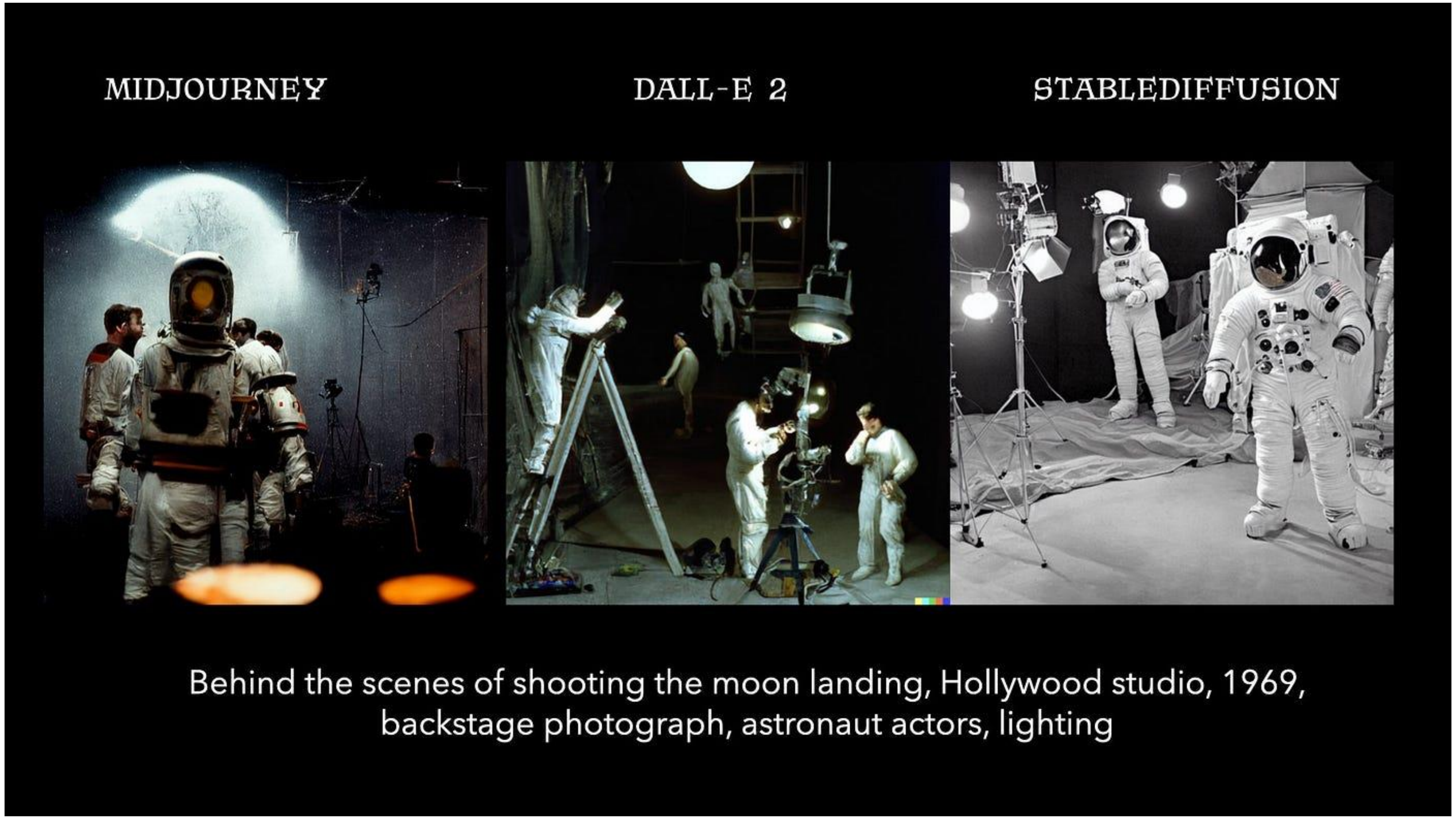


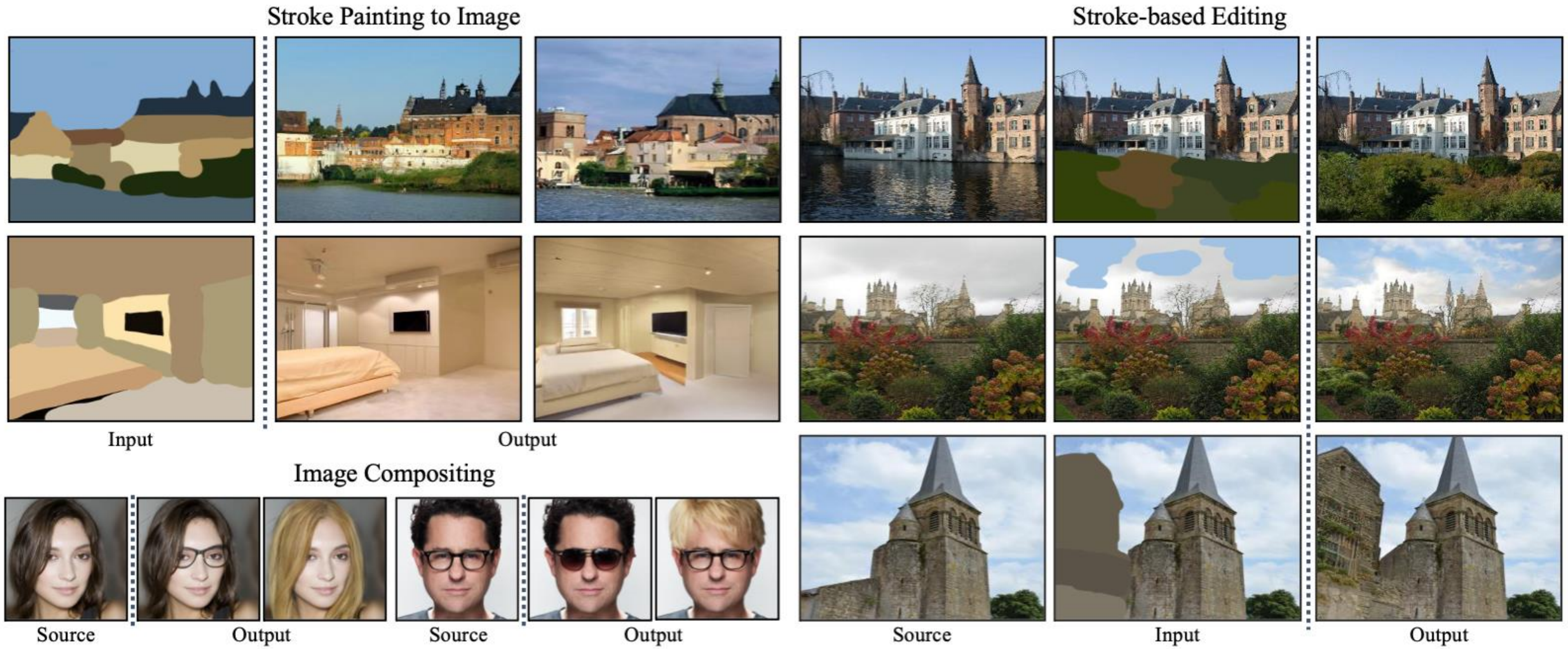
Image generation



Medium.

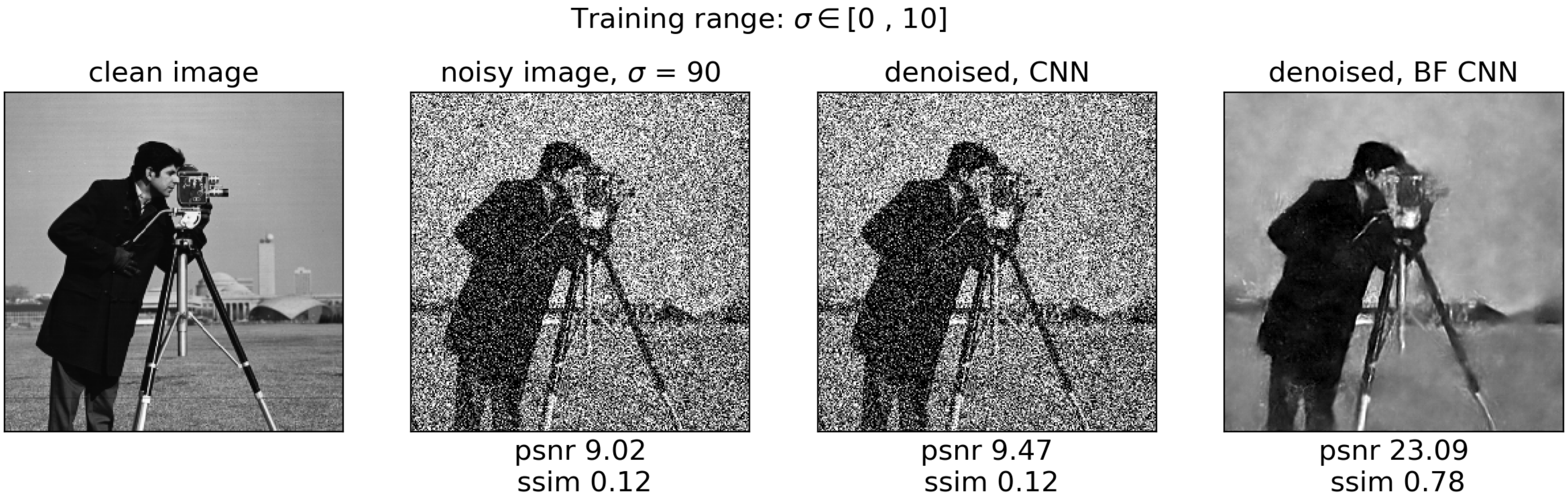
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
BESTART

Image editing



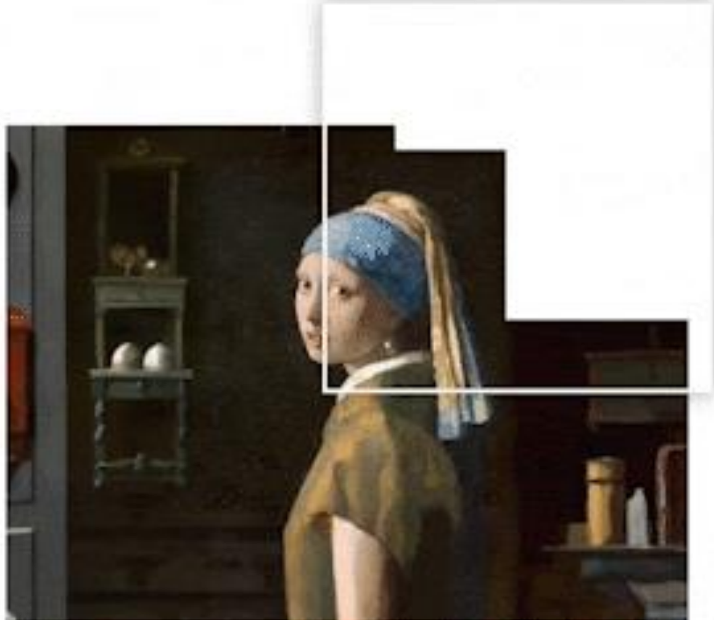
C. Meng, Y. He, Y. Song, J. Song, J. Wu, J.-Y. Zhu, S. Ermon, SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations, in *Int. Conf. on Learning Representations* (ICLR) 2022.

Image denoising



[Medium](#).

Inpainting

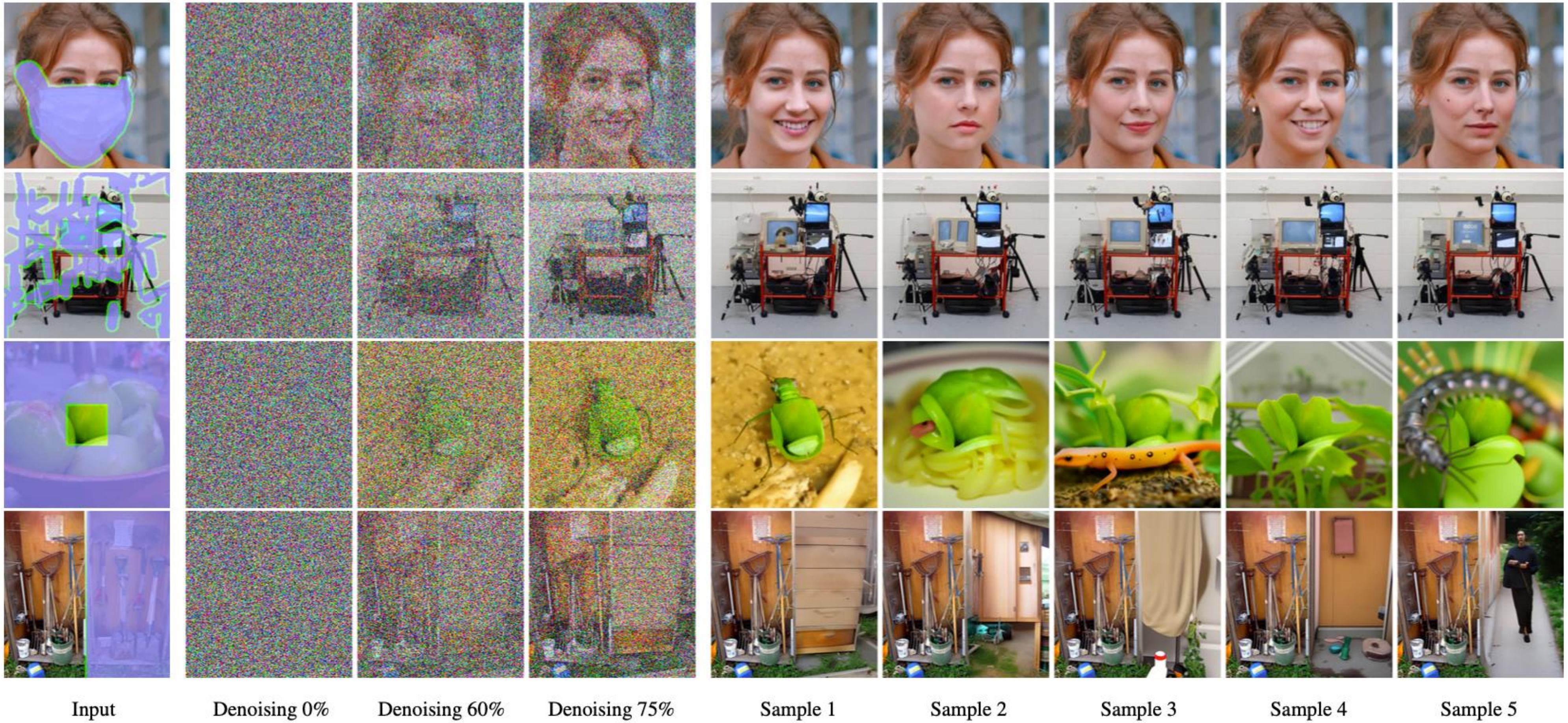


Medium.

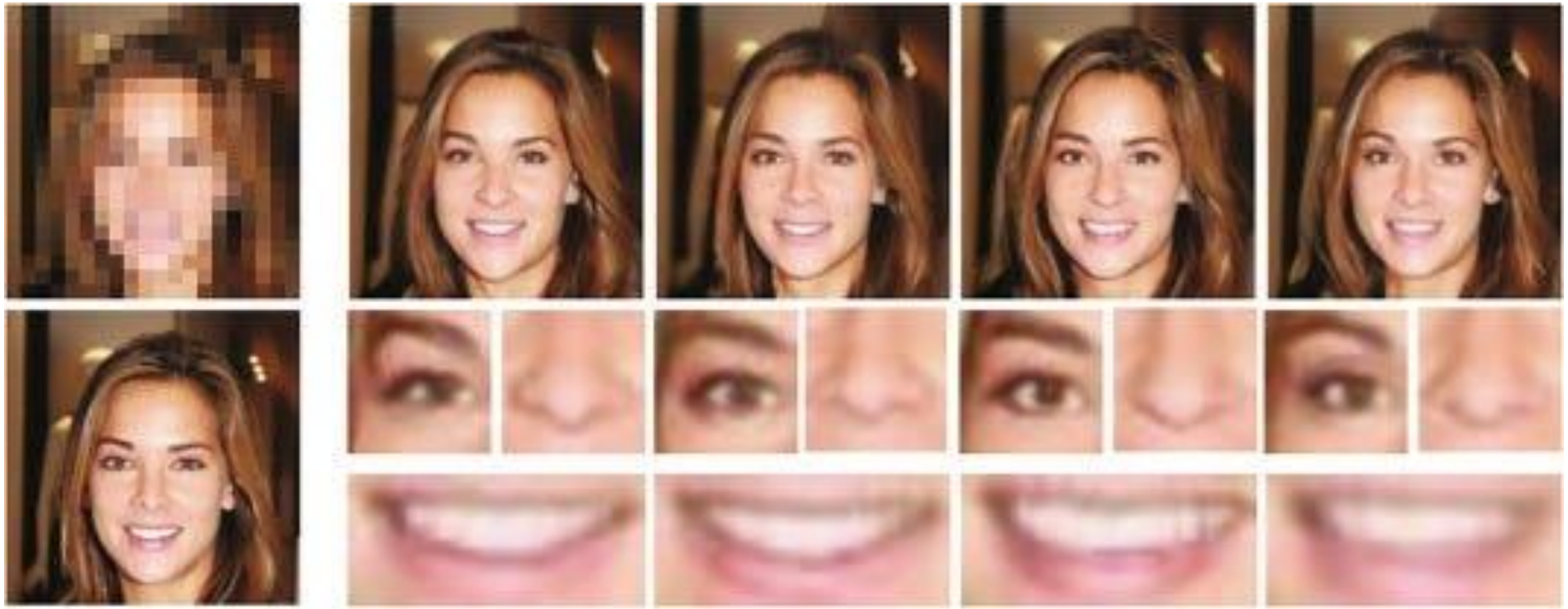
Input Image.

A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, L. Van Gool, RePaint: Inpainting Using Denoising Diffusion Probabilistic Models, Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), , pp. 11461-11471, 2022

Inpainting

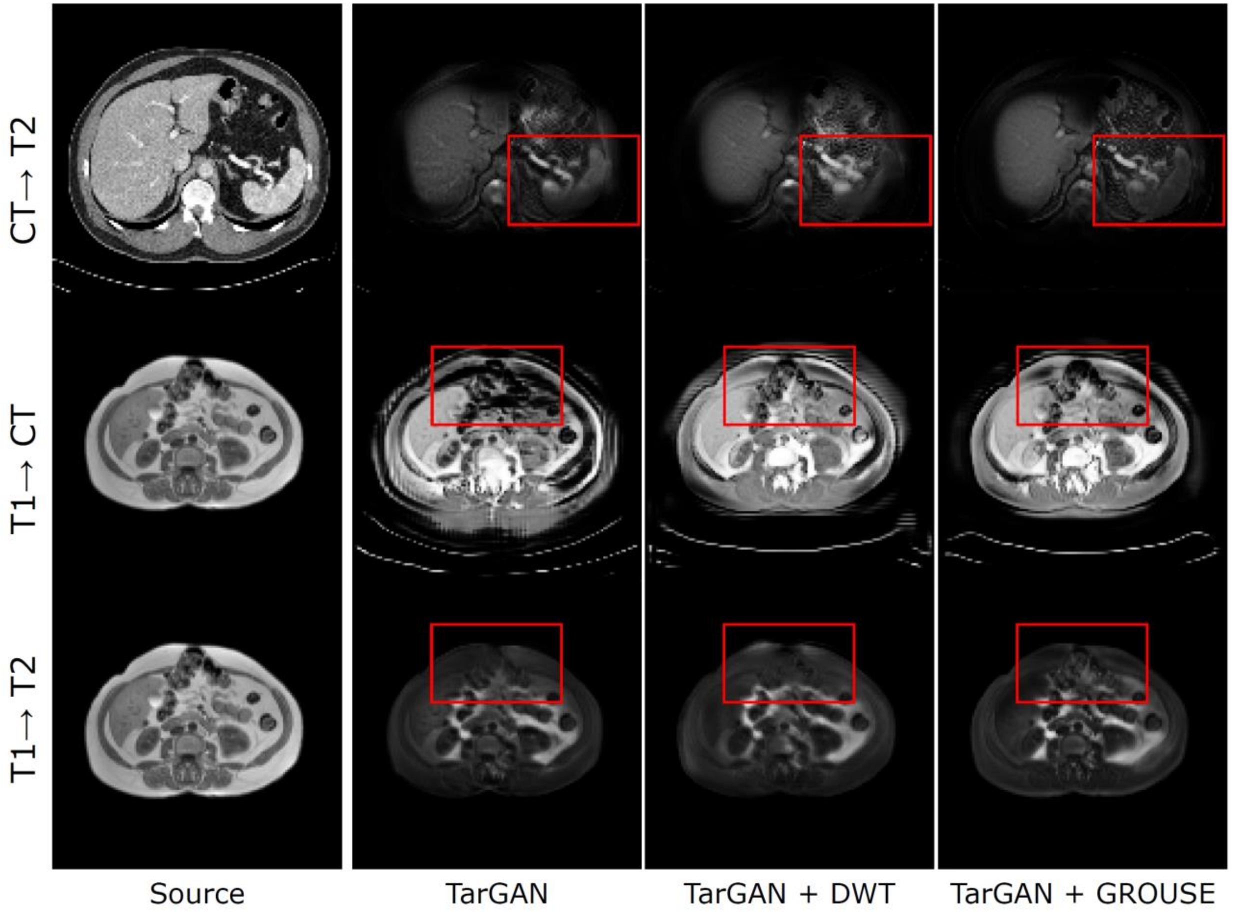

Input
Denoising 0%
Denoising 60%
Denoising 75%
Sample 1
Sample 2
Sample 3
Sample 4
Sample 5

Super-resolution



Medium.

Image Modality Translation



E. Grassucci, L. Sigillo, A. Uncini, D. Comminiello, GROUSE: A Task and Model Agnostic Wavelet-Driven Framework for Medical Imaging, to Appear in *IEEE Signal Process. Lett.*, 2023

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
BESTART

Image Modality Translation

L. Sigillo, E. Grassucci, D. Comminiello, StawGAN: Structural-Aware Generative Adversarial Networks for Infrared Image Translation, in *IEEE Int. Symp. On Circuits and Systems* (ISCAS), Monterey, CA, May 2023.

Image Modality Translation

L. Sigillo, E. Grassucci, D. Comminiello, StawGAN: Structural-Aware Generative Adversarial Networks for Infrared Image Translation, in *IEEE Int. Symp. On Circuits and Systems* (ISCAS), Monterey, CA, May 2023.

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
BESTART

Image Modality Translation

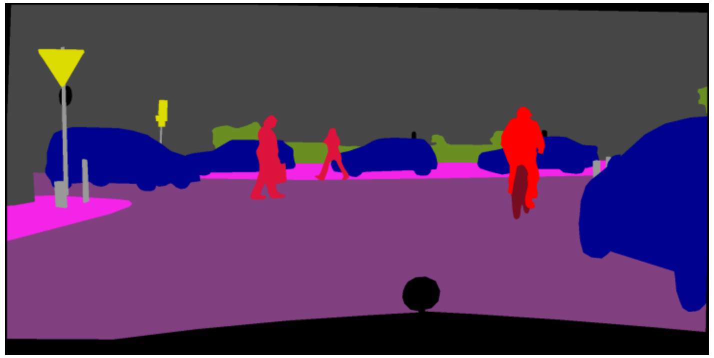
L. Sigillo, E. Grassucci, D. Comminiello, StawGAN: Structural-Aware Generative Adversarial Networks for Infrared Image Translation, in *IEEE Int. Symp. On Circuits and Systems* (ISCAS), Monterey, CA, May 2023.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
BESTART

Image Modality Translation

L. Sigillo, E. Grassucci, D. Comminiello, StawGAN: Structural-Aware Generative Adversarial Networks for Infrared Image Translation, in *IEEE Int. Symp. On Circuits and Systems* (ISCAS), Monterey, CA, May 2023.

Generative Semantic Communication

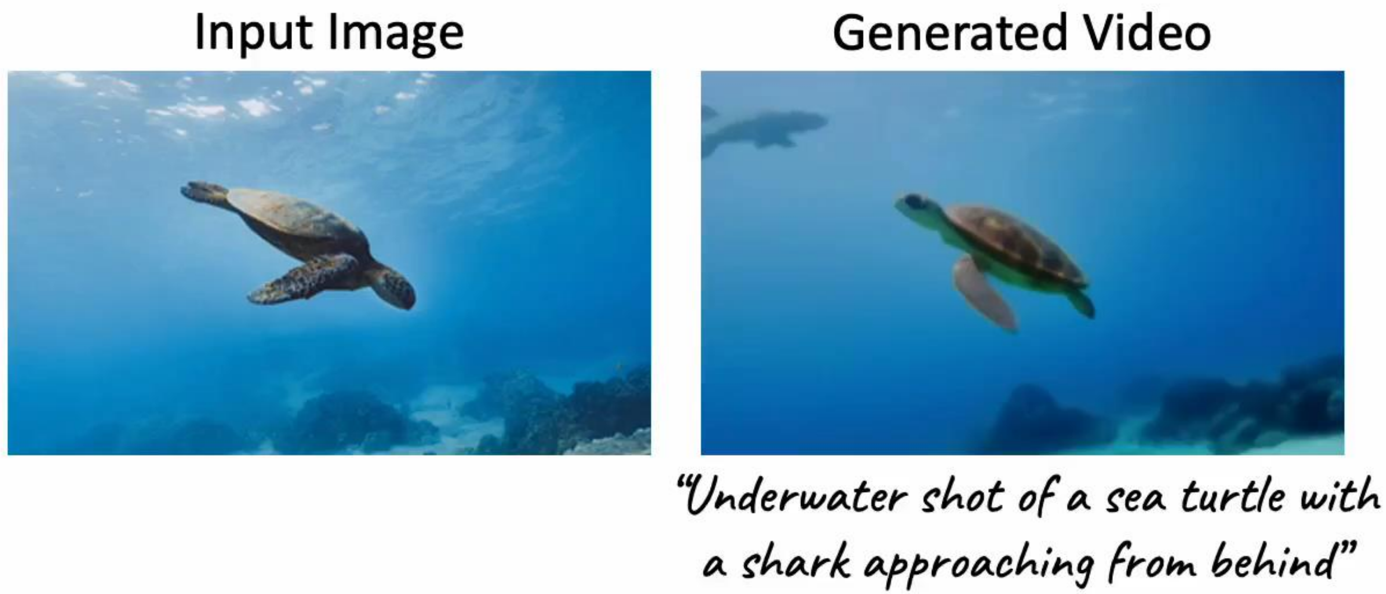


Semantic Generation

Different samples with preserved semantic scenario

Images generated with: E. Grassucci, S. Barbarossa, D. Comminiello, Generative Semantic Communication: Diffusion Models Beyond Bit Recovery, *arXiv preprint arXiv:2306.04321*, 2023.

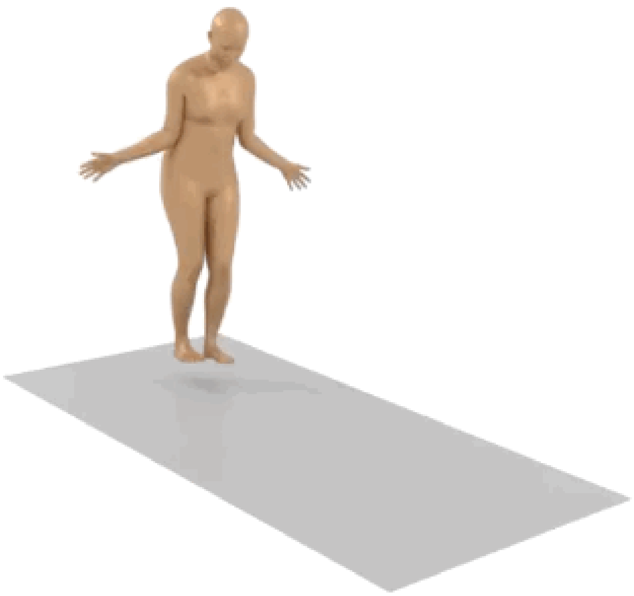
Video generation



E. Molad, E. Horwitz, D. Valevski, A. R. Acha, Y. Matias, Y. Pritch, Y. Leviathan, Y. Hoshen, Dreamix: Video Diffusion Models are General Video Editors, *arXiv preprint: arXiv:2302.01329*, 2023.

Computer graphics and animation

"A person punches in a manner consistent with martial arts."



"A person is skipping rope."

"a man kicks with something or someone with his left leg."

G. Tevet, S. Raab, B. Gordon, Y. Shafir, D. Cohen-Or, A. H. Bermano, Human Motion Diffusion Model, *arXiv preprint arXiv:2209.14916*, 2022.

Generative AI for digital twin

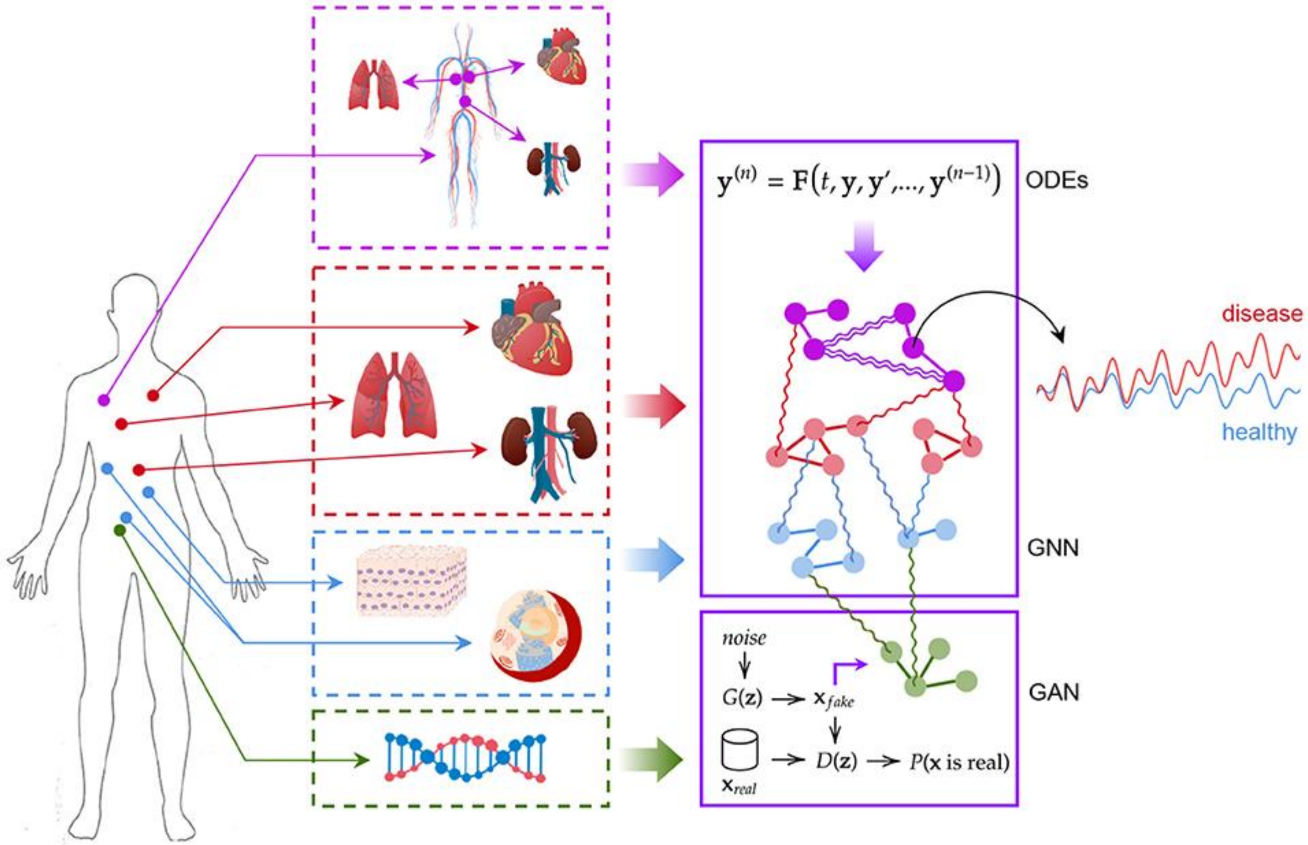


Image credit: Pietro Barbiero et al, University of Cambridge, UK

Generative design technology

- Structural topology optimization optimizes the material layout within given design spaces for a given set of functional requirements and constraints.
- This reduces the amount of material required to meet product requirements and reduces material waste and cost.
- Combining topology optimization and additive manufacturing gives manufacturers the capability to produce complex shapes that were previously impossible, and it accelerates the production of finished parts.

7 Conclusion and Future Directions

Training data matter

Semantic Generation →

Data dependence

Semantic Generation

Inference time

Diffusion models training is stabler with respect to GANs and VAEs.

The diffusion model sampling needs several steps of a Markov chain.

Communication systems need fast inference, while diffusion models suffer from **slow sampling**.

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

BESTART

Generative models are huge

To obtain impressive results in real-world datasets and scenarios, deep generative models require a huge **number of learnable parameters**.

Large **storage memory demand** just for storing the learned weights of a pretrained generative model (even only for doing inference).

Stable Diffusion

890M learnable parameters

3.5GB storage memory required

Sustainability and resource management

Deep generative models require large **computational resources** and consequently very high **energy consumption**.

Needs GPUs with **large VRAM** and often takes days of computation, producing a **massive amount of CO_2**.

Also the inference needs **GPUs** that are hard to embed on smaller devices.

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
RESTART

What we can do and future directions

Reduce the computations or make network blocks reusable by considering pruning, quantization, and modular networks.

Release code and pretrained models to make them accessible and available for future researchers.

Fine-tune the released models without retraining them.

Develop **faster sampling** strategies tailored for communication systems.

References

- Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, B. Poole, Score-Based Generative Modeling through Stochastic Differential Equations, *Int. Conf. On Learning Representations* (ICLR), 2021.
- D. Foster, Generative Deep Learning – Teaching Machines to Paint, Write, Compose and Play, 2nd ed. O'Reilly Media, Inc., May 2023.
- A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, Dive Into Deep Learning, 2020.
- I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative Adversarial Networks, *arXiv preprint: arXiv:1406.2661v1*, 2014.
- A. Radford, L. Metz, and S. Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, *arXiv preprint: arXiv:1511.06434v2*, pp. 1–16, Jan. 2016.
- C. Luo, Understanding diffusion models: A unified perspective, *arXiv preprint: arXiv:2208.11970*, pp. 1–23, Aug. 2022.
- J. M. Tomczak, Deep Generative Modeling. Springer Nature Switzerland AG, 2022.
- D. P. Kingma and M. Welling, Auto-Encoding Variational Bayes, *arXiv Preprint: arXiv:1312.6114v10*, May 2014.
- D. J. Rezende, L. Metz, and S. Chintala, Stochastic Backpropagation in Approximate Inference in Deep Generative Models, in *Int. Conf. on Machine Learning* (ICML), Beijing, China, Jun. 2014, pp. 1278–1286.

References

- J. Ho, A. Jain, and P. Abbeel, Denoising Diffusion Probabilistic Models, in *Conference on Neural Information Processing Systems* (NeurIPS), vol. 33, 2020, pp. 6840–6851.
- D. Kingma, B. Poole, and J. Ho, Variational Diffusion Models, in *Conference on Neural Information Processing Systems* (NeurIPS), vol. 34, 2021, pp. 21 696–21 707.
- C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. Denton, S. K. S. Ghasemipour, B. K. Ayan, S. S. Mahdavi, R. G. Lopes, T. Salimans, J. Ho, D. J. Fleet, and M. Norouzi, Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding," in *Conference on Neural Information Processing Systems* (NeurIPS), 2022.
- Y. Song and S. Ermon, Generative modeling by estimating gradients of the data distribution, in *Conference on Neural Information Processing Systems* (NeurIPS), vol. 32, 2019.
- J. Ho, C. Saharia, W. Chan, D. J. Fleet, M. Norouzi, and T. Salimans, Cascaded Diffusion Models for High Fidelity Image Generation, *Journal of Machine Learning Research*, pp. 23–47, 2022.
- C. Meng, Y. He, Y. Song, J. Song, J. Wu, J.-Y. Zhu, S. Ermon, SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations, in *Int. Conf. on Learning Representations* (ICLR) 2022.
- A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, L. Van Gool, RePaint: Inpainting Using Denoising Diffusion Probabilistic Models, *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition* (CVPR), pp. 11461-11471, 2022

References

- E. Grassucci, L. Sigillo, A. Uncini, D. Comminiello, GROUSE: A Task and Model Agnostic Wavelet-Driven Framework for Medical Imaging, to Appear in *IEEE Signal Process. Lett.*, 2023.
- L. Sigillo, E. Grassucci, D. Comminiello, StawGAN: Structural-Aware Generative Adversarial Networks for Infrared Image Translation, in *IEEE Int. Symp. On Circuits and Systems* (ISCAS), Monterey, CA, May 2023.
- E. Grassucci, S. Barbarossa, D. Comminiello, Generative Semantic Communication: Diffusion Models Beyond Bit Recovery, *arXiv preprint arXiv:2306.04321*, 2023.
- E. Molad, E. Horwitz, D. Valevski, A. R. Acha, Y. Matias, Y. Pritch, Y. Leviathan, Y. Hoshen, Dreamix: Video Diffusion Models are General Video Editors, *arXiv preprint: arXiv:2302.01329*, 2023.
- G. Tevet, S. Raab, B. Gordon, Y. Shafir, D. Cohen-Or, A. H. Bermano, Human Motion Diffusion Model, *arXiv preprint arXiv:2209.14916*, 2022.

https://danilocomminiello.site.uniroma1.it

danilo.comminiello@uniroma1.it

www.ispamm.it