Diffusion Models summary

D. Tsitselas, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr

Version 3.0

Artificial Intelligence &
Information Analysis Lab

Diffusion Models

- **Generative Modeling**
- Diffusion Models Methods
- Diffusion Models Applications on CV
- Future Directions

Generative Modeling

- Classic ML/DL methods:
  - Regression
  - Classification
  - Clustering

- Generative Modeling

  Unsupervised learning

  Model the dataset’s underlying structure

  Create new datapoints that resemble the original data

Generative Modeling - Applications

- Computer Vision
  - Unconditional image generation
  - Super-resolution, Inpainting, Image restoration
  - Semantic image segmentation
  - Video generation
  - Point cloud completion - generation
- Natural Language Processing
  - Composition of meaningful text

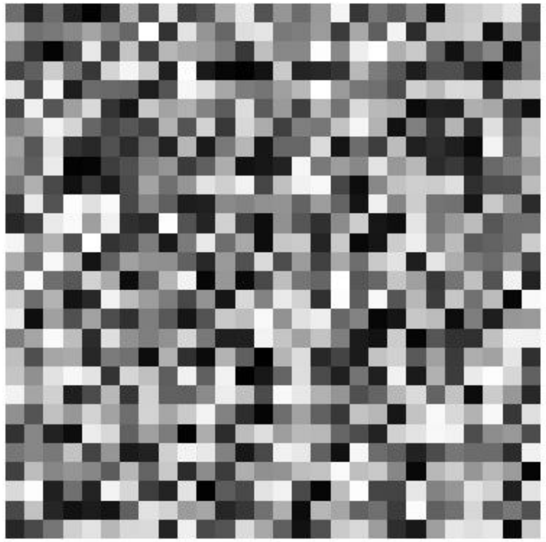
Generative Modeling - Applications

- Temporal Data Modeling
  - Time series forecasting
  - Audio synthesis
- Multi-Modal Learning
  - Text to Image
  - Text to Audio
  - …
- Others
  - Drug design
  - Inverse problems (e.g. medical imaging)

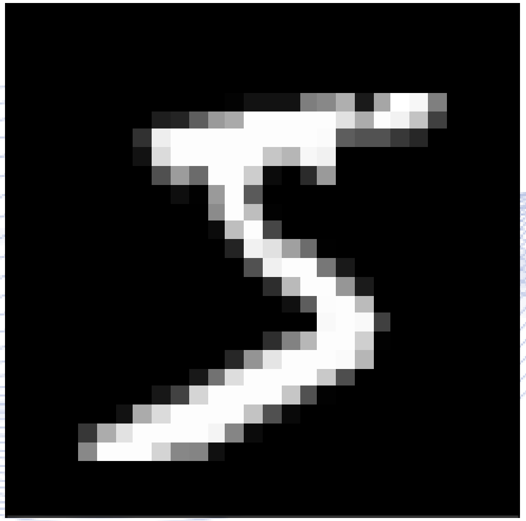
Generative Modeling

- Generative models view the data as samples of an unknown distribution
- Task:

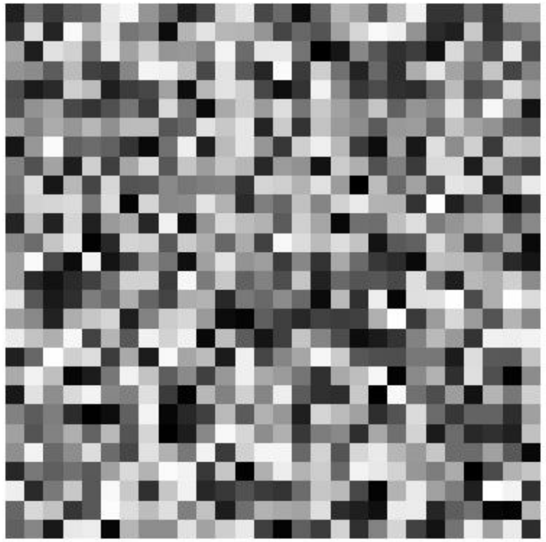
Find the distribution, the dataset was sampled from.



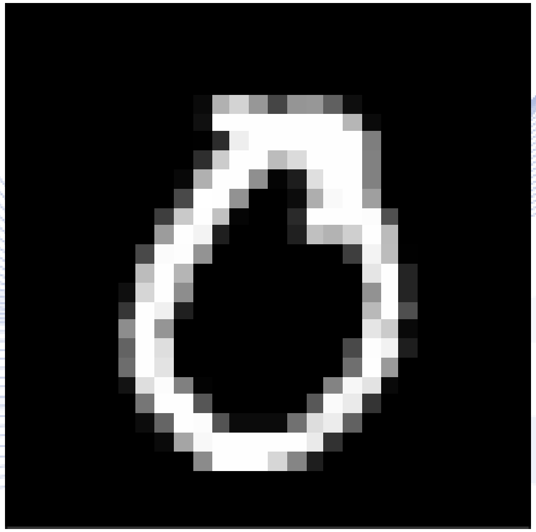
Not MNIST



MNIST



Not MNIST



MNIST

Diffusion Models

- Generative Modeling
- **Diffusion Models Methods**
- Diffusion Models Applications on CV
- Future Directions

Variational Auto-Encoder (VAE)

Generative modeling method based on

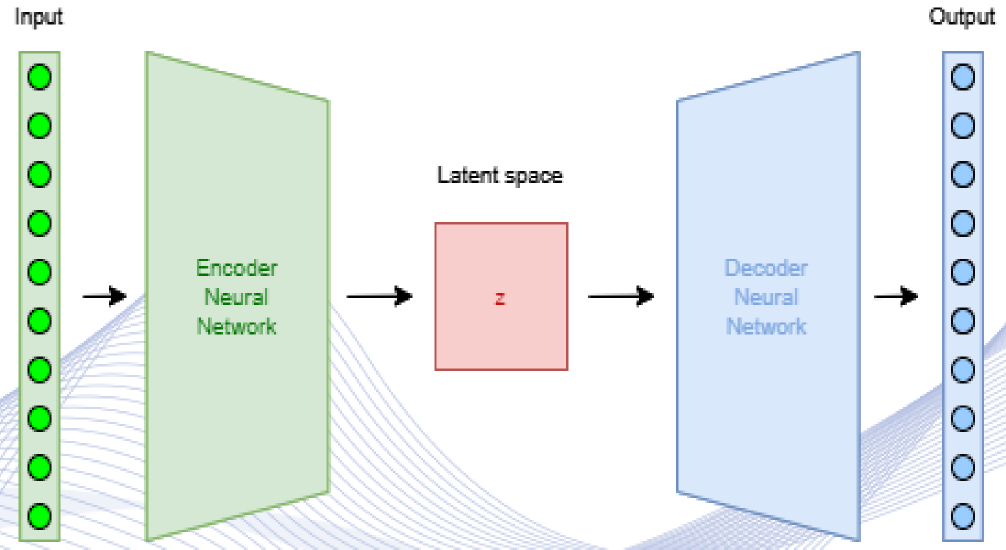
Auto-Encoder architecture

<u>Auto-Encoder</u>

Encodes input $\boldsymbol{x} \in \mathbb{R}^N$ to a latent variable $\boldsymbol{z} \in \mathbb{R}^M$

Decodes $\boldsymbol{z}$ to $\hat{\boldsymbol{x}}$ with loss

$$l(\mathbf{x}, \hat{\mathbf{x}}) = \| \hat{\mathbf{x}} - d(e(\mathbf{x})) \|^2$$



Autoencoder architecture

Typically, they are implemented with a U-Net.

Latent space's distribution is unknown not regularized

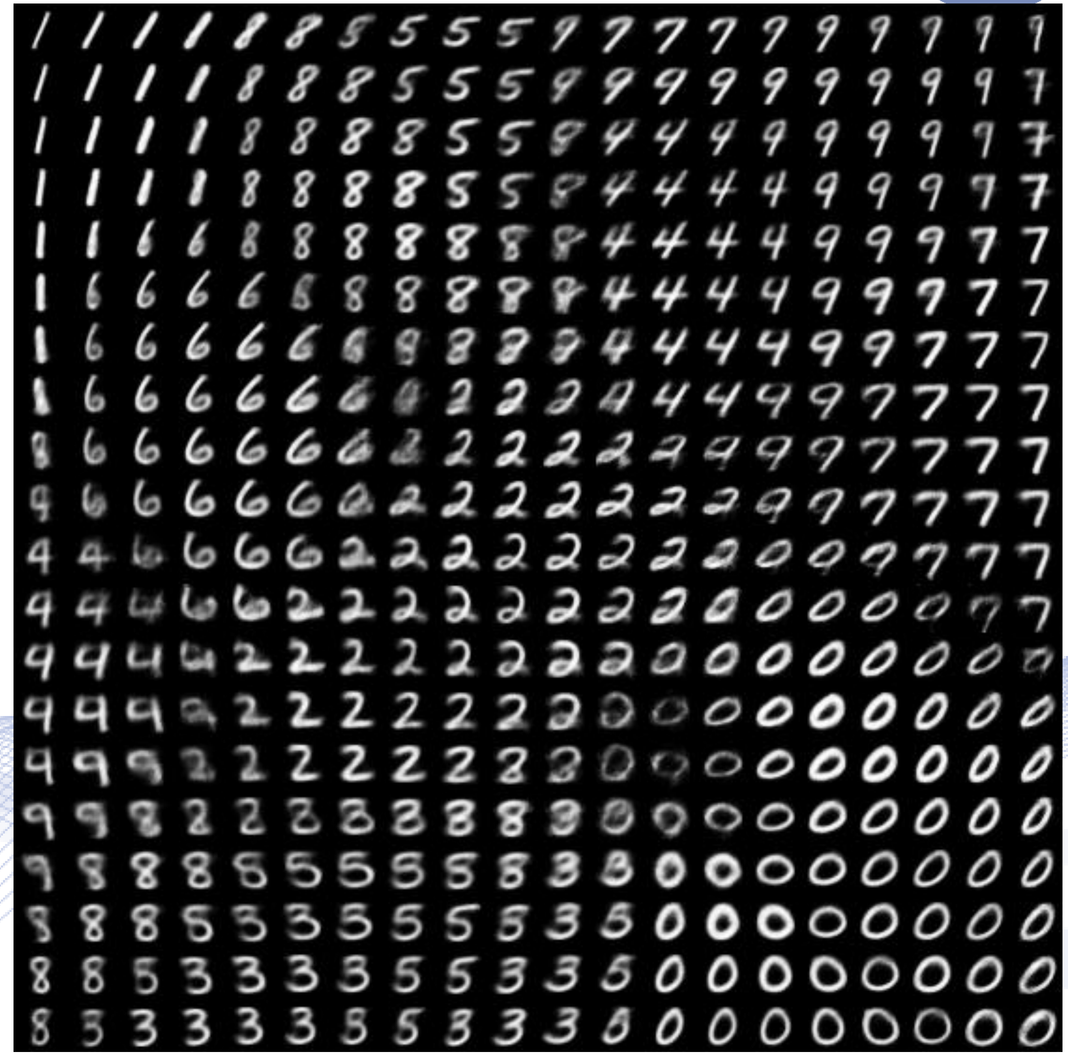
Variational Auto-Encoder (VAE)

For this image $\boldsymbol{z} \in \mathbb{R}^2$

The center of the grid is $(0, 0)$

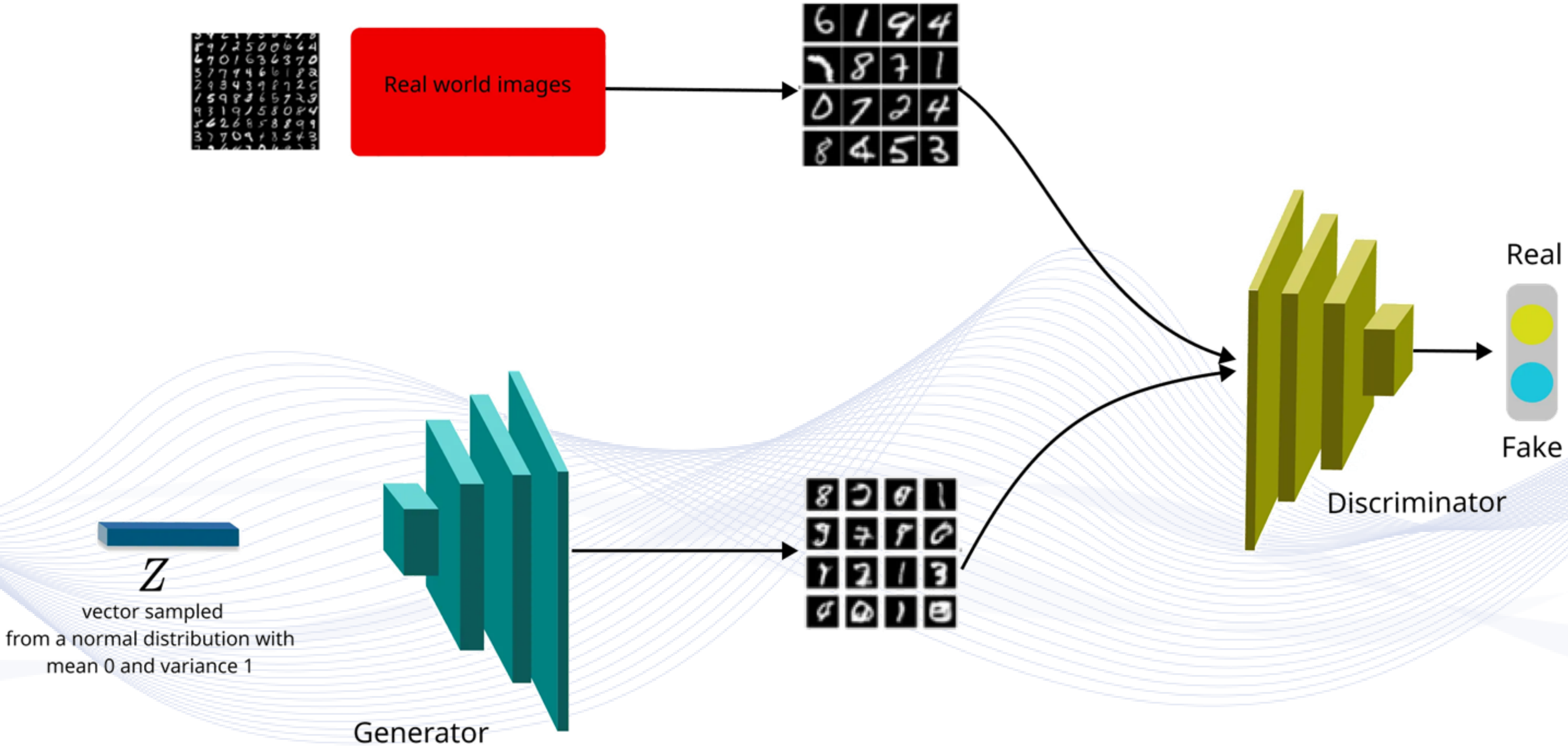
Note that continuous change is the
Latent results in continuous change in
The final image.

This might be a problem – VQ-VAE as a solution



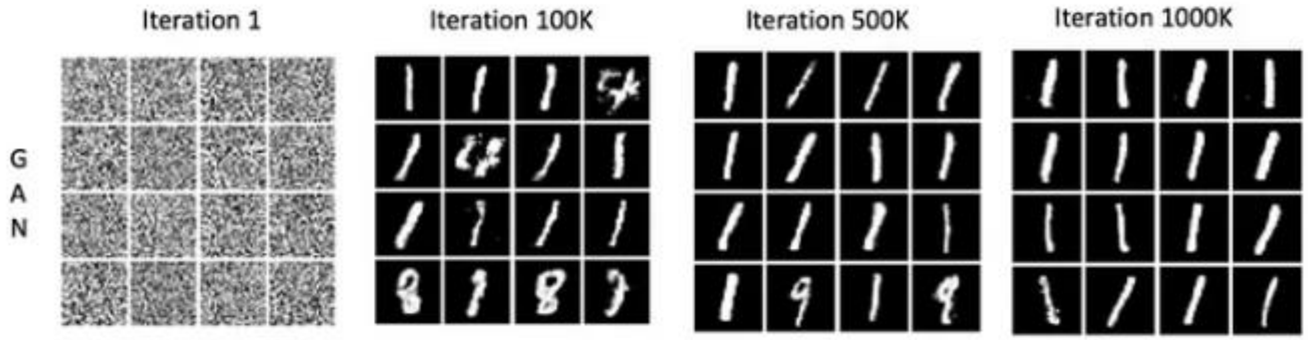
VAE latent space visualization

Generative Adversarial Networks (GAN)



Limitations

- Often we have unnormalized distributions. Trying to optimize the log-likelihood (VAE) can be challenging, due to the unknown normalization factor.
- GANs exhibit mode collapse

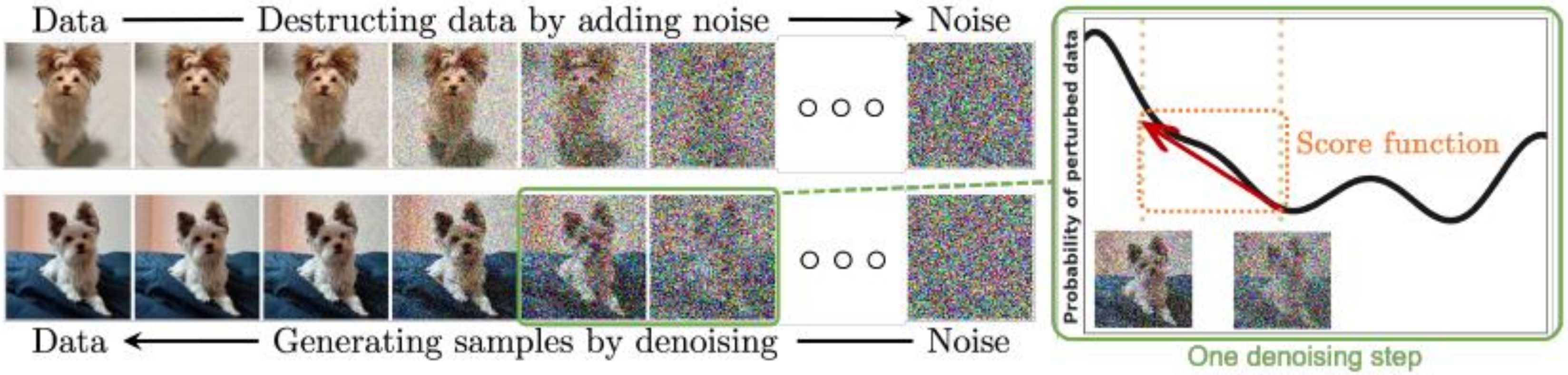


- Very difficult to train

Diffusion Models

Diffusion Models consist of three main processes

1. The forward process that adds noise to the dataset. Most of the time is handcrafted
2. The reverse process that learns how to generate images from noise
3. Sampling that applies the trained reverse process to generate samples



Diffusion Model Process [1]

Diffusion Models

They come in three formulations

1. Denoising Diffusion Probabilistic Models (DDPMs)
2. Score-based Generative Models (SGMs)
3. Score – Based Generative Model with Stochastic Differential Equations (SDEs)

DDPM

- Both the forward and the reverse process are modeled as Markov Chains

$$P(\mathbf{x}_n, \mathbf{x}_{n-1}, \dots, \mathbf{x}_1, \mathbf{x}_0) = P(\mathbf{x}_0) \prod_{i=1}^{n} P(\mathbf{x}_i \mid \mathbf{x}_{i-1})$$

- The forward process has Gaussian transition kernels with predetermined parameters

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I}\right)$$

- It gradually perturbs the data, in T steps, with the goal to produce

$$\mathbf{x}_T = \mathcal{N}(0; \mathbf{I})$$

DDPM

Algorithm 1 Training

repeat

$\mathbf{x}_0 \sim q(\mathbf{x}_0)$

$t \sim \mathrm{Uniform}(\{1, \ldots, T\})$

$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

Take gradient descent step on

$$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$$

until converged

Algorithm 2 Sampling

$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

for $t = T, \ldots, 1$ **do**

$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right) + \sigma_t \mathbf{z}$$

end for

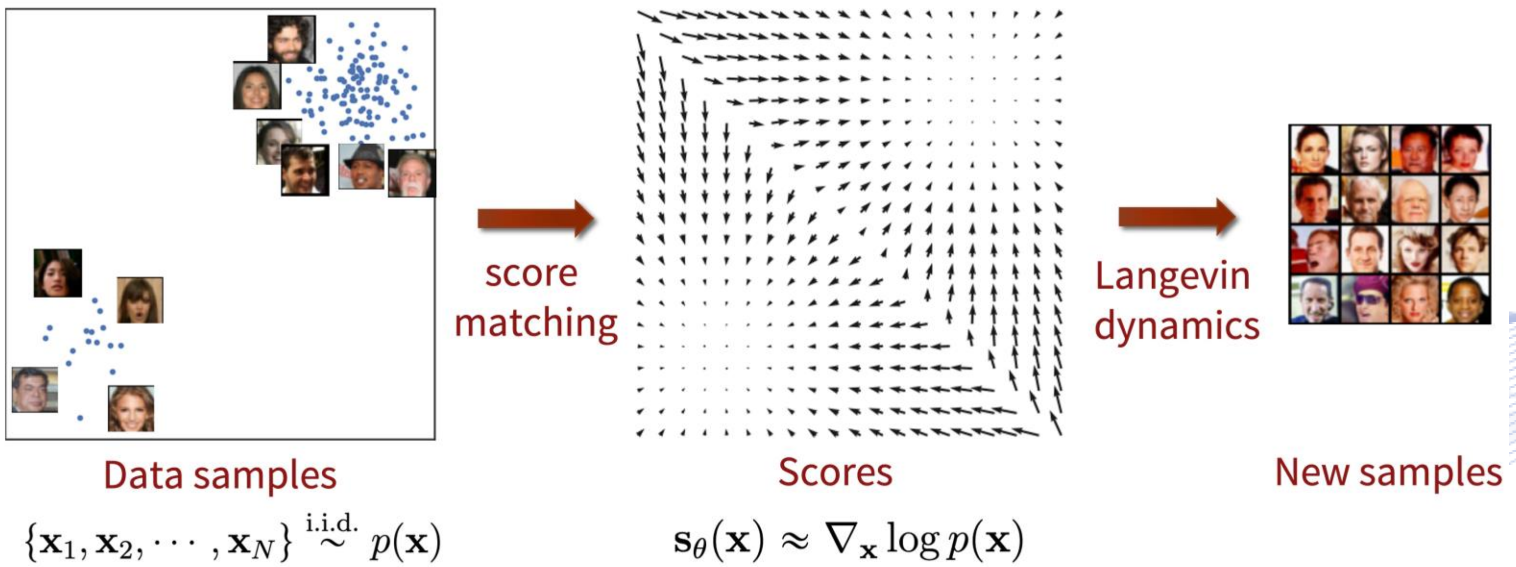
return $\mathbf{x}_0$

DDPM Training and Sampling algorithms [6]

SGM

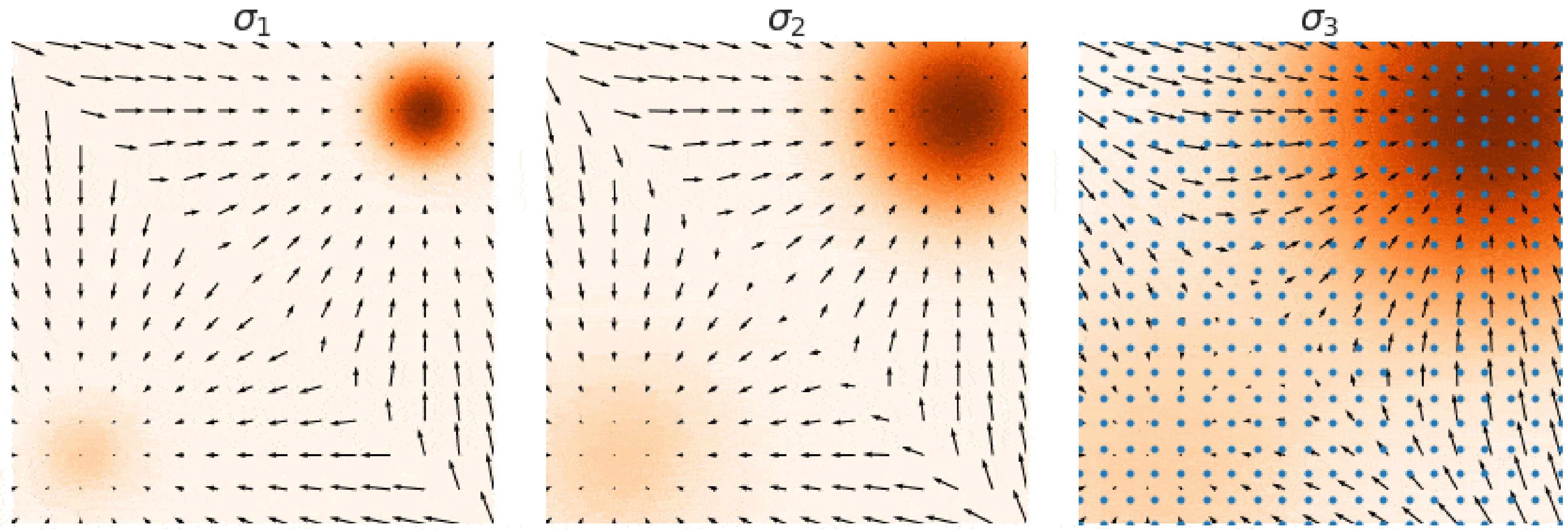
Overview of the method

1. Learn the field of scores using a score matching method
2. Begin from a random position
3. Using an iterative method, follow the gradients and end up on an image



Naïve score matching on the original distribution [1]

SGM



Score based diffusion model [1]

SGM - SDE

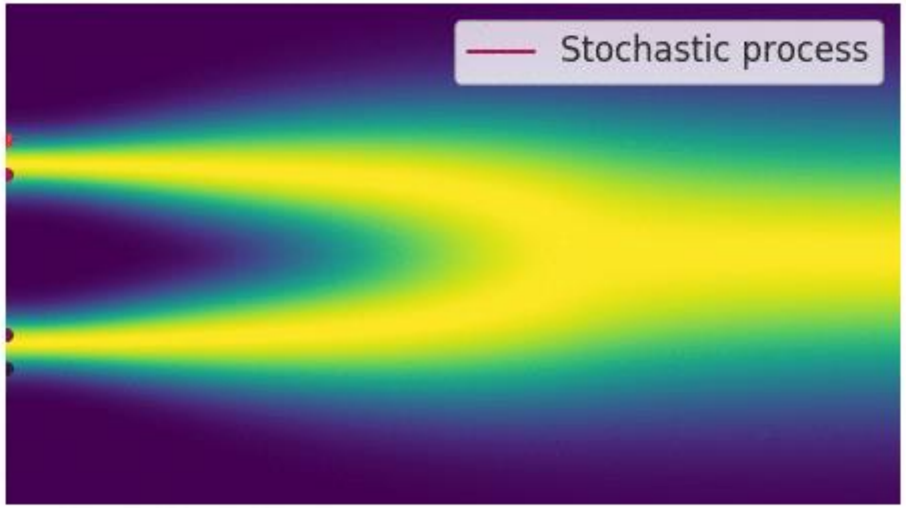
- Both previous methods use discrete time formulations to describe their forward and reverse processes.
- DDPMs and SGMs are special cases of more general formulation that uses Stochastic Differential Equations (SDEs).

$$d\mathbf{x} = f(\mathbf{x}, t)dt + g(t)d\overline{\mathbf{w}}$$

where $f(x, t)$ and $g(t)$ are handpicked and $d\overline{\mathbf{w}}$ can be viewed as infinitesimal white noise

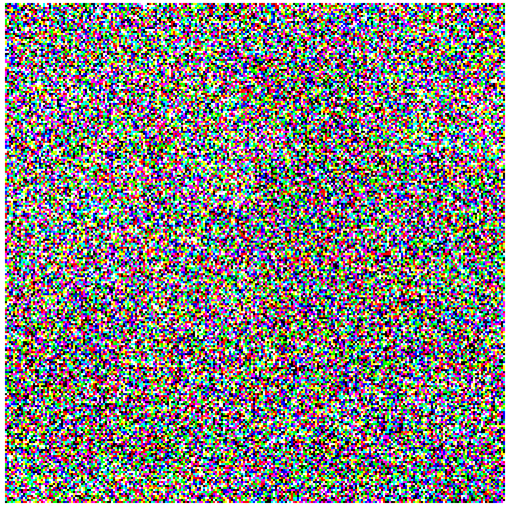
SGM - SDE

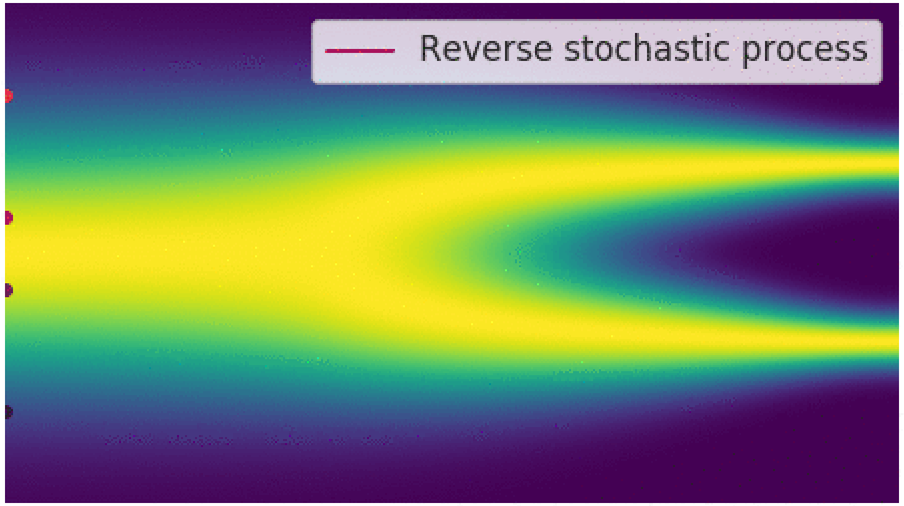
Forward process [1]



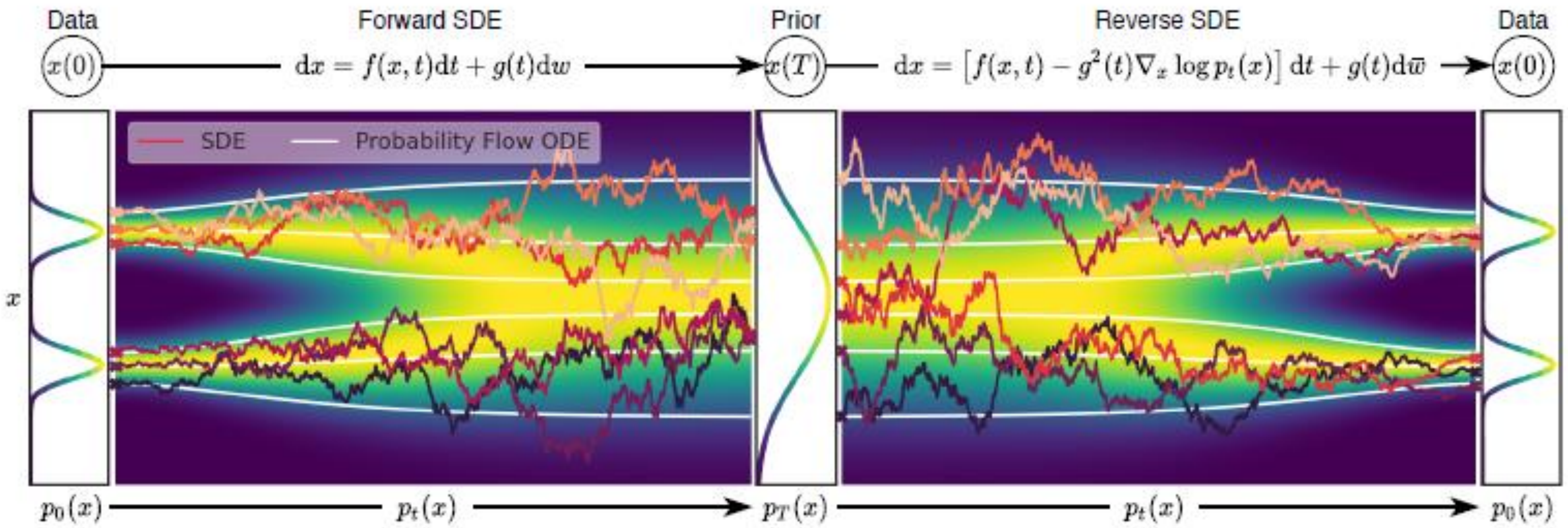
SGM - SDE

Reverse process [1]





SGM - SDE



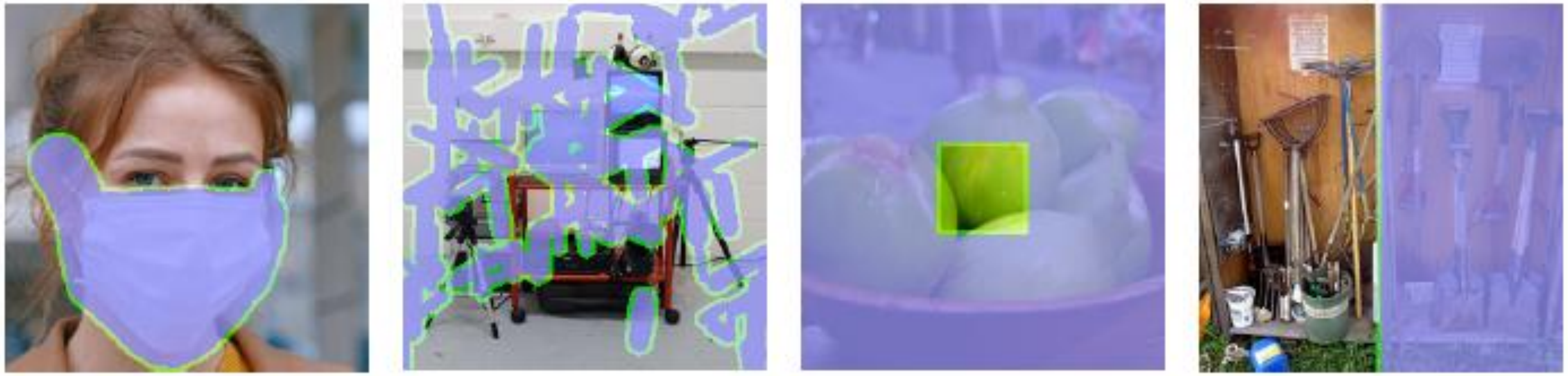
Data
Forward SDE
Prior
Reverse SDE
Data
$x(0)$
$\mathrm{d}x = f(x,t)\mathrm{d}t + g(t)\mathrm{d}w$
$x(T)$
$\mathrm{d}x = \left[f(x,t) - g^2(t)\nabla_x \log p_t(x)\right]\mathrm{d}t + g(t)\mathrm{d}\bar{w}$
$x(0)$
SDE
Probability Flow ODE
x
$p_0(x)$
$p_t(x)$
$p_T(x)$
$p_t(x)$
$p_0(x)$

Diffusion Models

- Generative Modeling
- Diffusion Models Methods
- **Diffusion Models Applications on CV**
- Future Directions

RePaint

RePaint is a method that was developed to solve inpainting tasks



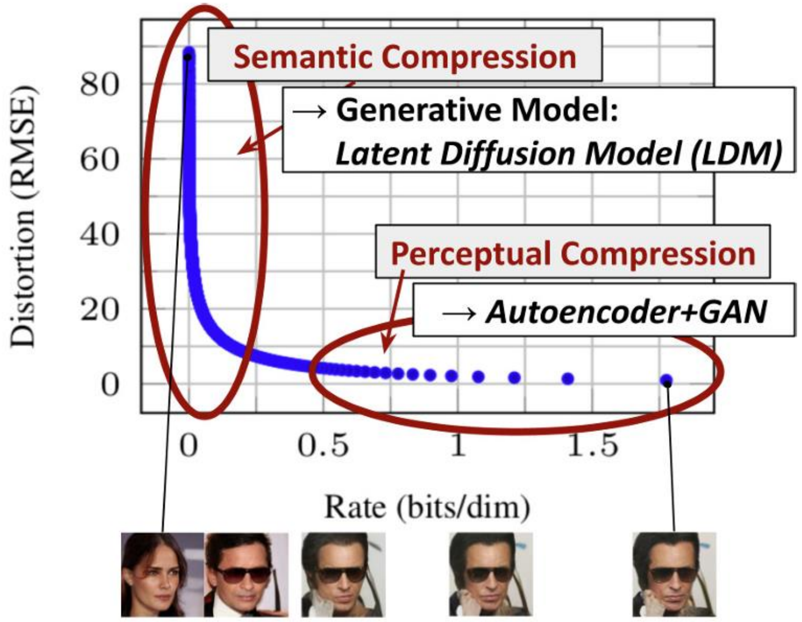
Input

It tries to fill missing parts of an image using an unconditionally trained DDPM.

Latent Diffusion Models

Scaling diffusion models to higher dimensions is hard. It is computationally infeasible and needs many more data than other methods (e.g. GANs)

Latent Diffusion Models propose to train the DM on the latent space of a pretrained Autoencoder. This allows the DM to learn only the crucial parts of an image, such as the semantic and conceptual composition, while the refinement is left for the Autoencoder



Tradeoff between compression rate and distortion [3]

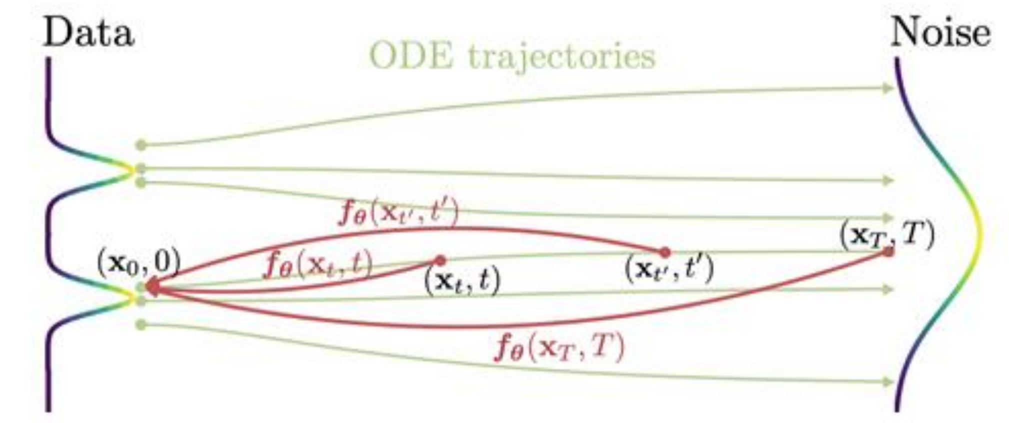
Diffusion Models

- Generative Modeling
- Diffusion Models Methods
- Diffusion Models Applications on CV
- **Future Directions**

Future Directions

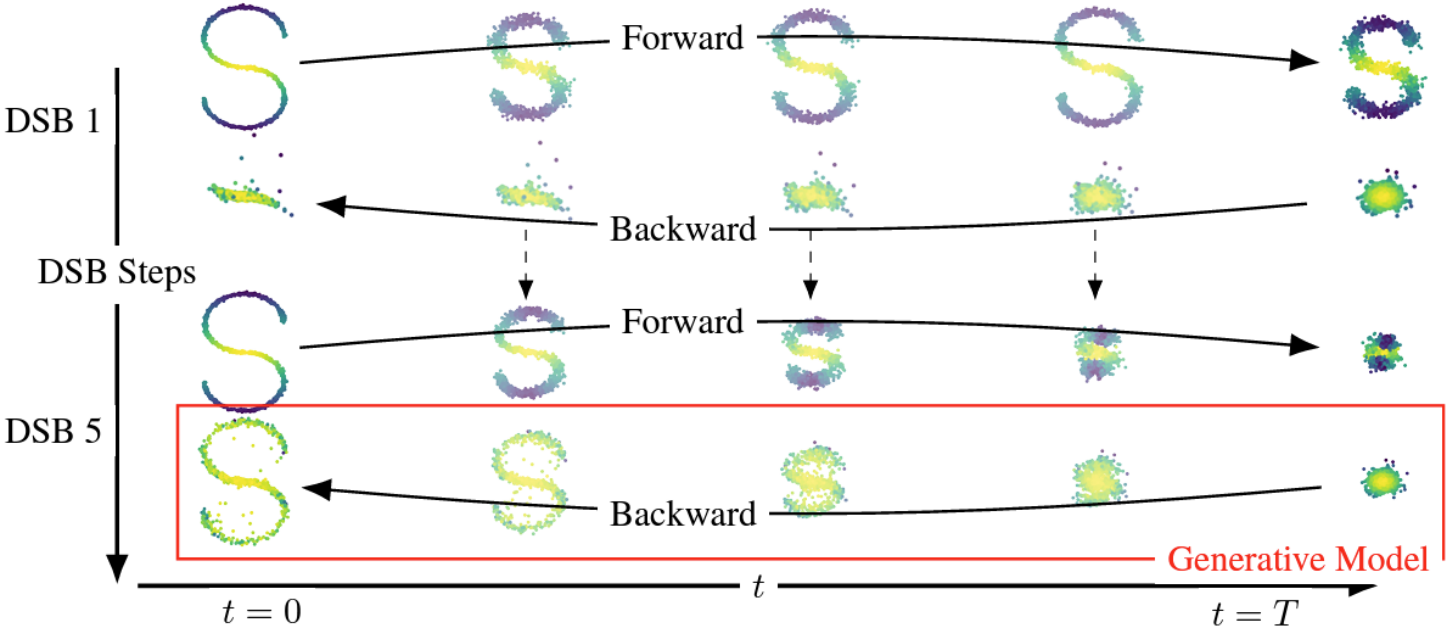
The main challenge with diffusion models is their slow inference time. Some proposed methods that try to tackle this problem involve

1. Consistency Models. These are models based on DMs, that try to learn intermediate points of the probability flow ODE.



Future Directions

2. Schrodinger bridges. Using a handcrafted PDE imposes strict restrictions on the process of moving from the data's distribution to the prior, and vice-versa. Methods that use Schrodinger bridges theory try to connect the two distributions in an optimal manner. This can reduce the necessary diffusion time, producing better results for fixed time T.



Bibliography

[1] I. Pitas, "Artificial Intelligence Science and Society Part A: Introduction to AI Science and Information Technology", Amazon/Kindle Direct Publishing, 2022,

https://www.amazon.com/dp/9609156460?ref_=pe_3052080_397514860

[2] I. Pitas, "Artificial Intelligence Science and Society Part B: AI Science, Mind and Humans", Amazon/Kindle Direct Publishing, 2022, https://www.amazon.com/dp/9609156479?ref_=pe_3052080_397514860

[3] I. Pitas, "Artificial Intelligence Science and Society Part C: AI Science and Society", Amazon/Kindle Direct Publishing, 2022, https://www.amazon.com/dp/9609156487?ref_=pe_3052080_397514860

[4] I. Pitas, "Artificial Intelligence Science and Society Part D: AI Science and the Environment", Amazon/Kindle Direct Publishing, 2022,

https://www.amazon.com/dp/9609156495?ref_=pe_3052080_397514860

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