Denoising-Diffusion Models

another way to see my sample's representation.

CAS AML 2023

Dr. Mykhailo Vladymyrov

Introduction

- Denoising Diffusion Models (DDMs) are a class of generative models that have gained significant attention in recent years.
- They play a crucial role in the field of generative modeling, offering a unique approach to image generation and data synthesis.
- DDMs are composed of two fundamental components: the <u>diffusion</u> process and the <u>generation</u> process. These components work together to create high-quality and realistic samples.

Diffusion Process

- The diffusion process is a fundamental component of Denoising Diffusion Models (DDMs). It serves as the foundation for generating high-quality samples.
- In the diffusion process, noise is incrementally added to an input image. This step-by-step addition of noise gradually transforms the clean image into a noisy version.

Diffusion Process

sequence of noisy sample

gaussian noise

as we're adding gaussian noise we can always reconstruct sample from prev. step

$$x_{t+1} = \sqrt{\beta} x_t + \epsilon_t \,, \epsilon_t \in N(0, \beta)$$

$$x_t = \alpha_t \, x_0 + \epsilon, \epsilon \in N(0, \sigma_t)$$

$$\alpha_t = \sqrt{\prod_{i=1..t} (1 - \beta_i)}$$

$$\sigma_t = \sqrt{1 - \alpha_t^2}$$

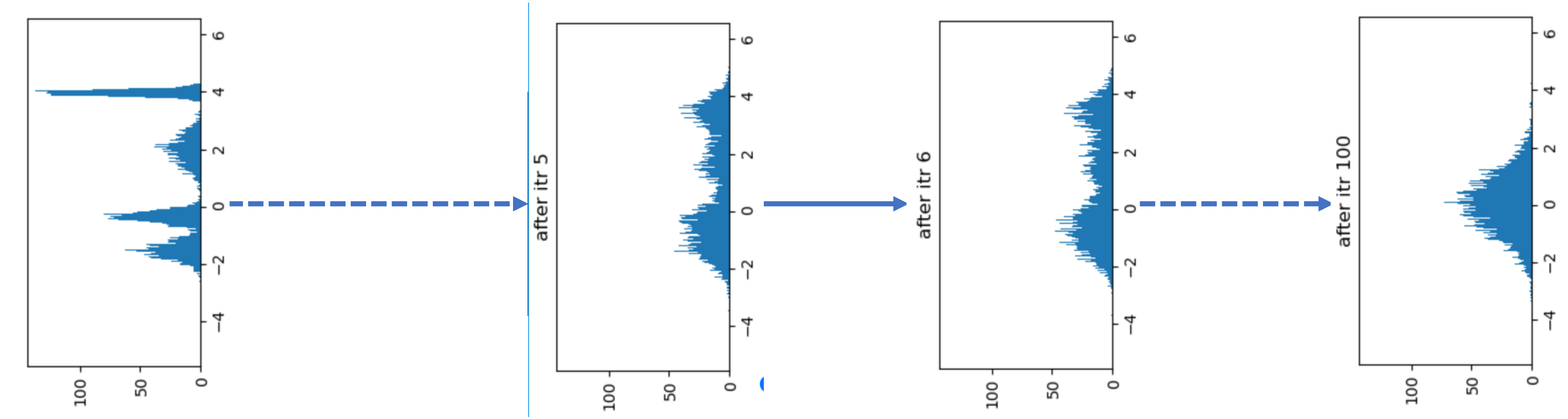
each sample is rescale of the previous sample + noise $\beta = \text{std}$

Diffusion Process

each sample is one point

$$x_{t+1} = \sqrt{\beta} x_t + \epsilon_t \,, \epsilon_t \in N(0, \beta)$$

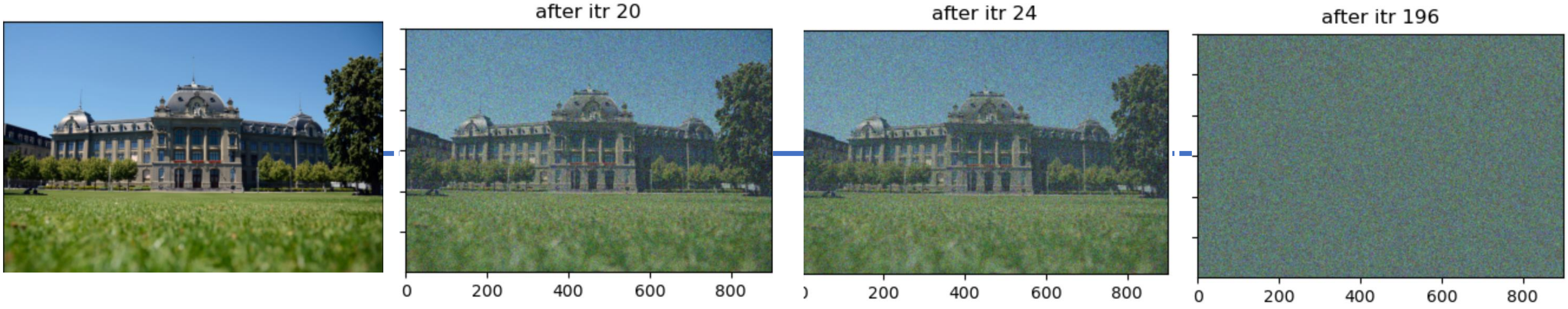
going on like that each sample will become normally distributed



Diffusion Process

this is what happened from the slide before visually

$$x_{t+1} = \sqrt{\beta} x_t + \epsilon_t \, , \epsilon_t \in N(0, \beta)$$



Which noise was added to get this noisy image?

Basically we're trying to train a model that from noise can reconstruct the original image.

Denoising Model

The denoising model is then trained to predict the added noise on each step.

random sample

estimation from my model

$$\min_{\boldsymbol\theta} \mathbb{E}_{t\sim U\{1,T\},\mathbf{x}_0\sim p(\mathbf{x}_0),\boldsymbol\epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})}\left[w(t)||\boldsymbol\epsilon-\boldsymbol\epsilon_{\boldsymbol\theta}(\alpha_t\mathbf{x}_0+\sigma_t\boldsymbol\epsilon,t)||_2^2\right], \quad w(t)=\frac{\beta_t^2}{2\rho_t^2(1-\beta_t)(1-\alpha_t^2)},$$

random t

random number

t steps of diffusion

samples close to my original samples

this is repeated many times

$$\mathbf{x}_{t-1}=\frac{1}{\sqrt{1-\beta_t}}\left(\mathbf{x}_t-\frac{\beta_t}{\sqrt{1-\alpha_t^2}}\boldsymbol\epsilon_{\boldsymbol\theta}(\mathbf{x}_t,t)\right)+\rho_t\boldsymbol\eta,$$

rescaled

subtract what the model predicted

Generation Process

- In DDMs, the generation process is where high-quality samples are created from noisy images generated during the diffusion process.
- The generation process involves a generative network that transforms noisy data into realistic samples.
- The processed is sequentially predicting the noise to be subtracted at each denoising iteration

Generation Process

latent space of a VAE → sampling → denoising

(not in the image space but in the latent space)

- **Generative Network:** A neural network, often a deep neural network or a variational autoencoder (VAE), is employed to map noisy images to high-quality samples.
- **Gradual Improvement:** The generative network progressively refines the noisy images, reducing noise and enhancing details at each step.
- **Sampling:** During training, the generative network samples from a noise distribution to generate various versions of the same image.

* To where the sample will be moved to that original position (yellow)

EXAMPLE

each bed have coordinates

we shift each model (bed) until when we have noise only

we train the model to predict the red line (from noise to go back to original position)

what the model basically learn is where to denoise (move) the samples so that they assume a position similar to the original position.

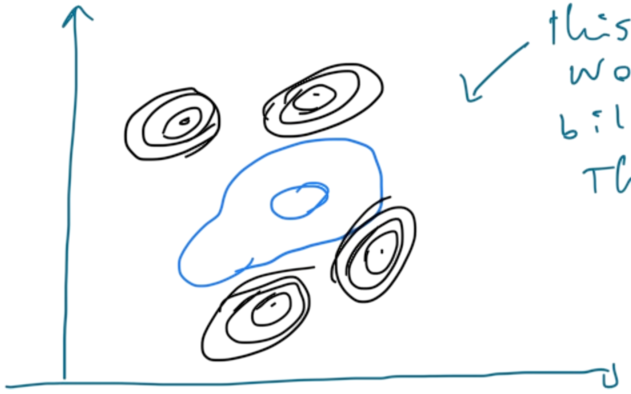
what the model does (simplified) is to assign area *

Generation Process

the model does not predict which random no did you made up, but which random no. did you made up to add noise to that specific sample

- **Generative Network:** A neural network, often a deep neural network or a variational autoencoder (VAE), is employed to map noisy images to high-quality samples.
- **Gradual Improvement:** The generative network progressively refines the noisy images, reducing noise and enhancing details at each step.
- **Sampling:** During training, the generative network samples from a noise distribution to generate various versions of the same image.

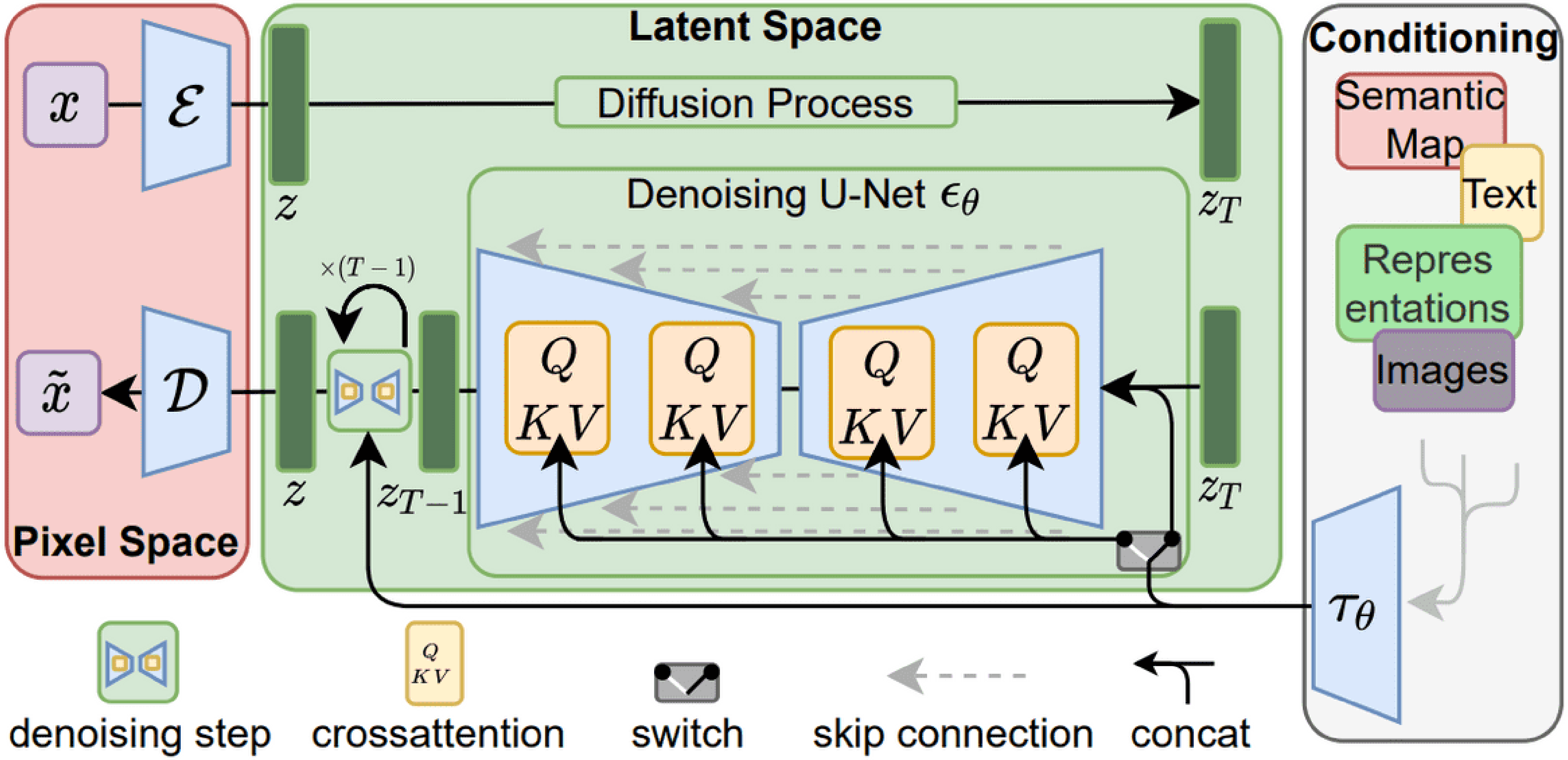
EXAMPLE (continuation):



this is how the model would see 流: as a probability to find the sample there

when the model is sampling, it is doing it randomly: we can though condition the sampling.

Generation Process



Rombach et al

$\underbrace{\text{text} \rightarrow \text{embedded}}_{\text{prompt}}$

⇓

depending on the prompt the model will learn where should it go.

matematically speaking, this field is a gradient of likelihood

Conditioned on time, prompt, etc

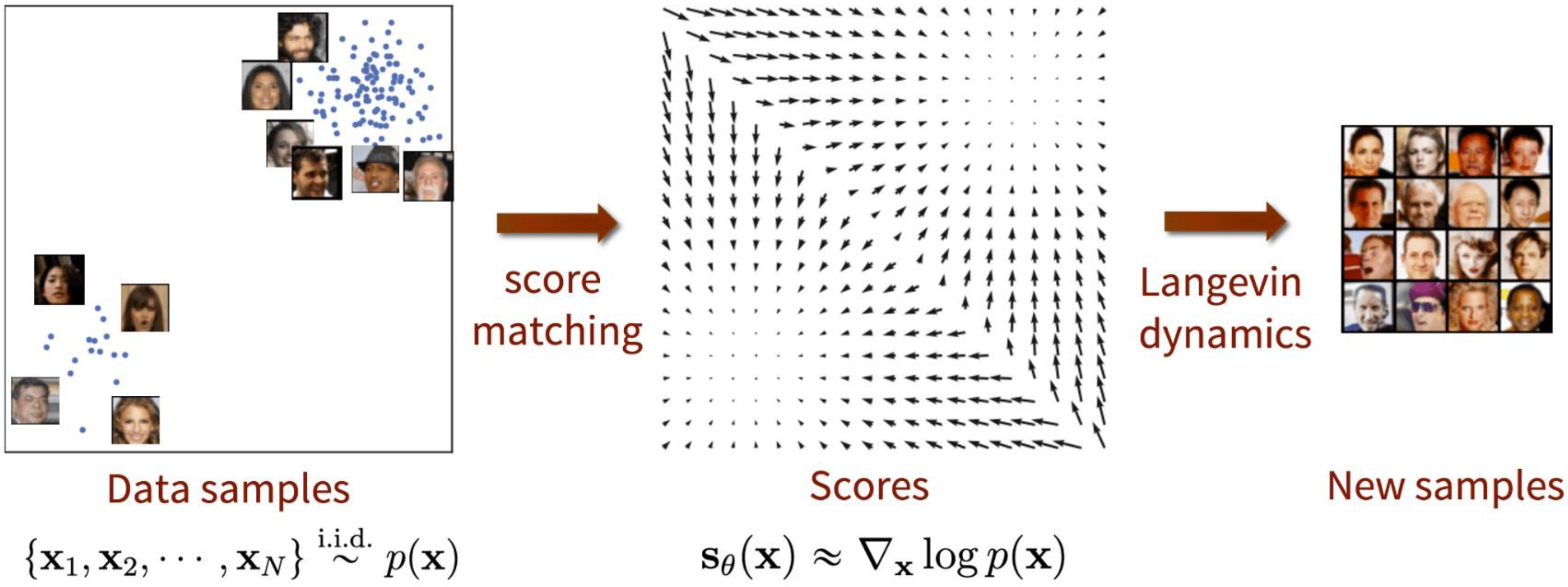
The good thing of this whole process is that we don't need labelled data.

Challenges in Training DDMs

- **Computational Complexity:** DDMs can be computationally intensive, requiring substantial resources for training and inference. ← inference itself is computationally intensive
- **Hyperparameter Tuning:** Properly configuring hyperparameters is crucial for DDMs, and finding the right settings can be challenging.
- **Data Size:** DDMs often require large datasets for effective training, which may not be readily available in some domains.

Advances in DDMs

- **Improved Sampling Methods:** Recent research has introduced more efficient sampling techniques, reducing the computational burden of DDMs.
- **Transfer Learning:** Applying pre-trained DDMs to new tasks or domains has become more accessible, thanks to advances in transfer learning.
- **Scalability:** Efforts to make DDMs more scalable have led to the development of smaller, more efficient models suitable for various applications.



Score-based generative modeling with score matching + Langevin dynamics. Source: Generative Modeling by Estimating Gradients of the Data Distribution