

Deep Generative Models

Lecture 14

Roman Isachenko

Moscow Institute of Physics and Technology
Yandex School of Data Analysis

2024, Autumn

Outline

1. Conditional flow matching

Linear interpolation

Rectified flows

2. Latent models

Score-based models

Autoregressive models

3. The worst course overview

Outline

1. Conditional flow matching

Linear interpolation

Rectified flows

2. Latent models

Score-based models

Autoregressive models

3. The worst course overview

Outline

1. Conditional flow matching

Linear interpolation

Rectified flows

2. Latent models

Score-based models

Autoregressive models

3. The worst course overview

Flow Matching

Conditional Flow Matching

$$\mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})} \mathbb{E}_{\mathbf{x} \sim p_t(\mathbf{x}|\mathbf{z})} \|\mathbf{f}(\mathbf{x}, \mathbf{z}, t) - \mathbf{f}_\theta(\mathbf{x}, t)\|^2 \rightarrow \min_\theta$$

Let choose $\mathbf{z} = (\mathbf{x}_0, \mathbf{x}_1)$. Then $p(\mathbf{z}) = p_0(\mathbf{x}_0)p_1(\mathbf{x}_1)$.

$$p_t(\mathbf{x}) = \int p_t(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) p_0(\mathbf{x}_0) p_1(\mathbf{x}_1) d\mathbf{x}_0 d\mathbf{x}_1$$

We need to ensure boundary conditions:

$$\begin{cases} p_0(\mathbf{x}) = p(\mathbf{x}) = \mathcal{N}(0, \mathbf{I}); \\ p_1(\mathbf{x}) = \pi(\mathbf{x}). \end{cases} \Rightarrow \begin{cases} p_0(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_0); \\ p_1(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_1). \end{cases}$$

Gaussian conditional probability path

$$p_t(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \mathcal{N}(\boldsymbol{\mu}_t(\mathbf{x}_0, \mathbf{x}_1), \boldsymbol{\sigma}_t^2(\mathbf{x}_0, \mathbf{x}_1)); \quad \mathbf{x}_t = \boldsymbol{\mu}_t(\mathbf{x}_0, \mathbf{x}_1) + \boldsymbol{\sigma}_t^2(\mathbf{x}_0, \mathbf{x}_1) \odot \mathbf{x}_0$$

Flow Matching

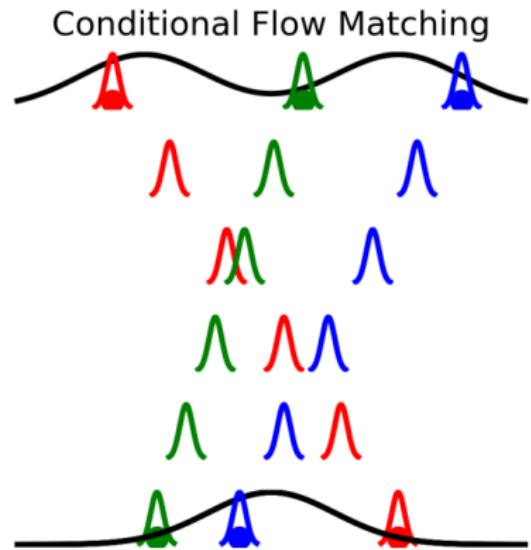
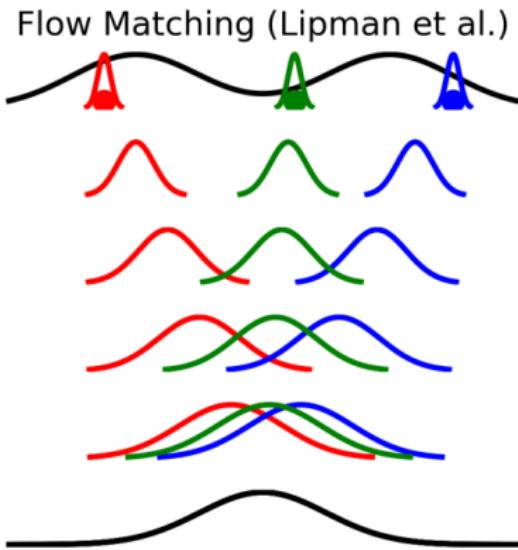
$$\begin{cases} p_0(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_0); \\ p_1(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_1). \end{cases} \Rightarrow \begin{cases} \mu_0(\mathbf{x}_0, \mathbf{x}_1) = \mathbf{x}_0, & \sigma_0(\mathbf{x}_0, \mathbf{x}_1) = 0 \\ \mu_1(\mathbf{x}_0, \mathbf{x}_1) = \mathbf{x}_1, & \sigma_1(\mathbf{x}_0, \mathbf{x}_1) = 0 \end{cases}$$

Let consider straight conditional paths

$$\begin{cases} \mu_t(\mathbf{x}_0, \mathbf{x}_1) = t\mathbf{x}_1 + (1 - t)\mathbf{x}_0; \\ \sigma_t(\mathbf{x}_0, \mathbf{x}_1) = 0. \end{cases} \Rightarrow \begin{cases} p_t(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \mathcal{N}(t\mathbf{x}_1 + (1 - t)\mathbf{x}_0, 0) \\ \mathbf{x}_t = t\mathbf{x}_1 + (1 - t)\mathbf{x}_0. \end{cases}$$

$$\frac{d\mathbf{x}}{dt} = \mathbf{x}_1 - \mathbf{x}_0.$$

Flow Matching



Outline

1. Conditional flow matching

Linear interpolation

Rectified flows

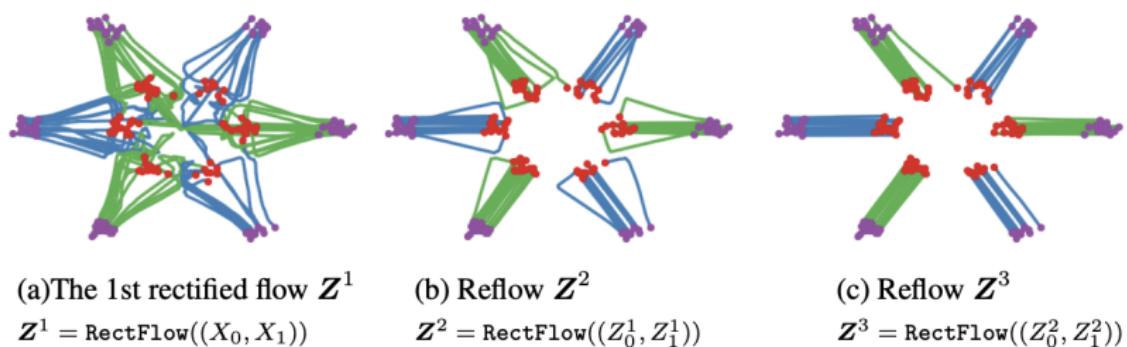
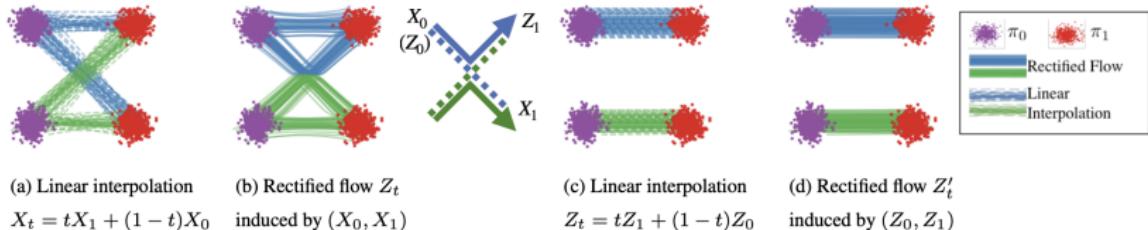
2. Latent models

Score-based models

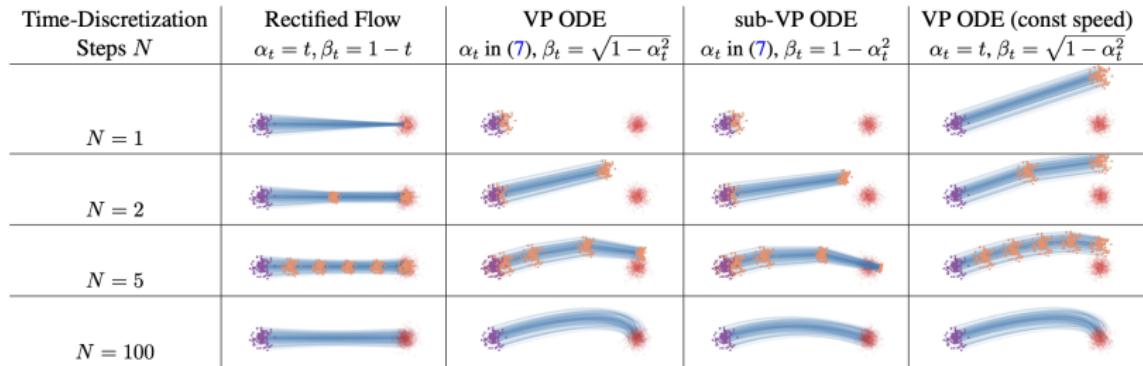
Autoregressive models

3. The worst course overview

Rectified flows



Rectified flows



Rectified flows



Outline

1. Conditional flow matching

Linear interpolation

Rectified flows

2. Latent models

Score-based models

Autoregressive models

3. The worst course overview

Outline

1. Conditional flow matching

Linear interpolation

Rectified flows

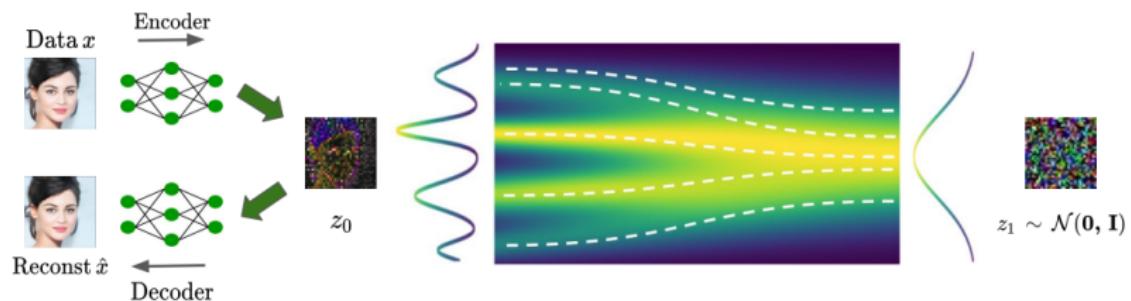
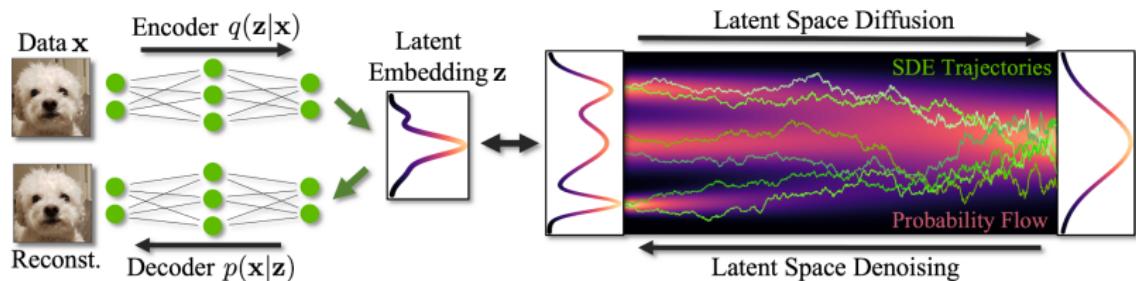
2. Latent models

Score-based models

Autoregressive models

3. The worst course overview

Latent models



Dao Q. et al. *Flow Matching in Latent Space*, 2023

NeurIPS 2023 Tutorial: Latent Diffusion Models: Is the Generative AI Revolution Happening in Latent Space?

Outline

1. Conditional flow matching

Linear interpolation

Rectified flows

2. Latent models

Score-based models

Autoregressive models

3. The worst course overview

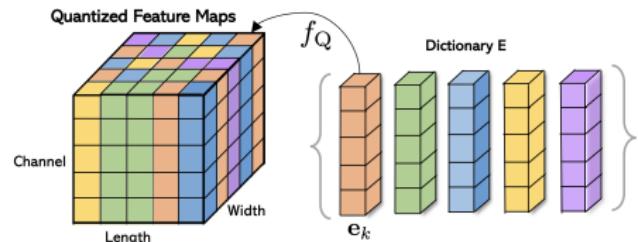
Vector Quantized VAE (VQ-VAE)

Vector quantization

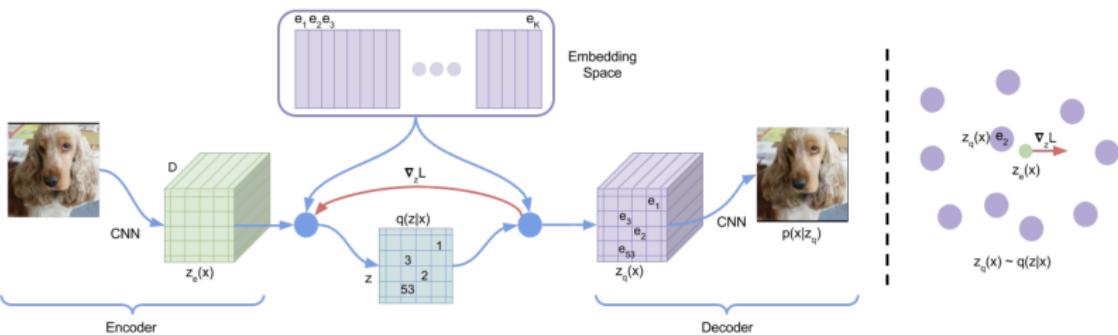
Define the dictionary space $\{\mathbf{e}_k\}_{k=1}^K$, where $\mathbf{e}_k \in \mathbb{R}^C$, K is the size of the dictionary.

$$\mathbf{z}_q = \mathbf{q}(\mathbf{z}) = \mathbf{e}_{k^*}$$

$$\text{Here } k^* = \arg \min_k \|\mathbf{z} - \mathbf{e}_k\|.$$



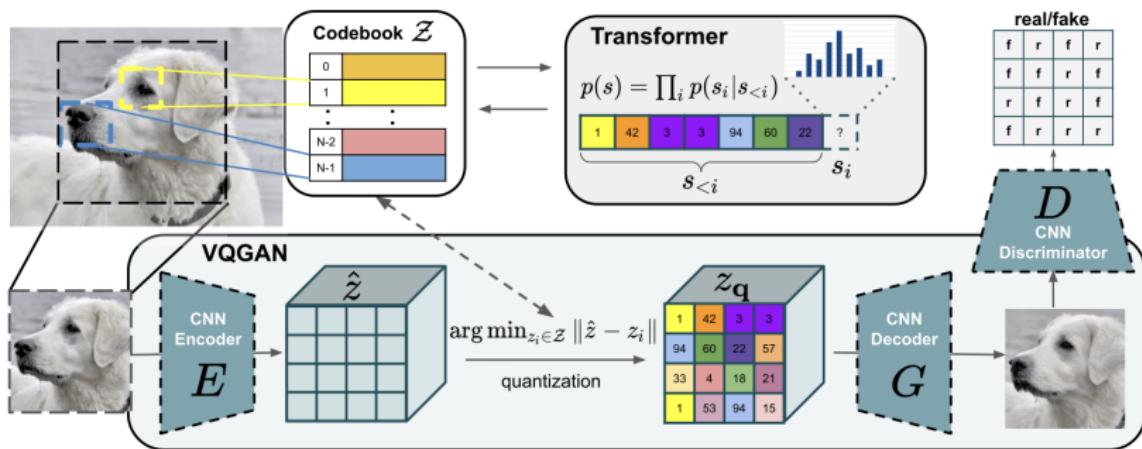
$$\mathcal{L}_{\phi, \theta}(\mathbf{x}) = \log p(\mathbf{x} | \mathbf{z}_q, \theta) - \log K$$



Zhao Y. et al. Feature Quantization Improves GAN Training, 2020

Oord A., Vinyals O., Kavukcuoglu K. Neural Discrete Representation Learning, 2017

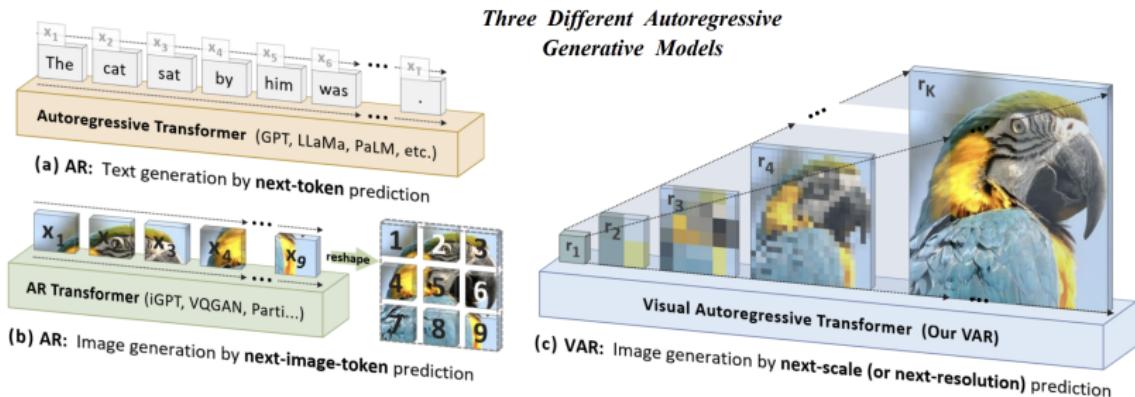
Vector Quantized GAN



LlamaGen

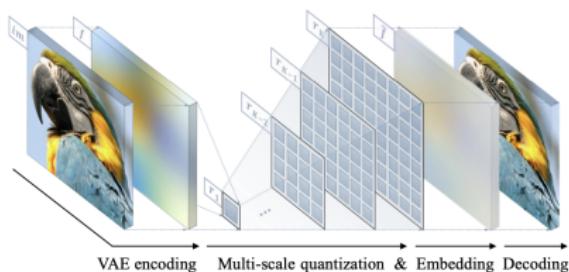


Visual Autoregressive Modeling (VAR)

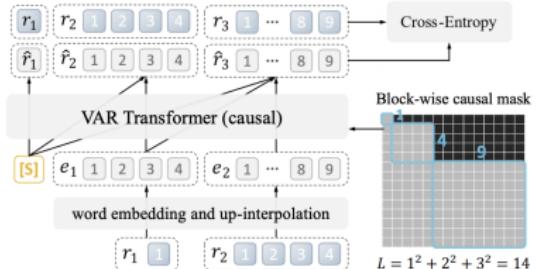


Visual Autoregressive Modeling (VAR)

Stage 1: Training multi-scale VQVAE on images
(to provide the ground truth for training Stage 2)



Stage 2: Training VAR transformer on tokens
($[S]$ means a start token with condition information)



Outline

1. Conditional flow matching

Linear interpolation

Rectified flows

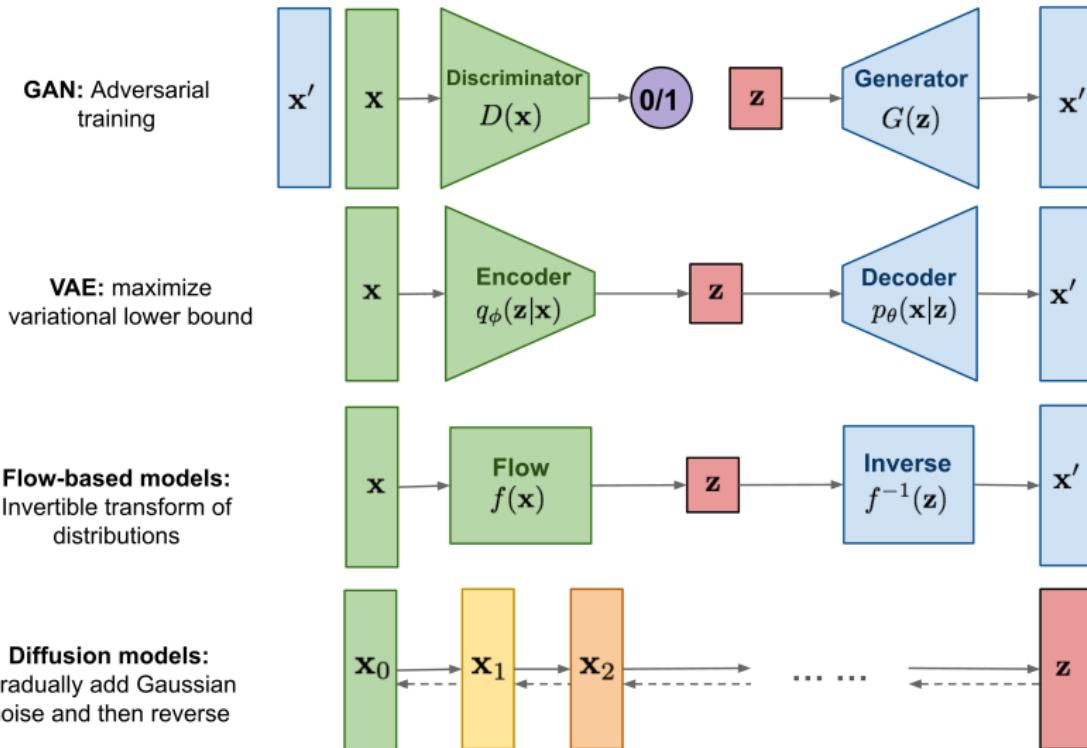
2. Latent models

Score-based models

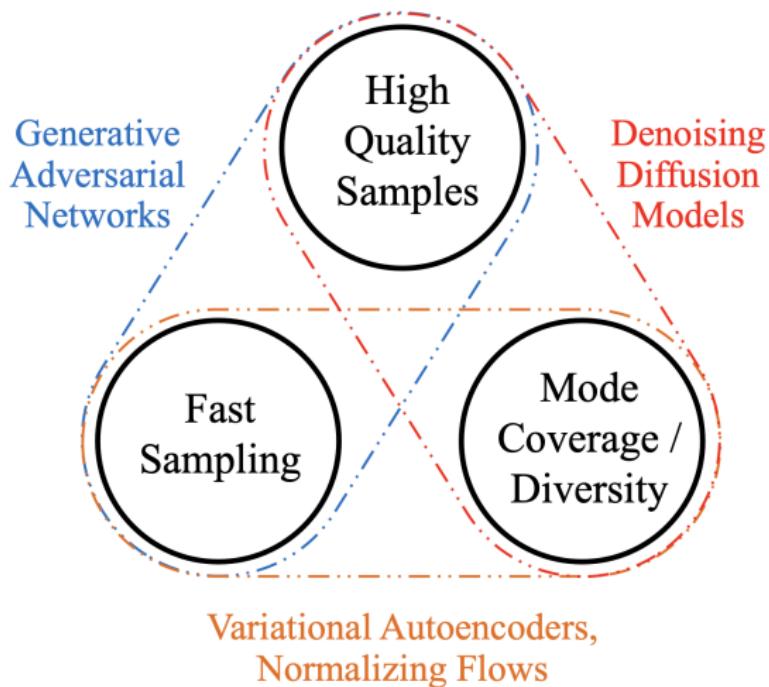
Autoregressive models

3. The worst course overview

The worst course overview :)



The worst course overview :)



The worst course overview :)

Model	Efficient	Sample quality	Coverage	Well-behaved latent space	Disentangled latent space	Efficient likelihood
GANs	✓	✓	✗	✓	?	n/a
VAEs	✓	✗	?	✓	?	✗
Flows	✓	✗	?	✓	?	✓
Diffusion	✗	✓	?	✗	✗	✗

Summary

