

# Deep Generative Models

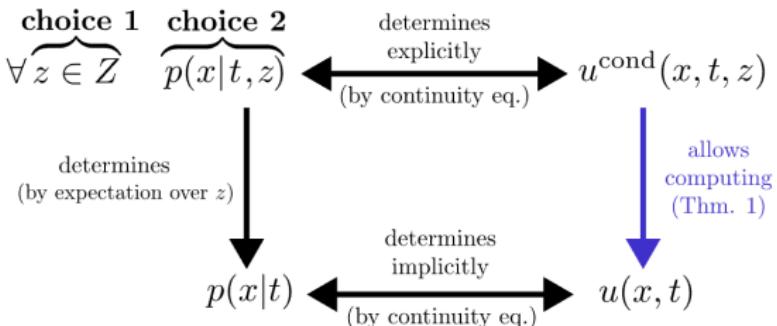
## Lecture 13

Roman Isachenko

Moscow Institute of Physics and Technology  
Yandex School of Data Analysis

2025, Autumn

# Recap of Previous Lecture



## Constraints

$$p(\mathbf{x}) = \mathcal{N}(0, \mathbf{I}) = \mathbb{E}_{p(\mathbf{z})} p_0(\mathbf{x}|\mathbf{z}); \quad p_{\text{data}}(\mathbf{x}) = \mathbb{E}_{p(\mathbf{z})} p_1(\mathbf{x}|\mathbf{z}).$$

- ▶ How should we choose the conditioning latent variable  $\mathbf{z}$ ?
- ▶ How can we define  $p_t(\mathbf{x}|\mathbf{z})$  so that it meets the constraints?

## Gaussian Conditional Probability Path

$$p_t(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\boldsymbol{\mu}_t(\mathbf{z}), \boldsymbol{\sigma}_t^2(\mathbf{z}))$$

$$\mathbf{x}_t = \boldsymbol{\mu}_t(\mathbf{z}) + \boldsymbol{\sigma}_t(\mathbf{z}) \odot \mathbf{x}_0, \quad \mathbf{x}_0 \sim p_0(\mathbf{x}) = \mathcal{N}(0, \mathbf{I})$$

## Recap of Previous Lecture

### Gaussian Conditional Probability Path

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

$$\mathbf{f}(\mathbf{x}, \mathbf{z}, t) = \boldsymbol{\mu}'_t(\mathbf{z}) + \frac{\boldsymbol{\sigma}'_t(\mathbf{z})}{\boldsymbol{\sigma}_t(\mathbf{z})} \odot (\mathbf{x} - \boldsymbol{\mu}_t(\mathbf{z}))$$

### Conditioning Latent Variable

Let's choose  $\mathbf{z} = \mathbf{x}_1$ . Then  $p(\mathbf{z}) = p_1(\mathbf{x}_1)$ .

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

We must ensure the boundary constraints:

$$\begin{cases} p(\mathbf{x}) = \mathbb{E}_{p(\mathbf{z})} p_0(\mathbf{x}|\mathbf{z}); (= \mathcal{N}(0, \mathbf{I})) \\ p_{\text{data}}(\mathbf{x}) = \mathbb{E}_{p(\mathbf{z})} p_1(\mathbf{x}|\mathbf{z}). \end{cases} \Rightarrow \begin{cases} p_0(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(0, \mathbf{I}); \\ p_1(\mathbf{x}|\mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_1). \end{cases}$$

## Recap of Previous Lecture

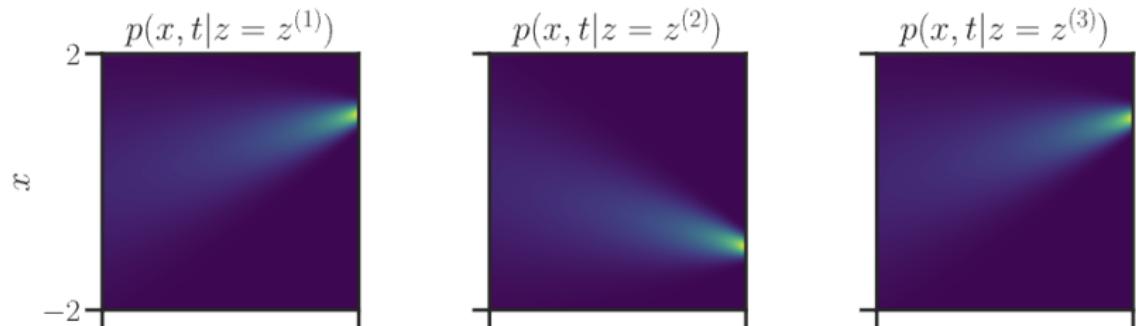
$$p_0(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(0, \mathbf{I}); \quad p_1(\mathbf{x}|\mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_1).$$

## Gaussian Conditional Probability Path

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

Let's consider straight conditional paths:

$$\begin{cases} \boldsymbol{\mu}_t(\mathbf{x}_1) = t\mathbf{x}_1; \\ \boldsymbol{\sigma}_t(\mathbf{x}_1) = 1 - t. \end{cases} \Rightarrow \begin{cases} p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(t\mathbf{x}_1, (1-t)^2\mathbf{I}); \\ \mathbf{x}_t = t\mathbf{x}_1 + (1-t)\mathbf{x}_0. \end{cases}$$



# Outline

## 1. Latent Space Models

Score-Based Models

Autoregressive Models

## 2. The Worst Course Overview

# Outline

## 1. Latent Space Models

Score-Based Models

Autoregressive Models

## 2. The Worst Course Overview

# Outline

## 1. Latent Space Models

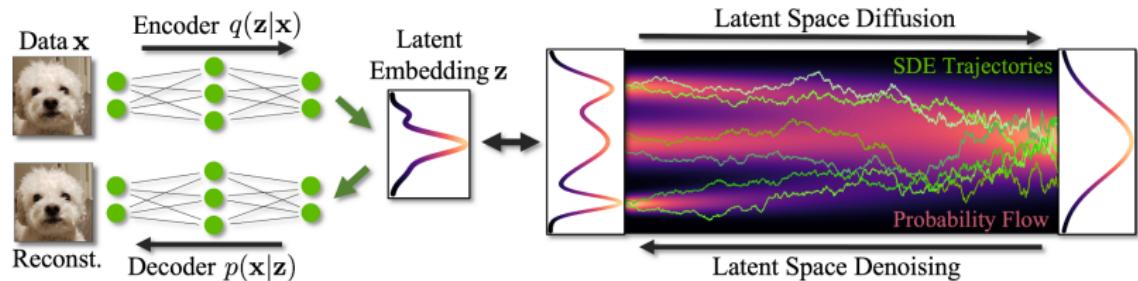
Score-Based Models

Autoregressive Models

## 2. The Worst Course Overview

# Latent Space Models

## Score-Based Models (Diffusion)

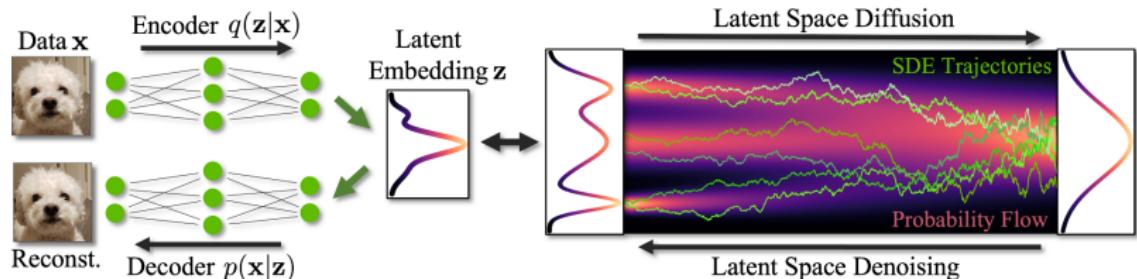


Dao Q. et al. *Flow Matching in Latent Space*, 2023

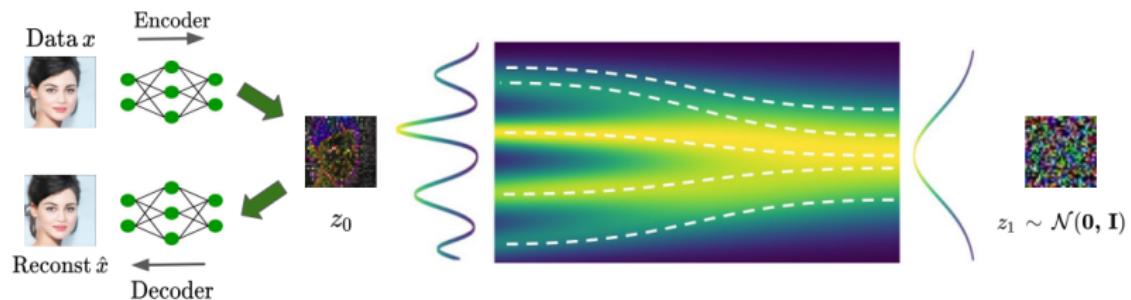
NeurIPS 2023 Tutorial: Latent Diffusion Models: Is the Generative AI Revolution Happening in Latent Space?

# Latent Space Models

## Score-Based Models (Diffusion)



## Flow Matching



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. Latent Space Models

Score-Based Models

Autoregressive Models

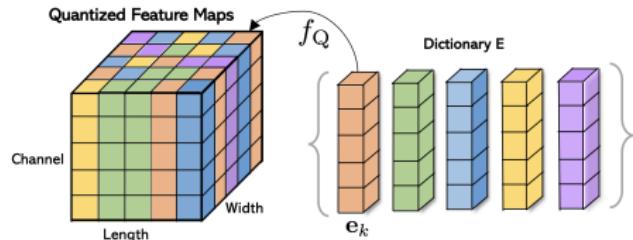
## 2. The Worst Course Overview

# Vector Quantized VAE (VQ-VAE)

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

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

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

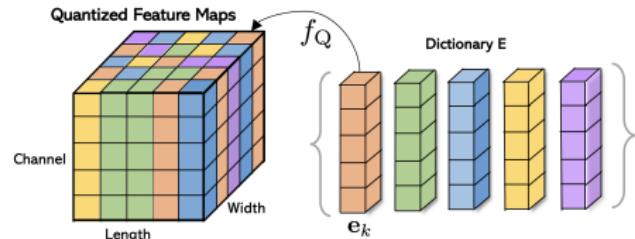


# Vector Quantized VAE (VQ-VAE)

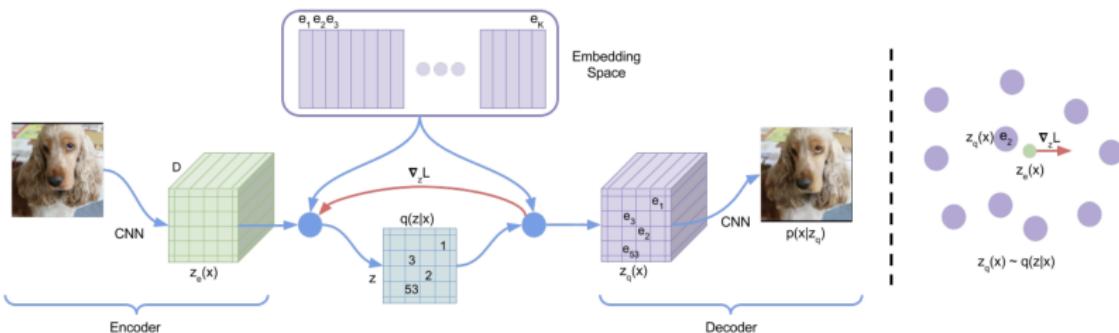
Define a dictionary space  $\{\mathbf{e}_k\}_{k=1}^K$ , where  $\mathbf{e}_k \in \mathbb{R}^C$  and  $K$  is the dictionary's size.

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

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



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

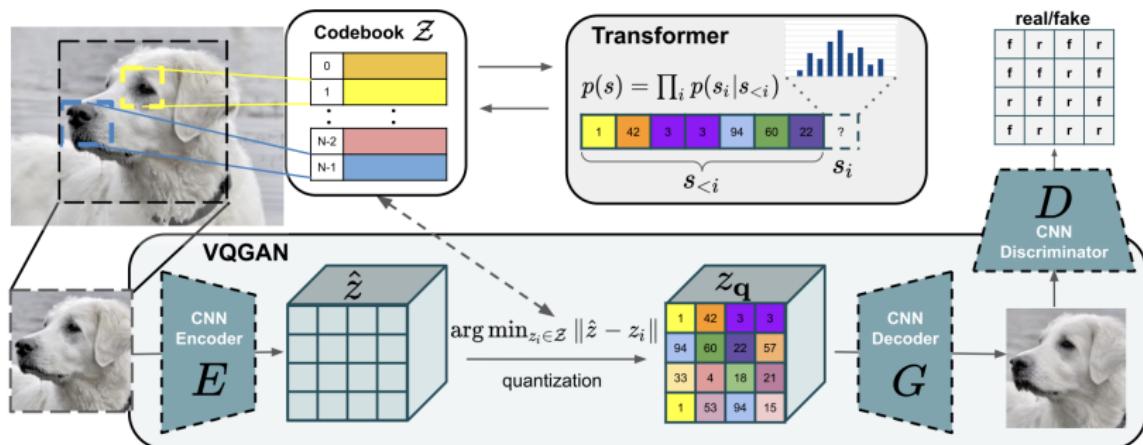


Zhao Y. et al. Feature Quantization Improves GAN Training, 2020

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

# Vector Quantized GAN

- ▶ We use a VQ-VAE model and its objective.
- ▶ We add an adversarial loss between generated and real images to further improve the visual quality of reconstructions.

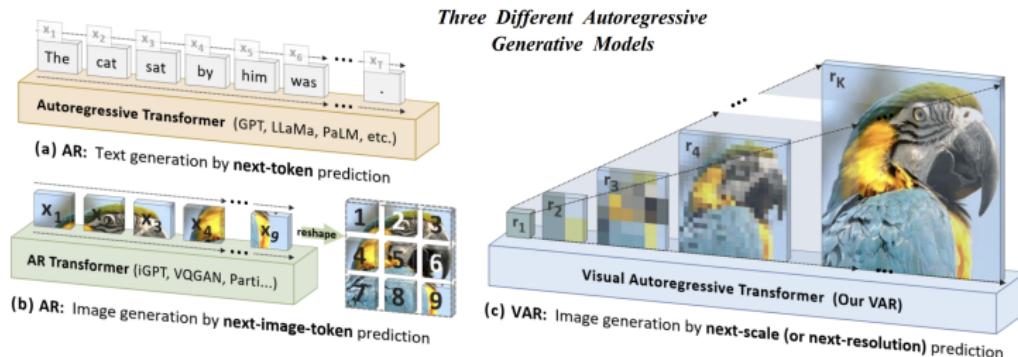


# LlamaGen: Pure Autoregression

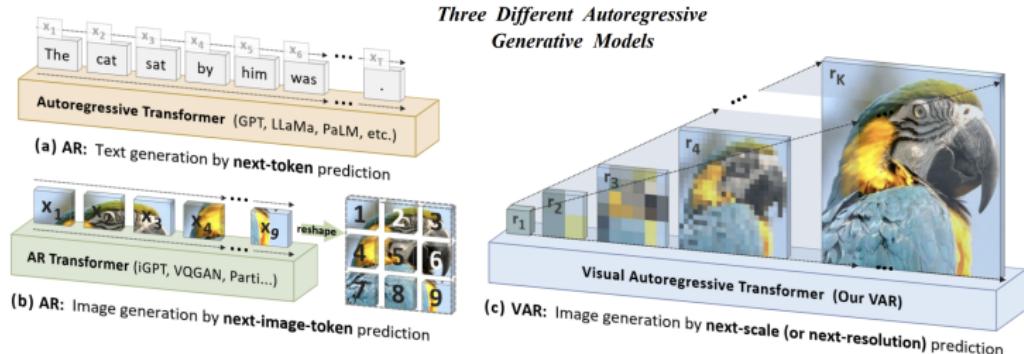
- ▶ Use a VQ-GAN encoder for mapping images into the discrete latent space (codebook vectors).
- ▶ Train a pure autoregressive model (Llama-based) in the latent space.
- ▶ Use the VQ-GAN decoder to map discrete tokens back to image space.



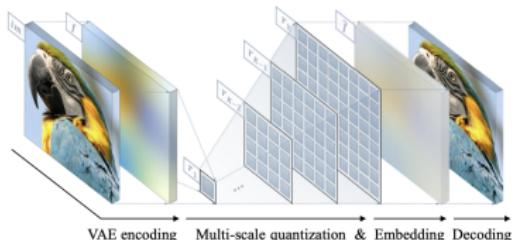
# Visual Autoregressive Modeling (VAR)



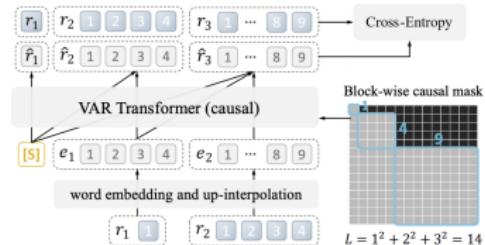
# Visual Autoregressive Modeling (VAR)



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



**Stage 2: Training VAR transformer on tokens**  
( $\$$  means a start token with condition information)



# Outline

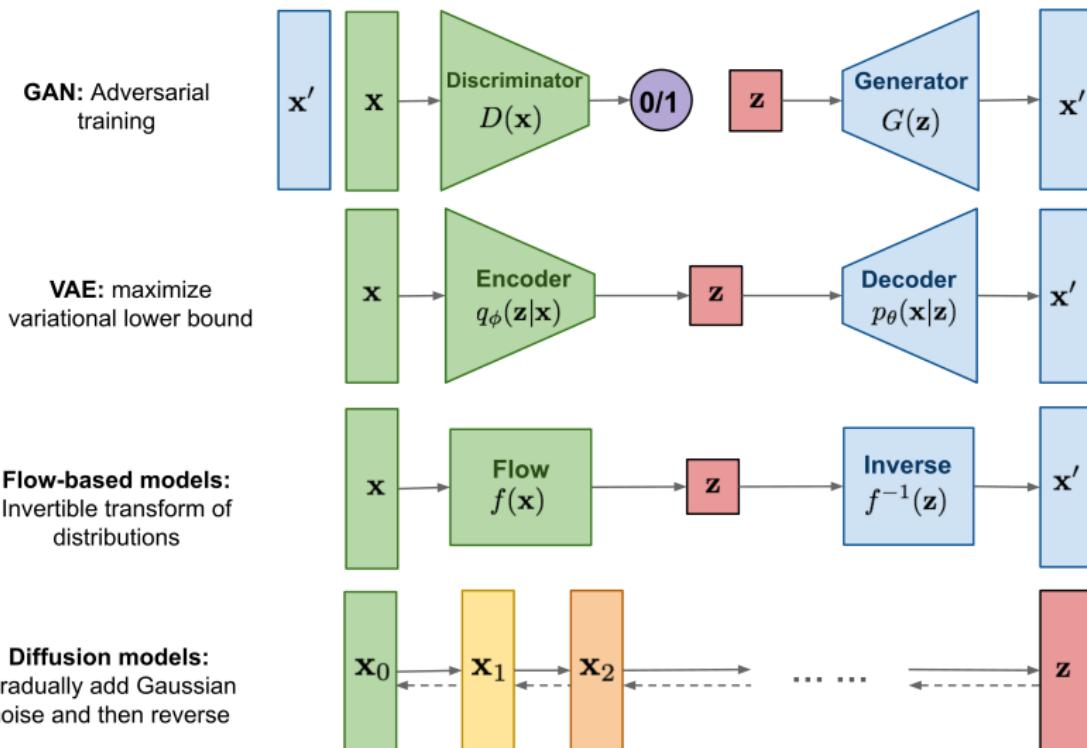
## 1. Latent Space Models

Score-Based Models

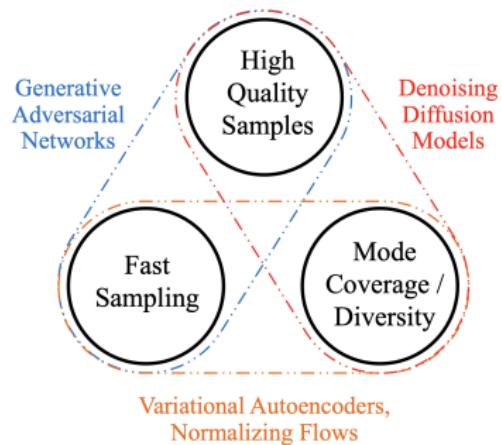
Autoregressive Models

## 2. The Worst Course Overview

# The Worst Course Overview :)



# The Worst Course Overview :)

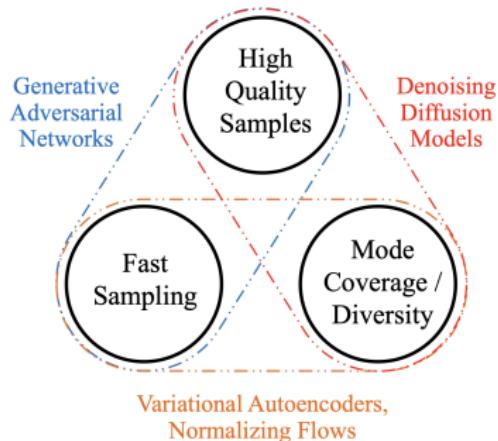


---

Xiao Z., Kreis K., Vahdat A. *Tackling the generative learning trilemma with denoising diffusion GANs*, 2021

Simon J.D. Prince. *Understanding Deep Learning*, 2023

# The Worst Course Overview :)



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

Xiao Z., Kreis K., Vahdat A. *Tackling the generative learning trilemma with denoising diffusion GANs*, 2021

Simon J.D. Prince. *Understanding Deep Learning*, 2023

## Summary

- ▶ Most state-of-the-art generative models are latent variable models with either continuous or discrete latent spaces.