Deep Generative Models

Lecture 14

Roman Isachenko

Moscow Institute of Physics and Technology
Yandex School of Data Analysis

2025, Autumn

Recap of Previous Lecture

Outline

Outline

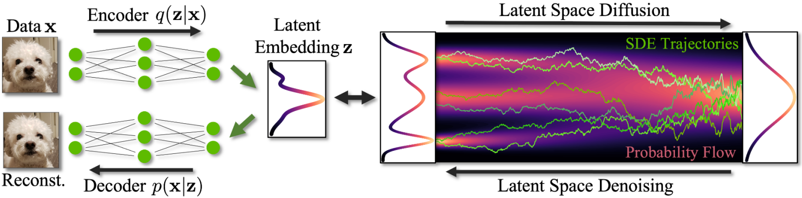
1. Latent Space Models
   - Score-Based Models
   - Autoregressive Models

2. The Worst Course Overview

Outline

Latent Space Models

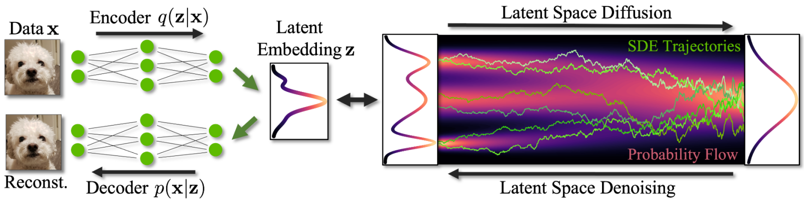
Score-Based Models (Diffusion)



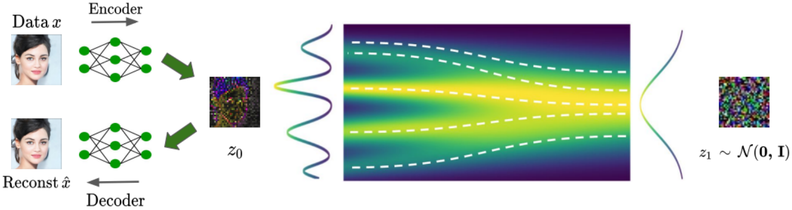
Dao Q. et al. Flow Matching in Latent Space, 2023
NeurIPS 2023 Tutorial: Latent Diffusion Models: Is the Generative AI Revolution Happening in Latent Space?

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?

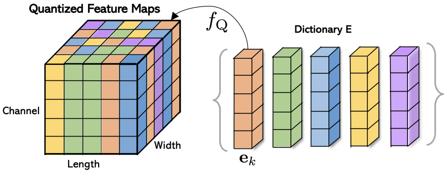
Outline

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.

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

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



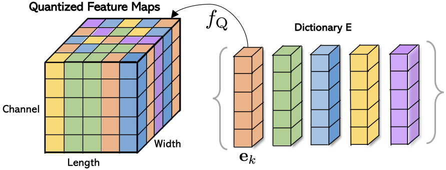
Zhao Y. et al. *Feature Quantization Improves GAN Training*, 2020
Oord A., Vinyals O., Kavukcuoglu K. Neural Discrete Representation Learning, 2017

Vector Quantized VAE (VQ-VAE)

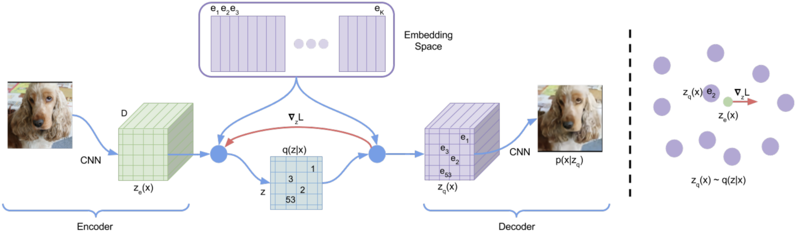
Define a dictionary space $\{\mathbf{e}_k\}_{k=1}^K$, where $\mathbf{e}_k \in \mathbb{R}^C$ and K is the dictionary's size.

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

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



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

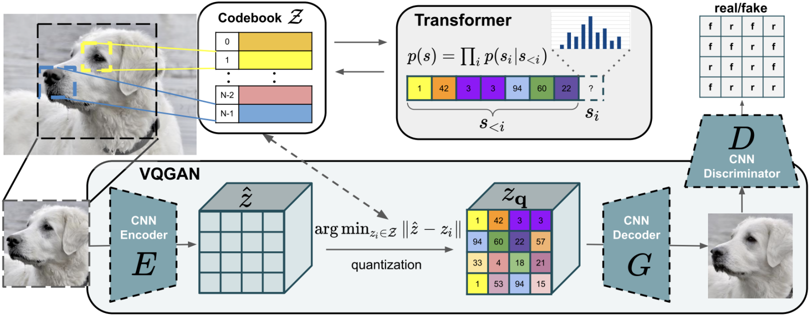


Zhao Y. et al. *Feature Quantization Improves GAN Training*, 2020

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.



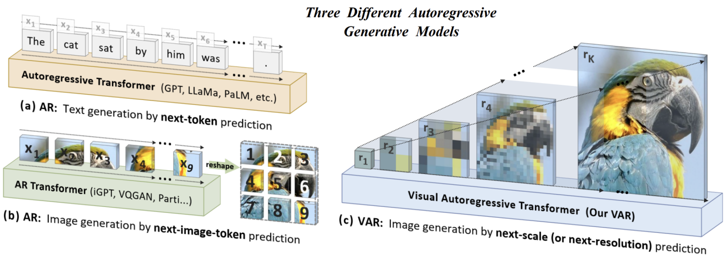
Esser P. et al. *Taming Transformers for High-Resolution Image Synthesis, 2020*

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.

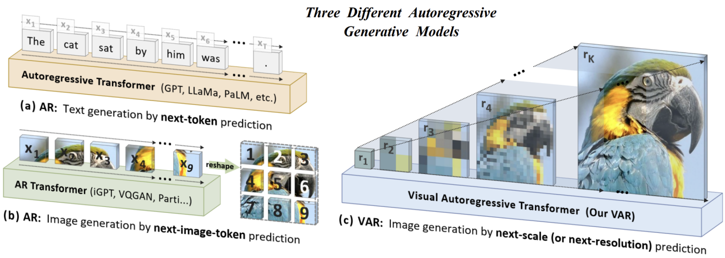
Sun P. et al. Autoregressive Model Beats Diffusion: Llama for Scalable Image Generation, 2024

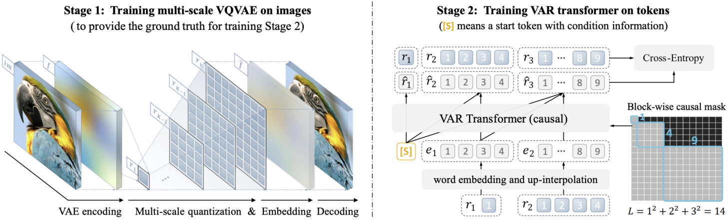
Visual Autoregressive Modeling (VAR)



Tean K. et al. Visual Autoregressive Modeling: Scalable Image Generation via Next-Scale Prediction, 2024

Visual Autoregressive Modeling (VAR)

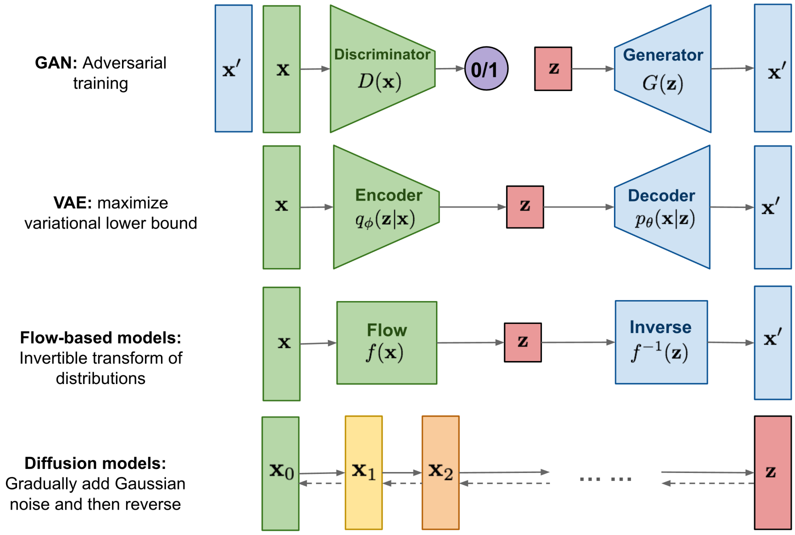




Tean K. et al. *Visual Autoregressive Modeling: Scalable Image Generation via Next-Scale Prediction*, 2024

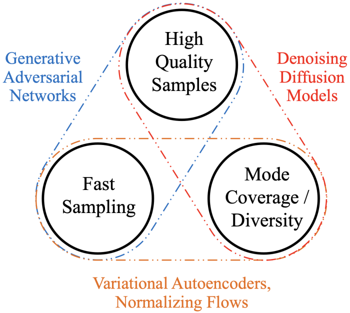
Outline

The Worst Course Overview :)



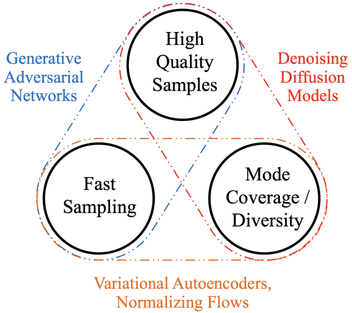
Weng L. What are Diffusion Models?, blog post, 2021

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.