

# 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

Conical gaussian paths (continued)

Linear interpolation

Link with score-based models

## 2. Latent space models

Score-based models

Autoregressive models

## 3. The worst course overview

## Recap of previous lecture

### Discrete-in-time objective

$$\mathbb{E}_{\pi(\mathbf{x}_0)} \mathbb{E}_{t \sim U\{1, T\}} \mathbb{E}_{q(\mathbf{x}_t | \mathbf{x}_0)} \| \mathbf{s}_{\theta, t}(\mathbf{x}_t) - \nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t | \mathbf{x}_0) \|_2^2$$

### Continuous-in-time objective

$$\mathbb{E}_{\pi(\mathbf{x}(0))} \mathbb{E}_{t \sim U[0, 1]} \mathbb{E}_{q(\mathbf{x}(t) | \mathbf{x}(0))} \| \mathbf{s}_{\theta}(\mathbf{x}(t), t) - \nabla_{\mathbf{x}(t)} \log q(\mathbf{x}(t) | \mathbf{x}(0)) \|_2^2$$

## NCSN

$$q(\mathbf{x}(t) | \mathbf{x}(0)) = \mathcal{N} (\mathbf{x}(0), [\sigma^2(t) - \sigma^2(0)] \cdot \mathbf{I})$$

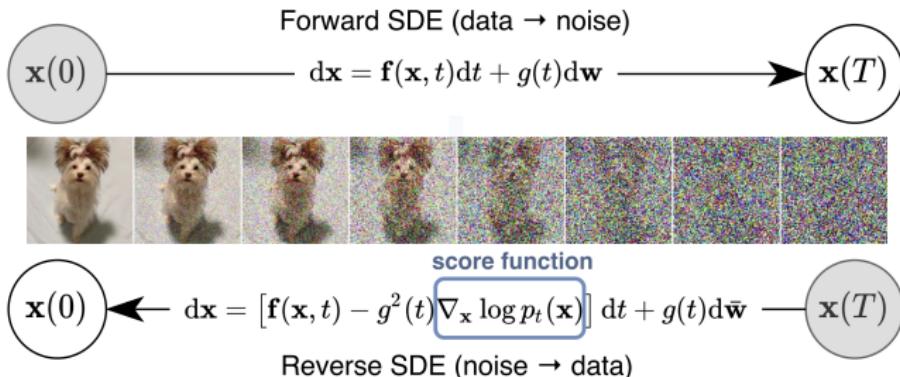
## DDPM

$$q(\mathbf{x}(t) | \mathbf{x}(0)) = \mathcal{N} \left( \mathbf{x}(0) e^{-\frac{1}{2} \int_0^t \beta(s) ds}, \left( 1 - e^{-\int_0^t \beta(s) ds} \right) \cdot \mathbf{I} \right)$$

# Recap of previous lecture

## Sampling

Solve reverse SDE using numerical solvers (SDESolve).



- ▶ Discretization of the reverse SDE gives us the ancestral sampling.
- ▶ If we use probability flow instead of SDE than the reverse ODE gives us the DDIM sampling.

## Recap of previous lecture

Let consider ODE dynamic  $\mathbf{x}(t)$  in time interval  $t \in [0, 1]$  with  $\mathbf{x}_0 \sim p_0(\mathbf{x}) = p(\mathbf{x})$ ,  $\mathbf{x}_1 \sim p_1(\mathbf{x}) = \pi(\mathbf{x})$ .

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, t), \quad \text{with initial condition } \mathbf{x}(0) = \mathbf{x}_0.$$

## KFP theorem (continuity equation)

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = -\operatorname{div}(\mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x})) \Leftrightarrow \frac{d \log p_t(\mathbf{x}(t))}{dt} = -\operatorname{tr}\left(\frac{\partial \mathbf{f}(\mathbf{x}(t), t)}{\partial \mathbf{x}(t)}\right)$$

Solving the continuity equation using the adjoint method is complicated and unstable process.

## Flow Matching

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

## Recap of previous lecture

Let's introduce the latent variable  $\mathbf{z}$ :

$$p_t(\mathbf{x}) = \int p_t(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z}$$
$$\frac{\partial p_t(\mathbf{x}|\mathbf{z})}{\partial t} = -\text{div}(\mathbf{f}(\mathbf{x}, \mathbf{z}, t)p_t(\mathbf{x}|\mathbf{z})).$$

- ▶  $p_t(\mathbf{x}|\mathbf{z})$  is a **conditional probability path**;
- ▶  $\mathbf{f}(\mathbf{x}, \mathbf{z}, t)$  is a **conditional vector field**.

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, t) \quad \Rightarrow \quad \frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, \mathbf{z}, t)$$

### Theorem

The following vector field generates the probability path  $p_t(\mathbf{x})$ .

$$\mathbf{f}(\mathbf{x}, t) = \mathbb{E}_{p_t(\mathbf{z}|\mathbf{x})}\mathbf{f}(\mathbf{x}, \mathbf{z}, t) = \int \mathbf{f}(\mathbf{x}, \mathbf{z}, t) \frac{p_t(\mathbf{x}|\mathbf{z})p(\mathbf{z})}{p_t(\mathbf{x})} d\mathbf{z}$$

## Recap of previous lecture

### Flow Matching (FM)

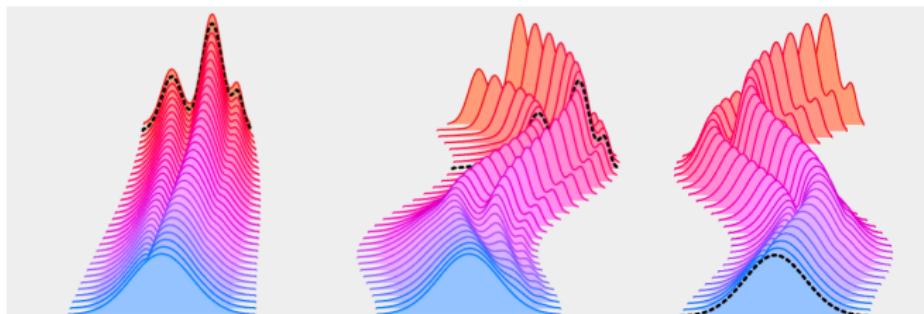
$$\mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{x} \sim p_t(\mathbf{x})} \|\mathbf{f}(\mathbf{x}, t) - \mathbf{f}_\theta(\mathbf{x}, t)\|^2 \rightarrow \min_{\theta}$$

### Conditional Flow Matching (CFM)

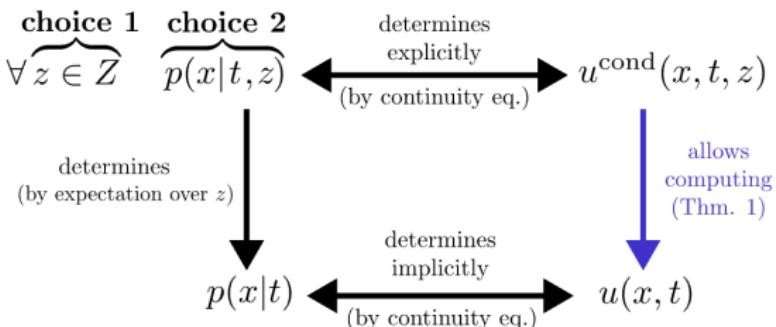
$$\mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})} \mathbb{E}_{\mathbf{x} \sim p_t(\mathbf{x}|\mathbf{z})} \|\mathbf{f}(\mathbf{x}, \mathbf{z}, t) - \mathbf{f}_\theta(\mathbf{x}, t)\|^2 \rightarrow \min_{\theta}$$

### Theorem

If  $\text{supp}(p_t(\mathbf{x})) = \mathbb{R}^m$ , then the optimal value of FM objective is equal to the optimal value of CFM objective.



# Recap of previous lecture



## Constraints

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

- ▶ How to choose the conditioning latent variable  $z$ ?
- ▶ How to define  $p_t(\mathbf{x}|z)$  which follows the constraints?

## Gaussian conditional probability path

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

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

## 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 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 need to ensure boundary constraints:

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

## Recap of previous lecture

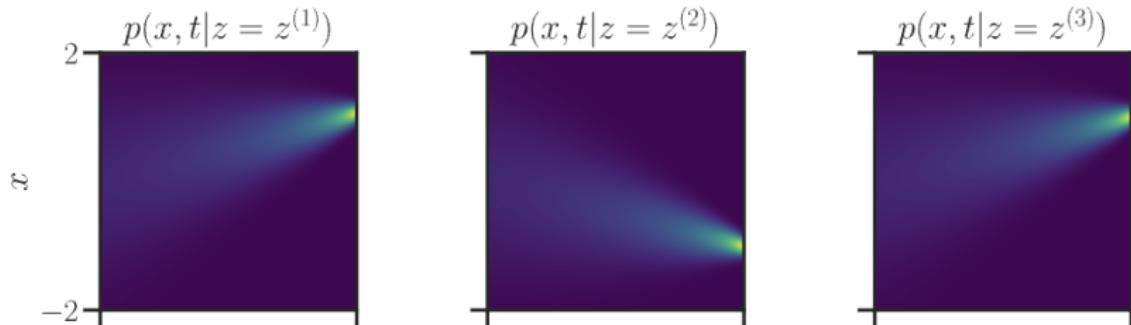
$$p_0(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(0, \mathbf{I}); \quad p_1(\mathbf{x}|\mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_1).$$

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 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. Conditional flow matching

Conical gaussian paths (continued)

Linear interpolation

Link with score-based models

## 2. Latent space models

Score-based models

Autoregressive models

## 3. The worst course overview

# Outline

## 1. Conditional flow matching

Conical gaussian paths (continued)

Linear interpolation

Link with score-based models

## 2. Latent space models

Score-based models

Autoregressive models

## 3. The worst course overview

## Conical gaussian paths

$$p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(t\mathbf{x}_1, (1-t)^2 \mathbf{I}); \quad \mathbf{x} = t\mathbf{x}_1 + (1-t)\mathbf{x}_0.$$

## Conditional vector field

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, \mathbf{x}_1, t) = \mu'_t(\mathbf{x}_1) + \frac{\sigma'_t(\mathbf{x}_1)}{\sigma_t(\mathbf{x}_1)} \odot (\mathbf{x} - \mu_t(\mathbf{x}_1))$$

$$\begin{aligned}\mathbf{f}(\mathbf{x}, \mathbf{x}_1, t) &= \mathbf{x}_1 - \frac{1}{1-t} \cdot (\mathbf{x} - t\mathbf{x}_1) = \frac{\mathbf{x}_1 - \mathbf{x}}{1-t} = \\ &= \frac{\mathbf{x}_1 - t\mathbf{x}_1 - (1-t)\mathbf{x}_0}{1-t} = \mathbf{x}_1 - \mathbf{x}_0\end{aligned}$$

## Conditional Flow Matching

$$\begin{aligned}\mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})} \mathbb{E}_{\mathbf{x} \sim p_t(\mathbf{x}|\mathbf{z})} \|\mathbf{f}(\mathbf{x}, \mathbf{z}, t) - \mathbf{f}_\theta(\mathbf{x}, t)\|^2 &= \\ \mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{x}_1 \sim \pi(\mathbf{x})} \mathbb{E}_{\mathbf{x} \sim p_t(\mathbf{x}|\mathbf{x}_1)} \left\| \left( \frac{\mathbf{x}_1 - \mathbf{x}}{1-t} \right) - \mathbf{f}_\theta(\mathbf{x}, t) \right\|^2 &= \\ \mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{x}_1 \sim \pi(\mathbf{x})} \mathbb{E}_{\mathbf{x}_0 \sim \mathcal{N}(0, \mathbf{I})} \|(\mathbf{x}_1 - \mathbf{x}_0) - \mathbf{f}_\theta(t\mathbf{x}_1 + (1-t)\mathbf{x}_0, t)\|^2\end{aligned}$$

# Conditional Flow Matching

$$\mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{x}_1 \sim \pi(\mathbf{x})} \mathbb{E}_{\mathbf{x}_0 \sim \mathcal{N}(0, \mathbf{I})} \|(\mathbf{x}_1 - \mathbf{x}_0) - \mathbf{f}_\theta(\mathbf{x}, t)\|^2 \rightarrow \min_{\theta}$$

We fit straight lines between noise distribution  $p(\mathbf{x})$  and the data distribution  $\pi(\mathbf{x})$ .

## Training

1. Get the sample  $\mathbf{x}_1 \sim \pi(\mathbf{x})$ .
2. Sample timestamp  $t \sim U[0, 1]$  and  $\mathbf{x}_0 \sim \mathcal{N}(0, \mathbf{I})$ .
3. Get noisy image  $\mathbf{x}_t = t\mathbf{x}_1 + (1 - t)\mathbf{x}_0$ .
4. Compute loss  $\mathcal{L} = \|(\mathbf{x}_1 - \mathbf{x}_0) - \mathbf{f}_\theta(\mathbf{x}, t)\|^2$ .

## Sampling

1. Sample  $\mathbf{x}_0 \sim \mathcal{N}(0, \mathbf{I})$ .
2. Solve the ODE to get  $\mathbf{x}_1$ :

$$\mathbf{x}_1 = \text{ODESolve}_f(\mathbf{x}_0, \theta, t_0 = 0, t_1 = 1).$$

# Flow Matching

$$\mathbf{x}_t = t\mathbf{x}_1 + (1 - t)\mathbf{x}_0$$

- ▶ The conditional probability path  $p_t(\mathbf{x}|\mathbf{z})$  is an **optimal transport path** from  $p_0(\mathbf{x}|\mathbf{z})$  to  $p_1(\mathbf{x}|\mathbf{z})$  (in terms of the conditional trajectories straightness).
- ▶ The marginal path  $p_t(\mathbf{x})$  is not in general an optimal transport path from the standard normal  $p_0(\mathbf{x})$  to the data distribution  $p_1(\mathbf{x})$ .



image credit: <https://mlg.eng.cam.ac.uk/blog/2024/01/20/flow-matching.html>

# Outline

## 1. Conditional flow matching

Conical gaussian paths (continued)

**Linear interpolation**

Link with score-based models

## 2. Latent space models

Score-based models

Autoregressive models

## 3. The worst course overview

# Pair conditioning

## Conditional Flow Matching

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

## Conditioning latent variable

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

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

We need to ensure boundary constraints:

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

# Linear interpolation

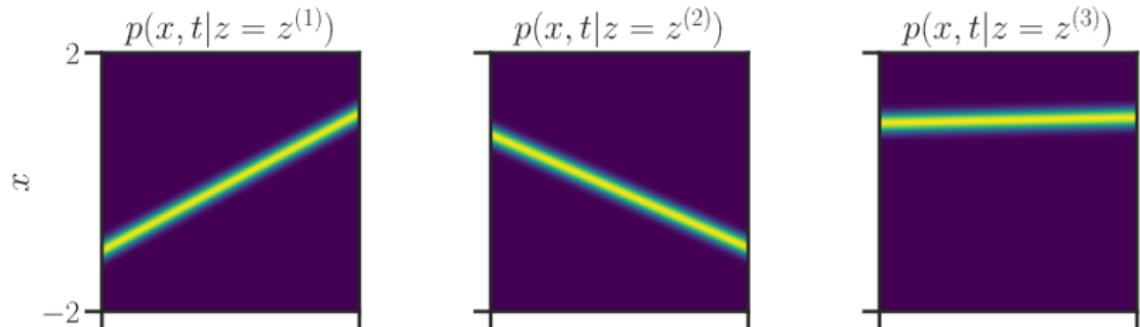
$$p_0(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_0); \quad p_1(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_1).$$

Gaussian conditional probability path

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

Let consider straight conditional paths

$$\begin{cases} \mu_t(\mathbf{x}_1) = t\mathbf{x}_1 + (1-t)\mathbf{x}_0; \\ \sigma_t(\mathbf{x}_1) = \epsilon. \end{cases} \Rightarrow \begin{cases} p_0(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_0); \\ p_1(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \delta(\mathbf{x} - \mathbf{x}_1). \end{cases}$$



# Flow Matching: conical paths vs linear interpolation

$$z = x_1$$

$$p_t(x|x_1) = \mathcal{N}(tx_1, (1-t)^2 I)$$

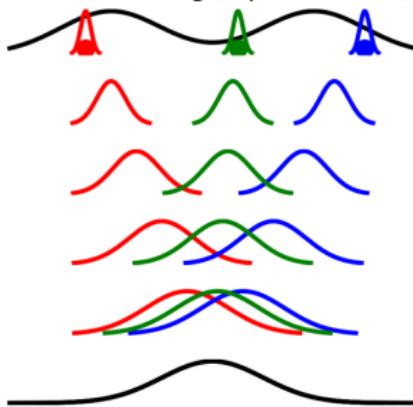
$$x_t = tx_1 + (1-t)x_0.$$

$$z = (x_0, x_1)$$

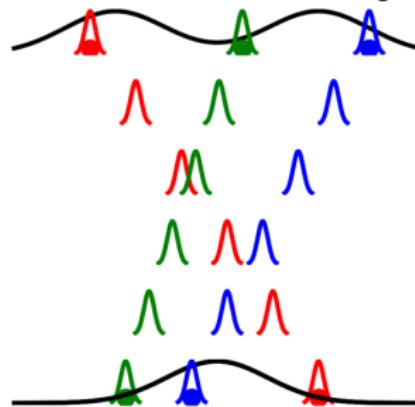
$$p_t(x|x_0, x_1) = \mathcal{N}(tx_1 + (1-t)x_0, \epsilon^2 I)$$

$$x_t = tx_1 + (1-t)x_0.$$

Flow Matching (Lipman et al.)



Conditional Flow Matching



## Linear interpolation

$$p_t(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1) = \mathcal{N}\left(t\mathbf{x}_1 + (1-t)\mathbf{x}_0, \epsilon^2 \mathbf{I}\right); \quad \mathbf{x}_t = t\mathbf{x}_1 + (1-t)\mathbf{x}_0.$$

## Conditional vector field

$$\begin{aligned} \frac{d\mathbf{x}}{dt} &= \mathbf{f}(\mathbf{x}, \mathbf{x}_0, \mathbf{x}_1, t) = \boldsymbol{\mu}'_t(\mathbf{x}_0, \mathbf{x}_1) + \frac{\boldsymbol{\sigma}'_t(\mathbf{x}_0, \mathbf{x}_1)}{\boldsymbol{\sigma}_t(\mathbf{x}_0, \mathbf{x}_1)} \odot (\mathbf{x} - \boldsymbol{\mu}_t(\mathbf{x}_0, \mathbf{x}_1)) \\ \mathbf{f}(\mathbf{x}, \mathbf{x}_0, \mathbf{x}_1, t) &= \mathbf{x}_1 - \mathbf{x}_0 \end{aligned}$$

## Conditional Flow Matching

$$\begin{aligned} \mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})} \mathbb{E}_{\mathbf{x} \sim p_t(\mathbf{x}|\mathbf{z})} \|\mathbf{f}(\mathbf{x}, \mathbf{z}, t) - \mathbf{f}_\theta(\mathbf{x}, t)\|^2 &= \\ \mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{(\mathbf{x}_0, \mathbf{x}_1) \sim p(\mathbf{x}_0, \mathbf{x}_1)} \mathbb{E}_{\mathbf{x} \sim p_t(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_1)} \|(\mathbf{x}_1 - \mathbf{x}_0) - \mathbf{f}_\theta(\mathbf{x}_t, t)\|^2 & \end{aligned}$$

- ▶ We got the same procedure as for conical paths!
- ▶ Now we do not have the constraint that  $p_0(\mathbf{x})$  should be  $\mathcal{N}(0, \mathbf{I})$ .

## Conditional Flow Matching

- ▶ We could use this conditioning for transferring any distribution  $p_0(\mathbf{x})$  to any distribution  $p_1(\mathbf{x})$ .
- ▶ It is possible to use this approach for paired problems (style transfer).

### Training

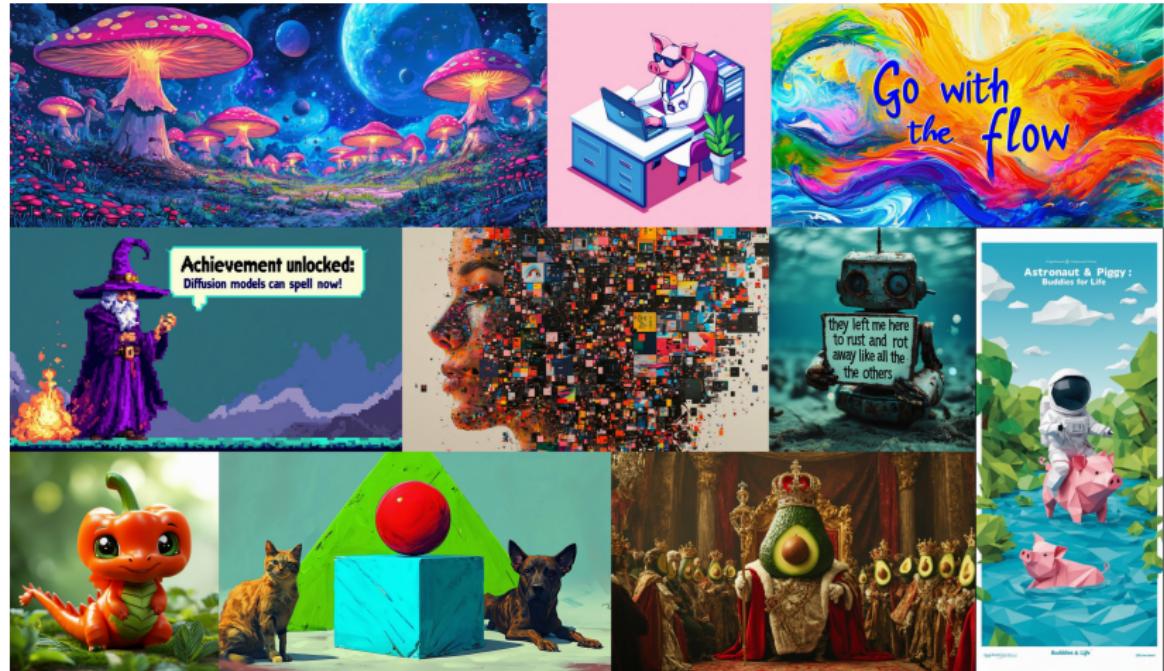
1. Get the sample  $(\mathbf{x}_0, \mathbf{x}_1) \sim p(\mathbf{x}_0, \mathbf{x}_1)$ .
2. Sample timestamp  $t \sim U[0, 1]$ .
3. Get noisy image  $\mathbf{x}_t = t\mathbf{x}_1 + (1 - t)\mathbf{x}_0$ .
4. Compute loss  $\mathcal{L} = \|(\mathbf{x}_1 - \mathbf{x}_0) - \mathbf{f}_{\theta}(\mathbf{x}, t)\|^2$ .

### Sampling

1. Sample  $\mathbf{x}_0 \sim p_0(\mathbf{x})$ .
2. Solve the ODE to get  $\mathbf{x}_1$ :

$$\mathbf{x}_1 = \text{ODESolve}_f(\mathbf{x}_0, \boldsymbol{\theta}, t_0 = 0, t_1 = 1).$$

# Stable Diffusion 3: scalable flow matching



# Outline

## 1. Conditional flow matching

Conical gaussian paths (continued)

Linear interpolation

Link with score-based models

## 2. Latent space models

Score-based models

Autoregressive models

## 3. The worst course overview

# Score-based generative models through SDEs

## Training

$$\mathbb{E}_{\pi(\mathbf{x}(0))} \mathbb{E}_{t \sim U[0,1]} \mathbb{E}_{q(\mathbf{x}(t)|\mathbf{x}(0))} \| \mathbf{s}_\theta(\mathbf{x}(t), t) - \nabla_{\mathbf{x}(t)} \log q(\mathbf{x}(t)|\mathbf{x}(0)) \|_2^2$$

## Variance Exploding SDE (NCSN)

$$q(\mathbf{x}(t)|\mathbf{x}(0)) = \mathcal{N}(\mathbf{x}(0), [\sigma^2(t) - \sigma^2(0)] \cdot \mathbf{I}), \quad \sigma(0) = 0.$$

## Variance Preserving SDE (DDPM)

$$q(\mathbf{x}(t)|\mathbf{x}(0)) = \mathcal{N}(\mathbf{x}(0)\alpha(t), (1 - \alpha(t)^2) \cdot \mathbf{I}); \quad \alpha(t) = e^{-\frac{1}{2} \int_0^t \beta(s) ds}$$

Flow matching uses the reverse time direction.

$$\textbf{NCSN: } p(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(\mathbf{x}_1, \sigma_{1-t}^2 \cdot \mathbf{I})$$

$$\textbf{DDPM: } p(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(\alpha_{1-t}\mathbf{x}_1, (1 - \alpha_{1-t}^2) \cdot \mathbf{I})$$

## Flow matching vs score-based SDE models

### Flow matching probability path

$$p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N} \left( t\mathbf{x}_1, (1-t)^2 \mathbf{I} \right); \quad \mathbf{f}(\mathbf{x}, \mathbf{x}_1, t) = \frac{\mathbf{x}_1 - \mathbf{x}}{1-t}$$

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, \mathbf{x}_1, t) = \boldsymbol{\mu}'_t(\mathbf{x}_1) + \frac{\boldsymbol{\sigma}'_t(\mathbf{x}_1)}{\boldsymbol{\sigma}_t(\mathbf{x}_1)} \odot (\mathbf{x} - \boldsymbol{\mu}_t(\mathbf{x}_1))$$

### Variance Exploding SDE probability path

$$p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N} \left( \mathbf{x}_1, \sigma_{1-t}^2 \mathbf{I} \right) \quad \Rightarrow \quad \mathbf{f}(\mathbf{x}, \mathbf{x}_1, t) = -\frac{\boldsymbol{\sigma}'_{1-t}}{\boldsymbol{\sigma}_{1-t}} \cdot (\mathbf{x} - \mathbf{x}_1)$$

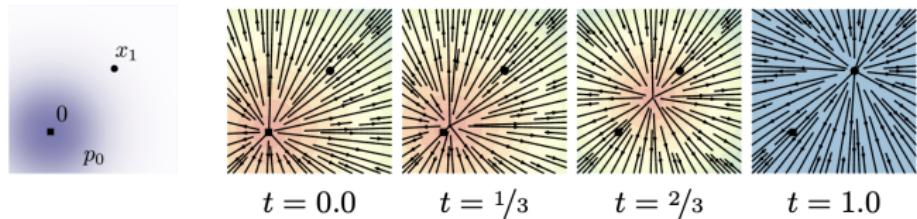
### Variance Preserving SDE probability path

$$p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N} \left( \alpha_{1-t} \mathbf{x}_1, (1 - \alpha_{1-t}^2) \mathbf{I} \right) \Rightarrow \mathbf{f}(\mathbf{x}, \mathbf{x}_1, t) = \frac{\alpha'_{1-t}}{1 - \alpha_{1-t}^2} \cdot (\alpha_{1-t} \mathbf{x} - \mathbf{x}_1)$$

# Flow matching vs score-based SDE models

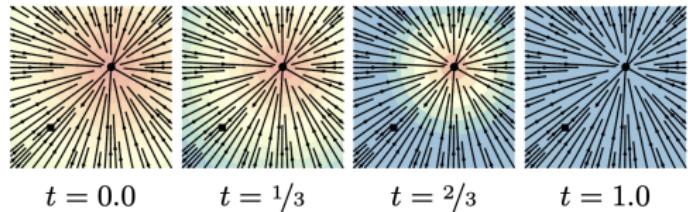
SDE vector field

$$\mathbf{f}(\mathbf{x}, \mathbf{x}_1, t) = \frac{\alpha'_{1-t}}{1 - \alpha_{1-t}^2} \cdot (\alpha_{1-t} \mathbf{x} - \mathbf{x}_1)$$



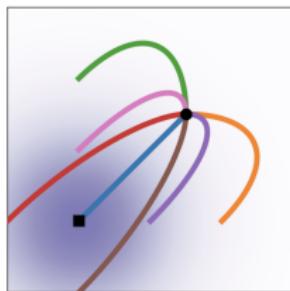
FM vector field

$$\mathbf{f}(\mathbf{x}, \mathbf{x}_1, t) = \frac{\mathbf{x}_1 - \mathbf{x}}{1 - t}$$

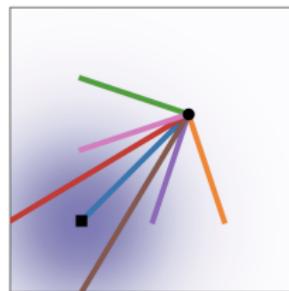


# Flow matching vs score-based SDE models

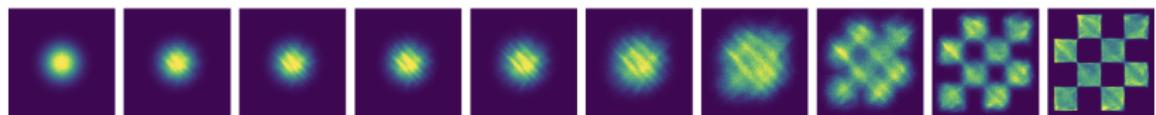
## Trajectories



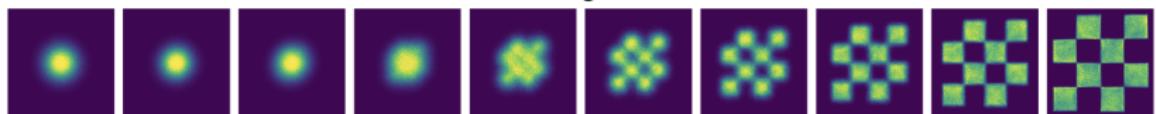
Diffusion



OT



Score matching w/ Diffusion



Flow Matching w/ OT

# Outline

## 1. Conditional flow matching

Conical gaussian paths (continued)

Linear interpolation

Link with score-based models

## 2. Latent space models

Score-based models

Autoregressive models

## 3. The worst course overview

# Outline

## 1. Conditional flow matching

Conical gaussian paths (continued)

Linear interpolation

Link with score-based models

## 2. Latent space models

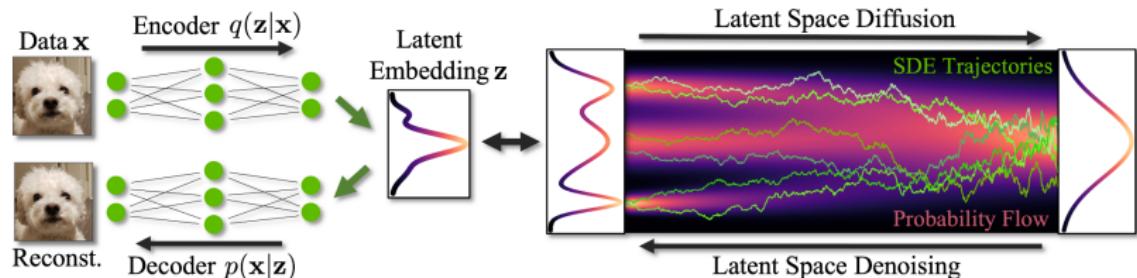
Score-based models

Autoregressive models

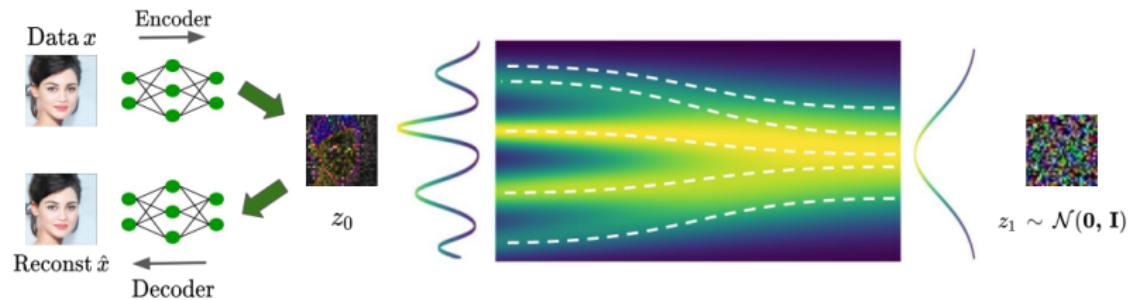
## 3. The worst course overview

# 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. Conditional flow matching

Conical gaussian paths (continued)

Linear interpolation

Link with score-based models

## 2. Latent space models

Score-based models

Autoregressive models

## 3. The worst course overview

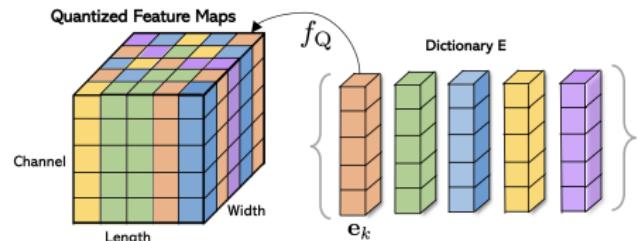
# Vector Quantized VAE (VQ-VAE)

## Vector quantization

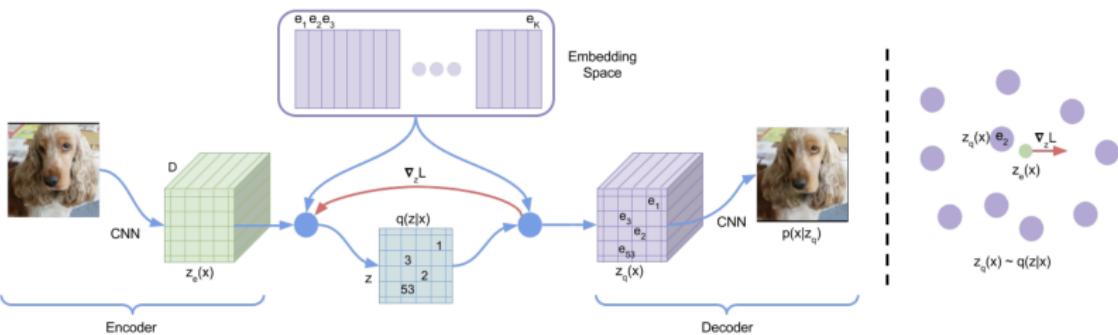
Define the dictionary space  $\{\mathbf{e}_k\}_{k=1}^K$ , where  $\mathbf{e}_k \in \mathbb{R}^C$ ,  $K$  is the size of the dictionary.

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

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



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

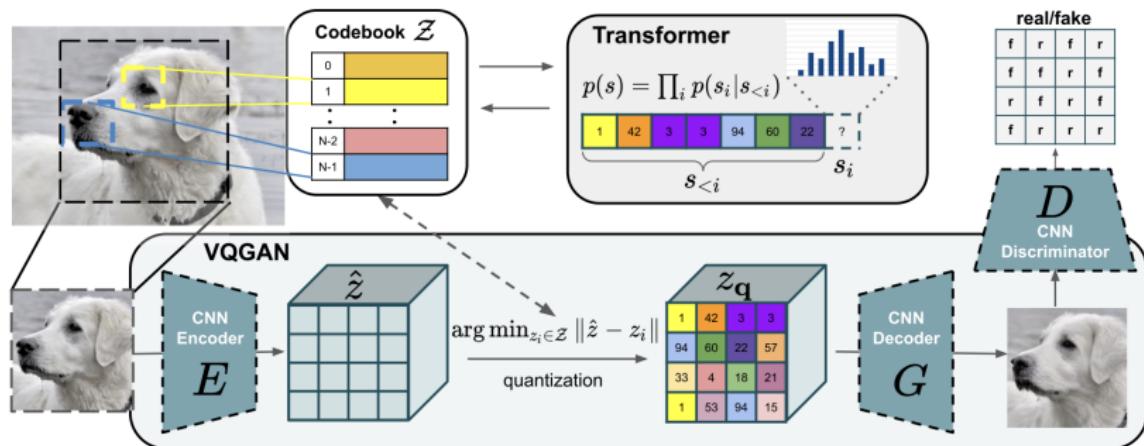


Zhao Y. et al. Feature Quantization Improves GAN Training, 2020

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

# Vector Quantized GAN

- ▶ Use VQVAE model and objective.
- ▶ Add adversarial loss between generated and real images to improve the visual quality of the reconstructions.

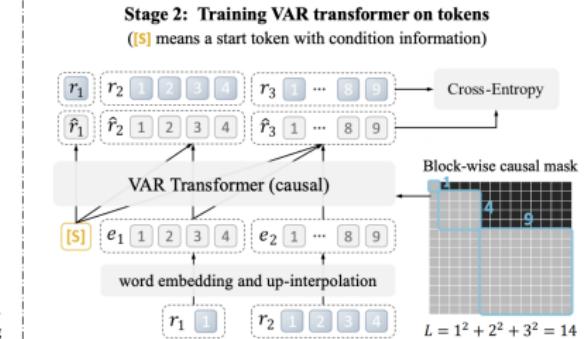
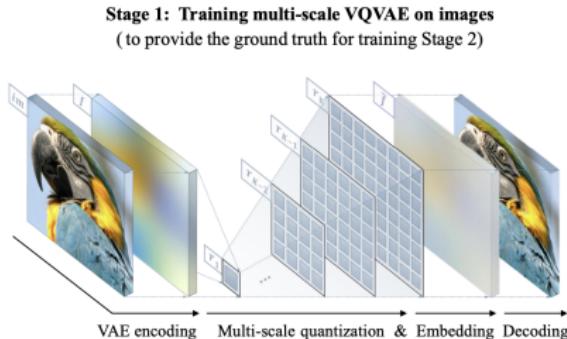
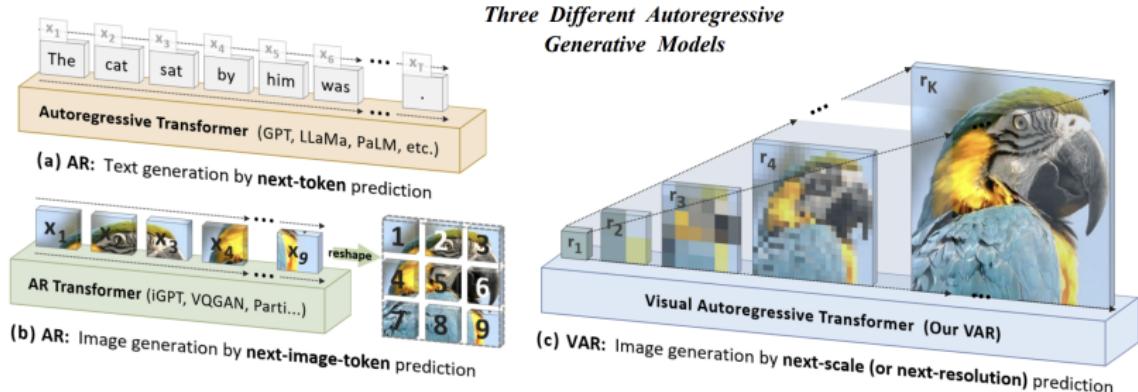


## LlamaGen: pure autoregression

- ▶ Use VQGAN encoder to map images to the discrete latent space of the codebook vectors.
- ▶ Learn the pure autoregression model (Llama-based) in the latent space.
- ▶ Use VQGAN decoder to map the discrete space tokens to the image space.



# Visual Autoregressive Modeling (VAR)



# Outline

## 1. Conditional flow matching

Conical gaussian paths (continued)

Linear interpolation

Link with score-based models

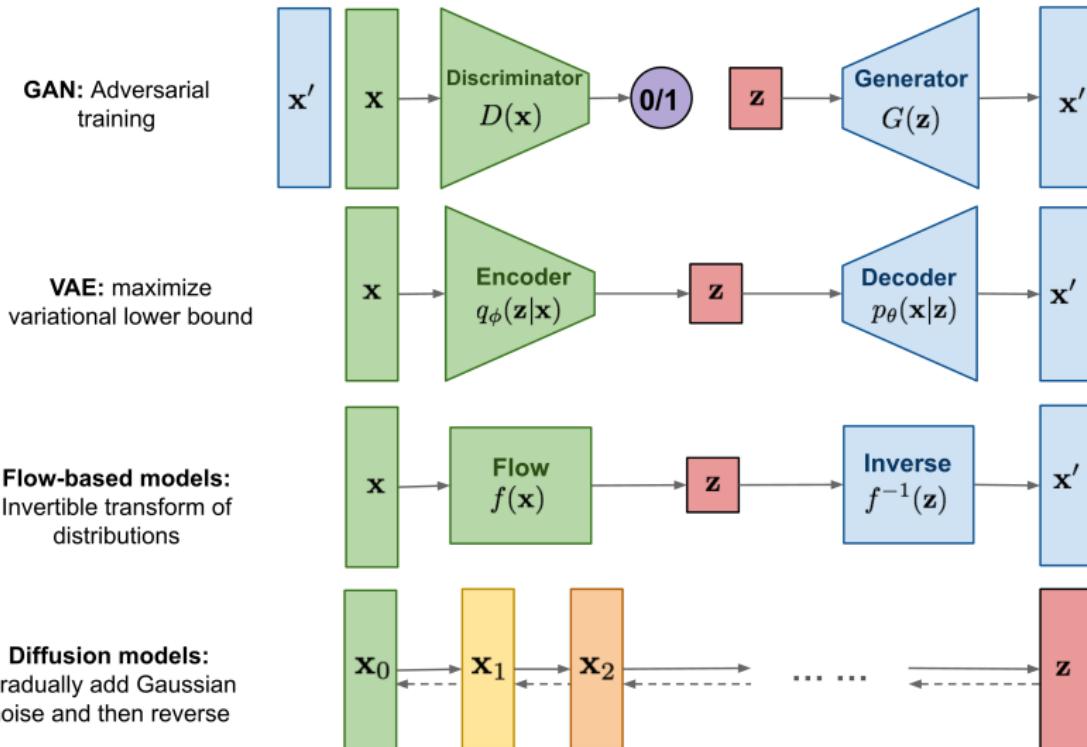
## 2. Latent space models

Score-based models

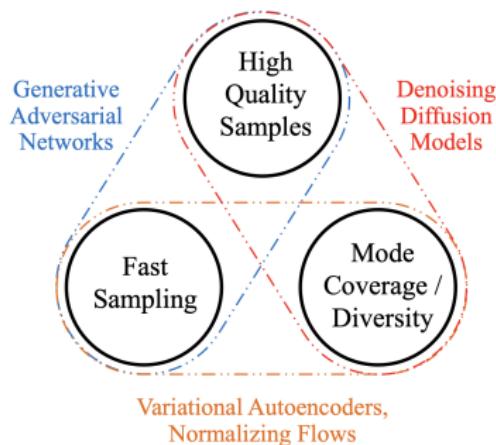
Autoregressive models

## 3. 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	✗	✓	?	✗	✗	✗

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

- ▶ Conical gaussian paths is the example of the effective FM technique.
- ▶ Pair conditioning gives the same procedure, but it is more general.
- ▶ Diffusion and score-based model are the special case of flow matching approach with curved trajectories.
- ▶ Most of the state-of-the-art generative models are latent models with continuous or discrete latent space.