

Siddhant Mishra-Sharma (MIT/AI FI) | AI FI Summer School

166

3

5

Variational diffusion models



Align the forward and reverse distributions; variational bound (ELBO)

$$L = \left\langle \log \frac{q \left(x, z_1, z_2, \cdots, z_T \right)}{p \left(x, z_1, z_2, \cdots, z_T \right)} \right\rangle_{q(x)}$$



Some calculations later

[checkexpectations.vrt]

$$L = \underbrace{\left\langle \log p_{\theta}(x | z_1) \right\rangle_{q(z_1|x)}}_{\text{Reconstruction}} - \underbrace{D_{\text{KL}}\left(q(z_T | x) \parallel p(z_T)\right)}_{\text{Prior regularization}} - \underbrace{\sum_{t=2}^T \left\langle D_{\text{KL}}\left(q(z_{t-1} | z_t, x) \parallel p_{\theta}(z_{t-1} | z_t)\right) \right\rangle_{q(z_t|x)}}_{\text{Denoising matching}}$$

Reconstruction

(Noise model; no trainable parameters)

Prior regularization

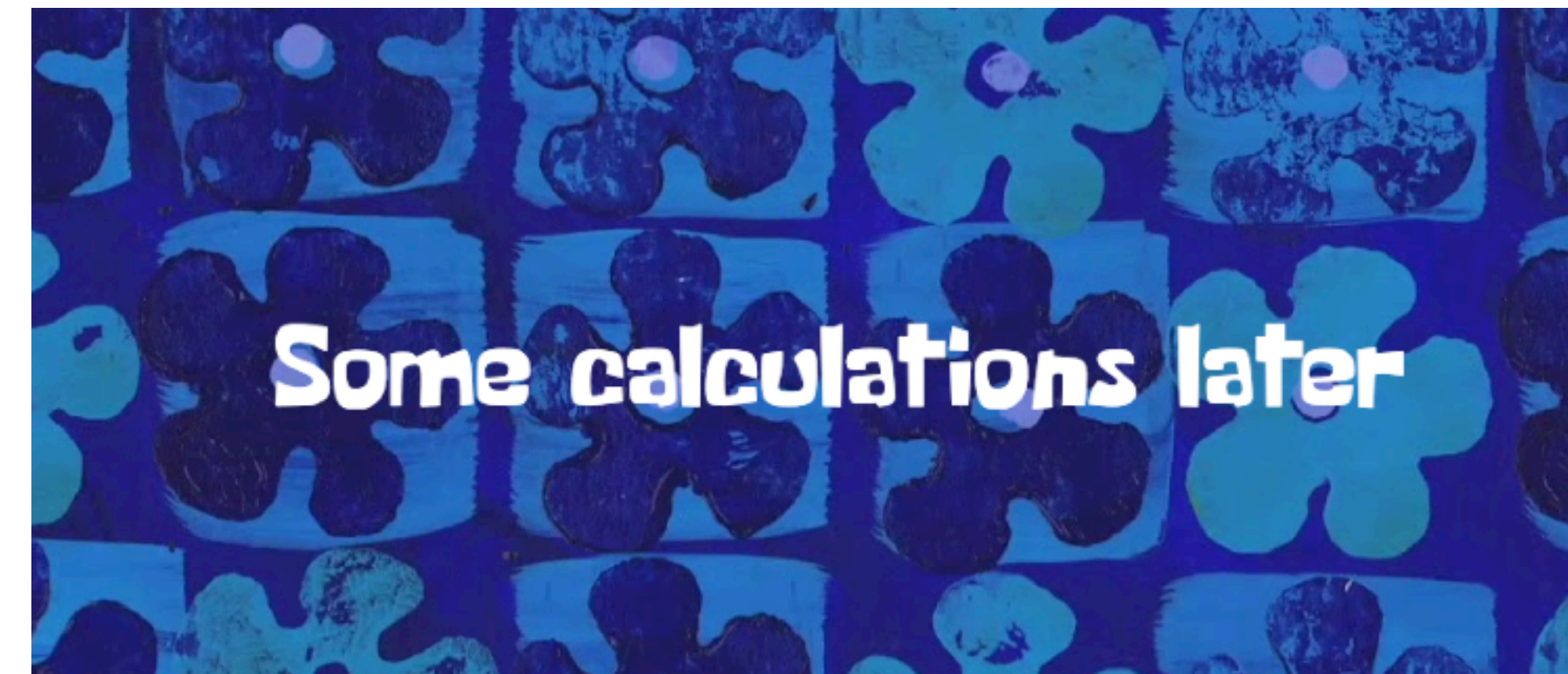
(No trainable parameters)

Denoising matching

Variational diffusion models

Align the forward and reverse distributions; variational lower bound (ELBO) as before

$$L = \left\langle \log \frac{q(x, z_1, z_2, \dots, z_T)}{p(x, z_1, z_2, \dots, z_T)} \right\rangle_{q(x)}$$



$$L = \left\langle \log p_{\theta}(x | z_1) \right\rangle_{q(z_1|x)} - D_{\text{KL}} \left(q(z_T | x) \parallel p(z_T) \right) - \sum_{t=2}^T \left\langle D_{\text{KL}} \left(q(z_{t-1} | z_t, x) \parallel p_{\theta}(z_{t-1} | z_t) \right) \right\rangle_{q(z_t|x)}$$

Reconstruction

(Noise model; no trainable parameters)

Prior regularization

(No trainable parameters)

Denoising matching

The forward process and diffusion kernel

Predict arbitrary timestep without Markovian sampling

$$\sum_{t=2}^T \left\langle D_{\text{KL}} \left(q(z_{t-1} | z_t, x) \parallel p_{\theta}(z_{t-1} | z_t) \right) \right\rangle_{q(z_t|x)}$$

