

Towards diffusion: (Markovian) hierarchical VAEs

Diffusion models can be seen as hierarchical VAEs with a few restrictions:

- The forward (*encoding*) distribution prescribed as a Markov chain of Gaussians; it is not learned

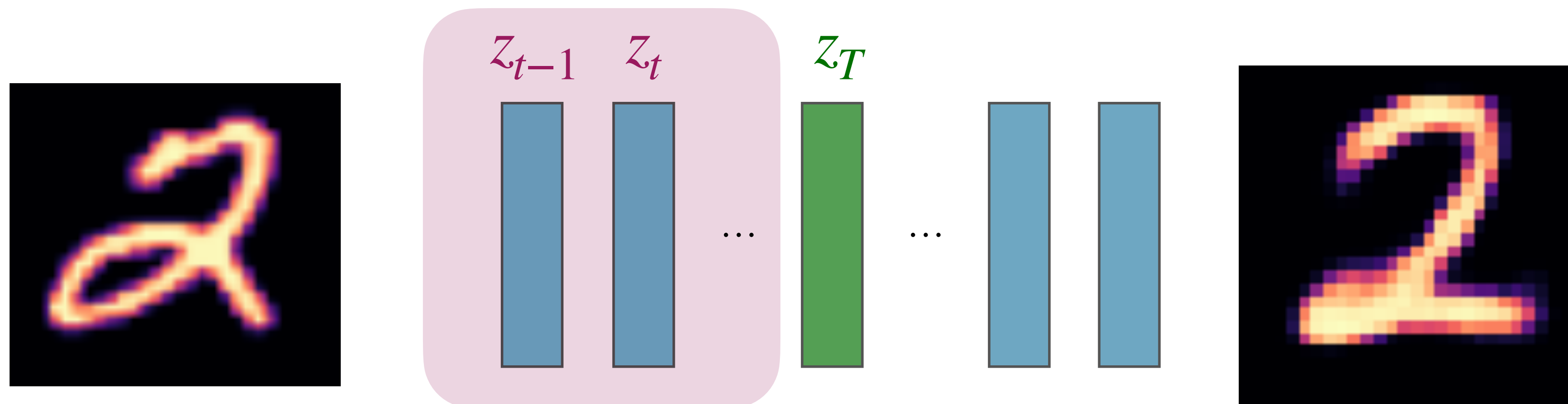
$$q(z_t | z_{t-1}) = \mathcal{N}(z_t; \alpha_t z_{t-1}, \beta_t)$$

- Distributions of latents at the final timestep T is a standard (unit) Gaussian

$$q(z_T | z_{T-1}, \dots, x) = \mathcal{N}(z_T; 0, \mathbb{I})$$

- The dimensionality of latents is the same as the data dimensionality

$$\dim(z_t) = \dim(x)$$



Variational diffusion models

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

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