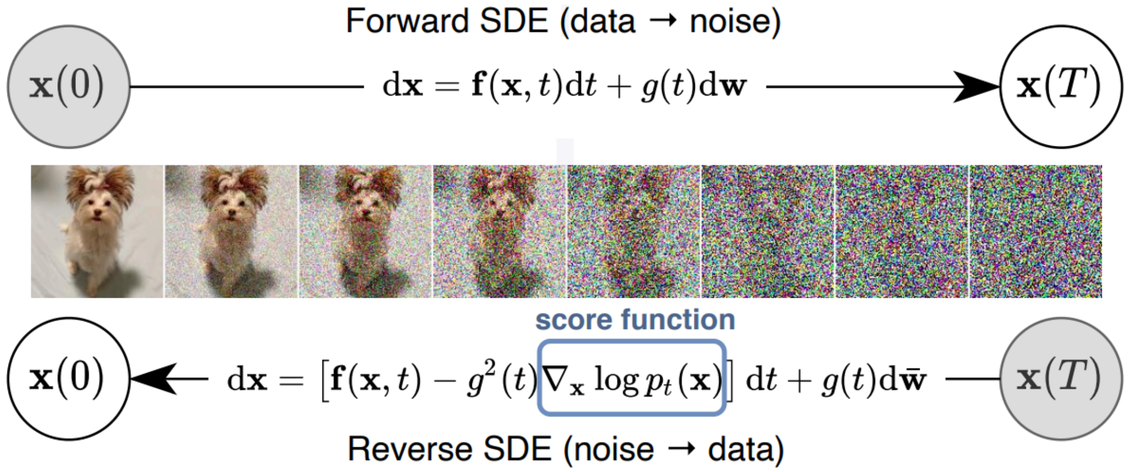
Diffusion models in latent space

02.12.2022

Denoising Score-Matching

At first glance



- Marginal kernel: $q(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t \mid \alpha_t \mathbf{x}_0, \sigma_t^2 I)$
- Minimisation objective:

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{t\sim\mathcal{U}[0,1]} \left[\lambda(t) \mathbb{E}_{q(\mathbf{x}_0)} \mathbb{E}_{q(\mathbf{x}_t \mid \mathbf{x}_0)} \left[\left\| \nabla_{\mathbf{x}_t} \log q\left(\mathbf{x}_t\right) - \nabla_{\mathbf{x}_t} \log p_{\boldsymbol{\theta}}\left(\mathbf{x}_t\right) \right\|_2^2 \right] \right]$$

Denoising Score-Matching

More details

- What is $q\left(\mathbf{x}_t\right)$? All, we know - $q\left(\mathbf{x}_0\right), q\left(\mathbf{x}_t \mid \mathbf{x}_0\right)$
- So training objective can be rewritten as:

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{t \sim \mathcal{U}[0,1]}\left[\lambda(t) \mathbb{E}_{q\left(\mathbf{x}_0\right)} \mathbb{E}_{q\left(\mathbf{x}_t \mid \mathbf{x}_0\right)}\left[\left\| \nabla_{\mathbf{x}_t} \log q\left(\mathbf{x}_t \mid \mathbf{x}_0\right)-\nabla_{\mathbf{x}_t} \log p_{\boldsymbol{\theta}}\left(\mathbf{x}_t\right)\right\|_2^2\right]\right]+C$$

- It's not just a random constant, we don't care about (in app.):

$$C=\mathbb{E}_{t \sim \mathcal{U}[0,1]}\left[\lambda(t) \mathbb{E}_{q\left(\mathbf{x}_0\right)} \mathbb{E}_{q\left(\mathbf{x}_t \mid \mathbf{x}_0\right)}\left[\left\| \nabla_{\mathbf{x}_t} \log q\left(\mathbf{x}_t\right)\right\|_2^2-\left\| \nabla_{\mathbf{x}_t} \log q\left(\mathbf{x}_t \mid \mathbf{x}_0\right)\right\|_2^2\right]\right]$$

- Complex derivation from Song in "Maximum likelihood training of score-based diffusion models":

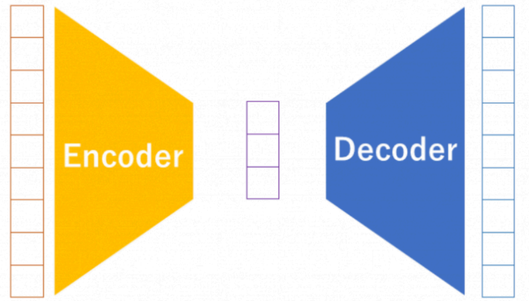
$$\mathrm{KL}\left(q\left(\mathbf{x}_0\right) \| p_{\boldsymbol{\theta}}\left(\mathbf{x}_0\right)\right) \leq \mathbb{E}_{t \sim \mathcal{U}[0,1]}\left[\frac{g(t)^2}{2} \mathbb{E}_{q\left(\mathbf{x}_0\right)} \mathbb{E}_{q\left(\mathbf{z}_t \mid \mathbf{x}_0\right)}\left[\left\| \nabla_{\mathbf{x}_t} \log q\left(\mathbf{x}_t\right)-\nabla_{\mathbf{x}_t} \log p_{\boldsymbol{\theta}}\left(\mathbf{x}_t\right)\right\|_2^2\right]\right]$$

Variational AutoEncoder

Cause it’s likelihood-based generative model

- Root idea: $\log p(x) = ELBO + KL(q(z|x)\|p(z|x))$

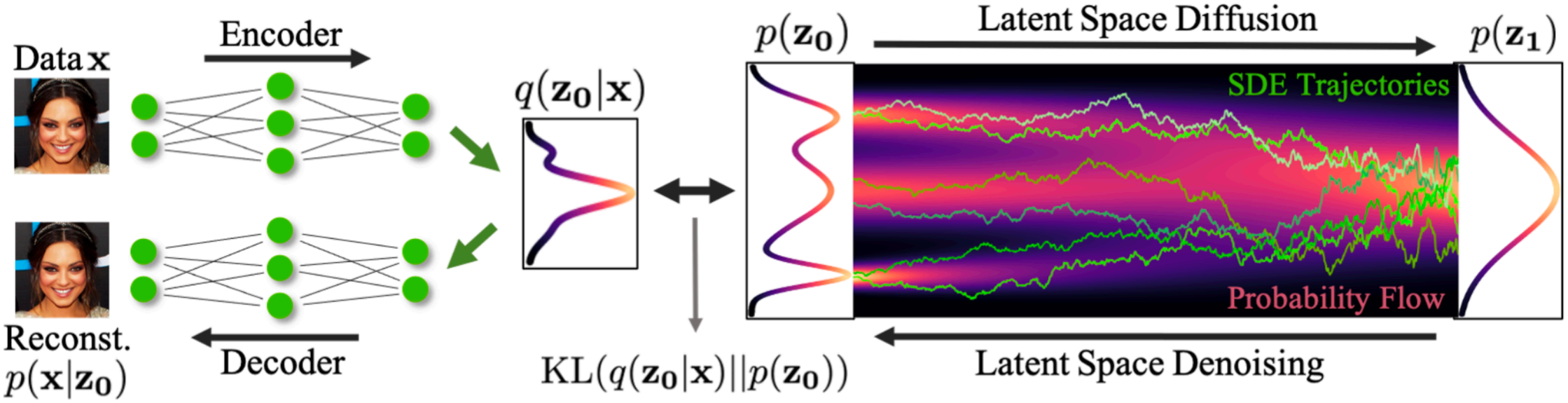
- $$ELBO = \mathbb{E}_{q(z|x)}\left[\log p(x,z) - \log q(z|x)\right] =$$
$$= \mathbb{E}_{q(z|x)}\left[\log p(x|z) + \log p(z) - \log q(z|x)\right]$$
$$= \mathbb{E}_{q(z|x)} \log p(x|z) - KL(q(z|x)\|p(z))$$



- Training objective looks like: $\mathbb{E}_{q_\phi(z|x)}\left[-\log p_\psi(x|z)\right] + KL(q_\phi(z|x)\|p(z))$

Latent Score-based Generative Model

…instead of a thousands equations

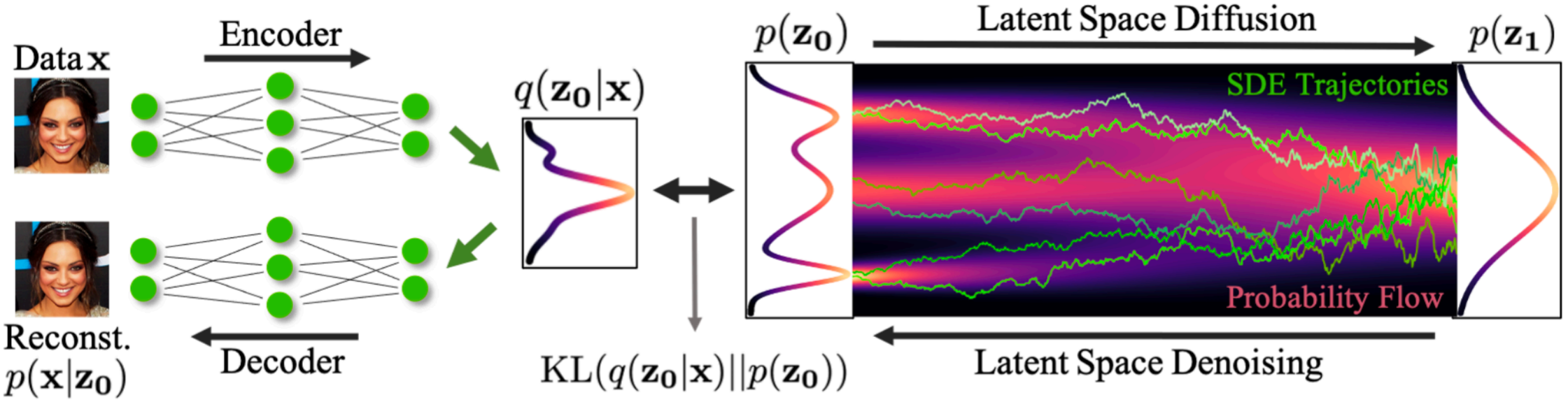


$$\mathbb{E}_{q_\phi(\mathbf{z}_0|\mathbf{x})}\left[-\log p_\psi(\mathbf{x}\,|\,\mathbf{z}_0)\right] + KL(q_\phi(\mathbf{z}_0\,|\,\mathbf{x})\|p_\theta(\mathbf{z}_0))$$

$$\mathrm{KL}\left(q\left(\mathbf{z}_0\,|\,\mathbf{x}\right)\|p_{\boldsymbol{\theta}}\left(\mathbf{z}_0\right)\right) \leq \mathbb{E}_{t\sim\mathcal{U}[0,1]}\left[\frac{g(t)^2}{2}\mathbb{E}_{q(\mathbf{z}_0\,|\,\mathbf{x})}\mathbb{E}_{q(\mathbf{z}_t\,|\,\mathbf{z}_0)}\left[\left\|\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right) - \nabla_{\mathbf{z}_t}\log p_{\boldsymbol{\theta}}\left(\mathbf{z}_t\right)\right\|_2^2\right]\right]$$

Latent Score-based Generative Model

…instead of a thousands equations



$$\mathbb{E}_{q_\phi(\mathbf{z}_0|\mathbf{x})}\left[-\log p_\psi(\mathbf{x}\,|\,\mathbf{z}_0)\right] + KL(q_\phi(\mathbf{z}_0\,|\,\mathbf{x})\|p_\theta(\mathbf{z}_0))$$

$$\mathrm{KL}\left(q\left(\mathbf{z}_0\,|\,\mathbf{x}\right)\|p_{\boldsymbol{\theta}}\left(\mathbf{z}_0\right)\right) \leq \mathbb{E}_{t\sim\mathcal{U}[0,1]}\left[\frac{g(t)^2}{2}\mathbb{E}_{q(\mathbf{z}_0\,|\,\mathbf{x})}\mathbb{E}_{q(\mathbf{z}_t\,|\,\mathbf{z}_0)}\left[\left\|\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right) - \nabla_{\mathbf{z}_t}\log p_{\boldsymbol{\theta}}\left(\mathbf{z}_t\right)\right\|_2^2\right]\right]$$

Latent Score-based Generative Model

Objective's derivation - I

$$\begin{aligned}\mathscr{L}(\mathbf{x}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\psi}) &= \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}_0 \mid \mathbf{x})}\left[-\log p_{\boldsymbol{\psi}}\left(\mathbf{x} \mid \mathbf{z}_0\right)\right] + \mathrm{KL}\left(q_{\boldsymbol{\phi}}\left(\mathbf{z}_0 \mid \mathbf{x}\right) \| p_{\boldsymbol{\theta}}\left(\mathbf{z}_0\right)\right) \\ &= \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}_0 \mid \mathbf{x})}\left[-\log p_{\boldsymbol{\psi}}\left(\mathbf{x} \mid \mathbf{z}_0\right)\right] + \mathbb{E}_{q_{\phi}(\mathbf{z}_0 \mid \mathbf{x})}\left[\log q_{\phi}\left(\mathbf{z}_0 \mid \mathbf{x}\right)\right] + \mathbb{E}_{q_{\phi}(\mathbf{z}_0 \mid \mathbf{x})}\left[-\log p_{\boldsymbol{\theta}}\left(\mathbf{z}_0\right)\right]\end{aligned}$$

Aim - derive last CE term suitable for optimisation:

$$CE\left(q\left(\mathbf{z}_0 \mid \mathbf{x}\right) \| p\left(\mathbf{z}_0\right)\right) = \mathbb{E}_{t \sim \mathscr{U}[0,1]}\left[\frac{g(t)^2}{2} \mathbb{E}_{q\left(\mathbf{z}_t, \mathbf{z}_0 \mid \mathbf{x}\right)}\left[\left\| \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t \mid \mathbf{z}_0\right) - \nabla_{\mathbf{z}_t} \log p\left(\mathbf{z}_t\right) \right\|_2^2\right]\right] + \frac{D}{2} \log\left(2 \pi e \sigma_0^2\right)$$

LSGM's objective - I

Fokker-Plank equation recap

- Consider forward SDE: $d\mathbf{z} = \mathbf{f}(\mathbf{z}, t)dt + g(t)d\mathbf{w}$
- Fokker-Plank equation:

$$\frac{\partial q\left(\mathbf{z}_t\right)}{\partial t} = \nabla_{\mathbf{z}_t}\left(-\mathbf{f}(\mathbf{z}, t)q\left(\mathbf{z}_t\right) + \frac{1}{2}g^2(t)q\left(\mathbf{z}_t\right)\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)\right)$$

$$= \nabla_{\mathbf{z}_t}\left(\mathbf{h}_q\left(\mathbf{z}_t, t\right)q\left(\mathbf{z}_t\right)\right)$$

- Backward SDE: $d\mathbf{z} = [\mathbf{f}(\mathbf{z}, t) - \nabla_{\mathbf{z}}\log q\left(\mathbf{z}\right)]dt + g(t)d\mathbf{w}$

LSGM's objective - I

Fokker-Plank equation recap

- Consider forward SDE: $d\mathbf{z} = \mathbf{f}(\mathbf{z}, t)dt + g(t)d\mathbf{w}$
- Fokker-Plank equation:

$$\frac{\partial q\left(\mathbf{z}_t\right)}{\partial t} = \nabla_{\mathbf{z}_t}\left(-\mathbf{f}(\mathbf{z}, t)q\left(\mathbf{z}_t\right) + \frac{1}{2}g^2(t)q\left(\mathbf{z}_t\right)\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)\right)$$

$$= \nabla_{\mathbf{z}_t}\left(\mathbf{h}_q\left(\mathbf{z}_t, t\right) q\left(\mathbf{z}_t\right)\right)$$

- Backward SDE: $d\mathbf{z} = [\mathbf{f}(\mathbf{z}, t) - \nabla_{\mathbf{z}}\log q\left(\mathbf{z}\right)]dt + g(t)d\mathbf{w}$
- $\mathbf{h}_q\left(\mathbf{z}_t, t\right) = -\mathbf{f}(\mathbf{z}, t) + \frac{1}{2}g^2(t)\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)$

LSGM's objective - I

CE's partial derivative - I

- Rewrite CE with Newton-Leibnitz Formula:

$$\mathrm{CE}\left(q\left(\mathbf{z}_0\right)\|p\left(\mathbf{z}_0\right)\right)=\mathrm{CE}\left(q\left(\mathbf{z}_1\right)\|p\left(\mathbf{z}_1\right)\right)+\int_1^0\frac{\partial}{\partial t}\mathrm{CE}\left(q\left(\mathbf{z}_t\right)\|p\left(\mathbf{z}_t\right)\right)dt=\mathrm{H}\left(q\left(\mathbf{z}_1\right)\right)-\int_0^1\frac{\partial}{\partial t}\mathrm{CE}\left(q\left(\mathbf{z}_t\right)\|p\left(\mathbf{z}_t\right)\right)dt$$

- Assume, that at $t=1, q\left(\mathbf{z}_1\right)=p\left(\mathbf{z}_1\right)$

$$\frac{\partial}{\partial t}\mathrm{CE}\left(q\left(\mathbf{z}_t\right)\|p\left(\mathbf{z}_t\right)\right)=-\frac{\partial}{\partial t}\int q\left(\mathbf{z}_t\right)\log p\left(\mathbf{z}_t\right)d\mathbf{z}=-\int\left[\frac{\partial q\left(\mathbf{z}_t\right)}{\partial t}\log p\left(\mathbf{z}_t\right)+\frac{q\left(\mathbf{z}_t\right)}{p\left(\mathbf{z}_t\right)}\frac{\partial p\left(\mathbf{z}_t\right)}{\partial t}\right]d\mathbf{z}$$

- $$=-\int\left[\nabla_{\mathbf{z}_t}\left(\mathbf{h}_q\left(\mathbf{z}_t,t\right)q\left(\mathbf{z}_t\right)\right)\log p\left(\mathbf{z}_t\right)+\frac{q\left(\mathbf{z}_t\right)}{p\left(\mathbf{z}_t\right)}\nabla_{\mathbf{z}_t}\left(\mathbf{h}_p\left(\mathbf{z}_t,t\right)p\left(\mathbf{z}_t\right)\right)\right]d\mathbf{z}$$

- $\mathbf{h}_q\left(\mathbf{z}_t,t\right)=-\mathbf{f}(\mathbf{z},t)+\frac{1}{2}g^2(t)\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)$

LSGM's objective - I

CE's partial derivative - II

$$\frac{\partial}{\partial t}\mathrm{CE}\left(q\left(\mathbf{z}_t\right)\|p\left(\mathbf{z}_t\right)\right) = -\int\left[\nabla_{\mathbf{z}_t}\left(\mathbf{h}_q\left(\mathbf{z}_t,t\right)q\left(\mathbf{z}_t\right)\right)\log p\left(\mathbf{z}_t\right)+\frac{q\left(\mathbf{z}_t\right)}{p\left(\mathbf{z}_t\right)}\nabla_{\mathbf{z}_t}\left(\mathbf{h}_p\left(\mathbf{z}_t,t\right)p\left(\mathbf{z}_t\right)\right)\right]d\mathbf{z}$$

$$= \int\left[\mathbf{h}_q\left(\mathbf{z}_t,t\right)^\top q\left(\mathbf{z}_t\right)\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)+\mathbf{h}_p\left(\mathbf{z}_t,t\right)^\top p\left(\mathbf{z}_t\right)\nabla_{\mathbf{z}_t}\frac{q\left(\mathbf{z}_t\right)}{p\left(\mathbf{z}_t\right)}\right]d\mathbf{z}$$

$$= \int\left[\mathbf{h}_q\left(\mathbf{z}_t,t\right)^\top q\left(\mathbf{z}_t\right)\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)+\mathbf{h}_p\left(\mathbf{z}_t,t\right)^\top p\left(\mathbf{z}_t\right)\frac{p\left(\mathbf{z}_t\right)\nabla_{\mathbf{z}_t}q\left(\mathbf{z}_t\right)-q\left(\mathbf{z}_t\right)\nabla_{\mathbf{z}_t}p\left(\mathbf{z}_t\right)}{p\left(\mathbf{z}_t\right)^2}\right]d\mathbf{z}$$

$$= \int q\left(\mathbf{z}_t\right)\left[\mathbf{h}_q\left(\mathbf{z}_t,t\right)^\top\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)+\mathbf{h}_p\left(\mathbf{z}_t,t\right)^\top\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)-\mathbf{h}_p\left(\mathbf{z}_t,t\right)^\top\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)\right]d\mathbf{z}$$

LSGM's objective - I

CE's partial derivative - III

$$\frac{\partial}{\partial t}\mathrm{CE}\left(q\left(\mathbf{z}_t\right)\|p\left(\mathbf{z}_t\right)\right)=\int q\left(\mathbf{z}_t\right)\left[\mathbf{h}_q\left(\mathbf{z}_t,t\right)^{\top}\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)+\mathbf{h}_p\left(\mathbf{z}_t,t\right)^{\top}\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)-\mathbf{h}_p\left(\mathbf{z}_t,t\right)^{\top}\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)\right]d\mathbf{z}$$

$$\mathbf{h}_q\left(\mathbf{z}_t,t\right)=-\mathbf{f}(\mathbf{z},t)+\frac{1}{2}g^2(t)\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)\qquad\mathbf{h}_p\left(\mathbf{z}_t,t\right)=-\mathbf{f}(\mathbf{z},t)+\frac{1}{2}g^2(t)\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)$$

$$\begin{aligned}\frac{\partial}{\partial t}\mathrm{CE}\left(q\left(\mathbf{z}_t\right)\|p\left(\mathbf{z}_t\right)\right)=\int q\left(\mathbf{z}_t\right)\Bigg[&-\mathbf{f}(\mathbf{z},t)^{\top}\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)+\frac{1}{2}g^2(t)\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)^{\top}\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)+\\&-\mathbf{f}(\mathbf{z},t)^{\top}\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)+\frac{1}{2}g^2(t)\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)^{\top}\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)+\\&+\mathbf{f}(\mathbf{z},t)^{\top}\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)-\frac{1}{2}g^2(t)\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)^{\top}\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)\Bigg]d\mathbf{z}\end{aligned}$$

LSGM's objective - I

CE's partial derivative - III

$$\frac{\partial}{\partial t}\mathrm{CE}\left(q\left(\mathbf{z}_t\right)\|p\left(\mathbf{z}_t\right)\right)=\int q\left(\mathbf{z}_t\right)\left[\mathbf{h}_q\left(\mathbf{z}_t,t\right)^{\top}\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)+\mathbf{h}_p\left(\mathbf{z}_t,t\right)^{\top}\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)-\mathbf{h}_p\left(\mathbf{z}_t,t\right)^{\top}\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)\right]d\mathbf{z}$$

$$\mathbf{h}_q\left(\mathbf{z}_t,t\right)=-\mathbf{f}(\mathbf{z},t)+\frac{1}{2}g^2(t)\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)\qquad\mathbf{h}_p\left(\mathbf{z}_t,t\right)=-\mathbf{f}(\mathbf{z},t)+\frac{1}{2}g^2(t)\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)$$

$$\begin{aligned}\frac{\partial}{\partial t}\mathrm{CE}\left(q\left(\mathbf{z}_t\right)\|p\left(\mathbf{z}_t\right)\right)=\int q\left(\mathbf{z}_t\right)\Big[&-\mathbf{f}(\mathbf{z},t)^{\top}\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)+\frac{1}{2}g^2(t)\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)^{\top}\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)+\\&-\mathbf{f}(\mathbf{z},t)^{\top}\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)+\frac{1}{2}g^2(t)\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)^{\top}\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)+\\&+\mathbf{f}(\mathbf{z},t)^{\top}\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)-\frac{1}{2}g^2(t)\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)^{\top}\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)\Big]d\mathbf{z}\end{aligned}$$

LSGM's objective - II

Score of $q\left(\mathbf{z}_t\right)$ - II

$$\mathrm{CE}\left(q\left(\mathbf{z}_0\right)\|p\left(\mathbf{z}_0\right)\right)=\mathrm{CE}\left(q\left(\mathbf{z}_1\right)\|p\left(\mathbf{z}_1\right)\right)+\int_1^0\frac{\partial}{\partial t}\mathrm{CE}\left(q\left(\mathbf{z}_t\right)\|p\left(\mathbf{z}_t\right)\right)dt=\mathrm{H}\left(q\left(\mathbf{z}_1\right)\right)+\int_0^1\frac{\partial}{\partial t}\mathrm{CE}\left(q\left(\mathbf{z}_t\right)\|p\left(\mathbf{z}_t\right)\right)dt$$

$$=\mathrm{H}\left(q\left(\mathbf{z}_1\right)\right)+\int_0^1\mathbb{E}_{q(\mathbf{z}_t)}\left[\frac{1}{2}g^2(t)\left\|\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)\right\|_2^2+2\mathbf{f}\left(\mathbf{z}_t,t\right)^\top\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)-g^2(t)\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)^\top\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)\right]dt$$

$$=\mathrm{H}\left(q\left(\mathbf{z}_1\right)\right)+\int_0^1\mathbb{E}_{q(\mathbf{z}_t)}\frac{1}{2}\left(g^2(t)\left[\left\|\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)\right\|_2^2-2\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)^\top\nabla_{\mathbf{z}_t}\log p\left(\mathbf{z}_t\right)\right]+2\mathbf{f}\left(\mathbf{z}_t,t\right)^\top\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)\right)dt$$

$$\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)=\frac{\nabla_{\mathbf{z}_t}q\left(\mathbf{z}_t\right)}{q\left(\mathbf{z}_t\right)}=\frac{\nabla_{\mathbf{z}_t}\int q\left(\mathbf{z}_0\right)q\left(\mathbf{z}_t\,|\,\mathbf{z}_0\right)d\mathbf{z}_0}{q\left(\mathbf{z}_t\right)}=\frac{\int q\left(\mathbf{z}_0\right)\nabla_{\mathbf{z}_t}q\left(\mathbf{z}_t\,|\,\mathbf{z}_0\right)d\mathbf{z}_0}{q\left(\mathbf{z}_t\right)}=\frac{\int q\left(\mathbf{z}_0\right)q\left(\mathbf{z}_t\,|\,\mathbf{z}_0\right)\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\,|\,\mathbf{z}_0\right)d\mathbf{z}_0}{q\left(\mathbf{z}_t\right)}$$

$$=\mathbb{E}_{q(\mathbf{z}_0\,|\,\mathbf{z}_t)}\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\,|\,\mathbf{z}_0\right)$$

$$\mathbb{E}_{q(\mathbf{z}_t)}\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\right)=\mathbb{E}_{q(\mathbf{z}_t)}\mathbb{E}_{q(\mathbf{z}_0\,|\,\mathbf{z}_t)}\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\,|\,\mathbf{z}_0\right)=\mathbb{E}_{q(\mathbf{z}_0)}\mathbb{E}_{q(\mathbf{z}_t\,|\,\mathbf{z}_0)}\nabla_{\mathbf{z}_t}\log q\left(\mathbf{z}_t\,|\,\mathbf{z}_0\right)$$

LSGM's objective - III

in search of elements that do not depend $p\left(\mathbf{z}_t\right)$

$$
\begin{aligned}
&\mathrm{CE}\left(q\left(\mathbf{z}_0\right) \| p\left(\mathbf{z}_0\right)\right) = \\
&= \mathrm{H}\left(q\left(\mathbf{z}_1\right)\right) + \int_0^1 \mathbb{E}_{q(\mathbf{z}_t)} \frac{1}{2}\left(g^2(t)\left[\left\| \nabla_{\mathbf{z}_t} \log p\left(\mathbf{z}_t\right) \right\|_2^2 - 2 \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t\right)^{\top} \nabla_{\mathbf{z}_t} \log p\left(\mathbf{z}_t\right)\right] + 2\mathbf{f}\left(\mathbf{z}_t, t\right)^{\top} \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t\right)\right) dt \\
&= \mathrm{H}\left(q\left(\mathbf{z}_1\right)\right) + \int_0^1 \mathbb{E}_{q(\mathbf{z}_t, \mathbf{z}_0)} \frac{1}{2}\left(g^2(t)\left[\left\| \nabla_{\mathbf{z}_t} \log p\left(\mathbf{z}_t\right) \right\|_2^2 - 2 \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t \mid \mathbf{z}_0\right)^{\top} \nabla_{\mathbf{z}_t} \log p\left(\mathbf{z}_t\right)\right] + 2\mathbf{f}\left(\mathbf{z}_t, t\right)^{\top} \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t \mid \mathbf{z}_0\right)\right) dt \\
&\int_0^1 \mathbb{E}_{q(\mathbf{z}_t, \mathbf{z}_0)} \frac{1}{2}\left(g^2(t) \left\| \nabla_{\mathbf{z}_t} \log p\left(\mathbf{z}_t\right) - \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t \mid \mathbf{z}_0\right) \right\|_2^2 + 2\mathbf{f}\left(\mathbf{z}_t, t\right)^{\top} \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t \mid \mathbf{z}_0\right) - g^2(t) \left\| \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t \mid \mathbf{z}_0\right) \right\|_2^2\right) dt \\
= &\int_0^1 \mathbb{E}_{q(\mathbf{z}_t, \mathbf{z}_0)} \frac{1}{2}\left(g^2(t) \left\| \nabla_{\mathbf{z}_t} \log p\left(\mathbf{z}_t\right) - \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t \mid \mathbf{z}_0\right) \right\|_2^2 + 2\mathbf{f}\left(\mathbf{z}_t, t\right)^{\top} \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t \mid \mathbf{z}_0\right) - g^2(t) \left\| \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t \mid \mathbf{z}_0\right) \right\|_2^2\right) dt
\end{aligned}
$$

LSGM's objective - IV

Constant even for encoder

$$\int_0^1 \mathbb{E}_{q(\mathbf{z}_t, \mathbf{z}_0)} \frac{1}{2} \left(2\mathbf{f}\left(\mathbf{z}_t, t\right)^\top \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t \,|\, \mathbf{z}_0\right) - g^2(t) \left\| \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t \,|\, \mathbf{z}_0\right) \right\|_2^2 \right) dt$$

$$d\mathbf{z} = \mathbf{f}(\mathbf{z}, t)dt + g(t)d\mathbf{w} = f(t)\mathbf{z}_t dt + g(t)d\mathbf{w} \qquad q(\mathbf{z}_t \,|\, \mathbf{z}_0) = \mathcal{N}(\mathbf{z}_t \,|\, \alpha_t \mathbf{z}_0, \sigma_t^2 I) \qquad \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t \,|\, \mathbf{z}_0\right) = -\frac{\epsilon}{\sigma_t}$$

$$\int_0^1 \mathbb{E}_{q(\mathbf{z}_t, \mathbf{z}_0)} \frac{1}{2} \left(2\mathbf{f}\left(\mathbf{z}_t, t\right) - g^2(t) \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t \,|\, \mathbf{z}_0\right) \right)^\top \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t \,|\, \mathbf{z}_0\right) dt$$

$$\int_0^1 \mathbb{E}_{q(\mathbf{z}_0), \epsilon} \frac{1}{2} \left(2f(t)(\alpha_t \mathbf{z}_0 + \sigma_t \epsilon) + g^2(t) \frac{\epsilon}{\sigma_t} \right)^\top (-\frac{\epsilon}{\sigma_t}) dt = \int_0^1 -\frac{f(t)}{\sigma_t} \underbrace{\mathbb{E}_{q(\mathbf{z}_0), \epsilon} \left[\alpha_t \left(\mathbf{z}_0\right)^T \epsilon \right]}_{=0} - \frac{2f(t)\sigma_t^2 + g(t)^2}{2\sigma_t^2} \underbrace{\mathbb{E}_\epsilon \left[\epsilon^T \epsilon \right]}_{=D} dt$$

$$= -\frac{D}{2} \int_0^1 \frac{2f(t)\sigma_t^2 + g(t)^2}{\sigma_t^2} dt = \frac{D}{2} \left(\log \sigma_0^2 - \log \sigma_1^2 \right)$$

LSGM's objective - V

CE - final

$$\mathrm{CE}\left(q\left(\mathbf{z}_0\right) \| p\left(\mathbf{z}_0\right)\right) = \mathrm{H}\left(q\left(\mathbf{z}_1\right)\right) + \int_0^1 \mathbb{E}_{q(\mathbf{z}_t, \mathbf{z}_0)} \frac{1}{2}\left(g^2(t) \left\| \nabla_{\mathbf{z}_t} \log p\left(\mathbf{z}_t\right) - \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t \mid \mathbf{z}_0\right) \right\|_2^2\right) dt + \frac{D}{2}\left(\log \sigma_0^2 - \log \sigma_1^2\right)$$

$$\mathrm{H}\left(q\left(\mathbf{z}_1\right)\right) = \int q\left(\mathbf{z}_1\right) \log q\left(\mathbf{z}_1\right) d\mathbf{z}_1 = \frac{D}{2} \log\left(2\pi e \sigma_1^2\right)$$

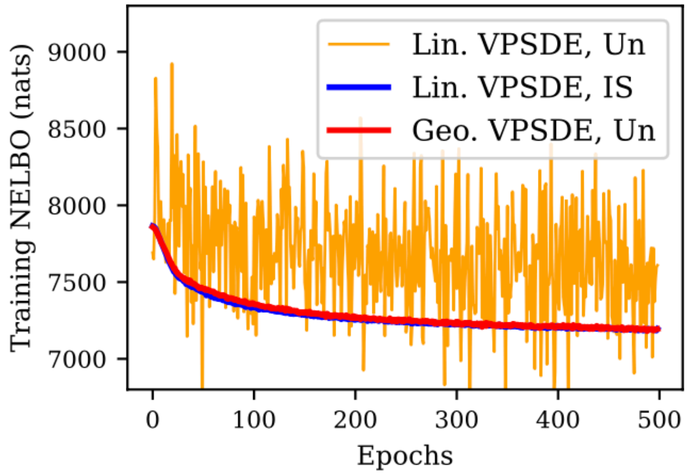
$$q\left(\mathbf{z}_1\right) = \mathcal{N}\left(\mathbf{z}_1 \mid 0, \sigma_1^2 I\right)$$

$$\mathrm{CE}\left(q\left(\mathbf{z}_0 \mid \mathbf{x}\right) \| p\left(\mathbf{z}_0\right)\right) = \mathbb{E}_{t \sim \mathcal{U}[0,1]}\left[\frac{g(t)^2}{2} \mathbb{E}_{q(\mathbf{z}_t, \mathbf{z}_0 \mid \mathbf{x})}\left[\left\| \nabla_{\mathbf{z}_t} \log q\left(\mathbf{z}_t \mid \mathbf{z}_0\right) - \nabla_{\mathbf{z}_t} \log p\left(\mathbf{z}_t\right) \right\|_2^2\right]\right] + \frac{D}{2} \log\left(2\pi e \sigma_0^2\right)$$

Things, I didn’t say

…or a thousand more formulas

- New parametrization for score-estimation: $p\left(z_t\right) \propto \mathcal{N}\left(z_t; 0,1\right)^{1-\alpha} p'_{\boldsymbol{\theta}}\left(z_t\right)^{\alpha}$
  - So new score: $\nabla_{z_t} \log p\left(z_t\right) = -(1-\alpha) z_t + \alpha \nabla_{z_t} \log p'_{\boldsymbol{\theta}}\left(z_t\right)$
- Several ways for reducing loss’s dispersion:
  - Importance sampling for weighting
  - New Geometric-SDEs



LSGM experiments

Table 2: Generative performance on CIFAR-10.

	Method	NLL↓	FID↓
Ours	LSGM (FID)	≤3.43	**2.10**
	LSGM (NLL)	≤**2.87**	6.89
	LSGM (balanced)	≤2.95	2.17
	VAE Backbone	2.96	43.18
VAEs	VDVAE [21]	**2.87**	-
	NVAE [20]	2.91	23.49
	VAEBM [76]	-	12.19
	NCP-VAE [56]	-	24.08
	BIVA [48]	3.08	-
	DC-VAE [77]	-	17.90
Score	NCSN [3]	-	25.32
	Rec. Likelihood [40]	3.18	9.36
	DSM-ALS [39]	3.65	-
	DDPM [1]	3.75	3.17
	Improved DDPM [26]	2.94	11.47
	SDE (DDPM++) [2]	2.99	2.92
	SDE (NCSN++) [2]	-	2.20

Table 3: Generative results on CelebA-HQ-256.

	Method	NLL↓	FID↓
Ours	LSGM	≤**0.70**	**7.22**
	VAE Backbone	**0.70**	30.87
VAEs	NVAE [20]	**0.70**	29.76
	VAEBM [76]	-	20.38
	NCP-VAE [56]	-	24.79
	DC-VAE [77]	-	15.80
Score	SDE [2]	-	**7.23**
Flows	GLOW [85]	1.03	68.93
Aut. Reg.	SPN [86]	0.61	-
GANs	Adv. LAE [87]	-	19.21
	VQ-GAN [64]	-	10.70
	PGGAN [88]	-	8.03

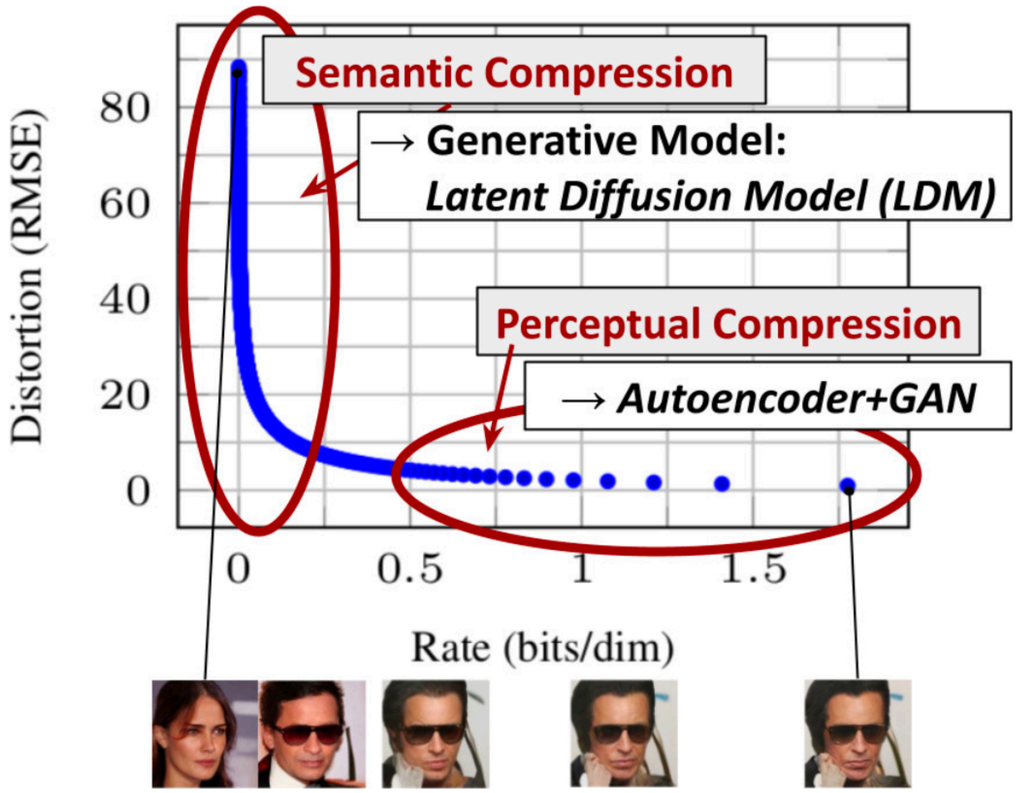
(b) CelebA-HQ-256

Towards stability

Key idea

Divide generative model's training into 2 stages:

- Train to restoring high-frequency details
- Train to generate various and hard semantics

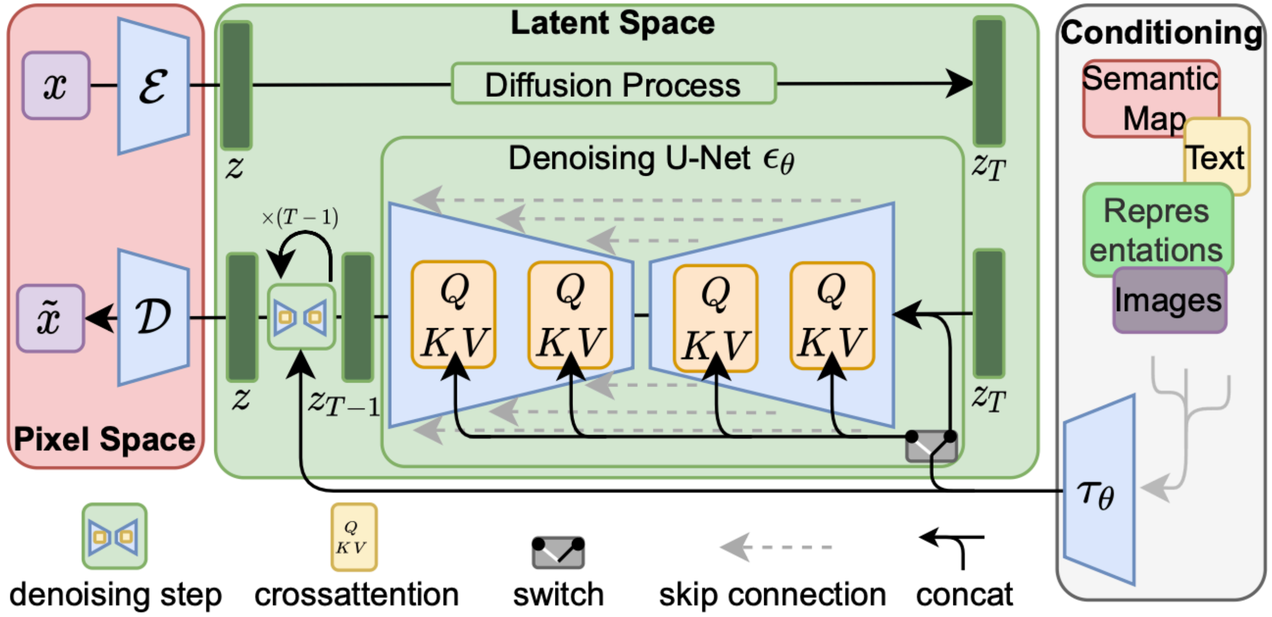


So let's use fixed latent space

- Fixed good encoder-decoder for reconstruction high-frequency details
- New more suitable "pixel" space for learning semantics

Whole scheme

For training authors used 256x256 images



$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right) \cdot V,$$

$$Q = W_Q^{(i)} \cdot \varphi_i\left(z_t\right),$$

$$K = W_K^{(i)} \cdot \tau_\theta(y),$$

$$V = W_V^{(i)} \cdot \tau_\theta(y)$$

$$L_{LDM} := \mathbb{E}_{\mathcal{E}(x),\epsilon\sim\mathcal{N}(0,1),t}\left[\|\epsilon - \epsilon_\theta(z_t, t)\|_2^2\right]$$

$$L_{LDM} := \mathbb{E}_{\mathcal{E}(x),y,\epsilon\sim\mathcal{N}(0,1),t}\left[\|\epsilon - \epsilon_\theta(z_t, t, \tau_\theta(y))\|_2^2\right]$$

High-resolution image synthesis

General approach for high-resolution generation -
Sliding window with 256x256 crops

- High-resolution layouts:
  - $\tau_\theta(y)$ - Identity
  - y - semantic mask
- Super-resolution
  - $\tau_\theta(y)$ - Identity
  - y - downsampled image

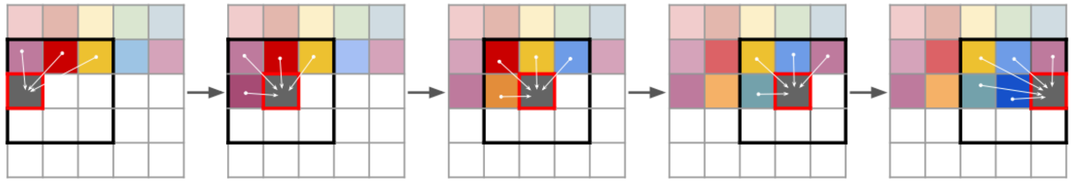


Figure 3. Sliding attention window.

Stable Diffusion's experiments

Text-to-image

Text-Conditional Image Synthesis				
Method	FID ↓	IS↑	Nparams	
CogView† [17]	27.10	18.20	4B	self-ranking, rejection rate 0.017
LAFITE† [109]	26.94	26.02	75M	
GLIDE* [59]	12.24	-	6B	277 DDIM steps, c.f.g. [32] $s = 3$
Make-A-Scene* [26]	**11.84**	-	4B	c.f.g for AR models [98] $s = 5$
LDM-KL-8	23.31	$20.03_{\pm 0.33}$	1.45B	250 DDIM steps
*LDM-KL-8-G**	12.63	$\mathbf{30.29}_{\pm \mathbf{0.42}}$	1.45B	250 DDIM steps, c.f.g. [32] $s = 1.5$

Table 2. Evaluation of text-conditional image synthesis on the 256 × 256-sized MS-COCO [51] dataset: with 250 DDIM [84] steps our model is on par with the most recent diffusion [59] and autoregressive [26] methods despite using significantly less parameters. †/*:Numbers from [109]/ [26]

Text-to-Image Synthesis on LAION. 1.45B Model.

Stable Diffusion’s experiments

Super-resolution (1024x1024)

bicubic

LDM-BSR

Conclusion

- Saw diffusion in latent space
  - With simultaneously training - derived non-trivial training objective for this approach
    - End-to-end
    - Iterative optimization
  - Saw what models in fixed latent space can achieve using crossattention with prompts