Diffusion : 어떤 데이터에 랜덤한 noise를 조금씩 뿌려서 완전한 noise를 만들 수 있다면 역으로 완전 복원도 가능!

DDPM

https://cvpr2022-tutorial-diffusion-models.github.io/

Gaussian Noise를 조금씩 추가 → 아주 짧은 시간 내에서 역도 Gaussian 분포

$$q(x_t|x_{t-1}) = N(x_t; \sqrt{1-\beta_t}\, x_{t-1}, \beta_t I)$$

β_t : 조금씩 증가하는 작은 상수

처음에는 $N(x_{t-1}, 0)$ ⇝ 완전한 가우시안 노이즈 $N(0, I)$

Gaussian Noise

Forward ↓

Data ⇄ Noise

$x_0 \quad x_1 \quad \cdots \quad x_T$

Reverse ↑

Gaussian Noise

$$P(x_t|x_{t-1}) = N(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2 I)$$

$\mu_\theta(x_t, t)$: trainable network ↓ denoising

* Forward Process : $q(x_T|x_{T-1:0})\, q(x_{T-1}|x_{T-2:0}) \cdots q(x_1|x_0)$

학습X

$$= \frac{q(x_{1:T})}{q(x_0)} = q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1})$$

$q(x_{1:T}|x_0)$: data가 주어졌을때 noise ⋯

$\prod$: Markov Chain

t에 따라 sequential할 필요X

x_t를 한번에 생성하도록

※ Reparameterization Trick

$x \sim N(\mu, \sigma^2)$

$\rightarrow x = \mu + \sigma\epsilon \; (\epsilon \sim N(0,1))$

$$N(x_t; \underbrace{\sqrt{1-\beta_t}}_{\alpha_t}\, x_{t-1}, \underbrace{\beta_t}_{1-\alpha_t} I) = N(x_t; \sqrt{\alpha_t}\, x_{t-1}, (1-\alpha_t) I)$$

$$\begin{aligned}
x_t &= \sqrt{\alpha_t}\, x_{t-1} + \sqrt{1-\alpha_t}\, \epsilon_{t-1} \\
&= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}\, x_{t-2} + \sqrt{1-\alpha_{t-1}}\, \epsilon_{t-2}\right) + \sqrt{1-\alpha_t}\, \epsilon_{t-1} \\
&= \sqrt{\alpha_t \alpha_{t-1}}\, x_{t-2} + \sqrt{\alpha_t(1-\alpha_{t-1})}\, \epsilon_{t-2} + \sqrt{1-\alpha_t}\, \epsilon_{t-1} \\
&= \sqrt{\alpha_t \alpha_{t-1}}\, x_{t-2} + \sqrt{\alpha_t(1-\alpha_{t-1}) + 1-\alpha_t}\, \bar{\epsilon}_{t-2} \\
&= \sqrt{\alpha_t \alpha_{t-1}}\, x_{t-2} + \sqrt{1-\alpha_t \alpha_{t-1}}\, \bar{\epsilon}_{t-2} \\
&= \cdots \\
&= \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon
\end{aligned}$$

$\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$

$$\therefore q(x_t|x_0) = N(x_t; \sqrt{\bar{\alpha}_t}\, x_0, (1-\bar{\alpha}_t) I)$$

※ Moment Generating Function

$\mu_n = \mathbb{E}\{X^n\}$

$$M(t) = \mathbb{E}\{e^{tx}\} = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$

$$= \mathbb{E}\{1 + tx + \frac{t^2x^2}{2!} + \cdots\} : \text{Taylor Series}$$

$$= 1 + t\,\mathbb{E}\{x\} + \frac{t^2}{2!}\mathbb{E}\{x^2\} + \cdots$$

$$= 1 + t\mu_1 + \frac{t^2}{2!}\mu_2 + \cdots$$

$X \sim M_X(t)$, $Y \sim M_Y(t)$

$M_{X+Y}(t) = M_X(t) M_Y(t)$

* Reverse Process $P_\theta(x_{0:T}) = P(x_T)\, P_\theta(x_{T-1}|x_T) \cdots P_\theta(x_0|x_1)$

train!

$$= P(x_T) \prod_{t=1}^{T} P_\theta(x_{t-1}|x_t)$$

$$P(x_T) = \mathcal{N}(x_T; 0, I)$$

$$P_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \underline{\mu_\theta(x_t, t)}, \Sigma_\theta(x_t, t))$$

trainable network

U-net. denoising autoencoder

⇒ Training

※ Variational Autoencoder의 loss (VAE)

$$\mathbb{E}_{z \sim q_\phi(z|x)} \{\log P_\theta(x)\} = \mathbb{E}_z \left\{ \log \frac{P_\theta(x|z) P_\theta(z)}{P_\theta(z|x)} \cdot \frac{q_\phi(z|x)}{q_\phi(z|x)} \right\}$$

$$= \mathbb{E}_z \left\{ \log P_\theta(x|z) - \log \frac{q_\phi(z|x)}{P_\theta(z)} + \log \frac{q_\phi(z|x)}{P_\theta(z|x)} \right\}$$

$$= \underbrace{\mathbb{E}_z \{\log P_\theta(x|z)\} - D_{KL}(q_\phi(z|x) \,\|\, P_\theta(z))}_{\text{계산가능} \to \text{maximize!} \; = \text{ELBO}} + D_{KL}(q_\phi(z|x) \,\|\, \underline{P_\theta(z|x)})$$

계산불가능

직접계산은 X
KLD≥0으로
생략 → ELBO ⇑

→ Negative Log Likelihood

$$\mathbb{E}_q(-\log P_\theta(X_0)) = \mathbb{E}_q \left\{ -\log \frac{P_\theta(X_{0:T})}{P_\theta(X_{1:T}|X_0)} \cdot \frac{q(X_{1:T}|X_0)}{q(X_{1:T}|X_0)} \right\}$$

$$= \mathbb{E}_q \left\{ -\log \frac{P_\theta(X_{0:T})}{q(X_{1:T}|X_0)} - \log \frac{q(X_{1:T}|X_0)}{P_\theta(X_{1:T}|X_0)} \right\}$$

$$= \mathbb{E}_q \left\{ -\log \frac{P_\theta(X_{0:T})}{q(X_{1:T}|X_0)} \right\} - \underbrace{D_{KL}(q(X_{1:T}|X_0) \,\|\, P_\theta(X_{1:T}|X_0))}_{KLD \geq 0}$$

$$\leq \mathbb{E}_q \left\{ -\log \frac{P_\theta(X_{0:T})}{q(X_{1:T}|X_0)} \right\}$$

$$= \mathbb{E}_q \left\{ -\log P_\theta(X_T) - \sum_{t=1}^{T} \log \frac{P_\theta(X_{t-1}|X_t)}{q(X_t|X_{t-1})} \right\}$$

$$= \mathbb{E}_q \left\{ -\log P_\theta(X_T) - \sum_{t=2}^{T} \log \frac{P_\theta(X_{t-1}|X_t)}{q(X_t|X_{t-1})} - \log \frac{P_\theta(X_0|X_1)}{q(X_1|X_0)} \right\}$$

$$q(X_t|X_{t-1}) = q(X_t|X_{t-1}, X_0) \to \frac{q(X_t|X_{t-1}, X_0)\, q(X_{t-1}|X_0)}{q(X_t|X_0)} = q(X_{t-1}|X_t)$$

$$= \mathbb{E}_q \left\{ -\log P_\theta(X_T) - \sum_{t=2}^{T} \log \frac{P_\theta(X_{t-1}|X_t)}{q(X_{t-1}|X_t)} \cdot \frac{q(X_{t-1}|X_0)}{q(X_t|X_0)} - \log \frac{P_\theta(X_0|X_1)}{q(X_1|X_0)} \right\}$$

$$= \mathbb{E}_q \left\{ -\log P_\theta(X_T) - \sum_{t=2}^{T} \log \frac{P_\theta(X_{t-1}|X_t)}{q(X_{t-1}|X_t)} - \log \frac{P_\theta(X_0|X_1)}{q(X_T|X_0)} \right\}$$

$$= \mathbb{E}_q \left\{ -\log \frac{P_\theta(X_T)}{q(X_T|X_0)} - \sum_{t>1} \log \frac{P_\theta(X_{t-1}|X_t)}{q(X_{t-1}|X_t)} - \log P_\theta(X_0|X_1) \right\}$$

얼마나 잘복원하는가

$$= \underbrace{\cancel{D_{KL}(q(X_T|X_0) \,\|\, P_\theta(X_T))}}_{L_T \text{ 고정!}} + \sum_{t>1} \underbrace{D_{KL}(q(X_{t-1}|X_t) \,\|\, P_\theta(X_{t-1}|X_t))}_{L_{t-1}} - \underbrace{\cancel{\mathbb{E}_q \{\log P_\theta(X_0|X_1)\}}}_{L_0 \text{ 매우작음}}$$

* Loss : $D_{KL}(\underbrace{q(x_{t-1}|x_t)}_{\text{평균과 분산을 바로 알수없다!}} \,\|\, p_\theta(x_{t-1}|x_t))$

평균과 분산을 바로 알수없다!

$q(x_t|x_{t-1}) = N(x_t; \sqrt{1-\beta_t}\,x_{t-1}, \beta_t I)$

$= N(x_t; \sqrt{\alpha_t}\,x_{t-1}, \beta_t I)$

$N(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

$q(x_t|x_0) = N(x_t; \sqrt{\bar\alpha_t}, (1-\bar\alpha_t)I)$

① $q(x_{t-1}|x_t)\,q(x_t|x_0) = q(x_t|x_{t-1})\,q(x_{t-1}|x_0)$

$$q(x_{t-1}|x_t) = q(x_t|x_{t-1})\frac{q(x_{t-1}|x_0)}{q(x_t|x_0)}$$

$$\propto \exp\left(-\frac{(x_t-\sqrt{\alpha_t}x_{t-1})^2}{2\beta_t}\right)\frac{\exp\left(\frac{(x_{t-1}-\sqrt{\bar\alpha_{t-1}}x_0)^2}{2(1-\bar\alpha_{t-1})}\right)}{\exp\left(\frac{(x_t-\sqrt{\bar\alpha_t}x_0)^2}{2(1-\bar\alpha_t)}\right)}$$

x_{t-1}에 대해 정리 & 나머지 상수처리

$$= \exp\left(-\frac{1-\bar\alpha_t}{2\beta_t(1-\bar\alpha_{t-1})}\left(x_{t-1}^2 - 2\left(\frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})}{1-\bar\alpha_t}x_t + \frac{\sqrt{\bar\alpha_{t-1}}\beta_t}{1-\bar\alpha_t}x_0\right)x_{t-1}\right) + C(x_t, x_0)\right)$$

$$= \exp\left(-\frac{1}{2\sigma^2}(x^2 - 2\mu x + \mu^2)\right)$$

$$= N(x_{t-1}; \tilde\mu_t(x_t,x_0), \tilde\beta_t I)$$

$$\begin{cases} \tilde\mu_t(x_t,x_0) = \frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})}{1-\bar\alpha_t}x_t + \frac{\sqrt{\bar\alpha_{t-1}}\beta_t}{1-\bar\alpha_t}x_0 \\ \tilde\beta_t = \frac{1-\bar\alpha_{t-1}}{1-\bar\alpha_t}\beta_t \end{cases}$$

$x_t = \sqrt{\bar\alpha_t}\,x_0 + \sqrt{1-\bar\alpha_t}\,\epsilon$

$x_0 = \frac{x_t - \sqrt{1-\bar\alpha_t}\,\epsilon}{\sqrt{\bar\alpha_t}}$

$\bar\alpha_t = \prod_{s=1}^{t}\alpha_s$

$\beta_t = 1-\alpha_t$

$$= \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{(1-\alpha_t)}{\sqrt{1-\bar\alpha_t}}\epsilon\right)$$

② $p_\theta(x_{t-1}|x_t) = N(x_{t-1}; \mu_\theta(x_t,t), \Sigma_\theta(x_t,t))$

목적: q에 근사!

Σ_θ: β_t or $\frac{1-\bar\alpha_{t-1}}{1-\bar\alpha_t}$

$$\mu_\theta(x_t,t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{(1-\alpha_t)}{\sqrt{1-\bar\alpha_t}}\boxed{\epsilon_\theta(x_t,t)}\right)$$

입력 고정값 학습

③ $D_{KL}(q\|p) = \int q(x)\frac{q(x)}{p(x)}dx \xrightarrow[\text{분포}]{\text{가우시안}} D_{KL}(q\|p) = \log\frac{\sigma_p}{\sigma_q} + \frac{\sigma_q^2 + (\mu_q-\mu_p)^2}{2\sigma_p^2} - \frac{1}{2}$

평균만 취하고 나머지는 상수

$$D_{KL}(q(x_{t-1}|x_t,x_0)\,\|\,p_\theta(x_{t-1}|x_t)) = \frac{1}{2\sigma_t^2}\|\tilde\mu(x_t,x_0) - \mu_\theta(x_t,t)\|^2$$

$$= \frac{1}{2\sigma_t^2}\left\|\frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar\alpha_t}}\epsilon\right) - \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar\alpha_t}}\epsilon_\theta(x_t,t)\right)\right\|^2$$

$$= \frac{(1-\alpha_t)^2}{2\sigma_t^2\alpha_t(1-\bar\alpha_t)}\|(\epsilon - \epsilon_\theta(x_t,t))\|^2$$

x_t ← $\sqrt{\bar\alpha_t}\,x_0 + \sqrt{1-\bar\alpha_t}\,\epsilon$

$$\mathcal{L}_{t-1} = \mathbb{E}_{x_0,\epsilon}\left\{\cancel{\frac{(1-\alpha_t)^2}{2\sigma_t^2\alpha_t(1-\bar\alpha_t)}}\left\|\epsilon - \boxed{\epsilon_\theta(\sqrt{\bar\alpha_t}\,x_0 + \sqrt{1-\bar\alpha_t}\,\epsilon, t)}\right\|^2\right\}$$

상수

t에서 input에 추가된 noise 예측