Paper summary

AI VISION Lab

1. 공부한 논문의 제목, 게재된 학회 혹은 저널 등 논문 기본 정보를 적으세요.
   A. 이름: **Denoising Diffusion Probabilistic Models**
   B. 저널: NeurIPS
   C. 도메인: Diffusion Probabilistic Model
   D. 출판연도: 2020
   E. 저자: Jonathan Ho et al.

2. 논문에서 제안한 알고리즘 및 프레임워크에 대해 본인이 이해한대로 다이어그램을 그려보세요. **논문 Figure를 그대로 따라 그리면 안됩니다.**
   A. **해당 논문은 DPM(Diffusion Probabilistic Model)의 형태를 따르고 있음**

   B. [선행연구] DPM에서는 Forward Process, Reverse Process, ELBO 형태의 Loss를 제안했고, DDPM에서는 해당 연구를 기반으로 발전시키는 구조임.

   C. [선행연구] Forward Process의 경우, Noise를 주입해가는 과정으로, $x_0 \rightarrow x_T$ 결과적으로는 x_0의 분포를 붕괴시키는 과정임. 이때, Noise는 ~ N(0, I)를 따르며 xT ~ N(0, 1) 형태를 만들어 냄
   .

   D. [선행연구] Reverse Process의 경우, 앞서 붕괴한 xT를 시작으로 x0를 Sampling하는 과정을 의미하는데, 해당과정은 Intractable하다보니, Jensen's Inequality 기반의 ELBO 형태로 Loss를 구성해서 학습을 수행함.

   E. [선행연구] DBM에서는 이 과정을 3가지로 나눠서 계산했는데,
      i. L0 $p(x_0 \mid x_1)$의 경우, Edge Effect를 고려해서, 고정된 형태로 계산
      ii. LT $p(x_T)$는 Closed Form이기 때문에 계산이 간단함. -> Entropy형태
      iii. LT-1, 0, T를 제외한 부분으로 ELBO형태

   F. DDPM: Forward Processing과정의 β에 대해서 상수항으로 고정(학습X, 뿐만 아니라, 이 형태의 경우, LT은 무시할 수 있음(상수취급)
   G. $q(x_t \mid x_0) = N\left(xt; \sqrt{\bar{\alpha}}x_0, 1-\bar{\alpha}\right), x_0 \rightarrow x_t$의 Closed Form Distribution 추론
   H. [표기] $1 - \beta_t = \alpha_t, \bar{\alpha} = \Pi_{s=1}^{\mathrm{t}} \alpha_s$

I. Loss Function: L0의 경우, Edge Effect 로 인해, 학습이 잘못된 방향으로 갈 수 있다는 점을 예방하고자, 실제 상황을 가정하고 실험함. (Image input에 대해서, 각 Pixel 마다 구분해서 값을 구하고 적분하는 형태로 계산) 이때, 결과적으로, 연속적인 분포의 경우와 값이 비슷함을 증명했음. 즉, 이산적인 계산으로 Likelihood 값과 동일함을 보임.

J. LT-1의 경우, 기존의 DPM에서 $D_{KL}(q(x_{t-1} \mid x_t, x_0) \| p_\theta(x_{t-1} \mid x_t))$를 계산하는 형태로, x_0를 조건으로 제시해야만, Closed Form형태의 Posterior 값을 구할 수 있었음.

K. 이때, DDPM의 경우, 해당 과정을 L2 Loss 형태로 표현할 수 있음을 보임.

L. $E_q(\frac{1}{2\sigma^2}[\tilde{\mu_t}(x_t, x_0) - \mu_t(x_t, t)]^2)$

M. 여기서 더 나아가, Reparameterization 과정을 결합, Mean에 대한 예측이 아닌, $x_t(x_0, \epsilon) = \sqrt{\bar{\alpha}}x_0 + \sqrt{1-\bar{\alpha}}\epsilon$ 형태를 적용해, x0(고정)과 $\epsilon \sim N(0, I)$만으로 표현하는 구조를 추론함. 즉(KL Divergence -> Mean -> noise 예측)

N. 최종적으로, Simple Loss(L0, LT 삭제 및 시간에 따라 달라지는 noise의 scheduler를 제거한 Loss로, noise의 크기와 상관없이 동일한 Loss 값을 부여하는 구조임) 가 우수함을 실험적으로 보임으로써, 간단한 구조의 Loss Function을 선보임.

3. 본인이 생각하는 이 논문의 장점이 무엇이라고 생각하나요? **논문 Contribution bullet을 그대로 따라 적으면 안됩니다.**
   A. **가장 큰 장점은 Diffusion Loss를 ELBO 형태에서, Epsilon(Noise)를 예측하는 문제로 전환시켰다는 점임. 기존의 DPM 모델의 형태, L0, LT, LT-1을 발전시켰음과 동시에 Simple Loss를 제시했고, 현재 대부분(Hugging Face, GitHub 등)에서는 Simple Loss를 사용하고 있다는 점에서 큰 의미가 있음.**

   B. 뿐만 아니라, $x_t|x_0$의 Closed Form을 보여줌으로써, Forward process를 빠르게 통과할 수 있도록 함.

4. 이 논문을 읽으면서 느낀 점, 혹은 배운 점이 있으면 적어보세요.
   A. Reparameterization의 적용과정에 대해서 깊게 파악가능

5. 이 논문의 한계점이 있다면 무엇이라고 생각하나요?
   A. 여전히, Reverse process를 T~1 (Simple Loss 기준)번 실행해야 한다는 점은 Sampling과정을 오랫동안 기다려야 한다는 단점으로 작용함.

6. 본인의 연구에 접목시켜볼 점이 있을지 생각하고 적어보세요.
   A. Diffusion 선행연구

7. 본 Summary를 작성하는 과정에서 생성형AI를 사용했나요?
   A. 아니요

날짜: 2025-06-30

이름: 신준원

06/20 Denoising ~~Diffusion~~ Diffusion probabilistic model

keyword : VI (variational inference), denoising Score Matching, Langevin

DPM (Diffusion Probabilistic model - Dickstein et al 2015)

- VI (variational Inference) → intractable 한 문제를 해결하기 위한 근사방법.
  - → ELBO 사용
- Markov chain
- finite step.
- forward + reverse.
- Noise (to Gaussian)

model probability

① $p(X_T) = \mathcal{N}(X_T; 0, I)$ → forward process의 output

$$p_\theta(x_0) = \int \underbrace{p_\theta(x_{0:T})}_{②} \, dx_{1:T}$$

$$② \; p_\theta(x_{0:T}) = \underbrace{p(x_T)}_{①} \prod_{t=1}^{T} \underbrace{p_\theta(x_{t-1}|x_t)}_{③}$$

③ reverse kernel

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

$$E[-\log p_\theta(x_0)] \le E_q\left[-\log \frac{p_\theta(0:T)}{q(x_{1:T}|x_0)}\right] = \boxed{E_q\left[-\log p(x_T) - \sum_{t\ge1} \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_t|x_{t-1})}\right] = L}$$

intractable → (Jensen's inequality) ELBO 유도. (negative log likelihood)

☆ **forward ~~process~~ process (β_t)**

(β_t가 충분히 작을때!)

β_t → Constant, → learning. → 결과적으로는 둘다 reverse process가 Gaussian 하도록 만듦.

→ $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$

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

계산과정 probability.

$\leftrightarrow p \rightarrow$ model!

일반적인 q 계산 (non - markov) ⇒ 계산 불가능

$$p(x_0, x_1, x_2, \ldots, x_T) = p(x_0) \cdot p(x_1 | x_0) \cdot p(x_2 | x_1, \cancel{x_0}) \cdots p(x_T | x_{T-1}, \ldots, \cancel{x_2}, \cancel{x_1}, \cancel{x_0})$$

그러나 Diffusion ⇒ markov process를 따름. ($x_t \rightarrow x_{t-1}$만 필요, 나머지 영향 X)

$$p(x_0, x_1, x_2, \ldots x_T) = p(x_0) \cdot p(x_1 | x_0) p(x_2 | x_1) \cdots p(x_T | x_{T-1})$$

따라서 $p(x_0, x_1, \ldots x_T) = p(x_0) \prod_{t=1}^{T} (p(x_t | x_{t-1}))$

이때, forward에서 $p(X_t | X_{t-1}) \sim N(X_t ; \sqrt{1-\beta_t} \cdot X_{t-1}, \beta_t I)$ 설계함

해당 분포는 backpropa 불가능.

따라서 "reparametrization trick" $\rightarrow X_t = \sqrt{1-\beta_t} \cdot X_{t-1} + \sqrt{\beta_t}$

2015 Dickstein $\rightarrow$ forwarding 1000 step

$\boxed{\text{2020 Jonathan ho}}$ $\rightarrow$ closed form 증명 $q(x_t | x_0)$

$\rightarrow$ reparametrization trick 활용.

$x = 1$ $\quad x_1 = \sqrt{1-\beta_1} \cdot X_0 + \sqrt{\beta_1}$

$$x_2 = \sqrt{1-\beta_2} \cdot \boxed{X_1} + \sqrt{\beta_2} \Rightarrow x_2 = \sqrt{1-\beta_2} \left(\sqrt{1-\beta_1} \cdot x_0 + \sqrt{\beta_1} \right) + \sqrt{\beta_2}$$

$$\rightarrow x_2 = \sqrt{1-\beta_2} \cdot \sqrt{1-\beta_1} \cdot x_0 + \sqrt{1-\beta_2} \cdot \sqrt{\beta_1} + \sqrt{\beta_2}$$

$$x_3 = \sqrt{1-\beta_3} \cdot \boxed{x_2} + \sqrt{\beta_3} \Rightarrow \sqrt{1-\beta_3} \cdot \sqrt{1-\beta_2} \cdot \sqrt{1-\beta_1} \cdot x_0 + \sqrt{1-\beta_3} \cdot \sqrt{1-\beta_2} \cdot \sqrt{\beta_1} + \sqrt{1-\beta_3} \sqrt{\beta_2} + \sqrt{\beta_3}$$

$$\rightarrow x_t = \sqrt{1-\beta_t} \cdot \sqrt{1-\beta_{t-1}} \cdots \sqrt{1-\beta_1} \cdot x_0 + \sqrt{\beta_t} + \sqrt{\beta_{t-1}} \cdot \sqrt{\beta_{t-2}} + \cdots \sqrt{\beta_{t-1}} \cdot \sqrt{\beta_{t-2}} \cdots \sqrt{\beta_1}$$

notation : $1 - \beta_t = \alpha_t$, $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$

$$\rightarrow x_t = \sqrt{\alpha_t} \cdot \sqrt{\alpha_{t-1}} \cdots \sqrt{\alpha_1} \, x_0 + \sqrt{1-\alpha_t} + \sqrt{1-\alpha_t} \cdot \sqrt{1-\alpha_{t-1}} \cdots$$

$$x_t = \sqrt{\bar{\alpha}_t} \cdot x_0 + \prod_{s=1}^{t} \sqrt{1-\alpha_t}$$

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

대체 Loss function.

■ DDPM

앞서 확인한 Loss.

$$\mathbb{E}_q\left[-\log p(x_T) - \sum_{t\geq 1} \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_t|x_{t-1})}\right]$$

↓ 최적화 하면

해당 내용은 이미
2015 DPM에서 증명완료.

$$\mathbb{E}_q\Big[\underbrace{D_{KL}(q(x_T|x_0)\,||\,p(x_T))}_{L_T} + \underbrace{\sum_{t>1} D_{KL}(q(x_{t-1}|x_t,x_0)\,||\,p_\theta(x_{t-1}|x_t))}_{L_{T-1}} \underbrace{- \log p_\theta(x_0|x_1)}_{L_0}\Big]$$

edge effect 제외 실제 의미 있는 부분 $= L_{T-1}$. → x_0 Condition.

(feat. markov + 식 변환)

$\boxed{\text{DDPM에서는 } D_{KL}}$ → $q(x_{t-1}|x_t, x_0)\,||\,p_\theta(x_{t-1}|x_t)$ 를 $\boxed{\text{직접 비교}}$하는 형태

방향 동일,,

둘다 reverse 형태.

① markov process 이므로

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

정답 (방향성 제시)

② reverse 한 형태로 변형시.

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

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

대입시.

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

③ x_0 필요 이유

$q(x_{t-1}, x_t)$ 를 구할수 없음 → $\boxed{x_0}$ + 있어야만 closed-form 형태 계산 가능

noise x.

~~analytic (손쉽, 2배) 계산 가능~~

④ Gaussian distribution의 곱

앞서 구한 $q(x_{t-1} | x_t, x_0) = q(x_{t-1} | x_t)$ (markov process.)

$$\Rightarrow \frac{q(x_t | x_{t-1}) \cdot q(x_{t-1}, x_0)}{q(x_t | x_0)}$$

$\rightarrow 1-\alpha_t$

① $q(x_t | x_{t-1}) \sim N(x_t ; \sqrt{\alpha_t} x_{t-1}, \beta_t I)$ — notation 적용

② $q(x_{t-1} | x_0) \sim N(x_{t-1} ; \sqrt{\bar{\alpha}_{t-1}} x_0, (1-\bar{\alpha}_{t-1}) I)$

③ $q(x_t | x_0) \sim N(x_t ; \sqrt{\bar{\alpha}_t} x_0, (1-\bar{\alpha}_t) I)$ — 앞서 구한 과정.

$$\frac{① \times ②}{③} \sim \frac{x_t \cdot x_{t-1}}{x_t} \sim \boxed{N(x_{t-1} ; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t I)}$$

normalizing

$$\tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t$$

$$\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}$$

DBM에서는...
〃 이후 다시 등장 예정

DDPM

Diffusion → 제한도 있어 보이나 자유로운 모델.

그러나. 필수적으로 정하고 넘어가야하는 부분도 있음

① Variance β_t

② model Architecture

③ Gaussian distribution parameterization of the reverse process

①,②,③ → denoising score matching, Diffusion 결합

⇒ 최종적으로, 단순(simple)한 objective fuction 제안
(Variational Bound)

recap) DBM은 Loss $= L_T + L_{T-1} + L_0$ 사용 ☆

DDPM은 β_T (noise) fix 함 → L_T 무시

→ 추정적으로 ($\mathcal{N}(x_0 ; \mu_\theta(x_1, 1), \sigma_1^2 I)$)

fix → answer x_0로.

$\boxed{L_0}$

DBA → edge effect ($t=1$ → $t=0$ process).

DDPM → $p_\theta(x_0|x_1)$에 대해서 pixel 별, 계산 진행

⇒ 결과적으로 Scailing X. / noise 추가 X → Variational Bound. ⇒ lossless code length

→ 이산적인 → 연속적으로.

실제분포 = Continuous. X ([0, 255], image])

→ 픽셀별 계산 ⇒ likelihood 증가 (실제 Scailing, Jacobian 필요X)

$\boxed{L_{T-1}}$ (오른쪽 참고)

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \underbrace{\mu_\theta(x_t,t)}_{\text{reverse process}}, \Sigma_\theta(x_t,t))$$

$$\Sigma_\theta(x_t,t) = \sigma_t I \;/\; \sigma_t^2 = \beta_t \;/\; \tilde{\sigma}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$$

보수 비슷한 성능

upper bound ← $x_0 \sim \mathcal{N}(0,1)$ 일때 최적

lower bound ← $x_0 \sim$ fix 일때 최적

$$L_{T-1} : E_q\Big[\frac{1}{2\sigma_t^2}\|\underbrace{\tilde{\mu}_t(x_t,x_0)}_{q\text{의 분포}} - \underbrace{\mu_\theta(x_t,t)}_{p\text{의 분포}}\|^2\Big] + \underbrace{C}_{\text{무시}}$$

develop

→ $\mathcal{N}(x_{t-1}; \mu_\theta(x_t,x_0), \Sigma_\theta(x_t,t))$

forward process posterior mean.

$$L_{T-1} - C = E_{x_0,\epsilon}\Big[\frac{1}{2\sigma_t^2}\Big\|\tilde{\mu}_t\Big(x_t(x_0,\epsilon), \frac{1}{\sqrt{\bar{\alpha}_t}}\big(x_t(x_0,\epsilon) - \sqrt{1-\bar{\alpha}_t}\,\epsilon\big)\Big) - \mu_\theta(x_t(x_0,\epsilon),t)\Big\|^2\Big]$$

$$= E_{x_0,\epsilon}\Big[\frac{1}{2\sigma_t^2}\Big\|\frac{1}{\sqrt{\alpha_t}}\Big(x_t(x_0,\epsilon) - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon\Big) - \mu_\theta(x_t(x_0,\epsilon),t)\Big\|^2\Big]$$

L_{T-1} 유도 과정

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

$\rightarrow$ $x_t(x_0, \epsilon) = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\cdot\epsilon$ $\quad (\epsilon \sim \mathcal{N}(0, I))$

$\rightarrow$ x_t 를 sampling 이 아닌, $\boxed{x_0, \epsilon \text{ sampling}}$

x_0 : deterministic $\quad$ ϵ : stochastic (random)

L_{T-1} : μ_θ, model

(Langevin dynamic 와 비슷)

$\epsilon_\theta(x_t, t)$ 예측

$\rightarrow$ μ_θ 를 구할 수 있음

$$\mu_\theta(x_t, t) = \tilde{\mu}_t\left(x_t, \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(x_t)\right)$$

$\rightarrow$ σ 는 fix

$$p_\theta(x_{t-1}|x_t) = \boxed{\frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boxed{\epsilon_\theta(x_t, t)}\right)} + \sigma_t^2 \quad x_{t-1} \sim p_\theta(x_{t-1}|x_t)$$

sampling

$$x_{t-1} \Rightarrow \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right) + \sigma_t z \quad z \sim (0, I)$$

앞서 구한 $E_{x_0, \epsilon}$ 에 적용하면

(denoising score matching 과 비슷)

$$E_{x_0, \epsilon}\left[\boxed{\frac{\beta_t^2}{2\sigma^2\alpha_t(1-\bar{\alpha}_t)}}\|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}}\,x_0 + \sqrt{1-\bar{\alpha}_t}\,\epsilon, t)\|^2\right]$$

$\Rightarrow$ epsilon 회귀식 = VI 회귀식과 동일

~~x_0 → 생성~~

제거 version → sigma 기준 모두 동일하게 계산 = Unweight

작은 noise는 작게, 큰 noise는 크게 학습

$$L_{simple}(\theta) = E_{t, x_0, \epsilon}\left[\|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}\cdot x_0 + \sqrt{1-\bar{\alpha}_t}\,\epsilon, t)\|^2\right]$$

$t \rightarrow [1, T]$, 심지어 L_0 생략 시 성능 ↑

→ evaluation 시에만 계산 / 학습 이용 X