Paper summary

AI VISION Lab

1. 공부한 논문의 제목, 게재된 학회 혹은 저널 등 논문 기본 정보를 적으세요.
   - **A. 이름: Diffusion Models Beat GANs on Image Synthesis**
   - **B. 저널: NeurIPS**
   - **C. 도메인: Diffusion**
   - **D. 출판연도: 2021**
   - **E. 저자: Prafulla Dhariwal, Alex Nichol**

2. 논문에서 제안한 알고리즘 및 프레임워크에 대해 본인이 이해한대로 다이어그램을 그려보세요. **논문 Figure를 그대로 따라 그리면 안됩니다.**
   - A. **[선행연구] GANs의 한계: 학습 불안정, Tuning의 어려움**
   - B. **[선행연구] Likelihood Based Model의 한계 및 장점**
     - i. **장점: GAN에서 분포를 Implicit하게 표현하는 점을 해결가능 + 학습 안정적임**
     - ii. **단점: 성능(Fidelity)이 GAN에 비해 좋지 않음**
   - C. **[선행연구] Diffusion의 한계: Distribution Coverage가 넓고, 성능적 측면도 이미 증명함(예: DDPM -> CIFAR10, MNIST Dataset)**
   - D. 그러나 아직까지, 복잡한 분포를 가진 데이터셋에서는 성능이 GAN에 비해 상대적으로 낮음. 따라서, 해당 논문에서는 ImageNet Dataset에 대해서, GAN대비 높은 성능을 뽑을 수 있는 Diffusion Model을 연구하고자 함.
   - E. 따라서, 해당 연구에서는 GAN의 Truncation Trick처럼, Diversity-Fidelity Trade-off가 가능한 Model Architecture을 제시함(Classifier Guidance).
   - F. 해당 Classifier Guidance의 경우, Architecture Layer에서 class Label을 Normalization과정의 Bias로 추가해줌.
   - G. 뿐만 아니라, 모델의 Distribution에 대해서, y를 conditional Distribution으로 Tractable하게 계산할 수 있는 구조를 보임. (+ Scaling)

3. 본인이 생각하는 이 논문의 장점이 무엇이라고 생각하나요? **논문 Contribution bullet을 그대로 따라 적으면 안됩니다.**
   A. Class label을 Condition으로 제공함으로써, 기존에 이미 GANs의 성능을 뛰어넘었음을 보였던 상대적으로 간단한 데이터셋이 아닌, 복잡한 데이터셋(예: LSUN, ImageNet)에서 GAN의 성능을 압도함으로써, Diffusion의 시대임을 확실하게 했다는 장점을 가지고 있음.

   B. 또한, 실험 기반의 연구를 보임으로써 설득력을 높였다는 점이 장점이라 판단됨.
   C. 뿐만 아니라, 다양한 Ablation 실험을 토대로 UNet기반의 Resblock Based Diffusion Model의 Architecture을 제시함. (이미 UNet은 사용되고 있었으나, Tuning과정을 실험적으로 보였다는 점에서 의의가 있음)

4. 이 논문을 읽으면서 느낀 점, 혹은 배운 점이 있으면 적어보세요.
   A. Classifier Guidance에 대한 개념을 재정리할 수 있었고, 향후 Classifier Free Guidance 논문을 읽는 과정에 큰 도움이 될 것이라 판단됨.

   B. 또한, Appendix 과정이 상세히 설명되어 있다는 점에서, Diffusion 선행연구를 Review하는 측면에서도 좋은 논문이라고 생각함.

5. 이 논문의 한계점이 있다면 무엇이라고 생각하나요?
   A. 해당 논문은 ImageNet Dataset의 성능을 향상시키는 방법을 탐구하고자 한 연구임. 결과적으로 Class Label을 condition으로 제공했을 때, 성능의 향상을 확인할 수 있었으나, Label이 없는 Dataset에 대해서는 해당 연구의 장점을 적용할 수 없다는 한계를 가지고 있음.

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

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

날짜: 2025-07-10

이름: 신준원

Diffusion Models Beat GANs on Image Synthesis 07.10

option.

① L_{hybrid} → $L_{vlb} + L_{simple}$ (Σ_θ 학습)

② sampling → 적은 step : DDIM. 많은 step : DDPM.

③ Metrics

→ IS : 전체분포 차이 비교함 (특정 class 갯수가 많은 경우 → 영향↑) 사용X

→ FID : diversity & fidelity 측정가능

→ Recall : diversity 측정

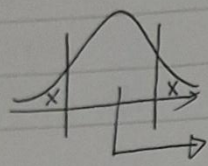
→ Precision : fidelity 측정

Architecture (Imagenet)

- Unet
- Attention (기준 : 16×16 → 전부 (32×32, 16×16, 8×8))
- Attention head (channel)
- Adaptive Group Normalization (AdaGN)
  - ↳ $AdaGN(h, y) = y_s \, GroupNorm(h) + y_b$
  - (y_s, y_b → class, timestep)

Classifier Guidance

GAN 에서 Truncation 의 trick



⇒ 해당 분포에서 sampling 함.

⇒ diversity 감소 + fidelity 증가

목표 ⇒ GAN 의 trade-off ⇒ 현실적인 샘플링

① GAN ⇒ class-label 사용 → Diffusion + class-label

↳ Classifier 없이, 성능향상 목표.

↓

Gradient of a classifier 사용

reverse kernel

$$p_{\theta,\phi}(x_t \mid x_{t+1}, y) = Z p_\theta(x_t \mid x_{t+1}) p_\phi(y \mid x_t)$$

label condition transition kernel

$$\approx \hat{q}(x_t \mid x_{t+1}, y) = \frac{\hat{q}(x_t, x_{t+1}, y)}{\hat{q}(x_{t+1}, y)}$$

$$= \frac{\hat{q}(x_t, x_{t+1}, y)}{\hat{q}(y \mid x_{t+1}) \cdot \hat{q}(x_{t+1})}$$

$$= \frac{\hat{q}(x_t \mid x_{t+1}) \hat{q}(y \mid x_t, x_{t+1}) \cdot \hat{q}(x_{t+1})}{\hat{q}(y \mid x_{t+1}) \cdot \hat{q}(x_{t+1})}$$

$$= \frac{\hat{q}(x_t \mid x_{t+1}) \cdot \hat{q}(y \mid x_t)}{\hat{q}(y \mid x_{t+1})}$$

$$= \frac{\hat{q}(x_t \mid x_{t+1}) \cdot \hat{q}(y \mid x_t)}{\hat{q}(y \mid x_{t+1})} \approx \frac{\hat{q}(x_t \mid x_{t+1}) \cdot p_\phi(y \mid x_t)}{Z}$$

↳ x_t 에 의존 ⇒ "Z"

Gradient of a classifier

$x_{t+1} = \mu \cdot x_t + \Sigma$

$p_\theta(x_t | x_{t+1}) \sim N(\mu, \Sigma)$

log 취함.

multivariant log

μ, σ

Reverse kernel : $p_\theta(x_t | x_{t+1}) = N(\mu, \Sigma)$

① $\log p_\theta(x_t | x_{t+1}) = -\frac{1}{2}(x_t - \mu)^T \Sigma^{-1}(x_t - \mu) + C.$

② $p_\phi(y | x_t)$

↳ ▷ taylor expansion, $\Sigma \to 0$

$\log p_\phi(y | x_t) \approx \log p_\phi(y | x_t)|_{x_t = \mu} + (x_t - \mu)\nabla_{x_t} \log p(y | x_t)|_{x_t = \mu}$

$= (x_t - \mu) g + C_1$

$g = \nabla_{x_t} \log p_\phi(y | x_t)|_{x_t = \mu}$

합치면, $\log(p_\theta(x_t | x_{t+1}) \cdot p_\phi(y | x_t)) \approx \underbrace{-\frac{1}{2}(x_t - \mu)^T \Sigma^{-1}(x_t - \mu)}_{①} + \underbrace{(x_t - \mu) g}_{②} + C_2$ (①+②)

$\to -\frac{1}{2}(x_t - \mu - \Sigma g)^T \Sigma^{-1}(x_t - \mu - \Sigma g) + \frac{1}{2} g^T \Sigma g + C_2.$

$\to -\frac{1}{2}(x_t - \mu - \Sigma g)^T \Sigma^{-1}(x_t - \mu - \Sigma g) + C_3$

$\to \log p(z) + C_4, \quad z \sim N(\mu + \Sigma g, \Sigma)$

↳ $z \propto p_\theta(x_t | x_{t+1}) p_\phi(y | x_t)$

$x_{t-1} \leftarrow$ sample from $N(\mu + s\Sigma \nabla_{x_t} \log p_\phi(y | x_t), \Sigma)$

↳ with condition.

s: Scale factor.

확률 → SDE based.

Conditional sampling for deterministic methods (DDIM)

$$\nabla_{x_t} \log p_\theta(x_t) = -\frac{1}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(x_t)$$

$$\nabla_{x_t} \log\left(p_\theta(x_t) \cdot p_\phi(y|x_t)\right) = \nabla_{x_t} \log p_\theta(x_t) + \nabla_{x_t} \log p_\phi(y|x_t)$$

$$= -\frac{1}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(x_t) + \nabla_{x_t} \log p_\phi(y|x_t)$$

$$\rightarrow \quad \hat{\epsilon}(x_t) = \epsilon_\theta(x_t) - \sqrt{1-\bar{\alpha}_t} \cdot \nabla_{x_t} \log p_\phi(y|x_t)$$

$$x_{t-1} \leftarrow \sqrt{\bar{\alpha}_{t-1}} \left(\frac{x_t - \sqrt{1-\bar{\alpha}_t}\hat{\epsilon}}{\sqrt{\bar{\alpha}_t}} \right) + \sqrt{1-\bar{\alpha}_{t-1}}\hat{\epsilon}$$

Scaling Classifier Gradients

① unconditional Imagenet model
→ Classifier : 50 % 사용 (Scale of "1") → FID : 33.0
문제: 성능이 낮음 unmatched

② Conditional Imagenet model
→ classifier : 100 % 사용 (Scale of "10") → FID : 12.0

$$s \cdot \nabla_x \log p(y|x) = \nabla_x \log \frac{1}{Z} p(y|x)^s$$

효과 ① $p(y|x) \longrightarrow p(y|x)^s$: **sharpen.**
⇒ classifier 강화됨 → fidelity ↑
diversity ↓

확산 모델 → $p(x|y)$ – Conditional
→ $p(x)$ – unconditional

sampling
→ Guidance. (Scale : hyperparameter)