



[5주차] 논문리뷰1 (Introduction)

DDPM

1. Introduction

논문이 다루는 분야

해당 task에서 기존 연구 한계점

논문의 contributions

Reference

DDPM

1. Introduction



논문에서 다루고 있는 주제가 무엇인지와 해당 주제의 필요성이 무엇인가

논문에서 제안하는 방법이 기존 방법의 문제점에 대응되도록 제안 되었는가

논문이 다루는 분야

- Diffusion 계열 생성형 모델 (발전 역사가 궁금해서 GPT 도움 받아 표로 정리)

연도	모델/기술 이름	주요 기여 또는 특징	관련 논문/출처
2015	Diffusion Probabilistic Models (DDPM 초기 개념)	확률적 그래픽 모델 기반의 초기 Diffusion 개념 정립	Sohl-Dickstein et al. (2015)
2020	DDPM (Denoising Diffusion Probabilistic Models)	Gaussian noise를 점진적으로 제거하는 방식, 이미지 생성 성능에서 GAN 과 경쟁	Ho et al. (2020), NeurIPS
2021	DDIM (Denoising Diffusion Implicit Models)	DDPM보다 더 빠른 샘플링 가능, non-Markovian 구조 도입	Song et al. (2021), ICLR

연도	모델/기술 이름	주요 기여 또는 특징	관련 논문/출처
2021	Score-based Generative Modeling (SDE 기반)	확률적 미분방정식(SDE) 기반으로 연속적인 시간 표현	Song et al. (2021), ICLR
2022	GLIDE	텍스트 조건 기반 생성 + Diffusion, CLIP을 활용한 조건 강화	Nichol et al. (2022), OpenAI
2022	Latent Diffusion Models (LDM)	고해상도 이미지 생성을 위한 latent 공간에서의 Diffusion 수행 → 계산 효율성 향상	Rombach et al. (2022), CVPR
2022	Imagen	T5 기반의 텍스트 인코더 + Diffusion 조합, 매우 높은 FID 성능 기록	Saharia et al. (2022), Google Brain
2022	Stable Diffusion	오픈소스 LDM 기반 모델, 커스터마이징과 Fine-tuning에 최적화	Rombach et al. (2022), CompVis
2023	ControlNet	조건 기반 제어 기능 (포즈, 윤곽 등)을 Diffusion에 통합	Lvmin Zhang et al. (2023), Stanford
2023	SDXL (Stable Diffusion XL)	안정성과 해상도 향상, 텍스트 이해력 강화	Stability AI (2023)
2024	Sora (Video Diffusion, 발표됨)	OpenAI에서 개발한 비디오 생성용 Diffusion 모델 (텍스트→비디오)	OpenAI Sora (2024, 발표 기준)

해당 task에서 기존 연구 한계점

- 언급된 기존 연구 (좋은 성능을 보여왔음)
 - Generative adversarial networks (GANs)
 - autoregressive models
 - flows
 - variational autoencoders (VAEs)
 - energy-based modeling
 - score matching

- 디퓨전 모델은 그동안 좋은 품질의 이미지를 생성할 수 있다는 근거가 부족했음 → 이 모델에서 좋은 품질 가능성을 보이게 됨

논문의 contributions

- 논문의 작업 그래프로 정리
 - 노이즈를 추가해서 데이터를 조금씩 손상시키고
 - 반대로 노이즈를 제거하며 원래 데이터를 복원하는 모델을 학습
 - ⇒ 노이즈의 크기가 작고, 가우시안 분포를 따르면 복원과정 역시 조건부 가우시안 형태로 설계할 수 있어서 신경망으로 쉽게 구현할 수 있음

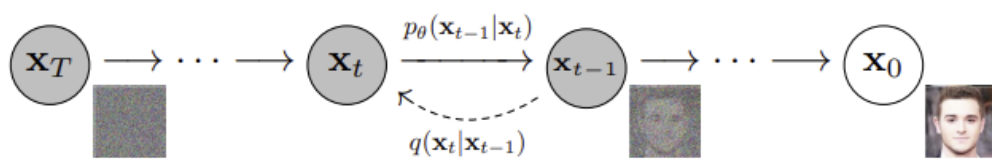


Figure 2: The directed graphical model considered in this work.

- **(핵심기여)** 특정한 방식으로 디퓨전 모델을 설계(parameterize)하면,
 - 학습 과정은 denoising score matching 기법처럼 되고
 - 샘플링(생성) 과정은 annealed Langevin dynamics 방법처럼 작동함을
 - 수학적으로 발견함. 요 때 가장 좋은 품질의 샘플 생성.
- 품질 좋은 이미지 만들어도 **로그 가능도(log-likelihood)** 지표에서는 기존의 디퓨전 모델들과 비교해 경쟁력이 낮은 경향이 있음.
 - ⇒ 그래도 energy-based modeling나 score matching 보다는 좋음
- 무손실 압축길이의 대부분이 눈치채기 힘든 미세한 디테일 설명에 쓰임을 발견했음
 - ⇒ 그래서 손실 압축의 관점에서 분석했음
 - ⇒ 모델이 픽셀 순서대로 복원하는 게 아니라, 노이즈의 크기(정보의 정밀도) 정도에 따른 순서대로 이미지를 단계적으로 복원함을 발견했음
 - 무손실 압축 : 데이터를 원본 그대로 복원할 수 있도록 압축하는 방법
 - 무손실 압축 실이 : 무손실 압축 시 필요한 최소 비트 수
 - 무손실 압축에 쓰인 비트 중 상당수가 지나치게 미세한 정보 표현에 사용되어서 그럴 필요가 없다는 생각으로 손실 압축 관점에서 다시 분석한 것.
 - progressive decoding : 데이터를 점점 단계적으로 정교하게 복원해나가는 방식
 - bit ordering : 정보를 복원할 때 어떤 순서로 데이터를 처리할 지 정의하는 방식

- 기존 autoregressive model : 이전까지 복원된 정보들을 기반으로 다음 정보를 한 단계씩 예측. 픽셀 순서대로.
- 디퓨전 모델 : 노이즈 크기 순서대로(정밀도가 낮는 순서대로. 큰 윤곽 → 디테일)
- 디퓨전 모델이 기존 오토리그레시브 모델처럼 데이터를 순차적으로 복원하긴 하지만, 픽셀 순이 아닌 노이즈 크기(정밀도 정도)에 따라 복원한다는 점에서 더 일반적이고 강력한 구조를 가진다는 것.

Reference

- <https://www.youtube.com/watch?v=H45IF4sUgiE&t=206s>