

Latent Diffusion Model

[Diffusion model의 문제](#)

[학습단계](#)

[LDM\(Latent Diffusion Models\)](#)

[방법](#)

[Perceptual Image Compression](#)

[Latent Diffusion Models](#)

[LDM 정리](#)

Diffusion model의 문제

DM을 최적화하는데 대량의 GPU와 시간이 필요하다는 것이었다(결국 너무 시간이 많이 걸린다)

→ 학습 및 샘플링 모두에 대한 계산 복잡성을 줄여야 함.

이 문제를 pretrained auto encoder의 latent space를 적용하여 제한된 컴퓨팅 리소스로 학습이 가능하게 한다.

학습단계

1. 픽셀 공간에서 이미 학습된 diffusion model의 분석:
 - Perceptual compression: high-frequency detail들을 제거하지만 의미(semantic)는 거의 학습x 단
 - Semantic compression: 실제 생성 모델이 데이터의 의미론적(semantic) 구성과 개념적(conceptual) 구성을 학습하는 단계

LDM(Latent Diffusion Models)

▼ 특징

1. 범용 autoencoding 단계를 한 번만 학습하면 되므로 여러 diffusion model 학습에 재사용하거나 완전히 다른 task를 탐색가능→ 시간down, 효율적인 탐색 가능
2. 학습 과정에서 patch-based adversarial objective를 사용하여 이미지의 디테일을 잘 보존

▼ autoencoder를 학습

▼ 데이터 space와 perceptual하게 동일한 저차원 representational space로 보내는것이 autoencoder

▼ 학습된 잠재 공간에서 diffusion model을 학습

이전 모델들과 달리 과도한 space 압축에 의존할 필요x

▼ 효율적인 이미지 생성

감소된 복잡성으로 인해 네트워크를 한 번만 통과

방법

▼ LDM 모델: 이미지 공간과 perceptual하게 동일한 공간을 학습하지만 계산 복잡성을 크게 줄이는 autoencoding 모델을 활용

cf) diffusion model이 해당 손실 항을 적게 샘플링하는 방법은 detail 무시 가능하지만 계산량 많은 단점이 있어 이 방법은 사용 안함

▼ 장점

1. 고차원 이미지 space를 남겨두어 저차원 space에서 샘플링→계산효율 up
2. UNet 아키텍처에서 상속된 diffusion model의 학습편향을 활용하는 공간 구조 데이터에 효과적
3. Latent space가 여러 생성 모델을 훈련하는 데 사용 가능 & downstream task에도 활용 가능한 범용 압축 모델 얻을 수 있음

Perceptual Image Compression

Perceptual Image Compression 모델 동작 방법:

인코더가 입력 RGB 이미지를 latent representation

z로 인코딩 → 디코더가 z로부터 이미지를 n이 정규분포를 따르도록 유도 → 분산을 제어

KL 정규화: latent representation이 정규분포를 따르도록 유도, latent space의 분산이 커지는 것을 막

VQ 정규화: 벡터 양자화를 이용하여 latent space를 정규화, 양자화 layer가 디코더에 흡수되어 작동(latent space 구조 유지, 압축 효율, 모델 안정)

결국 학습된 latent space z 의 2차원 구조를 유지하면서 압축을 수행 → 높은 재구성 품질

Latent Diffusion Models

결국 그래서 이것 가지고 무엇을 하느냐?

:

학습된 perceptual compression 모델을 사용해서 감지할 수 없는 고주파의 detail이 추상화되는 효율적이고 낮은 차원의 latent space에 접근가능(계산적으로 훨씬 더 효율적인 더 낮은 차원에서 학습)

이러면 complexity reduction, detail-preservation간 near-optimal point를 처음으로 찾아낼 수 있음

cross-attention layer를 model architecture 도입으로

1. 텍스트나 bounding box를 DM이 유연하게 변경가능
2. 고해상도는 convolution으로 가능

따라서 Latent Diffusion Model은 이미지 복원, class-conditioned image synthesis에서 최고수준을 달성했고, 문자에서 이미지 합성, 조건 없이 이미지 생성, super-resolution에서도 컴퓨팅 요구를 줄이면서 픽셀에 기반한 DM대비 높은 퍼포먼스를 보인다.

성능을 저하시키지 않으면서 계산 요구량을 줄이는 것이 핵심

LDM 정리

따라서 결국 LDM은 low dimension인 공간에서 학습하여 계산량이 적어 시간이 오래걸리지도 않고 LDM은 압축 모델로 perceptual image compression을 사용하는데 이는 2차원 구조로 디자인 되어 x 의 디테일을 1차원보다 더 잘 보존하고 likelihood based 생성 모델에 적합(데이터의 중요하고 의미적인 bit에 집중 가능)하다는 특징이 있다. 따라서 기존의 DM의 단점을 보완하고 이미지 합성에서 높은 퍼포먼스를 보이는 모델이다.

