카테고리 없음

[Euron] 5week_Denoising Diffusion Probabilistic Models

yejjji 2025. 4. 7. 18:16

0. Abstract

본 논문은 Diffusion Probabilistic Models를 활용한 고품질의 이미지 합성 결과를 보여준다.
이는 nonequilibrium thermodynamics로부터 영감받은 잠재 변수 모델들의 클래스로 이루어져있다.

이 논문의 최고 결과는 Langevin Dynamics에서 **diffusion probabilistic models**와 **denoising score matching** 간의
새로운 연결들로 이루어진 가중된 variational bound로 train되어 얻어진 결과값이다.
또, 본 논문의 모델들은 자연스럽게 **autoregressive decoding**의 일반화된 버전으로 해석될 수 있는 점진적 손실 압축 해제 방식을 받아들인다.

점진적 손실 압축 해제 방식 (progressive lossy decompression scheme) 이란 ?
: 데이터를 점진적으로 복원하는 방식의 손실 압축 기법

1) 손실 압축 : 데이터의 일부분을 영구적으로 제거함
2) 점진적 해제 : 일부 데이터만으로 먼저 복원시킴 으로 이루어져 있다.
즉, 압축된 데이터를 순차적으로 해제시켜 저품질의 데이터에서 점점 고품질의 데이터로 복원시키는 방식을 의미한다.

무제한의 CIFAR 데이터셋에서, Inception score로 9.46점, 최고 성능 FID score로 3.17점을 얻었다.
256 X 256 크기의 LSUN에서는 ProgressiveGAN과 비슷한 성능의 샘플을 얻을 수 있었다.

1. Introduction

모든 종류의 심층 생성 모델들은 넓은 범위의 데이터들에서 고품질의 샘플 이미지들을 보여준다.
GAN, 자기회귀모델(Autoregressive Model), variational 오토인코더는 눈에 띄는 이미지와 오디오 샘플들을 합성하고,
GAN모델과 비교하여 이미지들을 생성하는 에너지 기반의 모델링과 scored matching 측면에서 눈에 띄는 발전이 있었다.

본 논문은 Diffusion Probabilistic Models에서 어떤 부분이 발전되었는지를 제시한다.
Diffusion Probabilistic Models는 유한한 시간 이후에 **데이터와 맞는 샘플들을 생성하기 위해서**
variance inference를 사용하여 훈련된 파라미터화된 Markov Chain의 형태이다.

Variance Inference (변분 추론)란 ?
: 확률분포의 복잡한 Posterior 분포를 근사적으로 계산하는 과정이다.

Markov Chain의 전이들은 **diffusion 과정을 역으로 되돌리도록 학습**되는데, 이
과정은 샘플링과 반대 방향으로 작용하는 Markov Chain이며, **데이터에 점진적으
를 추가**하여 결국에는 signal을 완전히 파괴하는 과정이다.

Diffusion이 약간의 Gaussian noise들로 구성되어있을 때, 샘플링 과정의 전이
즉 diffusion 과정 역시 조건부 Gaussian 분포로 설정하는 것으로 충분하다.
이 과정으로 **신경망의 파라미터값들이 단순화**될 수 있다.

Diffusion model들은 정의하기가 쉽고 훈련 측면에서 효율적이나, 기존에 관련 연구들에서는 Diffusion model이
고품질의 데이터를 생성 가능하다는 명확한 증거가 없었다.

본 논문은 **Diffusion Model이 실제로 고품질의 샘플들을 생성해낼 수 있으며**,
가끔씩은 다른 생성 모델들의 기존 결과들보다 우수하다는 것도 보인다.

게다가, Diffusion Model이 특정 파라미터값으로 파라미터화되면, 훈련 동안의 다중 noise 레벨들에서의 **Denoising Score Matching과 샘플링 동안의 Annealed Langevin Dynamicss와 동등하다**는 것을 밝혀냈다.
다시 말해서, 기존의 확률적 샘플링 기법들과 깊은 연관성을 가지고 있다는 것이다.

파라미터화 과정을 통해 고품질의 샘플을 얻을 수 있었고, **Denoising Score Matching과 샘플링 동안의 Annealed Langevin Dynamicss와 동등성을 이 연구의 주요한 기여 중 하나로** 간주하고자 한다.

샘플 품질에도 불구하고, Diffusion Model들은 다른 likelihood 기반 모델들과 비교하였을 때, 경쟁적인 log likelihood를 가지지는 않는다.하지만, 에너지 기반 모델과 Score Matching을 대상으로 Annealed Importance Sampling 과정으로 얻은 **기존의 추정치보다는 더 높은 가능성을 가진다**는 사실을 알 수가 있었다.

모델들의 lossless codelength는 감지가 불가능한 이미지의 디테일들을 묘사하는 데에 상당수가 사용된다는 사실을 알 수 있었고,

Loseless Codelength(무손실 코드길이)란 ?: 정보를 손실 없이 압축하기 위

비트 수

본 논문은 이를 **손실 압축**으로 더 자세하게 설명하고 Diffusion Model의 샘플링 과정이 일종의 **progressive decoding과정**임을 보인다.이 과정은 Autoregressive Decoding 과정에서 가능한 부분을 훨씬 일반화시킨 비트 순서를 따르는 Progressive Decoding과 비슷하다.즉, Autoregressive Decoding보다 더 유연한 방식으로 데이터 생성이 가능하다.

2. Background

$p_\theta(\mathbf{x}_0) := \int p_\theta(\mathbf{x}_{0:T})\, d\mathbf{x}_{1:T}$ 의 형태의 잠재변수 모델을 Diffusion Model들은 가진다.
즉, 관측된 데이터를 여러 개의 잠재 변수들로 설명하는 것이다.

$p_\theta(\mathbf{x}_{0:T})$ 라는 Joint Distribution은 Reverse Process라고도 불리우며, 마지막 노이즈 값인

$p(\mathbf{x}_T) = \mathcal{N}(\mathbf{x}_T; \mathbf{0}, \mathbf{I})$ 에서 시작하여 학습된 Gaussian 전이값으로 이루어진 Markov Chain으로 정의된다.

Diffusion 모델이 다른 종류의 잠재변수 모델들과 구별되는 점은 **대략적인 Posterior값**이다.
이는 Forward Process, Diffusion Process라고도 불리우고 **variance값에 따라 데이터에 점진적인 Gaussian 노이즈를 추가**한 Markov Chain의 형태이다.

Training 과정은 negative log likelihood에서 **variational bound를 최적화**한다.

β_t 라는 Forward Process의 분산값은 학습에 의해 설정될 수 있고, 하이퍼파라미터값으로 고정시킬 수도 있다.

$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ 라는 Reverse Process의 표현은 Gaussian 조건부 선택에 의

다.
그 이유는

β_t 가 작을 때 Forward, Reverse Process가 같은 형태의 수학적 구조를 가지기 때문이다.

Forward Process의 중요한 특징은 임의의 시점 t에서의 상태 xt를 직접 닫힌 형태로 샘플링할 수 있다는 것이다.

효율적인 training을 위해서는 **확률적 경사 하강법으로 L의 랜덤함을 최적화시키는 과정이 필요**하다.
L을 다른 형태로 재구성할 경우, **variance reduction이 가능해져 학습 성능이 좋아지는 특징**이 있다.

$$\mathbb{E}_q\Big[\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0)\parallel p(\mathbf{x}_T))}_{L_T}+\sum_{t>1}\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)\parallel p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}}\underbrace{-\log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0}\Big] \quad (5)$$

KL divergence 활용한 equation 5

결과적으로, 위 equation 5에서 모든 KL divergence들은 Gaussian으로 비교되고,
높은 Monte Carlo 분산 추정치값 대신에 가까운 형태의 값으로 계산이 가능하다.

3. Diffusion models and denoising autoencoders

Diffusion model들은 제한적인 구조의 잠재 변수 모델처럼 보일 수 있으나, 실제로는 구현상으로 다양한 자유도값을 가진다.
Forward Process에서는 분산값을 선택해야하고, Reverse Process에서는 신경망 구조와 Gaussian 파라미터화 방식을 선택해야 한다.

분산값 선택을 돕기 위해, 본 논문은 Diffusion model과 Denoising Score Matching 간의 새로운 연관성을 제시했다.

이는 Diffusion model들에 단순화되고, 가중화된 variational bound값으로 이

Denoising Score Matching 이란?

: 확률값을 직접 학습하지 않고 확률의 로그값의 그래디언트를 예측하는 방법이다.

궁극적으로, 이러한 모델의 디자인이 이론적으로 간단하고, 경험적으로도 효과적이라는 것이 입증되었다.

3.1 Forward process and LT

이론적으로는 Forward Process의 분산값을 재파라미터화 과정으로 학습이 가능하나,
본 연구에서는 분산값들을 상수값으로 고정하였다.
따라서, q는 학습 가능한 파라미터가 없고, LT는 항상 일정한 값으로 학습 과정에서 무시 가능하다.

3.2 Reverse process and L1:T-1

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)) \text{ for } 1 < t \leq T$$

에 대해서 알아보고자 한다.

첫 번째로,

$\boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t) = \sigma_t^2 \mathbf{I}$ 라는 모델이 예측하는 역방향 과정의 분산 행렬을 학습하지 않고,

시간에 따라 변하는 상수값으로 설정하였다.

실험적으로,

$\sigma_t^2 = \tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$ 는 원래 데이터의 분산값과도 유사한 결과값

$\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 일 때 최적인 경우와 1개의 고정된 값일 때 최적인 경우, 이렇게 2가지 경우가 있는데.
이 두 경우 모두 각 차원마다 분산값이 1인 데이터에 대해 역방향 과정의 엔트로피의 상한, 하한값을 나타내는 극단적인 경우를 의미한다.

두 번째로,

$\boldsymbol{\mu}_\theta(\mathbf{x}_t, t)$ 를 제시하기 위해, Lt의 다음과 같은 분석을 통해 유래된 특정 파라미터화 과정을 제안한다.

$$L_{t-1} = \mathbb{E}_q \left[\frac{1}{2\sigma_t^2} \|\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t, \mathbf{x}_0) - \boldsymbol{\mu}_\theta(\mathbf{x}_t, t)\|^2 \right] + C$$

가장 직관적인 평균값의 파라미터값을 구하는 방법은 평균값의 바 형태를 예측하는 모델로 보는 것이다.

하지만,

$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$ 를 재파라미터화하여 Forward 과정에서 특정 시간 t의 샘플 xt값을

$\mathbf{x}_t(\mathbf{x}_0, \boldsymbol{\epsilon}) = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$ for $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 로 표현이 가능하다.

즉, 노이즈값으로 xt를 직접 샘플링 가능하며 Forward 과정의 Posterior 공식 적용이 가능하다는 것이다.

$$\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t, \mathbf{x}_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathbf{x}_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t \quad \text{and} \quad \tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$$

Posterior 공식

$$\mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\frac{1}{2\sigma_t^2}\left\|\frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t(\mathbf{x}_0,\boldsymbol{\epsilon})-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}\right)-\boldsymbol{\mu}_\theta(\mathbf{x}_t(\mathbf{x}_0,$$

Equation 10

Equation 10은 평균값이 주어진 xt에 대해

$\frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}\right)$ 를 예측한다는 것을 보여준다.

xt가 모델의 input값으로 이용 가능하므로 우리는 파라미터값으로 평균값을 선택한다.

전체 샘플링 과정은 Langevin Dynamics와 유사하며, 입실론값은 데이터 밑도 함수의 학습된 그래디언트 역할을 가진다.

Algorithm 1 Training

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \text{Uniform}(\{1,\dots,T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$
5: Take gradient descent step on
$\nabla_\theta \left\|\boldsymbol{\epsilon}-\boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0+\sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon},t)\right\|^2$
6: **until** converged

Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0},\mathbf{I})$
2: **for** $t=T,\dots,1$ **do**
3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$ if $t>1$, else $\mathbf{z}=\mathbf{0}$
4: $\mathbf{x}_{t-1}=\frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t-\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\right)+\sigma_t\mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

3.3 Data scaling, reverse process decoder, and L0

이미지 데이터는 0~255 사이의 정수들로 구성되어 있다. 이를 선형 변환하여 [-1, 1] 범위로 정규화하고자 한다.
즉, 신경망의 더 안정적인 학습이 가능해지도록 한다는 것이다.

이는 신경망 Reverse Process가 결과적으로 표준 정규분포 P(xT)에서 시작되어, 스케일링된 일정한 Input값을 다룰 수있도록 보장한다

Discrete log Likelihood를 얻기 위해, Reverse Process 과정의 마지막 단계를 독립적인 이산 디코더로 설정한다.

이때, 이 디코더는 정규분포를 기반으로 한다.

VAE 디코더와 Autoregressive Model에서 사용되는 이산화 연속 분포 방시과 유
본 연구에서는 variational bolund가 이산 데이터의 손실 없는 코드길이가 되도

샘플링의 마지막 부분에서, 노이즈 없이 평균값을 표시했다.

3.4 Simplified training objective

$$\mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\frac{\beta_t^2}{2\sigma_t^2\alpha_t(1-\bar{\alpha}_t)}\left\|\boldsymbol{\epsilon}-\boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0+\sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon},t)\right\|^2\right]$$

Equation 12

$$p_\theta(\mathbf{x}_0|\mathbf{x}_1)=\prod_{i=1}^{D}\int_{\delta_-(x_0^i)}^{\delta_+(x_0^i)}\mathcal{N}(x;\mu_\theta^i(\mathbf{x}_1,1),\sigma_1^2)\,dx$$

$$\delta_+(x)=\begin{cases}\infty & \text{if } x=1\\ x+\frac{1}{255} & \text{if } x<1\end{cases}\qquad \delta_-(x)=\begin{cases}-\infty & \text{if } x=-1\\ x-\frac{1}{255} & \text{if } x>-1\end{cases}$$

Equation 13

위에서처럼 Reverse Process와 Decoder가 사용되면, variational bound가 식 12, 13에서 유도된 항들로 구성되며, 세타값에 대해 명확히 미분 가능하므로 훈련할 준비가 되어 훈련을 시작할 수 있다.

하지만, t가 1부터 T까지 균일하게 확률이 선택될 때 그대로 variational bound를 학습하는 것보다,
변형된 형태를 사용하는 것이 품질 개선이 도움이 된다.

$$L_{\text{simple}}(\theta):=\mathbb{E}_{t,\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\left\|\boldsymbol{\epsilon}-\boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0+\sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon},t)\right\|^2\right]$$

Equation 14

Equation 14의 단순한 목표는 Equation 12의 가중치를 제거하는 것이다.
이를 통해 기존의 variational bound는 모든 t값을 동일하게 고려해야 한다는 것이다.

t가 1일 때, L0와 동일하며 이산 디코더의 적분값을 분산과 경계 효과는 무시한 채로
Gaussiam Probability Density function와 구간의 너비를 곱하는 방식으로 근

t가 1보다 큰 경우에는, 위의 Equation 12의 가중치가 없는 버전과 동일하며 NC
유사하다.

이 항들은 작은 양의 노이즈가 추가된 데이터를 복원하도록 학습하는데,
이를 줄이면 네트워크가 더 어려운 복원 작업에 집중이 가능하다.

4. Experiments

샘플링 과정에서 필요한 신경망 평가수가 이전의 연구와 같도록 T의 값은 1000으로 맞춰 실험을 진행하였다.
Forward Process의 분산을 10 ^ -4에서 0.02까지로 선형적으로 증가하는 상수값으로 설정해주었다.

이 상수값들은 [-1, 1] 사이로 스케일링된 데이터에 관련성이 적고,
Forward Process와 Reverse Process가 xT에서 신호, 노이즈 비율이 각 차원마다 가능한 한 작도록 유지할 때와
대략적으로 같은 함수 형태를 가지도록 한다.

Reverse Process를 나타내기 위해, 그룹 정규화로 유사한 U-Net Backbone을 사용하였다.
파라미터들은 시간에 따라 공유되고, Transformer sinusoidal position embedding을 사용한 네트워크로 이루어져 있다.
16 X 16 featuring에서 self-attention을 활용하였다.

4.1 Sample quality

Table 1: CIFAR10 results. NLL measured in bits/dim.

Model	IS	FID	NLL Test (Train)
Conditional			
EBM [11]	8.30	37.9	
JEM [17]	8.76	38.4	
BigGAN [3]	9.22	14.73	
StyleGAN2 + ADA (v1) [29]	**10.06**	**2.67**	
Unconditional			
Diffusion (original) [53]			≤ 5.40
Gated PixelCNN [59]	4.60	65.93	3.03 (2.90)
Sparse Transformer [7]			**2.80**
PixelIQN [43]	5.29	49.46	
EBM [11]	6.78	38.2	
NCSNv2 [56]		31.75	
NCSN [55]	8.87 ± 0.12	25.32	
SNGAN [39]	8.22 ± 0.05	21.7	
SNGAN-DDLS [4]	9.09 ± 0.10	15.42	
StyleGAN2 + ADA (v1) [29]	$\mathbf{9.74} \pm 0.05$	3.26	
Ours (L, fixed isotropic $\mathbf{\Sigma}$)	7.67 ± 0.13	13.51	≤ 3.70 (3.69)
Ours (L_{simple})	9.46 ± 0.11	**3.17**	≤ 3.75 (3.72)

Table 1 : CIFAR10 데이터셋에서의 결과값들

Table 1은 CIFAR10 데이터셋에서의 Inception score, FID score, negative log likelihood값을 보여준다.

3.17의 FID score는 연구의 비조건적 모델이 클래스 조건 모델을 포함하여 문헌에 기존에 보고된 대부분의 모델들에서

더 나은 샘플 품질을 달성했다는 사실을 보여준다.

또, **FID score는 표준적인 관행에 따라 training set을 기준으로 계산**되는데, test set을 기준으로 계산했을 때에도 5.24로,

역시 문헌에 보고된 많은 training set 기반의 FID score들보다 더 뛰어난 값을 가진다는 사실을 알 수가 있었다.

진짜 variational bound값을 기준으로 훈련시킬 경우 더 나은 코드 길이를 얻을 수 있으나,

단순화된 목적 함수로 훈련시켰을 때의 샘플 품질이 가장 좋았다.

4.2 Reverse process parameterization and training objective ablation

Training objective ablation 이란?

: 학습 목표의 구성 요소를 1개씩 제거하거나 변경하며 그 구성 요소의 영향력을 분석하는 실험

Table 2: Unconditional CIFAR10 reverse process parameterization and training objective ablation. Blank entries were unstable to train and generated poor samples with out-of-range scores.

Objective	IS	FID
$\tilde{\mu}$ prediction (baseline)		
L, learned diagonal Σ	7.28 ± 0.10	23.69
L, fixed isotropic Σ	8.06 ± 0.09	13.22
$\|\|\tilde{\mu} - \tilde{\mu}_\theta\|\|^2$	–	–
ϵ prediction (ours)		
L, learned diagonal Σ	–	–
L, fixed isotropic Σ	7.67 ± 0.13	13.51
$\|\|\tilde{\epsilon} - \epsilon_\theta\|\|^2$ (L_{simple})	$\mathbf{9.46 \pm 0.11}$	**3.17**

Table 2 : 샘플 품질에 대한 영향(by Reverse Process 파라미터화, 손실함수)

평균의 바값을 예측하는 기본적인 옵션은 가중치가 없는 MSE와 간단한 akin 대신에 진짜 variational bound에서 훈련되었을 경우에만 잘 작용한다는 것을 알 수가 있다.
또한, Reverse Process의 분산들이 불안정한 학습과 좋지 못한 품질의 샘플로 이끈다는 사실도 알 수 있다.

분산을 고정한 상태에서 variational bound로 훈련될 때 입실론값을 예측하는 것은 평균의 바값을 예측하는 것만큼 성능이 비슷하게 좋으나, **간단한 목표함수로 훈련될 때에는 성능이 평균의 바값을 예측할 때보다 훨씬 좋다.**

4.3 Progressive coding

위의 Table 1은 CIFAR 데이터셋의 모델의 코드 길이도 함께 보여준다.
train과 test 데이터의 차이는 차원마다 최대 0.03 비트에 불구하고, 이는 다른 likelihood 기반 모델에서 보고된 차이와 비슷한 수준이다.

그럼에도 불구하고, 이 연구의 손실값이 없는 코드 길이는 **에너지 기반 모델이나 Annealed Importance Sampling을 사용한 Score Matching에서 보고된 큰 추정치보다 더 나은 성능**을 보이지만, **다른 종류의 likelihood 기반의 생성 모델들과 비교했을 때에는 경쟁**

하다.

하지만, 이 샘플들이 여전히 고품질이고, 손실 압축기로서 우수한 귀납적 편향값을 가진다고 결론을 내렸다.
Variational Bound의 항들을 L1 + L2 + ... + LT로 보고— 이중 rate는 L1, L0는 distortion이라고 가정했을 때,
CIFAR 모델 중 가장 고품질의 데이터를 생성한 모델은 차원당 1.78 비트의 rate와 1.97 비트/차원의 distortion을 기록하였다.
이는 0~255 범위에서 RMSE가 0.95에 해당한다.

Progressive lossy compression

Progressive lossy compression: 처음엔 대략적인 저해상도 복원본을 사용하다가, 추가 데이터를 시간이 지남에 따라 받아가며 더 선명하게 이미지를 복원해나가는 과정

모델의 rate-distortion 특성을 더 깊이있게 탐색하기 위해 Progressive lossy compression 코드를 도입하였다.
이는 알고리즘 3, 4에서 설명되며, minimal random 코딩과 같은 절차에 접근 가능하다고 가정하였다.

Algorithm 3 Sending $\mathbf{x}_0$

1: Send $\mathbf{x}_T \sim q(\mathbf{x}_T|\mathbf{x}_0)$ using $p(\mathbf{x}_T)$
2: **for** $t = T-1, \ldots, 2, 1$ **do**
3: Send $\mathbf{x}_t \sim q(\mathbf{x}_t|\mathbf{x}_{t+1}, \mathbf{x}_0)$ using $p_\theta(\mathbf{x}_t|\mathbf{x}_{t+1})$
4: **end for**
5: Send $\mathbf{x}_0$ using $p_\theta(\mathbf{x}_0|\mathbf{x}_1)$

Algorithm 4 Receiving

1: Receive $\mathbf{x}_T$ using $p(\mathbf{x}_T)$
2: **for** $t = T-1, \ldots, 1, 0$ **do**
3: Receive $\mathbf{x}_t$ using $p_\theta(\mathbf{x}_t|\mathbf{x}_{t+1})$
4: **end for**
5: **return** $\mathbf{x}_0$

알고리즘 3, 4

$$D_{\mathrm{KL}}(q(\mathbf{x}) \parallel p(\mathbf{x}))$$

의 평균적인 비트 내로 샘플을 전송 가능한데, 이때 송신자와 수신자 중 수신자만이 p 값을 사전에 알고 있다.
이를 x0에 적용하면, 알고리즘 3과 4는 xT, ..., x0을 순차적으로 전송하고, [illegible]
코드 길이는 Equation 5에 해당한다.

임의의 시간 t에서, 수신자는 xt의 부분적인 정보를 완전히 이용 가능하고, [illegible]
x0을 추정해낸다.

Stochastic reconstruction x0도 유효하나, 여기에서는 사용하지 않았다.
이는 평가하기에 distortion을 더 어렵게 만들기 때문이다.

Figure 5가 CIFAR10 test 데이터셋에서 rate-distortion 그래프를 보여준다.

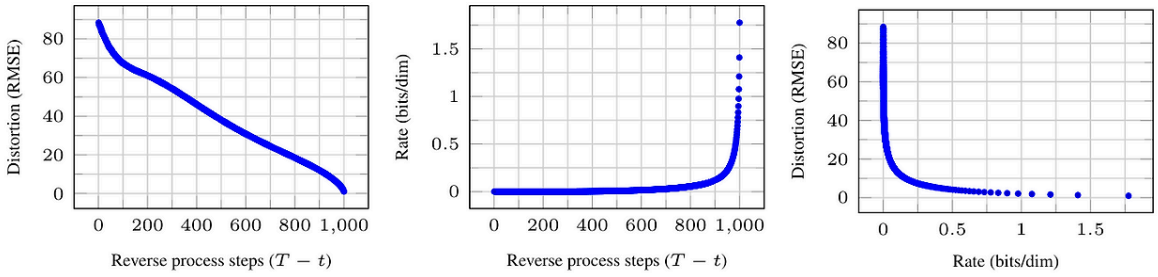


Figure 5: Unconditional CIFAR10 test set rate-distortion vs. time. Distortion is measured in root mean squared error on a $[0, 255]$ scale. See Table 4 for details.

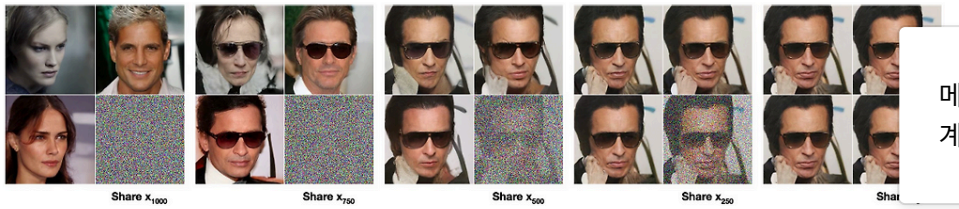
Figure 5

각 시간 t에 대해 distortion은 RMSE로 계산되고, rate는 시간 t부터 지금까지 수신한 누적된 비트 수로 계산된다.
distortion은 rate-distortion 그래프의 low-rate 영역에 대해서 급격히 감소하는데,
이는 **비트의 상당수가 실제로 감지 불가능한 distortion들에 할당되어 있기 때문**이다.

Progressive generation

본 연구는 또한 랜덤 비트들로부터 Progressive depression이 주어진 점진적 비조건 생성 과정을 수행한다.
다시 말하자면, Reverse Process x0 바의 결과를 예측하고, 이는 알고리즘 2를 사용하여 Reverse Process과정으로부터 샘플링 과정을 수행한다는 것이다.

큰 스케일의 이미지 feature가 가장 먼저 나타나고, detail들은 나중에 나타나게 된다.



메뉴바를 클릭하면
계정을 관리할 수 있어요

Figure 7: When conditioned on the same latent, CelebA-HQ 256×256 samples share high-level attributes. Bottom-right quadrants are $\mathbf{x}_t$, and other quadrants are samples from $p_\theta(\mathbf{x}_0|\mathbf{x}_t)$.

Figure 7 : 4사분면 부분이 xt / 다른 사분면들은 p로부터의 샘플들

Figure 7은 xt가 고정된 상태에서 다양한 t값에 대해 확률적 예측 x0를 보여준다.
t가 크면 오직 큰 스케일의 feature들만 보존되지만,
t가 작은 경우에는 반대로 모든 디테일들 중에서도 괜찮은 디테일들만 보존된다.

Connection to autoregressive decoding

Diffusion Process의 길이 T를 데이터 차원에 맞추어 설정한다고 가정한다.
이때의 Forward Process는 q가 x0에서 처음 t개의 자료를 마스킹한 상태로 모든 확률 질량 값들을 두도록 정의된다.

p(xt)를 모든 질량을 빈 이미지에 두도록 설정되었다.
연구에서는 논의의 편의를 위해 p가 완전 표현력 있는 조건부 분포라고 가정한다.

Gaussion Diffusion Model을 일종의 autoregressive model로 해석될 수 있다고 본다.
이때, 데이터 좌표의 순서들을 바꾸는 것으로 표현될 수없는 일반화된 비트 순서를 따른다.

따라서 Gaussian Diffusion도 이와 유사한 역할을 수행한다고 추정하며, 특히 **마스킹이 자연스러운 이미지에**
Gaussian 노이즈를 더하는 방식이 더 적합하다는 점에서 효과적일 수 있다.
또한, Gaussian Diffusion의 단계수가 무조건 데이터 차원과 같을 필요는 없다.

Gaussion Diffusion 과정을 **더 짧게 설정하여 빠른 샘플링과 모델 표현력이 더 높아지도록 활용**도 가능하다.

4.4 Interpolation

확률적 인코더 q를 사용하여 소스 이미지들을 잠재 공간에서 보간할 수 있고, 그 ㅈ

으로 보간된 잠재 벡터

$$\bar{\mathbf{x}}_t = (1-\lambda)\mathbf{x}_0 + \lambda\mathbf{x}'_0$$

를 Reverse Process로 디코딩하여 이미지 공간으로 되돌릴 수가 있다.

이는 잠재 공간에서의 선형 보간을 통해 이미지들 간의 자연스러운 변화 생성이 가능하다는 것이다.

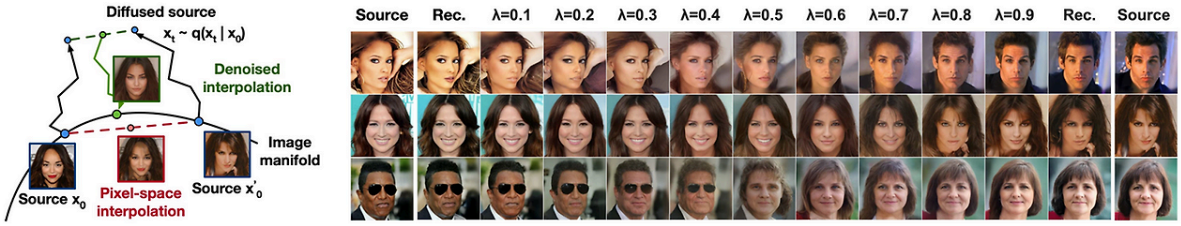


Figure 8: Interpolations of CelebA-HQ 256x256 images with 500 timesteps of diffusion.

Figure 8

실제로, Figure 8에서 묘사된 것처럼 소스 이미지들의 **손상된 부분을 선형 보간하면서 생긴 노이즈들을 제거하기 위해**
Reverse Process를 활용한다.
다른 람다값으로 노이즈를 고정시켜 xt와 xt'를 동일하게 유지시킨다.

Figure 8의 오른쪽 부분을 보면 원본 이미지들의 보간과 재구성 과정을 보여준다.
Reverse Process는 고품질의 재구성과 그럴듯한 보간 결과들을 생성한다.
(피부색, 포즈와 같은 특성은 잘 나타내나 눈동자 같은 특성은 잘 나타내지 못 한다.)

5. Related Work

q의 경우 파라미터가 없고, xT는 고차원의 잠재변수를 의미한다.
데이터 x0에 대해 상호 정보량이 0으로, x0과 xT가 서로 완전히 독립적이라는 것을 의미한다.

입실론 prediction은 노이즈 자체를 예측하여 예측한 값을 기반으로 원래 이미지 x0을 복원한다.

Diffusion Model들은 직관적인 log likelihood를 평가하고,
훈련 과정에서는 variational inference를 이용하여 Langevin dynamics를 훈련시킨다.

이러한 연관성은 Score Matching의 특정 가중화된 형태와 분산 추론과 동일하다

Score Matching 과 에너지 기반 모델링 간의 잘 알려진 연관성을 통해, 연구는 에너지 기반 모델들에 관한
다른 최근의 연구들을 함축한다.
Rate-distortion curves의 경우 variational bound의 평가에서 시간에 따라 계산되는 부분으로,
어떻게 계산될지는 Annealed Importance Sampling의 **단일 실행 과정에서 각 단계에서의 손실값을 측정**하는 것이다.
이때, 각 단계에서의 손실값은 과정에서 나오는 중간 샘플들을 활용하여 distortion 정도를 측정한 값이다.

저차원의 계산들에서 더 일반적인 디자인을 가지고, Autoregressive 모델들에 샘플링 전략들로 이끌어낼 수 있다.

6. Conclusions

본 연구에서는 Diffusion Model들을 활용한 고품질의 이미지 샘플들을 제시하고,
Diffusion Model과 Markov Chain을 훈련하기 위한 분산 추론, Denosing Score Matching과 Progressive lossy compression
간의 연관성들도 찾을 수 있었다.
Diffusion Model들은 이미지 데이터에 대해 매우 뛰어난 귀납적 편향을 갖고 있기 때문에,
앞으로 다른 데이터 형태나 다른 유형의 생성 모델, 머신러닝 시스템의 구성요소로 활용될 가능성을 기대할 수 있었다.

Broader Impact

사회적 측면에서 부정적으로 활용될 가능성이 있다.
가장 먼저, 생성 모델은 고위 인사들의 가짜 이미지·영상 제작 등에 악용될 수 있다. 특히,
CNN 기반의 생성 이미지는 아직까지는 감지 가능한 결함이 있지만, 기술이 더 발전하면 탐지

하는 데에 어려움이 커질 수 있다.

두 번째로, 편향이 재생산될 수 있다. 대규모 데이터셋은 종종 인터넷에서 자동 수
향된 데이터를 포함하게 되는데, 이러한 데이터로 학습된 생성 모델이 인터넷에 더
면, 편향이 증폭될 수 있다.

메뉴바를 클릭하면
계정을 관리할 수 있어요

하지만, 긍정적인 부분도 존재한다.
데이터 압축 부분에서, 인터넷 트래픽이 증가하고 고해상도 데이터가 늘어나면서, 데이터의 접근 가능성을 보장할 수가 있다.

공감

'카테고리 없음'의 다른글

이전글 [Euron] 3week_ Generative Adversarial Nets

현재글 : [Euron] 5week_Denoising Diffusion Probabilistic Models

댓글 0

yejjji
내용을 입력하세요.