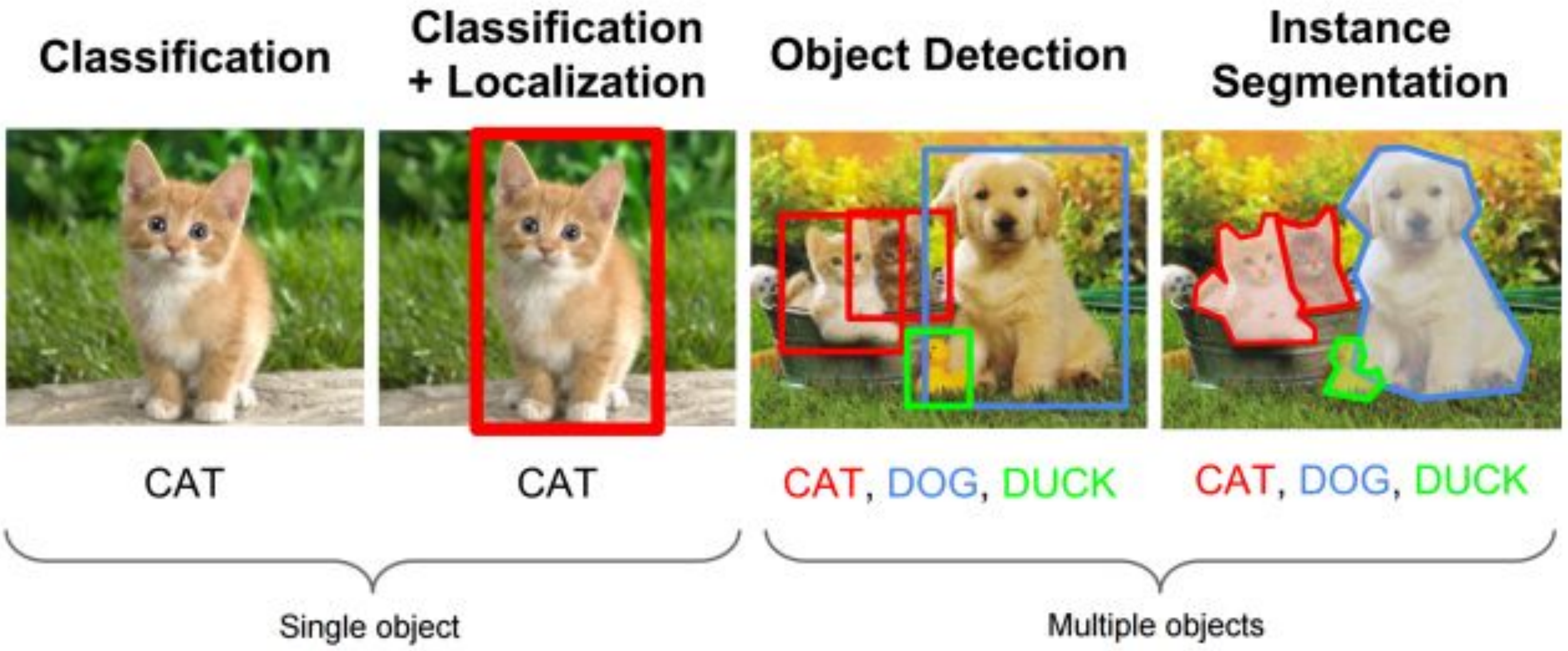
Diffusion Models

Apoorv Agnihotri
Deep Learning Researcher
Rephrase AI

What can AI do?



[1]

Can it modify?

Gatys, Leon & Ecker, Alexander & Bethge, Matthias. (2015). A Neural Algorithm of Artistic Style. arXiv. 10.1167/16.12.326.

Generation?

Karras, Tero & Laine, Samuli & Aila, Timo. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks. 4396-4405. 10.1109/CVPR.2019.00453.

Contents

- What is generative AI?
  - Examples
  - Problem statement
- What are Generative Models
  - VAE
  - GANs
  - Flow - based model
  - Diffusion Models
- Diffusion Models
  - Idea - Thermodynamics 💡
  - Previous works
  - Overview
  - Connection to VAEs
    - Diffusion model $==t$ latent variable hierarchical VAE
    - Into some math:
      - ELBO pt. 1
      - Reconnecting to our objective
      - Simplifying ELBO
    - Takeaways
  - In sum
  - Problems
  - Advancements
    - Text conditioning
- Code
- References

Examples

- *teddy bears working on new AI research on the moon in the 1980s*

generated from dalle-2

Examples

- *Avocados dancing, drinking, singing and partying at a Hawaiian luau*

generated from dalle-2

Examples

- prompt given

generated from [midjourney]()

Problem Statement

- Given a dataset coming from distribution: ***p(x)***, where ***x*** is a datapoint, we want a model ***f***, that can create new objects with the same distribution
- ***f(z)***→***x***, s.t ***x*** comes from ***p(x)***
  - Unconditional
    - seed **z** sampled from noise
  - Conditional
    - **z** sampled from another distribution ***p(z)***

Generative Models

- VAEs
- GANs
- Flow-based Models
- **Diffusion Models**

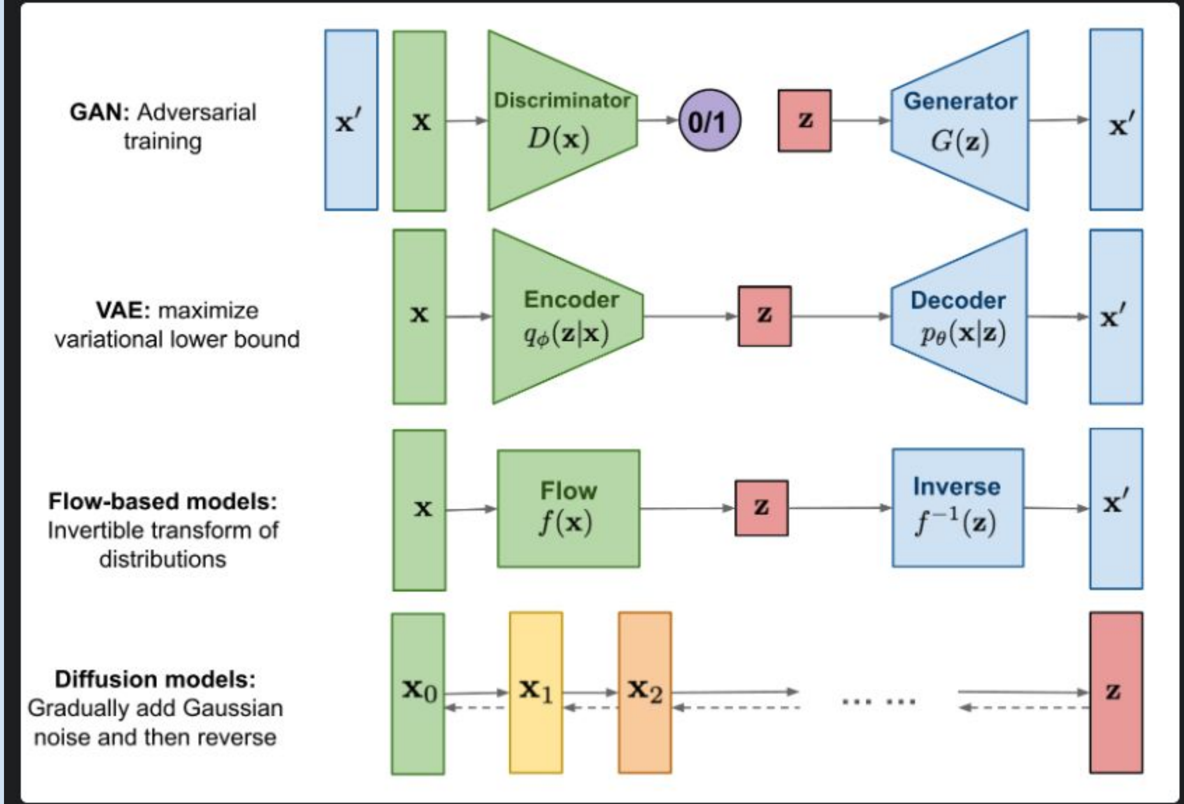


Fig. 1. Overview of different types of generative models.

Weng, Lilian. (Jul 2021). What are diffusion models? Lil'Log. https://lilianweng.github.io/posts/2021-07-11-diffusion-models/.

VAEs

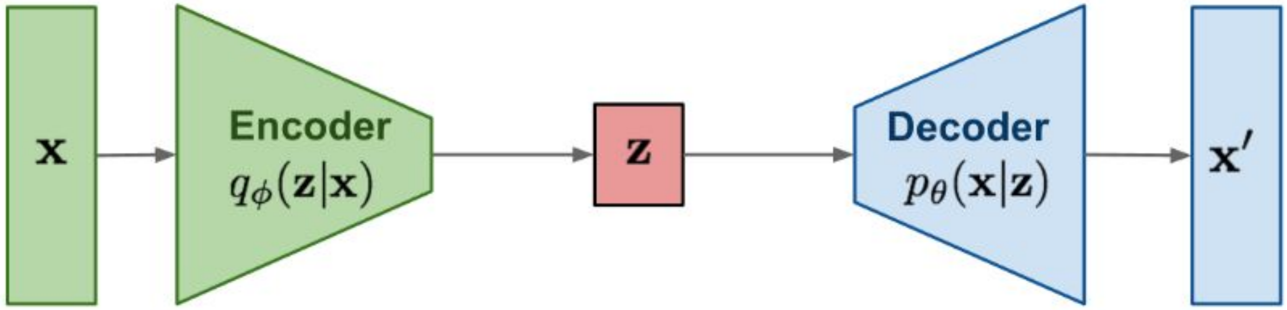
Pros

- Access and easy manipulation of latent space.
- Good for controlling outputs
- Fast inference

Cons

- The fidelity of generated points it low.

VAE: maximize variational lower bound



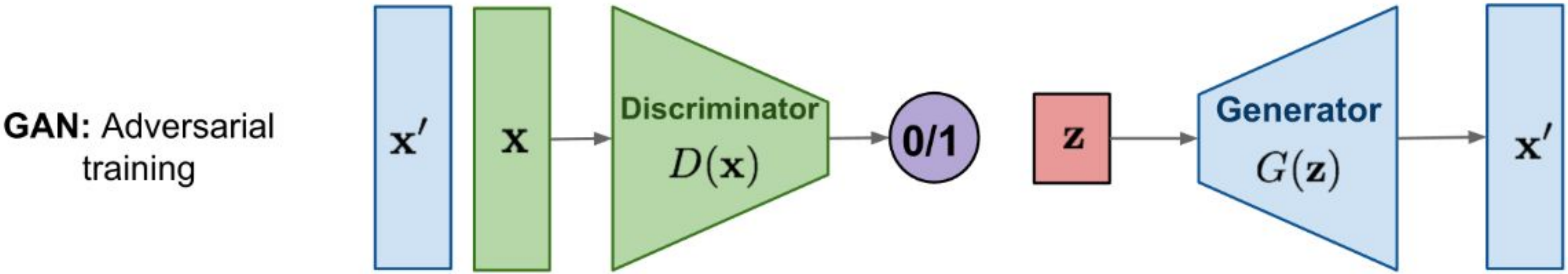
GANs

Pros

- High fidelity images.
- Fast inference

Cons

- Tricky to train
- Latent space isn't accessible, difficult to manipulate



Normalizing Flows

Pros

-

Cons

-

[1]

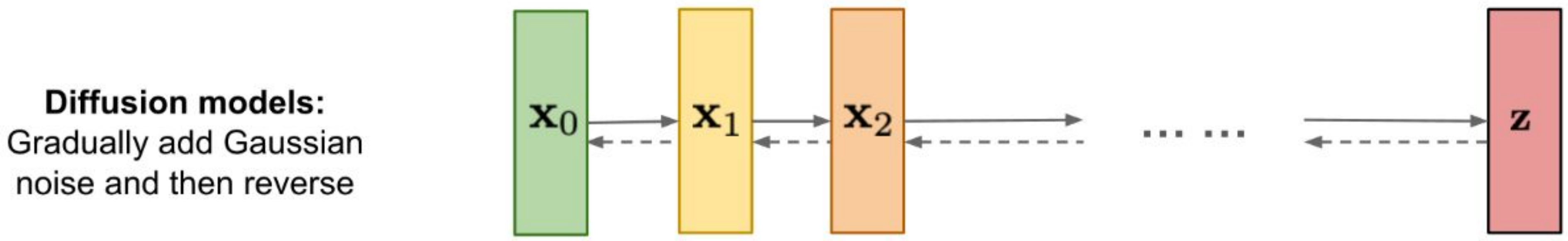
Diffusion Models

Pros

- High fidelity. Even better than GANs.

Cons

- Slow Inference
- Costly to train



[1]

Contents

- What is generative AI?
  - Examples
  - Problem statement
- What are Generative Models
  - VAE
  - GANs
  - Flow - based model
  - Diffusion Models
- Diffusion Models
  - Idea - Thermodynamics 💡
  - Previous works
  - Overview
  - Connection to VAEs
    - Diffusion model $==t$ latent variable hierarchical VAE
    - Into some math:
      - ELBO pt. 1
      - Reconnecting to our objective
      - Simplifying ELBO
    - Takeaways
  - In sum
  - Problems
  - Advancements
    - Text conditioning
- Code
- References

Diffusion Models

Idea

- **Forward Process (perturbing) (~Increasing entropy) →**
- Reverse Process (denoising)

Image: Moussa / Public Domain

Idea

- Forward Process (perturbing)
- **Reverse Process (denoising) (~decreasing noise)**

  →

Image: Moussa / Public Domain

the reverse process seems impossible?

- at the macroscopic level yes, seems impossible
- The idea is that entropy increases in a system on a macroscopic level.
    - example: Energy dissipates from hot food.
    - counterexample: A cold dish, spontaneously turns hot.

Getty Images/iStockphoto

the reverse process is possible, but improbable.

- Second law broken | Nature
- ^ entropy decreased and was observed in **microscopic** experiments. “breaking the second law of thermodynamics”

Go over: Entropy Explained, With Sheep

Getty Images/iStockphoto

Previous works

Dickstein, et al (**2015**) (Stanford)

Ho, et al (**2020**) (Berkeley)

arXiv:1503.03585v8 [cs.LG] 18 Nov 2015

Deep Unsupervised Learning using Nonequilibrium Thermodynamics

Jascha Sohl-Dickstein JASCHA@STANFORD.EDU
Stanford University
Eric A. Weiss EAWEISS@BERKELEY.EDU
University of California, Berkeley
Niru Maheswaranathan NIRUM@STANFORD.EDU
Stanford University
Surya Ganguli SGANGULI@STANFORD.EDU
Stanford University

Abstract

A central problem in machine learning involves modeling complex data-sets using highly flexible families of probability distributions in which learning, sampling, inference, and evaluation are still analytically or computationally tractable. Here, we develop an approach that simultaneously achieves both flexibility and tractability. The essential idea, inspired by non-equilibrium statistical physics, is to systematically and slowly destroy structure in a data distribution through an iterative forward diffusion process. We then learn a reverse diffusion process that restores structure in data, yielding a highly flexible and tractable generative model of the data. This approach allows us to rapidly learn, sample from, and evaluate probabilities in deep generative models with thousands of layers or time steps, as well as to compute conditional and posterior probabilities under the learned model. We additionally release an open source reference implementation of the algorithm.

1. Introduction

Historically, probabilistic models suffer from a tradeoff between two conflicting objectives: *tractability* and *flexibility*. Models that are *tractable* can be analytically evaluated and easily fit to data (e.g. a Gaussian or Laplace). However, these models are unable to aptly describe structure in rich datasets. On the other hand, models that are *flexible* can be molded to fit structure in arbitrary data. For example, we can define models in terms of any (non-negative) function $\phi(\mathbf{x})$ yielding the flexible distribution $p(\mathbf{x}) = \frac{\phi(\mathbf{x})}{Z}$, where Z is a normalization constant. However, computing this normalization constant is generally intractable. Evaluating, training, or drawing samples from such flexible models typically requires a very expensive Monte Carlo process.

A variety of analytic approximations exist which ameliorate, but do not remove, this tradeoff–for instance mean field theory and its expansions (T, 1982; Tanaka, 1998), variational Bayes (Jordan et al., 1999), contrastive divergence (Welling & Hinton, 2002; Hinton, 2002), minimum probability flow (Sohl-Dickstein et al., 2011b;a), minimum KL contraction (Lyu, 2011), proper scoring rules (Gneiting & Raftery, 2007; Parry et al., 2012), score matching (Hyvärinen, 2005), pseudolikelihood (Besag, 1975), loopy belief propagation (Murphy et al., 1999), and many, many more. Non-parametric methods (Gershman & Blei, 2012) can also be very effective[1].

1.1. Diffusion probabilistic models

We present a novel way to define probabilistic models that allows:

1. extreme flexibility in model structure,
2. exact sampling,

Proceedings of the 32^{nd} *International Conference on Machine Learning*, Lille, France, 2015. JMLR: W&CP volume 37. Copyright 2015 by the author(s).

[1] Non-parametric methods can be seen as transitioning smoothly between tractable and flexible models. For instance, a non-parametric Gaussian mixture model will represent a small amount of data using a single Gaussian, but may represent infinite data as a mixture of an infinite number of Gaussians.

arXiv:2006.11239v2 [cs.LG] 16 Dec 2020

Denoising Diffusion Probabilistic Models

Jonathan Ho UC Berkeley jonathanho@berkeley.edu
Ajay Jain UC Berkeley ajayj@berkeley.edu
Pieter Abbeel UC Berkeley pabbeel@cs.berkeley.edu

Abstract

We present high quality image synthesis results using diffusion probabilistic models, a class of latent variable models inspired by considerations from nonequilibrium thermodynamics. Our best results are obtained by training on a weighted variational bound designed according to a novel connection between diffusion probabilistic models and denoising score matching with Langevin dynamics, and our models naturally admit a progressive lossy decompression scheme that can be interpreted as a generalization of autoregressive decoding. On the unconditional CIFAR10 dataset, we obtain an Inception score of 9.46 and a state-of-the-art FID score of 3.17. On 256x256 LSUN, we obtain sample quality similar to ProgressiveGAN. Our implementation is available at `https://github.com/hojonathanho/diffusion`.

1 Introduction

Deep generative models of all kinds have recently exhibited high quality samples in a wide variety of data modalities. Generative adversarial networks (GANs), autoregressive models, flows, and variational autoencoders (VAEs) have synthesized striking image and audio samples [14, 27, 3, 58, 38, 25, 10, 32, 44, 57, 26, 33, 45], and there have been remarkable advances in energy-based modeling and score matching that have produced images comparable to those of GANs [11, 55].

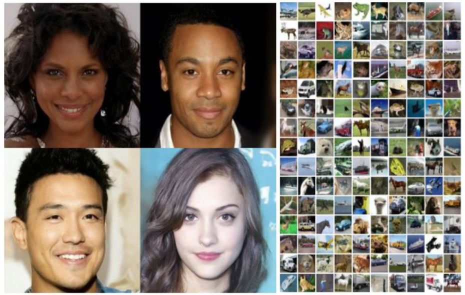


Figure 1: Generated samples on CelebA-HQ 256 × 256 (left) and unconditional CIFAR10 (right)

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

Previous works

Nichol and Dhariwal (2021) (OpenAI)

Nichol and Dhariwal (2021) (OpenAI)

arXiv:2102.09672v1 [cs.LG] 18 Feb 2021

Improved Denoising Diffusion Probabilistic Models

Alex Nichol [*1] Prafulla Dhariwal [*1]

Abstract

Denoising diffusion probabilistic models (DDPM) are a class of generative models which have recently been shown to produce excellent samples. We show that with a few simple modifications, DDPMs can also achieve competitive log-likelihoods while maintaining high sample quality. Additionally, we find that learning variances of the reverse diffusion process allows sampling with an order of magnitude fewer forward passes with a negligible difference in sample quality, which is important for the practical deployment of these models. We additionally use precision and recall to compare how well DDPMs and GANs cover the target distribution. Finally, we show that the sample quality and likelihood of these models scale smoothly with model capacity and training compute, making them easily scalable. We release our code at https://github.com/openai/improved-diffusion.

1. Introduction

Sohl-Dickstein et al. (2015) introduced diffusion probabilistic models, a class of generative models which match a data distribution by learning to reverse a gradual, multi-step noising process. More recently, Ho et al. (2020) showed an equivalence between denoising diffusion probabilistic models (DDPM) and score based generative models (Song & Ermon, 2019; 2020), which learn a gradient of the log-density of the data distribution using denoising score matching (Hyvärinen, 2005). It has recently been shown that this class of models can produce high-quality images (Ho et al., 2020; Song & Ermon, 2020; Jolicoeur-Martineau et al., 2020) and audio (Chen et al., 2020b; Kong et al., 2020), but it has yet to be shown that DDPMs can achieve log-likelihoods competitive with other likelihood-based models such as autoregressive models (van den Oord et al., 2016c) and VAEs (Kingma & Welling, 2013). This raises various questions, such as whether DDPMs are capable of capturing all the modes of a distribution. Furthermore, while Ho et al. (2020) showed extremely good results on the CIFAR-10 (Krizhevsky, 2009) and LSUN (Yu et al., 2015) datasets, it is unclear how well DDPMs scale to datasets with higher diversity such as ImageNet. Finally, while Chen et al. (2020b) found that DDPMs can efficiently generate audio using a small number of sampling steps, it has yet to be shown that the same is true for images.

In this paper, we show that DDPMs can achieve log-likelihoods competitive with other likelihood-based models, even on high-diversity datasets like ImageNet. To more tightly optimise the variational lower-bound (VLB), we learn the reverse process variances using a simple reparameterization and a hybrid learning objective that combines the VLB with the simplified objective from Ho et al. (2020).

We find surprisingly that, with our hybrid objective, our models obtain better log-likelihoods than those obtained by optimizing the log-likelihood directly, and discover that the latter objective has much more gradient noise during training. We show that a simple importance sampling technique reduces this noise and allows us to achieve better log-likelihoods than with the hybrid objective.

After incorporating learned variances into our model, we surprisingly discovered that we could sample in fewer steps from our models with very little change in sample quality. While DDPM (Ho et al., 2020) requires hundreds of forward passes to produce good samples, we can achieve good samples with as few as 50 forward passes, thus speeding up sampling for use in practical applications. In parallel to our work, Song et al. (2020a) develops a different approach to fast sampling, and we compare against their approach, DDIM, in our experiments.

While likelihood is a good metric to compare against other likelihood-based models, we also wanted to compare the distribution coverage of these models with GANs. We use the improved precision and recall metrics (Kynkäänniemi et al., 2019) and discover that diffusion models achieve much higher recall for similar FID, suggesting that they do indeed cover a much larger portion of the target distribution.

Finally, since we expect machine learning models to consume more computational resources in the future, we evaluate the performance of these models as we increase model size and training compute. Similar to (Henighan et al.,

*Equal contribution [1]OpenAI, San Francisco, USA. Correspondence to: <alex@openai.com>, <prafulla@openai.com>.

arXiv:2105.05233v4 [cs.LG] 1 Jun 2021

Diffusion Models Beat GANs on Image Synthesis

Prafulla Dhariwal* OpenAI prafulla@openai.com

Alex Nichol* OpenAI alex@openai.com

Abstract

We show that diffusion models can achieve image sample quality superior to the current state-of-the-art generative models. We achieve this on unconditional image synthesis by finding a better architecture through a series of ablations. For conditional image synthesis, we further improve sample quality with classifier guidance: a simple, compute-efficient method for trading off diversity for fidelity using gradients from a classifier. We achieve an FID of 2.97 on ImageNet 128×128, 4.59 on ImageNet 256×256, and 7.72 on ImageNet 512×512, and we match BigGAN-deep even with as few as 25 forward passes per sample, all while maintaining better coverage of the distribution. Finally, we find that classifier guidance combines well with upsampling diffusion models, further improving FID to 3.94 on ImageNet 256×256 and 3.85 on ImageNet 512×512. We release our code at https://github.com/openai/guided-diffusion.

1 Introduction

Figure 1: Selected samples from our best ImageNet 512×512 model (FID 3.85)

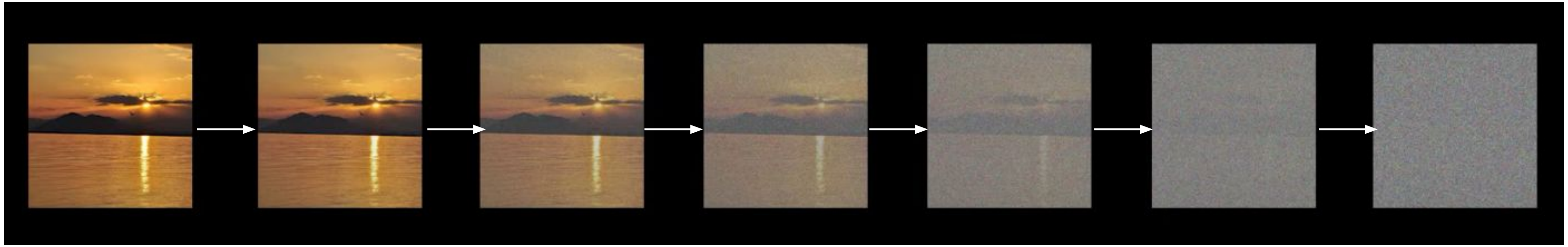
Over the past few years, generative models have gained the ability to generate human-like natural language [6], infinite high-quality synthetic images [5, 28, 51] and highly diverse human speech and music [64, 13]. These models can be used in a variety of ways, such as generating images from text prompts [72, 50] or learning useful feature representations [14, 7]. While these models are already

*Equal contribution

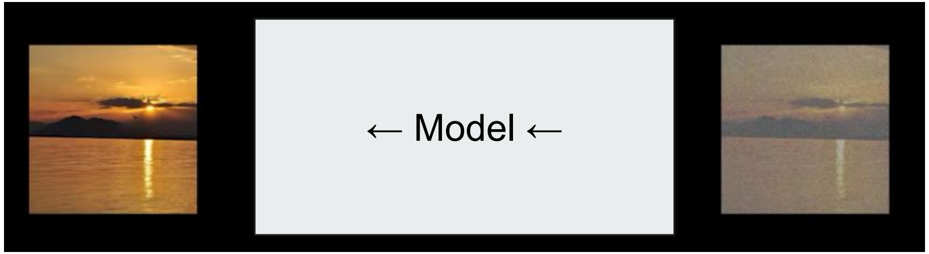
Overview

- Our objective is to model the reverse process (in a small timesteps).

Forward process: Adding noise iteratively



Reverse Process: Recover image from the noised image

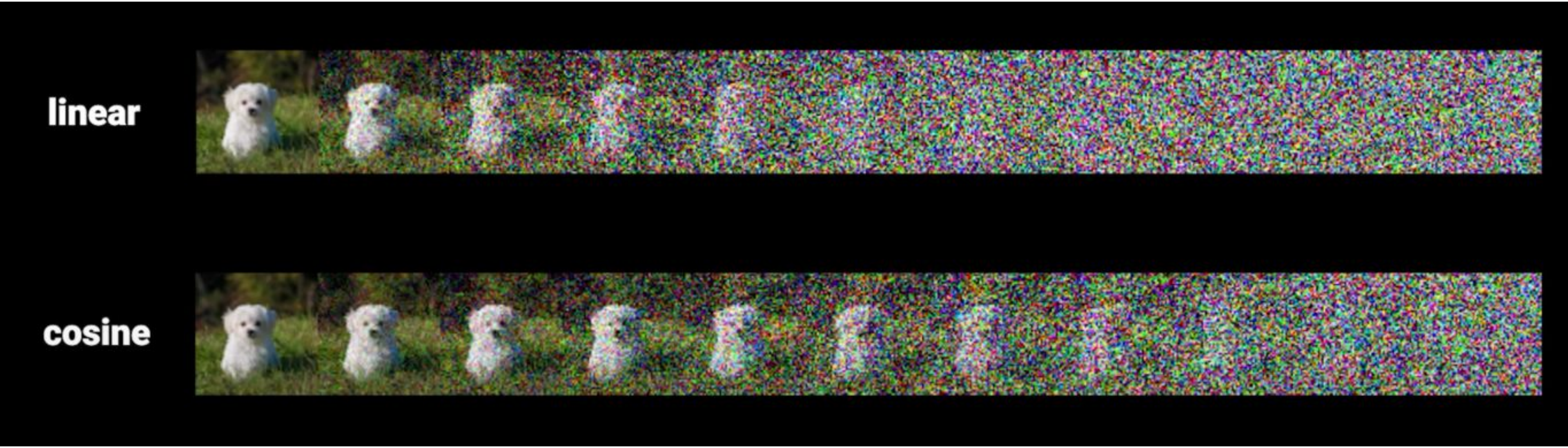


https://www.youtube.com/watch?v=HoKDTa5jHvg

Noise addition

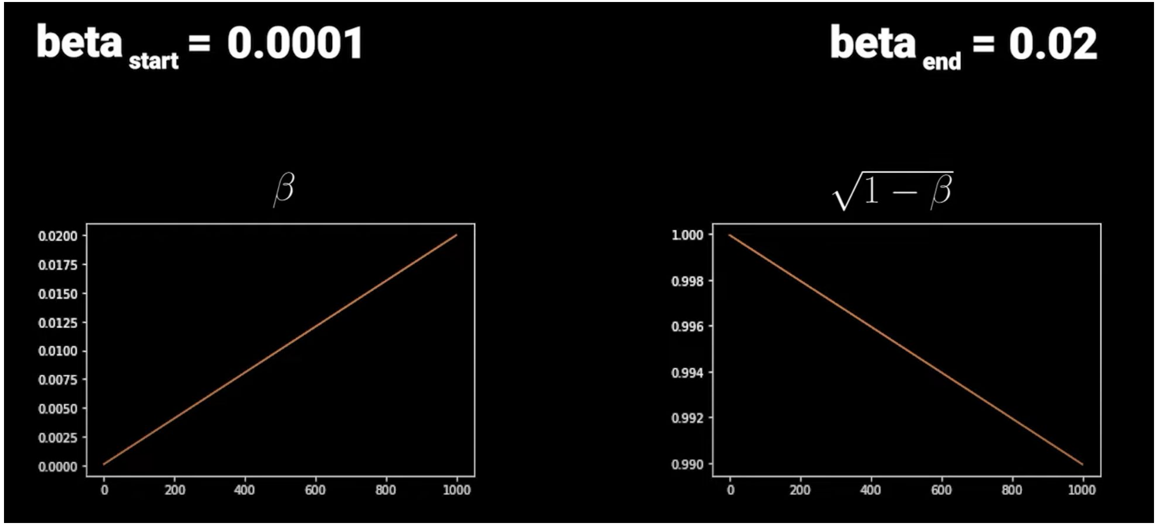
Linear-2020 Paper: Too harsh and last few steps look redundant.

Cosine-2021 Paper: Gently destroying the structure, making it easier for modelling the reverse process



https://www.youtube.com/watch?v=HoKDTa5jHvg

Linear Schedule



$$q(x_t|x_{t-1}) = \mathcal{N}(x_t, \sqrt{1-\beta_t}x_{t-1}, \beta_t I)$$

https://www.youtube.com/watch?v=HoKDTa5jHvg

Linear Schedule - trick

Suppose I want to get t^{th} image after forward process, we can use the fact that addition of t isotropic gaussian is an isotropic gaussian.

$$
\begin{aligned}
q(x_t|x_{t-1}) &= \mathcal{N}(x_t, \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \\
&= \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t}\epsilon \\
&= \sqrt{\alpha_t}\, x_{t-1} + \sqrt{1-\alpha_t}\ \epsilon \\
&= \sqrt{\alpha_t\alpha_{t-1}}\, x_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\ \epsilon \\
&= \sqrt{\alpha_t\alpha_{t-1}\alpha_{t-2}}\, x_{t-3} + \sqrt{1-\alpha_t\alpha_{t-1}\alpha_{t-2}}\ \epsilon \\
&= \sqrt{\alpha_t\alpha_{t-1}\ldots\alpha_1\alpha_0}\, x_0 + \sqrt{1-\alpha_t\alpha_{t-1}\ldots\alpha_1\alpha_0}\ \epsilon \\
&\boxed{= \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\ \epsilon}
\end{aligned}
$$

Notation

$$\alpha_t = 1 - \beta_t$$

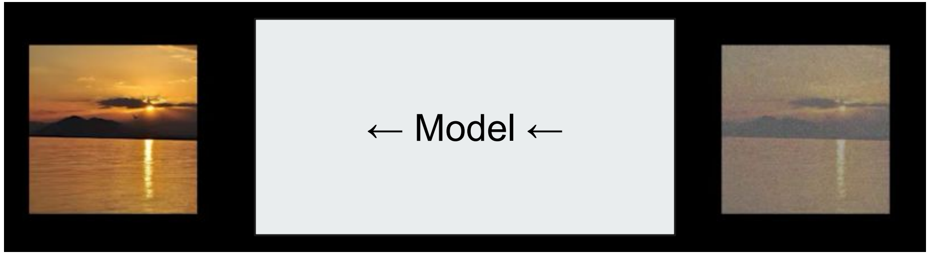
$$\bar{\alpha}_t = \prod_{s=1}^{t} a_s$$

How to recover the image?

Since: The destructive process add isotropic gaussian with varying mean and covariance at each step.

$$N(\mu, \sigma^2)$$

Therefore: In reverse process, the NN models the last *added noise* as a function of x_t and timestep *t*.



- μ_t and Σ_t (of the noise added) modelled by the network

https://www.youtube.com/watch?v=HoKDTa5jHvg

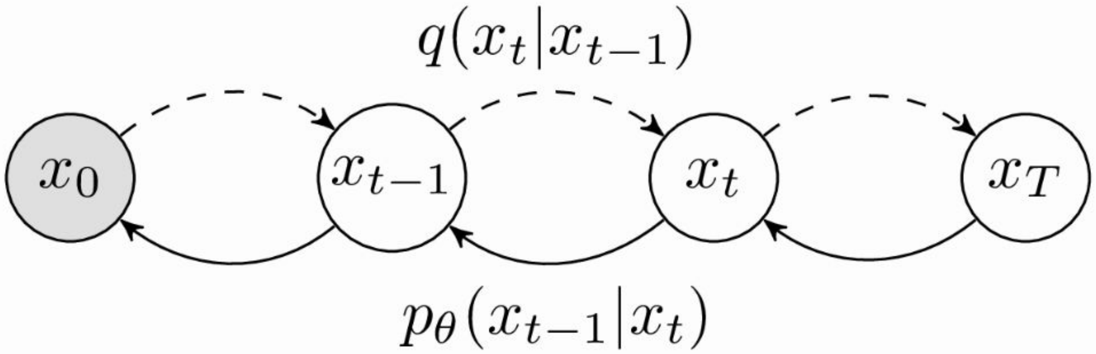
Contents

- What is generative AI?
    - Examples
    - Problem statement
- What are Generative Models
    - VAE
    - GANs
    - Flow - based model
    - Diffusion Models
- Diffusion Models
    - Idea - Thermodynamics 💡
    - Previous works
    - Overview
    - Connection to VAEs
        - Diffusion model == *t* latent variable hierarchical VAE
        - Math:
            - ELBO pt. 1
            - Reconnecting to our objective
            - Simplifying ELBO
        - Takeaways
    - In sum
    - Problems
    - Advancements
        - Text conditioning
- Code
- References

Into the math

We can think of diffusion models as PGMs

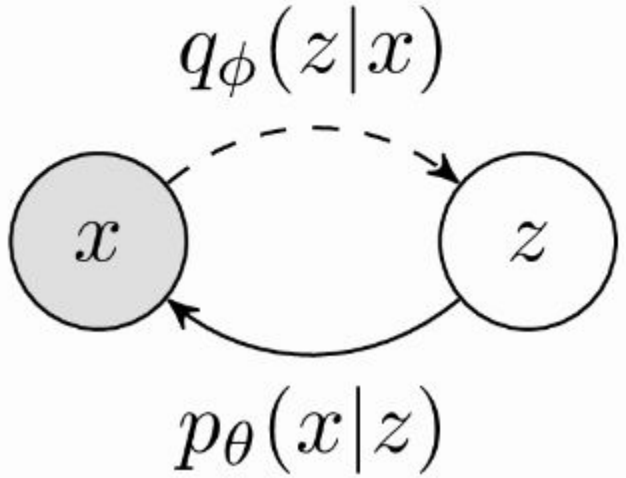
Figure 3 - Diffusion Probabilistic Model



Connections to VAE

Diffusion models: a special VAE.

Figure 1 - Graphical Model for VAE

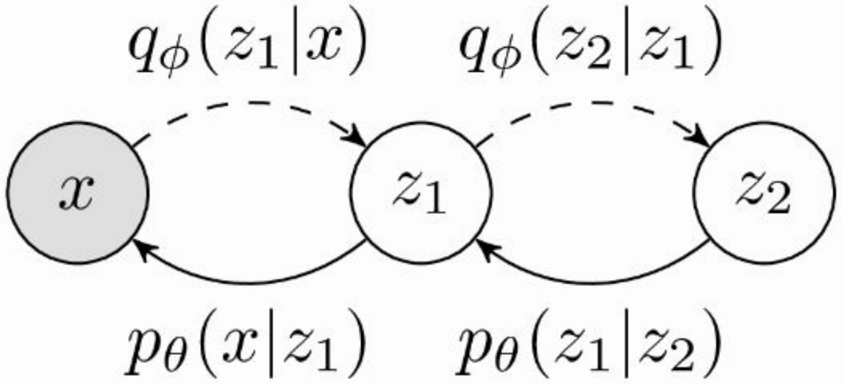


2 latent variable VAE

If ***x*** came from two latents variables?

$$p(x) = \int_{z_1} \int_{z_2} p_\theta(x, z_1, z_2) dz_1, dz_2$$

Figure 2 - A Hierarchical VAE



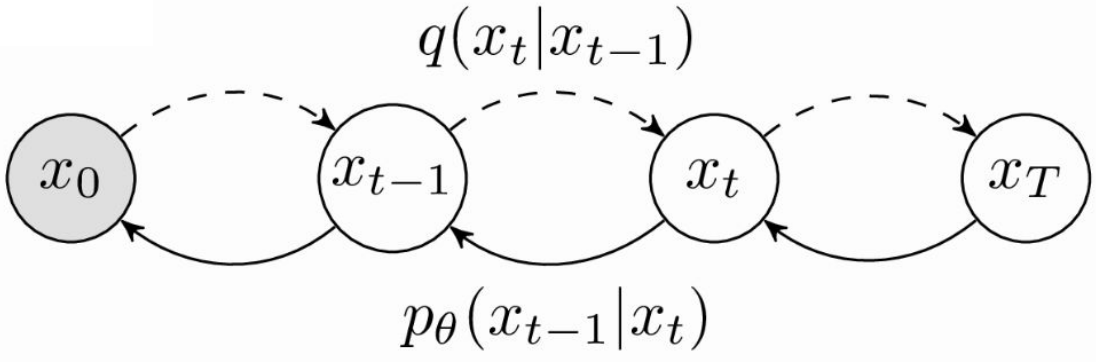
Angus Turner (2021), Diffusion Models as a kind of VAE. https://angusturner.github.io/generative_models/2021/06/29/diffusion-probabilistic-models-I.html

T latent variable VAE

extrapolating it to *T* latent variables

$$p(x_0) = \int_{x_1} \dots \int_{x_T} p_\theta(x_0, x_1, \dots, x_T) dx_1, , \dots, dx_T$$

Figure 3 - Diffusion Probabilistic Model



Angus Turner (2021), Diffusion Models as a kind of VAE. https://angusturner.github.io/generative_models/2021/06/29/diffusion-probabilistic-models-I.html

Contents

- What is generative AI?
    - Examples
    - Problem statement
- What are Generative Models
    - VAE
    - GANs
    - Flow - based model
    - Diffusion Models
- Diffusion Models
    - Idea - Thermodynamics 💡
    - Previous works
    - Overview
    - Connection to VAEs
        - Diffusion model == *t* latent variable hierarchical VAE
        - Math:
            - ELBO pt. 1
            - Reconnecting to our objective
            - Simplifying ELBO
        - Takeaways
    - In sum
    - Problems
    - Advancements
        - Text conditioning
- Code
- References

Refresher: MLE on VAE → ELBO

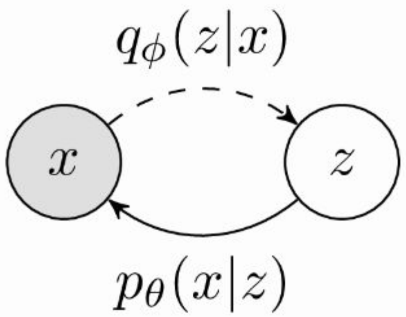
- To learn the parameters θ, let us try to perform maximum likelihood estimation on the generator.

$$\log p_{\boldsymbol{\theta}}(\mathbf{x}) = \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[\log p_{\boldsymbol{\theta}}(\mathbf{x})\right] \tag{2.5}$$

$$= \underbrace{\mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[\log\left[\frac{p_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})}{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\right]\right]}_{\substack{=\mathcal{L}_{\boldsymbol{\theta},\boldsymbol{\phi}}(\mathbf{x}) \\ \text{(ELBO)}}} + \underbrace{\mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[\log\left[\frac{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}{p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})}\right]\right]}_{=D_{KL}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})||p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x}))} \tag{2.8}$$

Evidence Lower BOund

Figure 1 - Graphical Model for VAE



Diederik P. Kingma and Max Welling (2019), "An Introduction to Variational Autoencoders", Foundations and Trends® in Machine Learning: Vol. 12: No. 4, pp 307-392. http://dx.doi.org/10.1561/2200000056

Refresher: MLE on VAE → ELBO

- To learn the parameters θ, let us try to perform maximum likelihood estimation on the generator.

$$\log p_{\theta}(\mathbf{x}) = \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})}\left[\log p_{\theta}(\mathbf{x})\right] \tag{2.5}$$

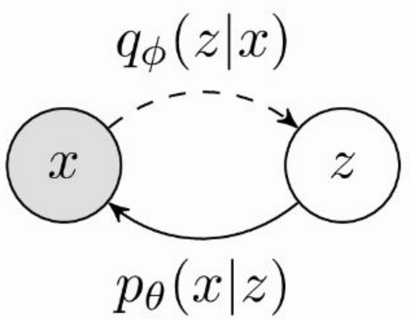
$$= \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})}\left[\log \left[\frac{p_{\theta}(\mathbf{x},\mathbf{z})}{p_{\theta}(\mathbf{z}|\mathbf{x})}\right]\right] \tag{2.6}$$

$$= \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})}\left[\log \left[\frac{p_{\theta}(\mathbf{x},\mathbf{z})}{q_{\phi}(\mathbf{z}|\mathbf{x})}\frac{q_{\phi}(\mathbf{z}|\mathbf{x})}{p_{\theta}(\mathbf{z}|\mathbf{x})}\right]\right] \tag{2.7}$$

$$= \underbrace{\mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})}\left[\log \left[\frac{p_{\theta}(\mathbf{x},\mathbf{z})}{q_{\phi}(\mathbf{z}|\mathbf{x})}\right]\right]}_{\substack{=\mathcal{L}_{\theta,\phi}(\mathbf{x}) \\ \text{(ELBO)}}} + \underbrace{\mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})}\left[\log \left[\frac{q_{\phi}(\mathbf{z}|\mathbf{x})}{p_{\theta}(\mathbf{z}|\mathbf{x})}\right]\right]}_{=D_{KL}(q_{\phi}(\mathbf{z}|\mathbf{x})||p_{\theta}(\mathbf{z}|\mathbf{x}))} \tag{2.8}$$

Evidence Lower BOund

Figure 1 - Graphical Model for VAE



Diederik P. Kingma and Max Welling (2019), "An Introduction to Variational Autoencoders", Foundations and Trends® in Machine Learning: Vol. 12: No. 4, pp 307-392. http://dx.doi.org/10.1561/2200000056

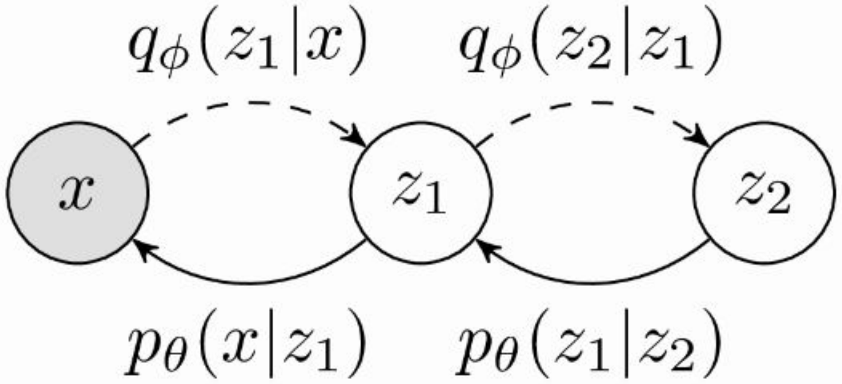
2 latent variable VAE

Marginal for 2 latent variables becomes:

$$p(x) = \int\int q_\phi(z_1, z_2|x) \frac{p_\theta(x, z_1, z_2)}{q_\phi(z_1, z_2|x)}$$

$$p(x) = \mathbb{E}_{z_1, z_2 \sim q_\phi(z_1, z_2|x)} \left[\frac{p_\theta(x, z_1, z_2)}{q_\phi(z_1, z_2|x)} \right]$$

Figure 2 - A Hierarchical VAE



Angus Turner (2021), Diffusion Models as a kind of VAE. https://angusturner.github.io/generative_models/2021/06/29/diffusion-probabilistic-models-I.html

2 latent variable VAE

$$\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log\left[\frac{p_\theta(\mathbf{x},\mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x})}\right]\right]$$

2 latent variable VAE

$$\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log\left[\frac{p_\theta(\mathbf{x},\mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x})}\right]\right]$$

New factorizations:

$$p(x, z_1, z_2) = p(x|z_1)p(z_1|z_2)p(z_2)$$

$$q(z_1, z_2|x) = q(z_1|x)q(z_2|z_1)$$

2 latent variable VAE

$$\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log\left[\frac{p_\theta(\mathbf{x},\mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x})}\right]\right]$$

New factorizations:

$$p(x, z_1, z_2) = p(x|z_1)p(z_1|z_2)p(z_2)$$

$$q(z_1, z_2|x) = q(z_1|x)q(z_2|z_1)$$

Substituting them in ELBO:

$$\mathcal{L}(\theta, \phi) = \mathbb{E}_{q(z_1,z_2|x)}\left[\log p(x|z_1) - \log q(z_1|x) + \log p(z_1|z_2) - \log q(z_2|z_1) + \log p(z_2)\right]$$

2 latent variable VAE

Substituting them in ELBO:

$$\mathcal{L}(\theta, \phi) = \mathbb{E}_{q(z_1, z_2|x)} [\log p(x|z_1) - \log q(z_1|x) + \log p(z_1|z_2) - \log q(z_2|z_1) + \log p(z_2)]$$

can be written as:

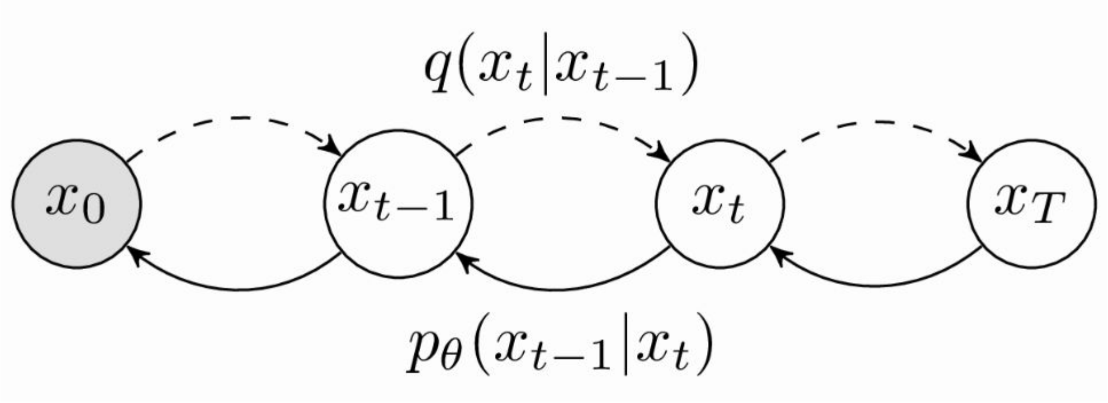
$$\mathcal{L} = \mathbb{E}_q(z_1, z_2|x) \left[\log p(z_2) + \sum_{i \geq 1}^{2} \log \frac{p(z_{i-1} \mid z_i)}{q(z_i \mid z_{i-1})} \right]$$

$$z_0 = x$$

Angus Turner (2021), *Diffusion Models as a kind of VAE. https://angusturner.github.io/generative_models/2021/06/29/diffusion-probabilistic-models-I.html*

t latent variable VAE

Simply extrapolating last equation:



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

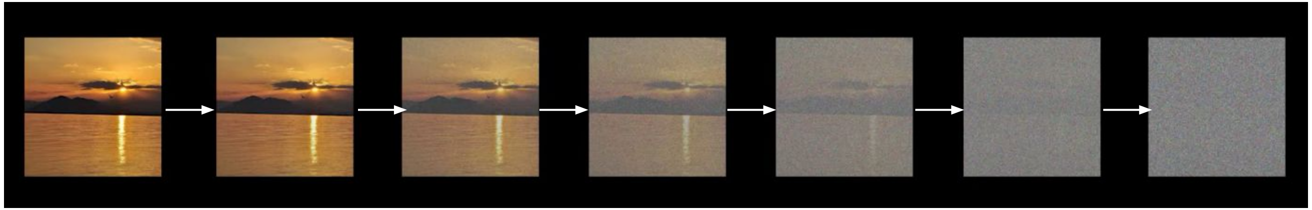
Angus Turner (2021), Diffusion Models as a kind of VAE. https://angusturner.github.io/generative_models/2021/06/29/diffusion-probabilistic-models-I.html

With me?

- We showed diffusion models are special VAEs.
- The special hierarchical VAE has *t* latent variables.
- Wanted to maximize the likelihood for the data
- Reusing ELBO in *Special VAEs* (read diffusion models)
    - Next: Simplify ELBO (with markov property in latents)

Encoder (Forward Diffusion)

Forward process == Encoder
Adding noise iteratively (markov property)



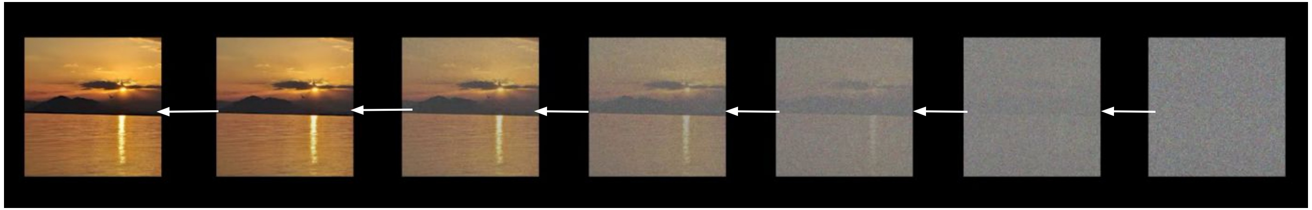
x_0 x_T

$$q(x_t|x_{t-1}) = \mathcal{N}(x_T\ ;\ x_{t-1}\sqrt{1-\beta_t}, \beta_t I)$$

Angus Turner (2021), Diffusion Models as a kind of VAE. https://angusturner.github.io/generative_models/2021/06/29/diffusion-probabilistic-models-I.html

Decoder (Reverse Diffusion)

Reverse process == Decoder (Generator)
Removing noise iteratively (modelled using a NN) (markov property)



x_0 x_T

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

Angus Turner (2021), Diffusion Models as a kind of VAE. https://angusturner.github.io/generative_models/2021/06/29/diffusion-probabilistic-models-I.html

Simplifying ELBO

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

Angus Turner (2021), Diffusion Models as a kind of VAE. https://angusturner.github.io/generative_models/2021/06/29/diffusion-probabilistic-models-I.html

Simplifying ELBO

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

We know (via markov property):

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

Angus Turner (2021), Diffusion Models as a kind of VAE. https://angusturner.github.io/generative_models/2021/06/29/diffusion-probabilistic-models-I.html

Simplifying ELBO

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

We know (via markov property):

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

With Bayes rule:

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

Angus Turner (2021), Diffusion Models as a kind of VAE. https://angusturner.github.io/generative_models/2021/06/29/diffusion-probabilistic-models-I.html

Simplifying ELBO

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

We know (via markov property):

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

With Bayes rule:

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

$$\mathbb{E}_q \left[-\log p(x_T) - \log \frac{p_\theta(x_0|x_1)}{q(x_1|x_0)} - \sum_{t>1}^{T} \left(\log \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t, x_0)} + \log \frac{q(x_{t-1}|x_0)}{q(x_t|x_0)} \right) \right]$$

Angus Turner (2021), Diffusion Models as a kind of VAE. https://angusturner.github.io/generative_models/2021/06/29/diffusion-probabilistic-models-I.html

Simplifying ELBO

$$\mathbb{E}_q\left[-\log p(x_T) - \log \frac{p_\theta(x_0|x_1)}{q(x_1|x_0)} - \sum_{t>1}^{T}\left(\log \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t, x_0)} + \log \frac{q(x_{t-1}|x_0)}{q(x_t|x_0)}\right)\right]$$

We see that the conditionals get cancelled upon expanding.

$$\cancel{\log q(x_1|x_0)} - \cancel{\log q(x_1|x_0)} + \cdots + \log q(x_T|x_0)$$

Angus Turner (2021), Diffusion Models as a kind of VAE. https://angusturner.github.io/generative_models/2021/06/29/diffusion-probabilistic-models-I.html

Simplifying ELBO

$$L := \mathbb{E}_q \left[\underbrace{-\log p(x_T) + \log q(x_T|x_0)}_{L_T} - \underbrace{\log p_\theta(x_0|x_1)}_{L_0} - \underbrace{\sum_{t>1}^{T} \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t, x_0)}}_{L_{t-1}} \right]$$

- L_T has no parameters. (T = last step, x_T is assumed standard normal)
- Both L_T and L_{t-1} are KL divergence between gaussians. Easy (analytical) calculation.
- L_0 reconstruction loss.

Angus Turner (2021), Diffusion Models as a kind of VAE. https://angusturner.github.io/generative_models/2021/06/29/diffusion-probabilistic-models-I.html

Congratulations 🥳

We just derived the loss for original the diffusion models (2015)

Contents

- What is generative AI?
  - Examples
  - Problem statement
- What are Generative Models
  - VAE
  - GANs
  - Flow - based model
  - Diffusion Models
- Diffusion Models
  - Idea - Thermodynamics 💡
  - Previous works
  - Overview
  - Connection to VAEs
    - Diffusion model == *t* latent variable hierarchical VAE
    - Math:
      - ELBO pt. 1
      - Reconnecting to our objective
      - Simplifying ELBO
    - Takeaways
  - In sum
  - Problems
  - Advancements
    - Text conditioning
- Code
- References

Takeaways - Connection to VAE

- Diffusion models can be understood as special VAE models.
- We have a similar objective to maximize the data likelihood
- The encoder (diffusion process := $q(x_t|x_{t-1})$) doesn't involve any learning.
- The latent variables remain the same shape as inputs.
- The whole model is a decoder ($p(x_{t-1}|x_t)$).
- Both encoder and decoder possess markov properties.

Angus Turner (2021), Diffusion Models as a kind of VAE. https://angusturner.github.io/generative_models/2021/06/29/diffusion-probabilistic-models-I.html

In sum

Diffusion Models (Decoder := $p(x_{t-1} | x_t)$) take 2 inputs:

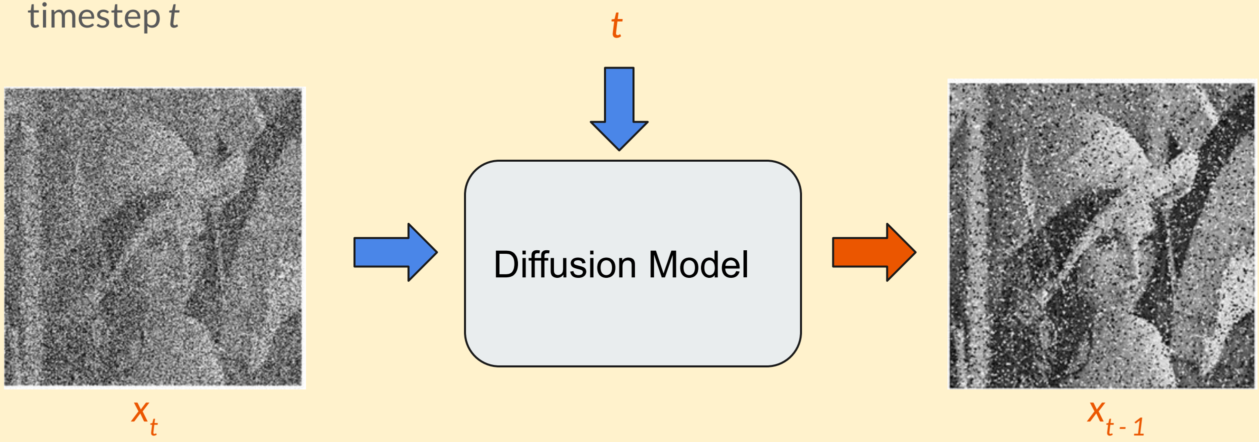
- timestep *t*
- noisy image at timestep *t*

Outputs:

- noise added at time *t - 1* to generate noisy image at timestep *t*

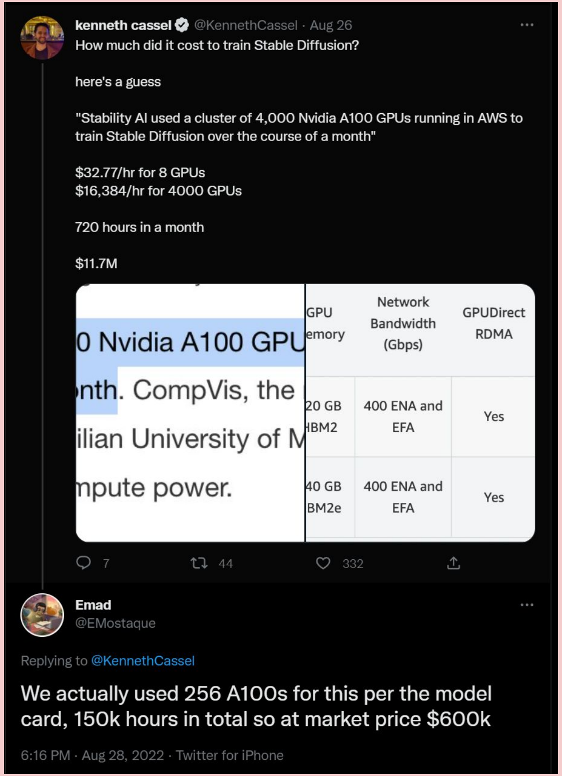
While Optimizing:

ELBO *(previous slide)*



Diffusion Models - Problems

- Too costly to train:
  Stable Diffusion: The model was trained using 256 Nvidia A100 GPUs on Amazon Web Services for a total of 150,000 GPU-hours, at a cost of $600,000.



https://twitter.com/EMostaque/status/1563870674111832066?s=20&t=4Fvxch05FN6JqrAU8hcLkQ

Diffusion Models - Problems

- Too costly to train:
  Stable Diffusion: The model was trained using 256 Nvidia A100 GPUs on Amazon Web Services for a total of 150,000 GPU-hours, at a cost of $600,000.

- We don't completely understand them:
  Primary idea used throughout diffusion literature is adding "noise" to destroy data. A paper called, *Cold Diffusion* used **deterministic** image transformations to create similar results.

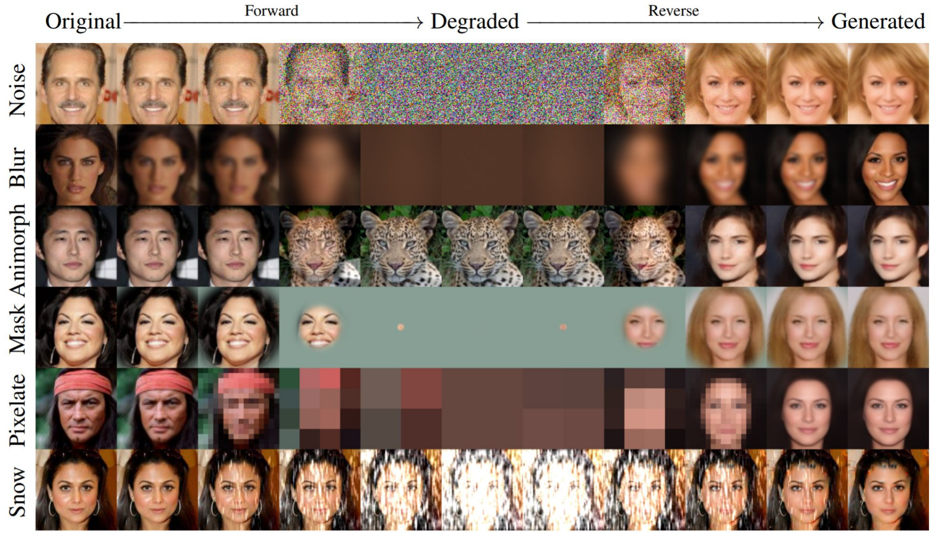
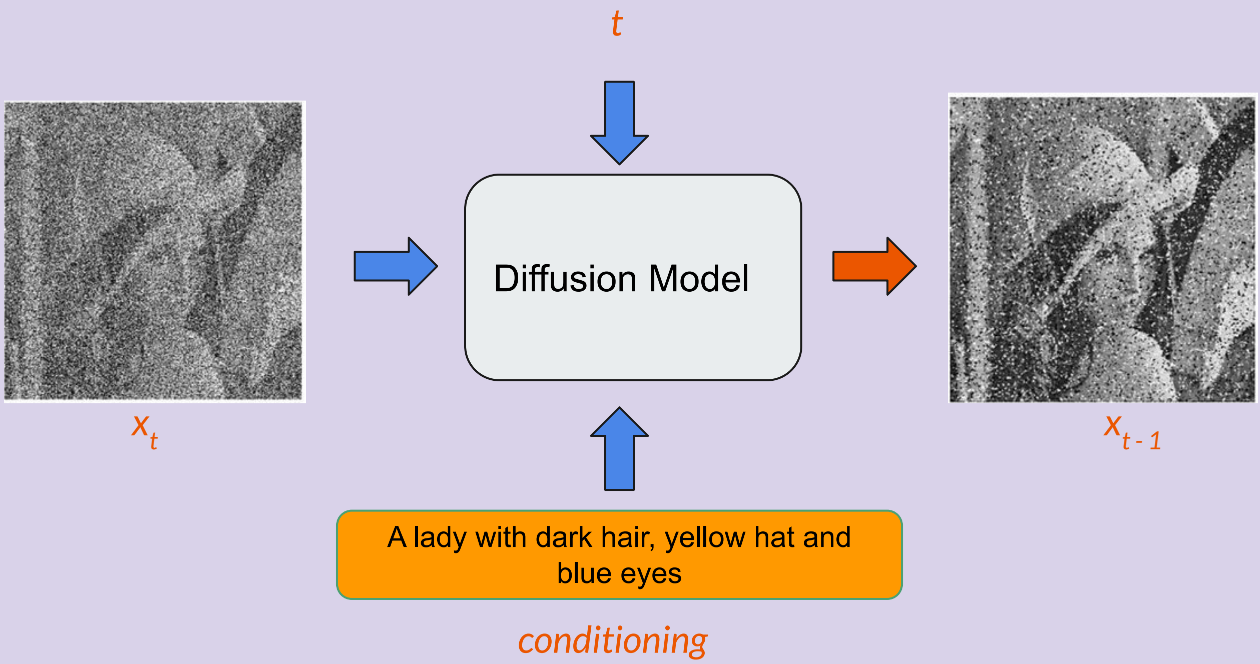


Figure 1: Demonstration of the forward and backward processes for both hot and cold diffusions. While standard diffusions are built on Gaussian noise (top row), we show that generative models can be built on arbitrary and even noiseless/cold image transforms, including the ImageNet-C *snowification* operator, and an *animorphosis* operator that adds a random animal image from AFHQ.

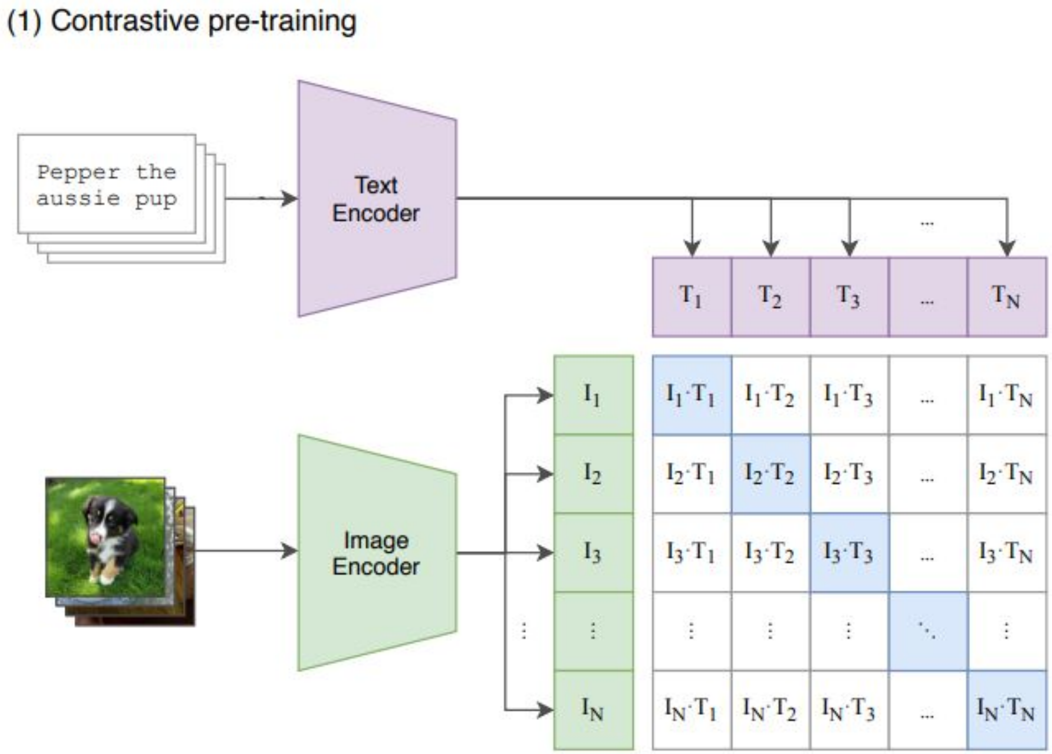
Bansal, Arpit, et al. "Cold diffusion: Inverting arbitrary image transforms without noise." arXiv preprint arXiv:2208.09392 (**2022**).

Advancements - Conditioning



Nichol, Alex & Dhariwal, Prafulla & Ramesh, Aditya & Shyam, Pranav & Mishkin, Pamela & McGrew, Bob & Sutskever, Ilya & Chen, Mark. (2021). GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models.

Clip - Training Image and Text encoders



Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever: Learning Transferable Visual Models From Natural Language Supervision. ICML 2021: 8748-8763

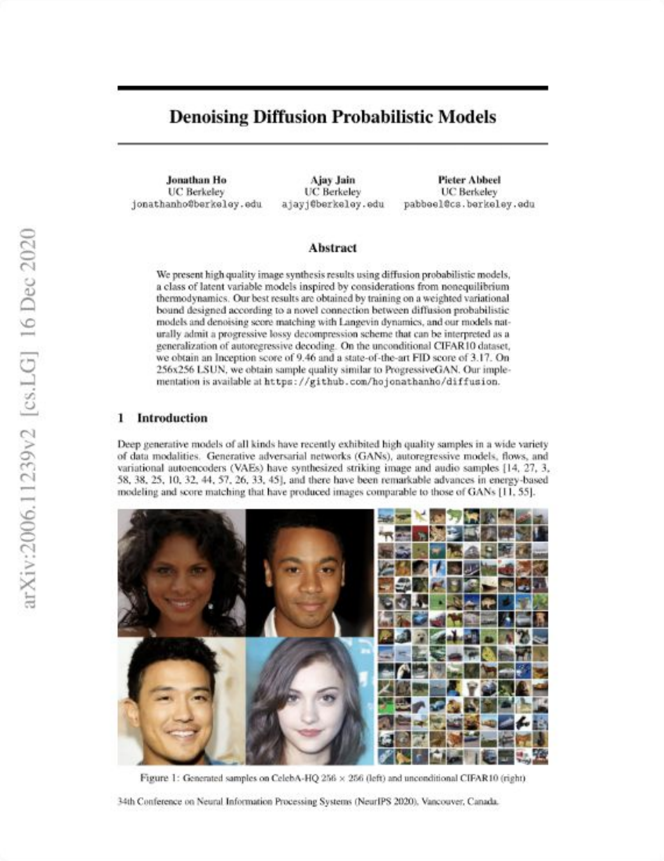
Code

Talks about theory while parallely implementing a full blown diffusion model.

Check it out ↓

[The Annotated Diffusion Model]

```python
from IPython.display import Image
Image(filename='assets/78_annotated-diffusion/ddpm_paper.png')
```



Denoising Diffusion Probabilistic Models

Jonathan Ho, UC Berkeley, jonathanho@berkeley.edu
Ajay Jain, UC Berkeley, ajayj@berkeley.edu
Pieter Abbeel, UC Berkeley, pabbeel@cs.berkeley.edu

arXiv:2006.11239v2 [cs.LG] 16 Dec 2020

Abstract

We present high quality image synthesis results using diffusion probabilistic models, a class of latent variable models inspired by considerations from nonequilibrium thermodynamics. Our best results are obtained by training on a weighted variational bound designed according to a novel connection between diffusion probabilistic models and denoising score matching with Langevin dynamics, and our models naturally admit a progressive lossy decompression scheme that can be interpreted as a generalization of autoregressive decoding. On the unconditional CIFAR10 dataset, we obtain an Inception score of 9.46 and a state-of-the-art FID score of 3.17. On 256x256 LSUN, we obtain sample quality similar to ProgressiveGAN. Our implementation is available at https://github.com/hojonathanho/diffusion.

1 Introduction

Deep generative models of all kinds have recently exhibited high quality samples in a wide variety of data modalities. Generative adversarial networks (GANs), autoregressive models, flows, and variational autoencoders (VAEs) have synthesized striking image and audio samples [14, 27, 3, 58, 38, 25, 10, 32, 44, 57, 26, 33, 45], and there have been remarkable advances in energy-based modeling and score matching that have produced images comparable to those of GANs [11, 55].

Figure 1: Generated samples on CelebA-HQ 256 × 256 (left) and unconditional CIFAR10 (right)

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

That's all

Any questions? Reach out to me. :)

ApoorvAgnihotr2

www.linkedin.com/in/apoorvagni

apoorvagni@gmail.com

References

[1] - What are Diffusion Models? | Lil'Log
[2] - https://artificialintelligence.oodles.io/wp-content/uploads/2020/08/computer-vision.png\
[3] - Diffusion Models as a kind of VAE | Angus Turner
[4] - Entropy Explained, With Sheep
[5] - Diffusion Models | Paper Explanation | Math Explained