An Implementation of Denoising Diffusion Probabilistic Models

Yug Ajmera
Athrva Pandhare

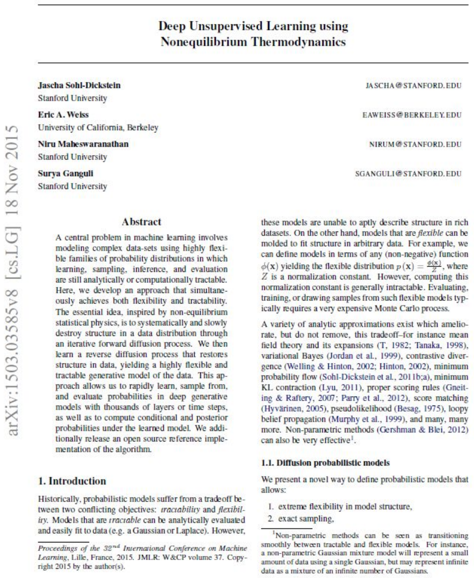
Penn Engineering

Motivation

Recent breakthrough in Generative AI: Open AI's DALL-E 2 and GLIDE, Google's Imagen, and Stability AI's Stable Diffusion.

Penn Engineering

Literature Survey



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

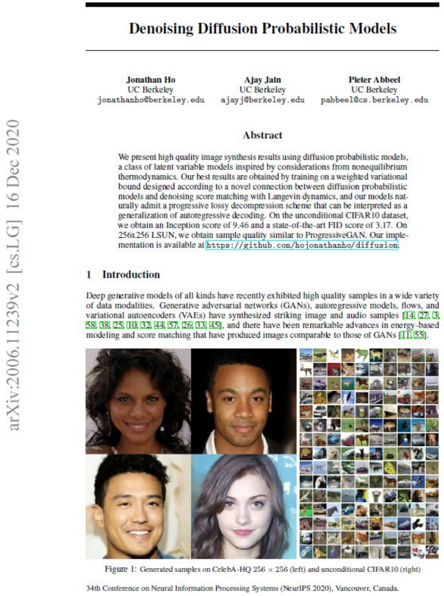
We present a novel way to define probabilistic models that allows:

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

Proceedings of the 32nd 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.

Sohl-Dickstein (2015)



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)

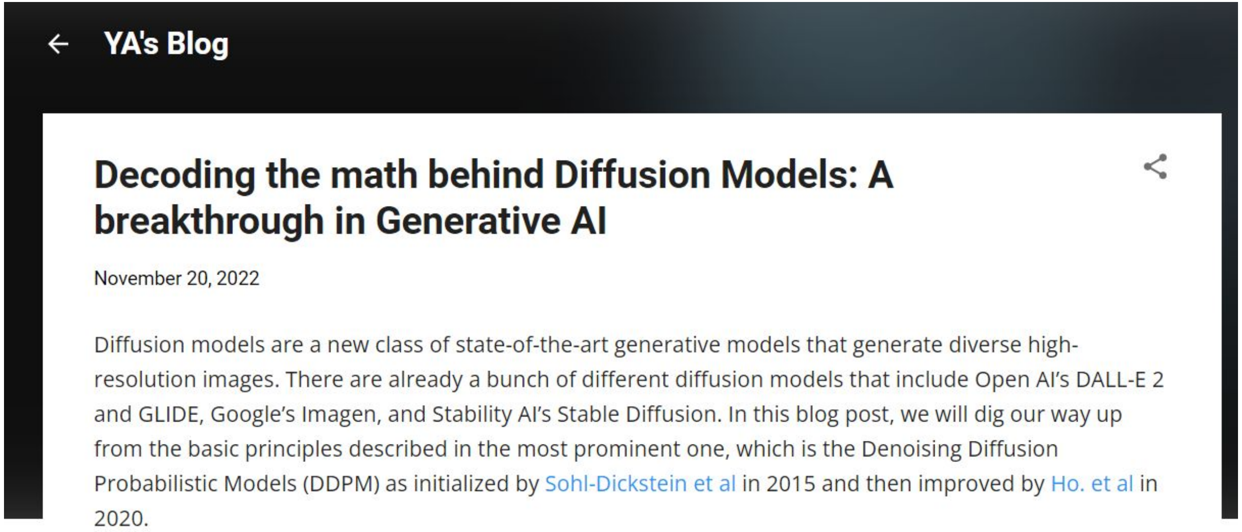
34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

Ho et al. (2020)

Penn Engineering

Full Derivation

https://yainnoware.blogspot.com/2022/11/decoding-math-behind-diffusion-models.html



YA's Blog

Decoding the math behind Diffusion Models: A breakthrough in Generative AI

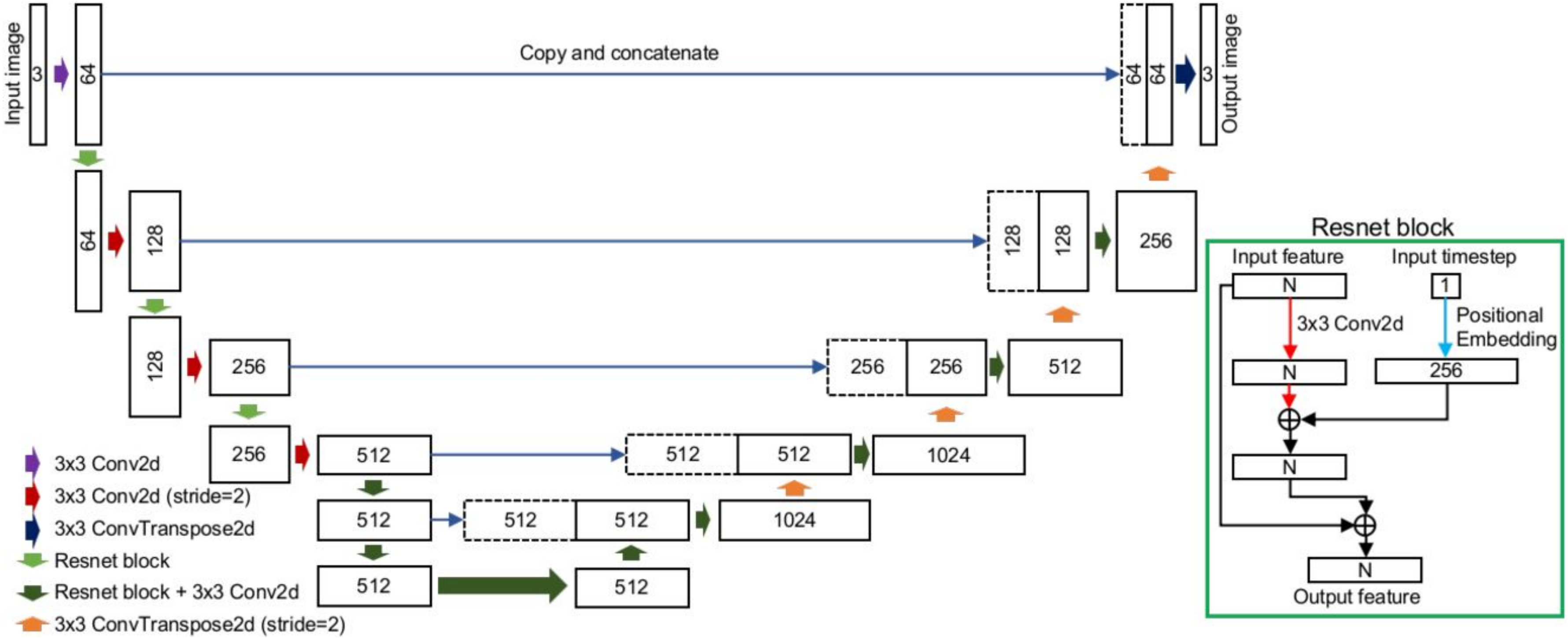
November 20, 2022

Diffusion models are a new class of state-of-the-art generative models that generate diverse high-resolution images. There are already a bunch of different diffusion models that include Open AI's DALL-E 2 and GLIDE, Google's Imagen, and Stability AI's Stable Diffusion. In this blog post, we will dig our way up from the basic principles described in the most prominent one, which is the Denoising Diffusion Probabilistic Models (DDPM) as initialized by Sohl-Dickstein et al in 2015 and then improved by Ho. et al in 2020.

UNet Architecture

1. This implementation is called the Conditional UNet, primarily because of the time embedding.
2. The architecture contains ResNet blocks near the bottleneck interleaved with attention blocks.
3. Group normalization is used before the Attention blocks.
4. Time embedding (also called the position embedding due to similarities with the position embedding in the Transformer) allows the neural network to produced outputs *conditioned* on the current timestep.

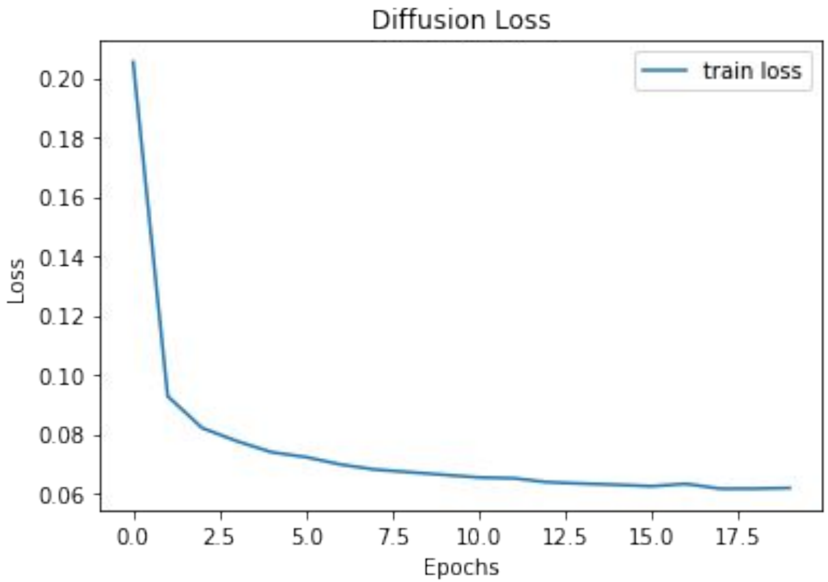
Architecture



Ablation study

Schedule	Loss	Inception Score (IS)
Linear	L1	3.3707
Cosine	L1	1.614
Quadratic	L1	3.6521
Sigmoid	L1	3.304
Linear	L2	3.3808
Cosine	L2	2.6078
Quadratic	L2	3.4592
Sigmoid	L2	3.3833
Linear	Smooth-L1	3.4503
Cosine	Smooth-L1	2.1523
Quadratic	Smooth-L1	3.878
Sigmoid	Smooth-L1	3.6121

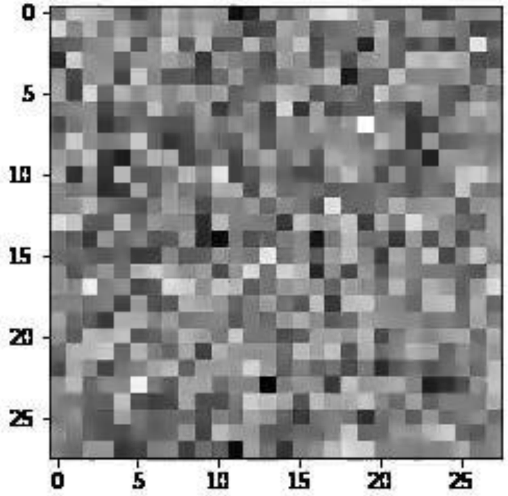
Table 1. Results of the Ablation Study



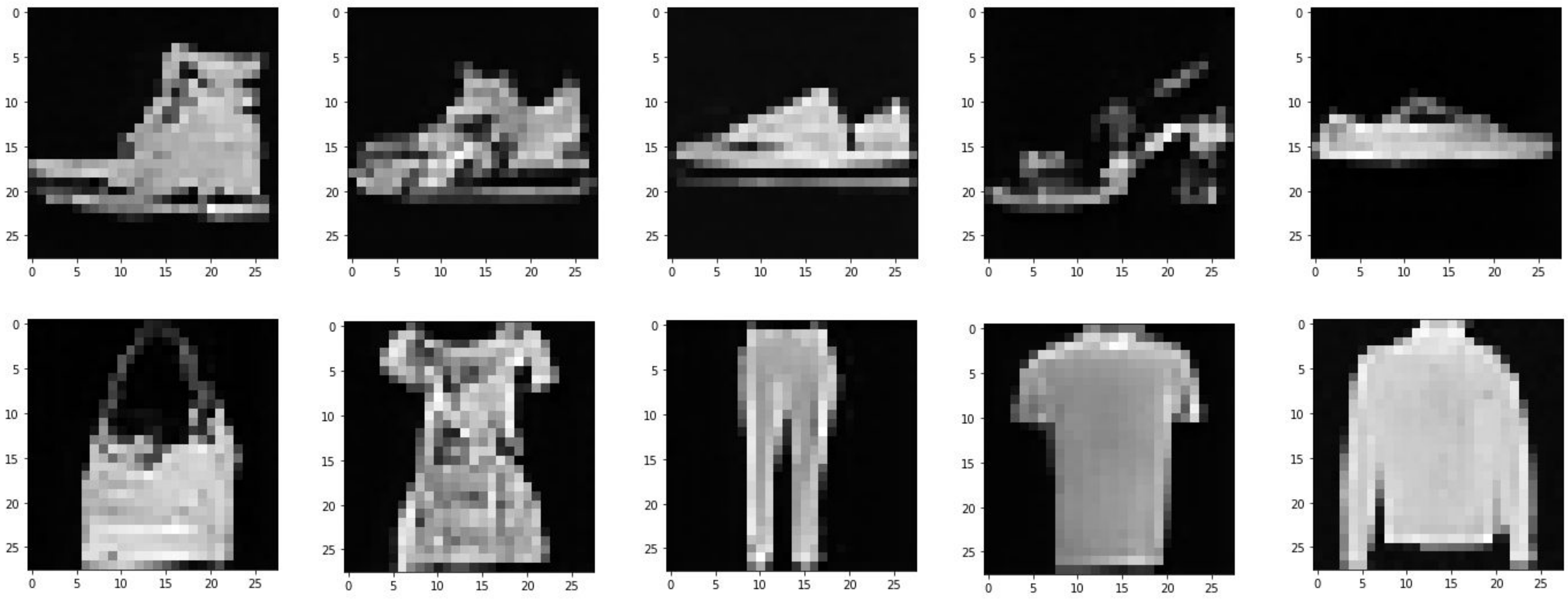
Results

Model	Inception Score (IS)
Real Test Set	4.0885
Huber + Quadratic (20 epochs)	4.1842

Table 2. Results with test set



Visualizations on MNIST Fashion dataset



Extending the DDPM to Complex Dataset

We extend the implemented DDPM to a more complex dataset (Stanford Cars dataset).

Model Changes :

1. Increased the depth of the UNet architecture
2. Increased the number of channels in the UNet architecture.

Implementation Details :

1. Used a linear schedule for beta.
2. The loss between the predicted noise and the true noise was L2-loss.

Results of Training on the Cars Dataset

We think that using a Larger model / using Latent Diffusion can produce better results.

Penn Engineering

Thank You!