Innovative Class-Conditional Image Generation: A Comparative Study

Ethan Vertal
Statistics and Machine Learning
Carnegie Mellon University
Pittsburgh, PA 15213
evertal@andrew.cmu.edu

Mudita Sai
Statistics and Machine Learning
Carnegie Mellon University
Pittsburgh, PA 15213
mksai@andrew.cmu.edu

Supriya Shingade
Statistics and Machine Learning
Carnegie Mellon University
Pittsburgh, PA 15213
sshingad@andrew.cmu.edu

1 Introduction

We will perform class-conditional image generation using ImageNet data with the intention of comparing experimental results of these new methods to the standard Denoising Diffusion Probabilistic Models or shortly, DDPM (Ho et. al., 2020). We propose a novel implementation of a Reverse Diffusion Monte Carlo (rdMC) sampler and re-implementation of a DDPM and Score-based Generative diffusion model via stochastic differential equations. Our hope is that the Monte Carlo sampling-based approximation methods will achieve near or better performance on this task, as measured by Fréchet Inception Distance (FID).

2 Dataset and Task

We will be using a subset of the ImageNet Large Scale Visual Recognition Challenge (ILVRC) dataset. We will first begin with using a training dataset subsetted from a single class to evaluate the model's performance on a singular class-level basis and if time and memory permits, we will extend our training dataset to span over multiple classes. The reason we are subsetting the dataset as such is because we are aiming to experiment with various Monte Carlo sampling techniques in place of the UNet in the reverse diffusion process. That being said, we will first be focusing on a class-conditional image generation task of cats.

3 Related Work

There exist four broad principal categories of generative models (Song et. al., 2021), which are as follows. The first of which are GANs that perform in a min-max adversarial fashion. Secondly, Variational Autoencoders (VAEs), autoencoders, and normalizing flows belong to the category rooted in likelihood-based approaches. The third category encompasses energy-based modeling, wherein the distribution is modeled as an energy function that is subsequently normalized. The final category includes score-based matching where the score of the energy based model is learned as a neural net. Current state-of-the-art diffusion models, such as DDPMs, leverage denoising score matching with annealed Langevin dynamics and a parameterized Markov chain trained via variational inference to generate samples approximating the underlying data distribution (Ho et al., 2020). This falls under our second category of generative models. However, it is noteworth that these various generative models more or less make use of similar objective functions. Current literature in diffusion models showcase three equivalent objectives that optimize the model by learning a neural network to 1. predict the original image x_0, 2. predict the source noise ϵ_0 and 3. compute the score of the image, $\nabla \log p(x_t)$ (Luo, 2022). Nonetheless, a crucial element in reversing the learned diffusion model hinges upon the efficiency of convergence from any complex distribution to a normal distribution. Apart from the 3 methods mentioned above, there exists a non-parametric method that claims to achieve better performance as well as computational efficiency (Dong et. al., 2024). Termed the reverse diffusion Monte Carlo (rdMC) algorithm, this approach samples data from an unnormalized distribution and is designed in such a way as to avoid learning score functions. As a result, this algorithm does not require a parameterized diffusion model to be trained. The rdMC diverges from conventional Markov Chain Monte Carlo (MCMC) techniques in utilizing a reverse Stochastic Differential Equation (SDE) Ornstein-Uhlenbeck (OU) process, in contrast to Langevin dynamics employed in MCMC, for sampling from the score function vector field.

4 Approach

Since the forward process of a diffusion model has no learnable parameters and can be computed efficiently in one pass with the help of the reparametrization trick, our focus will be on the more interesting reverse process which in essence is the denoising process and can be implemented in a multitude of ways. And as such, we will try to implement the new reverse diffusion Monte Carlo (RdMC) sampling technique. The basis for vanilla diffusion models relies on parameterizing the presumed posterior distribution instead of directly sampling from the target distribution. It uses the underlying assumptions of a Markov Chain and Gaussian distributions, along with a UNet to estimate noise, to provide a closed form solution to the reverse process. So as our first step, we will simply be re-implementing DDPM for class-conditional image generation. However, an rdMC estimates this reverse process with an Ornstein-Uhlenbeck process which also models the noise from a standard normal to the target distribution. This is a relatively new approach and will be a novel implementation, to our knowledge. So while we might be able to find some code to build upon, it might become a from-scratch implementation. Our third approach would be a score-matching generative diffusion model, where we estimate $\nabla \log p(x_t)$, as discussed above. We'll build upon an existing implementation, aiming to reproduce its results and attempt to compare the performance of this model with the other two models mentioned earlier.

5 Expected Outcomes

We anticipate that this novel algorithm will effectively sample the target distribution within a reasonable range of accuracy and to be computationally more efficient than the MCMC algorithms currently in use even in the case of multi-modal distributions. In practice, we will evaluate the performance by generating images corresponding to a particular class and judge by eye, as well as using the Fréchet Inception Distance (FID) metric. Our aim is to produce images that are highly realistic in appearance and achieve a low FID score.

6 Plan

- Major milestones:
  - Collect data
  - Set up work environment
  - Data pipeline
  - Class structures organized
  - DDPM working
  - Create score based generative model using the online tutorials/works: colab notebook
  - rdMC written, debugged
  - compare the images generated from all models
- The goal is to have the DDPM working by the midway report make significant progress on the other models. Ideally, we would like to get some form of working model for all three methods.

- We plan on working on the major programming parts as a group to ensure mutual understanding of the concepts involved. We are all interested in these methods and would value the experience working on them.
- However, should we run into issues scheduling meeting times, we will all work on setting up the environment and getting set up to develop the models as a group. We will divide the work for the models evenly, but with the goal of finishing the DDPM first as we feel this will be the least amount of new code.
- We will each complete two sections of the Midway Summary.

Bibliography

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models (arXiv:2006.11239). arXiv. `http://arxiv.org/abs/2006.11239`

Huang, X., Dong, H., Hao, Y., Ma, Y.-A., & Zhang, T. (2024). Reverse Diffusion Monte Carlo (arXiv:2307.02037). arXiv. `http://arxiv.org/abs/2307.02037`

Luo, C. (2022). Understanding Diffusion Models: A Unified Perspective (arXiv:2208.11970). arXiv. `http://arxiv.org/abs/2208.11970`

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2021). Score-Based Generative Modeling through Stochastic Differential Equations (arXiv:2011.13456). arXiv. `http://arxiv.org/abs/2011.13456`