Reading Skills Assignment (DS 899)
on
Score-Based Generative Modeling through Stochastic Differential Equations

Jimut Bahan Pal
22D1594

March 31, 2023

1 Review of the paper using 5C method

Category: This is a statistical deep-learning paper that provides a new method for sampling and generating data samples by the use of stochastic differential equations.

Context: There are cutting-edge papers that came during/before the publication of this paper related to Denoising Diffusion Probabilistic Models (DDPM) [1], which use Langevin Dynamics for obtaining samples from a probability density. The current paper relies on solving reverse time Stochastic Differential Equations (SDE) for generating samples that will have lesser sampling time compared to DDPMs.

Correctness: Of course, the assumption is valid both theoretically and when applied in the real world with a wide range of ablation studies and experiments. The paper takes ideas from Neural Ordinary Differential Equations (ODEs) [2], Variational Auto-Encoders (VAEs) [3], DDPMs etc. The code base is available, and this is a published paper at ICLR 2021, which has a rigorous peer review in terms of code and concept presented in the paper.

Contributions: The main contributions of the paper are they present a framework for score-based generative modeling, which is better than DDPMs since this family of models does not suffer from slow sampling and also achieves better quality images. Other important contributions of this work are it allows fast adaptive sampling via a black-box ODE solver, flexible manipulation of latent code, and also uniquely identifiable encoding for each training procedure with exact likelihood computation.

Clarity: The paper is well written, with enough background information for each of the methods they have proposed and described. They have also been successfully able to provide the motivation for proposing such a novel architecture and methodology for efficient sampling.

2 Third Pass

Overall

What was the article type?: The article was a research article focusing on advanced deep learning and statistical methods, which concentrated on theory and extensive analytical results to approach the conclusion.

What was the title?: The title of the paper was "Score-Based Generative Modeling through Stochastic Differential Equations."

Who were the authors?: The authors were in the following order, Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole.

Introduction

What was the research problem?: The main research problem was to create data from noise by the use of generative modeling. The author presented a stochastic differential equation (SDE) that smoothly transforms a complex data distribution (like image data distribution) to noise by slowly injecting noise and a corresponding reverse-time SDE that does the reverse process by transforming the noise to data distribution; hence, generating the original data distribution (e.g., images) from noise.

Was there any mention of previous studies on this topic?: Several topics can be thought of as previous studies of this topic that deal with the generation of data via noise. The oldest one being Generative Adverserial Networks (GANs) and Variational Auto-Encoders (VAEs). There has been recent work in this direction, like Normalizing Flows and DDPMs, all of which have been thoroughly compared and studied as benchmarks in this paper in terms of comparison of benchmarks.

What were the aims and objectives of the study?: The main aim of this study was to create a different form of sampling technique which can be used to generate data from noise, which is better than GANs, i.e., they don't suffer mode collapse and faster than DDPMs in sampling. Now there is a trade off between sampling and generating high quality diverse samples, so some techniques which generate high quality diverse samples are very hard to train and samples slowly from the model, whereas faster sampler like GANs sufferers from mode collapse issue, which means they will not be able to produce diverse samples. This study tries to mitigate the two constraints which is optimizing all those tradeoffs and doing its best.

Materials and Methods

Which variables were measured?: Computer vision poses a great variety of interesting problems for Generative networks, where the networks train on a set of images, and produces samples from the same distribution. Here, the images that the model were trained on belonged to the category of natural images. Hence the most popular metrics that are required to compare two models are Frechet Inception Distance (FID) and Inception Score (IS). The authors have also measured the Negative Log-Likelihood (NLL) on the test dataset. A low FID and NLL means the generated data is close to the natural imagery, hence it is better, and a high IS is better for this task.

What equipment/instruments were used for data collection? Were they appropriate?: The data were collected from standard repository which are widely used and cited in the Deep Learning literature. Those images are collected from the web and are very well curated for research purposes.

What statistical methods/tests were employed? Were they apt for evaluation?: Training of generative model and collecting results using the standard metrics (discussed above) were done for this study. This is usually common in case of comparing generative networks which generates images from the domain of natural imagery.

Results

What were the key findings?: The key finding of this study are-

- Exact likelihood computation, which was previously not possible in any architectures.
- Manipulating latent representations and uniquely identifiable latent code for any dataset, which was previously not provided by other methods.
- Efficient sampling, which produces high-quality samples from a given dataset.

Were the results reliable?: The codebase is open-sourced, and the community has rigorously tested their hypothesis and claims. This is a fundamental contribution in terms of generative models, which provides efficient and reliable sampling.

Which results were statistically significant?: Table 2 of the paper shows the results obtained in the test set of CIFAR 10 which is a widely used dataset in the computer vision community. Their ablation studies show that the Negative Log Likelihood and Frechet Inception Distance is the lowest when compared with other state-of-the-art competitive models. The unconditional version of the model shows significant improvement in the Inception Score and Frechet Inception Distance.

Discussions

Did the results answer the research question?: The results showed that stochastic differential equations can be used for producing high-quality data samples when trained using the DDPM framework, which surpassed the previous state-of-the-art architectures.

How were these results different/similar when compared to other studies?: The results obtained from the set of experiments were unified with the results obtained by the other models/architectures, by using the same set of metrics which showed that the results were better when compared with the related models.

What were the limitations of the study?: One of the major limitations of the study is the sampling is better and faster than DDPM models, but slower than GANs. This is because, GANs have other limitations like mode collapse which are mitigated by the training and sampling strategy proposed in this study.

Conclusions

What were the conclusions?: The main conclusion drawn from this study is that the method performs significantly better than GANs and DDPMs in terms of generating high-quality image samples when trained on standard datasets.

Was the study worth doing?: Yes. A new framework has been implemented which will result in other derivative work in terms of generating samples in this direction.

Does the reader have any questions unanswered by the article?: The reader (Jimut) is highly unqualified to answer this question since the reader does not have the necessary background to fully understand the complex equations presented in this study. Of course, no paper is perfect, and maybe 5 years down the line, he will be able to point out the unanswered questions that could be done in an unknown research area by just going through the paper a few times. For now, I (Jimut) think they (the authors) could have a follow-up paper that explains how the quality of samples might further increase when increasing the breadth of the samplers, as discussed in the conclusion section, which might be future work for this study.

3 What did you find difficult during the reading article in class (and in general)

Personal Note: These are my personal views, and these should not be taken to reflect the views of any organization I'm affiliated with.

The article has an in-depth derivation of mathematical equations, which might be very challenging for newcomers in the field. Machine Learning and its allied research areas are an ever-changing field where thousands of new ideas creep in over the course of time. An idea today might be outdated a few days after it has been published since there is a lot of competition in the field, and every idea provides a new perspective on a problem. It is becoming increasingly difficult to get the hang of all the research areas, so the focus should be a specialized small subset of areas where the student thinks he can work and contribute.

What do you find difficult about reading at university?:
The main thing that I find difficult reading at the university is there is a high knowledge gap in me when I try to learn an advanced deep learning paper. Sometimes I often find myself clueless as to where to start and where to focus. I personally think that all the research work that was created by an amalgamation of concepts in a period of 40-50 years needs time to understand. Until things are not coded up, I don't feel like I have digested the concepts just by reading theoretical derivations until I see my code reflecting the stuff that is proposed in the papers. I feel this is common to other students too since no one is perfect and somewhere or other people have to start from square one. I also feel the amount of pressure the newcomers have to face, probably; these extremely difficult contents might be simplified by them, but in the coming decades, the young researchers have to catch up a lot. It is always good to get prior knowledge in

mathematics and continue to look for new papers which are difficult to read for creating seminal work in the field.

Finding enough time and energy: Finding time and energy is very crucial, and I have not yet fully mastered this. This is because there are a lot of distractions and other pressures that are unavoidable in today's culture. One thing that can be done is to cut off everything (Like the monks of Ramakrishna Mission, whom I have personally seen during my masters) and focus on the topic at hand by visiting reading rooms etc. Still, often people are demotivated when they have done a lot of work studying, and there is not much improvement in the results. I presume at Ph.D., marks should not matter either, until and unless the researcher/student is able to produce high-quality research papers, but there is a criterion that needs to be fulfilled and students find it often difficult to do research and coursework parallelly. In short, the focus should be in understanding and doing good research and not solving enormous mathematical calculations in a short period of time. Things need time, and with constant dedication effort, and the attitude of not giving up, one will be able to overcome every obstacle in life.

Maintaining concentration: Maintenance of concentration might be possible if a student has small goals which can be achieved and are essentially rewarding.

Improving speed: The more paper you read, the more speed you pick up when reading unrelated areas.

Managing vocabulary: It is often difficult to manage vocabulary since there are a lot of terms that come up when reading a new field. One way to mitigate this is to maintain a notebook or digital document, but I think this is a waste of time since things are available at your fingertips whenever you open the internet. Another thing might be do be ultra-nerd and study always, but that might cause health-related problems. There are always tradeoffs and one needs to balance these to attain a meaningful life.

Selecting what to focus on in texts: Going over the paper at once often makes me feel what things might be necessary for me to focus on and give a bit more time to understanding the related literature.

Understanding new, theoretical or detailed information: Understanding new theoretical information might be a pain if we are not using pen and paper to fully understand the things that are written/proposed. It is often necessary to have basic knowledge and revisit the topic when a researcher is familiar with the topic.

Evaluating evidence: Evidence can be evaluated when I am focused on reproducing the study by digging into their codebase and providing some minor modifications to see how things are changing. (This often takes time and commitment which is only possible during long summer breaks.)

4 Short summary of the quality of the paper

The current paper is well written. It builds up context right from the start from where the first seeds of innovation started. It actually points out and refers to the paper by Anderson (1982) which introduced the theoretical concepts of reverse time SDE.

Of course, no paper is perfect, but in its current state, the reader (Jimut) is not able to comment on the limitations of the paper, especially those lines of thought that it failed to address. To me (Jimut), I feel like there is a lot of appendix section which derives the theoretical concepts from scratch, but I have not yet found time to go through them individually. The interesting thing that the reviewers and community found by checking their code are, it really performs better and mitigates the trade-offs which are caused by other generative networks.

Note

The current paper is not one of my favorite papers. This paper is essential in the current scenario since this is one of the fundamental contributions toward score-based generative modeling through stochastic differential equations, which provides a new way to sample data points. The reason I am selecting this paper is we need to have a thorough understanding of the material for working in cutting-edge research, which will be useful in the Advanced Machine Learning course. Also, all the materials presented here might be taken in a humorous way.

Acknowledgements

The author is thankful to Prof. Janani Srree Murallidharan for the stimulating discussions she has shared for reading scientific documents.

References

[1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

[2] Ricky T. Q. Chen, Brandon Amos, and Maximilian Nickel. Learning neural event functions for ordinary differential equations. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021.

[3] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In Yoshua Bengio and Yann LeCun, editors, *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

[4] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial networks. *Commun. ACM*, 63(11):139–144, 2020.

[5] George Papamakarios, Eric T. Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *J. Mach. Learn. Res.*, 22:57:1–57:64, 2021.