Thesis Project Portfolio

Investigation into the Efficiency and Effectiveness of Diffusion Modeling in Predicting Calorimeter Particle Showers

(Technical Report)

Irreproducibility in the Sciences

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Luke Ostyn

Spring, 2024

Department of Computer Science

Contents of Portfolio

Executive Summary

Science is capable of bringing about tremendous advancements in all areas of life. Within physics, and more specifically particle physics, calorimetry is of great importance. It allows us to gauge the energy associated with various particles, helping us to better understand the world around us. In order to take such a measurement, it is necessary that the particle break down into numerous lower energy particles in a so-called particle shower. Traditionally, the modeling of such a process has been infeasible, the computational and monetary cost bordering on prohibitive. More recently, various deep learning methods have been proposed as more efficient techniques by which to fill this void. In this case as well as in many others, science is clearly a force for good, allowing us to explore new frontiers and enrich the human experience. Of course, this is not always the case. Namely, when published results prove to be irreproducible, science as an institution can be quite damaging. When we attach ourselves to ideas that are false, it can cause us to act in ways which are not consonant with the actual truth of the world around us. Recently, there has been an uptick of studies which falter under repetition. At the forefront of the crisis is the field of psychology where the issue has gotten so bad that it is actually referred to as the replication crisis. Naturally, if we desire to course correct and remedy the current situation of science, then it is paramount for us to first understand the mechanisms that have brought it on.

Within my technical paper, I investigated the utility of normalizing flows, a type of deep learning model, in predicting particle showers. The primary manner in which I did this was by tweaking the CaloScore diffusion model of Mikuni and Nachman (2023). Because of hardware constraints, this tweaking consisted of adjusting hyperparameters and then assessing the loss statistics of the model after one or two training epochs. The motivation of these changes were twofold: decreasing both loss and training time. Overall, six different adjustments were tried. As

far as reducing loss, the most efficacious of these modifications were increasing the number of residual layers within the U-Net architecture and increasing the number of convolutional layers within the Resnet. Both reduced loss statistics while only slightly increasing training time. As far as decreasing training time, elevating the learning rate and reducing the number of timesteps during sample creation showed the most promise. Understandably, decreasing the amount of noise added during training and lowering the number of sampling layers were meritless in both capacities. Further work should look to go deeper and explore the results of such altered models after a greater number of epochs.

For my STS paper I evaluated the reproducibility crisis through the lens of Co-production of Science and Social Order, trying to understand the social factors and bad incentives which have led to the production of such a high volume of flawed research. The methodology I employed was that of a literature review, compiling a wealth of viewpoints on the issue and then assembling them into a cohesive assessment of the situation. The overarching theme throughout much of the discourse was the idea that scientists routinely slack in strict adherence to the scientific method. For instance, scientists might selectively choose the trials they wish to report on or p-hack in order to create a significant result. They are, however, not the only entities to blame. Much of this blame is to be shouldered by the journals which choose what to publish. They care about significant and surprising results, so naturally researchers strive to meet those goals. In order to rectify these issues, there are various movements that are pushing for a myriad of different solutions. Among these are data transparency, prepublication review, open-access journals, and greater social control measures.

As far as the technical project, I achieved much of what was possible given my hardware handicap but less than what I had hoped. I failed to make any substantial change to the model

and instead merely adjusted hyperparameters. That said, I felt like the results I got were interesting, and I would like to see if they would hold given more extensive training. For my STS paper, I felt that I did achieve all that I set out to do and that I managed to successfully capture the sociotechnical situation which impelled the replication crisis into being. Moreover, I firmly believe that this is a problem in need of urgent fixing and that the prescriptions I laid out are a critical part of the path there. Moving forward, I think important work would be evaluating the effectiveness of these measures post their implementation.