Deep generative models for fast shower simulation in ATLAS

Aishik Ghosh
on behalf of the ATLAS Collaboration
LAL, Univ. Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay, France
E-mail: aishik.ghosh at cern.ch

Abstract. The need for large scale and high fidelity simulated samples for the ATLAS experiment motivates the development of new simulation techniques. Building on the recent success of deep learning algorithms at interpolation as well as image generation, Variational Auto-Encoders and Generative Adversarial Networks are investigated for modeling the response of the electromagnetic calorimeter for photons in a central calorimeter region over a range of energies. The synthesized showers are compared to showers from a full detector simulation using Geant4. This study demonstrates the potential of using such algorithms for fast calorimeter simulation for the ATLAS experiment in the future.

ATL-SOFT-PROC-2019-007
07/10/2019

1. Introduction

The ATLAS experiment [1] relies on Monte Carlo (MC) simulation as a basis for data analysis. Simulations of the deposition of energy in the calorimeter due to developing showers are slow because they require the modeling of interactions of particles with matter at the microscopic level, as implemented using the Geant4 toolkit [2]. ATLAS already relies on fast calorimeter simulation techniques based on thousands of individual parametrizations of the calorimeter response [3].

In recent years, deep generative algorithms such as Variational Auto-Encoders (VAEs) [4, 5] and Generative Adversarial Networks (GANs) [6] have been demonstrated to accurately model the underlying distributions of data from various domains, including the response of an ATLAS-like calorimeter [7, 8, 9]. Crucially, deep learning based models have demonstrated the ability to interpolate on untrained parameter spaces, allowing for smartly curated training datasets that do not exhaustively encompass all possible input combinations.

This document summarises [10, 11] the first application of a VAE, a GAN and an updated version of the GAN for fast simulation of the calorimeter response of the ATLAS detector for photons over a range of energies in the central region of the electromagnetic calorimeter.

The central region of the electromagnetic (EM) calorimeter is segmented into a matrix in the r/z, η, ϕ space, and its structure is presented in Fig. 1.

2. Monte Carlo Samples and Preprocessing

Samples of single unconverted photons are simulated using Geant4 10.1.patch03.atlas02, the standard MC16 RUN2 ATLAS geometry (`ATLAS-R2-2016-01-00-01`) with the conditions tag `OFLCOND-MC16-SDR-14`. The samples are generated for nine discrete particle energies

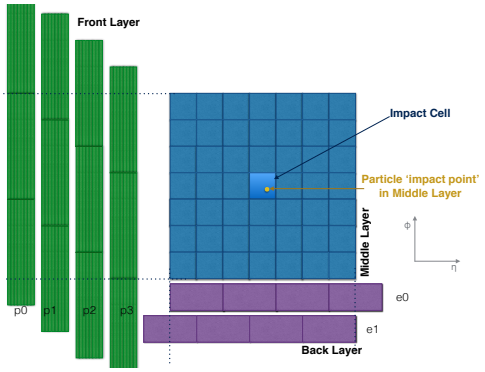


Figure 1: Various calorimeter layer alignments caused by difference in cell widths between the layers. Each layers is cropped with respect to the impact cell, to have the same number of cells regardless of the alignment. p0 through p3 are four possible alignments in ϕ for the front layer, (left, showing a 8×3 portion of the 56×3 cell image), and e0 & e1 are the two possible alignments in η for the back layer, (bottom, showing a 4×1 portion of the 4×7 cell image) with respect to the middle layer (center, showing the full 7×7 image). The calorimeter layers are actually one behind another in the third dimension. [10]

logarithmically spaced in the range between approximately 1 and 260 GeV and uniformly distributed in $0.20 < |\eta| < 0.25$, totaling approximately 90000 events. The truth particles are generated on the calorimeter surface. The generated samples do not include displacements corresponding to the expected beam spread, electronic noise, cross talk between neighbouring cells or dead cells. Cell energies are required to be positive.

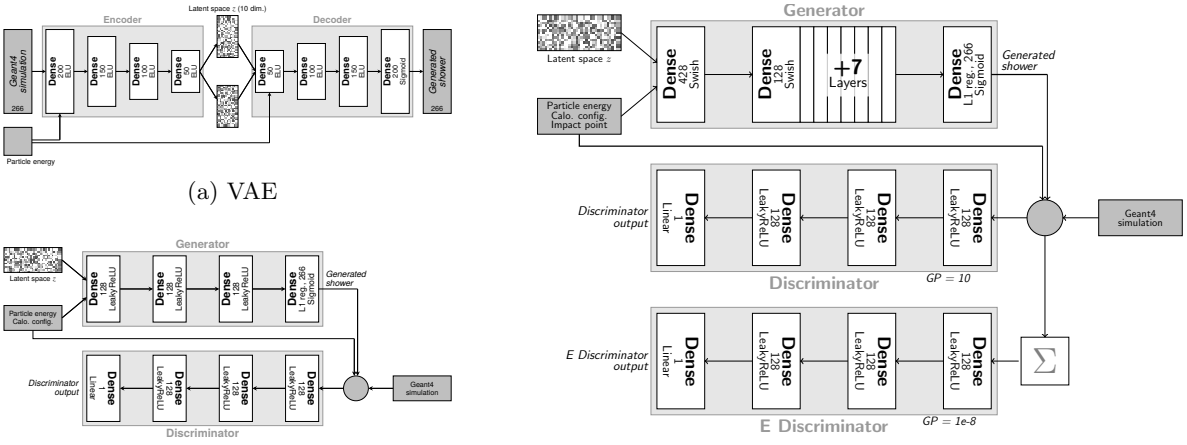
For each layer of the calorimeter, the energy deposits within a rectangular region are selected with respect to the impact cell, defined as the cell in the middle layer closest to the extrapolated position of the photon. The dimensions for the presampler, front, middle, and back layers are 7×3, 56×3, 7×7, and 4×7, respectively, totalling 266 cells. 99% of the total energy deposited is within this selection. The calorimeter cells' η and ϕ referred to in this document are the raw values, i.e. not taking into account corrections accounting for imperfections of the detector.

3. Algorithms

The explored VAE (Fig. 2a) is composed of two stacked neural networks, acting as encoder and decoder, both conditioned on the energy of the incident particle.

The VAE training uses an *RMSProp* optimiser. To ensure the latent space is continuous and allows for smooth interpolations between the encoded instances, a negative Kullback-Leibler divergence between $q_\theta(z|x)$ and the prior probability density function $p(z)$, is included in the loss function in addition to the mean-squared error reconstruction loss. Every additional loss term is associated with a scalar valued hyperparameter that specifies its relative importance to the total loss. The model is implemented in Keras 2.0.8 [12] using TensorFlow 1.3.0 [13] as the backend. The last epoch of the training is used for synthesizing the presented showers.

The first GAN discussed in this document (Fig. 2b) is composed of two neural networks, a generator and a discriminator, trained with a Wasserstein loss function [15] as proposed by [16]. Both networks are conditioned on the energy of the incident particle and the alignments of the different calorimeter layers with the middle layer (Fig. 1). The training was done with an *Adam* optimizer on 4% of the available samples (due to unstable training on bigger training datasets). The discriminator estimates a function that maximally separates the true and synthesized showers which must lie in the space of 1-Lipschitz functions [17, 18], a constraint softly enforced with a double sided penalty on the distance of the norm of the discriminator's gradients from



(a) VAE

(b) First GAN

(c) Updated GAN

Figure 2: Schematic representation of the architectures of (a) VAE, (b) the first GAN, [10] (c) updated GAN (with a trainable swish[14] activation), used in this study.

1 when evaluated on an image that is a randomly weighted average of an image from the real (Geant4) and generated (from the generative network's) distribution. The hyperparameter associated with the additional loss term is known as the *gradient penalty weight* (λ_{GP}). The model is implemented in Keras 2.0.8 [12] using TensorFlow 1.3.0 [13] as the backend.

An updated GAN architecture (Fig. 2c) is later used to improve the energy resolution (discussed in section 4). It was observed that lower *gradient penalty weights* improved the energy resolution at the expense of other physics distributions, therefore two discriminators with different *gradient penalty weights* are used. While the original discriminator has a *gradient penalty weight* of 10, the additional discriminator has a smaller *gradient penalty weight* of 10^{-8}, and it receives the total energy of an image as the input, instead of the image itself. The updated GAN is further conditioned on the extrapolated position of the particle inside the impact cell (Fig. 1). Changing the optimiser from *Adam* to $RMSProp$ allowed for the training to remain stable on 50% of the dataset, which further improved results.

An L1 activity regularizer is applied on the generator output to encourage sparsity of cell energies. When performing an optimization of the hyperparameters, four GANs are trained with different random seeds and their average performance is compared to avoid picking up random fluctuations. As a control, average performance of four GANs is compared with two other sets of four GANs, all trained with the same hyperparemters but different random seeds.

The training takes 80 h for 15000 epochs on 50% of the dataset for the updated GAN, performed on an NVIDIA® Kepler™ GK210 GPU with a processing power of 2496 cores, each clocked at 562 MHz. The card has a video RAM size of 12 GB with a clock speed of 5 GHz.

4. Results

This section presents comparisons between physics properties of the synthesised showers from the generative models and the full simulation.

The energy deposited in the middle layer of the calorimeter, is shown in Fig. 3a for photons with an energy of approximately 65 GeV in the range $0.20 < |\eta| < 0.25$. Both VAE and GAN accurately describe the bulk of the energy deposits but with reduced agreement in the tails of the distribution. They reproduce the energy weighted average η of the middle calorimeter layer, shown in Fig. 3b, to a large extent. The shower depth, $d = \frac{1}{E}\sum_{i\in\text{layers}} E_i d_i$, calculated from the energy weighted mean of the longitudinal center of all calorimeter layers, is shown in Fig. 3c.

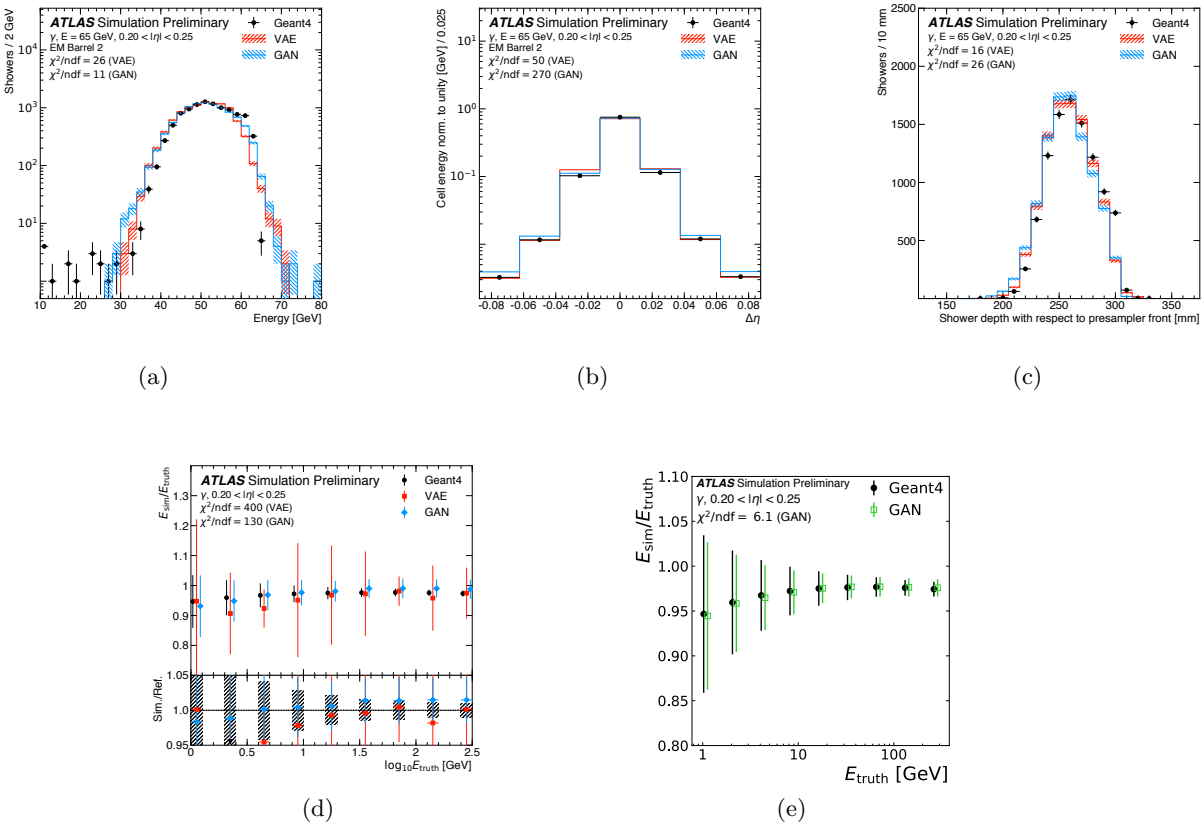


Figure 3: (a) The energy deposited, (b) the average η distribution in the middle calorimeter layer, and (c) the shower depth for photons with 65 GeV energy in the range $0.20 < |\eta| < 0.25$. The full simulation (black markers) is compared to a VAE (solid red line) and a GAN (solid blue line). The underflow and overflow is included in the first and last bin of each distribution, respectively. (d) Energy response of the calorimeter as function of the true photon energy, where the shown error bars indicate the resolution of the simulated energy deposits, and (e) the same for the updated GAN (green) [10, 11]

Both VAE and GAN reproduce the shape of shower depth simulated by Geant4, but with a slight shift.

Fig. 3d shows the simulated energy as a function of true photon energy. The modeling of the total energy response reflects the modeling of the underlying distributions, i.e. the energy deposited in the calorimeter layers, and enhances the mismodeling of the tails due to underestimating the underlying correlations observed in these. Both generative models simulate a wider spread of energies than Geant4. The updated GAN architecture was conceived to improve the modeling of this distribution, and Fig. 3e shows considerable improvement in the updated GAN.

5. Conclusion

This document presents the first application of generative models for simulating particle showers in the ATLAS calorimeter. Two algorithms, a VAE and a GAN, have been used to learn the response of the EM calorimeter for photons with energies between approximately 1 and 260 GeV in the range $0.20 < |\eta| < 0.25$. The properties of synthesized showers show promising agreement

with showers from a full detector simulation using Geant4. The only distribution these algorithms completely failed to model, the energy response of the calorimeter as a function of the true photon energy, has been addressed with the updated GAN. The updated architecture could be extended to improve other distributions of particular concern in a similar fashion. This study demonstrates the feasibility of using such algorithms for fast calorimeter simulation for the ATLAS experiment in the future and opens the possibility to complement current techniques. More useful assessments of the performance of these algorithms will come from a direct comparison with current fast simulation techniques used within ATLAS, by plugging the models into the ATLAS software. These comparisons will reveal where further improvements are required. An advantage of such models based on neural networks is their ability to interpolate on untrained parameter spaces, which allows to train them without a dataset exhaustively encompassing all possible input combinations. It is therefore imperative to verify that these models are able to also simulate correctly photon showers at untrained energy points within the ATLAS software. Incorporating other regions of the calorimeter as well as other types of particles are the next logical steps for this effort. In addition, the algorithms must either be conditioned on the direction of the incident particle, or be trained on more granular datasets which can benefit from parameterisations already present in fast calorimter simulation techniques [3] used in ATLAS.

References

[1] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008. doi:10.1088/1748-0221/3/08/S08003.

[2] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl.Instrum.Meth.*, A506:250–303, 2003. doi:10.1016/S0168-9002(03)01368-8.

[3] ATLAS Collaboration. The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim. ATL-PHYS-PUB-2010-013, 2010. URL: https://cds.cern.ch/record/1300517.

[4] D. P Kingma and M. Welling. Auto-Encoding Variational Bayes. December 2013. arXiv:1312.6114.

[5] D. Jimenez Rezende et al. Stochastic Backpropagation and Approximate Inference in Deep Generative Models. January 2014. arXiv:1401.4082.

[6] I. J. Goodfellow et al. Generative Adversarial Networks. June 2014. arXiv:1406.2661.

[7] Luke de Oliveira et al. Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis. *Comput. Softw. Big Sci.*, 1(1):4, 2017. arXiv:1701.05927, doi:10.1007/s41781-017-0004-6.

[8] Paganini et al. Accelerating Science with Generative Adversarial Networks: An Application to 3D Particle Showers in Multilayer Calorimeters. *Phys. Rev. Lett.*, 120(4):042003, 2018. arXiv:1705.02355, doi:10.1103/PhysRevLett.120.042003.

[9] Paganini et al. CaloGAN : Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks. *Phys. Rev.*, D97(1):014021, 2018. arXiv:1712.10321, doi:10.1103/PhysRevD.97.014021.

[10] ATLAS Collaboration. Deep generative models for fast shower simulation in ATLAS. ATL-SOFT-PUB-2018-001, 2018. URL: https://cds.cern.ch/record/2630433.

[11] ATLAS Collaboration. Energy resolution with a GAN for Fast Shower Simulation in ATLAS. ATLAS-SIM-2019-004, 2019. URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/SIM-2019-004/.

[12] François Chollet et al. Keras, 2015. URL: https://keras.io.

[13] Martín Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL: https://www.tensorflow.org/.

[14] Prajit Ramachandran et al. Searching for activation functions. *CoRR*, abs/1710.05941, 2017. URL: http://arxiv.org/abs/1710.05941.

[15] M. Arjovsky et al. Wasserstein GAN. January 2017. arXiv:1701.07875.

[16] Ishaan Gulrajani et al. Improved training of wasserstein gans. 2017. arXiv:1704.00028.

[17] Frank L. Hitchcock. The distribution of a product from several sources to numerous localities. *Journal of Mathematics and Physics*, 20(1-4):224–230, 1941. doi:10.1002/sapm1941201224.

[18] L.N. Vaserstein. Markovian processes on countable space product describing large systems of automata. *Probl. Peredachi Inf.*, 5(3):64–72, 1969.