

Validation of Deep Convolutional Generative Adversarial Networks for High Energy Physics Calorimeter Simulations

Florian Rehm^{1,2}, Sofia Vallecorsa¹, Kerstin Borras^{2,3}, Dirk Krücker³

¹ CERN, Switzerland

² RWTH Aachen University, Germany

³ DESY, Germany

florian.matthias.rehm@cern.ch

Abstract

In particle physics the simulation of particle transport through detectors requires an enormous amount of computational resources, utilizing more than 50% of the resources of the CERN Worldwide Large Hadron Collider Grid. This challenge has motivated the investigation of different, faster approaches for replacing the standard Monte Carlo simulations. Deep Learning Generative Adversarial Networks are among the most promising alternatives. Previous studies showed that they achieve the necessary level of accuracy while decreasing the simulation time by orders of magnitudes. In this paper we present a newly developed neural network architecture which reproduces a three-dimensional problem employing 2D convolutional layers and we compare its performance with an earlier architecture consisting of 3D convolutional layers. The performance evaluation relies on direct comparison to Monte Carlo simulations, in terms of different physics quantities usually employed to quantify the detector response. We prove that our new neural network architecture reaches a higher level of accuracy with respect to the 3D convolutional GAN while reducing the necessary computational resources. Calorimeters are among the most expensive detectors in terms of simulation time. Therefore we focus our study on an electromagnetic calorimeter prototype with a regular highly granular geometry, as an example of future calorimeters.

Introduction

Today, particle detectors are simulated using Monte Carlo-based methods such as the Geant4 toolkit (Agostinelli et al. 2003). Due to complex detector geometries and manifold diverse physical processes, Monte Carlo simulations have a high computational cost and require, currently, more than half of the Worldwide Large Hadron Collider (LHC) Grid resources (Albrecht et al. 2019). In the future High Luminosity phase of the LHC (HL-LHC) around 100 times more simulated data are expected (Apollinari et al. 2017). Even when taking the technological improvements into account, this will exceed considerably the expected computational resources at the HL-LHC regime.

In order to overcome this obstacle, intense research is ongoing for developing faster alternatives to the standard

Monte Carlo simulation. The main difficulty in searching for a faster simulation approach is given by the high level of accuracy required. Based on deep generative models, several prototypes demonstrated an enormous potential by achieving similar accuracy than the traditional simulation (de Oliveira et al. 2017; Paganini et al. 2018; Ghosh 2019; DiSipio et al. 2019; Chekalina et al. 2019). Generative Adversarial Networks (GANs) represent an example of such models: so far, they have been employed primarily for simulating calorimeters (Carminati et al. 2018). In this paper we describe a novel GAN architecture which uses two-dimensional convolutional layers (Conv2D) for generating the three-dimensional volume of an electromagnetic calorimeter. We compare our architecture to a previous three-dimensional convolutional (Conv3D) GAN model developed for the same detector use case (Carminati et al. 2018). Results are compared to the corresponding Monte Carlo simulation in terms of multiple physics validation metrics. The Conv2D model improves the physics accuracy while reducing the computation time.

In the following we provide a review of related work, followed by a brief description of the electromagnetic calorimeter and the training data set. Next, we introduce our Conv2D prototype and we evaluate its performance in terms of physics accuracy and computation time. The last section summarizes our conclusions and plan for future work.

Related Work

GANs belong to the group of generative models and are nowadays used for a large variety of different tasks (Alqahatani 2019). The whole GAN model consists of two deep neural networks: a generator network, which generates images from a latent vector consisting of random numbers, and a discriminator, trained to evaluate and distinguish between the generated images of the generator and the training or validation images. GANs are among the most promising approaches, which are being successfully investigated for replacing traditional Monte Carlo for simulating detectors and, in particular, calorimeters. Several studies demonstrate that GANs achieve similar levels of accuracy as Monte Carlo, while considerably decreasing the simulation time (Carminati et al. 2018) (Khattak et al. 2019) (Erdmann et al. 2019).

In this research we compare our results with the Conv3D GAN model (Carminati et al. 2018) developed for the same

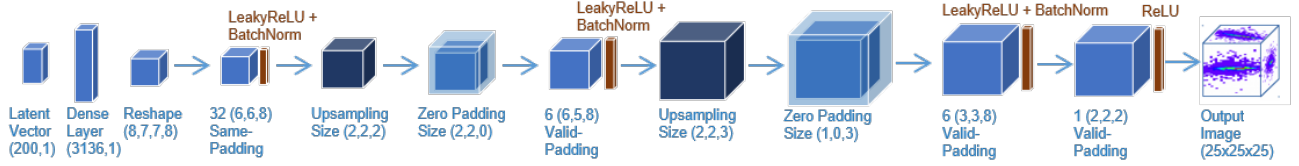


Figure 1: Convolutional 3D generator neural network architecture from (Carminati et al. 2018). Input is a latent vector and output is the 3-dimensional shower image.

calorimeter prototype. The model is available on GitHub (Vallecorsa et al. 2020) and its generator architecture is shown in figure 1 for reference. This model achieved in (Carminati et al. 2018) a large speed up (4 ms simulation time for a single electron compared to 17 s simulation time using Geant4 on a Intel Xeon processor) while reaching a good agreement with Monte Carlo in terms of physics (Carminati et al. 2018).

(Quast 2020) used a different Conv2D network architecture for simulating three dimensional calorimeter shower images. This was achieved by generating 12 two dimensional 15x7 images and stacking them to 3D images together. In our studies we have tested a similar architecture, based on stacking a set of two dimensional images, each representing one of the calorimeter layers. However this approach resulted in a worse performance, in terms of physics and speed up, with respect to the model we suggest in this research.

Electromagnetic Calorimeters

By measuring particles energy, calorimeters are one of the core components of High Energy Physics experiments (Brown and Cockerill 2012). The primary particles enter and interact with the material of the calorimeter and deposit their energies by creating secondary particles as they pass through the detector. The secondary particles, in turn, create other particles by the same mechanisms leading to the generation of showers. While the shower evolves, the energy of the particles gradually reduces as it is absorbed or measured by the calorimeter sensors. In our work, we simulate electromagnetic calorimeters, which are specially designed to measure energies of electrons, positrons and photons that interact in the detector volume via electromagnetic interactions (mainly bremsstrahlung for electrons and pair production for photons). As stated above, calorimeters represent computationally the most demanding simulations and are, therefore, the starting point for accelerating the simulation process. By representing each of the calorimeter sensors as a pixel in an image and the corresponding energy measurement as the pixel intensity, it is possible to interpret the detector output as an image and apply to it computer vision strategies and methods. It should be noted, however, that such "shower images" exhibit important differences with respect of typical RGB pictures, notably their high level of sparsity and large pixel intensity dynamic range, which can span several orders of magnitude.

Our training and test data sets contain particle shower im-

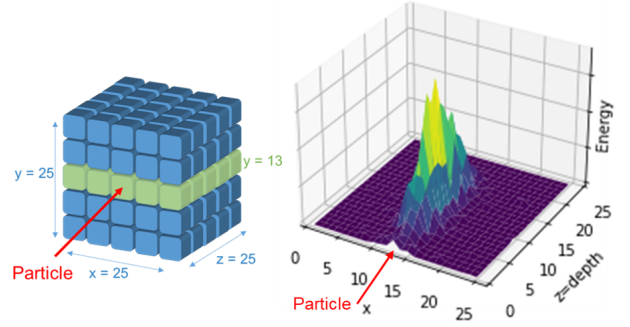


Figure 2: Representation of the 3D space (left) and an example shower development at $y = 13$ (right).

ages recorded by the calorimeter, simulated using Geant4 for a future electromagnetic calorimeter prototype. For our study, we use 200 000 electron shower images with initial energies in the range of 2-500 GeV and a 3-dimensional shape of 25x25x25 pixels. Details on the calorimeter architecture and the data set can be found in (Carminati et al. 2018). Figure 2 shows an example shower image cutout in the central layer ($y = 13$), where the particle enters the calorimeter perpendicular to the x - y -plane at position $x = 13, y = 13$ and $z = 0$, the shower is evolving in the z -direction. Particle shower images are parameterised according to the energy of the primary particle E_p , which is the energy of the electron entering the calorimeter volume and the total energy measured by the calorimeter, which we indicate as E_{cal} (for Electromagnetic CALorimeter). The relation between E_{cal} and E_p depends on the type, geometry and material of the calorimeter.

3D Generative Adversarial Networks

Replacing Conv3D layers, in neural networks, with Conv2D layers reduces in general the computational complexity and therefore, computing resources and training time per epoch. However a non-trivial strategy is necessary to combine Conv2D layer in order to reproduce correctly a three-dimensional volume. Our Conv2D generator architecture, shown in figure 3, uses a mix of 2D convolutional layers and transposed 2D convolutional (Conv2D-transpose) layers (to increase the image size to the desired dimension). Additionally, we use batch normalization (BatchNorm) (Ioffe et al. 2015), a mix of rectified linear units (ReLU) and leaky rectified linear units activation functions (LeakyReLU) (Maas

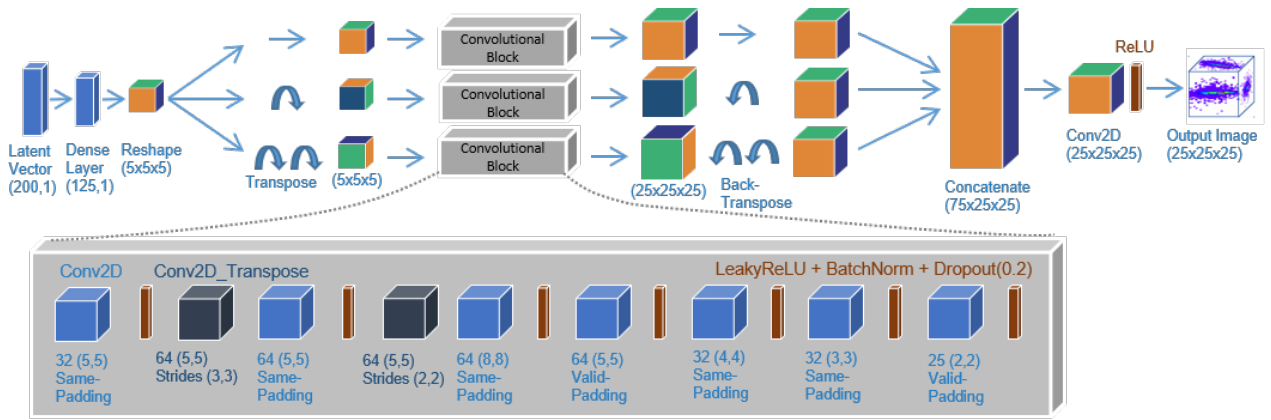


Figure 3: Conv2D Generator neural network architecture. Input is a latent vector and output is the 3-dimensional shower image.

2013), in order to account for the image sparsity, and dropout layers (Dropout) (Nitish et al. 2014) to prevent mode collapse. The convolutional layers are organised in three branches in order to enable the model to correctly learn pixel-level correlations along the three canonical image dimensions. Three-dimensional images, with $25 \times 25 \times 25$ shapes are generated and parameterised (conditioned) using the input particle energy E_p . The hyper parameters, learning rate and learning rate decay, which we use for our model are found through a hyperparameter search with the Optuna tool (Akiba et al. 2019).

Comparing the total number of parameters for the Conv3D and the Conv2D generator (table 1), it can be noted that the Conv2D model has more than twice the number of parameters as the Conv3D model. Moreover, table 1 shows that the Conv3D network has most of its parameters (630k) in the first dense layer. This choice could represent a limitation, of the Conv3D model, in terms of representational power. On the other hand, the Conv2D model contains only a small fraction of its parameters in the dense layer. The advantage of having more parameters in the convolutional layers (meaning more convolutional layers, more channels and/or bigger filter sizes) is that convolutional layers perform better in learning images correlations. For this reason, we expect the Conv2D model to have a higher learning capability and to be able to solve more sophisticated tasks.

The architecture differences have a direct impact on the training and inference time. We bench-marked the inference process by running 20 warm-up steps and then 100 inference steps including 20 batches in each step. As shown in table 1, the Conv2D inference time is much shorter, despite the larger number of parameters. The Conv2D model is also faster in terms of training time: with only 105s per epoch, instead of 291s for the Conv3D. All GAN runs are performed on a NVIDIA Tesla V100-SXM2 GPU using TensorFlow version 2.0, CUDA version 10.2 and a batch size of 128. Unfortunately, Geant4 does not fully support GPU architectures at this time, its simulation time is measured on a Intel Xeon processor in (Carminati et al. 2018).

Table 1: Total number of parameters of the Conv2D and Conv3D generator networks and the number of parameters within the first dense layer. Inference times on GPU and the respective speed up compared to Monte Carlo.

Model	Total	Dense	Inference	Speed Up
Conv2D	2.05M	0.12M	5.4 s	8 000x
Conv3D	0.87M	0.63M	7.0 s	6 200x

Evaluation

In this section we evaluate the performance of our new Conv2D model against the Conv3D model in (Carminati et al. 2018) and the Monte Carlo approach (Geant4).

Performance evaluation of generative models remains a difficult task for which, depending on the specific applications, several methods have been proposed (Borji 2018). For evaluating the physics we measure the energy patterns across the calorimeter volume focusing on specific quantities that are typically used to characterize calorimeter properties.

Firstly, we build the particle shower shapes along the y - and z -axes for 20 000 samples within an energy range of 2-500 GeV. The shower development along the x -axis is similar to the y -axis and is therefore not shown. These shower shapes are shown in figure 4 (left) on a linear energy axis scale and (right) on a logarithmic scale. The GAN distributions agree with the expected Monte Carlo values within the uncertainties. In particular, previous studies had indicated poor performance at the edges of the simulated volume (the first and last cells along the x - and y -axes) (Khattak et al. 2019). In figure 4 we can see that both GAN models perform worse when the energy depositions became very small at the edges of the sensitive volume. However, the overall Kolmogorov-Smirnov (two-samples) test (KS- 2008), yields probability above 80% for most distributions and primary particle energies.

In order to simplify the evaluation during training and hyper-parameter tuning, we build a Mean Squared Error (MSE) based on the above 2D shower shapes comparing the

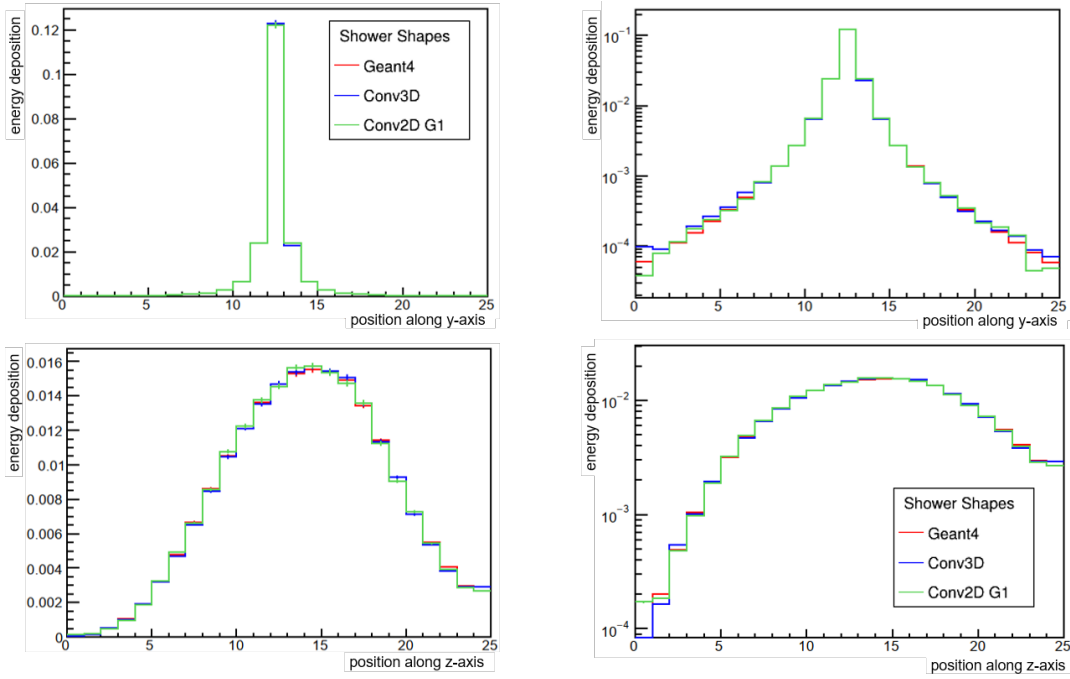


Figure 4: Particle shower distribution along the y - and z -axes on a linear energy axis scale (left) and on a logarithmic scale (right).

GAN distribution with Monte Carlo. We measured for the Conv2D model a MSE of 0.044 and for the Conv3D model 0.065, indicating a better accuracy (the lower the MSE the better), in agreement with the results from the shower shape plots from figure 4.

The sampling fraction is shown in figure 5 as the ratio between the total measured energy E_{cal} and the initial particle energy E_p . It represents the energy response of the generator network over the full energy range. We can see that the Conv2D model maintains a exact agreement to Geant4 over the whole energy range, whereas the Conv3D model performs worse for E_p energies below 100 GeV.

For calorimeter simulations it is important that the single cell energy response is correctly reproduced by the GAN models. Figure 6 shows the single cell energy deposition for 20 000 showers with input energies in the range of 2-500 GeV. The GAN description is not exact, in particular for small cell energy depositions, below 1 MeV. Additional research is required in order to improve the agreement. However, it should be noted, that we applied a minimal energy threshold of 10^{-6} GeV to take the detector energy resolution into account.

Figures 7 and 8 compare the mean and standard deviations of single cell energy distributions in GAN and Monte Carlo samples for 1 580 showers with input energy $E_p = 250 \pm 2$ GeV. For simplicity, we focus on the cells that constitute the core of the energy showers, where the larger energy depositions occur, and we analyse the cells with coordinates $x, y, z \in [10, 16]$, for a total of $7 \cdot 7 \cdot 7 = 343$ cells. The expected ratio = 1 is drawn as a red line in both figures, for comparison.

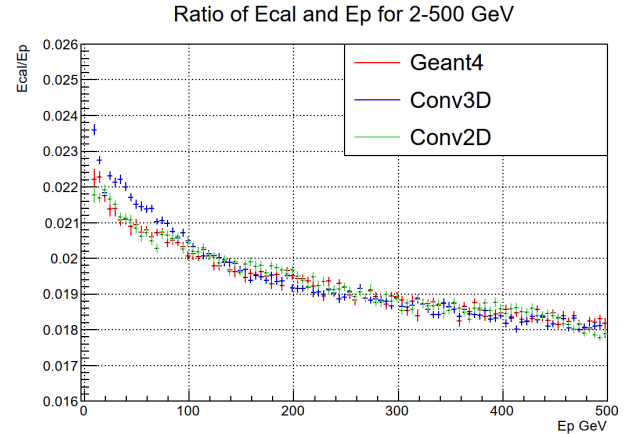


Figure 5: Sampling fractions comparing the Conv3D and Conv2D response to Geant4 over the full energy range.

Figure 7 shows that the mean energy deposition per cell is extremely close to Geant4 for the Conv2D model, their ratio averaging at 1.00. On the other hand, the Conv3D model slightly underestimates the mean cell energy deposition with respect to Geant4, the average ratio close to 0.93. Similarly, figure 8 shows the GAN/Geant4 ratios in terms of the single cell energy distribution standard deviations (STD), which we use as a proxy for the single cell energy resolution. While both models underestimate this quantity, the Conv2D architecture exhibits values that are closer to the Monte Carlo prediction, with average ratios around 0.90 against the 0.86 ob-

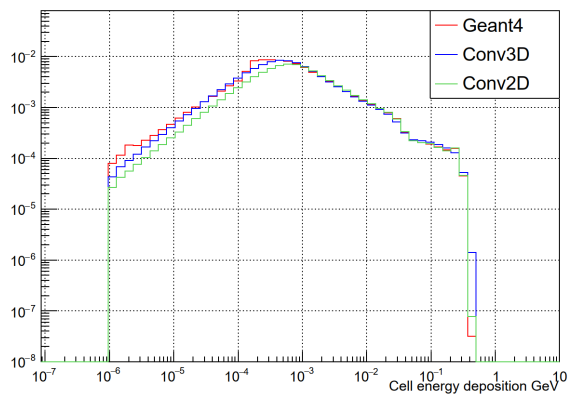


Figure 6: Histogram of cell energy deposition.

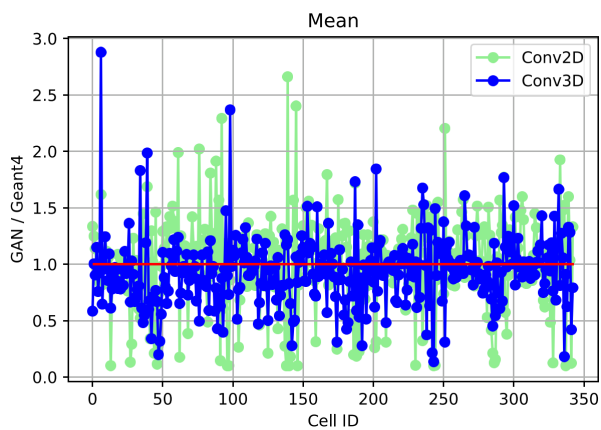


Figure 7: Single cell (pixel) energy distribution mean. The Conv2D model has an overall mean of 1.00 and the Conv3D model an overall mean of 0.93.

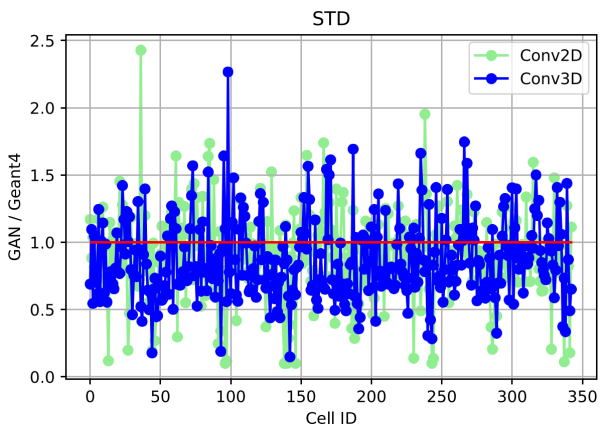


Figure 8: Single cell (pixel) energy standard deviations. The Conv2D model has an overall standard deviation of 0.90 and the Conv3D model an overall standard deviation of 0.86.

tained for Conv3D. In addition it is worthwhile for a further study to look also in to the tails of the shower to demonstrate

the level of agreement.

The original GAN paper (Goodfellow et al. 2014) suggests that GANs can learn any target distribution if sufficiently large networks, training samples, and computation time are given. However, the theoretical analysis conducted in (Arora et al. 2017) showed that the training objective can approach its optimum value even though the generated distribution is far from the target distribution. Moreover, GAN models suffer from well-known problems, such as mode dropping or mode collapse (Yazici et al. 2020), which affect the quality of the generated sample. Understanding whether GANs can reproduce the same level of similarity exhibited by the original Monte Carlo data set, is essential in the context of scientific simulations. In order to better understand the level of similarity/diversity among the images, we measure the Structural Similarity Index (SSIM) (Kumar et al. 2013). The SSIM estimates the perceptual difference between similar images, a value of 1 indicating two identical images. The SSIM is adjusted by the parameter L , which accounts for the pixel intensity dynamic range, and stabilizes for $L = 10^{-4}$. We calculate the SSIM building 2D windows inside the image and sliding them across the calorimeter $x - y$ plane, while the z axis is treated as the image channel dimension. We measure the index for random image pairs from the Monte Carlo data (labelled as "MC vs MC"), the GAN data ("GAN vs GAN") or both sets ("MC vs GAN"). Figure 9 shows the results, calculated for $L = 10^{-4}$. Overall, we observe that the SSIM of generated images vs. generated images of both GAN models is clearly higher than the corresponding Monte Carlo value, indicating that, on average, GAN images are more similar to each other than Monte Carlo images are. Or in other words, that the GANs fail to reproduce the image diversity present in the Monte Carlo sample. The Conv2D model does however perform better with lower SSIM (green, in figure 9). It is equally interesting to note that, by measuring the similarity between GAN and Monte Carlo images (light blue and gray) we can conclude that both GAN models produce images that resemble Monte Carlo's with the equivalent level of similarity as the Monte Carlo data set itself.

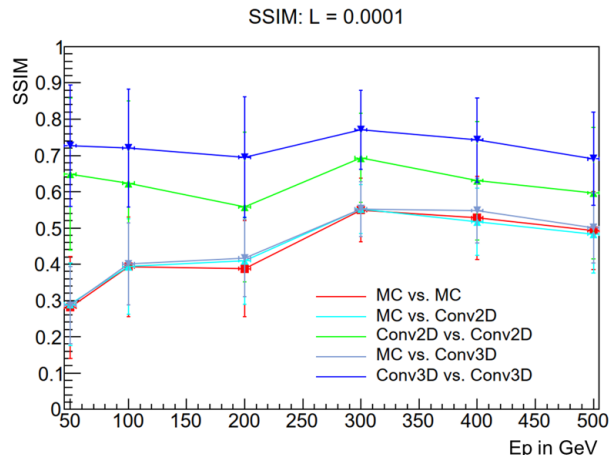


Figure 9: SSIM between the different image sets.

While we plan to continue our investigation in order to better quantify GAN performance in terms of support space size, sample diversity and mode dropping effects, the current result is well within the precision level required from fast simulation models. The largest discrepancy is observed for single cell energies, shown in figure 6, which we can adjust by tuning the initial image pre-processing step.

Conclusion and Future Work

In this paper we introduce a novel Conv2D neural network architecture for simulating high energy physics calorimeter detector outputs. We compare the Conv2D to a prior Conv3D model and validate the results against Monte Carlo simulations. The result evaluation is performed in terms of several physics quantities, such as energy resolution, shower shapes or sampling fraction and image analysis metrics such as the SSIM. All things considered, the Conv2D model achieves an extremely satisfactory level of agreement with respect to Monte Carlo simulations. At the same time, the SSIM index hints to a slightly more limited diversity among the generated images. A full characterisation of the consequences of this difference, with respect to typical simulated data set use cases is needed. Together with a detailed study to understand, whether image similarity can be correlated to any specific bias in the generated data. On the computational side, despite a larger number of parameters in the Conv2D model, we obtain a significant speed up in terms both the inference and training time.

ACKNOWLEDGEMENTS

This work has been sponsored by the Wolfgang Gentner Programme of the German Federal Ministry of Education and Research.

References

2008. *Kolmogorov–Smirnov Test*, 283–287. New York, NY: Springer New York. ISBN 978-0-387-32833-1.
- Agostinelli; et al. 2003. GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A* 506: 250–303.
- Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; and Koyama, M. 2019. Optuna: A Next-generation Hyperparameter Optimization Framework.
- Albrecht; et al. 2019. A Roadmap for HEP Software and Computing R&D for the 2020s. *Computing and Software for Big Science* 3. doi:10.1007/s41781-019-0031-6.
- Alqahtani, e. a. 2019. Applications of Generative Adversarial Networks (GANs): An Updated Review. *Archives of Computational Methods in Engineering* .
- Apollinari; et al. 2017. High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1 4/2017.
- Arora; et al. 2017. Generalization and Equilibrium in Generative Adversarial Nets (GANs).
- Borji, A. 2018. Pros and Cons of GAN Evaluation Measures.
- Brown, R.; and Cockerill, D. 2012. Electromagnetic calorimetry. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 666: 47 – 79.
- Carminati; et al. 2018. Three dimensional Generative Adversarial Networks for fast simulation. *Journal of Physics: Conference Series* 1085: 032016.
- Chekalina; et al. 2019. Generative Models for Fast Calorimeter Simulation: the LHCb case. *EPJ Web of Conferences* 214: 02034.
- de Oliveira; et al. 2017. Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis. *Computing and Software for Big Science* 1(1).
- DiSipio; et al. 2019. DijetGAN: a Generative-Adversarial Network approach for the simulation of QCD dijet events at the LHC. *Journal of High Energy Physics* 2019(8).
- Erdmann; et al. 2019. Precise Simulation of Electromagnetic Calorimeter Showers Using a Wasserstein Generative Adversarial Network. *Comput Softw Big Sci* 3 .
- Ghosh, A. 2019. Deep generative models for fast shower simulation in ATLAS. Technical report, CERN, Geneva.
- Goodfellow; et al. 2014. Generative Adversarial Networks.
- Ioffe; et al. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift .
- Khattak; et al. 2019. Particle Detector Simulation using Generative Adversarial Networks with Domain Related Constraints. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*.
- Kumar; et al. 2013. Development of improved SSIM quality index for compressed medical images. 251–255. doi:10.1109/ICIP.2013.6707593.
- Maas, A. L. 2013. Rectifier Nonlinearities Improve Neural Network Acoustic Models.
- Nitish; et al. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* 15(56): 1929–1958.
- Paganini; et al. 2018. CaloGAN: Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks. *Physical Review D* 97(1).
- Quast, T. 2020. Qualification, performance validation and fast generative modelling of beam test calorimeter prototypes for the CMS calorimeter endcap upgrade 236.
- Vallecorsa; et al. 2020. 3Dgan. <https://github.com/svalleco/3Dgan/tree/master/keras>.
- Yazici, Y.; Foo, C.-S.; Winkler, S.; Yap, K.-H.; and Chandrasekhar, V. 2020. Empirical Analysis of Overfitting and Mode Drop in GAN Training.