Evaluating Generative Adversarial Networks for particle hit generation in a cylindrical drift chamber using Fréchet Inception Distance

Irene Andreou[1], **Noam Mouelle**[2]

Imperial College London, Blackett Laboratory, UK

E-mail: [1]`irene.andreou18@imperial.ac.uk` [2]`ndm33@cam.ac.uk`

Abstract. We use Fréchet Inception Distance (FID) measured in the latent spaces of pre-trained, fine-tuned and custom-made inception networks to evaluate Generative Adversarial Networks (GANs) developed by the COherent Muon to Electron Transition (COMET) collaboration to generate sequences of background hits in a Cylindrical Drift Chamber (CDC). We validate the convergence of the GANs' training and show that the use of self-attention layers reduces FID. Our method enables the use of FID as an evaluation metric even when an application-specific inception network is not readily available, making it transferable to other GAN applications in High Energy Physics.

1. Introduction

The COherent Muon to Electron Transition (COMET) experiment [1] is an experiment based at J-PARC in Japan, which will probe for neutrinoless muon to electron ($\mu - e$) conversion, which constitutes an observation of Charged Lepton Flavour Violation (CLFV). Neutrino oscillations render $\mu - e$ conversions possible, albeit with a very small cross-section. Observing $\mu - e$ conversions would be a clear indicator of new physics. The COMET experiment is a two-phase experiment with Phase-I aiming to achieve a single event sensitivity (SES) of 7×10^{-15}, and Phase-II aiming for a SES of 2.6×10^{-17} [2], improving on the current experimental limit [3] by a factor of 2.7×10^4. In Phase-I, a proton beam will be directed onto a graphite target, producing pions which decay into muons, which are transported to an Aluminium target forming muonic Al^{13} atoms. The aluminium target is surrounded by a Cylindrical Detector system [1], consisting of trigger hodoscope arrays and a Cylindrical Drift Chamber (CDC). A particle entering one of the CDC cells triggers a hit. Each CDC hit has four features: the energy deposit (EDEP), the hit time, the distance of closest approach to the sense wire (DOCA) and the ID of the cell triggered (wire ID). Hits are used for track reconstruction and to extract kinematics. The signature of a $\mu - e$ conversion is a 105 MeV electron [1].

2. Motivation

COMET Phase-I will produce 1.5×10^{16} captured muons [1]. Monte Carlo (MC) methods, which were used by the collaboration to develop a simulation of the experiment, are too intensive to simulate the number of background hits present in a full-size dataset. Thus, the COMET collaboration developed Generative Adversarial Networks [4, 5] (GANs) to approximate the

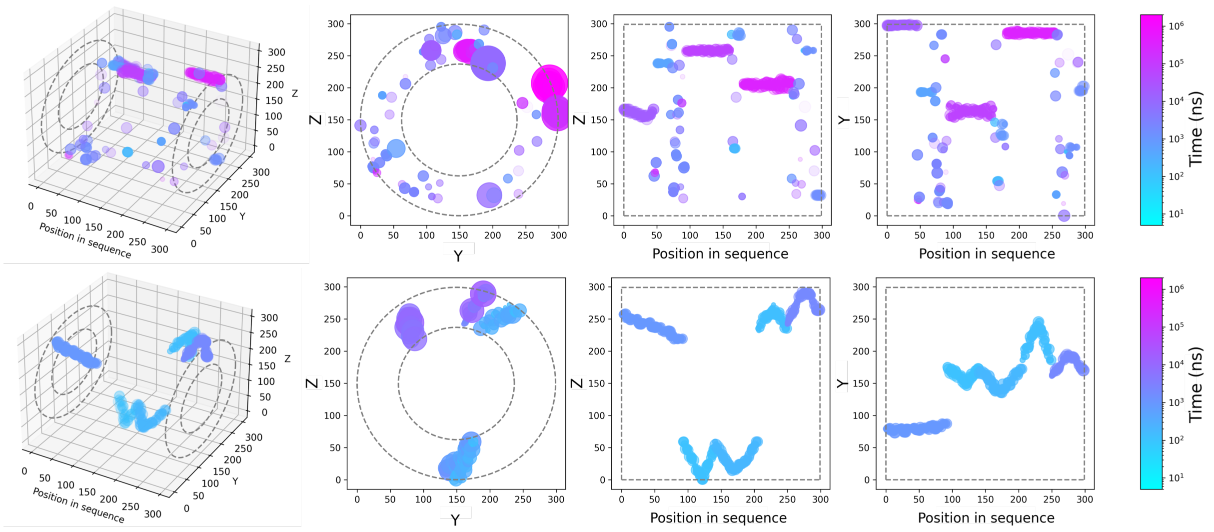

Figure 1. Plots of the 3d representations, and of their three 2d projections, of a sequence of (top) noise hits and (bottom) reconstructible hits in the COMET cylindrical drift chamber. Colour indicates time, size indicates the distance of closest approach, and transparency indicates the energy deposit.

mapping between a multivariate normal distribution, which can be efficiently sampled, and the background hit sequences distribution. Evaluating the GAN is crucial to ensure that it can be used to augment the simulated data and enable accurate sensitivity estimates.

In previous works, deep generative models were evaluated by measuring the similarity of real and generated distributions of explicitly or implicitly learned features [6, 7, 8], using the accuracy of a classifier as a proxy [6, 9] or by measuring Fréchet Inception Distance [10] (FID) in a latent space [7, 11]. The latter method has the advantage of directly comparing real and fake samples but necessitates a fully trained inception network, which requires many resources. In this work, we measure FID using pre-trained, fine-tuned and custom-made inception networks to evaluate GANs developed by the COMET collaboration. Our methods are transferable and, in the case of pre-trained models, could alleviate the need to develop and train an inception network from scratch in HEP applications.

3. Method

We aim to use deep neural networks as maps from the hit sequences space to a latent space in which a meaningful metric can be constructed.

Firstly, we use Inception v3 (Iv3) [12] pre-trained on the ImageNet [13]. Provided that our data is re-shaped to exhibit geometrical patterns, Iv3 should be able to map similar hit sequences, with respect to those patterns, to similar regions of the latent space. The data is composed of particle hits ordered in time. We segment the data into shorter sequences of L consecutive hits and use the wire ID feature to place each hit on a 2-dimensional $N \times N$ grid before stacking them together to obtain 3-dimensional images (see Fig. 1, top). The remaining features (DOCA, EDEP, time) are channels of these images. The input of Iv3 must have a shape of $3 \times 299 \times 299$ [12], so we set $L = N = 299$ and reduce the images to two dimensions by taking their projections along each of the three axes.

While being an out-of-the-box solution, Iv3 was not explicitly trained to form a good representation of the CDC data. Hence, we also used a fine-tuned version of Iv3 (FTIv3) and a custom 3-dimensional Convolutional Neural Network (3D CNN), both trained on CDC hits.

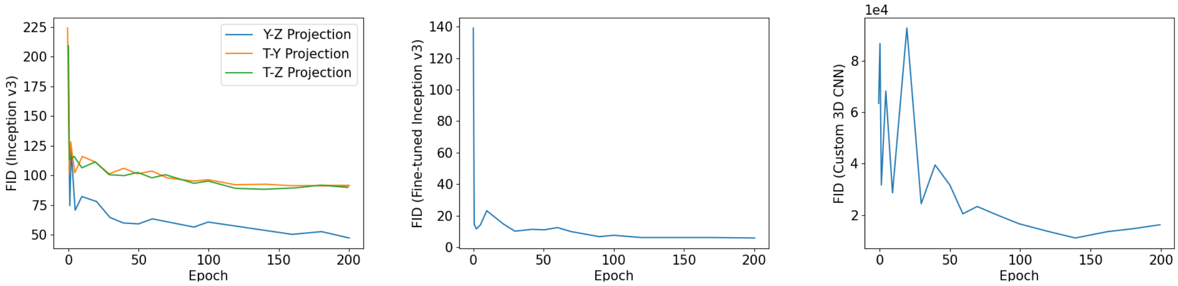

Figure 2. Evolution, during the self-attention GAN training, of FID measured in the latent spaces of (left) Inception v3 (centre) Fine-tuned Inception v3 and (right) a custom 3d CNN.

FTIv3 and the custom 3D CNN were trained to classify sequences of noise hits from sequences of reconstructible hits. Reconstructible hits, in the MC simulation, are defined as hits caused by particles with $p_T > 50\text{MeV/c}$, which are more likely to leave tracks in the CDC (see Fig. 1, bottom) than noise hits ($p_T < 50\text{MeV/c}$). FTIv3 was made by simply adding three new layers to Iv3. One convolutional layer before the original input, which maps from $\mathbb{R}^{3\times3\times299\times299}$ to $\mathbb{R}^{3\times299\times299}$, allowing the network to consider the three 2-dimensional projections at the same time, and two linear layers at the output, one which maps from the original latent space $\mathbb{R}^{1000}$ to a new latent space $\mathbb{R}^{512}$ and one which maps from that latent space to the output space $\mathbb{R}^{2}$ (two classes). During the training of FTIv3, the weights of Iv3 are frozen, and only the new layers are trained. The 3D CNN was designed to take as an input the 3D images, and all the layers were trained on the classification task. To augment the size of the training dataset, we reduced the size of the sequences using $L = N = 150$. For FTIv3 and the 3D CNN, the latent space is accessed by removing the last linear layer of the neural networks. For all inception networks, the means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, as well as the covariances $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ of the distributions of the GAN and MC-generated hit sequences, respectively, are measured to compute the FID between the two distributions in a given latent space [10]:

$$FID^2 = ||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2||_2^2 + \text{Tr}\Big(\boldsymbol{C}_1 + \boldsymbol{C}_2 - 2(\boldsymbol{C}_1\boldsymbol{C}_2)^{1/2}\Big), \tag{1}$$

thereby providing a measure of the similarity between the GAN and MC-generated distributions in that space.

4. Results

4.1. Training convergence

The evolution of the FIDs in the latent spaces of Iv3, FTIv3 and the custom 3D CNN during the training of the GAN is shown in Fig. 2. All plots suggest that the FID decreases during training and converges within the training time. While the evolution of the FID in the Iv3 and FTIv3 latent spaces seems well-behaved, the FID in the latent space of the 3D CNN reaches values of $\mathcal{O}(10^4)$ and decreases less monotonically. The training of the 3D CNN was found to be unstable, with exploding activations after a few epochs of training, which explains the large FID scores. The spikes in this FID plot are, therefore, likely due to a poorly constructed latent space resulting from the unstable training. Fundamentally, this could be due to the high sparsity of the data, making convolutions in $\mathbb{R}^3$ sub-optimal. In the future, an alternative would be to use convolutions on a graph adapted to the detector's geometry.

4.2. Comparison of two GAN architectures

We used the FID in the latent spaces of Iv3, FTIv3 and the custom 3D CNN to compare two GAN architectures developed by the COMET collaboration: one with self-attention layers [14, 5] (SAGAN) and one without (GAN). Values are shown in Table 1. While the FIDs in all latent spaces suggest the superiority of the SAGAN architecture, one must account for the fact that FID has a statistical bias which scales with $1/N$ [15], where N is the number of samples used to compute the metric, with an unknown prefactor which can differ between two sets of samples. In other words, the FID measured from a finite number of samples cannot directly be compared across GANs. To validate the comparison between models, we also measure the effectively unbiased FID [15] in the latent spaces, with results shown in Fig. 3 for the Iv3 latent space. Plots clearly show that the limit of the FID as $N \to \infty$ is lower for the SAGAN than the GAN.

Table 1. FID between distributions of Monte Carlo simulated hits, GAN generated hits, and SAGAN generated hits measured in the latent spaces of 3 inception networks.

	Inception v3			Fine-tuned	3D CNN
Projection	**Y-Z**	**T-Y**	**T-Z**		
FID, MC-MC	4.6	6.4	6.4	0.64	8,408
FID, GAN-MC	69.5	95.5	92.8	17.8	54,678
FID, SAGAN-MC	37.3	72.1	71.3	4.5	15,150

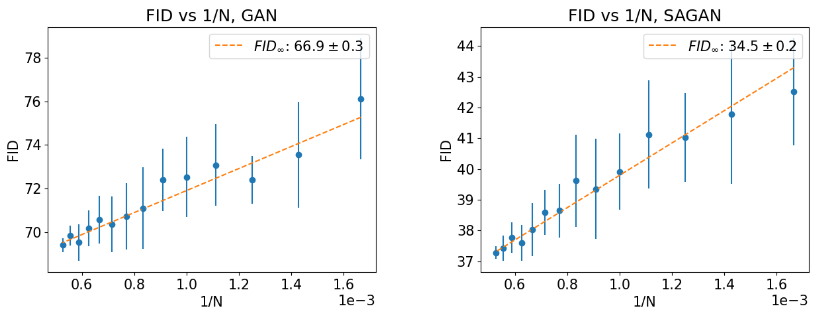

Figure 3. Measurement of the effectively unbiased FID (FID_∞) between GAN and MC hits and SAGAN and MC hits measured in the latent space of Inception v3. FID_∞ is the limit of FID as the number of samples tends to infinity.

5. Conclusion

The COMET experiment aims to improve the state-of-the-art single-event sensitivity to neutrino-less $\mu - e$ conversions by a factor of 2.7×10^4. However, accurately estimating the sensitivity of COMET and other very high background experiments is challenging using Monte Carlo simulations. Therefore, the COMET collaboration developed GANs to approximate and sample the distribution of background hit sequences in the experiment's Cylindrical Drift Chamber (CDC). Fréchet Inception Distance is the standard metric to evaluate GANs in computer vision but is often impractical in HEP applications since it requires a fully trained inception network. In this work, we described how we used pre-trained, fine-tuned and custom inception networks to evaluate the GANs developed by the COMET collaboration.

We found that the custom inception network was unable to form a well-constructed latent space, highlighting the importance of carefully considering the architecture and training process of custom inception networks. On the other hand, we have shown that pre-trained and fine-tuned inception networks can be valuable tools for evaluating the performance of our GANs. Our study has contributed to the COMET collaboration by validating the convergence of the GANs' training and providing guidance on the choice of GAN architectures for future work by demonstrating the advantages of using self-attention GANs to improve the generated hit sequence distributions.

Acknowledgments
We would like to express our gratitude to Dr Matthias Dubouchet, who developed the GAN models evaluated in this work, and whose valuable contributions significantly enhanced the quality of our research. We would also like to thank our MSci project supervisor, Professor Yoshi Uchida, and the rest of Imperial College's COMET group for their guidance and support throughout this exciting project. Finally, we acknowledge the COMET Physics and Software Group, which developed the Monte Carlo code that forms the basis of this work.

References

[1] The COMET Collaboration (COMET) 2020 *PTEP* **2020** 033C01 (*Preprint* 1812.09018)
[2] Lee M 2018 *Frontiers in Physics* **6** ISSN 2296-424X
[3] van der Schaaf A 2021 *SciPost Phys. Proc.* **5** 008
[4] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 Generative adversarial nets *Advances in Neural Information Processing Systems* vol 27
[5] Dubouchet M 2023 (to be published) Sensitivity and Background Estimates towards Phase-I of the COMET Muon-to-Electron Conversion Search phd thesis
[6] Lu Y, Collado J, Whiteson D and Baldi P 2021 *Phys. Rev. D* **103**(3) 036012
[7] Kansal R, Duarte J, Su H, Orzari B, Tomei T, Pierini M, Touranakou M, Vlimant J R and Gunopulos D 2021 (*Preprint* 2106.11535)
[8] Rehm F, Vallecorsa S, Borras K and Krücker D (*Preprint* 2103.13698)
[9] Dohi K (*Preprint* 2009.04842)
[10] Heusel M, Ramsauer H, Unterthiner T, Nessler B and Hochreiter S 2017 Gans trained by a two time-scale update rule converge to a local nash equilibrium *Proceedings of the 31st International Conference on Neural Information Processing Systems* NIPS'17 (Red Hook, NY, USA) p 6629–6640 ISBN 9781510860964
[11] Kansal R, Duarte J, Su H, Orzari B, Tomei T, Pierini M, Touranakou M, vlimant j r and Gunopulos D *Advances in Neural Information Processing Systems* pp 23858–23871
[12] Szegedy C, Vanhoucke V, Ioffe S, Shlens J and Wojna Z 2015 (*Preprint* 1706.08500)
[13] Deng J, Dong W, Socher R, Li L J, Li K and Fei-Fei L 2009 ImageNet: A large-scale hierarchical image database *2009 IEEE Conference on Computer Vision and Pattern Recognition* pp 248–255
[14] Zhang H, Goodfellow I, Metaxas D and Odena A 2018 (*Preprint* 1805.08318)
[15] Chong M and Forsyth D 2020 Effectively unbiased fid and inception score and where to find them *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Los Alamitos, CA, USA) pp 6069–6078