Diversity Balancing Generative Adversarial Networks for fast simulation of the Zero Degree Calorimeter in the ALICE experiment at CERN

Jan Dubiński
Warsaw University of Technology
jan.dubinski.dokt@pw.edu.pl

Kamil Deja
Warsaw University of Technology

Sandro Wenzel
CERN

Przemysław Rokita
Warsaw University of Technology

Tomasz Trzciński
Warsaw University of Technology
Jagiellonian University
Tooploox

Abstract

Generative Adversarial Networks (GANs) are powerful models able to synthesize data samples closely resembling the distribution of real data, yet the diversity of those generated samples is limited due to the so-called mode collapse phenomenon observed in GANs. Conditional GANs are especially prone to mode collapse, as they tend to ignore the input noise vector and focus on the conditional information. Recent methods proposed to mitigate this limitation increase the diversity of generated samples, yet they reduce the performance of the models when similarity of samples is required. To address this shortcoming, we propose a novel method to control the diversity of GAN-generated samples. By adding a simple, yet effective regularization to the training loss function we encourage the generator to discover new data modes for inputs related to diverse outputs while generating consistent samples for the remaining ones. More precisely, we reward or penalize the model for synthesising diverse images, matching the diversity of real and generated samples for a given conditional input. We show the superiority of our method on simulating data from the Zero Degree Calorimeter of the ALICE experiment in LHC, CERN.

1 Introduction

Generative Adversarial Networks (GANs) [1] constitute a gold standard for synthesizing complex data distributions and they are, therefore, widely used across various applications [2, 3, 4]. Among others, they are employed in high energy physics experiments at the Large Hadron Collider (LHC) at CERN, where they allow to speed up the process of simulating particle collisions [5, 6, 7, 8]. In this context, GANs are used to generate samples of possible detectors' responses resulting from a collision of particles described with a series of physical parameters.

Although by using conditional GANs (cGANs) [9] we can obtain more context-dependent generations that are closer to the values observed in real experiments, these models are considered more vulnerable to the so-called mode collapse phenomenon [10], observed as a tendency to generate a limited number of different outputs per each conditional prior. This, in turn, significantly reduces the effectiveness of employing GANs for particle collision simulations, as alignment of generated samples with the real data distribution is fundamental for drawing correct conclusions from the performed experiments.

Machine Learning and the Physical Sciences workshop, NeurIPS 2022.

To address the above-mentioned limitations of cGANs, recent methods [11, 12, 13] attempt to increase the diversity of generated samples by modifying the associated cost function. However, they do not consider conditioning the diversity on the input conditioning values, assuming a uniform distribution of diversity across all of them. This assumption is rarely observed in practical applications, for instance in particle collision simulations at CERN diversity of generated samples highly depends on a set of conditioning variables.

In this work, we identify this shortcoming of existing models and propose a simple, yet effective method to balance the diversity of GAN-generated samples. In principle, we introduce a regularization method that enforces GANs to follow the diversity observed in the original dataset for a given conditional value. More exactly, we penalise or reward the model for generating diverse samples, matching the diversity observed in the real and generated data for a given conditional input. Our approach, dubbed DivBal-GAN, is readily applicable for conditional image synthesis models and does not require any modification of the baseline GAN architecture.

We apply our method to a challenging task of simulating data from the Zero Degree Calorimeter of the ALICE experiment in LHC, CERN outperforming competing approaches. The main contribution of this paper is a novel method for a controlled increase of the diversity of GAN-generated results.

2 Related work

2.1 Generative simulations:

The need for simulating complex processes exists across many scientific domains. In recent years, solutions based on generative machine learning models have been proposed as an alternative to existing methods in cosmology [14] and genetics [15]. However, one of the most profound applications for generative simulations is in the field of High Energy Physics, where machine learning models can be used as a resource-efficient alternative to classic Monte Carlo-based [16] approaches.

Recent attempts [5, 8, 17] leverage solutions based on Generative Adversarial Networks [1] or Variational Autoencoders [18]. Although those methods offer considerable speed-up of the simulation process, they also suffer from the limitations of existing generative models. Controlling the diversity of simulated results while maintaining the high fidelity of the simulation is one of the challenges of using generative models for such applications.

2.2 Mode collapse and sample diversity in cGAN:

The authors of MS-GAN [12] address the mode collapse problem in cGANs by introducing a regularization term that maximizes the dissimilarity between two images generated from two different latent codes. DS-GAN [11] tries to tackle the problem with a similar approach. In DivCo [13] the authors use contrastive learning to achieve diverse conditional image synthesis. They introduce a latent-augmented contrastive loss which encourages images generated from distant latent codes to be dissimilar and those generated from close latent codes to be similar. The similarity of images is measured using their latent representations extracted from the discriminator network.

Our approach shares a similar method of calculating the diversity of images with [12] and [11]. However, contrary to those approaches we do not base our measure of diversity on pixels of generated images. Instead we operate on image representation, similarly to [13].

In principle, all previously described approaches do not account for different levels of variance of samples corresponding to different conditional inputs and instead maximize the diversity of the results generated for all conditional inputs.

3 Method

Traditional conditional GANs are trained using adversarial loss $L_{adv}(G, D)$. This loss function encourages the generator to produce realistic data, but as observed by [13] it does not directly promote the diversity of synthesised samples. To alleviate this problem Mao et al. [12] propose adding a regularization term that penalizes low diversity of generated samples. More precisely, the introduced method maximizes the ratio of the distance between two images generated from two different latent codes z_1, z_2 and the same conditioning value c with respect to the distance between those latent codes.

$$\mathcal{L}_{\mathrm{ms}} = \left(\frac{d_{\mathbf{I}}\left(G\left(c, \mathbf{z}_1\right), G\left(c, \mathbf{z}_2\right)\right)}{d_{\mathbf{z}}\left(\mathbf{z}_1, \mathbf{z}_2\right)} \right)^{-1} \tag{1}$$

Although this approach successfully forces the generator to produce dissimilar examples it does not account for different levels of sample diversity for different conditioning input c. To address this issue we propose a simple yet effective modification of the regularization term by introducing a balancing term $f(div_{real}(c), div_{gen}(c))$.

$$f(div_{real}(c), div_{gen}(c)) = tanh\Big(d_{\mathbf{I}}\left(X_c^1, X_c^2\right) - d_{\mathbf{I}}\big(G\left(c, \mathbf{z}_1\right), G\left(c, \mathbf{z}_2\right)\big)\Big) \tag{2}$$

where X_c^1, X_c^2 are 2 samples taken from the training dataset that correspond to the conditioning prior c. To account for varying levels of sample diversity for each conditioning input c we multiply the regularization term defined by Eq. 1 by the balancing term $f(div_{real}(c), div_{gen}(c))$.

$$\mathcal{L}_{\mathrm{div}} = f\big(div_{real}(c), div_{gen}(c)\big) * \left(\frac{d_{\mathbf{I}}\left(G\left(c, \mathbf{z}_1\right), G\left(c, \mathbf{z}_2\right)\right)}{d_{\mathbf{z}}\left(\mathbf{z}_1, \mathbf{z}_2\right)} \right)^{-1} \tag{3}$$

Introducing the balancing term forces the generator to match the diversity observed in the training data with respect to conditional values. When the dissimilarity of generated samples is lower than the dissimilarity of real samples, the loss function encourages diverse generations. At the same time, when the dissimilarity observed in the produced samples is higher than in the training data samples, the model is penalized for generating diverse results. The overall objective of training DivBal-GAN is given by Eq.4 where λ_{div} is a hyperparameter controlling the strength of the regularization.

$$\mathcal{L} = \mathcal{L}_{\mathrm{adv}}(G, D) + \lambda_{\mathrm{div}}\mathcal{L}_{\mathrm{div}}(G) \tag{4}$$

Additionally, we propose to base the distance d_I on the dissimilarity of latent representations of images rather than the dissimilarity of pixels. We measure the distance between two generated images by calculating the L_1 metric between their latent representations obtained from the penultimate layer of the discriminator during training. This change shifts the focus of the generator from the visual dissimilarity of images to the difference in their underlying characteristics extracted by the encoder.

4 Experiments

We evaluate our method on simulating data from the Zero Degree Calorimeter from the ALICE experiment. We compare our approach to conditional DC-GAN [19], MS-GAN [12] and DivCo [13].

4.1 Zero Degree Calorimeter

The task of simulating the response of the Zero Degree Calorimeter (ZDC) offers a challenging benchmark for generative models. The dataset consists of 295867 samples obtained from the GEANT4 [16] simulation tool. Each response is created by a single particle described with 9 attributes (mass, energy, charge, momenta, primary vertex).

During the simulation process, the particle is propagated through the detector for over 100 meters while simulation tools must account for all of its interactions with the detector's matter. The end result of the simulation is the energy deposited in the calorimeter's fibres, which are arranged in a grid with 44 × 44 size. We treat the calorimeter's response as a 1-channel image with 44 × 44 pixels, where pixel values are the number of photons deposited in a given fibre. To create the dataset the simulation was run multiple times for the same input particles. For that reason, multiple possible outcomes correspond to the same particle properties. We refer to this dataset as HEP.

Although the process that governs the propagation of the particles is non-deterministic by nature, the majority of particles create consistent ZDC responses. However, a subset of particles produces highly diverse results and allows for multiple possible calorimeter responses. In the top row of Fig. 1 we present sampled simulations for two particles corresponding to different conditional values.

4.2 Results

The most common method for evaluating GANs utilizes Frechet Inception Distance (FID) [20]. However, for the HEP dataset, we propose a domain-specific evaluation scheme that better measures the quality of the simulation. Following the calorimeter's specification [21] we base our evaluation procedure on 5 channels calculated from the pixels of generated images that reflect the physical properties of simulated collision. To measure the quality of the simulation we compare the distribution of channels for the original and generated data using Wasserstein distance [22].

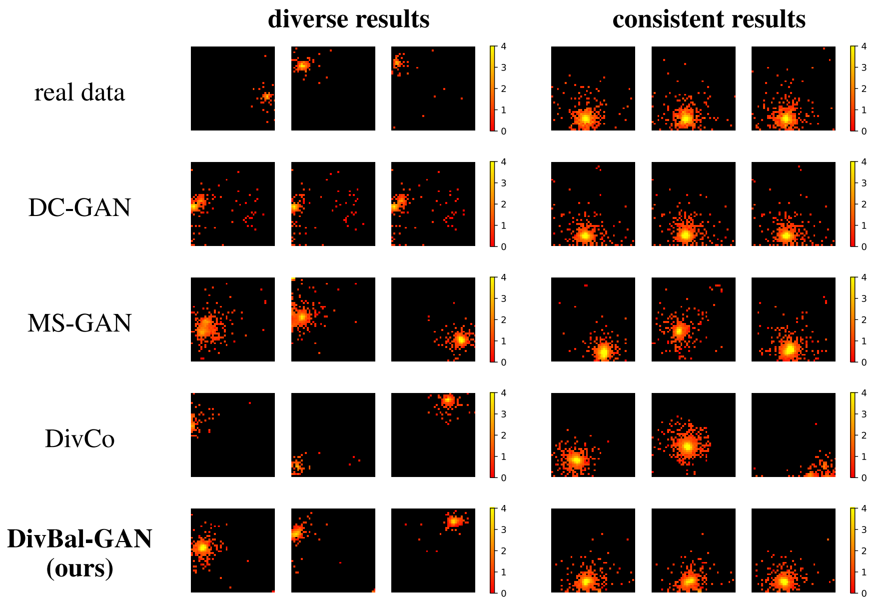

Figure 1: Examples of calorimeter response simulations with different methods. DC-GAN works well for particles with consistent responses but fails to generate diverse outcomes when needed. Although MS-GAN and DivCo successfully increase the diversity of generated samples those models do not distinguish between particles that should produce diverse or consistent showers. Our method is able to generate diverse results while producing consistent responses for appropriate particles.

As presented in Tab. 1, our approach outperforms other solutions on the HEP datasets. In Fig. 1 we demonstrate that our method is able to generate diverse results for a specific subset of particles while keeping consistent responses for the remaining conditional inputs. The positive impact of this approach on the distribution of the generated samples is further confirmed by Fig. 2 where we compare channel distribution for DivBal-GAN and conditional DC-GAN. Our method increases the fidelity of the simulation by smoothing the distribution of generated responses and covering the whole range of possible outputs.

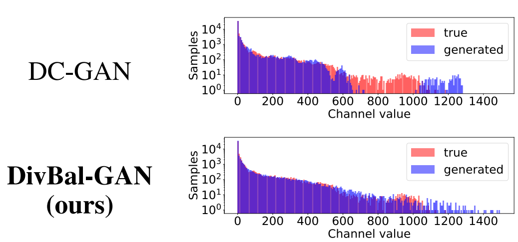

Figure 2: Comparison of channel values distribution for a selected channel. Our method decreases the differences between the distribution of original and generated data and smooths the distribution of the synthesised results.

Table 1: Results comparison for the HEP datasets. DivBal-GAN achieves the lowest Wasserstein distance between channels calculated from original and generated data.

	Wassesrstein ↓
Real	-
DC-GAN	7.6
MS-GAN	21.7
DivCo	14.3
DivBal-GAN	**4.4**

The additional regularization term for training of DivBal-GAN does not influence the inference speed of the model. In our initial experiments, we observe a speed-up of simulations of two orders of magnitude when compared to the standard Monte-Carlo approach. With DivBal-GAN this computation boost is observed without degradation in simulation quality. We leave a detailed analysis of this performance gain and the influence of fast simulations on physical experiments for future work.

5 Conclusions

In this work, we introduce a simple, yet effective modification of the loss function for conditional generative adversarial networks. Our solution enforces increased sample diversity for a subset of conditional data without affecting samples that are characterised by conditional values associated with consistent responses.

We show that our solution outperforms other comparable approaches on the challenging practical dataset of calorimeter response simulations in the ALICE experiment at CERN.

Acknowledgments

This research was funded by National Science Centre, Poland grant no 2020/39/O/ST6/01478, grant no 2018/31/N/ST6/02374 and grant no 2020/39/B/ST6/01511.

Broader impact

As the Large Hadron Collider located at CERN has recently undergone extensive upgrades, the amount of data recorded during the experiments has increased greatly. Therefore, there exists a need for novel simulation methods for high energy physics (HEP). Moreover, advancements in generative machine learning benefit other domains of science with the demand for generative simulations, such as nuclear medicine or cosmology. Finally, by improving HEP experiments, we indirectly contribute to disciplines that benefit from the results of HEP research. Outside the context of physical sciences, increasing the diversity of generated samples helps to mitigate any potential discriminatory biases learned by the model during training.

References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[2] Saman Motamed and Farzad Khalvati. Inception augmentation generative adversarial network. *CoRR*, abs/2006.03622, 2020. URL `https://arxiv.org/abs/2006.03622`.

[3] Shengyu Zhao, Jonathan Cui, Yilun Sheng, Yue Dong, Xiao Liang, Eric I-Chao Chang, and Yan Xu. Large scale image completion via co-modulated generative adversarial networks. *CoRR*, abs/2103.10428, 2021. URL `https://arxiv.org/abs/2103.10428`.

[4] Maciej Zieba, Piotr Semberecki, Tarek El-Gaaly, and T. Trzciński. Bingan: Learning compact binary descriptors with a regularized gan. *NeurIPS*, 2018.

[5] Michela Paganini, Luke de Oliveira, and Benjamin Nachman. Calogan: Simulating 3d high energy particle showers in multi-layer electromagnetic calorimeters with generative adversarial networks. *CoRR*, abs/1705.02355, 2017.

[6] Kamil Deja, Jan Dubiński, Piotr Nowak, Sandro Wenzel, Przemysław Spurek, and Tomasz Trzcinski. End-to-end sinkhorn autoencoder with noise generator. *IEEE Access*, 9:7211–7219, 2021. doi: 10.1109/ACCESS.2020.3048622.

[7] Jan Dubiński, Kamil Deja, Sandro Wenzel, Przemysław Rokita, and Tomasz Trzciński. Selectively increasing the diversity of gan-generated samples, 2022. URL `https://arxiv.org/abs/2207.01561`.

[8] Raghav Kansal, Javier Duarte, Hao Su, Breno Orzari, Thiago Tomei, Maurizio Pierini, Mary Touranakou, Jean-Roch Vlimant, and Dimitrios Gunopulos. Particle Cloud Generation with Message Passing Generative Adversarial Networks. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2021.

[9] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets, 2014. URL `https://arxiv.org/abs/1411.1784`.

[10] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, Xi Chen, and Xi Chen. Improved techniques for training gans. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL `https://proceedings.neurips.cc/paper/2016/file/8a3363abe792db2d8761d6403605aeb7-Paper.pdf`.

[11] Dingdong Yang, Seunghoon Hong, Yunseok Jang, Tianchen Zhao, and Honglak Lee. Diversity-sensitive conditional generative adversarial networks. In *Proceedings of the International Conference on Learning Representations*, 2019.

[12] Qi Mao, Hsin-Ying Lee, Hung-Yu Tseng, Siwei Ma, and Ming-Hsuan Yang. Mode seeking generative adversarial networks for diverse image synthesis. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[13] Rui Liu, Yixiao Ge, Ching Lam Choi, Xiaogang Wang, and Hongsheng Li. Divco: Diverse conditional image synthesis via contrastive generative adversarial network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021.

[14] Ruichen Rong, Shuang Jiang, Lin Xu, Guanghua Xiao, Yang Xie, Dajiang J Liu, Qiwei Li, and Xiaowei Zhan. MB-GAN: Microbiome Simulation via Generative Adversarial Network. *GigaScience*, 10(2), 02 2021. ISSN 2047-217X. URL `https://doi.org/10.1093/gigascience/giab005`.

[15] Andres C. Rodríguez, Tomasz Kacprzak, Aurelien Lucchi, Adam Amara, Raphaël Sgier, Janis Fluri, Thomas Hofmann, and Alexandre Réfrégier. Fast cosmic web simulations with generative adversarial networks. *Computational Astrophysics and Cosmology*, 5(1), Nov 2018. ISSN 2197-7909. URL `http://dx.doi.org/10.1186/s40668-018-0026-4`.

[16] Sebastien Incerti, Ioanna Kyriakou, MA Bernal, MC Bordage, Z Francis, Susanna Guatelli, V Ivanchenko, M Karamitros, N Lampe, Sang Bae Lee, et al. Geant4-dna example applications for track structure simulations in liquid water: A report from the geant4-dna project. *Medical physics*, 45(8):e722–e739, 2018.

[17] Martin Erdmann, Jonas Glombitza, and Thorben Quast. Precise simulation of electromagnetic calorimeter showers using a wasserstein generative adversarial network. *Computing and Software for Big Science*, 3 (1), Jan 2019. ISSN 2510-2044.

[18] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *CoRR*, abs/1312.6114, 2013.

[19] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[20] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in neural information processing systems*, pages 6626–6637, 2017.

[21] G Dellacasa, X Zhu, M Wahn, FM Staley, V Danielian, TL Karavicheva, DP Mikhalev, N Carrer, M Gheata, G Stefanek, et al. Alice technical design report of the zero degree calorimeter (zdc). Technical report, ALICE, 1999.

[22] Ilya Tolstikhin, Olivier Bousquet, Sylvain Gelly, and Bernhard Schoelkopf. Wasserstein auto-encoders. *arXiv preprint arXiv:1711.01558*, 2017.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes]
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [Yes]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [No]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [No]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]