Event Generation in Particle Physics with the B-VAE

Sydney Otten
University of Amsterdam, Radboud University and straiqr.ai
sydney@straiqr.ai

Shankha Banerjee
Durham University

Melissa van Beekveld
Radboud University

Sascha Caron
Radboud University and Nikhef

Luc Hendriks
Radboud University

Caspar van Leeuwen
SURFsara

Damian Podareanu
SURFsara

Roberto Ruiz de Austri
University of Valencia

Michael Spannowsky
Durham University

Rob Verheyen
University College London

Philip Waite
Durham University

Abstract

We present a study for the generation of events from a physical process with generative deep learning. To simulate physical processes it is not only important to produce physical events, but also to produce these events with the right frequency of occurrence (density). We investigate the feasibility to learn the event generation and the frequency of occurrence with standard and modified Variational Autoencoders (VAEs) to produce events like Monte Carlo generators with up to 20 final state objects. We study $pp \to t\bar{t}$ including the decay of the top quarks and a simulation of the detector response. By buffering density information of Monte Carlo events in latent space given the encoder of a VAE and by introducing a smudge factor α we are able to construct a prior for the sampling of new events from the decoder that yields distributions that are in very good agreement with real Monte Carlo events and are generated 10^8 times faster.

1 Introduction

The simulation of physical and other statistical processes is typically done with the help of sampling quasi-random numbers as the input of a simulation. The simulation turns random numbers into observable physical events. This is known as the Monte Carlo (MC) method. A fundamental problem with the numerical calculations simulating physical processes today is their immense need for computational resources which restricts the corresponding scientific progress regarding its velocity, budget and therefore general availability. As an example, the full pipeline of the MC event generation in particle physics experiments including the detector response may take up to 10 minutes per event [1–8] and largely depends on MC sampling algorithms such as VEGAS [9]. Accelerating the event generation pipeline with the help of machine learning will also provide a significant speed up for signal studies allowing e.g. broader searches for signals of new physics. Another issue is the probabilistic nature underlying event generation: the inability to exactly specify the event that ought to be generated. Data analysis often requires to generate events which are kinematically similar to events seen in the data. Current event generators typically require the generation of many events and then, after generation, to select the interesting events with a low efficiency. Furthermore, currently the simulation itself cannot be learned directly from real detector data.

Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019), Vancouver, Canada.

In this paper we outline an alternative approach with deep generative machine learning (ML) models and perform a feasibility study of our method. The main problem we tackle in this paper is to create a generator that learns to sample from distributions that are in good agreement with the distributions found in the training data. However, most efforts of the machine learning community regarding generative models are typically not aimed at learning the correct frequency of occurrence. So far, applications of generative ML approaches in particle physics focused on image generation [10–13] due to the recent successes in unsupervised machine learning with generative adversarial networks (GANs) [14–16] to generate realistic images according to human judgement [17, 18]. Here, we investigate deep generative models, namely Variational Autoencoders (VAEs) [19] and provide proof for the feasibility of expanding the use of generative models in the form of VAEs beyond images for applications in particle physics.

To test our setup, we study the $t\bar{t}$ production from proton collisions, where at least one of the top quarks is required to decay leptonically. Additionally we study $t\bar{t}$ production with up to 20 final state objects. We find that by using a density information buffer with the decoder of a VAE we are able to produce a realistic collection of events that follows the distributions present in the MC event data.

2 Methodology

2.1 Event Generation

We have generated 0.5 million events of $pp \to t\bar{t}$, where at least one of the top-quarks is required to decay leptonically. We used MG5_aMC@NLO v6.3.2 [1] for the matrix element generation, using the NNPDF PDF set [20]. Madgraph is interfaced to Pythia 8.2 [2], which handles the showering of the matrix element level generated events. The matching with the parton shower was done using the MLM merging prescription [21]. Then a quick detector simulation was done with Delphes 3 [3, 4], using the ATLAS detector card.

For the dataset with up to twenty final state objects, we implement the leading order (LO) matrix element (ME) generation of $t\bar{t}$ events in MG5_aMC@NLO v2.6.5 [1], using the NNPDF PDF set (NNPDF2.1 LO) [20]. We allow the tops to decay both leptonically and hadronically, yielding a mixed sample of fully leptonic, semi-leptonic and fully hadronic final states. The ME events are then showered using Pythia 8 [2]. An analysis is performed with the resulting hepmc files using the following selection cuts:

- $p_T^{\ell} > 10$ GeV,
- $p_T^{j} > 30$ GeV,
- $|y^{\ell,b,\tau}| < 2.5$.

Here, ℓ, j, b are the leptons (e, μ), light-jets and $b-$tagged jets and p_T, y denote the transverse momentum and the rapidity of an object. The τ-leptons are allowed to decay both leptonically and hadronically. The jets are constructed using the anti-k_t jet clustering algorithm [**?**] within the `FastJet` framework [4]. For the b-tagging, the B-mesons are identified and are matched with the jets with which it has $R < 0.2$ [1]. These jets are further required to be within the aforementioned p_T and y ranges. Finally, we require require 70% of such jets to be b-tagged. For the τ-tagged jets, we require that the τ-hadrons also lie within $\Delta R = 0.2$ of some jet and require them to abide by the p_T, y requirements. The leptons are isolated by demanding that the total hadronic activity around $\Delta R < 0.2$ of that lepton is less than 10% of its p_T. Finally, we also demand a flat 70% tagging efficiency to these τ-tagged jets. Then, the CSV file that is used for training the B-VAE contains the following observables: `Event weight`, E_T , ϕ_{E_T} , $p_T^{obj_i}$, y^{obj_i} , ϕ^{obj_i} , E^{obj_i} , index. Here, E_T denoted the the missing transverse energy. The energy of a visible object i is denoted by E, where i runs from 1 to maximally 20. This means that we store a maximum of 20 objects per event, but there also may be less objects in a single event.

[1] $R = \sqrt{\Delta y^2 + \Delta\phi^2}$, where ϕ is the azimuthal angle.

2.2 Generative Models

We model the generation of collider events $\boldsymbol{x}$ by assuming that the events we want to sample follow an unknown distribution $p(\boldsymbol{x})$ and that $p(\boldsymbol{x})$ depends on some set of hidden latent variables $\boldsymbol{z}$. Practically, we treat this as a two step process where we first sample from a parametrized prior $p_\theta(\boldsymbol{z})$ and then sample $\boldsymbol{x}$ from the conditional probability $p_\theta(\boldsymbol{x}|\boldsymbol{z})$. Variational Autoencoders [19] are explicit probabilistic models that allow us to efficiently learn the parameters θ to sample from an approximation of the marginalized likelihood $p_\theta(\boldsymbol{x}) = \int d\boldsymbol{x}\, p_\theta(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z})$ which is otherwise intractable in many cases, especially when $p_\theta(\boldsymbol{x}|\boldsymbol{z})$ is a deep neural network. VAEs consist of two neural networks: 1) the probabilistic encoder $q_\phi(\boldsymbol{z}|x)$, an approximation to the true, intractable posterior $p(\boldsymbol{z}|\boldsymbol{x})$, that creates a distribution over latent variables for a true observation and 2) the probabilistic decoder $p_\theta(\boldsymbol{x}|\boldsymbol{z})$ that creates a distribution over possible $\boldsymbol{x}$ events given some latent code z. The two neural networks with parameters ϕ and θ are jointly trained by maximizing a practical estimator of the lower bound of the marginal likelihood

$$\mathcal{L} = -D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x})\|p_\theta(\boldsymbol{z})) + \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log(p_\theta(\boldsymbol{x}|\boldsymbol{z})\right], \tag{1}$$

where D_{KL} is the Kullback-Leiber divergence that is a distance measure for probability distributions, i.e. the first term of the objective incentivizes that the distribution over latent variables is trained to be similar to some chosen prior, which in our case is $p_\theta(\boldsymbol{z}) = \mathcal{N}(0, 1)$. The second term corresponds to the negative reconstruction error of $p_\theta(\boldsymbol{x}|\boldsymbol{z})$ given some input $\boldsymbol{z}$ that is sampled from $q_\phi(\boldsymbol{z}|\boldsymbol{x})$. Thus, the loss function of the VAE that is minimized is proportional to $D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x})\|p_\theta(z))$ and the mean squared error between a true observation $\boldsymbol{x}$ and its reconstructed output $\boldsymbol{x'}$. Additionally, we introduce a factor $B \ll 1$ [22] to the D_{KL} term to emphasize good reconstructions, such that we can write the loss function of the VAE as

$$L = \frac{1}{M}\sum_{i=1}^{M}(1-B)\cdot MSE + B\cdot D_{KL}, \tag{2}$$

where M is the batch-size. Training this and then constructing the prior

$$q_{\phi,X}(\boldsymbol{z}) = \frac{1}{N}\sum_{i=1}^{N} q_\phi(\boldsymbol{z}|\boldsymbol{x}_i)p(\boldsymbol{x}_i) \tag{3}$$

by aggregating the encodings of true observations $\boldsymbol{x}$ yields the B-VAE.

It turns out that the quality of reconstructions deteriorates the larger the highest multiplicity present in the data set. This is related to the increasing sparsity of $\boldsymbol{x}$: because all events have the same dimensionality in our setup, independent of their actual number of final state objects, many entries with the value 0 are implied. Assuming $\#f_{max}$ is the maximum number of final state objects present in our data, then for event i, the number of 0 entries is $4\cdot(\#f_{max} - \#f_i)$ - and our data contains events with $\#f_i = \{0, \ldots, 20\}$. In principle the B-VAE is also capable of learning the marginalized likelihood for high-multiplicity events but it turns out that the performance can be improved by incorporating additional terms into the loss function in eq. 2 that classify the type and number of final state objects.

3 Results

In Fig. 1, the results for the more complicated $t\bar{t}$ production with a subsequent semi-leptonic decay are shown. We show results using the two different values for the latent space dimension ($\dim(\mathrm{z})$), as well as using different values for α and β. We train the B-VAE on events that have a maximum of four jets and two leptons in the final state. For simplicity we do not discriminate between b-jets and light-flavored jets, nor between different kinds of leptons. A jet is defined as a clustered object that has a minimum transverse momentum (p_T) of 20 GeV in the Monte Carlo simulation. Regarding $\dim(\mathrm{z})$ we conclude that both 16 and 32 can deliver satisfactory results, although we get slightly better reconstructions for $\dim(\mathrm{z}) = 32$. The results also show that we require a low value of β to get satisfying distributions. Fig. 1 shows a very good agreement for the invariant mass of all four jets, the MET and METϕ. Finally, generating 10^7 $t\bar{t}$ events with the VAE has taken 177.5 seconds on an Intel i7-4790K and is therefore 10^8 faster than the traditional MC methods.

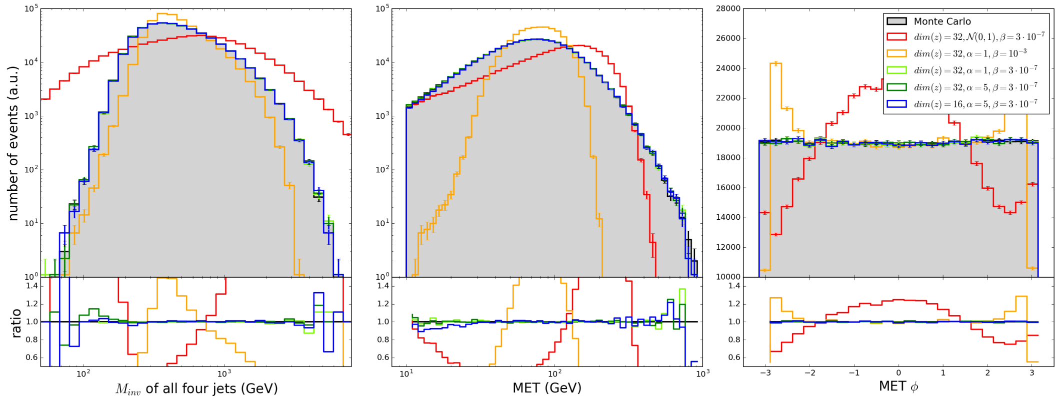

Figure 1: Events that are generated by the Monte Carlo generator for the $pp \to t\bar{t}$ process (gray), by the standard VAE (red line) and by the B-VAE (rest) for different values of α, β and $\dim(z)$. Shown from left to right is the invariant mass distributions for the four jets, the missing transverse energy (MET) and the azimuthal angle ϕ of the MET vector. All energies are given in GeV.

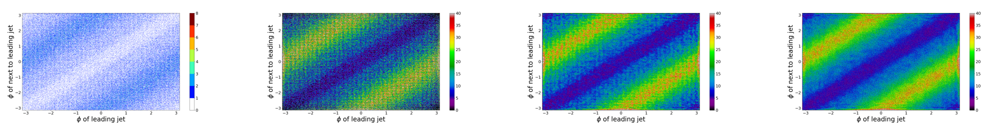

Figure 2: Histograms of events that are generated by the Monte Carlo generator with ≈ 475000 events and the B-VAE for $\alpha = 1$, $\alpha = 5$ and $\alpha = 10$ with ≈ 15 million events each for the full range $[-\pi, \pi] \times [-\pi, \pi]$. We show the histograms for the full ranges of ϕ. The full range is subdivided into 1000×1000 bins.

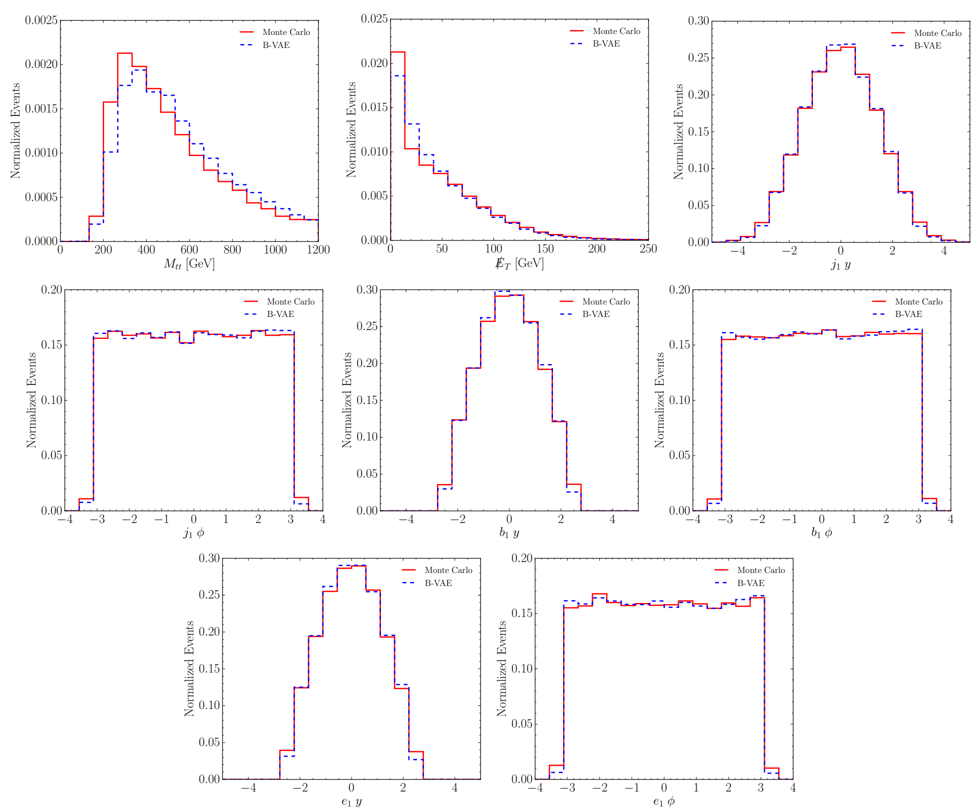

Figure 3: The ditop mass, MET, as well as y and ϕ for the leading jet, b-jet and electron for the dataset with up to 20 final state objects.

Fig. 2 shows four histograms for the training data (≈ 475000 events) and ≈ 15 million generated events from the B-VAE with three different $\alpha = \{1, 5, 10\}$ for ϕ of the leading jet vs. ϕ of the next to leading jet (ϕ_1 vs ϕ_2). We observe that for $\alpha = 1$, most of the blank space in (ϕ_1 vs. ϕ_2) is already filled with frequencies that globally follow the density pattern of the Monte Carlo data. However, we note that this pattern is not learned smoothly but in a way that resembles egg cartons with local peaks that seem to come from random seeds in the training data whose latent space representations have a small width and still contains some holes. Increasing α alleviates this issue: for $\alpha = 5$ almost all holes disappear but the landscape in (ϕ_1, ϕ_2) is still mountaineous which is smoothened out for $\alpha = 10$.

Fig. 3 shows preliminary results for the B-VAE that is extended with classification loss terms and trained on data containing up to 20 final state objects. We see that y and ϕ are learned perfectly, while there is some deviation for the MET and the ditop mass system. The results prove that it is also possible to learn the generation of events that can have low and very high multiplicities at the same time.

4 Conclusion

We have provided more evidence for the capability of deep generative models to learn the laws of physics with VAEs and in particular with our modification of it, the B-VAE. By creating a density information buffer of real MC events in the latent space of a VAE and by introducing a smudge factor α, i.e. with the B-VAE, we presented a way to generate events whose probabilistic characteristics are in very good agreement with those found in data simulated with Monte Carlo event generators for $pp \to t\bar{t}$ events. Adding a classification term to the B-VAE loss allows to reliably model the event generation for low and high multiplicities with up to 20 final state objects. These results indicate not only usefulness for particle physics but beyond that for all branches of science that involve computationally expensive Monte Carlo simulations, the interest to create a generative model from experimental data or the need to sample from high-dimensional and complex probability distributions.

References

[1] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5 : Going Beyond. *JHEP*, 06:128, 2011.

[2] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.

[3] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *JHEP*, 02:057, 2014.

[4] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur. Phys. J.*, C72:1896, 2012.

[5] G. Corcella, I. G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. H. Seymour, and B. R. Webber. HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes). *JHEP*, 01:010, 2001.

[6] T. Gleisberg, Stefan. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter. Event generation with SHERPA 1.1. *JHEP*, 02:007, 2009.

[7] Alexander Belyaev, Neil D. Christensen, and Alexander Pukhov. CalcHEP 3.4 for collider physics within and beyond the Standard Model. *Comput. Phys. Commun.*, 184:1729–1769, 2013.

[8] Wolfgang Kilian, Thorsten Ohl, and Jurgen Reuter. WHIZARD: Simulating Multi-Particle Processes at LHC and ILC. *Eur. Phys. J.*, C71:1742, 2011.

[9] G. Peter Lepage. VEGAS: AN ADAPTIVE MULTIDIMENSIONAL INTEGRATION PROGRAM. 1980.

[10] Michela Paganini, Luke de Oliveira, and Benjamin Nachman. CaloGAN : Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks. *Phys. Rev.*, D97(1):014021, 2018.

[11] Martin Erdmann, Jonas Glombitza, and Thorben Quast. Precise simulation of electromagnetic calorimeter showers using a Wasserstein Generative Adversarial Network. 2018.

[12] Luke de Oliveira, Michela Paganini, and Benjamin Nachman. Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis. *Comput. Softw. Big Sci.*, 1(1):4, 2017.

[13] Harold Erbin and Sven Krippendorf. GANs for generating EFT models. 2018.

[14] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks. *arXiv e-prints*, page arXiv:1406.2661, June 2014.

[15] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv e-prints*, page arXiv:1511.06434, November 2015.

[16] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large Scale GAN Training for High Fidelity Natural Image Synthesis. *arXiv e-prints*, page arXiv:1809.11096, September 2018.

[17] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved Techniques for Training GANs. *arXiv e-prints*, page arXiv:1606.03498, June 2016.

[18] 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. *arXiv e-prints*, page arXiv:1706.08500, June 2017.

[19] Diederik P Kingma and Max Welling. Auto-Encoding Variational Bayes. *arXiv e-prints*, page arXiv:1312.6114, December 2013.

[20] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J.*, C77(10):663, 2017.

[21] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, Roberto Pittau, and Antonio D. Polosa. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 07:001, 2003.

[22] Christopher P. Burgess, Irina Higgins, Arka Pal, Loic Matthey, Nick Watters, Guillaume Desjardins, and Alexander Lerchner. Understanding disentangling in β-VAE. *arXiv e-prints*, page arXiv:1804.03599, April 2018.