Mass conditioned and constrained normalising flows for jet generation

B Käch[1], D Krücker[1], I Melzer-Pellmann[1], M Scham[1,2,3], S Schnake[1,2] and A Verney-Provatas[1,2]

[1]Deutsches Elektronen-Synchrotron DESY, Notkestrasse 85, Germany
[2]RWTH Aachen University, III. Physikalisches Institut A, Aachen, Germany
[3]Institute for Advanced Simulation - Jülich Supercomputing Centre, Jülich, Germany

E-mail: benno.kaech@desy.de

Abstract. Fast data generation based on Machine Learning has become a major research topic in particle physics. This is mainly because the Monte Carlo simulation approach is computationally challenging for future colliders, which will have a significantly higher luminosity. The generation of collider data is similar to point cloud generation with complex correlations between the points. In this study, the generation of jets with up to 30 constituents with Normalising Flows using Rational Quadratic Spline coupling layers is investigated. Without conditioning on the jet mass, our Normalising Flows are unable to model all correlations in data correctly, which is evident when comparing the invariant jet mass distributions between Monte Carlo simulated and model-generated data. Using the invariant mass as a condition for the coupling transformation enhances the performance on all tracked metrics. In addition, we demonstrate how to sample the original mass distribution by interpolating the empirical cumulative distribution function. Similarly, the variable number of constituents is taken care of by introducing an additional condition on the number of constituents in the jet. Furthermore, we study the usefulness of including an additional mass constraint in the loss term. On the `JetNet` dataset, our model yields a stable and quick training and performs comparably to the state-of-the-art.

1. Introduction

High-Energy Physics (HEP) has benefited from the advances in Machine Learning (ML) since the analysis of HEP data is a high-dimensional multivariate problem. ML has been applied in multiple ways ranging from classification to regression tasks [1]. In HEP, detailed simulations of the physical processes are commonly available, which describe the details of the experimental measurement with high precision. These Monte Carlo simulations (MC) provide labelled data and are needed in large numbers to cover the extreme areas of the physical phase space. The simulations for the CMS detector at the Large Hadron Collider (LHC), for example, require about 50% [2] of the current CMS computing budget. An even larger number of simulations will be needed for the upcoming high-luminosity phase of the LHC [3]. Therefore, generative modelling with Deep Learning (DL), which might well be able to reduce the computational time by more than an order of magnitude, sparked great interest in the HEP community.

In this study, the generation of jets is investigated using different Normalising Flow (NF) [4, 5, 6, 7] architectures, which are motivated by their stable convergence. A powerful summary statistic used in this study is the invariant mass of the jet, which depends on the

features of all constituents forming the jet. Thus the distribution for the mass calculated from ML-generated jets should describe the one calculated from MC simulation very well. This paper first provides a summary of related work in the field of generative models for jet generation in Section 2, which also introduces the dataset on which this research was conducted. This is followed by a description of the architecture of the proposed models in Section 3. In Section 4 the various visualisations of the generated data are presented and compared to holdout data for different architectures using the same procedure as [8]. The results are then compared to state-of-the-art results from the field and are briefly discussed in Section 5. Finally, a conclusion is given in Section 6, which highlights the important aspects of this approach and plans for future work.

2. Related Work

Data of particle clouds with up to 30 particles and 3 features each is provided in Ref. [9] for bench-marking different generative models. The authors also propose the message-passing GAN [8](MPGAN) architecture which outperforms state-of-the-art generative models by orders of magnitude on nearly every metric. This study contributes to a comparison of different Normalizing Flow (NF) approaches to this problem, whereas the original study [10] only considered GANs. However, GANs suffer from unstable training [11] and generally need a strong inductive bias in their architecture to perform well on the problem as seen in [8].

2.1. Discrete Normalising Flows

Discrete Normalising Flows [12] are a relatively new subject in HEP, a conclusive overview is given in Refs. [6] and [7]. Invertible and differentiable transformations f are sought, which transform the training data distribution X to a Normal Gaussian distribution Z. If the model can transform the data to a Gaussian sufficiently well, the data distribution is sampled by first drawing from a normal distribution and applying the inverted transformations. This study considers only NFs with coupling layers, as auto-regressive models suffer from either slow training or generation [6].

2.2. Dataset

In this study, the `JetNet` [8] datasets are used. There are 3 different datasets available, each containing jets with an energy of about 1 TeV, with each jet containing up to 30 constituents. The difference in the datasets lies in the jet-initiating parton. Datasets for top quark, light quark and gluon initiated jets are studied. The jet constituents are considered to be massless and can therefore be described by their 3-momenta or equivalently by transverse momentum p_T, pseudorapidity η, and azimuthal angle ϕ. In the `JetNet` dataset, these variables are given relative to the jet axs: $\eta_i^{\text{rel}} = \eta_i^{\text{particle}} - \eta^{\text{jet}}$, $\phi_i^{\text{rel}} = (\phi_i^{\text{particle}} - \phi^{\text{jet}}) \bmod 2\pi$, and $p_{T,i}^{\text{rel}} = p_{T.i}^{\text{particle}} / p_T^{\text{jet}}$, where i runs over the particles in a jet. It is worth noting that the jet axis is determined using all particles generated with PYTHIA [13] initially, but due to the dataset's constraint of 30 particles, the jets in the dataset are not properly centered. The invariant mass m_{jet} of a jet is an essential high-level feature containing important physics information. It is a global variable that depends on the correlations between the single jet constituents and provides therefore an important metric for the performance of the generative model. For the relative quantities above, we can define the relative jet mass as $(m^{\text{rel}})^2 = (\sum_i {E_i}^{\text{rel}})^2 - (\sum_i \vec{p}_i^{\,\text{rel}})^2 = m_{\text{jet}}^2 / p_{\text{T,jet}}^2$.

3. Architecture

The NFs used in this study consist of multiple stacked coupling layers, where each coupling layer splits the input features into two (in this study equally sized) sets. The former is mapped with an identity transformation, while to the latter a non-trivial parameterised element-wise

transformation is applied. We studied the feasibility of rational quadratic monotonic splines [14] a transformation.

3.1. Mass Conditioning

Normalising flows are conditioned by adding variables to the input of the Neural Networks (NNs) predicting the coupling transformations' parameters. These additional variables are not transformed. Conditioning enhances the expressivity of the transformation, as more information is given to the network predicting the parameters of the coupling layers. In this study, the invariant jet mass m and the number of particles per jet n were tried for conditioning. These variables also need to be supplied when the NF is used in the generating direction and thus need to be generated independently. For one conditioning variable, this is done by transforming the variable with its cumulative distribution function to a uniform distribution. This one-dimensional transformation is then interpolated with monotonic piece-wise cubic Hermite polynomials [15], which is invertible. If two conditions are used an autoregressive approach is used: first the number of particles is sampled from its empirical probability mass function and the previous approach is applied to the mass distribution conditioned on the number particles. This autoregressive approach is especially viable here because the number of particles per jet is discrete and thus there are a finite number of conditional mass distributions. In the following, the variables that are used to condition are given in brackets after the name of the model.

3.2. Mass Constraint

To further improve the mass modelling, a mass constraint is introduced. Here, another loss term is added: the L_2 norm between the mass calculated from a generated jet and the value of the mass the generated jet has been conditioned with:

$$L_{\text{MSE}} = |m_{\text{cond}} - m_{\text{gen}}(x_{\text{gen}})|^2 \tag{1}$$

It is important to note that this means that the NF is used in both directions during one training step.

3.3. Variable Sized Particle Clouds

As NFs need to be invertible, there are major constraints on the dimension of the generated data space, disallowing the generation of a variable amount of particles. In this study, point clouds with fewer than 30 particles were zero-padded. Noise in the order of $O(10^{-7})$ was added to the zero-padded particles as otherwise, it would be an ill-posed problem in the context of normalising flows. For the case where the number of particles is given as a condition, the particles coming after the value of the condition are set to zero.

4. Results

In this section results for 3 different architectures are presented. For each architecture, an equally sized random search was conducted. Results are presented for all 3 datasets. The model needs 9.77 ± 0.04 μs for the sampling of a jet on an NVIDIA P100 when using a batch size of 50000. The training time for all models ranges from 1 to 2 hours on an NVIDIA P100.

4.1. Evaluation Metrics

The Wasserstein distance is a powerful metric for comparing the generated distribution with the real data distributions, but it is not tractable in more than one dimension. Thus to evaluate the performance of the generative model the Wasserstein distances W_1 between the inclusive distributions of $(\eta^{\text{rel}}, \phi^{\text{rel}}, p_T^{\text{rel}})$ of the MC-simulated and the flow-generated jets are considered. Additionally, the Wasserstein distance between energy flow polynomials [16] and between the

invariant mass of the jet are considered. Other metrics used include coverage (COV) [17] and minimum matching distance (MMD) [17] and Fréchet ParticleNet Distance (FPND) [8].

4.2. Performance comparison

The evaluation metrics are calculated as in [8], using batches of 10000 for Wasserstein distances and 50000 for FPND. Table 1 shows the results for the different architectures - including Vanilla NF (VNF), conditioned NF (CNF), and conditioned and constrained NF (CCNF) - and compared to the state-of-the-art [8] (MP-MP, MP-LFC). For architectures that rely on conditioning, the variables used for conditioning are given in the brackets.

Table 1. Comparison between the best performing models from [8], with 5 different model from this study. No systematic differences can be seen except for the FPND metric and W_1^{EFP} on the top-quark data set.

Jet Class	Model	$W_1^M \times 10^3$	$W_1^P \times 10^3$	$W_1^{EFP} \times 10^5$	FPND	COV ↑	MMD
Gluon	MP-MP	0.7 ± 0.2	0.9 ± 0.3	$\mathbf{0.7 \pm 0.7}$	**0.12**	**0.56**	0.037
	MP_LFC-MP	0.69 ± 0.07	1.8 ± 0.3	0.9 ± 0.2	0.20	0.54	0.037
	VNF	4.3 ± 0.2	2.2 ± 0.5	3 ± 1	1.82	0.53	**0.035**
	CNF (m)	0.9 ± 0.3	$\mathbf{0.5 \pm 0.2}$	0.8 ± 0.6	0.53	**0.56**	0.036
	CNF (m,n)	0.8 ± 0.3	1.2 ± 0.3	0.8 ± 0.8	0.31	0.55	**0.035**
	CCNF (m)	$\mathbf{0.6 \pm 0.2}$	0.6 ± 0.2	1.2 ± 0.9	0.90	0.54	0.036
	CCNF (m,n)	0.7 ± 0.4	0.9 ± 0.3	1.1 ± 0.6	0.54	**0.56**	0.036
Light Quark	MP-MP	$\mathbf{0.6 \pm 0.2}$	4.9 ± 0.5	$\mathbf{0.7 \pm 0.4}$	0.35	0.50	0.026
	MP_LFC-MP	0.7 ± 0.2	2.6 ± 0.4	0.9 ± 0.9	**0.08**	0.52	0.037
	VNF	2.8 ± 0.6	2.2 ± 0.4	1.9 ± 0.6	1.30	**0.54**	**0.024**
	CNF (m)	0.9 ± 0.3	1.1 ± 0.4	$\mathbf{0.7 \pm 0.3}$	0.49	0.53	**0.024**
	CNF (m,n)	1.0 ± 0.2	4.2 ± 0.6	$\mathbf{0.7 \pm 0.4}$	0.73	0.53	**0.024**
	CCNF (m)	0.7 ± 0.2	$\mathbf{0.8 \pm 0.5}$	$\mathbf{0.7 \pm 0.4}$	0.55	0.50	0.025
	CCNF (m,n)	0.7 ± 0.1	4.5 ± 0.9	1.0 ± 0.6	1.22	0.50	0.025
Top Quark	MP-MP	0.6 ± 0.2	2.3 ± 0.3	$\mathbf{2 \pm 1}$	**0.37**	0.57	0.071
	MP_LFC-MP	0.9 ± 0.3	2.2 ± 0.7	$\mathbf{2 \pm 1}$	0.93	0.56	0.073
	VNF	6.6 ± 0.6	2.2 ± 0.5	15 ± 1	7.76	**0.59**	**0.070**
	CNF (m)	1.8 ± 0.5	$\mathbf{1.2 \pm 0.4}$	3 ± 1	2.62	0.57	**0.070**
	CNF (m,n)	$\mathbf{0.55 \pm 0.08}$	1.7 ± 0.3	3 ± 1	2.19	0.57	0.071
	CCNF (m)	0.7 ± 0.2	12.2 ± 0.5	8 ± 3	8.52	0.44	0.076
	CCNF (m,n)	1.1 ± 0.5	1.3 ± 0.4	4.9 ± 0.7	2.27	0.56	0.073

The inclusive distribution of the particle features as well as the invariant mass distribution for the top-quark dataset are presented in Fig. 1.

As mentioned previously, the invariant mass is a powerful variable to test the correlation between the individual jet constituents. In Fig. 2 the mass distribution is shown for the different configurations. Significant differences can be observed, which are especially evident in the case of the top-quark sample, where there is more structure in the mass distribution. This structure is an artefact from the imperfect anti-k_T clustering, as in some cases the b quark in the top decay chain is not added to the jet cone and thus we only see a peak of the W boson. While the vanilla NF does not learn this feature, the other two models reproduce the ground truth reasonably well.

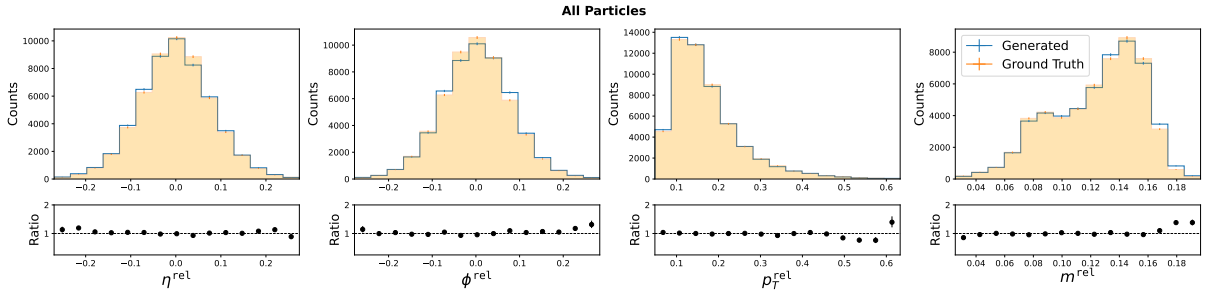

Figure 1. Inclusive histograms of $(\eta^{rel}, \phi^{rel}, p_T^{rel}, m^{rel})$ for all particles. Shown are examples for the top-quark dataset which was trained with the CCNF architecture using 2 conditions.

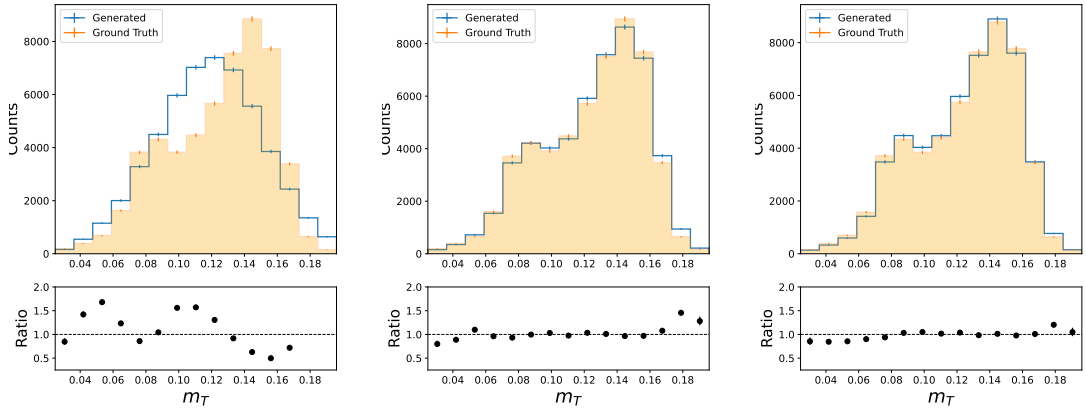

Figure 2. Mass distribution for the different models on the top-quark dataset. On the left is VNF, in the middle CNF (m) and to the right CCNF (m).

5. Discussion

The results indicate that Vanilla NF, while producing acceptable results for individual features and inclusive distributions, fail to model more complex correlations between particles. This may be due to the construction of the coupling layers, as only 50% of all features are seen by the networks in these layers. Using the mass as a condition and constraining it improves the results, but raises questions about the importance of the latent representation. The results of the model are comparable to the state-of-the-art on all but the FPND metric. The training of the NF is significantly simpler and more stable compared to the one of GANs, and the NF seem to perform well although their architecture is general without a problem specific inductive bias.

6. Conclusion

In this study 3 different architectures for the generation of variable-sized jets are presented. Whilst the most basic rational quadratic spline flows are unable to model the correlations present in the training data, conditioning and constraining the flow on the invariant mass of the jet gives results comparable to the state of the art. As the conditions need to be supplied during sampling, an additional simple model for sampling a two-dimensional distribution is presented as well. The model is evaluated on 6 different metrics, and various control plots between the "real" Monte Carlo simulated data and the flow-generated data are shown.

Acknowledgments

Benno Käch is funded by Helmholtz Association's Initiative and Networking Fund through Helmholtz AI (grant number: ZT-I-PF-5-64).

Moritz Scham is funded by Helmholtz Association's Initiative and Networking Fund through Helmholtz AI (grant number: ZT-I-PF-5-3).

Alexi Verney-Provatas is partly funded by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Sklodowska-Curie COFUND scheme with grant agreement No. 101034267.

This research was supported in part through the Maxwell computational resources operated at Deutsches Elektronen-Synchrotron DESY (Hamburg, Germany). The authors acknowledge support from Deutsches Elektronen-Synchrotron DESY (Hamburg, Germany), a member of the Helmholtz Association HGF.

References

[1] Feickert M and Nachman B 2021 A living review of machine learning for particle physics URL https://arxiv.org/abs/2102.02770

[2] CMS 2022 CMS offline and computing public results URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/CMSOfflineComputingResults

[3] Apollinari G, Brüning O, Nakamoto T and Rossi L 2015 High luminosity large hadron collider hl-lhc URL https://cds.cern.ch/record/2120673

[4] Tabak E G and Turner C V 2012 *Communications on Pure and Applied Mathematics* **66** 145–164

[5] Rezende D and Mohamed S 2015 Variational inference with normalizing flows *Proceedings of the 32nd International Conference on Machine Learning* (*Proceedings of Machine Learning Research* vol 37) (PMLR) pp 1530–1538 URL https://proceedings.mlr.press/v37/rezende15.html

[6] Kobyzev I, Prince S J and Brubaker M A 2021 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **43** 3964–3979 (*Preprint* 1908.09257) URL https://arxiv.org/abs/1908.09257

[7] Papamakarios G, Nalisnick E, Rezende D J, Mohamed S and Lakshminarayanan B 2019 URL https://arxiv.org/abs/1912.02762

[8] Kansal R, Duarte J M, Su H, Orzari B, Tomei T, Pierini M, Touranakou M, Vlimant J and Gunopoulos D 2021 *CoRR* **abs/2106.11535** (*Preprint* 2106.11535) URL https://arxiv.org/abs/2106.11535

[9] Kansal R jetnet library URL https://github.com/jet-net/JetNet

[10] Kansal R, Duarte J, Su H, Orzari B, Tomei T, Pierini M, Touranakou M, Vlimant J R and Gunopulos D 2021 *Zenodo*

[11] Kodali N, Abernethy J, Hays J and Kira Z 2017 On convergence and stability of gans URL https://arxiv.org/abs/1705.07215

[12] Rezende D and Mohamed S 2015 Variational inference with normalizing flows *Proceedings of the 32nd International Conference on Machine Learning* (*Proceedings of Machine Learning Research* vol 37) ed Bach F and Blei D (Lille, France: PMLR) pp 1530–1538 URL https://proceedings.mlr.press/v37/rezende15.html

[13] Sjöstrand T, Ask S, Christiansen J R, Corke R, Desai N, Ilten P, Mrenna S, Prestel S, Rasmussen C O and Skands P Z 2015 *Comput. Phys. Commun.* **191** 159 (*Preprint* 1410.3012)

[14] Durkan C, Bekasov A, Murray I and Papamakarios G 2019 Neural spline flows URL https://arxiv.org/abs/1906.04032

[15] Fritsch F N and Butland J 1984 *SIAM Journal on Scientific and Statistical Computing* **5** 300–304 URL https://doi.org/10.1137/0905021

[16] Komiske P T, Metodiev E M and Thaler J 2018 *Journal of High Energy Physics* **2018** URL https://doi.org/10.1007%2Fjhep04%282018%29013

[17] Achlioptas P, Diamanti O, Mitliagkas I and Guibas L 2018 Learning representations and generative models for 3D point clouds *Proceedings of the 35th International Conference on Machine Learning* (*Proceedings of Machine Learning Research* vol 80) ed Dy J and Krause A (PMLR) pp 40–49 URL https://proceedings.mlr.press/v80/achlioptas18a.html