Variational AutoEncoders for Anomaly Detection in VBS events within an EFT framework

G Boldrini[1,2], S Gennai[2], P Govoni[1,2], G Lavizzari[1,2]

[1] University of Milano-Bicocca, Ed. U2, Piazza della Scienza 3, - 20126 Milano, IT
[2] INFN Milano-Bicocca, Ed. U2, Piazza della Scienza 3, - 20126 Milano, IT

E-mail: `g.lavizzari1@campus.unimib.it`

Abstract. We present a machine-learning based method to detect deviations from a reference model, in an almost independent way with respect to the theory assumed to describe the new physics responsible for the discrepancies.

The analysis is based on an Effective Field Theory (EFT) approach: under this hypothesis the Lagrangian of the system can be written as an infinite expansion of terms, where the first ones are those from the Standard Model (SM) Lagrangian and the following terms are higher dimension operators. The presence of the EFT operators impacts the kinematic distributions by producing deviations from the shapes expected when the SM Lagrangian alone is considered.

We use a Variational AutoEncoder (VAE) trained on SM processes to identify EFT contributions as anomalies. While the output of the VAE when evaluated on the SM sample is very similar to the input, the events modified by EFT operators show output distributions very different from the inputs and hence they accumulate in the tails of the loss function. Since the training of the model does not depend on any specific new physics signature, the proposed strategy does not make specific assumptions on its nature.

In this talk we will discuss in detail the above-mentioned method using generator level VBS events produced at LHC and assuming, in order to estimate the sensitivity to possible new physics contributions, an integrated luminosity of $350 fb^{-1}$.

1. Introduction

The discovery of the Higgs boson by the Atlas and CMS collaborations [1, 2] in 2012 marked a major milestone in the validation of the Standard Model (SM) of particle physics. Nevertheless, many questions still remain unanswered: despite its success in providing accurate experimental predictions and theoretical explanations for many phenomena, the SM cannot be considered as a complete theory of fundamental interactions.

While the CERN Large Hadron Collider (LHC) has collected an unprecedented amount of data, no significant deviations from the SM have been observed yet. A possible explanation is that searches for Beyond Standard Model (BSM) phenomena are typically highly dependent on the specific theoretical model they target. Since the number of possible BSM theories and their variations in terms of free parameters is extremely large, and only few of them can be tested, it is possible that so far we have simply chosen the wrong theories to look for. An alternative approach is to build analyses that are as independent as possible from any underlying assumption on the new physics models.

In order to design an effective analysis of such kind, a general but still predictive theory that can regroup the largest possible number of BSM processes is needed. With this in mind, we chose to work within an Effective Field Theory (EFT) approach, where the SM is seen as the low energy approximation of a more general theory. The low energy footprints of such theory can be parametrized as higher order operators to be added to the SM Lagrangian, altering the expected kinematic distributions of a given process.

Our analysis strategy relies on unsupervised learning. The idea is to train the model on known physics, and then use it to detect EFT events as outliers, i.e. data following different patterns than the ones the model was trained to recognize. The model we chose is a Variational AutoEncoder (VAE) [3].

The physics process we used to test this strategy is the scattering of vector bosons (VBS), an event that takes place at the LHC when two partons of the incoming protons radiate vector bosons, which in turn interact. VBS is deeply connected to the electroweak symmetry breaking mechanism as a Higgs-less SM theory leads to a divergence of the VBS cross-section with increasing center of mass energy. Therefore, this process is a probe of the SM that is sensitive to modifications of the electroweak sector, which makes it an ideal field for the searches for BSM physics. In particular, we studied the set of VBS processes that leads to a final state with two same-sign W bosons. The decay products of the vector bosons produce a clean signature in the detector characterized by two jets with a large invariant mass in the forward region, two same-sign charged leptons and missing transverse energy.

This study has been conducted neglecting any background process that populates the aforementioned final state. We employed generator-level observables related to the kinematic of the charged leptons and of the final state partons originating from the initial scattering.

2. The SM as an Effective Field Theory

In an EFT interpretation of the SM, a more complete theory is expected to involve new matter content at energies well beyond the LHC scale, while its effects at low energy are parametrized as the additional terms obtained from the expansion of the SM Lagrangian:

$$\mathcal{L}_{EFT} = \mathcal{L}_{SM} + \sum_{i,d>4} \frac{c_i}{\Lambda^{d-4}} \mathcal{O}^{(d_i)} \tag{1}$$

where $\mathcal{O}_i^d$ is a set of dimension d operators, c_i are the so-called Wilson Coefficients that gauge the intensity of the operators and Λ represents the new physics energy scale. The first non-zero term following the SM Lagrangian is the set of dimension-six operators, since odd-dimensional operators would violate the accidental symmetries of the SM and therefore they are not taken into account. We employed the SMEFTsim package [4] to generate the EFT predictions at leading order in the U35 flavour scheme and m_W input scheme for the following operators: $\mathcal{Q}_W$, $\mathcal{Q}_{Hq}^{(1)}$, $\mathcal{Q}_{HW}$, $\mathcal{Q}_{qq}^{(1)}$, $\mathcal{Q}_{qq}^{(1,1)}$, $\mathcal{Q}_{qq}^{(3)}$, $\mathcal{Q}_{qq}^{(3,1)}$. The events were generated at parton level via MadGraph5_aMC@NLO [5].

The terms in the Lagrangian can be rewritten in terms of probability amplitudes, considering one operator at a time, as:

$$A_{BSM} = A_{SM} + \frac{c_\alpha}{\Lambda^2} \cdot \mathcal{A}_{\mathcal{Q}_\alpha} \tag{2}$$

By squaring A_{BSM} it is possible to obtain a quantity proportional to the probability of an event:

$$|A_{BSM}|^2 = |A_{SM}|^2 + \frac{c_\alpha}{\Lambda^2} \cdot 2Re(A_{SM} A^\dagger_{\mathcal{Q}_\alpha}) + \frac{c_\alpha^2}{\Lambda^4} \cdot |\mathcal{A}_{\mathcal{Q}_\alpha}|^2 \tag{3}$$

here the first term represents the pure SM contribution, while the following terms introduce a linear (LIN) and quadratic (QUAD) dependence on the EFT amplitudes. The overall effect of these operators is a modification of the kinematic distributions of a given process, as shown in figure 1. The entity of such modifications depends on the relative weight of the operators.

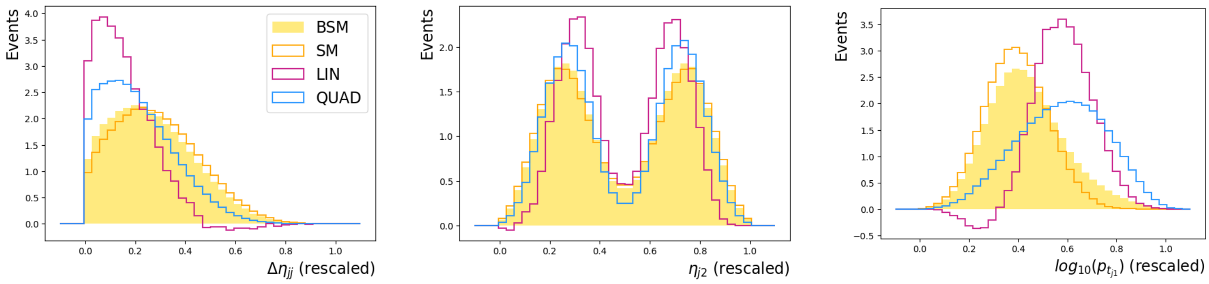

Figure 1. SM, LIN and QUAD distributions for $\delta\eta_{jj}$, η_{j2} and $\log_{10}(p_{T,j1})$, normalised to area 1. The weighted sum of the three contributions gives the BSM distribution.

3. Isolating EFT events with a VAE

First we scaled the input distributions between 0 and 1 and we computed the logarithm of the kinematic variables to reduce their dynamic range. We split the SM sample in two subsets comprising 80% and 20% of the total 900000 events, respectively used for training and testing. The whole strategy is implemented through the scikit-learn [6] and TensorFlow [7] libraries.

3.1. Simple VAE model

The VAE model is shown in figure 2. We studied a set of kinematic and angular distributions relevant for SSWW: the invariant masses of the dilepton/dijet systems (m_{ll}, m_{jj}), the transverse momentum of the two leptons/jets ($p_{t_{l1,2}}$, $p_{t_{j1,2}}$, the transverse momentum of the dilepton system ($p_{t_{ll}}$), the missing transverse momentum (MET), the pseudorapidities of leptons/jets ($\eta_{l1,2}$, $\eta_{j1,2}$, and the pseudorapidity and azimuthal angle differences between the two jets ($\Delta\eta_{jj}$, $\Delta\Phi_{jj}$). The inputs are first encoded as distributions in the latent space, from which a point is then sampled and decoded. Encoder and decoder are two separate networks comprising several densely connected layers, later combined into an end-to-end model. The model is trained by minimizing a sum of the Mean Squared Error (MSE) between input and output, to ensure a good reconstruction of the inputs, and the so-called Kullback-Leibler Divergence (KLD), which accounts for a regularization of the latent space by forcing the latent distributions to be close to gaussians. The Adam algorithm is used to optimize the training.

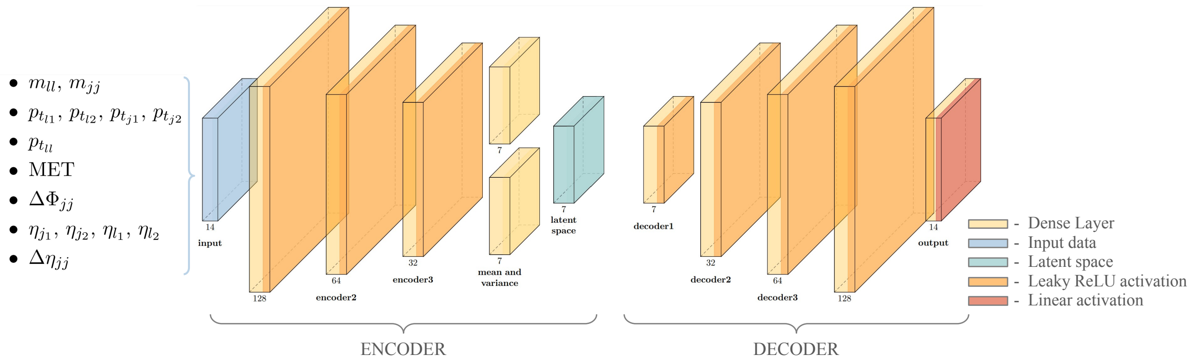

Figure 2. The VAE model employed.

The VAE model is trained only on SM events, therefore it learns the underlying patterns of known physics. Once trained, the model is evaluated also on BSM events: since the underlying mechanisms are different from the ones the model learnt during training, the output distributions present differences with respect to the input ones (figure 3). To assess the quality of the reconstruction of an event we compute the MSE between input and output, averaged on all the observables. By selecting the events for which the MSE is greater than a certain threshold (figure 4), we are able to identify an anomaly-enriched region.

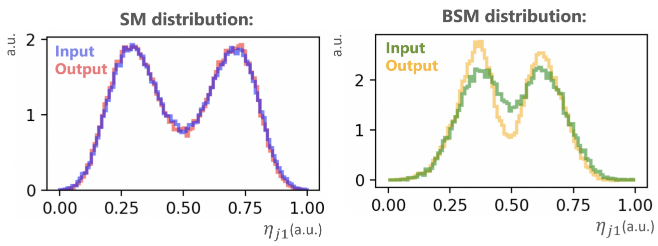

Figure 3. Comparison between the input and output of the model in the case of SM and BSM distributions for the pseudorapidity of the leading jet.

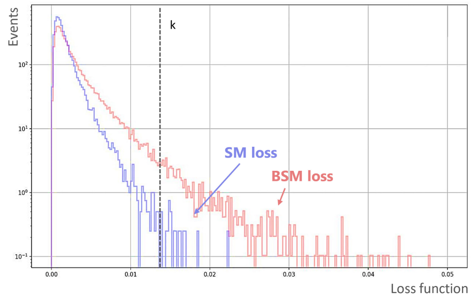

Figure 4. MSE for SM(blue) and BSM(red) events. Anomalous events lie in the tail of the loss function.

3.2. Embedding a classification step in the training to optimize for discrimination
The main limitation of this strategy is that even though our goal is the isolation of EFT events, during the training the model is only optimized to reconstruct a SM sample and the choices that improve the SM reconstruction are not always optimal for discrimination. For example, during our studies we have noticed that a larger latent space produces a better reconstruction for SM events but also decreases the discrimination power against EFT perturbations. This can be interpreted given the fact that a more complex latent space allows for the model to better learn the underlying data structure, and therefore the SM reconstruction improves. However, the model's extrapolation capabilities improve as well, thus degrading its discrimination power.

In order to overcome such limitation, we introduced a classification step in the training procedure by adding a third component to the model, namely a two-layers neural network that works as a classifier.

This new model is trained both on SM and EFT events: the VAE is trained to reconstruct the SM subset via the minimization of the MSE and KLD, then both the SM and EFT samples are run through it. The resulting MSE and KLD losses are then given as inputs to the classifier, which is trained by minimizing a binary cross-entropy. This strategy allows for embedding the discrimination process within the training, but at the price of gaining an additional dependence on the modelization of the new physics contribution. To keep the strategy as model independent as possible, we decided to use a single operator during training, and to later evaluate the performances of the model on the other ones as well.

3.3. Results
To quantify the discrimination power of the VAE and VAE+NN models we defined a proxy metric for the significance σ, which depends on the Wilson coefficients of the operator considered while testing the model:

$$\sigma(c_{op}) = \frac{|BSM(c_{op}) - SM|}{\sqrt{SM}} = \frac{|LIN(c_{op}) + QUAD(c_{op}^2)|}{\sqrt{SM}} \tag{4}$$

The last equality stems from the fact that, given our theoretical framework, the amount of events predicted following a BSM prior minus the events predicted considering only the SM contributions equals the amount of events predicted considering the LIN and QUAD operators. A model is considered sensitive to a given operator if $\sigma = 3$ for some value of c_{op} (table 1). The first conclusion we can draw is that it is possible to separate EFT contributions from known physics by means of the VAE model. Furthermore, the embedding of the classifier in the model improves the discrimination without impinging excessively the generality of the approach: the performances improve significantly for $\mathcal{Q}_W$, on which the model was trained, but also for $\mathcal{Q}_{qq}^{(1)}$, $\mathcal{Q}_{qq}^{(1,1)}$ and deteriorate only slightly for $\mathcal{Q}_{qq}^{(3)}$, $\mathcal{Q}_{qq}^{(3,1)}$ and $\mathcal{Q}_{HW}$. Furthermore, the new model is sensitive to $\mathcal{Q}_{Hq}^{(1)}$, which was not detected by the simple VAE.

Table 1. The value of c_{op} for which $\sigma(c_{op}) = 3$, considering an integrated luminosity of 350 fb^{-1}. The results for the VAE+NN refer to a model trained to reconstruct SM VBS EWK events and to discriminate them from a sample comprising contribution from the $\mathcal{Q}_W$ operator: during training c_W is set to 1, then the events are properly weighted by the correct value of c_{op}.

model	c_W	c_{qq}^{1}	$c_{qq}^{1,1}$	c_{qq}^{3}	$c_{qq}^{3,1}$	c_{Hq}^{1}	c_{HW}
VAE	0.34	0.56	0.29	0.04	0.06	-	0.41
VAE+NN	0.13	0.17	0.18	0.11	0.11	0.61	0.65

4. Conclusions and future perspectives

This study demonstrates the feasibility of isolating an EFT-enriched region by means of an unsupervised machine learning model. By training the model to recognise only the known SM physics patterns, we were able to detect BSM phenomena in a model-independent fashion. Such procedure was tested using the SSWW VBS process, considering an integrated luminosity of 350 fb^{-1}. This result meets the expectations intended for such kind of strategy, aimed at providing a region enriched with potentially interesting events rather than at enhancing the signal selection efficiency for specific BSM models.

While this study proved that unsupervised models can be instrumental for the detection of EFT perturbations with respect to the SM VBS SSWW signature, a detailed analysis including other sources of backgrounds is needed in order to assess the concrete sensitivity reach of the proposed strategy. Further studies will be needed in this direction including the major sources of background for the $2l2\nu + 2j$ final state, namely events with nonprompt or fake leptons and the small QCD induced production.

References

[1] ATLAS Collaboration, Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, *Phys.Lett.B* **716** (2012) 1-29, arXiv:1207.7214v2.

[2] CMS Collaboration, Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC, *Phys.Lett.B* **716** (2012) 30-61, arXiv:1207.7235v2.

[3] Kingma D P, Welling M, An Introduction to Variational Autoencoders, *Foundations and Trends in Machine Learning*: Vol. **12** (2019): No. 4, arXiv:1906.02691v3.

[4] Brivio I, SMEFTsim 3.0 – a practical guide, *JHEP 04 (2021)* **073**, arXiv:2012.11343v3.

[5] Alwall J et al., The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, *JHEP 07 (2014)* **079**, arXiv:1405.0301v2.

[6] Pedregosa F et al., Scikit-learn: Machine Learning in Python, *JMLR* **12**, pp. 2825-2830, 2011, arXiv:1201.0490v4.

[7] Abadi M et al., TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, arXiv:1603.04467v2.