PROCEEDINGS OF SCIENCE

Unsupervised tagging of semivisible jets with normalized autoencoders in CMS

Florian Eble[a,*] **on behalf of the CMS collaboration**

[a] *ETH Zürich*

E-mail: florian.eble@cern.ch

A particularly interesting application of autoencoders (AE) for High Energy Physics is their use as anomaly detection (AD) algorithms to perform a signal-agnostic search for new physics. This is achieved by training the AE on standard model physics and tagging potential signal events as anomalies. The use of an AE as an AD algorithm relies on the assumption that the network better reconstructs examples it was trained on than ones drawn from a different probability distribution, i.e. anomalies. Using the search for non resonant production of semivisible jets as a benchmark, we demonstrate the tendency of AEs to generalize beyond the dataset they are trained on, hindering their performance. We show how normalized AEs, specifically designed to suppress this effect, give a sizable boost in performance. We further propose a different loss function and signal-agnostic training stopping condition to reach the optimal performance.

PoS(EPS-HEP2023)491

The European Physical Society Conference on High Energy Physics (EPS-HEP2023)
21-25 August 2023
Hamburg, Germany

[*] Speaker

1. Introduction

The standard model (SM) of particle physics has been tested experimentally to very high precision and always found in excellent agreement with theoretical calculations. However, it does not explain the existence of dark matter, suggested by astronomical observation such as the cosmic microwave background power spectrum [1]. Despite an extensive search program for specific new physics models, there is so far no hint of dark matter at collider experiments. This motivates to search for new physics in a signal-agnostic way. In this report, using the search for non resonant production of semivisible jets as a benchmark, we demonstrate how normalized autoencoders (NAE) can be trained in a fully signal model agnostic way to search for new physics in the CMS experiment [2]. Compared to the first application of NAE to detect anomalous jets [3], which uses jet images, the loss function and regularization proposed in Ref. [4], we use jet substructure variables, a different loss function and a different strategy to define the best training point to use for inference.

2. Benchmark new physics model: semivisible jets

Semivisible jets [5] (SVJ) are a new physics signature arising in Hidden Valley theories where the dark sector is made of dark quarks interacting via a confining $SU(N)$ force, dark quantum chromodynamics (QCD). In this family of models, dark quarks are expected to hadronize in the dark sector, forming dark bound states. A fraction of them is unstable and promptly decays back to SM quarks, which then hadronize in the SM sector. Stable dark hadrons escape detection. The different jet substructure of SVJs, due to the double hadronization step and invisible dark hadrons, can be exploited to classify them versus SM jets. The signal model is parametrized by the mass of the t-channel mediator, m_{Φ}, and the invisible fraction of the jet, r_{inv}, defined as the ratio of the number of invisible dark hadrons and the total number of dark hadrons.

3. Autoencoders

3.1 Theoretical description

Autoencoders (AE) are neural networks composed of two parts: an encoder, which maps the input features space to a lower dimensional latent space, and a decoder which maps the latent space to an output space with same dimension as the input space. AEs are trained to minimize the reconstruction error between input and output, such that examples out of the training distribution, i.e. anomalies, have a higher reconstruction error. Trained on SM data, AEs can thus perform signal-agnostic searches for new physics [6, 7]. In the case of SVJs, AEs are trained on SM jets.

3.2 The problem of out-of-distribution reconstruction

AEs were proven to well perform anomalous detection of SVJs versus background QCD jets [8]. This study thus considers a top-quark jets dataset, which contains a mix of b-jets, light quark jets from resolved hadronic top decay, boosted W decay, boosted hadronic top decay and initial state radiation gluon jets, from $t\bar{t}$ production production. Simulated events are used. Both $t\bar{t}$ background and SVJ signal events are generated at parton level with `MadGraph5_amc@nlo` [9], final state quarks are hadronized with PYTHIA [10], and events are reconstructed with the CMS detector using GEANT4 [11]. For SVJ events, the Hidden Valley module implemented in PYTHIA version 8.2 is employed for the hadronization of the dark quarks. The detailed event selection can be found in Ref. [12]. 20 000 background events were used for training (85%) and evaluation (15%), and 10 000 to 20 000 signal events, depending on the signal hypothesis, were used for evaluation. Signal events are not used for the training.

Ten independent AEs, to reduce statistical fluctuations from one training to another and provide average performance, were trained on this top-quark jets dataset until minimal validation loss (no decrease for 100 epochs). They take 8 jet substructure input features, mapped to a normal distribution: major and minor jet axes, first energy flow polynomial, energy correlation function $C_2^{\beta=0.5}$, transverse momentum dispersion, softdrop mass, 2- and 3-subjettiness. The architecture is a fully connected neural network with 10, 10, 6, 10, 10 neurons. Other technical details can be found in Ref. [12]. The signal (SVJ) and background (top jet) reconstruction errors are shown in Fig. 1, together with the average Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) for multiple SVJ signals. The AEs generalize (reconstruct with low error) out of the training phase-space (out-of-distribution, OOD): the signal and background reconstruction errors have same value at the minimal validation loss. An illustration of OOD is given in Fig. 2.

Any other metric than the minimal validation loss, that would make use of the signal sample, e.g. the signal reconstruction error or AUC, would be biased towards the signals utilized and thus not signal-agnostic.

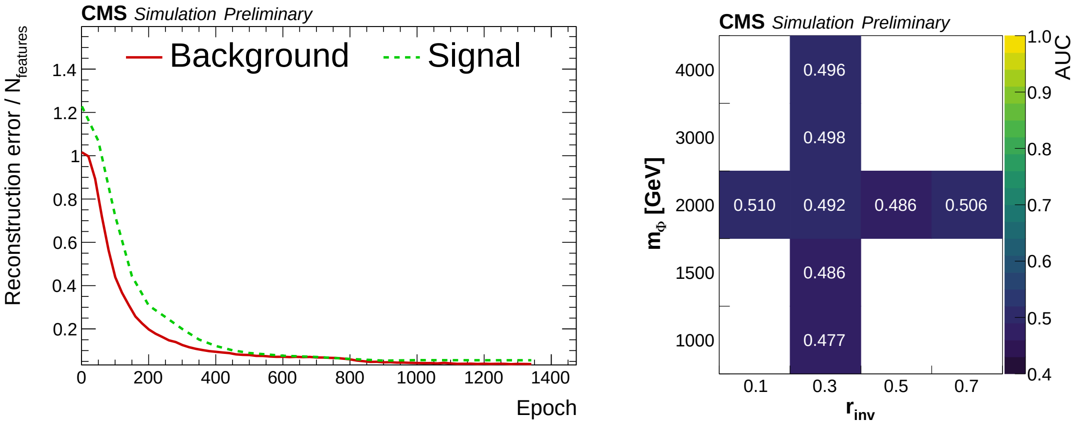

Figure 1: (Left) Average background and signal (m_Φ = 2 TeV, r_{inv} = 0.3) reconstruction errors for one representative training. (Right) Average Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) for all 10 trainings. The epoch with lowest validation loss was used for evaluation. The AUC is calculated for several mediator masses m_Φ with fixed invisible fraction r_{inv}, and several r_{inv} for fixed m_Φ for illustration purposes, but the full 2D parameter space could be explored.

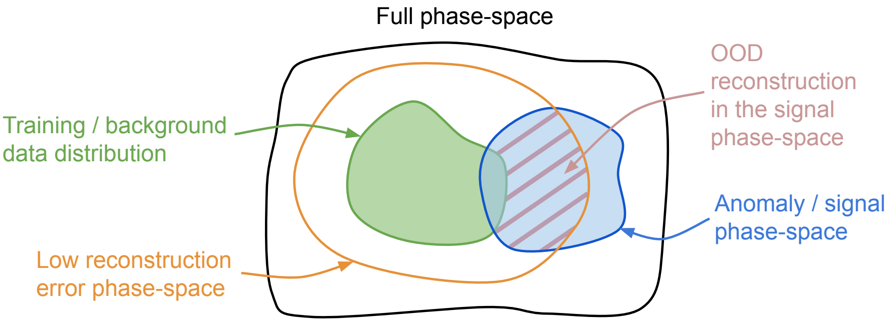
Full phase-space
OOD reconstruction in the signal phase-space
Training / background data distribution
Anomaly / signal phase-space
Low reconstruction error phase-space

Figure 2: Illustration of out-of-distribution reconstruction.

4. Normalized autoencoders

4.1 Theoretical description

Normalized autoencoders [4] suppress OOD reconstruction by learning the training data probability distribution p_{data}. The NAE model probability p_θ is defined using the reconstruction error E_θ. It assigns high probability to low reconstruction error examples:

$$p_\theta(x) = \frac{1}{\Omega_\theta} \exp\left(-E_\theta(x)\right)$$

where Ω_θ is a normalization constant ensuring that the sum of all probabilities is 1. Examples following p_θ are obtained by sampling via a Langevin Markov Chain Monte Carlo (MCMC) ("negative examples"). The loss function proposed in Ref. [4] is the difference between the reconstruction error of the training ("positive") examples and of the negative examples, aiming to achieve $p_\theta = p_{\text{data}}$:

$$\mathbb{E}_{x\sim p_{\text{data}}}\left[L_\theta(x)\right] = \underbrace{\mathbb{E}_{x\sim p_{\text{data}}}\left[E_\theta(x)\right]}_{\text{positive energy } E_+} - \underbrace{\mathbb{E}_{x'\sim p_\theta}\left[E_\theta(x')\right]}_{\text{negative energy } E_-}$$

The loss function proposed in this work prevents the divergence of negative energy and minimizes the positive energy while the energy difference is close to 0:

$$L = \log\left(\cosh\left(E_+ - E_-\right)\right) + \alpha E_+ \qquad \alpha > 0, \text{ hyper-parameter}$$

4.2 Application to semivisible jet tagging

Ten independent NAEs were trained on the same top-quark jets dataset. In order to measure the distance between positive and negative samples, the Energy Mover's Distance (EMD) was computed on the validation set between the positive and negative samples in the input feature space. Figure 3 shows reconstruction errors, EMD and AUC for a representative training.

The different loss function efficiently prevents OOD reconstruction as, after the positive and negative energies are equal, the reconstruction error of the signal is stable and higher than that of the background. As the positive energy is minimized beyond a certain value, the EMD increases:

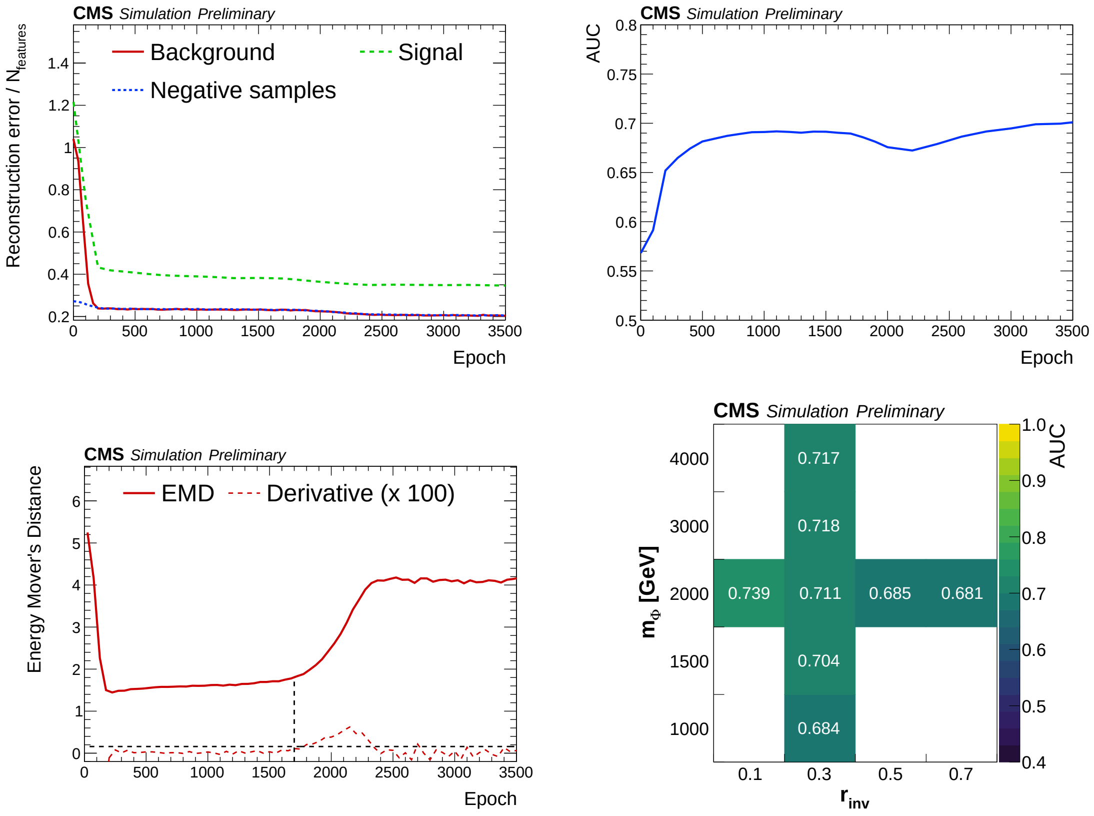

Figure 3: The background, signal (m_Φ = 2 TeV, r_{inv} = 0.3) and negative samples reconstruction errors (top left), the average AUC (top right) computed over an ensemble of signal hypotheses (the same as in the bottom right plot), and the EMD, calculated on the validation set between negative and positive samples (bottom left), as a function of the epoch number, for one representative training. Average AUC for all 10 independent NAE trainings at the best epoch (bottom right), corresponding to the epoch number before increase of the EMD, indicated by black dashed lines in the EMD plot.

the network cannot better reconstruct training examples and suppress OOD reconstruction at the same time. During and after the EMD increase, the AUC first decreases and then increases, because the low reconstruction error phase-space first overlaps more with signal-rich and then signal-depleted phase-spaces. Using the AUC to define the best epoch would therefore be biased towards the signals utilized for its calculation. The best epoch is thus just before the EMD increase: minimal OOD reconstruction and maximal training examples reconstruction. This is a fully signal-agnostic procedure to train a NAE, not using signal SVJs simulation. The NAE achieves sensible improvement in performance compared to the standard AE, as shown in the AUC table in Fig. 3.

5. Conclusion

This study demonstrated the tendency of AEs to generalize beyond the training dataset, resulting in drastically reduced anomaly detection performance, and how NAEs overcome this shortcoming. In addition, a fully signal-agnostic training definition to reach the optimal performance was provided.

References

[1] Planck collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [1807.06209].

[2] CMS collaboration, *The CMS Experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[3] B.M. Dillon, L. Favaro, T. Plehn, P. Sorrenson and M. Krämer, *A Normalized Autoencoder for LHC Triggers*, 2206.14225.

[4] S. Yoon, Y.-K. Noh and F. Park, *Autoencoding under normalization constraints*, in *ICML*, pp. 12087–12097, PMLR, 2021, https://arxiv.org/abs/2105.05735 [2105.05735].

[5] T. Cohen, M. Lisanti and H.K. Lou, *Semivisible Jets: Dark Matter Undercover at the LHC*, *Phys. Rev. Lett.* **115** (2015) 171804 [1503.00009].

[6] T. Heimel, G. Kasieczka, T. Plehn and J.M. Thompson, *QCD or What?*, *SciPost Phys.* **6** (2019) 030 [1808.08979].

[7] M. Farina, Y. Nakai and D. Shih, *Searching for New Physics with Deep Autoencoders*, *Phys. Rev. D* **101** (2020) 075021 [1808.08992].

[8] F. Canelli, A. de Cosa, L.L. Pottier, J. Niedziela, K. Pedro and M. Pierini, *Autoencoders for semivisible jet detection*, *JHEP* **02** (2022) 074 [2112.02864].

[9] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer 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 [1405.0301].

[10] T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [1410.3012].

[11] GEANT4 collaboration, *GEANT4–a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250.

[12] CMS collaboration, *Signal-agnostic Optimization of Normalized Autoencoders for Model Independent Searches*, 2023, https://cds.cern.ch/record/2871591.