Anomalous Jet Identification via Variational Recurrent Neural Network

Authors: Alan Kahn, Julia Gonski, Inês Ochoa*, Daniel Williams, Gustaaf Brooijmans
Nevis Laboratories, Columbia University, 136 S Broadway, Irvington NY 10533
*Laboratory of Instrumentation and Experimental Particle Physics, Lisbon, Portual

0.1 Method

The method described here employs a Variational Recurrent Neural Network (VRNN) to model jets as a sequence of constituents. A VRNN is a sequence-modeling architecture which replaces the standard encoder-decoder architecture of a Recurrent Neural Network with a Variational Autoencoder (VAE) [1]. This allows the VRNN to perform both sequence modeling in addition to variational inference, which has been shown to be a very powerful tool for anomaly detection [2]. A sequence-modeling architecture is well-motivated as it is capable of accommodating variable-length inputs, such as lists of constituent four-vectors in a jet, while suppressing the ability of the model to learn correlations with the jet's constituent multiplicity. By contrast, fixed-length architectures such as VAEs rely on a loss function that is computed between the input layer and the reconstructed layer. As a result, zero-padded inputs directly affect the value of the loss function, leading to correlations that are difficult to remove when using inputs that are naturally variable in length, but forced to work in a fixed-length framework.

Figure 1 shows a diagram of one VRNN cell. The VAE portion of the architecture is displayed on the top row of layers in the diagram, where a constituent's four-momentum components are input as a vector $x(t)$, which is encoded into a multivariate Gaussian distribution in the latent space z, and then decoded to produce a reconstruction of the same input constituent's components $y(t)$. The variable t refers to the *time-step*, which advances as the sequence is processed, and can be interpreted as the constituent number currently being processed by the model.

Inputs to the VRNN consist of sequences of jet four-vector constituent components p_T, η, and ϕ, where constituents are assumed to be massless. Jets are reconstructed using the anti-k_t algorithm with a radius parameter of 1.0 [3]. Before training, a pre-processing method is applied which boosts each jet to the same reference mass, energy, and orientation in $\eta - \phi$ space, such that all input jets differ only by their substructure. In addition, our pre-processing method includes a choice of *sequence ordering*, in which the constituent sequence input into the model is sorted by k_t-distance instead of by the typical constituent p_T. In more detail, the nth constituent in the list, c_n, is determined to be the constituent with the highest k_t-distance relative to the previous constituent, with the first constituent in the list being the highest p_T constituent.

$$c_n = max(p_{Tn}\Delta R_{n,n-1}) \tag{0.1}$$

This ordering is chosen such that non-QCD-like substructure, characterized by two or more separate prongs of constituents within in the jet, is more easily characterized by the sequence. When compared to p_T-sorted constituent ordering, the k_t-sorted sequence consistently travels between each prong, making their existence readily apparent and easy to model. As a result, a significant boost in performance is observed.

The loss function for each constituent, $\mathcal{L}(t)$, is very similar to that of an ordinary VAE. It consists of a mean-squared-error (MSE) loss between input constituents and generated output constituents as a reconstruction loss, as well as a weighted KL-Divergence from the latent space prior to the learned approximate posterior distribution. Since softer constituents contribute less to the overall classification of jet substructure, each KL-Divergence term, computed constituent-wise, is weighted by the constituent's p_T-fraction with respect to the jet's total p_T, averaged over all jets in the dataset to avoid correlations with constituent multiplicity. The weight coefficient of the KL-Divergence term is enforced as a hyperparameter, and has been optimized to a value of 0.1 in dedicated studies.

$$\mathcal{L}(t) = MSE + 0.1 \times \overline{p_T}(t) D_{KL} \tag{0.2}$$

After a jet is fully processed by the VRNN, a total loss function is computed as the average of the individual constituent losses over the jet: $\mathcal{L} = \frac{\Sigma \mathcal{L}(t)}{N}$.

The architecture is built with 16 dimensional hidden layers, including the hidden state, with a two-dimensional latent space. All hyperparameters used are determined by a hyperparameter optimization scan.

The model is trained on the leading and sub-leading jets of each event, where events are taken from the LHC Olympics datasets. After training, each jet in the dataset is assigned an *Anomaly Score*, defined in Equation 0.3, where D_{KL} is the KL-Divergence between the encoded posterior distribution and learned prior distribution.

$$\text{Anomaly Score} = 1 - e^{-\overline{D_{KL}}} \tag{0.3}$$

Since the LHC Olympics challenge entails searching for a signal on the event-level instead of the jet-level, an overall *Event Score* is determined by choosing the most anomalous score between the leading and sub-leading jets in an event. To ensure consistency between training scenarios, Event Scores are subject to a transformation in which the mean of the resulting distribution is set to a value of 0.5, and Event Scores closer to 1 correspond to more anomalous events.

0.2 Results on LHC Olympics

The performance of the VRNN was first assessed with the LHC Olympics R&D dataset, which includes a known signal of a beyond the Standard Model W' boson with a mass of 3500GeV which decays to two hadronically decaying X and Y particles, each reconstructed by a R=1.0 jet. This study was used as a validation of the method, with a goal of directly investigating
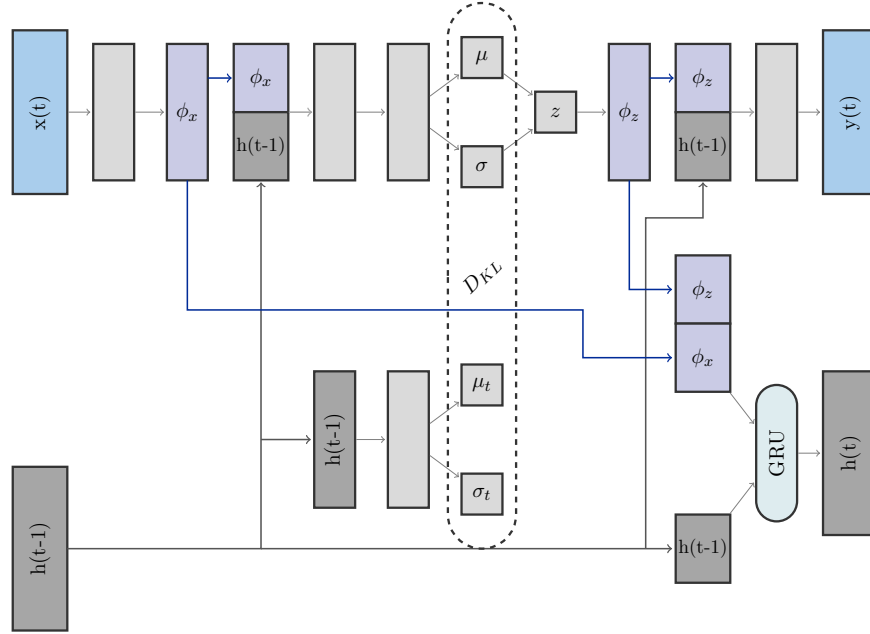
Figure 1. A Variational Recurrent Neural Network cell. The $x(t)$ and $y(t)$ layers represent respectively the input constituent and reconstructed constituents' four-momentum components p_T, η, and ϕ. The ϕ_x and ϕ_z layers are *feature-extracting layers* which encode a representation of the features in the input layer $x(t)$ and latent space z respectively. $h(t-1)$ represents the current time-step's hidden state, which is updated each iteration via a transition function between $h(t-1)$, ϕ_x, and ϕ_z carried out by a Gated Recurrent Unit (GRU). At each time-step, the prior distribution defined by μ_t and σ_t is determined from the current hidden state

the ability of the Event Score selection to reconstruct the W' mass. Therefore, no selections beyond those described in Section 0.1 are applied.

The VRNN was trained over a contaminated dataset consisting of 895113 background events and 4498 signal events, corresponding to a signal contamination level of 0.5%. A selection on the Event Score is applied as the sole discriminator, and the invariant mass m_{JJ} of the two jets is then scanned to assess the prominence of the W' mass peak. In this validation analysis, the Event Score is required to exceed a value of 0.65. This value is chosen to significantly enrich the fraction of anomalous jet events over the background, while retaining enough statistics in the background to display its smoothly falling behavior.

Figure 2 shows the dijet invariant mass distributions before and after the Event Score selection, along with the local significance of the signal computed in each bin using the BINOMIALEXPZ function from ROOSTATS with a relative background uncertainty of 15% [4]. Applying this selection dramatically increases the significance of the excess from 0.18σ to 2.2σ at a signal contamination of 0.5% without significantly sculpting the shape of the background.
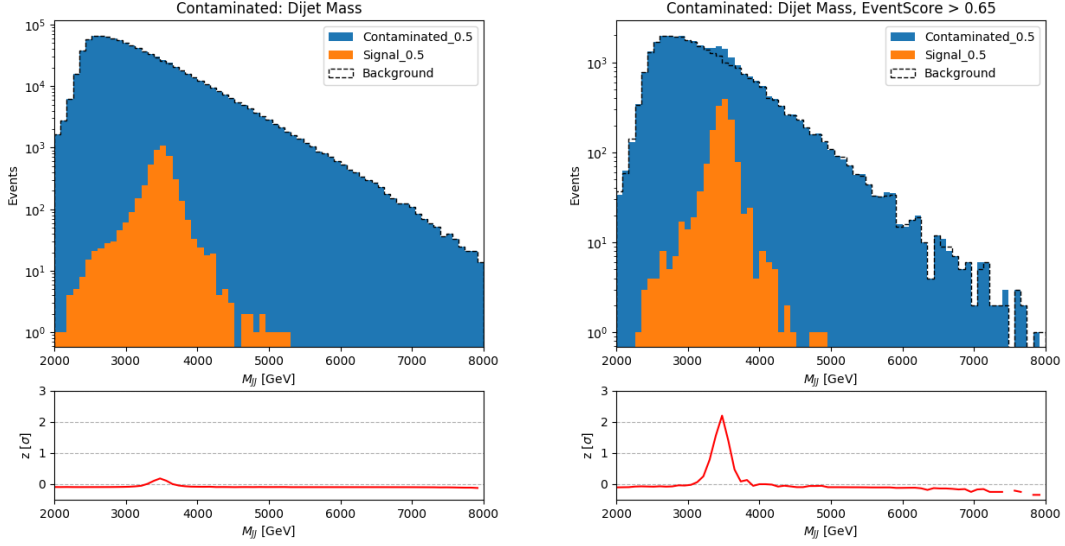
Figure 2. Dijet invariant mass distributions before (left) and after (right) a selection on the Event Score, with a two-prong W' signal contamination of 0.5%.

Once the method was validated in the R&D dataset, it was applied to Black Box 1, with a re-optimized tighter selection on the Event Score of 0.75, as well as a requirement on the pseudorapidity of the leading and sub-leading jets to be less than 0.75, to ensure that central, high momentum transfer events are considered. Figure 3 shows the dijet invariant mass for both Black Box 1 and Background datasets. The Event Score selection reveals an enhancement in m_{JJ} just below 4000 GeV. This is consistent with the Black Box 1 signal, which is a new Z' boson with a mass of 3800 GeV decaying to two new particles, each decaying hadronically.

The same method applied to Black Box 2, shown in Figure 4, results in no significant excesses in the invariant mass distribution. Additionally, the effect of the Event Score selection on the m_{JJ} shapes is similar between Black Box 2 and Background datasets. Black Box 2 does not contain any beyond the Standard Model events, and therefore these results are consistent with a QCD-only sample. It is important to note that the model was trained independently on each dataset, and the resulting Event Scores are from entirely unique sets of network weights.

Figure 5 shows results for Black Box 3. The signal in Black Box 3 consists of a new 4200 GeV particle, with varied final states beyond the two-prong large-R jets described earlier. As the model described here is specifically sensitive to substructure within a large-R jet, it is insensitive to the signal present in this Black Box.
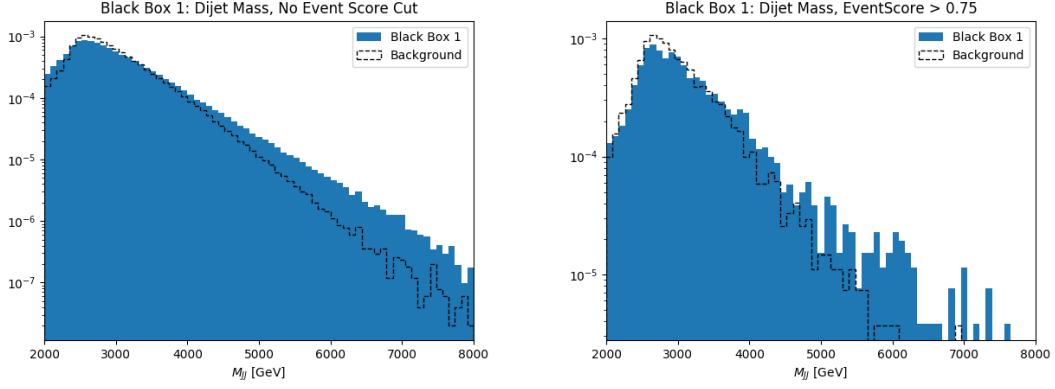
Figure 3. Dijet invariant mass distributions before (left) and after (right) a selection on the Event Score from the Black Box 1 dataset. The signal present is a Z' boson with a mass of 3800 GeV.
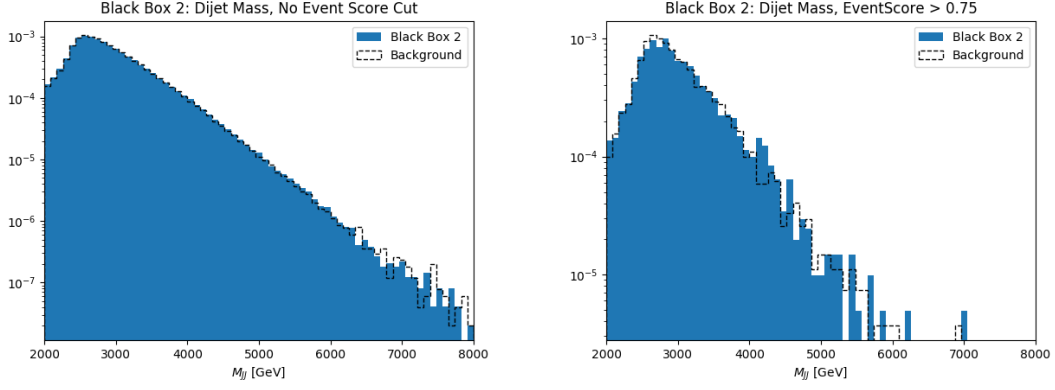

Figure 4. Dijet invariant mass distributions before (left) and after (right) a selection on the Event Score from the Black Box 2 dataset. No signal is present, and the dataset shown consists entirely of multijet background events.
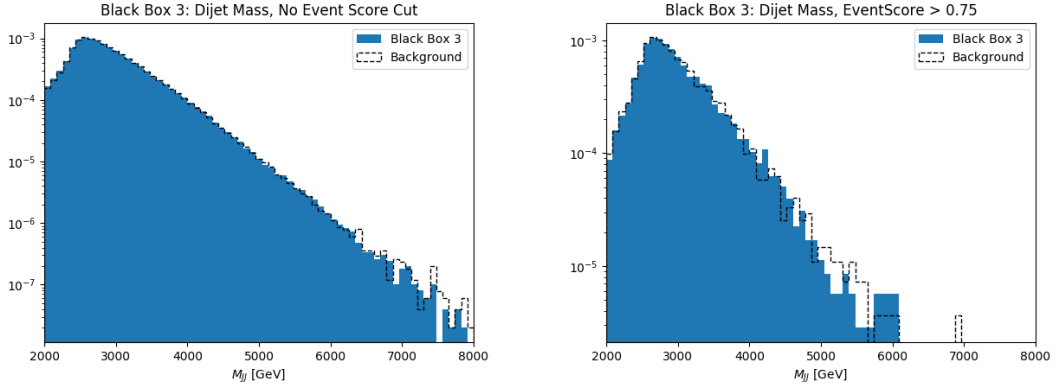

Figure 5. Dijet invariant mass distributions before (left) and after (right) a selection on the Event Score from the Black Box 3 dataset. The signal present is a new boson with a mass of 4200 GeV.

0.3 Lessons Learned

This challenge presented a highly useful avenue for development of our model. Results from the R&D and Black Box dataset analyses indicate that the VRNN is capable of identifying anomalous objects within a contaminated dataset as long as they can be characterized by sequential data. We learned that the pre-processing method is hugely influential on the performance of the model, in particular the choice of k_t-ordered sequencing. We feel that this is a generalizable conclusion from our study which can be applied to the understanding and use of jet substructure in future analyses.

Given these lessons, there are a variety of future opportunities with this application of the VRNN architecture to jet-level anomaly detection. Since the VRNN takes constituent information as input and learns jet substructure without explicit reliance on high level variables, it is expected to have less correlation with jet mass than standard substructure variables such as N-subjettiness. Further characterization of this point could reveal a key advantage in using such an approach in an analysis context. While we limited our scope in this study to be entirely unsupervised with no signal or background model information, the RNN and VAE elements of the VRNN give potential for accommodating more supervised training scenarios. Furthermore, a number of advancements to the architecture, such as a dedicated adversarial mass de-correlation network, or an additional input layer representing high-level features, are worthwhile avenues of exploration to enhance performance while minimizing unwanted correlations.

Acknowledgments

References

[1] J. Chung, K. Kastner, L. Dinh, K. Goel, A. Courville, and Y. Bengio, *A recurrent latent variable model for sequential data*, 2016.

[2] J. An and S. Cho, *Variational autoencoder based anomaly detection using reconstruction probability*, 2015.

[3] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-ktjet clustering algorithm*, Journal of High Energy Physics **2008** (Apr, 2008) 063–063.

[4] L. Moneta, K. Belasco, K. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf, *The roostats project*, 2011.