Published for SISSA by Springer

Received: March 21, 2021
Accepted: May 24, 2021
Published: June 4, 2021

Quasi anomalous knowledge: searching for new physics with embedded knowledge

Sang Eon Park,[a] **Dylan Rankin,**[a] **Silviu-Marian Udrescu,**[a] **Mikaeel Yunus**[a] **and Philip Harris**[a,b]

[a] *Laboratory for Nuclear Science, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, U.S.A.*

[b] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*

E-mail: sangeon@mit.edu, drankin@mit.edu, sudrescu@mit.edu, myunus@mit.edu, pcharris@mit.edu

Abstract: Discoveries of new phenomena often involve a dedicated search for a hypothetical physics signature. Recently, novel deep learning techniques have emerged for anomaly detection in the absence of a signal prior. However, by ignoring signal priors, the sensitivity of these approaches is significantly reduced. We present a new strategy dubbed Quasi Anomalous Knowledge (QUAK), whereby we introduce alternative signal priors that capture some of the salient features of new physics signatures, allowing for the recovery of sensitivity even when the alternative signal is incorrect. This approach can be applied to a broad range of physics models and neural network architectures. In this paper, we apply QUAK to anomaly detection of new physics events at the CERN Large Hadron Collider utilizing variational autoencoders with normalizing flow.

Keywords: Beyond Standard Model, Exotics, Jet substructure, Hadron-Hadron scattering (experiments), Jets

ArXiv ePrint: 2011.03550

https://doi.org/10.1007/JHEP06(2021)030

JHEP06(2021)030

Contents

JHEP06(2021)030

1 Introduction

With no evidence for new physical phenomena, many physicists at the CERN Large Hadron Collider (LHC) are asking themselves a critical question: *Am I searching for new physics in the right way?* Despite ten years of exhaustive research at the LHC, a rapidly growing cohort is becoming concerned that we have somehow missed a new fundamental physics discovery. Within particle physics and beyond, the identification of new physical phenomena has often been unexpected. With advances in deep learning (DL), a series of new approaches can improve the search for anomalous signatures. In this paper, we will present a new deep-learning-based anomaly search algorithm. This algorithm is broadly applicable to many fields of physics. Recent DL-based anomaly detection within high energy physics has largely focused on searching for anomalous signatures in the complete absence of a signal prior. In this scenario, two fundamental approaches have been considered:

- Isolate two distinct datasets that contain signal and background with different proportions, then try to find a deviation between them. [8–11]
- Embed our knowledge of known physics processes into simulation or a DL algorithm, such as an autoencoder, and then look for events with a low likelihood of being a known physics process. [12–16, 16–22]

In the first approach, colloquially referred to as classification without labels (CWoLA), conventional discrimination algorithms are used to separate the two datasets [8–11]. Care must be taken to ensure that selection biases are mitigated so that the only discernible difference within the discrimination algorithm is the presence of an unknown physics signal. The second approach attempts to embed into a likelihood discriminant a complete knowledge of physics processes within a selected region. An excess of events with low likelihood constitutes a new physics signature. This second method broadly comprises models that utilize deep learning autoencoders. However, when using large, high dimensional datasets, complete knowledge of all expected physical processes can become quite complicated. It can lead to reduced sensitivity [12–16, 16–21]. Recently, hybrid approaches have started to emerge, which aim to utilize aspects of both methods [23].

When comparing the two approaches, the CWoLA approach is often more sensitive, provided a signal-enriched region is present [11]. This enlarged sensitivity results from the implicit assumption on the signal properties; the signal is localized within a specific kinematic region. In other words, CWoLA assumes that we can find a signal enriched region. Signal priors frequently lead to enhanced sensitivity since they minimize the possibilities that must be explored when attempting to search for an anomaly. For many new physics models within HEP, several fundamental assumptions can be applied to all potential signals without loss in generality. However, these assumptions are often not embedded with a neural network aimed at anomaly detection. Thus, the network cannot infer whether an observed anomaly within the data violates fundamental symmetries of nature required for a new physics model. For example, when a massive particle decays, its decay products fall within a cone determined by the particle's energy and Lorentz invariance. When a generic DL algorithm attempts to probe data, it has no knowledge of Lorentz invariance [24–26]. By relying on *one* anomaly metric that measures any deviation, whether it be physical or not, we may miss the chance to apply fundamental physical laws about how new physics may appear, thus wasting our prior physics knowledge. If we can incorporate prior knowledge into the search, it should be possible to either improve the search's sensitivity or, at worst, restrict the network complexity. Within physics, several ideas have emerged to incorporate physical symmetries into the neural network. These ideas involve modifications to the network architecture so that networks implicitly respect physical laws. In this paper, we consider an alternative approach. Instead of modifying the network, we rely on extending the anomaly algorithm by exploiting a second self-learned space explicitly targeting a class of new physics signatures, i.e. an added signal prior. This space embeds the most critical symmetries respected by new physics signature through self-learning, thus allowing for detection of all anomalies that are broadly similar to the desired signal.

With our modified anomaly detection algorithm, we extend the use of signal priors to anomaly searches by developing a mechanism to add signal priors without degrading the sensitivity of a pre-existing model-independent search. Through our approach, signal priors, which may or may not be accurate signal descriptions, can be embedded within an anomaly search. Since priors are systematically added to construct information, we refer to this technique as Quasi-Anomalous Knowledge, or simply QUAK.

JHEP06(2021)030

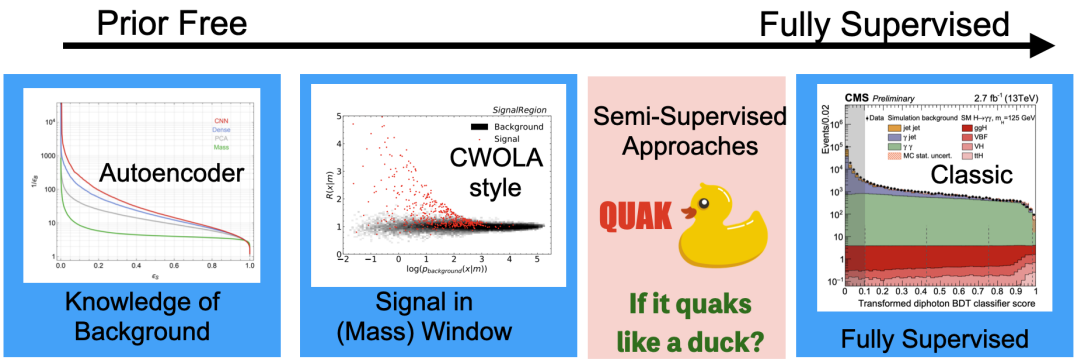

Figure 1. Overview of Deep Learning based methods used in high energy physics. The prior here indicates the amount of a signal prior used within the search. This paper presents QUAK, which falls into a region labeled by the red shaded area.

To understand the context of QUAK within the scope of search strategies at the LHC, consider figure 1. Currently, most searches for new physics at the LHC involve the search for a specific, well-motivated physics signature (rightmost region of figure 1). This is performed by considering a specific signal prior and optimizing selection towards this signal prior. A dedicated analysis is performed with each individual search to account for mismodeling and systematic effects.

The critical aspect of this type of search is that one constructs the selection, discriminator, and sensitivity estimate under a prior for what the signal model is assumed look like. As an example, consider a search for black holes. First, we hypothesize what a black hole signature would look like. From our hypothesis, we construct a Monte Carlo simulation of black hole signal events. We proceed to build an analysis around this assumption. Barring a discovery, our final search will give us bounds on the production of black hole processes under these assumptions. It may not give us a bound on any other process and it might not even cover all possible black hole signatures. However, it will have shown that, in the region of collision data where events are black hole like, we observe good modeling of predicted background processes with the data. While the choice of a specific signal signature is restrictive, it has the benefit that, for physics models that predict similar signatures to our signal, we can make powerful constraints.

With QUAK, we aim to embed this choice of a signal prior within a more generic search for an anomaly. Anomaly detection algorithms typically forego the signal prior, and, as a consequence, rely on some metric to determine what an anomaly is. With QUAK, we aim to create a space that allows us to interpolate from a dedicated search with a clear signal prior to a prior-free search that quantifies physical anomalies and implicitly emphasizes signal-like features. As a consequence, QUAK can be considered a semi-supervised approach.

In the following sections, we will introduce the QUAK algorithm. First, to show the generality of this approach, we demonstrate the usage of QUAK on the MNIST dataset [27]. Then, we present this work in the context of the LHC Olympics 2020 anomalous dataset [28].

JHEP06(2021)030

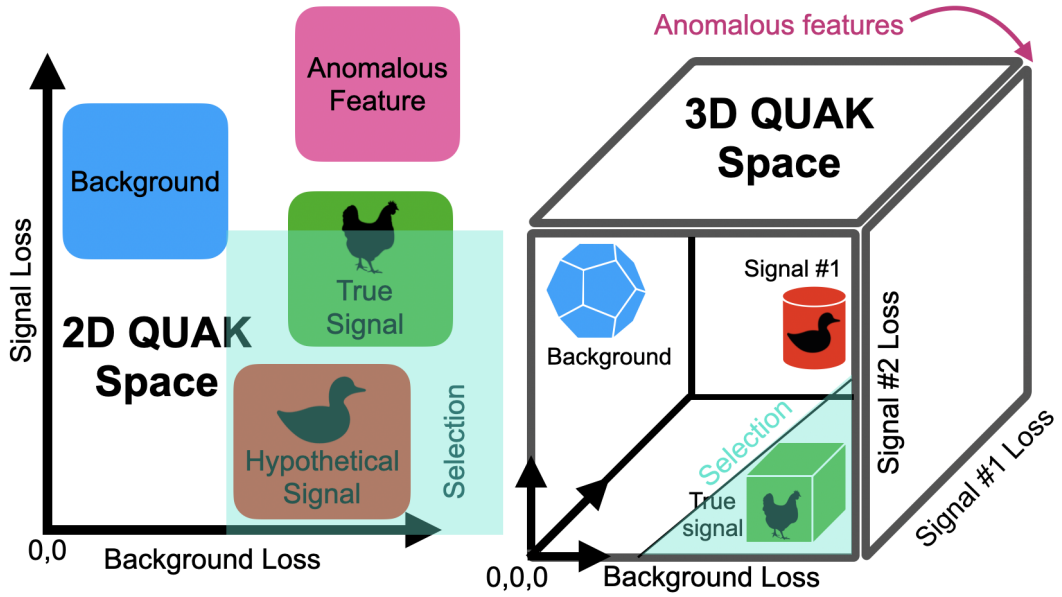

Figure 2. Illustration of the 2D QUAK space construction (Left) and the 3D QUAK space construction (Right). In each space, we list where background events would roughly be located (blue), signal events would roughly be located (red and green), and anomalous unphysical features would be located(magenta). We split the signal events between hypothetical, guessed signal, used to construct the space (red), and the location of a potential true signal(green). The selected region (cyan) indicates roughly where anomalous new physics signals are most likely to appear.

2 Concept

New models have emerged within the deep learning community that utilize semi-supervised learning to construct a physical, self-organized space [29, 30]. Semi-supervision applied to deep learning works by training on both labeled and unlabeled data. With the unlabeled data, an unsupervised network is trained (autoencoder). With the labeled data, a supervised classification is applied using the unsupervised network's intermediate latent space. The trainings are often done simultaneously. Other approaches can be considered semi-supervised learning, such as using analytic continuous to fill poorly understood spacer.

By performing the unsupervised training, the network constructs a latent space of self-organized patterns. By performing the supervised training, we label regions within the latent space with the labeled data. Consequently, objects present in the unlabeled data but not in the labeled data will be labeled by a superposition of its nearest labeled identifiers. Additionally, the construction of latent space from the unsupervised training using a variational approach by sampling Gaussians within the latent space can ensure that the latent space is continuous.

Semi-supervised constructions represent a different approach to training neural networks when compared with supervised and unsupervised. Semi-supervised networks are very robust to variations in the data, and, in some cases, these networks have been found to exceed the performance of supervised networks [31]. Within the context of anomaly

JHEP06(2021)030

detection, semi-supervision has been found to be effective for anomaly detection [32, 33], even very recently within physics [34]. QUAK builds on the concept of semi-supervision. However, the algorithm differs from other semi-supervised approaches in that we rely more on unsupervised networks when constructing our algorithm. To construct QUAK we:

- choose a set of N samples that reflect the anomaly search; this is typically a background and a set of signals, with the signals that capture the physical features of a potential new physics signature,
- train a separate unsupervised networks on each signal or background sample, yielding N unsupervised networks,
- construct an N-dimensional "QUAK" space consisting of the loss for each unsupervised network, and
- search within bins of QUAK space for anomalous signals.

The construction is semi-supervised in that we use the signal priors as labels for QUAK space construction. Figure 2 illustrates the concept of QUAK space. In each axis, we plot the loss of the unsupervised network trained on a specific sample. For a 1-dimensional QUAK where we train just on the background sample, QUAK reduces to a typical autoencoder based anomaly search. However, the advantage of QUAK is that the added signal dimensions help to disambiguate anomalies that are signal like from anomalies that are just anomalous, such as an anomalous event resulting from a detector glitch.

Within QUAK space, a background event will have low background loss and high signal loss. In the 2D QUAK plot, this corresponds to the top left of the region. An anomalous event with features similar to the chosen signal will have a low signal loss and large background loss. This will occur in the bottom right region of 2D QUAK space. An anomalous event different from both the chosen signal, and the background will have large signal and background loss. This region will include events coming from detector glitches. Extending this beyond 2 dimensions, we can add additional, different, signals. These additional signals can further improve the anomaly search since they help to disambiguate anomalies with features similar to the chosen signal. By searching in the region of large background loss and low signal loss, on all signal dimensions, we isolate a region of anomalies that have physical, signal-like features. This space greatly enhances the ability to find new physics signatures provided they have a sufficient set of features to be captured by the anomaly algorithm.

For many classes of new physics models at the LHC, there is a broad set of underlying physics features that can be assumed about any new signal (e.g. Lorentz invariance, lack of QCD color-flow with the event). These assumptions can restrict an anomaly search by highlighting, at low loss, features characteristic of a signal. With QUAK, we effectively embed these assumptions into our search while preserving much of the model-independence of the investigation.

For the network architecture of the unsupervised networks in QUAK, we utilize variational auto-encoders (VAEs) and normalizing flow VAEs. VAEs have been found to give a

JHEP06(2021)030

continuous latent space, which allows for the capture of physical properties within the latent space. Normalizing flows allow for the latent space to be irregular and not necessarily Gaussian. While the exact deep learning architecture is not critical, we focus on normalizing flow since it has been shown as a powerful tool for representing physical models [11, 35–45]. In the QUAK construction, our signal choices can be thought of as "approximate-priors" since they help direct the space of searches towards signals with similar features. This, in turn, leads to a potential model dependence in the anomaly search. However, as we will see with QUAK, the signal choice can significantly differ from the observed anomaly and give an enhanced sensitivity compared with other signal-less anomaly search algorithms.

Finally, we note that while we consider QUAK to be a semi-supervised network, QUAK deviates from other semi-supervised networks. We do not exploit a common latent space between the supervised and unsupervised component of the network.

3 Normalizing flows

To construct QUAK, we rely on normalizing-flow variational autoencoders (NF-VAEs). A variational autoencoder (VAE) is, in essence, an autoencoder that samples from a multidimensional Gaussian distribution in the first layer of the latent space. The loss is computed by constructing the cross-entropy between the input dataset and the output dataset. An additional Kullback-Leibler (KL) divergence term is added to the loss to encourage the VAE to approximate the posterior with a multidimensional Gaussian distribution [46].

VAEs have several limitations. With a VAE, we assume that the latent space can be approximated using a series of compounded linear transformations of a multidimensional Gaussian distribution. Consequently, VAEs work best when the input variables are approximately Gaussian. If the input variables adhere to a distribution that is clearly skewed or multimodal, we require a large decoder to transform the Gaussian latent vectors to vectors that can accurately reconstruct the input data. If the decoder is too large, the latent space will be rendered useless. This phenomenon is known as the "posterior collapse" [47].

To avoid the limitations of VAEs, one can use normalizing flows to transform a posterior distribution into a much more flexible distribution that is representative of the corresponding data [48]. We incorporate normalizing flows into our VAEs by using them in our latent space. Specifically, if z_0 is the latent vector obtained via Gaussian sampling, and $z_1, \cdots, z_k$ are the latent vectors that follow, a series of normalizing flows $f_1, \cdots, f_k$ will translate the posterior of the latent space as follows:

$$z_k = f_k \circ \cdots \circ f_1(z_0) \tag{3.1}$$

Furthermore, under the assumption that z_i belongs to a distribution $q_i(z_i)$ for all $i \in \{0, \cdots, k\}$, we can write

$$\log q_k(z_k) = \log q_0(z_0) - \sum_{i=1}^{k} \log \left| \det \left(\frac{\partial f_i}{\partial z_{i-1}} \right) \right| \tag{3.2}$$

Consequently, we have effectively transformed our space from the assumed Gaussian space utilized within the VAE. By transforming our Gaussian posterior (i.e. the latent

JHEP06(2021)030

space) into a more complex posterior, we can accurately account for jet variables whose underlying distributions are not necessarily Gaussian. This removes the need for a larger decoder, which inherently prevents a "posterior collapse" from occurring. Furthermore, we can discern from equation (3.2) that we must update our loss function to account for the introduction of a more expressive posterior z_k . This we perform by modifying the loss as follows:

$$\mathcal{L} = \mathcal{L}_{\text{reco}} + \mathcal{D}_{KL} - \sum_{i=1}^{k} \log \left| \det \left(\frac{\partial f_i}{\partial z_{i-1}} \right) \right|, \tag{3.3}$$

where $\mathcal{L}_{\text{reco}}$ corresponds to the conventional autoencoder loss, and $\mathcal{D}_{KL}$ corresponds to the KL divergence term. Our new loss function applies to a much wider range of posterior distributions. In order to determine the optimal normalizing flow algorithm, a scan of various normalizing flow algorithms was performed. With each, we required that $\det \left(\frac{\partial f_i}{\partial z_{i-1}} \right)$ be computed in linear time, along with the requirement that $f_1, \cdots, f_k$ be invertible. For this scan, we considered planar flows, radial flows, non-volume preserving flows, and masked autoregressive flow. We found that masked autoregressive flow gave the best performance for anomaly search. With masked autoregressive flow, we have that

$$f(x_i) = u_i \exp f_{\alpha_i}(z_{1-i-1}) + f_{\mu_i}(z_{1:i-1}) \text{ with,} \tag{3.4}$$

$$\left| \det \left(\frac{\partial f}{\partial z} \right) \right| = \exp \left(- \sum_i f_{\alpha_i} \right) \tag{3.5}$$

where u_i is a randomly sampled number over a Gaussian distribution of width 1, and the functions f_α and f_μ are applied on the previous 1 through i-th values of the input latent space. In this approach, the series of iterations taking the previous inputs z_i are used to progressively improve the ability of the model to construct a probabilistic distribution of the internal space used within the autoencoder. The choice of generating function, and the auto-regressive construction enable a broad range of distributions. Furthermore, the masking provides a scheme for efficient sampling and generation of events.

More advanced normalizing flow techniques have recently been developed by various researchers for use within physics [35, 36, 38, 49]. Exploration of these models is an interesting and exciting avenue for future work.

4 Results

To demonstrate the effectiveness of QUAK. We present two examples of how it can be applied to anomaly searches. First as a demonstration, we apply QUAK on a test dataset constructed from the MNIST dataset [27]. Then, we consider applying QUAK to the LHC Olympics 2020 dataset to find anomalous dijet signals [28].

4.1 MNIST

As a first example and to show that QUAK approach can broadly apply to many scenarios, we demonstrate the usage of QUAK on the MNIST dataset [27]. The MNIST dataset consists of a set of images of handwritten single digits ranging from 0 to 9, with explicit

JHEP06(2021)030

labels for each of the digits corresponding to their actual number. Standard VAEs without a normalizing flow layer are very effective at describing the MNIST dataset.

To emulate the hypothetical possibility of discovery, we split the MNIST dataset into "known" and "unknown" digits. For this example, we will choose the digit 9 as unknown. The dataset is split so that the digits 0 to 8 are known, but the digit 9 is not present anywhere in the training datasets. We then attempt to discover this digit by constructing a QUAK space classifier.

We construct QUAK space on MNIST through a variational autoencoder [46] with three dense layers on either end; we do not use a normalizing flow layer since we found it was not needed. The training for the VAE uses a loss function consisting of a binary cross-entropy of the input image with the output image. A KL divergence term with equal weight to that of the cross-entropy loss term is added to the loss function to constrain the sampled Gaussian means and widths to be close to unity. The QUAK loss is obtained by removing the KL divergence term. Figure 5 shows the deep learning architecture used.

4.1.1 Separation of digits with QUAK

To test QUAK on MNIST, we separately train a dedicated VAE for each of N known digits, yielding an N-dimensional QUAK space ranging from the digits 0 to 8. Within this space, we try to find 9 by choosing a proxy signal, 7, which is similar to the true signal digit. The choice of 7 is intended to be an educated guess for how to find a signal. If we did not know the digit beforehand, we could systematically scan each digit, 0–8, and use it as a proxy signal to search for an anomaly. Additionally, the use of a proxy signal allows us to deal with high dimensional QUAK spaces since we can compress an N-dimensional QUAK space into a single dimension through the training of an additional dense network aimed at separating the proxy signal from the background.

Figure 3 demonstrates the performance of various QUAK spaces in the separation of the digit 5 from the digit 9. In the language of physics searches, the digit 5 is the background and the digit 9 is the signal. We consider 3 different QUAK approaches:

- Scan of 2D QUAK space using 5 for L_{bkg} and 7 for L_{sig} and training a 3 layer dense network on the 2D QUAK space with 7 as the target signal and 5 as the background
- Scan of the 9D QUAK space using the VAE loss of all digits 0–8, and training a 3 layer dense network on the 9D QUAK space with 7 as the target signal and 5 as the background
- Scan of the 9D QUAK space using the VAE loss on all digits 0–8, and transforming the 9D space to a 1D discriminator using a linear discriminate trained on the 9D QUAK space with 7 as the target signal and 5 as the background

We contrast these approaches using a single autoencoder and a fully supervised network both trained to separate the digit 5 from 9.

We find near optimal performance when using a linear discriminant with a proxy signal to separate signal and background. We observe the QUAK methods approaches the fully

JHEP06(2021)030

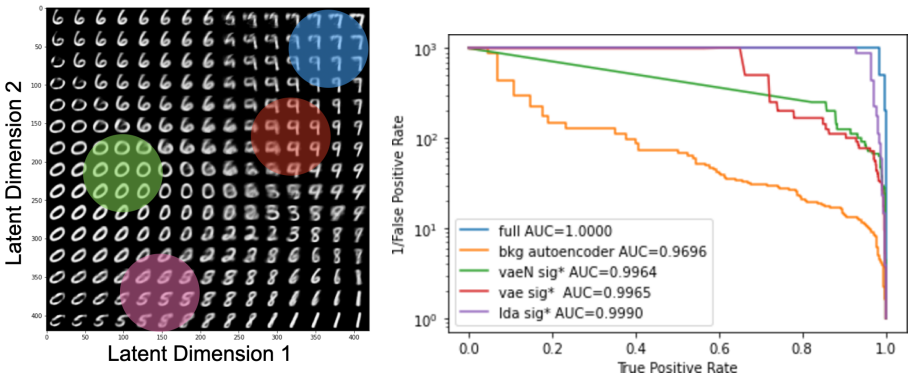

Figure 3. (Left) Illustrative plot of a 2-Dimensional latent space generated from a VAE trained on all of the digits 0-9. Colors indicate rough regions of interest for the digit 0 (green), 9 (red), 5 (purple), and 7 (blue). (Right) Exploration of different QUAK approaches as presented using the ROC applied on the separating of the digit 5 vs. the digit 9. We show the performance for (orange) a single autoencoder loss trained on the digit 5, (blue) a fully supervised network training the digit 5 vs. the digit 9, (red) a dense network trained to separate the digit 5 from 7 on a 2D QUAK space consisting of the loss of the digit 5 on one axis, and the loss of the digit 7 on another axis, (green) a dense network trained to separate the digit 5 from 7 on a 9D QUAK space consisting of the loss of each digit 0–8, (purple) a linear discriminant trained to separate the digit 5 from 7 on the same 9D QUAK space consisting of the loss of each digit 0–8.

supervised network performance as the dimension of QUAK space is increased. When compared with a signal autoencoder, we again observe a rejection factor that is many orders of magnitude better for the same signal efficiency (true positive rate).

4.1.2 Proxy signal choice

The choice of the proxy digit 7 to discriminate 9 is particularly appealing since both digits are quite similar. However, there is no guarantee that such an appealing choice would be present when searching for an anomaly. To illustrate the flexibility in proxy signal choice, we consider the instance where we use this same 7 digit proxy and dilute it within another less similar digit, 0. Following the diagram in the left of figure 3, the digit 0 is equally distant between the background digit 5 and signal digit 9 within a latent space generated on all digits. Figure 4 shows the performance in separating the 5 digit from the 9 digit with the diluted proxy. We again observe enhanced discrimination approaching the fully supervised limit as we inject more of the digit 7.

4.1.3 Latent vs loss space

In both the dijet example presented above and MNIST, we have chosen to construct QUAK from the losses of dedicated (NF)VAEs. Another approach is to not use the VAE loss and, instead, using a latent space generated from a single VAE trained on both signal and background priors together. This approach reduces the number of trainings from N-signal or background priors to a single training. However, it requires a significantly larger and more flexible network capable of capturing all different features.

JHEP06(2021)030

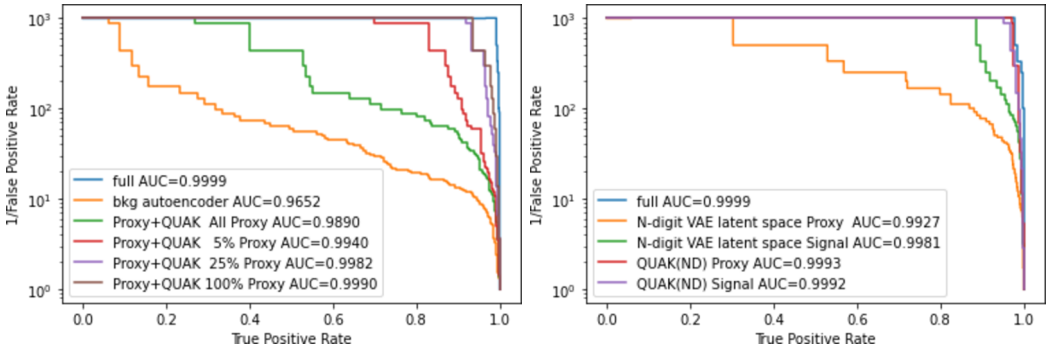

Figure 4. Performance in separating the digit 5 from 9 for a fully supervised network (full) with (Left) a linear discriminant on the space of digits 0-8 using a proxy signal that is X% digit 7 and 100-X% digit 0. (Right) Performance comparison of QUAK compared to an 9-dimensional latent space from a VAE trained on all digits 0–8 to separate out the digit 5 with the digit 9 with QUAK. To perform discrimination on the latent space a supervised network is trained on the latent space produced from the digit 5 against a latent space produced from the digits Signal(9)/Proxy(7). Performances are compared to QUAK space using a linear discriminant on the 0–8 digit loss space again using either Signal(9)/Proxy(7).

In figure 4, we compare the performance of a VAE with a 9-dimensional latent space trained on all 9-digits ranging from 0–8 simultaneously. We then use the latent space's N-dimensional output to train a three-layer multilayer perceptron to separate the digit 5 from 9 (signal) or 7 (proxy). We contrast this approach with the QUAK construction, whereby we train a linear discriminator on the 9-dimensional QUAK (loss) space. We observe QUAK significantly outperforms the latent space.

From the results obtained with MNIST, we can make substantive conclusions about the design and future use of QUAK. First, we observe that high-dimensional QUAK spaces are highly effective provided a proxy signal prior is used that is similar to the potential anomalous signal. This proxy signal is used to reduce the dimensionality of the anomaly search. Despite using a proxy signal that differs from the true signal, we find that a linear discriminant trained on this proxy signal is sufficient to separate the individual digits 5 vs. 9. Furthermore, the proxy signal can be diluted within other incorrect signals and still give a comparable performance improvement.

With our study on the MNIST dataset, we also observe that the latent space is not as effective as the N-dimensional QUAK (loss) space. We interpret this result as an indication that an individually trained VAE on each digit captures more information about a single-digit than a combined training intended to resolve all digits simultaneously. In the case where a VAE is sufficiently large and flexible to reproduce all digits as effectively as a VAE trained on a single digit, we expect that the use of the latent space in place of QUAK space should converge to the same level of performance.

4.1.4 Concluding observations

Finally, we would like to conclude that our studies with MNIST illustrate that QUAK can be extended in several directions towards more complex approaches to search for anomalies.

JHEP06(2021)030

There is a rich and growing set of semi-supervised approaches that are being developed along similar lines. We believe that the addition of QUAK can help complement these other approaches and provides new insight towards the construction of semi-supervised algorithms to perform anomaly detection.

As with the construction of QUAK space, we have effectively introduced a signal prior by separating digits or inserting signals. This will bias an anomaly search towards a specific region. However, we would like to stress that these added priors can deviate significantly from a true signal and still be an effective tool to search for anomalies. More generically, this concept lends itself to other types of physics analyses, such as measuring properties of a system where some, but not all, physical effects are known.

4.2 LHC olympics

Next, we perform an anomaly search with QUAK using the official LHC Olympics 2020 dataset [28]. The LHC Olympics 2020 consists of Pythia-simulated signal and background processes with detector smearing applied through the Delphes package [50–53]. A pure background sample consisting of QCD multijet(QCD) production using Pythia simulated dijet events is produced. This sample is used as a reference for simulated background events.

A simulated two-prong signal sample is generated from a decay chain composed of a heavy charged spin-1 mediator $W' \to XY, X \to \bar{q}q, Y \to \bar{q}q$, where X and Y are new physics mediators that decay to quark pairs and are light enough that their decay products fall within a single cone (i.e., a jet). Simulated three-prong signal events are generated from a heavy neutral spin-1 mediator decaying into a pair of heavy resonances that decay into three quarks. The intermediate resonances are also light enough that their decay products are reconstructed within single jets. A range of samples with different resonance masses and intermediary resonance masses are used to test various signals models' performance.

Lastly, as a part of the LHC Olympics, a series of BlackBox datasets (1-3) are constructed. The BlackBox datasets' goal is to emulate a true data sample where a new physics signal is hidden within the background. In BlackBox 1, a sample with similar topology to the $W' \to XY$ signal sample is hidden within a background sample. This sample is mixed into a QCD background sample using the same matrix element generation of the original Pythia sample, but with modified Pythia showering parameters so that the background does not match the pure background "simulation" sample. The variation in shower parameters is intended to emulate the observed deviation between QCD simulation and observed LHC multi-jet events. This means that data-driven techniques are needed within the BlackBox datasets to ensure the possibility of a signal.

In the following study, we treat BlackBox 1 as a hypothetical dataset. We use the simulated signal and background samples to construct QUAK space and calibrate our approach. We then apply QUAK to BlackBox 1 and use it to search for anomalous features. Finally, we consider alternative anomalous signatures and discuss how QUAK compares to other anomaly searches.

To perform an analysis on the LHC Olympic dataset, we first reconstruct jets using the anti-k_t algorithm with cone size parameter $\Delta R = 0.8$ [54, 55]. From the reconstructed jets, we select the two highest energy jets. High-level jet features of each jet is computed

JHEP06(2021)030

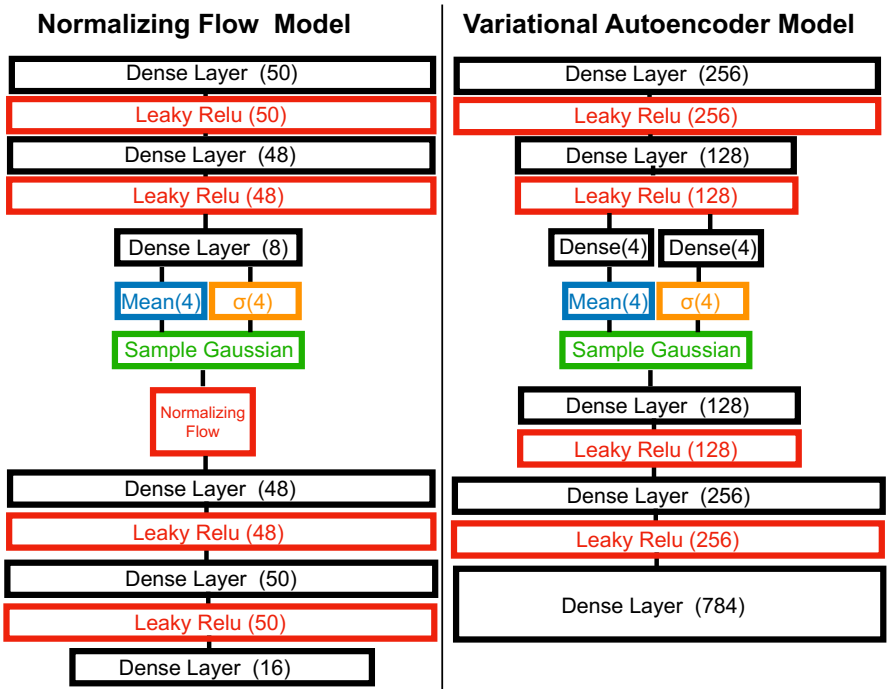

Figure 5. (Left) Illustration of the Normalizing Flow variational autoencoder network used for dijet identification. (Right) Illustration of the variational autoencoder used for the MNIST study. The number in parentheses indicates the number of output nodes of each layer.

and then these high level features are treated as inputs into the network for training. The high-level features consist of n-subjettines ratios ranging from 1-prong to 4-prong and the raw jet mass of the individual jets [56–59]. Training and testing are performed with 12 variables for each event (each jet: 4 n-subjettiness ratios, the total number of tracks, and the jet mass).

To perform the unsupervised training and to construct QUAK space, an optimization scan of network architectures and parameters is performed on the simulated background and a single two-prong signal dataset. A broad range of architectures is considered, and the network architecture with the highest area under a curve separation (AUC) between the chosen signal and background is used. In this scan, we considered several different normalizing flow VAEs including planar flow, radial flow, non-volume preserving flow, and masked autoregressive flow. The optimized network we found was a Masked Autoregressive normalizing flows [48, 60], with a latent $z_{dim} = 4$ with 3 dense layers on either end. The details of the network are shown in figure 5. We apply a loss metric of mean-squared reconstruction error on each of the 12 variables with a KL-divergence term to regularize the sampling parameters for each training.

As with other variational autoencoders, the KL-divergence term is added to constrain the samples mean and width of the Gaussian distributions to be near 0 and 1, respectively. This is added into the loss function with a tunable parameter β that characterizes the relative scale of the cross-entropy auto-encoder loss with the KL-divergence term [61]. We

JHEP06(2021)030

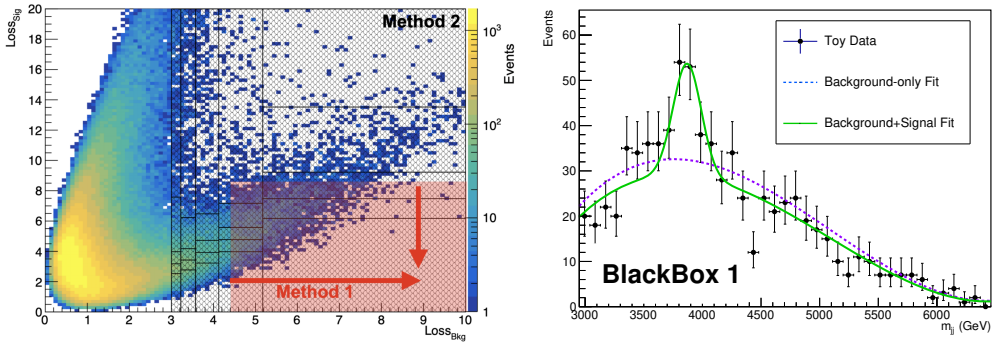

Figure 6. (Left) Illustration of the two methods used for the signal extraction. In Method 1, shown in red, we iteratively vary a selection on the signal loss L_{sig} and background loss L_{bkg} and select regions of low signal loss and high background loss. In Method 2, we separate the events by the black shaded boxes shown corresponding roughly to a uniform populations of events within each shaded region. (Right) Dijet mass fit after performing an optimized selection of $L_{\text{sig}} < 8$ and $L_{\text{bkg}} > 5.5$, a 3rd order Bernstein polynomial is used for the background while a Gaussian with a fixed width $\sigma/m_{jj} = 0.03$ is used for the signal model.

performed a hyperparameter scan for the β, and observed the optimal value of $\beta = 10$, and this value is used in the subsequent results. Finally, the QUAK space is constructed by computing the background loss and signal loss(es) with the KL-divergence term removed.

4.2.1 Example search on LHC Olympics BlackBox 1

First, we show the performance of QUAK on a search for a hidden signal in BlackBox 1 by constructing a 2-dimensional QUAK space constructed from loss on the simulated background sample with the loss of a single signal sample. For the signal loss, we use the loss trained on the "R&D" signal dataset, which consists of a $W' \to XY$ with the mass of the $W' = 3.5\,\text{TeV}$, and the mass of $X = 500\,\text{GeV}$, and $Y = 100\,\text{GeV}$. For the rest of this section, we will refer to L_{bkg} as the loss from the normalizing flow VAE trained on the background, and L_{sig} as the loss from the normalizing flow background trained on the chosen signal. The 2D QUAK space applied to the BlackBox 1 dataset is shown on the left of figure 6. While it was not known at the time of the LHC Olympics, a secret signal is hidden within the BlackBox 1 dataset. This signal consists of roughly 900 signal events injected into a background of 1 million events. The injected signal consists of a 3.8 TeV W' resonance decaying to two, two-prong resonances, X, and Y. In this case, the mass of X is 732 GeV, and the mass of the resonance Y is 378 GeV.

With the 2D QUAK space constructed, we consider two strategies to look for an excess within the space, method 1 and method 2. Both these strategies are shown in the left diagram of figure 6. In method 1, we systematically select events in a region of QUAK space and look for an excess of events. For this method, we require L_{sig} to be small, and L_{bkg} to be large. In method 2, we separate events within QUAK space into individual categories and perform a search for an anomalous resonance within each category separately. The individual categories are then combined, yielding a single anomaly

JHEP06(2021)030

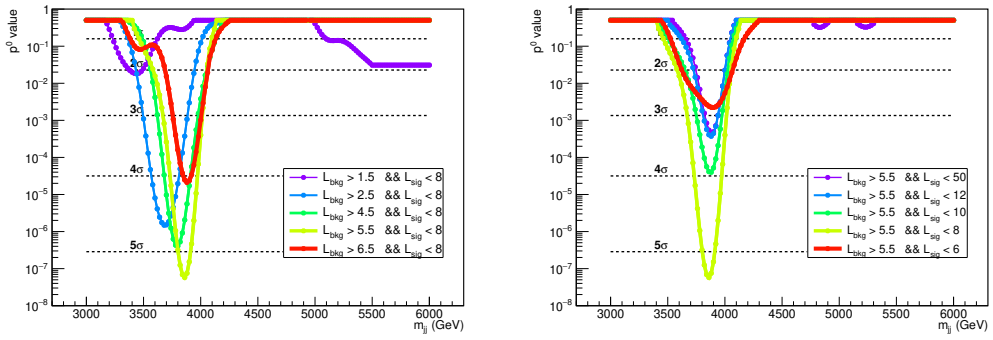

Figure 7. Scan of the p-value significance in the dijet mass performed on a fit to the dijet mass distribution for (a) a tight selection on the signal loss L_{sig} and varied selections on the background loss L_{bkg}, and for (b) a tight selection on the background loss L_{bkg} and a varied selection on signal loss L_{sig}. The significance is quoted is obtained from the asymptotic CLs method.

search. For each category/selection within QUAK space, a fit of the dijet mass m_{jj} is performed using a 3rd order Bernstein polynomial. The signal is assumed to be a Gaussian with a width of 3.0% $\times m_{jj}$, roughly equal to the detector resolution of the simulated sample. To compute the significance, we perform a binned likelihood fit and compute the significance through the likelihood ratio using the asymptotic CLs method [62]. The p-values of individual categories are combined using Fisher's method. The significance is computed at a fixed dijet mass with a fixed width, and the mass is scanned over a range of 3 TeV $< m_{jj} <$ 6 TeV.

To find an excess with method one, we systematically vary the selection on the signal loss requiring $L_{\text{sig}} < X$, with X varied. Additionally, we vary the selection, Y on the background loss requiring $L_{\text{bkg}} > Y$. We then construct a grid that is evenly spaced over the 2D space and systematically fit the dijet mass distribution of the selected events. From this scan, we observe a maximum significance exceeding 5 standard deviations for a cut of $L_{\text{bkg}} > 5.5$ and $L_{\text{sig}} < 8$. The right of figure 6 shows the resulting observed excess of events from the fit of the dijet mass after the optimized selection. In figure 7, we show the change in significance when we vary both the selection on the background loss and the signal loss. We observe that a selection on both signal and background loss variables is needed to obtain the large observed significance.

While method 1 is illustrative to find an excess, the quoted significance cannot be obtained from this approach since it is inherently biased towards finding large excesses. This results from the fact that the algorithm self-selects an excess in data by scanning variables and looking for an excess, whether it is real or not. In place of quoting the significance from method 1, we instead perform method 2. With method 2, we apply a preselection cut on the background loss, $L_{\text{bkg}} > 3$, roughly corresponding to a 2 percent background probability. After that preselection, we construct N uniformly populated categories by first dividing L_{bkg} into evenly populated regions and then within each region dividing along L_{sig} into evenly populated regions. The different regions used in this method are labeled as method

JHEP06(2021)030

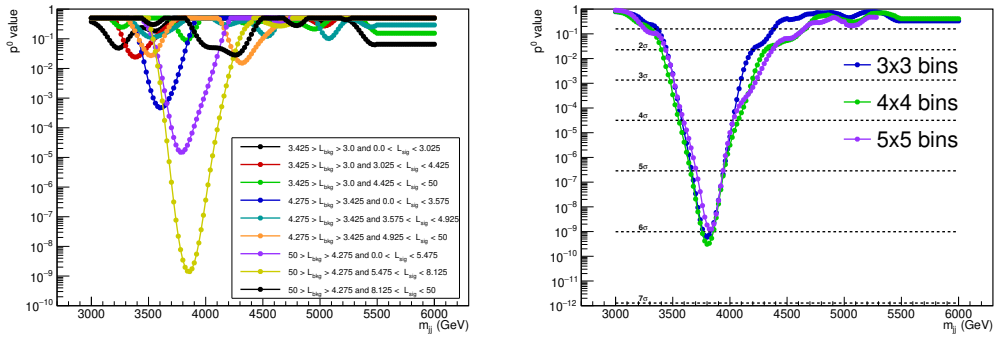

Figure 8. Scan of the significance of an excess quoted in the p-value significance of a Gaussian resonance on top of a fitted background using the asymptotic CLs method for (Left) individual categories in L_{sig} and L_{bkg}, and (Right) combined categories using a uniform populated set of 9 bins (3 in L_{sig}, 3 in L_{bkg}) 16 bins (4,4), and 25 bins (5,5).

2 in figure 6. With these uniformly populated bins, we perform a separate fit to the dijet mass distribution in each category. The uniform number of events allows for the same order Bernstein polynomial(3rd) to be used for every category of the fit. Additionally, we assume no correlation between fits. The categories are then combined into a single quoted p-value. To assess the performance of this method, we consider a 3x3,4x4,5x5 binned search in L_{bkg}xL_{sig} space. As a check of this method, we applied this method to a simulated sample with no signal injected. We found one excess above 2 standard deviations, consistent with the expected number of deviations expected from a random sampling.

Figure 8 shows the individual p-values obtained from method 2 for 9 separate bins; one region clearly dominates the p-value computation. The combined results yield an excess behyond 6 standard deviations. Additionally, we observe a consistent p-value profile when we perform the anomaly search in 9,16 and 25 bins. This approach effectively minimizes the look-elsewhere effect and allows us to quote a significance measurement within a broad, model independent, selected region.

In summary, we find an excess of events with a significant deviation from the background of over five standard deviations. The excess is found to be at 3.8 TeV, consistent with the hidden signal injected within the dataset. The excess of events is located in the region of low signal loss and high background loss, consistent with an anomalous signature similar to the test signal used to construct L_{sig}.

4.3 Performance of QUAK on different signals

Going beyond the example anomaly search performed with the BlackBox 1 dataset, in the following section, we consider the performance of QUAK on different signals and with higher dimensional QUAK space. Furthermore, to understand how effective QUAK is at finding new physics signatures, we characterize the performance of QUAK against fully supervised networks.

Figure 9 shows the performance of different QUAK methods in separating the background from a resonance decaying to two 3 prong resonances $Z' \to \bar{t}'t'$ with masses

JHEP06(2021)030

$m_{Z'} = 5\,\text{TeV}$, and $m_{t'} = m_{\bar{t}'} = 0.5\,\text{TeV}$. To understand the gain in adding signal loss to search for anomalies, we first consider anomaly detection performance with a single normalizing flow autoencoder trained on the background sample. In the following comparisons, we will label the single autoencoder as 1D QUAK since it is just a 1-dimensional QUAK space on L_{bkg}. To see the gain in using signal loss, we then construct 2D QUAK, which consists of 2D space with one axis being L_{bkg}, and the other axis L_{sig}. For L_{sig}, we choose as a signal prior the loss from a training on $W' \to XY$ with resonances X and Y decaying to two prongs , and the masses $m_{W'} = 4.5\,\text{TeV}$, $m_X = 500\,\text{GeV}$, and $m_Y = 150\,\text{GeV}$. Finally, we construct a 3D QUAK space by appending a 3rd loss L_{sig2} to the 2D QUAK space. Here L_{sig2} is computed from a network trained on the same Z' signal used for comparison.

To compare the performances in figure 9, Receiver Operator Characteristics (ROCs) are obtained by systematically scanning the 2D or 3D QUAK spaces linearly in each dimension of QUAK and requiring events to be selected from the region of minimum signal loss, $L_{\text{sig}_i} < X$, and maximum QCD loss, $L_{\text{bkg}} > Y$. This is done for the background sample and the signal sample; the resulting background and signal efficiencies for each selection are presented on the plot. The ROC characterizes the discrimination power for a systematic n-dimensional search to find anomalies.

The performance comparison when going from 1D to 2D to 3D QUAK is shown on the left in figure 9. In addition to QUAK, we compare QUAK to a supervised classifier trained on the same inputs, excluding the jet masses. The chosen signal prior to the supervised training is the same signal used in the ROC computation. The fully supervised network consists of a 4 layer multi-layer perceptron with batch normalization [63] and dropout [64]. Jet masses are excluded from the training since they have the ability to isolate specific kinematic features present in the samples that are not available in a normalizing flow autoencoder, which, by construction, is not strongly correlated with jet masses. Furthermore, in a realistic analysis scenario, similar supervised searches aim to decorrelate discrimination against the mass so that fitting and template based background methods can be used to extract the signal [65–82]. By comparing 1D, 2D, and 3D QUAK, we observe an increase in the search's sensitivity by adding more approximate signal priors. The addition of the approximate priors approaches, and in some places exceeds, the performance of a supervised discriminator computed by training the same inputs on the known signal. Interestingly, much of the gain in the separation of the 3-prong signal arises by adding the 2-prong signal prior despite these signals having manifestly different topologies. As a reference to existing studies, previously proposed single autoencoder searches would comprise 1D QUAK, so we see a large improvement over previous single autoencoder based searches given this modified search strategy.

With QUAK, we conclude that even if the signal priors are not accurate, we gain a sizable performance improvement. We theorize that the added information present in the signal loss helps isolate "signal-like" anomalies from other anomalous features present within the background. Through the construction of the QUAK space, we also demonstrate that incorrect signal priors, whether they result from inaccurate simulation or signal model choice, can still be a powerful discriminant when searching for new physics.

JHEP06(2021)030

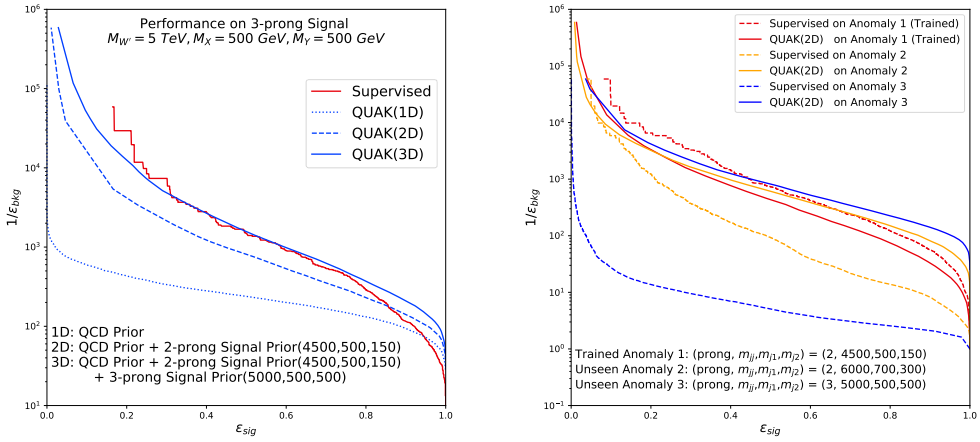

Figure 9. (Left) Receiver Operator Characteristic (ROC) for a resonant signal compared to QCD background; the resonant signal has mass $m_{jj} = 5\,\text{TeV}$ that decays into two three-pronged objects with mass $m_{j1} = m_{j2} = 500\,\text{GeV}$ against a background. Signal priors are labeled in the legend by (m_{jj}, m_{j1}, m_{j2}). Performances are shown in blue for the 1D QUAK (QCD prior only), 2D QUAK (QCD prior and a 2-prong signal prior) , and 3D QUAK (+ 3-prong signal prior), and shown in red for a fully supervised training on the true signal prior against background. Jet masses (m_{j1}, m_{j2}) are excluded in the training of the supervised classifier to mitigate model dependence. (Right) ROC for signal versus QCD background computed from a selection on *two* neural networks consisting of: QUAK(2D) (solid) for one network and a supervised network (dashed) for the other network. Both networks are trained once and constructed using the same 2-prong signal prior. The networks are then applied to different signal models. For both QUAK and the supervised network a signal prior of a resonance $W' \to XY$ with X and Y decaying to pairs of quarks having masses $(4500, 500, 150)$ is used in the training.

In the right part of figure 9, we contrast QUAK with a conventional new physics approach based on supervised learning. We train a supervised classifier to discriminate background against a specific two-pronged signal. We then apply this supervised classifier to a range of different signal models. The use of a supervised classifier is reflective of current supervised based machine learning searches used at the LHC since there is no guarantee that a signal model's simulation is consistent with a true signal. For the supervised training, we choose a $W' \to XY$ with resonances X and Y decaying to two prongs, and the masses $m_{W'} = 4.5\,\text{TeV}$, $m_X = 500\,\text{GeV}$, and $m_Y = 150\,\text{GeV}$. A fully-connected network is used for the supervised classifier (4 hidden layers with batch normalization [63] and dropout [64]). Both QUAK and the supervised classifier are trained on the same raw inputs, signal prior, and background prior (a QCD prior and a 2-prong prior) [63, 64], again exluding jet masses for the supervised network.

With the supervised network, we observe a general trend where the supervised classifier performs worse that QUAK as the signal deviates further from the chosen 2-prong prior used to train the supervised classifier. With the 3-prong signal, we find extremely poor performance with the supervised classifier. With QUAK, we observe a relatively stable

JHEP06(2021)030

separation between background and signal even as the test signal further deviates from the chosen signal prior. We conclude that QUAK incorporates signal priors in a more efficient way than supervised classifiers, and by using QUAK, we can do a more efficient scan of the possible space of new physics. For searches where the signal prior is partially known (to within uncertainties), QUAK has the potential to mitigate loss in sensitivity since deviations within this space have a smaller impact on the separation power when compared with supervised models.

We stress that the assumption that the signal is "approximately" correct is an assumption that is implicit in every new physics search at the LHC. All physics searches at LHC can be reduced to separating hypothetical signals from the background with a signal represented by a simulation that is our closest guess to what we believe the true signal is. With QUAK, we have constructed a space that attempts to mollify the differences between incorrect assumptions.

QUAK is a likelihood based approach, which aims to describe all of the variables of a specific process. As a consequence, our approach captures full distributions of the data. Whereas, supervised networks aim to optimize a decision boundary between a hypothetical signal, and a background under the assumption that a signal and background simulation are accurate representations. Methods like QUAK, which are sensitive to the full likelihood will be impacted less by deviations in simulated parameters from truth when compared with a supervised network. As a consequence, based on our observed performance, we believe QUAK and other likelihood based methods will be more robust to systematic variations, and uncertainties within the simulated signal and background modelling.

Minimizing the impact of systematic uncertainties on network design has been the focus of many studies. Previous approaches have used adversarial neural networks and other bounding principles to constrain the impact of systematic uncertainties on a network performance [83–86]. Based on our observed variations, the use of QUAK space in conjunction with these approaches can lead to further improvements in this area. In future LHC analyses, we expect that QUAK and other likelihood based approaches will be able to outperform supervised networks by limiting their sensitivity to inadequate simulations.

4.3.1 Concluding observations

In summary, we find that QUAK space's construction can be used to find anomalies over a broad range of signatures. It is robust against variations in signal choice. Additionally, incorrect signal priors within the QUAK space construction can still significantly enhance the detection of anomalies. Furthermore, QUAK is not strongly sensitive to model variations or incorrect signal choices, allowing for many new physics models to be probed by one search. Finally, we find that the QUAK construction can approach and sometimes exceed supervised networks.

5 Conclusions and outlook

In summary, we propose the exploration of a new algorithm, QUAK, to perform model independent searches. We demonstrate this work in the context of new physics search at

JHEP06(2021)030

the LHC. We observe that the addition of approximate priors to anomaly loss allows for enhanced identification of anomalies by providing generic "signal-like" or "background-like" features to help in identification. With QUAK, we have presented an approach that effectively adds these priors without degrading the sensitivity of a prior-free model. Furthermore, by relying on unsupervised learning techniques, we have allowed the neural networks to self-organize leading to a network performance that is robust against large variations. QUAK is broadly applicable to many different problems and can improve both anomaly searches and searches where large uncertainties are present on the signal modeling.

In particular, we have demonstrated QUAK as a new approach to construct anomaly searches using autoencoders and injecting signal priors. We demonstrate that even with incorrect priors, we can enhance anomaly searches by creating an anomalous space that is more conducive to searching for new physics models. Our work demonstrates a significant improvement in discrimination power on new physics models, when compared with previous single autoencoder approaches, even when the true signal deviates significantly from the assumed signal prior.

Furthermore, we observe that QUAK can approach or even exceed the sensitivity of supervised searches. Recently, a number of studies have observed that unsupervised and semi-supervised approaches are capable of exceeding the performance of direct supervised training. This has been attributed to the self assembly of the weights that occurs when training an unsupervised network. Lastly, while we have not directly compared this method to the classification without labels approach to anomaly searches, we would like to highlight that these approaches are mutually exclusive, and they can be used in concert.

We first characterized QUAK approach through an example using the MNIST dataset and a variational autoencoder architecture. We observe that QUAK can be extended to solve higher dimensional problems, provided a scheme to choose signal priors. Additionally, we observe that the QUAK construction is found to be more effective than a similar approach relying on the latent space.

To demonstrate the use of QUAK for anomalous features in LHC-like dijet data, we constructed QUAK space by using normalizing flow variational autoencoders trained on the n-subjettiness jet observable inputs. While we found both normalizing flows and n-subjettiness observables are effective for this approach; we would like to stress that this method can be used with other types of deep learning methods, and observable calculations. In particular, we think that recent work with energy flow polynomials, and earth movers distance can be used within QUAK in the construction of the space and training, respectively, to build a neural network with characteristic physical features [87, 88]. We see this as an exciting avenue for future study.

Based on the small variation across signal models, we observe that QUAK lends itself well to measurements, not just anomaly searches, where the data is high-dimensional, and where the signal model is not well known. This is true for many fields in physics where simulation is limited, including gravitational wave modeling, identifying astrophysical signatures, and quark-gluon plasma physics. With QUAK, we can naturally construct signal models that carry some of the signal features, allowing for enhanced anomaly identification under precepts of physical principles. This can potentially be used in high energy physics

JHEP06(2021)030

to perform measurements of new particle decays from existing known resonances, such as the Higgs Boson [89]. Outside of physics, this work builds on recent results that relate to semi-supervised anomaly detection and semi-supervision itself. We see this work as an initial step towards a broad range of new deep learning approaches that can have a significant impact in many new fields.

Acknowledgments

P. H., D. R., M. Y. are partially supported by NSF grants #1934700, #1931469, and the IRIS-HEP grant #1836650. We would like to thank Nhan Tran, Cristina Mantilla Suarez, Jesse Thaler, and Javier Duarte for useful comments. We thank Erik Katsavounidis, Tri Nguyen, and Alec Gunny for interesting discussions. Additionally, we would like to thank David Shih, Gregor Kasieczka, and Ben Nachman for their help with the LHC Olympics dataset.

References

[1] G. Kasieczka et al., *The LHC olympics* 2020*: a community challenge for anomaly detection in high energy physics*, arXiv:2101.08320 [INSPIRE].

[2] B. Bortolato, B.M. Dillon, J.F. Kamenik and A. Smolkovič, *Bump hunting in latent space*, arXiv:2103.06595 [INSPIRE].

[3] G. Stein, U. Seljak and B. Dai, *Unsupervised in-distribution anomaly detection of new physics through conditional density estimation*, in 34$^{\text{th}}$ *conference on neural information processing systems*, (2020) [arXiv:2012.11638] [INSPIRE].

[4] B.M. Dillon, D.A. Faroughy, J.F. Kamenik and M. Szewc, *Learning the latent structure of collider events*, *JHEP* **10** (2020) 206 [arXiv:2005.12319] [INSPIRE].

[5] V. Mikuni and F. Canelli, *ABCNet: an attention-based method for particle tagging*, *Eur. Phys. J. Plus* **135** (2020) 463 [arXiv:2001.05311] [INSPIRE].

[6] J.H. Collins, P. Martín-Ramiro, B. Nachman and D. Shih, *Comparing weak- and unsupervised methods for resonant anomaly detection*, arXiv:2104.02092 [INSPIRE].

[7] K. Benkendorfer, L.L. Pottier and B. Nachman, *Simulation-assisted decorrelation for resonant anomaly detection*, arXiv:2009.02205 [INSPIRE].

[8] E.M. Metodiev, B. Nachman and J. Thaler, *Classification without labels: learning from mixed samples in high energy physics*, *JHEP* **10** (2017) 174 [arXiv:1708.02949] [INSPIRE].

[9] J.H. Collins, K. Howe and B. Nachman, *Anomaly detection for resonant new physics with machine learning*, *Phys. Rev. Lett.* **121** (2018) 241803 [arXiv:1805.02664] [INSPIRE].

[10] J.H. Collins, K. Howe and B. Nachman, *Extending the search for new resonances with machine learning*, *Phys. Rev. D* **99** (2019) 014038 [arXiv:1902.02634] [INSPIRE].

JHEP06(2021)030

[11] B. Nachman and D. Shih, *Anomaly detection with density estimation*, *Phys. Rev. D* **101** (2020) 075042 [arXiv:2001.04990] [INSPIRE].

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

[13] M. Farina, Y. Nakai and D. Shih, *Searching for new physics with deep autoencoders*, *Phys. Rev. D* **101** (2020) 075021 [arXiv:1808.08992] [INSPIRE].

[14] O. Cerri, T.Q. Nguyen, M. Pierini, M. Spiropulu and J.-R. Vlimant, *Variational autoencoders for new physics mining at the Large Hadron Collider*, *JHEP* **05** (2019) 036 [arXiv:1811.10276] [INSPIRE].

[15] M. Kuusela, T. Vatanen, E. Malmi, T. Raiko, T. Aaltonen and Y. Nagai, *Semi-supervised anomaly detection — towards model-independent searches of new physics*, *J. Phys. Conf. Ser.* **368** (2012) 012032.

[16] T.S. Roy and A.H. Vijay, *A robust anomaly finder based on autoencoders*, arXiv:1903.02032 [INSPIRE].

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

[18] A. Blance, M. Spannowsky and P. Waite, *Adversarially-trained autoencoders for robust unsupervised new physics searches*, *JHEP* **10** (2019) 047 [arXiv:1905.10384] [INSPIRE].

[19] J. Hajer, Y.-Y. Li, T. Liu and H. Wang, *Novelty detection meets collider physics*, *Phys. Rev. D* **101** (2020) 076015 [arXiv:1807.10261] [INSPIRE].

[20] R.T. D'Agnolo, G. Grosso, M. Pierini, A. Wulzer and M. Zanetti, *Learning multivariate new physics*, *Eur. Phys. J. C* **81** (2021) 89 [arXiv:1912.12155] [INSPIRE].

[21] R.T. D'Agnolo and A. Wulzer, *Learning new physics from a machine*, *Phys. Rev. D* **99** (2019) 015014 [arXiv:1806.02350] [INSPIRE].

[22] M. Crispim Romão, N.F. Castro and R. Pedro, *Finding new physics without learning about it: anomaly detection as a tool for searches at colliders*, *Eur. Phys. J. C* **81** (2021) 27 [arXiv:2006.05432] [INSPIRE].

[23] O. Amram and C.M. Suarez, *Tag n' train: a technique to train improved classifiers on unlabeled data*, *JHEP* **01** (2021) 153 [arXiv:2002.12376] [INSPIRE].

[24] A. Butter, G. Kasieczka, T. Plehn and M. Russell, *Deep-learned top tagging with a Lorentz layer*, *SciPost Phys.* **5** (2018) 028.

[25] C. Choy, J. Gwak and S. Savarese, *4d spatio-temporal convnets: Minkowski convolutional neural networks*, arXiv:1904.08755.

[26] A. Bogatskiy, B. Anderson, J.T. Offermann, M. Roussi, D.W. Miller and R. Kondor, *Lorentz group equivariant neural network for particle physics*, arXiv:2006.04780 [INSPIRE].

[27] Y. LeCun and C. Cortes, *MNIST handwritten digit database*, http://yann.lecun.com/exdb/mnist/.

[28] G. Kasieczka, B. Nachman and D. Shih, *Official datasets for LHC olympics* 2020 *anomaly detection challenge*, Zenodo, (2019).

[29] T. Chen, S. Kornblith, K. Swersky, M. Norouzi and G. Hinton, *Big self-supervised models are strong semi-supervised learners*, arXiv:2006.10029.

JHEP06(2021)030

[30] Y. Ouali, C. Hudelot and M. Tami, *An overview of deep semi-supervised learning*, arXiv:2006.05278.

[31] D. Hendrycks, M. Mazeika, S. Kadavath and D. Song, *Using self-supervised learning can improve model robustness and uncertainty*, arXiv:1906.12340.

[32] L. Ruff et al., *Deep semi-supervised anomaly detection*, arXiv:1906.02694.

[33] D. Hendrycks, M. Mazeika and T. Dietterich, *Deep anomaly detection with outlier exposure*, arXiv:1812.04606.

[34] T. Cheng, J.-F. Arguin, J. Leissner-Martin, J. Pilette and T. Golling, *Variational autoencoders for anomalous jet tagging*, arXiv:2007.01850.

[35] D.J. Rezende et al., *Normalizing flows on tori and spheres*, arXiv:2002.02428 [INSPIRE].

[36] M.S. Albergo, G. Kanwar and P.E. Shanahan, *Flow-based generative models for Markov chain Monte Carlo in lattice field theory*, *Phys. Rev. D* **100** (2019) 034515 [arXiv:1904.12072] [INSPIRE].

[37] G. Kanwar et al., *Equivariant flow-based sampling for lattice gauge theory*, *Phys. Rev. Lett.* **125** (2020) 121601 [arXiv:2003.06413] [INSPIRE].

[38] J. Brehmer and K. Cranmer, *Flows for simultaneous manifold learning and density estimation*, arXiv:2003.13913 [INSPIRE].

[39] E. Bothmann, T. Janßen, M. Knobbe, T. Schmale and S. Schumann, *Exploring phase space with neural importance sampling*, *SciPost Phys.* **8** (2020) 069 [arXiv:2001.05478] [INSPIRE].

[40] C. Gao, S. Höche, J. Isaacson, C. Krause and H. Schulz, *Event generation with normalizing flows*, *Phys. Rev. D* **101** (2020) 076002 [arXiv:2001.10028] [INSPIRE].

[41] C. Gao, J. Isaacson and C. Krause, *i-flow: high-dimensional integration and sampling with normalizing flows*, *Mach. Learn. Sci. Tech.* **1** (2020) 045023 [arXiv:2001.05486] [INSPIRE].

[42] S. Choi, J. Lim and H. Oh, *Data-driven estimation of background distribution through neural autoregressive flows*, arXiv:2008.03636 [INSPIRE].

[43] Y. Lu, J. Collado, D. Whiteson and P. Baldi, *Sparse autoregressive models for scalable generation of sparse images in particle physics*, *Phys. Rev. D* **103** (2021) 036012 [arXiv:2009.14017] [INSPIRE].

[44] S. Bieringer et al., *Measuring QCD splittings with invertible networks*, arXiv:2012.09873 [INSPIRE].

[45] J. Hollingsworth, M. Ratz, P. Tanedo and D. Whiteson, *Efficient sampling of constrained high-dimensional theoretical spaces with machine learning*, arXiv:2103.06957 [INSPIRE].

[46] D.P. Kingma and M. Welling, *Auto-encoding variational Bayes*, arXiv:1312.6114 [INSPIRE].

[47] S.R. Bowman, L. Vilnis, O. Vinyals, A. Dai, R. Jozefowicz and S. Bengio, *Generating sentences from a continuous space*, in *Proceedings of the* 20$^{\text{th}}$ *SIGNLL conference on computational natural language learning*, Association for Computational Linguistics, (2016), pg. 10

[48] D.J. Rezende and S. Mohamed, *Variational inference with normalizing flows*, arXiv:1505.05770.

[49] D. Boyda et al., *Sampling using* SU(N) *gauge equivariant flows*, *Phys. Rev. D* **103** (2021) 074504 [arXiv:2008.05456] [INSPIRE].

JHEP06(2021)030

[50] DELPHES 3 collaboration, *DELPHES 3, a modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [arXiv:1307.6346] [INSPIRE].

[51] T. Sjöstrand et al., *An introduction to PYTHIA* 8.2, *Comput. Phys. Commun.* **191** (2015) 159 [arXiv:1410.3012] [INSPIRE].

[52] T. Sjöstrand, S. Mrenna and P.Z. Skands, *PYTHIA* 6.4 *physics and manual*, *JHEP* **05** (2006) 026 [hep-ph/0603175] [INSPIRE].

[53] T. Sjöstrand, S. Mrenna and P.Z. Skands, *A brief introduction to PYTHIA* 8.1, *Comput. Phys. Commun.* **178** (2008) 852 [arXiv:0710.3820] [INSPIRE].

[54] M. Cacciari and G.P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, *Phys. Lett. B* **641** (2006) 57 [hep-ph/0512210] [INSPIRE].

[55] M. Cacciari, G.P. Salam and G. Soyez, *FastJet user manual*, *Eur. Phys. J. C* **72** (2012) 1896 [arXiv:1111.6097] [INSPIRE].

[56] J. Thaler and K. Van Tilburg, *Identifying boosted objects with N-subjettiness*, *JHEP* **03** (2011) 015 [arXiv:1011.2268] [INSPIRE].

[57] J. Thaler and K. Van Tilburg, *Maximizing boosted top identification by minimizing N-subjettiness*, *JHEP* **02** (2012) 093 [arXiv:1108.2701] [INSPIRE].

[58] J. Thaler and K. Van Tilburg, *Identifying boosted objects with N-subjettiness*, *JHEP* **03** (2011) 015 [arXiv:1011.2268] [INSPIRE].

[59] K. Datta and A. Larkoski, *How much information is in a jet?*, *JHEP* **06** (2017) 073 [arXiv:1704.08249] [INSPIRE].

[60] G. Papamakarios, T. Pavlakou and I. Murray, *Masked autoregressive flow for density estimation*, arXiv:1705.07057.

[61] I. Higgins et al., *beta-vae: learning basic visual concepts with a constrained variational framework*, in *ICLR*, (2017).

[62] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554 [*Erratum ibid.* **73** (2013) 2501] [arXiv:1007.1727] [INSPIRE].

[63] S. Ioffe and C. Szegedy, *Batch normalization: accelerating deep network training by reducing internal covariate shift*, arXiv:1502.03167 [INSPIRE].

[64] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever and R.R. Salakhutdinov, *Improving neural networks by preventing co-adaptation of feature detectors*, arXiv:1207.0580.

[65] J. Dolen, P. Harris, S. Marzani, S. Rappoccio and N. Tran, *Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure*, *JHEP* **05** (2016) 156 [arXiv:1603.00027] [INSPIRE].

[66] I. Moult, B. Nachman and D. Neill, *Convolved substructure: analytically decorrelating jet substructure observables*, *JHEP* **05** (2018) 002 [arXiv:1710.06859] [INSPIRE].

[67] J. Stevens and M. Williams, *uBoost: a boosting method for producing uniform selection efficiencies from multivariate classifiers*, 2013 *JINST* **8** P12013 [arXiv:1305.7248] [INSPIRE].

[68] C. Shimmin et al., *Decorrelated jet substructure tagging using adversarial neural networks*, *Phys. Rev. D* **96** (2017) 074034 [arXiv:1703.03507] [INSPIRE].

JHEP06(2021)030

[69] L. Bradshaw, R.K. Mishra, A. Mitridate and B. Ostdiek, *Mass agnostic jet taggers*, *SciPost Phys.* **8** (2020) 011.

[70] ATLAS collaboration, *Performance of mass-decorrelated jet substructure observables for hadronic two-body decay tagging in ATLAS*, Tech. Rep. ATL-PHYS-PUB-2018-014, CERN, Geneva, Switzerland (2018).

[71] G. Kasieczka and D. Shih, *Robust jet classifiers through distance correlation*, *Phys. Rev. Lett.* **125** (2020) 122001 [arXiv:2001.05310] [INSPIRE].

[72] G. Kasieczka, B. Nachman, M.D. Schwartz and D. Shih, *Automating the ABCD method with machine learning*, *Phys. Rev. D* **103** (2021) 035021 [arXiv:2007.14400] [INSPIRE].

[73] CMS collaboration, *A multi-dimensional search for new heavy resonances decaying to boosted WW, WZ, or ZZ boson pairs in the dijet final state at 13 TeV*, *Eur. Phys. J. C* **80** (2020) 237 [arXiv:1906.05977] [INSPIRE].

[74] CMS collaboration, *Inclusive search for highly boosted Higgs bosons decaying to bottom quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **12** (2020) 085 [arXiv:2006.13251] [INSPIRE].

[75] CMS collaboration, *Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **01** (2018) 097 [arXiv:1710.00159] [INSPIRE].

[76] CMS collaboration, *Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev. D* **100** (2019) 112007 [arXiv:1909.04114] [INSPIRE].

[77] ATLAS collaboration, *Search for diboson resonances in hadronic final states in 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **09** (2019) 091 [*Erratum ibid.* **06** (2020) 042] [arXiv:1906.08589] [INSPIRE].

[78] CMS collaboration, *Search for high mass dijet resonances with a new background prediction method in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **05** (2020) 033 [arXiv:1911.03947] [INSPIRE].

[79] CMS collaboration, *Search for pair-produced resonances decaying to quark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev. D* **98** (2018) 112014 [arXiv:1808.03124] [INSPIRE].

[80] ATLAS collaboration, *Identification of boosted Higgs bosons decaying into b-quark pairs with the ATLAS detector at 13 TeV*, *Eur. Phys. J. C* **79** (2019) 836 [arXiv:1906.11005] [INSPIRE].

[81] CMS collaboration, *Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair*, *Phys. Rev. Lett.* **120** (2018) 071802 [arXiv:1709.05543] [INSPIRE].

[82] CMS collaboration, *Search for new physics in final states with an energetic jet or a hadronically decaying W or Z boson and transverse momentum imbalance at $\sqrt{s} = 13$ TeV*, *Phys. Rev. D* **97** (2018) 092005 [arXiv:1712.02345] [INSPIRE].

[83] S. Wunsch, S. Jörger, R. Wolf and G. Quast, *Reducing the dependence of the neural network function to systematic uncertainties in the input space*, *Comput. Softw. Big Sci.* **4** (2020) 5 [arXiv:1907.11674] [INSPIRE].

JHEP06(2021)030

[84] C. Englert, P. Galler, P. Harris and M. Spannowsky, *Machine learning uncertainties with adversarial neural networks*, *Eur. Phys. J. C* **79** (2019) 4 [arXiv:1807.08763] [INSPIRE].

[85] L.-G. Xia, *QBDT, a new boosting decision tree method with systematical uncertainties into training for high energy physics*, *Nucl. Instrum. Meth. A* **930** (2019) 15 [arXiv:1810.08387] [INSPIRE].

[86] G. Louppe, M. Kagan and K. Cranmer, *Learning to pivot with adversarial networks*, arXiv:1611.01046 [INSPIRE].

[87] P.T. Komiske, E.M. Metodiev and J. Thaler, *Energy flow polynomials: a complete linear basis for jet substructure*, *JHEP* **04** (2018) 013 [arXiv:1712.07124] [INSPIRE].

[88] P.T. Komiske, E.M. Metodiev and J. Thaler, *Metric space of collider events*, *Phys. Rev. Lett.* **123** (2019) 041801 [arXiv:1902.02346] [INSPIRE].

[89] P.C. Harris, D.S. Rankin and C. Mantilla Suarez, *An approach to constraining the Higgs width at the LHC and HL-LHC*, arXiv:1910.02082 [INSPIRE].

JHEP06(2021)030