Identifying the quantum properties of hadronic resonances using machine learning

Jakub Filipek[1], Shih-Chieh Hsu[1], John Kruper[1],
Kirtimaan Mohan[2] and Benjamin Nachman[3,4]

1 Department of Physics, University of Washington, Seattle, WA 98195, USA
2 Department of Physics and Astronomy Division, Michigan State University,
567 Wilson Rd, East Lansing, MI 48824, USA
3 Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
4 Berkeley Institute for Data Science, University of California, Berkeley, CA 94720, USA

⋆ balbok@uw.edu , † schsu@uw.edu , ‡ jk23@att.net ,
○ kamohan@msu.edu , § bpnachman@lbl.gov

Abstract

With the great promise of deep learning, discoveries of new particles at the Large Hadron Collider (LHC) may be imminent. Following the discovery of a new Beyond the Standard model particle in an all-hadronic channel, deep learning can also be used to identify its quantum numbers. Convolutional neural networks (CNNs) using jet-images can significantly improve upon existing techniques to identify the quantum chromodynamic (QCD) ('color') as well as the spin of a two-prong resonance using its substructure. Additionally, jet-images are useful in determining what information in the jet radiation pattern is useful for classification, which could inspire future taggers. These techniques improve the categorization of new particles and are an important addition to the growing jet substructure toolkit, for searches and measurements at the LHC now and in the future.

Contents

1 Introduction

Many theories of physics beyond the Standard Model (BSM) predict new particles that exclusively or at least predominantly decay hadronically. As a result, the experiments at the Large Hadron Collider (LHC) have dedicated significant resources to search for such particles across a wide kinematic range. If the masses of the new particles are at or below the electroweak scale, then they may be produced with significant Lorentz boost. Both ATLAS [1–5] and CMS [6–9] have performed searches for direct or indirect (via a cascade decay) boosted BSM resonances that are fully contained inside a single large-radius jet. If such a particle was to be discovered, a critical task will be to identify its quantum numbers, such as the spin, electric charge and color state.

The goal of this paper is to investigate the potential of tagging the quantum numbers of hadronically decaying resonances using all of the available information inside the resulting jet. There are now many applications of deep learning to jet substructure [10] with a variety of jet representations used for the learning (images [11–18], ordered lists [19–23], trees [24, 25], graphs [26–31], unordered sets [32–34], observable bases [35–39], etc. [40–42] - see Ref. [43] for additional references). Jet images [44] are used here for their interpretability.

Non-deep learning observables have already been used by the LHC experiments to identify quantum properties of hadronic resonances [10, 45–49]. For example, variations of the jet charge [50, 51] have been used to study quark and gluon jets [52, 53] as well as boosted hadronically decaying W and Z bosons [54]. These techniques were recently combined with deep learning approaches [55, 56] to show that additional discrimination power is possible when all of the charge-sensitive information is utilized. Another quantum property which has been well studied with non-deep learning techniques is the color representation. For example, the jet pull [57] has been used to study singlet versus octet radiation patterns from dijet resonances [58–60]. So far, there has not been a corresponding deep learning study to identify potential gains from using all of the available information in the jet radiation pattern.

The authors of Ref. [61] performed a first study to categorize various boosted hadronic resonances based on their quantum numbers. By using classical jet substructure observables, they were able to show well one can distinguish various possibilities.[1] The goal of this paper is to extend the analysis of Ref. [61] using deep learning techniques and quantify the pairwise distinguishability of BSM scenarios. Deep learning approaches are compared with more tra-

---

[1]Recently, Ref. [62] also showed that in the color singlet case, one can derive new observables using analytic insight.

ditional methods and various visualizations are used to identify the useful information within the jet radiation pattern. While the jet-by-jet classification performance is weak, even a small amount of separation can be used in a template fit to extract the quantum numbers. Thus, we envision that this approach will be used in a post-discovery phase to categorize the resonance without needing excellent per-event distinguishing power.

This paper is organized as follows. Section 2 introduces the BSM models and how they are simulated. The machine learning setup is documented in Sec. 3, starting with the jet image preparation and preprocessing. Results are presented in Sec. 4 and the dependence of these results on variations in the setup is explored in Sec. 5. The paper ends with conclusions and future outlook in Sec. 6.

2 Simulation setup

A set of leptohobic resonances are used as benchmarks and are described below. In all cases, the mass of the new resonance is set to 125 GeV. The model UFO files are generated using `FeynRules` [63]. In the list below we describe the different resonances and their interaction with standard model particles. We denote quark and and anti-quark fields by q and $\bar{q}$ respectively, the strong coupling constant by g_s and the gluonic field strength tensor as $G^a_{\mu\nu}$.

1. New gauge bosons such as Z-primes (Z') are a common occurrence in models with extended electro-weak gauge sectors. The benchmark considered here is a color singlet leptophobic Z'. The Z' interacts only with quarks through $\frac{g_B}{6}\bar{q}\gamma^\mu q Z'_\mu$ [64]. Here g_B is the Z' gauge coupling and determines the strength of its interactions to quarks.

2. New massive colored gauge bosons such as colorons (C_μ) are common in models with an extended gauge symmetry of the strongly interacting sector. The color octet colorons interact with quarks through $g_s \tan\theta \ \bar{q} \ T^a \gamma^\mu q C^a_\mu$ [64, 65]. Here θ is a dimensionless parameter that parametrizes mixing between the gauge and mass eigenbasis of the gluon and coloron.

3. Scalar diquarks appear in the 27-dimensional representation of string inspired grand-unified theories [66]. The benchmark in this paper is a color sextet diquark (Φ^γ_6) that decays to pairs of quarks through $\sqrt{2}(\bar{K}_6)^{ab}_\gamma \lambda_\Phi \Phi^\gamma_6 \bar{u}_{Ra} u_{Lb}$ [67]. Here $(\bar{K}_6)^{ab}$ is the relevant $SU(3)_c$ Clebsch-Gordon coefficient and λ_Φ is a coupling constant.

4. In composite models, quarks can be composite at a scale Λ. As a result color triplet excited quarks (q^*) are a prediction of such models and interact with quarks and gluons (as well as other gauge bosons) through the interaction term $\frac{1}{2\Lambda}\bar{q}^*_R \sigma^{\mu\nu}[g_S f_S \frac{\lambda^a}{2} G^a_{\mu\nu}]q_L$ [68]. Here q_L represents the doublet of left handed Standard Model quarks, λ^a are the Gell-Mann matrices, $\sigma^{\mu\nu}$ is the usual bilinear covariant and f_s is a coupling contant and Λ is the scale of validity of the effective field theory.

5. Massive Spin 2 objects ($X^{\mu\nu}$) may appear as Kaluza-Klein excitations of the gravitational field $h^{\mu\nu}$ in RS models. They are color singlets and interact with SM particles through the energy momentum tensor $T_{\mu\nu}$ as $\frac{1}{\Lambda}X^{\mu\nu}T_{\mu\nu}$ [69, 70], decaying to either quarks or gluons. Here Λ represents the scale of validity of the effective field theory [70–73].

6. Color octet scalars (S_8) that are electroweak singlets may arise in technicolor or extra-dimension models [74, 75]. They interact with gluons through the field strength tensor as $\frac{g_s d_{ABC} k_s}{\Lambda}S_8^A G^B_{\mu\nu} G^{C,\mu\nu}$. Here d_{ABC} are the usual $SU(3)$ structure constants and Λ represents the scale of validity of the effective field theory.

7. Two additional models are color singlet scalars that decay to a pair of gluons or a pair of quarks, like the SM Higgs boson. These particles are denoted by $H \to qq$ or $H \to gg$.

Proton-proton collisions involving the above models are generated with `MadGraph5_amc@NLO` [76], version 2.5.5, and `Pythia8` [77], version 8.235. Particles with a mean lifetime $\tau > 30$ ps, excluding neutrinos, are clustered into jets with `FastJet` 3 [78, 79], version 3.3.2, using the anti-k_t algorithm [80] with R = 1.0 and are required to be central with a pseudorapidity $|\eta| < 2.0$. Due to the benchmark mass of the resonance, the leading jet p_T is required to be between 300 GeV and 600 GeV to ensure that one large-radius jet is sufficient to capture most of the hadronic decay products. As in Ref. [11], jets are trimmed [81] by re-clustering the constituents into R = 0.3 k_t subjets and and dropping those which have $p_T^{\text{subjet}} < 0.05 \times p_T^{\text{jet}}$. Jet grooming is revisited in Sec. 5. In order for an event to be used for machine learning, the leading jet must have $m_{\text{jet}} \in [100, 150]$ GeV.

As stated in Sec. 1, the premise of this study is that a new resonance is identified and an event sample is investigated to identify its properties. The task of identifying the BSM in the first place from the generic quark and gluon jet background is beyond the scope of this goal.

3 Machine learning

3.1 Preprocessing

Jet images are constructed and preprocessed using a similar methodology as the one from Ref. [11] and briefly reviewed in the following. Images cover a range in η-ϕ that is $2R \times 2R$ where $R = 1$ is the radius of the jet. Jets are translated so that the average weighted on pixel intensity is at the origin. The images are pixelated with a size 65×65. The second subject is used to rotate the jet so that the vector connecting the leading and subleading subjets is at $-\pi/2$. In the absence of a second subject, the principle component axes are used instead. A last degree of freedom is removed by flipping the images horizontally so that the right side of the image has a higher sum p_T than the left side. The pixel intensities are then transformed with a logarithm and normalized to unity. The lossy nature of these preprocessing steps has been studied elsewhere [11, 82].

The average jet image for the event samples described in Sec. 2 is presented in Fig. 1. The leading subject is the bright spot in the center of the image and the elongated bright spot below it corresponds to the subleading subject. All eight of the average jet images in the top plot of Fig. 1 look similar by eye and so the bottom plot shows the ratio of the average jet image for a given sample to the one from the $H \to gg$ sample. The ratio images show systematic trends at the 5-10% level that illustrate the information that a neural network could use for distinguishing the various possibilities.

3.2 Low level tagger: Convolutional neural network on jet images

A Convolutional Neural Network (CNN) (Fig. 2) consisting of eight layers of convolutional filters, one dense layer with 128 neurons, and a final dense layer with a single output neuron is used. The first convolutional layer has eight 11×11 filters, all other convolutional layers have sixteen 3×3 filters. Each of the third, fifth, and seventh convolutional layers is followed by 2×2 max-pooling layers. The hidden dense layer is followed by a dropout layer with 50% dropout. All layers are followed by rectified linear unit (ReLU) activation.

The CNN training minimized the binary cross-entropy loss using the Nesterov Adam optimizer [83, 84] with a learning rate of 3×10^{-4} and a scheduled decay of 0.153. As the radiation pattern of the different models varies on the periphery of the jet, the jet mass distribution is not
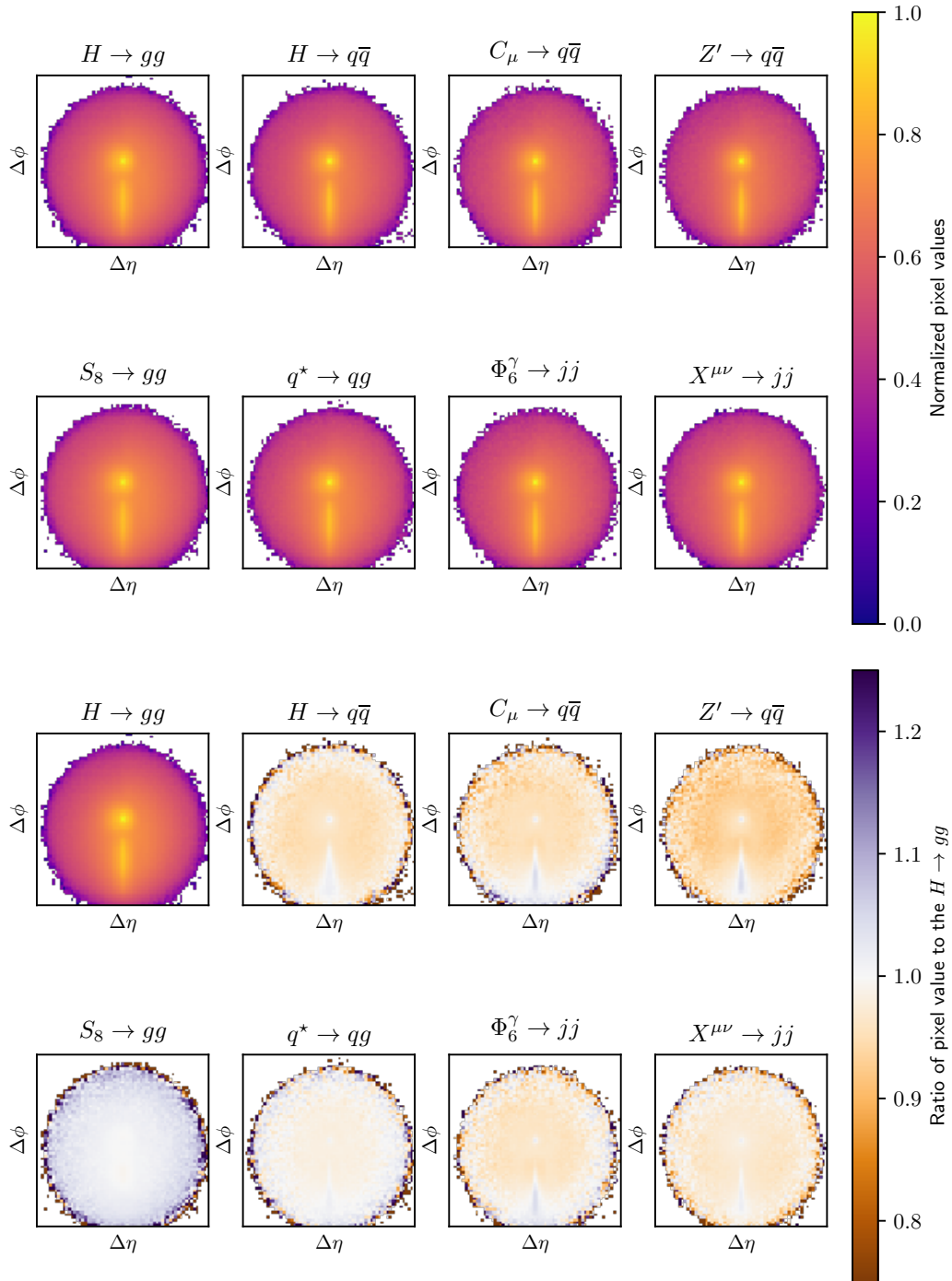
Figure 1: Top: Average images of jets generated by the processes considered in Sec. 2. Bottom: Per pixel ratios of the average images to the $H \to gg$ case. The upper-left image is left for reference, since the ratio of $H \to gg$ with itself would be all ones. Decays with the same final state have more similar jet images. While most processes go to either $q\bar{q}$ or gg, the excited quark ($q^\star$) goes to qg (or $\bar{q}g$). In this paper, classifiers are trained to distinguish these processes from each other.
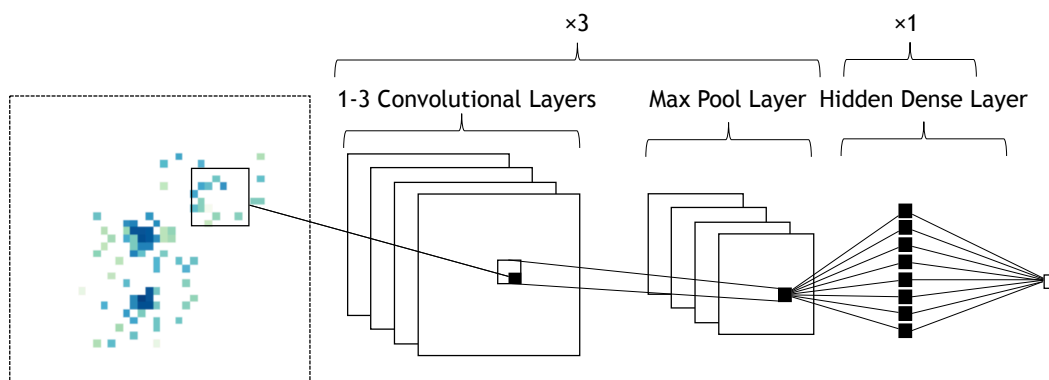
Figure 2: Architecture of the convolutional neural network. The input is a prepro-
cessed 65×65 jet-image. There are eight convolutional layers with three intermixed
max-pooling layers, followed by a dense layer. The output is a single number.

identical for all models. To remove the classification power of the jet mass, event are weighted
(using fine-binned histograms) for each binary classification task. The training is done with a
batch size of 1024 for 200 epochs. The best model according to the validation loss is saved.
Keras [85] built on TensorFlow [86] is used to train the CNN.

The Sequential Model-based Algorithm Configuration in python 3 (SMAC3) is used to op-
timize the CNN hyperparameters [87]. SMAC3 is based on Bayesian Optimization, and is used
to optimize arbitrary algorithms. The hyperparameters considered are the first layer filter size,
number of filters per convolutional layer, number of convolutional layers, dropout, activation
function, width of the dense layer(s), number of dense layer(s), optimizer type, learning rate,
and decay rate. A first layer filter size of 11×11 is considered because it can capture aspects of
both the leading and subleading subjet [11]. This is important for learning relations between
the sparse, non-zero parts of image. 11 pixels correspond roughly to the distance between the
two most energetic pixels in the 65×65 image. After the first filter, a series of 3×3 filters
follows, since they are able to handle similar transformations as larger filters without com-
putational overhead [88]. The Adam optimizer [83], the Nesterov Adam optimizer [83, 84],
stochastic gradient descent, and root mean squared propagation (RMSprop) are the CNN op-
timizers considered. SMAC3 chose the largest number of convolutional layers and the largest
dense layer width it considered.

3.3 High level tagger: Energy flow variables

Two high level taggers are trained to reduce the feature space and to compare to the CNN
performance. For the first tagger, ten high level observables (called 'E-flow') are constructed
using a method inspired by [89]. 10 linearly spaced rings are defined on the jet image with
diameters from 1 to 65 pixels. The pixel activations in each ring are summed. In all of the
figures, the observable corresponding to i^{th} ring will be refereed to as E_i. AdaBoost [90] from
scikit-learn [91] is used as the classifier. Adaboost takes a weighted sum of the outputs of base
classifiers to output a final classification. The base classifiers used are decision trees with a
maximum depth of 3. A maximum number of 200 decision trees are used.

Example distributions of the E_i observables are presented in Fig. 3. While it is possible
to see differences between the different types of color resonances in almost all the plots. The
most noticeable difference occurs in E_2. Here we see how the distribution of E_2 peaks closer to
zero for resonances that decay to quarks in the final state, such as Z' or C_μ, whereas resonances

that decay to gluonic final states, such as S_8 peak at larger values of E_2. Resonances that decay to a mixture of quarks and gluons, such as $q^\star$ peak at values larger than the purely quark and purely gluonic distributions. As expected, the energy distribution around a gluonic jet is more diffuse, and thus we see that E_2 is a good discriminator for distinguishing between quark and gluon final states. Similar patterns can also be observed for E_3 and E_4, possibly coming from the second subjet.

3.4 High level tagger: Jet pull

As discussed in Sec. 1, the jet pull angle [57] is an observable that has been demonstrated to be sensitive to the color representation of the new resonance [58–60]. The jet pull angle is computed from the subjet pull vector, which is a p_T weighted radial moment:

$$\sum_{i \in J} \frac{p_T^i |\vec{r}_i|}{p_T^J} \vec{r}_i, \tag{1}$$

where J is the given subjet, p_T^J is its transverse momentum, i iterates over the constituents of J, and $\vec{r}_i$ is the vector difference between the subjet axis and the ith constituent in rapidity-azimuthal angle space ($\Delta r_i = (\Delta y_i, \Delta \phi_i)$). The pull angle is defined for two subjets J_1 and J_2 as the angle formed between the pull vector of one of the jets and the vector difference between J_1 and J_2. Each jet has a corresponding pull angle, θ_{J_1,J_2} and θ_{J_2,J_1}.

The pull angle distribution is predicted to be more uniform when the two quarks from the resonance decay are not color-connected to each other and more peaked at zero if they are connected. The distributions of the two jet pull angles for each of the resonance models are shown in Fig. 4. As expected, the color singlets $Z' \to q\bar{q}$ and $H \to q\bar{q}$ have pull angles more closely distributed around 0, while the color octets $C_\mu \to gg$ and $S_8 \to gg$ have a more spread out distribution.

For each event, the two pull angles are used as high level features. The two angles are combined using the same AdaBoost algorithm as described in Sec. 3.3.

4 Results

For each binary classification task, 300,000 events are used with a train-test split of 80-20%. For the CNN, the training set is further sub-divided with a training-validation split of 80-20% and the CNN's loss function is weighted as described in Sec. 3.2.

4.1 Training results

The performance of the various methods and signal models are compared using significance improvement characteristic (SIC) curves and receiver operating characteristic (ROC) curves. For a given signal efficiency (true positive rate) ϵ, SIC curves show $\epsilon/\sqrt{\epsilon_b}$ and ROC curves show ϵ_b, where ϵ_b is the background efficiency (false positive rate). Note that all samples are BSM signals, but in each binary task, one is labeled as 'signal' and other as 'background'. The two classes are treated symmetrically in the loss function so this labeling scheme is irrelevant.

Figure 5 shows SIC and ROC curves for the three taggers in two different signal comparisons. The CNN has the strongest performance, followed by the E-flows and the jet pull. The curves on the left show stronger discrimination power than the ones on the right. This is expected, since the final state for the two BSM particles are different for the plots on the left and the same for the curves on the right. We also quote the peak SIC value, which is the highest value attained in the SIC curve, and area under curve (AUC), which is the area under
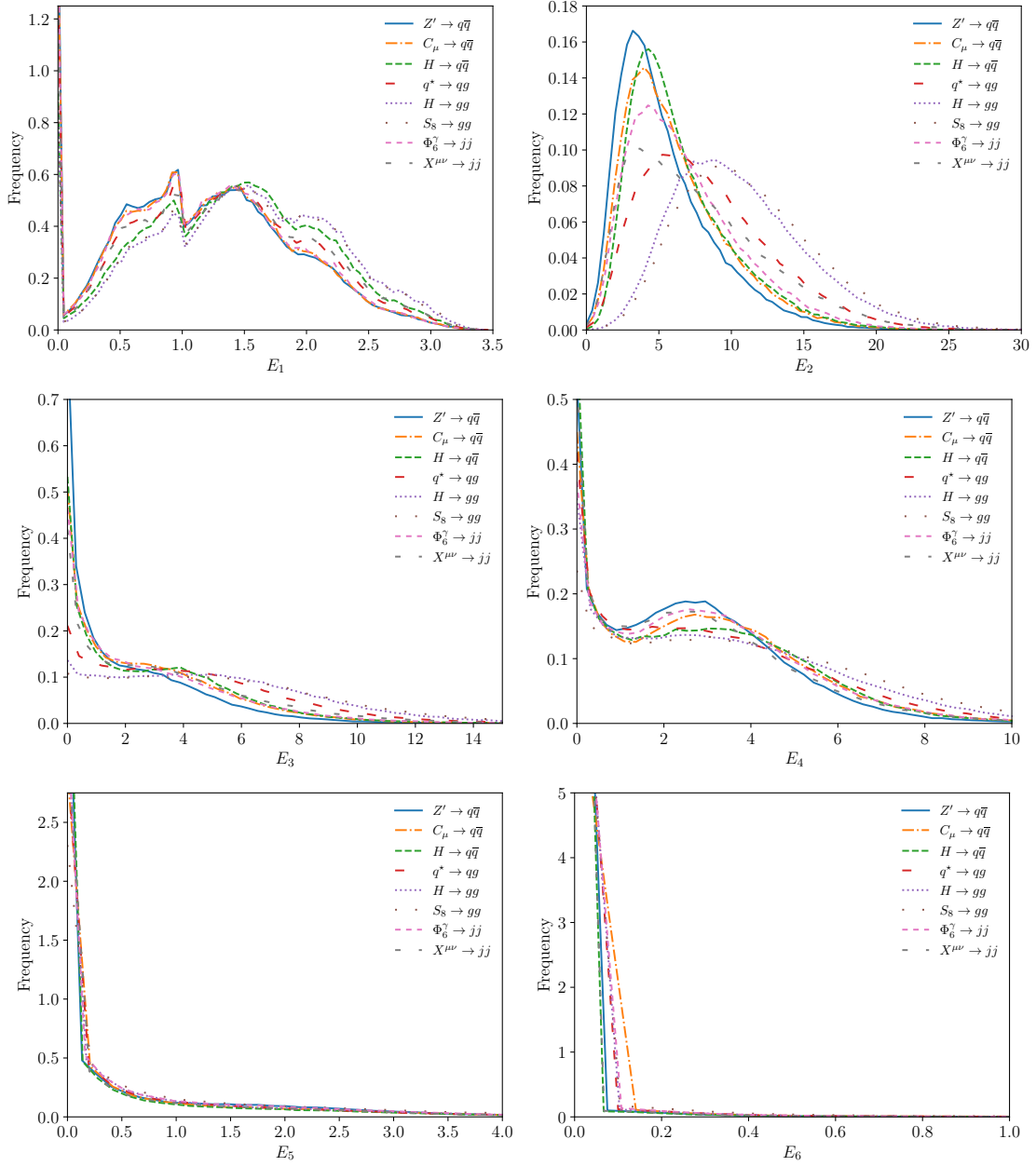
Figure 3: Distributions of the first six E-flow variables normalized to unity. The higher indices correspond to larger radii within the jet. Majority of distributions peak around 0, which shows that most of jets do not contain a lot of sub-jets. It interestingly occurs for E_1. Possible reason is that the leading subjet might occur in the pixels surrounding the center one, if the rest of subjets have high enough p_T.
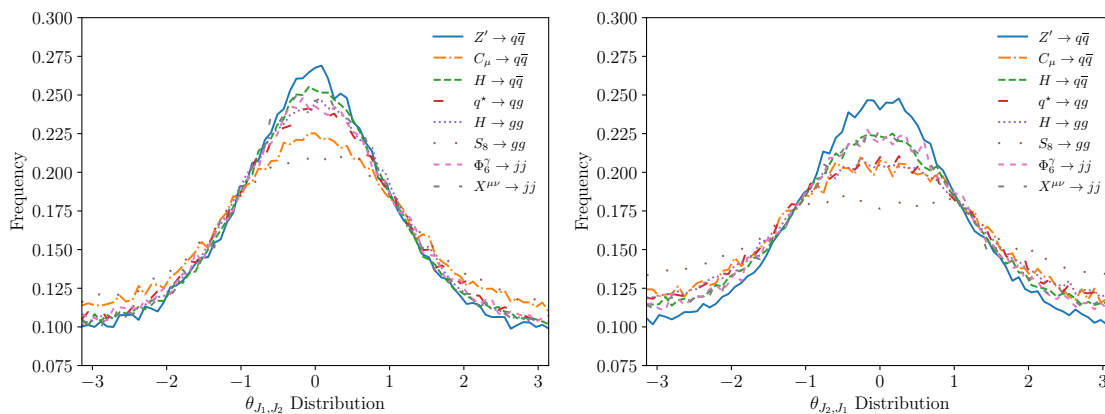
Figure 4: Distributions of both jet pull angles for each process.

the ROC curve. Both the peak SIC and the AUC can be used as single values which quantify performance.

We summarize the performance of the CNN and compare it to the performance of other high level taggers using the peak SIC and AUC values in Table 1. We see that the CNN performs just as well or better than the other observables in every combination of signals. The samples are ordered such that samples with similar final states are close together. The E-flow observables and CNN perform well when the final states are different. Jet pull is not a strong classifier for any combination. In general, it is easier to distinguish BSM particles which decay to pairs of gluons more easily from those that decay into quarks. On the other hand, BSM particles that decay into similar final states are harder to distinguish. For example, the AUC values of the $C_\mu \to q\bar{q}$ versus $H \to q\bar{q}$ are lower compared to $C_\mu \to q\bar{q}$ versus $H \to gg$. The AUC values further reduce when we compare BSM particles that decay into a similar final state (quarks or gluons) and also have the same spin quantum numbers. For example, $C_\mu \to q\bar{q}$ versus $Z' \to q\bar{q}$ has lower AUC values than $C_\mu \to q\bar{q}$ versus $H \to q\bar{q}$. C_μ and Z' are both vector resonances, but differ in the color quantum numbers, the former being a color octet and the later a color singlet. The only distinguishing feature therefore the small differences in radiation patterns that arise due to different color flow in these two samples. Similar observations can be made about $H \to gg$ versus $S_8 \to gg$ which differ only in their color representation. Finally we note that for the spin-2 resonance $X^{\mu\nu}$, the final state comprises of $X^{\mu\nu} \to gg$ and $X^{\mu\nu} \to q\bar{q}$ in roughly equal proportions. As expected this degrades the performance of the CNN, also indicating that the CNN performs better discriminating different final states, but not as well with distinguishing the spin quantum numbers of the BSM resonance.

4.2 Pearson correlation coefficient

To help understand how jet images are being used to classify the different BSM resonances, the Pearson Correlation Coefficient (PCC) is computed between the true labels and the pixel intensities for each pixel over all images. A PCC image is constructed, where each pixel corresponds to its correlation coefficient. Specifically, let I_{ij} be the distribution of the activations for some pixel (i, j), in the transformed rapidity-azimuthal angle space for every event; let y be the corresponding distribution of true labels. Then the PCC is calculated between these two distributions for that pixel. This is the intensity for that pixel in the new PCC image.

PCC images for the two combinations used as examples in Sec. 3 are shown in Fig. 6. On the left, there are two orange circles corresponding to the two prongs of the gg dijet, and two blue regions corresponding to the prongs of the $q\bar{q}$ dijet. These regions are a method for visualizing the discriminatory information in each pixel when using a linear model. The PCC images can
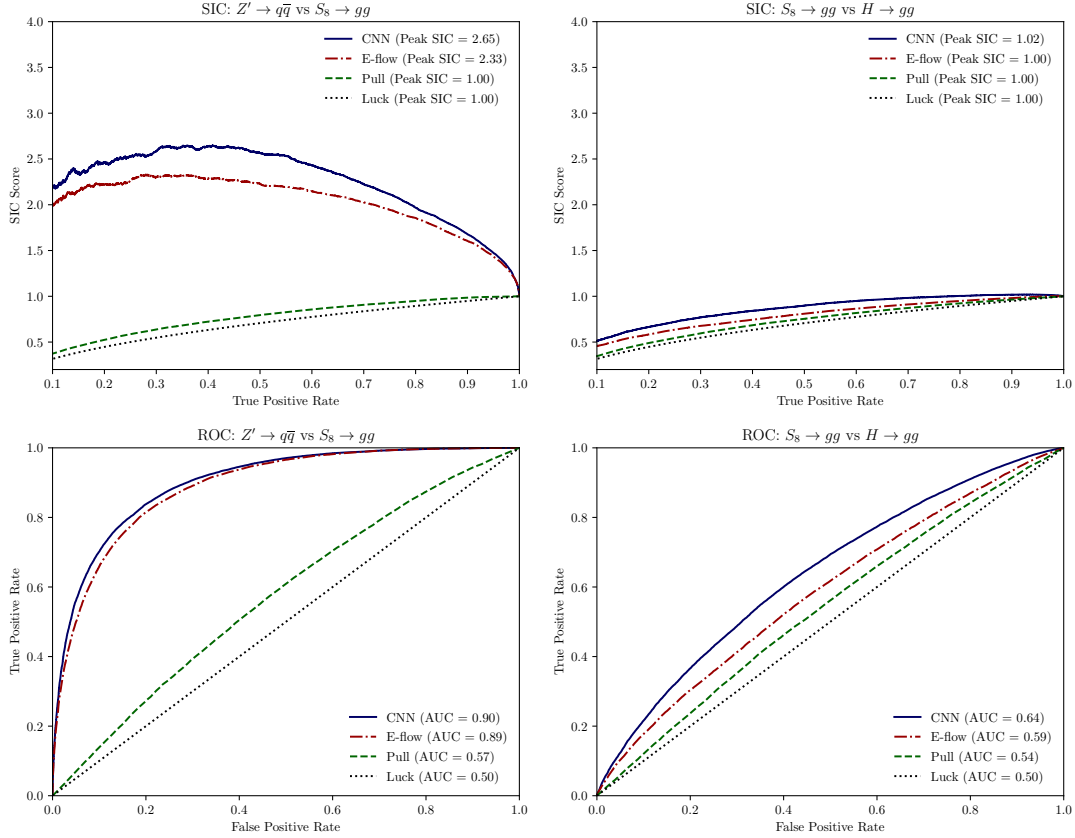
Figure 5: Significance improvement characteristic (SIC) curves (top) and Receiver Operating Characteristic (ROC) curves (bottom). CNN represents the convolutional neural network trained on the low level jet images. E-flows represents AdaBoost trained on the high level telescoping jet features. Pull represents AdaBoost trained on the high level jet pull angle angle features. Luck represents the scores that would be achieved by a random classifier. The classification task is $Z' \to q\bar{q}$ from $S_8 \to gg$ on the left and $S_8 \to gg$ from $H \to gg$ on the right.

be used as a computational inspiration for new classifiers. For example, the concentric circles of the E-flows should be effective when they separate the orange regions from the blue regions. On the right, there is a more difficult classification task as the final states are the same. Yet there is still blue and orange regions which could be used as inspirations for new classifiers. For completeness, we provide the PCCs constructed for all combinations of samples are found in the appendix in Fig. 9. Remember that PCC only capture linear correlations between individual pixels and the output, not non-linear correlations or correlations that are functions of multiple pixels.

5 Network dependence on the setup

5.1 Generator dependence

The features of the radiation pattern inside jets may not be well-modeled by the simulation. This is especially true for the case of colorflow information [59, 60]. While this may be a motivation for an analysis to differentiate various color configurations, it may also have an impact on the quantitative and qualitative conclusions from Sec. 4. The goal of this section is

Table 1: Peak SIC and ROC AUC achieved when comparing any two samples for each model. CNN, E-flow, and pull represent the three taggers as defined in Figure 5. The peak SICs and ROC AUC's which are largest for their combination of samples are shown in bold. The CNN has the strongest performance for all combinations. For all combinations with different final states, the E-flow has better performance than pull.

	Model	$Z' \to q\bar{q}$		$C_\mu \to q\bar{q}$		$H \to q\bar{q}$		$q^\star \to qg$		$H \to gg$		$S_8 \to gg$		$\Phi_6^\gamma \to jj$		$X^{\mu\nu} \to jj$	
	CNN	SIC	AUC	**1.05**	**0.67**	**1.16**	**0.72**	**1.36**	**0.78**	**2.31**	**0.89**	**2.65**	**0.90**	**1.12**	**0.71**	**1.02**	**0.67**
$Z' \to q\bar{q}$	E-flow			1.00	0.61	1.89	0.86	1.27	0.76	1.89	0.86	2.33	0.89	1.02	0.64	1.02	0.67
	Pull			1.00	0.54	1.00	0.53	1.00	0.53	1.00	0.53	1.00	0.57	1.00	0.53	1.00	0.52
	CNN	**1.05**	**0.67**	SIC	AUC	**1.24**	**0.76**	**1.24**	**0.76**	**1.73**	**0.85**	**2.03**	**0.88**	**1.00**	**0.62**	**1.03**	**0.66**
$C_\mu \to q\bar{q}$	E-flow	1.00	0.61			1.01	0.62	1.07	0.68	1.49	0.81	1.70	0.84	1.00	0.57	1.01	0.63
	Pull	1.00	0.54			1.00	0.54	1.00	0.52	1.00	0.52	1.00	0.53	1.00	0.52	1.00	0.53
	CNN	**1.16**	**0.72**	**1.24**	**0.76**	SIC	AUC	**1.29**	**0.77**	**1.73**	**0.85**	**2.20**	**0.89**	**1.37**	**0.78**	**1.13**	**0.72**
$H \to q\bar{q}$	E-flow	1.04	0.65	1.01	0.62			1.12	0.71	1.44	0.81	1.74	0.84	1.04	0.65	1.03	0.65
	Pull	1.00	0.52	1.00	0.54			1.00	0.52	1.00	0.52	1.00	0.56	1.00	0.51	1.00	0.51
	CNN	**1.36**	**0.78**	**1.24**	**0.76**	**1.29**	**0.77**	SIC	AUC	**1.05**	**0.68**	**1.12**	**0.71**	**1.04**	**0.68**	**1.03**	**0.67**
$q^\star \to qg$	E-flow	1.27	0.76	1.07	0.68	1.12	0.71			1.02	0.67	1.12	0.72	1.02	0.64	1.00	0.63
	Pull	1.00	0.53	1.00	0.52	1.00	0.52			1.00	0.50	1.00	0.54	1.00	0.51	1.00	0.51
	CNN	**2.31**	**0.89**	**1.73**	**0.85**	**1.73**	**0.85**	**1.05**	**0.68**	SIC	AUC	**1.02**	**0.64**	**1.59**	**0.79**	**2.20**	**0.73**
$H \to gg$	E-flow	1.89	0.86	1.49	0.81	1.44	0.81	1.02	0.67			1.00	0.59	1.52	0.78	1.58	0.72
	Pull	1.00	0.53	1.00	0.52	1.00	0.52	1.00	0.50			1.00	0.54	1.00	0.51	1.00	0.52
	CNN	**2.65**	**0.90**	**2.03**	**0.88**	**2.20**	**0.89**	**1.12**	**0.71**	**1.02**	**0.64**	SIC	AUC	**1.80**	**0.81**	**2.76**	**0.78**
$S_8 \to gg$	E-flow	2.33	0.89	1.70	0.84	1.74	0.84	1.12	0.72	1.00	0.59			1.86	0.81	2.16	0.76
	Pull	1.00	0.57	1.00	0.53	1.00	0.56	1.00	0.54	1.00	0.54			1.00	0.55	1.00	0.55
	CNN	**1.12**	**0.71**	**1.00**	**0.62**	**1.37**	**0.78**	**1.04**	**0.68**	**1.59**	**0.79**	**1.80**	**0.81**	SIC	AUC	**1.06**	**0.68**
$\Phi_6^\gamma \to jj$	E-flow	1.02	0.64	1.00	0.57	1.04	0.65	1.02	0.64	1.52	0.78	1.86	0.81			1.01	0.61
	Pull	1.00	0.53	1.00	0.52	1.00	0.51	1.00	0.51	1.00	0.51	1.00	0.55			1.00	0.51
	CNN	**1.02**	**0.67**	**1.03**	**0.66**	**1.13**	**0.72**	**1.03**	**0.67**	**2.20**	**0.73**	**2.76**	**0.78**	**1.06**	**0.68**	SIC	AUC
$X^{\mu\nu} \to jj$	E-flow	1.02	0.67	1.01	0.63	1.03	0.65	1.00	0.63	1.58	0.72	2.16	0.76	1.01	0.61		
	Pull	1.00	0.52	1.00	0.53	1.00	0.51	1.00	0.51	1.00	0.52	1.00	0.55	1.00	0.51		

to investigate the impact of simulation variations on the classification performance. Note that the performance of a network trained on one sample and tested on another may overestimate the impact of a simulation mismodeling - a tagger may identify the correct features even though they may be differently expressed in two different simulations. This was illustrated for the case of quark/gluon tagging in Ref. [92].

Five alternative simulations are prepared to study the modeling dependence. Three of these samples use the ATLAS A14 tune [93] and final state shower variations[2] (A14 Var2±) instead of the default Monash tune [94]. The most important parameter variation in the Var2 tune is the final state shower α_s, which is varied by about 10%. Two additional samples use alternative color reconnection models. One of these uses a simple where gluons can be shifted to reduce the total string length (called 'Gluon move') [95]. The other sample is more sophisticated and incorporates aspects of QCD (called 'QCD inspired') [96].

Figure 7 shows the impact of the model variations on the CNN performance. Overall, the variation is smaller than the spread in performance from Fig. 5, which builds confidence in the results from the previous section. The biggest differences are observed for the variations that modify the final state shower α_s.

---

[2]The most important difference between A14 and A14 Var2 is an increase or decrease in the final state shower α_s by about 10%.
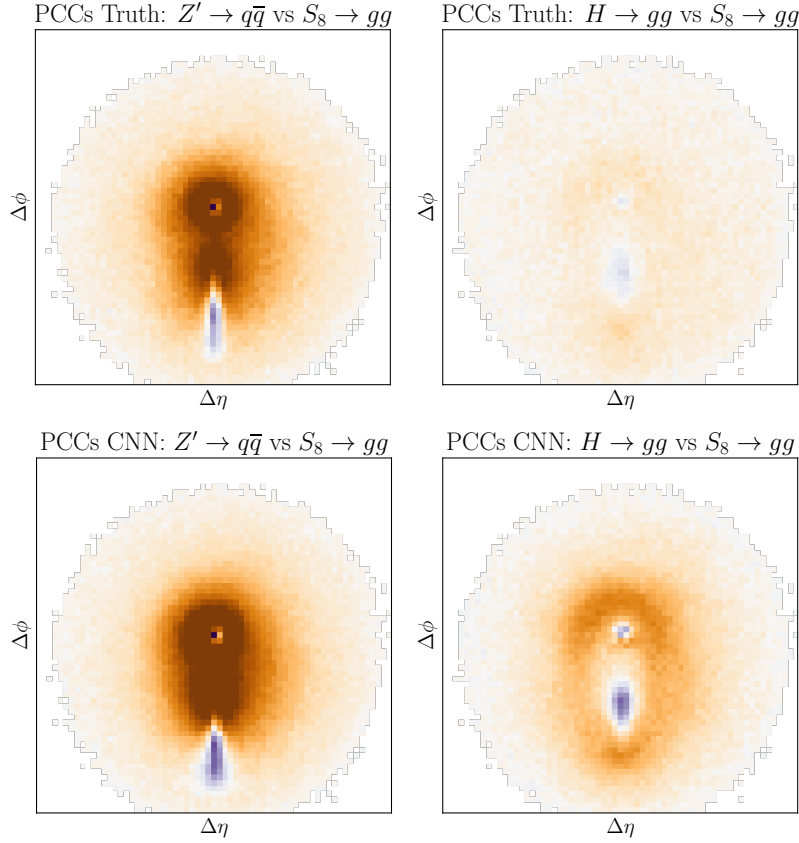
Figure 6: Top: Correlations of the pixel values with true labels. Bottom: Correlation of the pixels with the network labels. Left: PCC image comparing $Z' \to q\bar{q}$ from $S_8 \to gg$. Right: PCC image comparing $S_8 \to gg$ from $H \to gg$. Orange pixels correlate more the first process; blue pixels correlate more with the second. Note that the processes on the right have the same final state. The correlations of CNN and truth are identical in distribution, with CNN being much stronger.

5.2 Charged dependence

Thus far, the CNNs have been trained using images built from all final state particles. However, the energy and position resolution from a real calorimeter may not be suitable for this purpose. This section explores the impact of only using charged particles, which should be close to realistic information available from charged particle tracks reconstructed from the ATLAS and CMS inner detectors. Figure 8 compares the all-particles and charged-particles versions of the three taggers considered in Sec. 4. In general, the performance of the charged-only taggers are worse than the all-particles taggers, but the degradation is often smaller than the spread in performance between taggers and the order of tagger performance is preserved.

5.3 Image pixelation dependence

Various image pixelations of 25×25, 50×50, 65×65, and 100×100 were tried. It was found that increasing image pixelation leads to better results. For example, if performance was measured as the false positive rate (FPR) of the CNN when the true positive rate (TPR) is 0.5, there is a performance increase of $\sim 60\%$ when increasing image pixelation from 25×25 to 50×50. However, if the image pixelation is increased from 50×50 to 100×100, there is only an increase in performance of $\sim 20\%$. Additionally, larger images require larger networks and
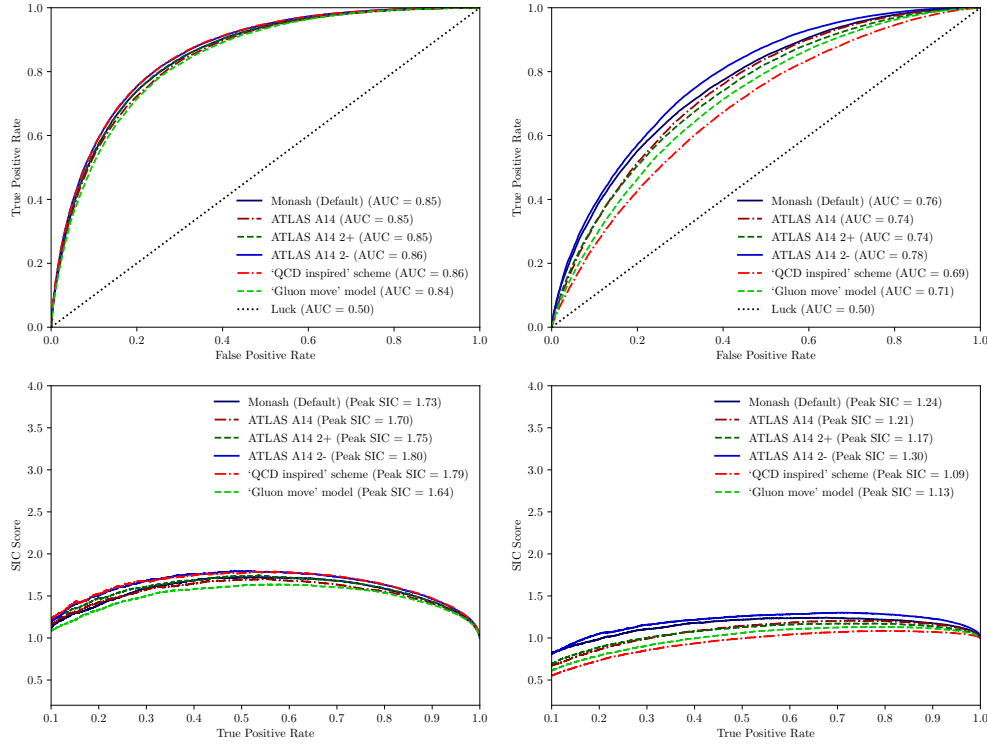
Figure 7: Left: ROC and SIC curves for $C_\mu \to q\bar{q}$ vs $H \to gg$. Right: ROC and SIC curves for $C_\mu \to q\bar{q}$ vs $H \to q\bar{q}$. All plots contain all different generators that have been checked for generalization of the study.

Table 2: Performance of different pre-processing steps, measured as the FPR of the CNN when the true positive rate TPR is 0.5. The columns show normalization techniques and the rows show whether the images are logged or not.

	l_1 normalization	l_2 normalization	Z-score normalization	Min-Max Scaling	No Normalization
No Logging	0.13	0.13	0.14	0.18	0.13
Logging	0.12	0.13	0.13	0.15	0.12

take longer to train. The image pixelation of 65×65 was chosen as a compromise between speed, network size, and performance. The hyper parameter optimization, as described in 3.2, was performed after the optimal pixel size was chosen. Future studies that rely on individual jet constituents may be able to benefit from other approaches (see Sec. 1) that do not rely on pixelation. All ATLAS and CMS detectors are segmented so this benefit will be most pronounced when the pixelation is larger than the detector pixelation as is nearly always the case for tracking detectors.

5.4 Preprocessing dependence

The impact of the preprocessing steps of rotating, normalizing, and logging the images was studied. The FPR of the CNN when the true positive rate TPR is 0.5 was used to measure performance. The rotation step, initially described in 3.1, did not change performance. Four normalization techniques: l_1 normalization, l_2 normalization, Z-score normalization (also called standardization), and min-max scaling were tried. Each of these on logged images were tried as well. The results are shown in Tab. 2 below. Based on the FPR values from the table, all jet-images used were rotated, logged, and l_1 normalized.
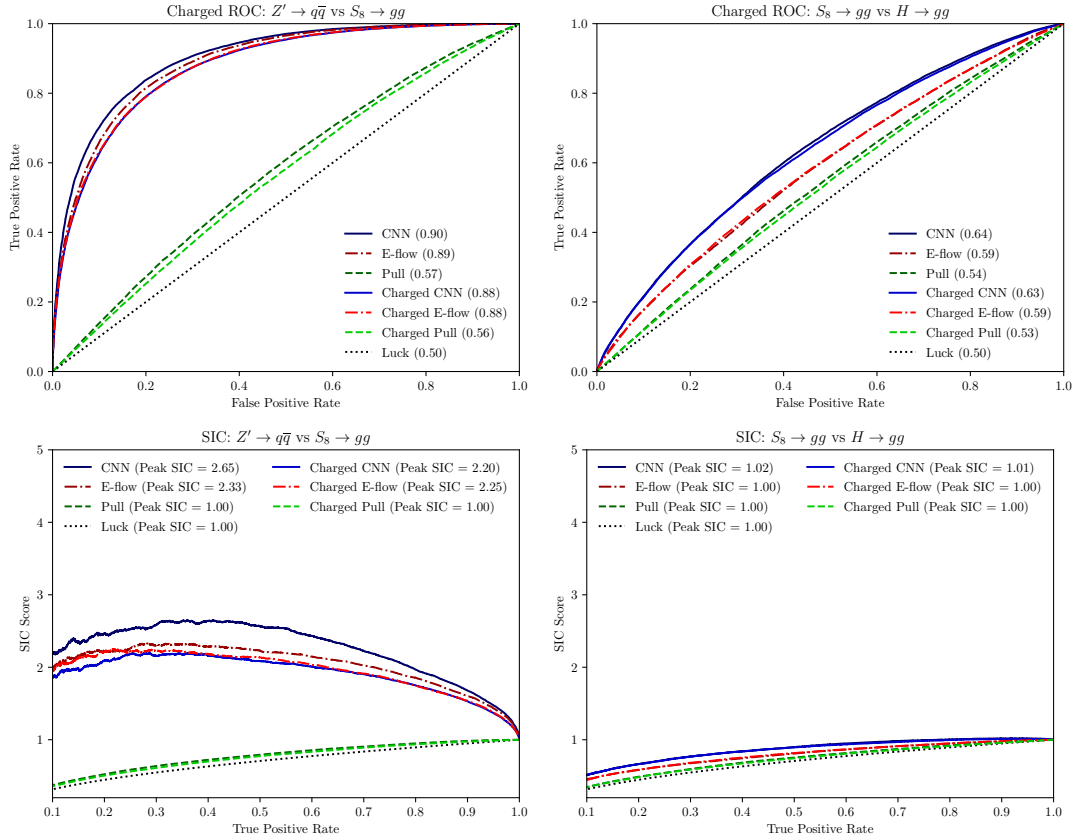
Figure 8: ROC and SIC curves using all-particles (default) or charged-only information for $Z' \to q\bar{q}$ from $S_8 \to gg$ and $S_8 \to gg$ from $H \to gg$ on the bottom.

5.5 Jet grooming dependence

Jets are trimmed by dropping subjets which have a transverse momentum fraction less than 5% of the jet. Grooming the jet in this way has a significant impact on both the CNN and high level taggers. Consider for example comparison of the color singlet Z' with the color octet coloron (C_μ). As mentioned earlier, these two resonances differ only in their color representation. The color flow of their decay products would be different resulting in different radiation patterns. However, grooming the jet would result in some information about these radiation patterns to be lost. On the contrary, we found that the performance of the CNN was better when using trimmed jets than when using untrimmed jets. This suggests that a general approach might not be sufficient and one might need to use a different CNN with possibly a different architecture to pick out these minor differences. Further, it may also happen that there is an optimal transverse momentum fraction cut that would aid in discriminating resonances of different color representation. These considerations are left for future deliberation and work.

6 Conclusions

Hypothesized BSM particles that interact with quarks and gluons may be light enough so that they are produced with large transverse momentum at the LHC. When decaying to SM quarks or gluons, the large transverse momentum implies production of a single fat jet. Analyzing the sub structure of this fat jet can help distinguish BSM and SM particles.

In this paper, we looked at a list of possible BSM particles with a variety of spins, charges and color quantum numbers that decay to pairs of SM particles comprising quarks or gluons. Deep learning was used to improve upon techniques for distinguishing such dijets based on their substructure. Jet images were used to train a CNN to classify and distinguish between these various types of dijets. The capabilities of CNNs trained on jet images were compared to other classifiers that used color information, such as Energy Flow Observables and jet pull. Table 1 demonstrates that the CNN classifier consistently outperforms traditional observables in this task and thus provides a useful method to categorize new particles and expand the jet substructure toolkit. The classifier performed well when distinguishing between BSM particles that decay to different final states (e.g. $Z' \rightarrow q\bar{q}$ against $H \rightarrow gg$), and also did reasonably well when trying to differentiate particles with different spin. On the other hand, its performance was not the strongest when it came to identifying only differences in color quantum numbers.

Further improvements in these classification tasks might be possible if using a point cloud representation, rather than a CNN. Another avenue that could be explored is the use of multi-class classification, which would allow us to combine various options into one model without qualitatively affecting performance.

Acknowledgments

A Appendix
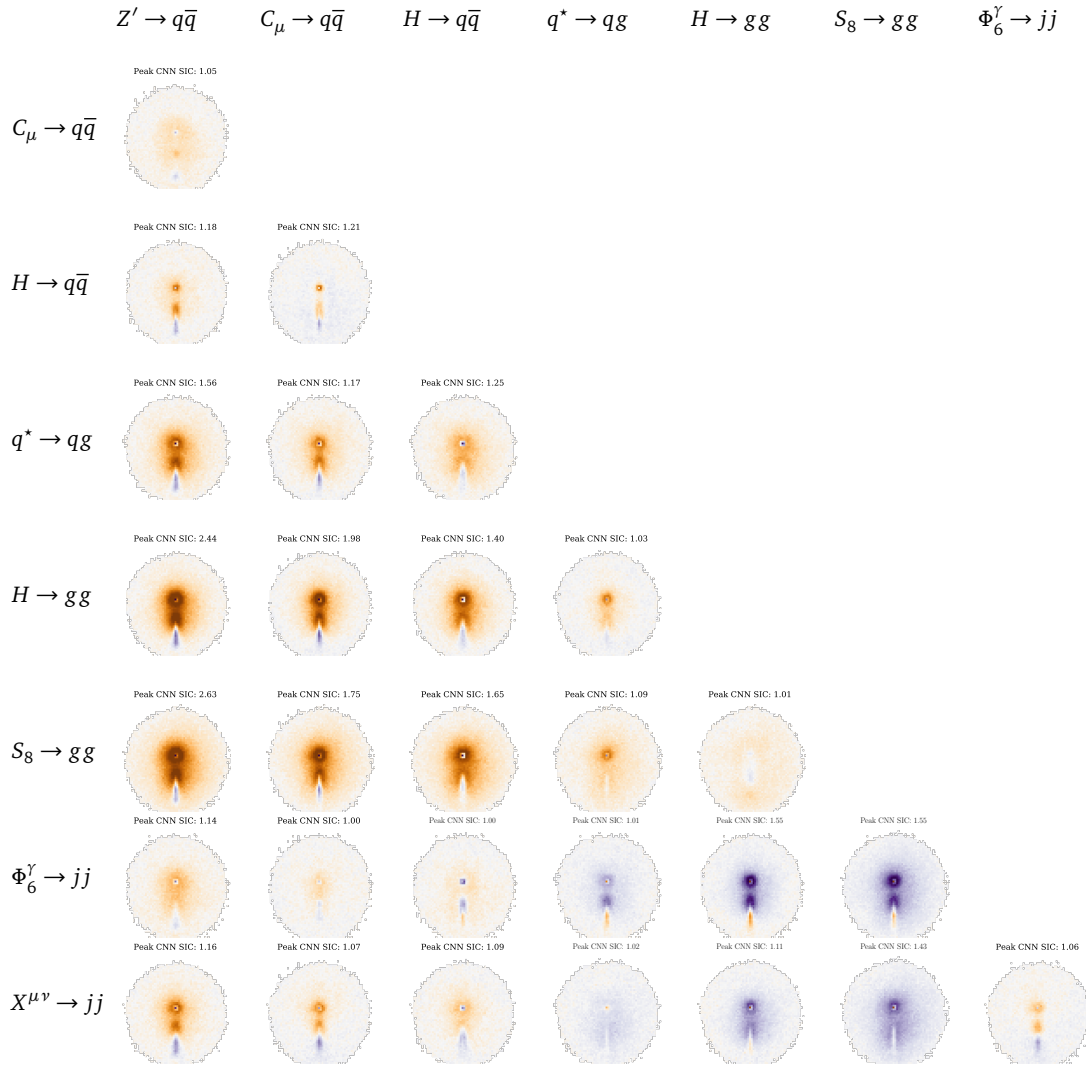
Figure 9: All PCC Images for pairwise comparisons.

References

[1] ATLAS collaboration: M. Aaboud et al., *Search for light resonances decaying to boosted quark pairs and produced in association with a photon or a jet in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **788**, 316 (2019), doi:10.1016/j.physletb.2018.09.062.

[2] ATLAS collaboration: M. Aaboud et al., *Search for a heavy Higgs boson decaying into a Z boson and another heavy Higgs boson in the $\ell\ell bb$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **783**, 392 (2018), doi:10.1016/j.physletb.2018.07.006.

[3] ATLAS collaboration, *Search for boosted resonances decaying to two b-quarks and produced in association with a jet at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2018-052, CERN, Geneva, Switzerland (2018), https://cds.cern.ch/record/2649081/.

[4] ATLAS collaboration: M. Aaboud et al., *A search for resonances decaying into a Higgs boson and a new particle X in the $XH \rightarrow qqbb$ final state with the ATLAS detector*, Phys. Lett. B **779**, 24 (2018), doi:10.1016/j.physletb.2018.01.042.

[5] ATLAS collaboration: G. Aad et al., *Dijet resonance search with weak supervision using $\sqrt{s} = 113$ TeV collisions in the ATLAS detector*, Phys. Rev. Lett. **125**, 131801 (2020), doi:10.1103/PhysRevLett.125.131801.

[6] CMS collaboration: A. M. Sirunyan et al., *Search for low mass vector resonances decaying to quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Rev. Lett. **119**, 111802 (2017), doi:10.1103/PhysRevLett.119.111802.

[7] CMS collaboration: A. M. Sirunyan et al., *Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV*, J. High Energy Phys. **01**, 097 (2018), doi:10.1007/JHEP01(2018)097.

[8] CMS collaboration: A. M. Sirunyan et al., *Search for low-mass quark-antiquark resonances produced in association with a photon at $\sqrt{s} = 13$ TeV*, Phys. Rev. Lett. **123**, 231803 (2019), doi:10.1103/PhysRevLett.123.231803.

[9] CMS collaboration: A. M. Sirunyan et al., *Search for low-mass resonances decaying into bottom quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Rev. D **99**, 012005 (2019), doi:10.1103/PhysRevD.99.012005.

[10] A. J. Larkoski, I. Moult and B. Nachman, *Jet substructure at the Large Hadron Collider: A review of recent advances in theory and machine learning*, Phys. Rep. **841**, 1 (2020), doi:10.1016/j.physrep.2019.11.001.

[11] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *Jet-images – Deep learning edition*, J. High Energy Phys. **07**, 069 (2016), doi:10.1007/JHEP07(2016)069.

[12] L. G. Almeida, M. Backovic, M. Cliche, S. J. Lee and M. Perelstein, *Playing tag with ANN: Boosted top identification with pattern recognition*, J. High Energy Phys. **07**, 86 (2015), doi:10.1007/JHEP07(2015)086.

[13] P. T. Komiske, E. M. Metodiev and M. D. Schwartz, *Deep learning in color: Towards automated quark/gluon jet discrimination*, J. High Energy Phys. **01**, 110 (2017), doi:10.1007/JHEP01(2017)110.

[14] P. Baldi, K. Bauer, C. Eng, P. Sadowski and D. Whiteson, *Jet substructure classification in high-energy physics with deep neural networks*, Phys. Rev. D **93**, 094034 (2016), doi:10.1103/PhysRevD.93.094034.

[15] J. Barnard, E. N. Dawe, M. J. Dolan and N. Rajcic, *Parton shower uncertainties in jet substructure analyses with deep neural networks*, Phys. Rev. D **95**, 014018 (2017), doi:10.1103/PhysRevD.95.014018.

[16] G. Kasieczka, T. Plehn, M. Russell and T. Schell, *Deep-learning top taggers or the end of QCD?*, J. High Energy Phys. **05**, 006 (2017), doi:10.1007/JHEP05(2017)006.

[17] J. Lin, M. Freytsis, I. Moult and B. Nachman, *Boosting $H \rightarrow b\bar{b}$ with machine learning*, J. High Energy Phys. **10**, 101 (2018), doi:10.1007/JHEP10(2018)101.

[18] Y.-L. Chung, S.-C. Hsu and B. Nachman, *Disentangling boosted Higgs boson production modes with machine learning*, J. Instrum. **16**, P07002 (2021), doi:10.1088/1748-0221/16/07/P07002.

[19] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban and D. Whiteson, *Jet flavor classification in high-energy physics with deep neural networks*, Phys. Rev. D **94**, 112002 (2016), doi:10.1103/PhysRevD.94.112002.

[20] ATLAS collaboration, *Identification of jets containing b-hadrons with recurrent neural networks at the ATLAS experiment*, Tech. Rep. ATL-PHYS-PUB-2017-003, CERN, Geneva, Switzerland (2017), https://cds.cern.ch/record/2255226.

[21] CMS collaboration, *Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13 TeV with phase 1 CMS detector*, Tech. Rep. CERN-CMS-DP-2018-058, CERN, Geneva, Switzerland (2018), https://cds.cern.ch/record/2646773.

[22] J. Pearkes, W. Fedorko, A. Lister and C. Gay, *Jet constituents for deep neural network based top quark tagging*, (arXiv preprint) doi:10.48550/arXiv.1704.02124.

[23] S. Egan, W. Fedorko, A. Lister, J. Pearkes and C. Gay, *Long short-term memory (LSTM) networks with jet constituents for boosted top tagging at the LHC*, (arXiv preprint) doi:10.48550/arXiv.1711.09059.

[24] G. Louppe, K. Cho, C. Becot and K. Cranmer, *QCD-aware recursive neural networks for jet physics*, J. High Energy Phys. **01**, 057 (2019), doi:10.1007/JHEP01(2019)057.

[25] T. Cheng, *Recursive neural networks in quark/gluon tagging*, Comput. Softw. Big Sci. **2**, 3 (2018), doi:10.1007/s41781-018-0007-y.

[26] I. Henrion, J. Brehmer, J. Bruna, K. Cho and K. Cranmer, *Neural message passing for jet physics*, in *Workshop on deep learning for physical sciences*, Long Beach, California, USA (2017), https://dl4physicalsciences.github.io/files/nips_dlps_2017_29.pdf.

[27] H. Qu and L. Gouskos, *Jet tagging via particle clouds*, Phys. Rev. D **101**, 056019 (2020), doi:10.1103/PhysRevD.101.056019.

[28] E. A. Moreno et al., *JEDI-net: A jet identification algorithm based on interaction networks*, Eur. Phys. J. C **80**, 58 (2020), doi:10.1140/epjc/s10052-020-7608-4.

[29] E. Bernreuther, T. Finke, F. Kahlhoefer, M. Krämer and A. Mück, *Casting a graph net to catch dark showers*, SciPost Phys. **10**, 046 (2021), doi:10.21468/SciPostPhys.10.2.046.

[30] V. Mikuni and F. Canelli, *ABCNet: An attention-based method for particle tagging*, Eur. Phys. J. Plus **135**, 463 (2020), doi:10.1140/epjp/s13360-020-00497-3.

[31] F. A. Dreyer and H. Qu, *Jet tagging in the Lund plane with graph networks*, J. High Energy Phys. **03**, 52 (2021), doi:10.1007/JHEP03(2021)052.

[32] P. T. Komiske, E. M. Metodiev and J. Thaler, *Energy flow networks: Deep sets for particle jets*, J. High Energy Phys. **01**, 121 (2019), doi:10.1007/JHEP01(2019)121.

[33] V. Mikuni and F. Canelli, *Point cloud transformers applied to collider physics*, Mach. Learn.: Sci. Technol. **2**, 035027 (2021), doi:10.1088/2632-2153/ac07f6.

[34] M. J. Dolan and A. Ore, *Equivariant energy flow networks for jet tagging*, Phys. Rev. D **103**, 074022 (2021), doi:10.1103/PhysRevD.103.074022.

[35] K. Datta and A. Larkoski, *How much information is in a jet?*, J. High Energy Phys. **06**, 073 (2017), doi:10.1007/JHEP06(2017)073.

[36] K. Datta and A. J. Larkoski, *Novel jet observables from machine learning*, J. High Energy Phys. **03**, 086 (2018), doi:10.1007/JHEP03(2018)086.

[37] K. Datta, A. Larkoski and B. Nachman, *Automating the construction of jet observables with machine learning*, Phys. Rev. D **100**, 095016 (2019), doi:10.1103/PhysRevD.100.095016.

[38] P. T. Komiske, E. M. Metodiev and J. Thaler, *Energy flow polynomials: A complete linear basis for jet substructure*, J. High Energy Phys. **04**, 013 (2018), doi:10.1007/JHEP04(2018)013.

[39] A. Chakraborty, S. H. Lim, M. M. Nojiri and M. Takeuchi, *Neural network-based top tagger with two-point energy correlations and geometry of soft emissions*, J. High Energy Phys. **07**, 111 (2020), doi:10.1007/JHEP07(2020)111.

[40] A. Butter, G. Kasieczka, T. Plehn and M. Russell, *Deep-learned top tagging with a Lorentz layer*, SciPost Phys. **5**, 028 (2018), doi:10.21468/SciPostPhys.5.3.028.

[41] A. Chakraborty, S. H. Lim and M. M. Nojiri, *Interpretable deep learning for two-prong jet classification with jet spectra*, J. High Energy Phys. **07**, 135 (2019), doi:10.1007/JHEP07(2019)135.

[42] F. A. Dreyer, G. P. Salam and G. Soyez, *The Lund jet plane*, J. High Energy Phys. **12**, 064 (2018), doi:10.1007/JHEP12(2018)064.

[43] M. Feickert and B. Nachman, *A living review of machine learning for particle physics*, (arXiv preprint) doi:10.48550/arXiv.2102.02770.

[44] J. Cogan, M. Kagan, E. Strauss and A. Schwarztman, *Jet-images: Computer vision inspired techniques for jet tagging*, J. High Energy Phys. **02**, 118 (2015), doi:10.1007/JHEP02(2015)118.

[45] R. Kogler et al., *Jet substructure at the Large Hadron Collider*, Rev. Mod. Phys. **91**, 045003 (2019), doi:10.1103/RevModPhys.91.045003.

[46] A. Altheimer et al., *Jet substructure at the Tevatron and LHC: New results, new tools, new benchmarks*, J. Phys. G: Nucl. Part. Phys. **39**, 063001 (2012), doi:10.1088/0954-3899/39/6/063001.

[47] A. Altheimer et al., *Boosted objects and jet substructure at the LHC. Report of BOOST2012, held at IFIC Valencia, 23rd-27th of July 2012*, Eur. Phys. J. C **74**, 2792 (2014), doi:10.1140/epjc/s10052-014-2792-8.

[48] S. Marzani, G. Soyez, and M. Spannowsky, *Looking inside jets: An introduction to jet substructure and boosted-object phenomenology*, Springer, Cham, Switzerland, ISBN 9783030157081 (2019), doi:10.1007/978-3-030-15709-8.

[49] A. Abdesselam et al., *Boosted objects: A probe of beyond the Standard Model physics*, Eur. Phys. J. C **71**, 1661 (2011), doi:10.1140/epjc/s10052-011-1661-y.

[50] R. D. Field and R. P. Feynman, *A parametrization of the properties of quark jets*, Nucl. Phys. B **136**, 1 (1978), doi:10.1016/0550-3213(78)90015-9.

[51] D. Krohn, M. D. Schwartz, T. Lin and W. J. Waalewijn, *Jet charge at the LHC*, Phys. Rev. Lett. **110**, 212001 (2013), doi:10.1103/PhysRevLett.110.212001.

[52] ATLAS collaboration: G. Aad et al., *Measurement of jet charge in dijet events from $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector*, Phys. Rev. D **93**, 052003 (2016), doi:10.1103/PhysRevD.93.052003.

[53] CMS collaboration: A. M. Sirunyan et al., *Measurements of jet charge with dijet events in pp collisions at $\sqrt{s} = 8$ TeV*, J. High Energy Phys. **10**, 131 (2017), doi:10.1007/JHEP10(2017)131.

[54] ATLAS collaboration: G. Aad et al., *A new method to distinguish hadronically decaying boosted Z bosons from W bosons using the ATLAS detector*, Eur. Phys. J. C **76**, 238 (2016), doi:10.1140/epjc/s10052-016-4065-1.

[55] K. Fraser and M. D. Schwartz, *Jet charge and machine learning*, J. High Energy Phys. **10**, 093 (2018), doi:10.1007/JHEP10(2018)093.

[56] Y.-C. J. Chen, C.-W. Chiang, G. Cottin and D. Shih, *Boosted W/Z tagging with jet charge and deep learning*, Phys. Rev. D **101**, 053001 (2020), doi:10.1103/PhysRevD.101.053001.

[57] J. Gallicchio and M. D. Schwartz, *Seeing in color: Jet superstructure*, Phys. Rev. Lett. **105**, 022001 (2010), doi:10.1103/PhysRevLett.105.022001.

[58] D0 collaboration: V. Abazov et al., *Measurement of color flow in $t\bar{t}$ events from $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Rev. D **83**, 092002 (2011), doi:10.1103/PhysRevD.83.092002.

[59] ATLAS collaboration: G. Aad et al., *Measurement of colour flow with the jet pull angle in $t\bar{t}$ events using the ATLAS detector at $\sqrt{s} = 8$ TeV*, Phys. Lett. B **750**, 475 (2015), doi:10.1016/j.physletb.2015.09.051.

[60] ATLAS collaboration: M. Aaboud et al., *Measurement of colour flow using jet-pull observables in $t\bar{t}$ events with the ATLAS experiment at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **78**, 847 (2018), doi:10.1140/epjc/s10052-018-6290-2.

[61] R. S. Chivukula, K. A. Mohan, D. Sengupta and E. H. Simmons, *Characterizing boosted dijet resonances with jet energy correlators*, J. High Energy Phys. **03**, 133 (2018), doi:10.1007/JHEP03(2018)133.

[62] A. Buckley, G. Callea, A. Larkoski and S. Marzani, *An optimal observable for color singlet identification*, SciPost Phys. **9**, 026 (2020), doi:10.21468/SciPostPhys.9.2.026.

[63] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr and B. Fuks, *FeynRules 2.0 – A complete toolbox for tree-level phenomenology*, Comput. Phys. Commun. **185**, 2250 (2014), doi:10.1016/j.cpc.2014.04.012.

[64] B. A. Dobrescu and F. Yu, *Coupling-mass mapping of dijet peak searches*, Phys. Rev. D **88**, 035021 (2013), doi:10.1103/PhysRevD.88.035021.

[65] R. S. Chivukula, A. G. Cohen and E. H. Simmons, *New strong interactions at the Tevatron?*, Phys. Lett. B **380**, 92 (1996), doi:10.1016/0370-2693(96)00464-9.

[66] J. L. Hewett and T. G. Rizzo, *Low-energy phenomenology of superstring inspired E_6 models*, Phys. Rep. **183**, 193 (1989), doi:10.1016/0370-1573(89)90071-9.

[67] R. S. Chivukula, P. Ittisamai, K. Mohan and E. H. Simmons, *Color discriminant variable and scalar diquarks at the LHC*, Phys. Rev. D **92**, 075020 (2015), doi:10.1103/PhysRevD.92.075020.

[68] U. Baur, I. Hinchliffe, and D. Zeppenfeld, *Excited quark production at hadron colliders*, Int. J. Mod. Phys. A **02**, 1285 (1987), doi:10.1142/S0217751X87000661.

[69] T. Han, J. D. Lykken and R.-J. Zhang, *Kaluza-Klein states from large extra dimensions*, Phys. Rev. D **59**, 105006 (1999), doi:10.1103/PhysRevD.59.105006.

[70] R. S. Chivukula, D. Foren, K. A. Mohan, D. Sengupta and E. H. Simmons, *Massive spin-2 scattering amplitudes in extra-dimensional theories*, Phys. Rev. D **101**, 075013 (2020), doi:10.1103/PhysRevD.101.075013.

[71] R. S. Chivukula, D. Foren, K. A. Mohan, D. Sengupta and E. H. Simmons, *Scattering amplitudes of massive spin-2 Kaluza-Klein states grow only as $\mathcal{O}(s)$*, Phys. Rev. D **101**, 055013 (2020), doi:10.1103/PhysRevD.101.055013.

[72] R. S. Chivukula, D. Foren, K. A. Mohan, D. Sengupta and E. H. Simmons, *Spin-2 Kaluza-Klein mode scattering in models with a massive radion*, Phys. Rev. D **103**, 095024 (2021), doi:10.1103/PhysRevD.103.095024.

[73] R. S. Chivukula, D. Foren, K. A. Mohan, D. Sengupta and E. H. Simmons, *Sum rules for massive spin-2 Kaluza-Klein elastic scattering amplitudes*, Phys. Rev. D **100**, 115033 (2019), doi:10.1103/PhysRevD.100.115033.

[74] C. T. Hill and E. H. Simmons, *Strong dynamics and electroweak symmetry breaking*, Phys. Rep. **381**, 235 (2003), doi:10.1016/S0370-1573(03)00140-6.

[75] B. A. Dobrescu, K. Kong and R. Mahbubani, *Massive color-octet bosons and pairs of resonances at hadron colliders*, Phys. Lett. B **670**, 119 (2008), doi:10.1016/j.physletb.2008.10.048.

[76] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, *MadGraph 5: Going beyond*, J. High Energy Phys. **06**, 128 (2011), doi:10.1007/JHEP06(2011)128.

[77] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178**, 852 (2008), doi:10.1016/j.cpc.2008.01.036.

[78] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72**, 1896 (2012), doi:10.1140/epjc/s10052-012-1896-2.

[79] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. B **641**, 57 (2006), doi:10.1016/j.physletb.2006.08.037.

[80] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, J. High Energy Phys. **04**, 063 (2008), doi:10.1088/1126-6708/2008/04/063.

[81] D. Krohn, J. Thaler and L.-T. Wang, *Jet trimming*, J. High Energy Phys. **02**, 084 (2010), doi:10.1007/JHEP02(2010)084.

[82] L. de Oliveira, M. Paganini and B. Nachman, *Learning particle physics by example: Location-aware generative adversarial networks for physics synthesis*, Comput. Softw. Big Sci. **1**, 4 (2017), doi:10.1007/s41781-017-0004-6.

[83] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, (arXiv preprint) doi:10.48550/arXiv.1412.6980.

[84] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, *On the importance of initialization and momentum in deep learning*, in *Proceedings of the 30th international conference on international conference on machine learning*, Proc. Mach. Learn. Res. **28**, 1139 (2013).

[85] F. Chollet, *Keras* (2015), https://github.com/keras-team/keras.

[86] M. Abadi et al., *TensorFlow: A system for large-scale machine learning*, (arXiv preprint) doi:10.48550/arXiv.1605.08695.

[87] M. Lindauer and F. Hutter, *Warmstarting of model-based algorithm configuration*, in *Thirty-second AAAI conference on artificial intelligence*, AAAI Press, Palo Alto, USA, ISBN 9781577358008 (2018), doi:10.1609/aaai.v32i1.11532.

[88] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, (arXiv preprint) doi:10.48550/arXiv.1409.1556.

[89] Y.-T. Chien, Z. Montague, A. Emerman, S.-C. Hsu and S. Meehan, *Isolating color-singlet boson jets at the LHC using telescoping jet substructure*, Phys. Rev. D **101**, 114006 (2020), doi:10.1103/PhysRevD.101.114006.

[90] R. E. Schapire and Y. Singer, *Improved boosting algorithms using confidence-rated predictions*, Mach. Learn. **37**, 297 (1999), doi:10.1023/A:1007614523901.

[91] F. Pedregosa et al., *Scikit-learn: Machine learning in Python*, J. Mach. Learn. Res. **12**, 2825 (2011).

[92] P. T. Komiske, E. M. Metodiev and M. D. Schwartz, *Deep learning in color: Towards automated quark/gluon jet discrimination*, J. High Energy Phys. **01**, 110 (2017), doi:10.1007/JHEP01(2017)110.

[93] ATLAS collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, Tech. Rep. ATL-PHYS-PUB-2014-021, CERN, Geneva, Switzerland (2014), https://cds.cern.ch/record/1966419.

[94] P. Skands, S. Carrazza and J. Rojo, *Tuning PYTHIA 8.1: The Monash 2013 tune*, Eur. Phys. J. C **74**, 3024 (2014), doi:10.1140/epjc/s10052-014-3024-y.

[95] S. Argyropoulos and T. Sjöstrand, *Effects of color reconnection on $t\bar{t}$ final states at the LHC*, J. High Energy Phys. **11**, 043 (2014), doi:10.1007/JHEP11(2014)043.

[96] J. R. Christiansen and P. Z. Skands, *String formation beyond leading colour*, J. High Energy Phys. **08**, 003 (2015), doi:10.1007/JHEP08(2015)003.