Published for SISSA by Springer

Received: *August 23, 2022*
Revised: *November 4, 2022*
Accepted: *November 21, 2022*
Published: *December 21, 2022*

Learning to identify semi-visible jets

Taylor Faucett,[a] Shih-Chieh Hsu[b] and Daniel Whiteson[a]

[a] *Department of Physics and Astronomy, University of California, Irvine, CA, U.S.A.*
[b] *Department of Physics, University of Washington, Seattle, WA, U.S.A.*

E-mail: tfaucett@uci.edu, schsu@uw.edu, daniel@uci.edu

Abstract: We train a network to identify jets with fractional dark decay (semi-visible jets) using the pattern of their low-level jet constituents, and explore the nature of the information used by the network by mapping it to a space of jet substructure observables. Semi-visible jets arise from dark matter particles which decay into a mixture of dark sector (invisible) and Standard Model (visible) particles. Such objects are challenging to identify due to the complex nature of jets and the alignment of the momentum imbalance from the dark particles with the jet axis, but such jets do not yet benefit from the construction of dedicated theoretically-motivated jet substructure observables. A deep network operating on jet constituents is used as a probe of the available information and indicates that classification power not captured by current high-level observables arises primarily from low-p_{T} jet constituents.

Keywords: Dark Matter at Colliders, Jets and Jet Substructure

ArXiv ePrint: 2208.10062

https://doi.org/10.1007/JHEP12(2022)132

JHEP12(2022)132

Contents

1 Introduction **1**

2 Semi-visible jets **2**

3 Sample generation and data processing **3**
3.1 High-level observables 4

4 Machine learning and evaluation **4**
4.1 Performance comparison 5

5 Finding new observables **7**
5.1 Search strategy 7

6 Guided iteration **8**
6.1 Analysis of the guided search results 9

7 Greedy search **11**

8 Exploration of p_{T} dependence **12**

9 Conclusions **14**

A Signal reweighting **15**

B Jet substructure observables **15**
B.1 Jet transverse momentum and mass 15
B.2 Generalized angularities 16
B.3 Energy correlation 17
B.4 N-subjettiness 18
B.5 Groomed momentum splitting fraction 19

C ML architectures **19**
C.1 Deep neural networks 19
C.2 High-level DNN 19
C.2.1 Deep neural networks 19
C.2.2 Particle-flow networks 20
C.3 Boosted learning models 20
C.3.1 LightGBM 20
C.3.2 XGBoost 20
C.4 Convolutional networks on jet images 20

JHEP12(2022)132

1 Introduction

The microscopic nature of dark matter (DM) remains one of the most pressing open questions in modern physics [1, 2], and a robust program of experimental searches for evidence of its interaction with the visible sector [3–11] described by the Standard Model (SM). These experiments typically assume that DM is neutral, stable and couples weakly to SM particles; in collider settings this predicts detector signatures in which weakly-produced DM particles are invisible, evidenced only by the imbalance of momentum transverse to the beam. No evidence of DM interactions has been observed to date.

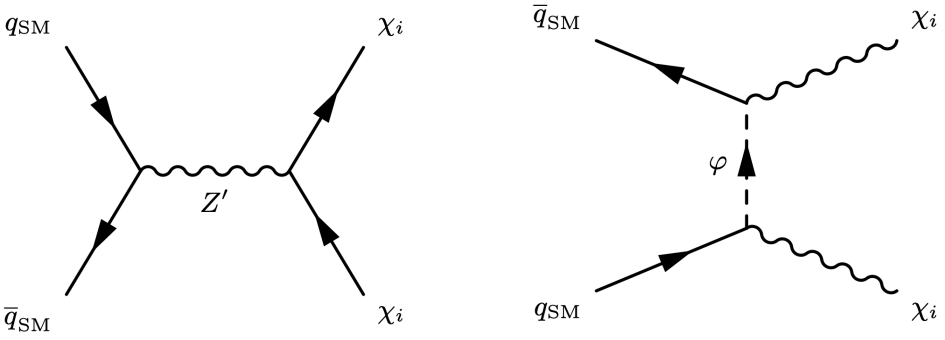
However, while these assumptions are reasonable, the lack of observation motivates the exploration of scenarios in which one or more of them are relaxed. Specifically, if DM contains complex strongly-coupled hidden sectors, such as appear in many Hidden-Valley models [12, 13], it may lead to the production of stable or meta-stable dark particles within hadronic jets [14–17]. Depending on the portion of the jet which results in dark-sector hadrons, it may be only *semi-visible* to detectors, leading to a unique pattern of energy deposits, or jet substructure.

A robust literature exists for the identification of jet substructure [18–21], with applications to boosted W-boson [22–24], Higgs boson [25] and top-quark tagging [24]. Typically, observables are designed to discriminate between the energy deposition patterns left by jets which result from a single hard parton and the patterns left by jets which result from several collimated hard partons, as can be produced from the decay of a massive boosted particle. While these observables have some power when adapted to the task of identifying semi-visible jets [16, 26], no observables have yet been specifically designed to be sensitive to the unique energy patterns of semi-visible nature of jets.

In parallel, the rapid development of machine learning to the analysis of jet energy depositions [19, 22, 27] demonstrated that jet tagging strategies, including those for semi-visible jets, can be learned directly from lower-level jet constituents without the need to form physics-motivated high level observables [17, 28]. Such learned models are naturally challenging to interpret, validate or quantify uncertainties, especially given the high-dimensional nature of their inputs. In the case of semi-visible jets, extra care must be taken when drawing conclusions from low-level details, many of which may depend on specific theoretical parameter choices as well as the approximations made during modeling of hadronization [26]. However, techniques have been recently demonstrated [29–32] to translate the learned model into interpretable high-level observables, which can provide guidance regarding the nature and robustness of the information being used for classification.

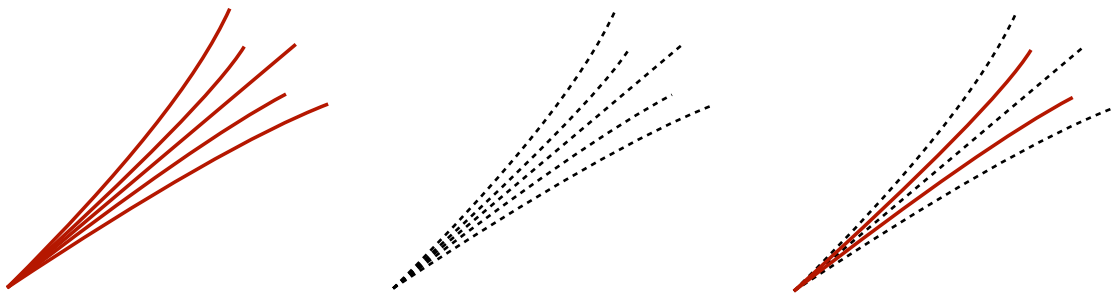
In this paper, we present a study of machine learning models trained to distinguish semi-visible jets from QCD background jets using the patterns of their low-level jet constituents. We compare the performance of models which use low-level constituents to those which use the set of existing high-level observables to explore where the existing HL observables do and do not capture all of the relevant information. Where gaps exist, we attempt to translate [29] low-level networks into networks which use a small set of interpretable observables which replicate their decisions and performance. Interpretation of these observables can yield insight into the nature of the energy deposition inside semi-visible jets.

JHEP12(2022)132



(a) s-channel (b) t-channel

Figure 1. Feynman diagrams for s-channel and t-channel pair-production of dark-sector quarks χ_i which lead to semi-visible jet production.



(a) Rapid Decay (Visible), $r_{\rm inv} = 0$. (b) Stable Dark Jet (Invisible), $r_{\rm inv} = 1$. (c) Fractional Decay (Semi-Visible), $r_{\rm inv} \in (0, 1)$.

Figure 2. Depictions of jets produced with varying visible (SM, solid red) and invisible (dark sector, dashed black) components.

2 Semi-visible jets

Following refs. [14, 16], we consider pair production of dark-sector quarks of several flavors ($\chi_i = \chi_{1,2}$) via a messenger sector which features a Z' gauge boson in an s-channel process (figure 1(a)) or scalar mediator φ in the t-channel process (figure 1(b)) that couples to both SM and DM sectors and leads to a dijet signature. The dark quarks produce QCD-like dark showers, involving many dark quarks and gluons which produce dark hadrons, some of which are stable or meta-stable and some of which decay into SM hadrons via an off-shell process.

The detector signature of the resulting semi-visible jet (SVJ) depends on the lifetime and stability of the dark hadrons, leading to several possibilities (see figure 2). Though the physics is complex and depends on the details of the dark sector structure, a description of the dark and SM hadrons produced by a DM model quark can be encapsulated in the

JHEP12(2022)132

quantity r_{inv}, the ratio of dark stable hadrons to all hadrons in the jet:

$$r_{\text{inv}} \equiv \left\langle \frac{\#\text{ of stable dark hadrons}}{\#\text{ of hadrons}} \right\rangle . \tag{2.1}$$

An invisible fraction of $r_{\text{inv}} = 0$ corresponds to a dark quark which produces a jet consisting of only visible hadrons, as depicted in the *Rapid Decay* example given in figure 2(a). Alternatively, an invisible fraction of $r_{\text{inv}} = 1$ describes a stable dark jet (figure 2(b)), in which the dark quark hadronizes exclusively in the dark sector. For any intermediate value of r_{inv}, jets contain a visible and invisible fraction, leading to $\not{E}_{\text{T}}$ along the jet axis (figure 2(c)).

3 Sample generation and data processing

Samples of simulated events with semi-visible jets are generated using the modified Hidden Valley [12] model described in [14, 15] for both an s-channel (figure 1(a)) and t-channel (figure 1(b)) process. Samples of simulated $pp \to Z' \to \chi_1\overline{\chi}_1$ and $pp \to \varphi \to \chi_1\overline{\chi}_1$ events are produced in proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ in MadGraph5 [33] (v2.6.7) with xqcut=100 and the NNPDF2.3 LO PDF set [34]. The mediator mass is set to 1.5 TeV and the dark quark mass to $M_{\chi_1} = 10\,\text{GeV}$. Up to two extra hard jets due to radiation are generated and MLM-matched [35], to facilitate comparison with existing studies. Invisible fraction, showering and hadronization are performed with Pythia8 v8.244 [36]. In particular, the following parameters are set: the dark confinement scale $\Lambda_d = 5\,\text{GeV}$; the number of dark colors $N_c = 2$; the number of dark flavors $N_f = 1$; and the intermediate dark meson ρ_d mass of $m_{\rho_d} = 20\,\text{GeV}$ and the dark matter π_d mass of $m_d = 9.9\,\text{GeV}$. Distinct sets were generated for invisible fractions of $r_{\text{inv}} \in [0.0, 0.3, 0.6]$ via configurations of the branching ratio of decay process $\rho_d \to \pi_d\pi_d$. Detector simulation and reconstruction are conducted in Delphes v3.4.2 [37] using the default ATLAS card. A sample of SM jets from a typical QCD process is generated from the $pp \to jj$ process. The same simulation chain as described above is applied to the SM jets.

Jets are built from calorimeter energy deposits and undergo the jet trimming procedure described in ref. [38] with the anti-k_{T} [39] clustering algorithm from pyjet [40], using a primary jet-radius parameter of $R = 1.0$ and subjet clustering radius of $R_{\text{sub}} = 0.2$ and $f_{\text{cut}} = 0.05$. The threshold on f_{cut} effectively removes constituents in subjets with p_{T} that is less than 5% of the jet p_{T}. Leading jets are required to have $p_{\text{T}} \in [300, 400]\,\text{GeV}$. For each event generated, the leading jet is selected and truth-matched to guarantee the presence of a dark quark within the region of $\Delta R < 1$. The fraction of events in which the leading jet satisfies this criterion is approximately 30% and is likely to be sensitive to the dark hadron shower modeling.

After all selection requirements, 2×10^6 simulated jets remain with a 50/50 split between SVJ signal and QCD background. To avoid inducing biases from artifacts of the generation process, signal and background events are weighted such that the distributions in p_{T} and η are uniform; see figure 7.

JHEP12(2022)132

s-channel	AUC		
Model	$r_{\text{inv}} = 0.0$	0.3	0.6
LightGBM	0.861	0.803	0.736
XGBoost	0.861	0.803	0.736
DNN	0.860	0.799	0.734
t-channel	AUC		
Model	$r_{\text{inv}} = 0.0$	0.3	0.6
LightGBM	0.808	0.755	0.683
XGBoost	0.806	0.755	0.681
DNN	0.801	0.726	0.656

Table 1. Summary of performance (AUC) in the SVJ classification task for several networks using high-level features, for the six simulated scenarios, three choices of invisible fraction r_{inv} for both the s-channel and t-channel processes. Statistical uncertainty in each case is less than ± 0.002 with a 95% confidence, measured using bootstrapping over 200 models.

3.1 High-level observables

A large set of jet substructure observables [20, 41, 42] have been proposed for tasks different from the focus of this study, that of identifying jets with multiple hard subjets. Nevertheless, these observables may summarize the information content in the jet in a way that is relevant for the task of identifying semi-visible jets [16], and so serve as a launching point for the search for new observables.

Our set of high-level (HL) observables includes 15 candidates: jet p_{T}, the Generalized Angularities [43] of p_{T}^{D} and Les Houches Angularity (LHA), N-subjettiness ratios $\tau_{21}^{\beta=1}$ and $\tau_{32}^{\beta=1}$ [44], and Energy Correlation function ratios [19] $C_2^{\beta=1}$, $C_2^{\beta=2}$, $D_2^{\beta=1}$, $D_2^{\beta=2}$, e_2, e_3, as well as jet width, jet e_{mass}, constituent multiplicity and the splitting function [18] z_g. In each case, observables are calculated from trimmed jet constituents described above. Definitions and distributions for each high-level observable are provided in appendix B.1 with re-weighting applied as described above.

4 Machine learning and evaluation

For both the low-level (LL) trimmed jet constituents and high-level jet substructure observables, a variety of networks and architectures are tested.

Due to their strong record in previous similar applications [17, 29–31], a deep neural network using dense layers is trained on the high-level observables. We additionally consider XGBoost [45], which has shown strong performance in training high-level classifiers with jet substructure [46, 47], as well as LightGBM [48] which has demonstrated power in class separation on high-level features. LightGBM has the strongest classification performance among these networks which use high-level features; see table 1.

In the case of classifiers which use low-level constituents, convolutional neural networks on jet images are considered [22, 29–31, 49, 50]. Given the specific task of classifying jet

JHEP12(2022)132

s-channel	AUC		
Model	$r_{\text{inv}} = 0.0$	0.3	0.6
PFN	0.866	0.822	0.776
EFN	0.849	0.795	0.735
CNN	0.855	0.792	0.740
t-channel	AUC		
Model	$r_{\text{inv}} = 0.0$	0.3	0.6
PFN	0.806	0.754	0.697
EFN	0.796	0.741	0.672
CNN	0.791	0.739	0.663

Table 2. Summary of performance (AUC) in the SVJ classification task for several networks using low-level constituents, for the six simulated scenarios, three choices of invisible fraction r_{inv} for both the s-channel and t-channel processes. Statistical uncertainty in each case is less than ± 0.002 with a 95% confidence, measured using bootstrapping over 200 models.

substructure observables and their use for a similar task in ref. [31], Energy Flow Networks (EFN) and Particle Flow Networks (PFN) are also applied [51], and found to significantly out-perform convolutional networks, with the PFN emerging as the most powerful network; see table 2.

Receiver operating characteristic (ROC) curves for both the PFN and LightGBM high-level models are given in figure 3. Additional details for the network training and hyper-parameter selections are provided in appendix C.

4.1 Performance comparison

We compare the SVJ classification performance of the most powerful networks based on high-level or low-level input features, through calculations of the Area under the ROC Curve (AUC); see figure 3 and table 3.

Note that the high-level (HL) observables are calculated directly from the low-level constituents with no additional information. If HL features extract all of the relevant information for the classification task, networks which use them as inputs should be able to match the performance of the networks which use the LL information directly, which we take as a probe of the power of the available information. If networks using LL information surpass the performance of those using HL information, it suggests that information exists in the LL constituents which is not being captured by the HL observables. One might consider directly applying networks based on LL information to the classification task [17], but this presents challenges in calibrating, validating and quantifying uncertainties on their high-dimensional inputs. Instead, a performance gap suggests the possibility that an additional HL observable might be crafted to take advantage of the overlooked information.

In each of the six cases explored here, the networks which use low-level information match or exceed the performance of networks which use high-level jet observables. Significant performance gaps are seen in the $r_{\text{inv}} = 0.6$ case, where the AUC between the LL and HL networks is 0.040 and 0.014 for the s-channel and t-channel cases, as well as in

JHEP12(2022)132

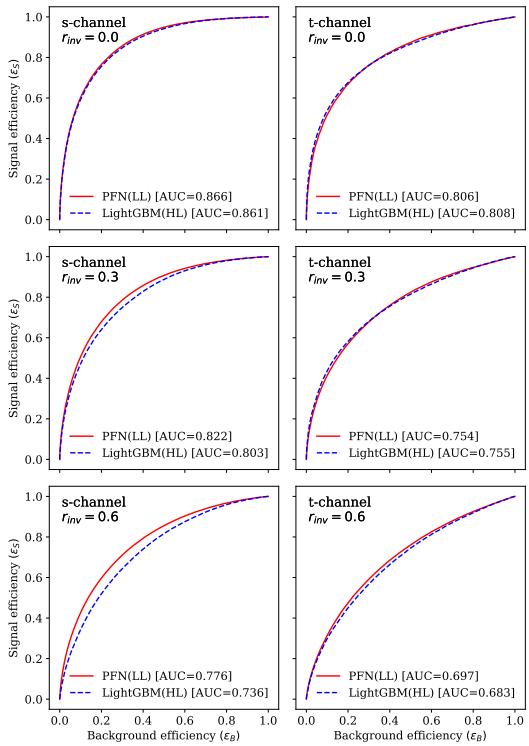


Figure 3. Comparison of performance between a network which uses low-level constituents (PFN in solid red) and one which uses high-level jet observables (LightGBM, dashed blue). Shown are the background rejection (inverse of background efficiency) versus signal efficiency for the six simulated cases: s-channel and t-channel production for three values of r_{inv}. Statistical uncertainty on AUC in each case is less than ± 0.002 at 95% confidence level, measured using bootstrapping over 200 models.

	s-channel					
Model [Features]	$r_{\text{inv}} = 0.0$		$r_{\text{inv}} = 0.3$		$r_{\text{inv}} = 0.6$	
	ADO $[\cdot, \text{PFN}]$	AUC	ADO $[\cdot, \text{PFN}]$	AUC	ADO $[\cdot, \text{PFN}]$	AUC
PFN [LL]	1	0.866	1	0.822	1	0.776
LightGBM [HL]	0.858	0.861	0.839	0.803	0.818	0.736
LL-HL gap		**0.005**		**0.019**		**0.040**

	t-channel					
Model [Features]	$r_{\text{inv}} = 0.0$		$r_{\text{inv}} = 0.3$		$r_{\text{inv}} = 0.6$	
	ADO $[\cdot, \text{PFN}]$	AUC	ADO $[\cdot, \text{PFN}]$	AUC	ADO $[\cdot, \text{PFN}]$	AUC
PFN [LL]	1	0.806	1	0.754	1	0.697
LightGBM [HL]	0.844	0.808	0.805	0.755	0.787	0.683
LL-HL gap		**−0.002**		**−0.001**		**0.014**

Table 3. Summary of performance (AUC and ADO) in the SVJ classification task for various network architectures and input features. Statistical uncertainty in each case is less than ± 0.002 at 95% confidence level, measured using bootstrapping over 200 models.

JHEP12(2022)132

the and $r_{\text{inv}} = 0.3$ invisible fraction in the s-channel process, where the gap is 0.019. In the $r_{\text{inv}} = 0.0$ s-channel scenario, a small gap is seen, though larger than the statistical uncertainty. In the other two cases, the LL and HL networks achieve essentially equivalent performance, strongly suggesting that the HL features have captured the relevant information in the LL constituents. As these observables were not designed for this task, it was not a priori clear that they would summarize all of the available and relevant information.

However, one can also assess the difference between the LL and HL networks using other metrics than AUC. The Average Decision Ordering (ADO) metric, see eq. (5.5), measures the fraction of input pairs in which two networks give the same output ordering. Even in cases where the AUC is equivalent, the ADO between the LightGBM and PFN (table 3) is well below 1, suggesting that while their classification performance is the same, they arrive at it using distinct strategies. This indicates that there may be room to improve the classification accuracy by considering a network which uses both sets of features.

5 Finding new observables

The studies above reveal that models which use low-level constituents as inputs provide the overall best classification performance. However, our objective is not merely to find the classifier with the optimal statistical performance with difficult-to-assess systematic uncertainties. Rather, we seek to understand the underlying physics used by the PFN and to translate this information into one or more meaningful physical features, allowing us to extract insight into the physical processes involved and assign reasonable systematic uncertainties. In this section, we search for these additional high-level observables among the Energy Flow Polynomials [27] (EFP), which form a complete basis for jet observables.

5.1 Search strategy

Interpreting the decision making of a black-box network is a notoriously difficult problem. For the task of jet classification, we apply the guided search technique utilized in past jet-related interpretability studies [21, 29–32]. In this approach, new HL observables are identified among the infinite set of Energy Flow Polynomials (EFPs) which exist as a complete linear basis for jet observables. In the EFP framework, one-dimensional observables are constructed as nested sums over measured jet constituent transverse momenta $p_{\text{T},i}$ and scaled by their angular separation θ_{ij}.

These parametric sums are described as the set of all isomorphic multigraphs where:

$$\text{each node} \Rightarrow \sum_{i=1}^{N} z_i, \tag{5.1}$$

$$\text{each } k\text{-fold edge} \Rightarrow (\theta_{ij})^k \,. \tag{5.2}$$

and where each graph can be further parameterized by values of (κ, β),

$$(z_i)^\kappa = \left(\frac{p_{\text{T},i}}{\sum_j p_{\text{T},j}} \right)^\kappa , \tag{5.3}$$

$$\theta_{ij}^\beta = \left(\Delta\eta_{ij}^2 + \Delta\varphi_{ij}^2 \right)^{\beta/2} . \tag{5.4}$$

JHEP12(2022)132

Here, $p_{\mathrm{T},i}$ is the transverse momentum of the trimmed jet for constituent i, and $\Delta\eta_{ij}$ ($\Delta\varphi_{ij}$) is pseudorapidity (azimuth) difference between constituents i and j. As the EFPs are normalized, they capture only the relative information about the energy deposition. For this reason, in each network that includes EFP observables, we include as an additional input the sum of p_{T} over all constituents, to indicate the overall scale of the energy deposition.

For the set of EFPs, infrared and collinear (IRC) safety requires that $\kappa = 1$. To more broadly explore the space, we consider examples with $\kappa \neq 1$ which generate more exotic observables. For the case of EFPs with IRC-unsafe parameters, we select all prime graphs with dimension $d \leq 5$ and κ and β values of $\kappa \in \left[-2, -1, 0, \frac{1}{2}, 1, 2, 4\right]$ and $\beta \in \left[\frac{1}{10}, \frac{1}{2}, 1, 2, 4\right]$. Given the form of eq. (5.1) and eq. (5.2), the size and relative sign of inputs chosen for (κ, β) will provide insights into utility of soft/hard p_{T} effects and narrow/wide angle energy distributions.

In principle, the EFP space is complete and any information accessible through constituent-based observables is contained in some combination of EFPs. However, there is no guarantee that an EFP representation will be compact. On the contrary, a blind search through the space can prove time and resource prohibitive. Rather than a brute force search of an infinite space of observables, we take the guided approach of ref. [29], which uses the PFN as a black-box guide and iteratively assembles a set of EFPs which provide the closest equivalent decision making in a compact and low-dimensional set of inputs. This is done by isolating the space of information in which the PFN and existing HL features make opposing decisions on the same inputs and isolates the EFP which most closely mimics the PFN in that subspace.

Here, the agreement between networks $f(x)$ and $g(x)$ is evaluated over pairs of inputs (x, x') by comparing their relative classification decisions, expressed mathematically as:

$$\mathrm{DO}[f, g](x, x') = \Theta\Big(\big(f(x) - f(x')\big)\big(g(x) - g(x')\big)\Big), \tag{5.5}$$

and referred to as *decision ordering* (DO). DO$= 0$ corresponds to inverted decisions over all input pairs and DO $= 1$ corresponds to the same decision ordering. As prescribed in ref. [29], we scan the space of EFPs to find the observable that has the highest average decision ordering (ADO) with the guiding network when averaged over disordered pairs. The selected EFP is then incorporated into the new model of HL features, HL_{n+1}, and the process is repeated until the ADO plateaus.

6 Guided iteration

For each of the four scenarios in which a gap is observed between the AUC performance of the HL LightGBM model and the PFN, a guided search is performed to identify an EFP observable which mimics the decisions of the PFN.

The search results, shown in table 4, identify in each of the four cases an EFP with $d \leq 3$ which boosts the classification performance of the LightGBM classifier when trained with the original 15 HL features as well as the identified EFP observable. In addition, the decision similarity (ADO) between the new HL model and the PFN is increased. Scans

JHEP12(2022)132

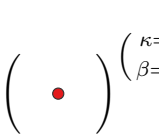
Process	r_{inv}	HL network AUC, ADO [.,PFN]	HL+EFP network EFP	κ	β	AUC, ADO [.,PFN]	PFN AUC
s-channel	0.0	0.861, 0.858		-2		0.864, 0.863	0.866
s-channel	0.3	0.803, 0.839		1	$\frac{1}{2}$	0.807, 0.840	0.822
s-channel	0.6	0.736, 0.818		-1	2	0.747, 0.821	0.776
t-channel	0.6	0.683, 0.787		-2	$\frac{1}{10}$	0.690, 0.792	0.697

Table 4. Results of a guided search for high-level (HL) EFP observables which mimic the decision ordering of the PFN, a network based on low-level constituents. For each of the four scenarios in which a gap is observed between the AUC performance of the HL network and the PFN, an EFP is selected to attempt to increase the average decision ordering (ADO) of the HL network with the PFN. Statistical uncertainty in each case is ± 0.002 with 95% confidence, measured using bootstrapping over 100 models.

for a second EFP do not identify additional observables which significantly increase performance or similarity. A guided search was also performed with an identical set of κ and β parameters for EFPs with dimension $d \leq 5$ but no improvements were seen in the case of higher dimensional graph structure. While the performance and similarity gaps have been reduced, they have not quite been erased.

6.1 Analysis of the guided search results

In the s-channel process, with invisible fraction $r_{\text{inv}} = 0.0$, the addition of a single EFP closes the small performance gap with the PFN within the statistical uncertainty. The identified EFP in this case is the *dot* graph with the IRC-unsafe energy exponent $\kappa = -2$, expressed as a sum over constituents in eq. (6.1).



$$= \sum_{a=1}^{N} \frac{1}{z_a^2} \,. \tag{6.1}$$

This graph is, in effect, simply a measure of the sum of the inverse p_{T}^2 of the jet constituents, and is sensitive to constituents with low p_{T}. Distribution of values of this observable for signal and background events are shown in figure 4, demonstrating good separation between signal and background.

In the remaining s-channel and t-channel examples, addition of the selected EFP improves performance but fails to match the PFN. Of the three existing gaps, the s-channel process with $r_{\text{inv}} = 0.3$ is the only result in which we see an IRC-safe EFP observable, where $\kappa = 1$. In the other cases, the EFP graphs again have $\kappa < 0$, making them sensitive to low-p_{T} information. The complete expression for each selected graph is given in eq. (6.2) and the distributions of each observable for signal and background are shown in figure 5.

JHEP12(2022)132

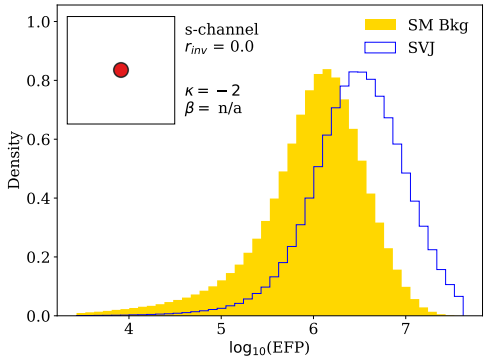
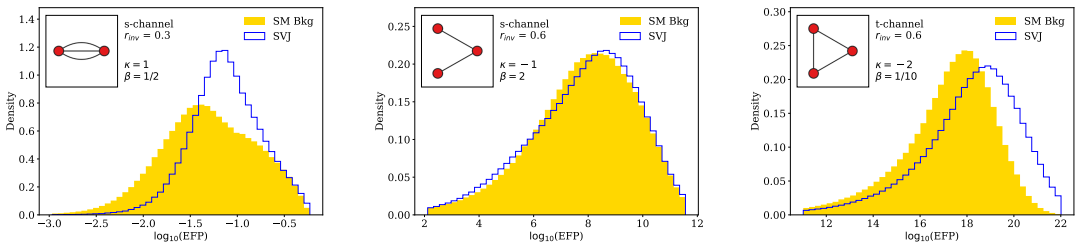


Figure 4. Distribution of the EFP observable selected by the guided search in the s-channel $r_{\text{inv}} = 0$ scenario, shown for semi-visible jets (SVJ) as well as QCD jets from the Standard Model background (SM Bkg). The inset pane shows the graph corresponding to the EFP. See text for additional details.



(a) s-channel / $r_{\text{inv}} = 0.3$. (b) s-channel / $r_{\text{inv}} = 0.6$. (c) t-channel / $r_{\text{inv}} = 0.6$.

Figure 5. Distribution of EFP observables selected by the guided search for the s-channel $r_{\text{inv}} = 0.3$, $r_{\text{inv}} = 0.6$ and t-channel $r_{\text{inv}} = 0.6$ scenarios, shown for semi-visible jets (SVJ) as well as QCD jets from the Standard Model background (SM Bkg). The inset panes show the graph corresponding to the selected EFP. See text for additional details.

The triangular graph selected in the case of the t-channel process with invisible fraction $r_{\text{inv}} = 0.6$ has the same structure as the energy correlation ratio (e_3), though with distinct κ and β values.

$$\left(\quad\right)^{\left(\substack{\kappa=1\\ \beta=\frac{1}{2}}\right)} = \sum_{a,b=1}^{N} z_a z_b \left(\theta_{ab}\right)^{3/2} \tag{6.2a}$$

$$\left(\quad\right)^{\left(\substack{\kappa=-1\\ \beta=2}\right)} = \sum_{a,b,c=1}^{N} \frac{\left(\theta_{ab}\theta_{ac}\right)^2}{z_a z_b z_c} \tag{6.2b}$$

$$\left(\quad\right)^{\left(\substack{\kappa=-2\\ \beta=\frac{1}{10}}\right)} = \sum_{a,b,c=1}^{N} \frac{\left(\theta_{ab}\theta_{bc}\theta_{ac}\right)^{\frac{1}{10}}}{z_a z_b z_c} \,. \tag{6.2c}$$

JHEP12(2022)132

Process	r_{inv}	HL AUC	Pass 1 Graph	κ	β	AUC	Pass 2 Graph	κ	β	AUC	Guided HL AUC	PFN AUC
s-channel	0.0	0.861		$\frac{1}{2}$	2	0.864		2	$\frac{1}{10}$	0.866	0.864	0.866
s-channel	0.3	0.803		4	2	0.807		-1	1	0.809	0.807	0.822
s-channel	0.6	0.736		4	4	0.744		-2	$\frac{1}{10}$	0.747	0.747	0.776
t-channel	0.6	0.683		-1	$\frac{1}{10}$	0.690		-2	4	0.692	0.690	0.697

Table 5. Summary of two passes of a greedy search through EFP space for additional observables which might capture the information used by the low-level network (PFN) and match its performance, as measured by AUC. For each of the four processes and r_{inv} scenarios in which we have identified a gap between performance of the PFN and the HL model, two passes are made to identify the EFP which most improves the AUC of a new HL model which incorporates the candidate EFP. Performance of the HL model deduced by the guided search (table 4) and the PFN are also given. Statistical uncertainty in each case is ± 0.002 with 95% confidence, measured using bootstrapping over 100 models.

7 Greedy search

Given the persistent gap between the performance of the LL networks and the HL models augmented by EFP observables, we consider whether the EFP space lacks the needed observables, or whether the guided search is failing to identify it. We examine this question by taking a more comprehensive look at the space of EFPs we consider. Similar to the technique described in ref. [29], we perform a greedy search in the same EFP space studied in the guided iteration approach explored above. In a greedy search, we explicitly train a new model for each candidate EFP, combining the EFP with the existing 15 HL features. Note that this is significantly more computationally intensive than evaluation of the ADO, as done in the guided search, and seeks to maximize AUC rather than to align decision ordering with the PFN. The candidate EFP which produces the best-performing model is kept as the 16th HL observable (Pass 1), and the process is repeated in search of a 17th (Pass 2), until a plateau in performance is observed.

The results of the greedy search across all choices of r_{inv} in s-channel and t-channel scenarios where a gap between HL and LL exists are given in table 5. Similar levels of performance are achieved as in the guided search, to within statistical uncertainties. In all cases, the HL and LL gap persists. Results are given for the IRC-unsafe selections with dimension $d \leq 3$. A similar greedy search was also performed on the IRC-unsafe $d \leq 5$ EFP set with no performance differences observed.

The greedy search selects similar EFP graphs as the guided search, with the exception of the 4-node graph selected in pass 2 for the $r_{\text{inv}} = 0.6$, t-chanel scenario. No IRC-safe graphs ($\kappa = 1$) are selected and we again see frequent sensitivity to low p_{T} parameters (i.e.

JHEP12(2022)132

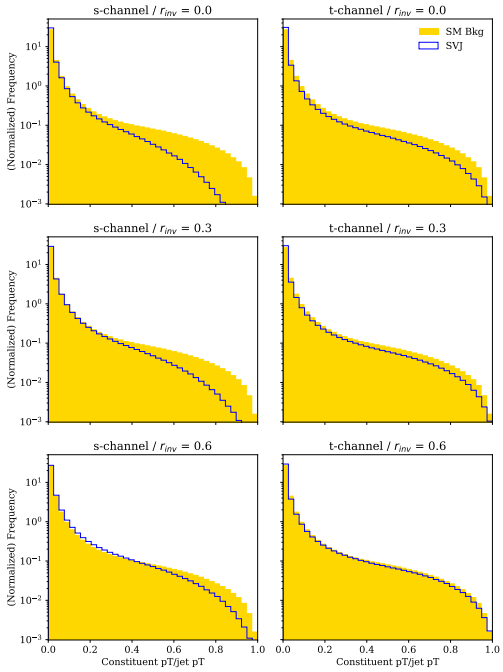


Figure 6. Distribution of constituent p_{T} relative to the jet p_{T} for the six scenarios: s-channel and t-channel with $r_{\mathrm{inv}} \in [0, 0.3, 0.6]$.

$\kappa = -2, -1, \frac{1}{2}$) and a variety of both narrow and wide angle features (i.e. $\beta = \frac{1}{10}, 2, 4$). A repeat of the greedy search with only IRC-safe observables achieves no performance improvements over the original HL features, suggesting that missing information may be strongly tied to IRC-unsafe representations of the information in the jet constituents.

8 Exploration of p_{T} dependence

The results of both the guided search and the greedy search strongly suggest that the full performance gap which persists between the HL and LL representation of the jet contents cannot be compactly expressed in terms of a small number of EFP observables in the set that have been considered. This raises the question of what feature of the LL constituents can be credited with this performance improvement and why that information does not translate compactly to our EFP observables. The clues from the guided and greedy search point to sensitivity to low-p_{T} constituents, as can be generated from soft radiation. Figure 6 shows the distribution of constituent p_{T} relative to the jet p_{T} for the six scenarios, in which SVJs appear to have more constituents at low relative p_{T}, as also seen in ref. [52]. Recall that jet constituents are trimmed if the subjet they belong to has a p_{T} below a threshold of 5% of the p_{T} of the jet, corresponding to $f_{\mathrm{cut}} = 0.05$.

To consider whether the distinguishing information is contained in these low-p_{T} constituents, we explore a broader range of thresholds, both lowering it to 0% and raising it to 10% and 15%. The HL features are re-evaluated on the newly trimmed constituents, and used as inputs to a new LightGBM model, whose performance is compared to a PFN

JHEP12(2022)132

	s-channel								
	$r_{\text{inv}} = 0.0$			$r_{\text{inv}} = 0.3$			$r_{\text{inv}} = 0.6$		
f_{cut}	PFN	LightGBM	LL-HL Gap	PFN	LightGBM	LL-HL Gap	PFN	LightGBM	LL-HL Gap
0.00	0.908	0.895	0.013	0.853	0.829	0.024	0.788	0.739	0.049
0.05	0.866	0.861	0.005	0.822	0.803	0.019	0.776	0.736	0.040
0.10	0.847	0.848	−0.001	0.790	0.790	0.000	0.746	0.721	0.025
0.15	0.838	0.843	−0.005	0.784	0.785	−0.001	0.738	0.717	0.021

	t-channel								
	$r_{\text{inv}} = 0.0$			$r_{\text{inv}} = 0.3$			$r_{\text{inv}} = 0.6$		
f_{cut}	PFN	LightGBM	LL-HL Gap	PFN	LightGBM	LL-HL Gap	PFN	LightGBM	LL-HL Gap
0.00	0.825	0.817	0.008	0.748	0.737	0.011	0.662	0.647	0.0015
0.05	0.806	0.808	−0.002	0.754	0.755	−0.001	0.697	0.683	0.014
0.10	0.741	0.742	−0.001	0.662	0.663	−0.001	0.595	0.597	−0.002
0.15	0.731	0.740	−0.009	0.655	0.661	−0.006	0.593	0.596	−0.003

Table 6. Comparison of the performance difference between a PFN operating on low-level constituent information and a LightGBM model using high-level summary quantities, for several values of the jet trimming parameter f_{cut}. Jet constituents belonging to a subjet whose fraction of the jet p_{T} is below f_{cut} are dropped, which has the effect of removing lower-p_{T} constituents. Shown is the AUC of each model for s- and t-channel processes under three r_{inv} scenarios. Statistical uncertainty in each case is ± 0.002 with 95% confidence, measured using bootstrapping over 100 models.

trained on the trimmed constituents. Results for networks with varying f_{cut} thresholds are shown in table 6.

In each case, raising the f_{cut} threshold decreases the classification performance, as might be expected due to the removal of low-p_{T} information. Perhaps more interesting is the variation in the gap between the performance of the HL LightGBM model and the PFN operating on low-level constituents. In nearly every case, the gap grows as more low-p_{T} information is included, supporting the hypothesis that this is the origin of most of the information missing from the HL models. The details of the low-p_{T} constituents are likely to be very sensitive to modeling uncertainties and subject to concerns about infrared and collinear safety.

However, even in for the most aggressive value of $f_{\text{cut}} = 0.15$, a persistent gap of $\Delta\text{AUC} = 0.021$ remains in the s-channel, $r_{\text{inv}} = 0.6$ scenario, which cannot be explained by low-p_{T} constituents. We therefore examine this in more detail.

First, we consider the EFPs selected in the study above where $f_{\text{cut}} = 0.05$, including EFPs selected by the guided search (table 4) and the greedy search (table 5). For all combinations of HL and identified EFPs, no performance gain is seen. Next, we perform a fresh guided search on models trained from the $f_{\text{cut}} = 0.15$ constituents. In contrast to previous cases, no large improvement in training performance is obtained on the first selected EFP. After 200 iterations, the gap is reduced to ΔAUC=0.010 with the addition of 200 EFPs. We note that this far exceeds the mean number of constituents of these trimmed jets, 60. We conclude that there does not appear to be a compact representation of the remaining information in the EFP space we have explored.

JHEP12(2022)132

9 Conclusions

We have analyzed the classification performance of models trained to distinguish background jets from semi-visible jets using the low-level jet constituents, and found them to offer stronger performance than models which rely on high-level quantities motivated by other processes, mostly those involving collimated hadronic decays of massive objects.

While models operating on the existing suite of HL quantities nearly match the performance of those using LL information, a significant gap exists which suggests that relevant information remains to be captured, perhaps in new high-level features. To our knowledge, this is the first study to compare the performance of constituent-based and high-level semi-visible-jet taggers, and to identify the existence of relevant information uncaptured by existing high-level features. Jets due to semi-visible decays are markedly different in energy distribution than those from massive objects, so it is not unexpected that existing features may not completely summarize the useful information.

Using a guided strategy, we identify a small set of new useful features from the space of energy-flow polynomials, but these do not succeed in completely closing the performance gap. In most cases, the remaining gap seems to be due to information contained in very low-p_{T} constituents, which is likely to be sensitive to modeling of showering and hadronization and may not be infrared and collinear safe. This highlights the importance of interpretation and validation of information used by constituent-based taggers. As demonstrated by ref. [26], the specific pattern of energy deposition may depend sensitively on both the parameters of the theoretical model as well as the settings chosen for the hadronization model. The results drawn from this study might not be universally generalizable to other dark shower models. It is therefore vital that the information being analyzed be interpreted before being applied to analysis of collider data.

In one case studied here, a gap persist between low- and high-level-based models even when low-p_{T} constituents are aggressively trimmed, suggesting the possibility that a new high-level feature could be crafted to capture this useful high-p_{T} information. Our efforts to capture this information with the simpler energy-flow polynomials was not successful, suggesting that more complex high-p_{T} observables may exist which provide useful discrimination between QCD and semi-visible jets. The studies presented here can inform and guide theoretical work to construct such observables specifically tailored to this category of jets. Whether such observables can be efficiently represented using alternative basis sets of observables, and whether they are robust to [53] or explicitly dependent on [54] uncertainties while providing power over large regions of theoretical parameter space is an important avenue for future work.

Acknowledgments

The authors thank Po-Jen Cheng for his assistance in sample validation. This material is based upon research supported by the Chateaubriand Fellowship of the Office for Science & Technology of the Embassy of France in the United States. T. Faucett and D. Whiteson are supported by the U.S. Department of Energy (DOE), Office of Science under Grant

JHEP12(2022)132

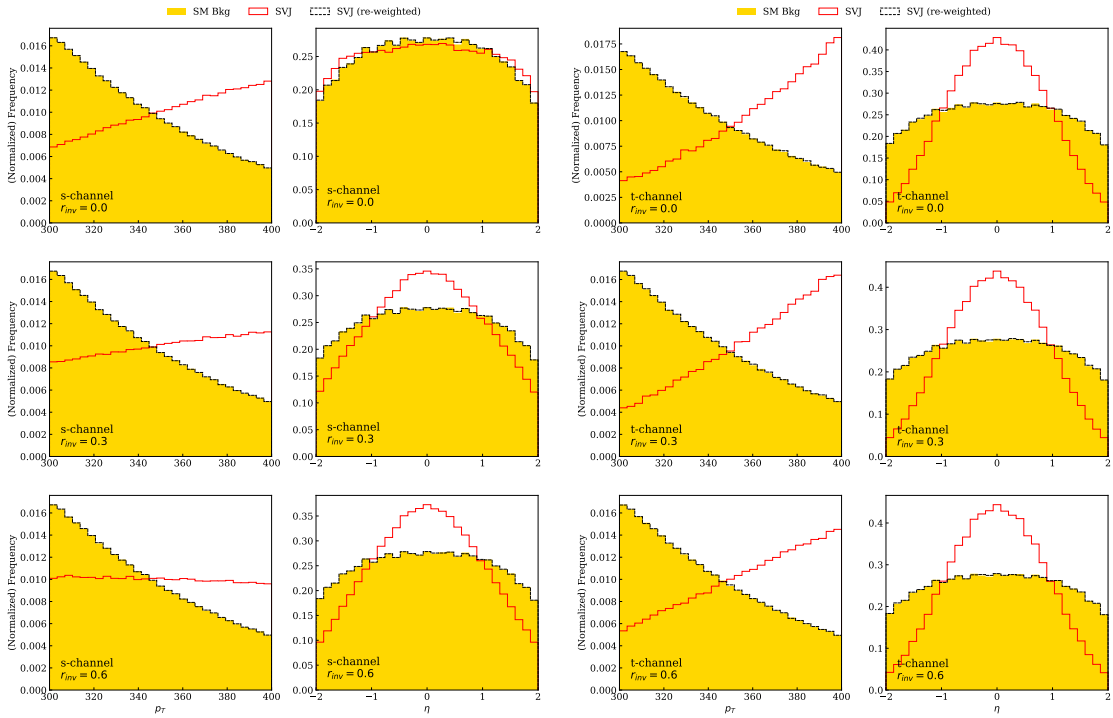


Figure 7. Distributions of jet transverse momentum (p_{T}) and pseudo-rapidity (η) shown for semi-visible jet (SVJ) and the standard model background (SM Bkg) for the six simulated scenarios, three choices of invisible fraction r_{inv} for both the s-channel and t-channel processes. The SVJ samples are reweighted to match the background distributions.

No. DE-SC0009920. S.-C. Hsu is supported by the National Science Foundation under Grant No. 2110963.

A Signal reweighting

The signal and background distributions have distinct transverse momentum and pseudo-rapidity distributions due to the processes used to generate them. We wish to learn to classify the signal and background independent of these quantities, and so reweight the signal events to match the background distribution, see figure 7.

B Jet substructure observables

High-level features used to discriminate between semi-visible and background jets are defined below.

B.1 Jet transverse momentum and mass

The sum of jet p_{T} constituents is included as a HL observable both in the initial HL inputs and along with EFPs to give ML algorithms a relative scale for dimensionless EFP features

JHEP12(2022)132

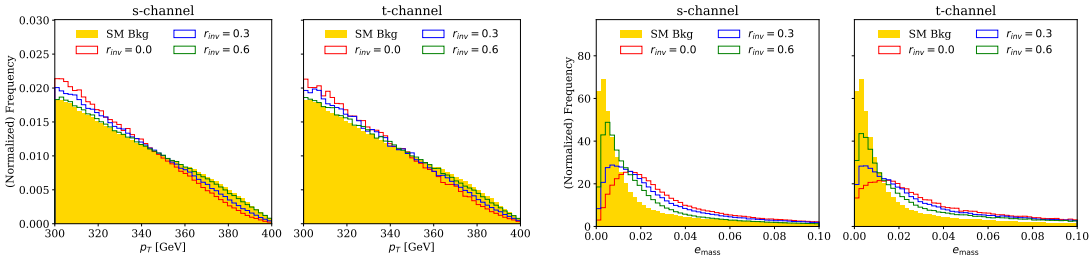


Figure 8. Distributions of jet transverse momentum (p_{T}) and e_{mass} shown for semi-visible jet and the standard model background (SM Bkg) for the six simulated scenarios, three choices of invisible fraction r_{inv} for both the s-channel and t-channel processes.

to train with. The jet p_{T} sum is calculated by

$$p_{\mathrm{T}} = \sum_{i \in \mathrm{jet}} p_{\mathrm{T},i} \,. \tag{B.1}$$

Distributions for jet p_{T} and e_{mass}, defined below, are shown in figure 8.

B.2 Generalized angularities

Multiple standard HL observables are defined by choices of κ and β parameters from the momentum fraction (z_i) and angular separation (θ_i) of a Generalized Angularity (GA) expression [43],

$$\lambda^{\kappa}_{\beta} = \sum_{i \in \mathrm{jet}} z_i^{\kappa} \theta_i^{\beta} \,. \tag{B.2}$$

The Les Houches Angularity (LHA) is defined from the GA expression with parameters ($\kappa = 1, \beta = 1/2$) and p_{T}^{D} with ($\kappa = 2, \beta = 0$). Written explicitly, these become

$$\mathrm{LHA} = \sum_{i \in \mathrm{jet}} z_i \theta_i^{1/2} \tag{B.3}$$

$$p_{\mathrm{T}}^{D} = \sum_{i \in \mathrm{jet}} z_i^2 \,. \tag{B.4}$$

Two additional values, e_{width} and e_{mass}, are produced by choices of ($\kappa = 1, \beta = 1$) and ($\kappa = 1, \beta = 2$), respectively

$$e_{\mathrm{width}} = \sum_{i \in \mathrm{jet}} z_i \theta_i \tag{B.5}$$

$$e_{\mathrm{mass}} = \sum_{i \in \mathrm{jet}} z_i \theta_i^2 \,. \tag{B.6}$$

Lastly, the multiplicity (although technically defined as simply the total number of constituents in the jet) can be expressed in this same generalized form for ($\kappa = 0, \beta = 0$)

$$\mathrm{multiplicity} = \sum_{i \in \mathrm{jet}} 1 \,. \tag{B.7}$$

Distributions for all GA observables are shown in figure 9

JHEP12(2022)132

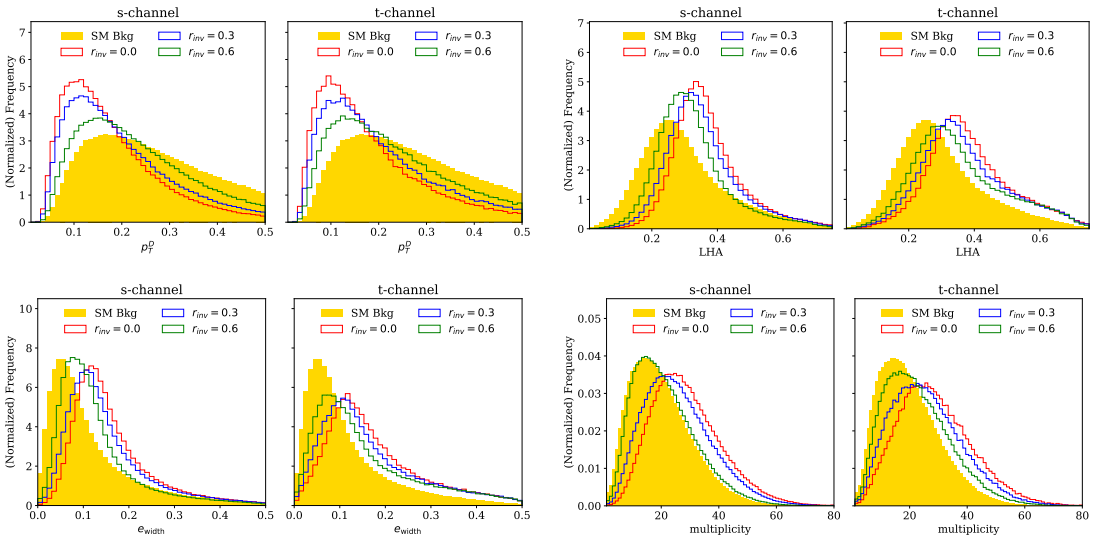


Figure 9. Distributions of jet p_T^D, LHA, width and multiplicity shown for semi-visible jet (red, green, blue) and the standard model background (SM Bkg, yellow) for the six simulated scenarios, three choices of invisible fraction $r_{\rm inv}$ for both the s-channel and t-channel processes.

B.3 Energy correlation

Energy Correlation Functions [19] and their corresponding ratios are computed via the functions: $\mathrm{ECF}_1, \mathrm{ECF}_2$ and ECF_3,

$$\mathrm{ECF}_1 = \sum_i p_{\mathrm{T},i} \tag{B.8}$$

$$\mathrm{ECF}_2^\beta = \sum_{i<j} p_{\mathrm{T},i}\, p_{\mathrm{T},j}\, (\theta_{ij})^\beta \tag{B.9}$$

$$\mathrm{ECF}_3^\beta = \sum_{i<j<k} p_{\mathrm{T},i}\, p_{\mathrm{T},j}\, p_{\mathrm{T},k}\, (\theta_{ij}\theta_{ik}\theta_{jk})^\beta \tag{B.10}$$

and the related ratios are given by,

$$e_2^\beta = \frac{\mathrm{ECF}_2^\beta}{(\mathrm{ECF}_1)^2} \tag{B.11}$$

$$e_3^\beta = \frac{\mathrm{ECF}_3^\beta}{(\mathrm{ECF}_1)^3} \tag{B.12}$$

from these ratios, we then compute the energy correlation ratios C_2 and D_2

$$C_2 = \frac{e_3}{(e_2)^2} \tag{B.13}$$

$$D_2 = \frac{e_3}{(e_2)^3}\,. \tag{B.14}$$

Distributions for all Energy Correlation observables are shown in figure 10

JHEP12(2022)132

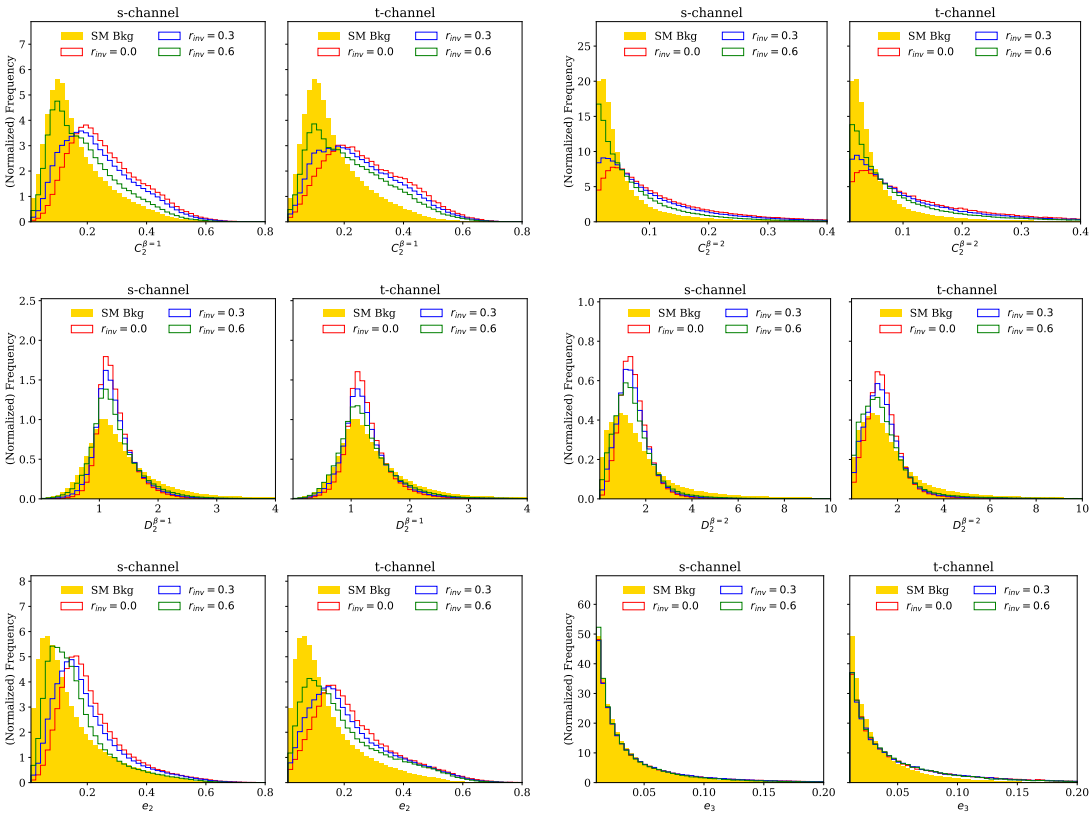


Figure 10. Distributions of jet energy correlation functions $\left(C_2^{\beta=1}, C_2^{\beta=2}, D_2^{\beta=1} \text{ and } D_2^{\beta=2}\right)$ and pairs e_2 and e_3 shown for semi-visible jet (red, green, blue) and the standard model background (SM Bkg, yellow) for the six simulated scenarios, three choices of invisible fraction r_{inv} for both the s-channel and t-channel processes.

B.4 N-subjettiness

Given subjets isolated via clustering, for N candidate subjets, the N-subjettiness (τ_N) [44] is defined as,

$$\tau_N = \frac{1}{d_0} \sum_k p_{\mathrm{T},k} \min\left(\Delta\theta_{1,k}, \Delta\theta_{2,k}, \ldots, \Delta\theta_{N,k}\right) \tag{B.15}$$

where we define the normalization factor d_0 by,

$$d_0 = \sum_k p_{\mathrm{T},k}\, R_0 \tag{B.16}$$

where R_0 is the characteristic jet radius used during clustering. Finally, the N-subjettiness ratios used are defined by

$$\tau_{21}^{\beta=1} = \frac{\tau_2}{\tau_1} \tag{B.17}$$

$$\tau_{32}^{\beta=1} = \frac{\tau_3}{\tau_2}\,. \tag{B.18}$$

Distributions for both N-subjettiness observables are shown in figure 11

JHEP12(2022)132

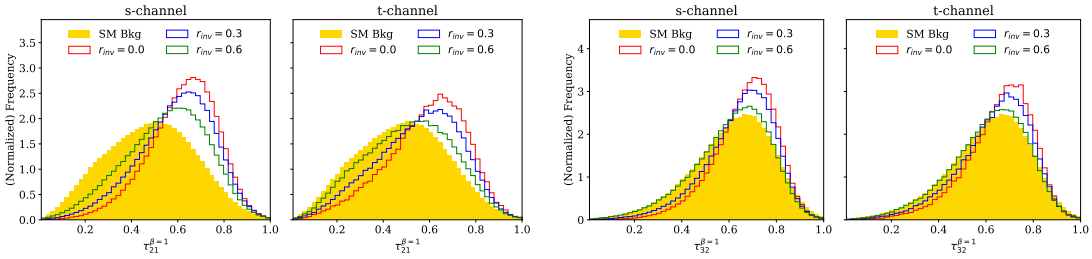


Figure 11. Distributions of jet N-subjettines $\left(\tau_{21}^{\beta=1} \text{ and } \tau_{32}^{\beta=1}\right)$ shown for semi-visible jet (red, green, blue) and the standard model background (SM Bkg, yellow) for the six simulated scenarios, three choices of invisible fraction r_{inv} for both the s-channel and t-channel processes.

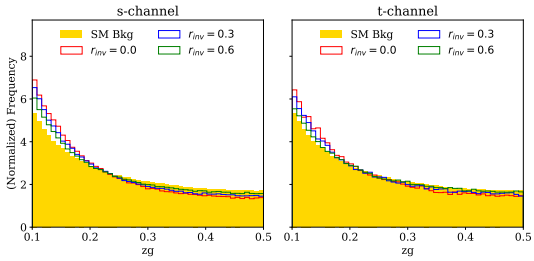


Figure 12. Distributions of jet splitting function z_g shown for semi-visible jet (red, green, blue) and the standard model background (SM Bkg, yellow) for the six simulated scenarios, three choices of invisible fraction r_{inv} for both the s-channel and t-channel processes.

B.5 Groomed momentum splitting fraction

The splitting fraction is described in terms of the Soft Drop grooming technique in ref. [18]. The feature is calculated using energyflow [27] with the Cambridge/Aachen algorithm using a jet radius of $R = 1$ and Soft Drop parameters of $\beta = 0$ and $z_{\text{cut}} = 0.1$.

A distribution for z_g is given in figure 12

C ML architectures

C.1 Deep neural networks

All deep neural networks were trained in Tensorflow [55] and Keras [56]. The networks were optimized with Adam [57] for up to 100 epochs with early stopping. For all networks, weights were initialized using orthogonal weights [58]. Hyperparameters were optimized using bayesian optimization with the Sherpa hyperparameter optimization library [59].

C.2 High-level DNN

All HL features are preprocessed with Scikit's Standard Scaler [60] before training.

C.2.1 Deep neural networks

Hyperparameters and network design for all Dense networks trained on HL or EFP features are selected via Sherpa optimization using between two and eight fully connected

JHEP12(2022)132

hidden layers and a final layer with a sigmoidal logistic activation function to predict the probability of signal or background.

C.2.2 Particle-flow networks

The Particle Flow Network (PFN) is trained using the energyflow package [51]. Input features are taken from the trimmed jet constituents and preprocessed by centering the constituents in $(\eta - \varphi)$ space to the average p_{T} and normalizing constituent values to 1. Both the EFN and PFN use this constituent information as inputs in the form of the 3-component `hadronic measure` measure option in EnergyFlow (i.e. p_{T}, η, φ).

The PFN uses 3 dense layers in the per-particle frontend module and 3 dense layers in the backend module. Both frontend and backend layers use 300 hidden nodes per layer with a latent and filter dropout of 0.2. Each layer uses relu [61] activation and glorot normal initializer. The final output layer uses a sigmoidal logistic activation function to predict the probability of signal or background. The Adam optimizer is used and trained with a batch size of 128 and a fixed learning rate of 0.001.

C.3 Boosted learning models

HL features are, again, preprocessed with Scikit's Standard Scaler before training. Except where indicated, default settings are used.

C.3.1 LightGBM

All applications of LightGBM are trained using regression for binary log loss classification using Gradient Boosting Decision Trees. Performance is measured by the AUC metric for a maximum of 5000 boosting rounds and early stopping set to 100 rounds against AUC improvements.

C.3.2 XGBoost

All applications of XGBoost are trained using the gradient tree booster and settings of: $\eta = 0.1$, subsample $= 0.5$, `base_score` $= 0.1$, $\gamma = 0.0$, and `max_depth` $= 6$. Performance is measured by the AUC metric for a maximum of 5000 boosting rounds and early stopping set to 100 rounds against AUC improvements.

C.4 Convolutional networks on jet images

The Convolutional Neural Networks used a jet image produced through EnergyFlow's `pixelate` function. Jet images were produced as a 32×32 pixel matrix with an image width of 1.0. Resulting jet images were then normalized to a range of values between $[0, 1]$. The network consisted of 3 hidden layers consisting of 300 nodes and used kernels of size 3×3 and strides of 1×1. Each layer uses relu [61] activation and glorot normal initializer. The final output layer uses a sigmoidal logistic activation function to predict the probability of signal or background. The Adam optimizer is used and trained with a batch size of 128 and a fixed learning rate of 0.001.

JHEP12(2022)132

References

[1] G. Bertone and D. Hooper, *History of dark matter*, *Rev. Mod. Phys.* **90** (2018) 045002 [arXiv:1605.04909] [INSPIRE].

[2] G. Bertone and T.M.P. Tait, *A new era in the search for dark matter*, *Nature* **562** (2018) 51 [arXiv:1810.01668] [INSPIRE].

[3] CDF collaboration, *A Search for dark matter in events with one jet and missing transverse energy in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Rev. Lett.* **108** (2012) 211804 [arXiv:1203.0742] [INSPIRE].

[4] ATLAS collaboration, *Search for dark matter candidates and large extra dimensions in events with a jet and missing transverse momentum with the ATLAS detector*, *JHEP* **04** (2013) 075 [arXiv:1210.4491] [INSPIRE].

[5] CMS collaboration, *Search for Dark Matter and Large Extra Dimensions in Monojet Events in pp Collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **09** (2012) 094 [arXiv:1206.5663] [INSPIRE].

[6] ATLAS collaboration, *Search for dark matter candidates and large extra dimensions in events with a photon and missing transverse momentum in pp collision data at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys. Rev. Lett.* **110** (2013) 011802 [arXiv:1209.4625] [INSPIRE].

[7] CMS collaboration, *Search for Dark Matter and Large Extra Dimensions in pp Collisions Yielding a Photon and Missing Transverse Energy*, *Phys. Rev. Lett.* **108** (2012) 261803 [arXiv:1204.0821] [INSPIRE].

[8] Y. Bai and T.M.P. Tait, *Searches with Mono-Leptons*, *Phys. Lett. B* **723** (2013) 384 [arXiv:1208.4361] [INSPIRE].

[9] F.J. Petriello, S. Quackenbush and K.M. Zurek, *The Invisible Z' at the CERN LHC*, *Phys. Rev. D* **77** (2008) 115020 [arXiv:0803.4005] [INSPIRE].

[10] D. Ferreira de Lima, P. Petrov, D. Soper and M. Spannowsky, *Quark-Gluon tagging with Shower Deconstruction: Unearthing dark matter and Higgs couplings*, *Phys. Rev. D* **95** (2017) 034001 [arXiv:1607.06031] [INSPIRE].

[11] LHCb collaboration, *Search for Dark Photons Produced in 13 TeV pp Collisions*, *Phys. Rev. Lett.* **120** (2018) 061801 [arXiv:1710.02867] [INSPIRE].

[12] M.J. Strassler and K.M. Zurek, *Echoes of a hidden valley at hadron colliders*, *Phys. Lett. B* **651** (2007) 374 [hep-ph/0604261] [INSPIRE].

[13] T. Han, Z. Si, K.M. Zurek and M.J. Strassler, *Phenomenology of hidden valleys at hadron colliders*, *JHEP* **07** (2008) 008 [arXiv:0712.2041] [INSPIRE].

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

[15] T. Cohen, M. Lisanti, H.K. Lou and S. Mishra-Sharma, *LHC Searches for Dark Sector Showers*, *JHEP* **11** (2017) 196 [arXiv:1707.05326] [INSPIRE].

JHEP12(2022)132

[16] D. Kar and S. Sinha, *Exploring jet substructure in semi-visible jets*, *SciPost Phys.* **10** (2021) 084 [arXiv:2007.11597] [INSPIRE].

[17] E. Bernreuther, T. Finke, F. Kahlhoefer, M. Krämer and A. Mück, *Casting a graph net to catch dark showers*, *SciPost Phys.* **10** (2021) 046 [arXiv:2006.08639] [INSPIRE].

[18] A.J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft Drop*, *JHEP* **05** (2014) 146 [arXiv:1402.2657] [INSPIRE].

[19] A.J. Larkoski, G.P. Salam and J. Thaler, *Energy Correlation Functions for Jet Substructure*, *JHEP* **06** (2013) 108 [arXiv:1305.0007] [INSPIRE].

[20] R. Kogler et al., *Jet Substructure at the Large Hadron Collider: Experimental Review*, *Rev. Mod. Phys.* **91** (2019) 045003 [arXiv:1803.06991] [INSPIRE].

[21] Y. Lu, A. Romero, M.J. Fenton, D. Whiteson and P. Baldi, *Resolving extreme jet substructure*, *JHEP* **08** (2022) 046 [arXiv:2202.00723] [INSPIRE].

[22] 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** (2016) 094034 [arXiv:1603.09349] [INSPIRE].

[23] Y.-C.J. Chen, C.-W. Chiang, G. Cottin and D. Shih, *Boosted W and Z tagging with jet charge and deep learning*, *Phys. Rev. D* **101** (2020) 053001 [arXiv:1908.08256] [INSPIRE].

[24] ATLAS collaboration, *Measurement of jet-substructure observables in top quark, W boson and light jet production in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **08** (2019) 033 [arXiv:1903.02942] [INSPIRE].

[25] Y.-L. Chung, S.-C. Hsu and B. Nachman, *Disentangling Boosted Higgs Boson Production Modes with Machine Learning*, *JINST* **16** (2021) P07002 [arXiv:2009.05930] [INSPIRE].

[26] T. Cohen, J. Doss and M. Freytsis, *Jet Substructure from Dark Sector Showers*, *JHEP* **09** (2020) 118 [arXiv:2004.00631] [INSPIRE].

[27] 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].

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

[29] T. Faucett, J. Thaler and D. Whiteson, *Mapping Machine-Learned Physics into a Human-Readable Space*, *Phys. Rev. D* **103** (2021) 036020 [arXiv:2010.11998] [INSPIRE].

[30] J. Collado, J.N. Howard, T. Faucett, T. Tong, P. Baldi and D. Whiteson, *Learning to identify electrons*, *Phys. Rev. D* **103** (2021) 116028 [arXiv:2011.01984] [INSPIRE].

[31] J. Collado, K. Bauer, E. Witkowski, T. Faucett, D. Whiteson and P. Baldi, *Learning to isolate muons*, *JHEP* **10** (2021) 200 [arXiv:2102.02278] [INSPIRE].

[32] L. Bradshaw, S. Chang and B. Ostdiek, *Creating simple, interpretable anomaly detectors for new physics in jet substructure*, *Phys. Rev. D* **106** (2022) 035014 [arXiv:2203.01343] [INSPIRE].

[33] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, *MadGraph 5: Going Beyond*, *JHEP* **06** (2011) 128 [arXiv:1106.0522] [INSPIRE].

[34] P. Skands, S. Carrazza and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 Tune*, *Eur. Phys. J. C* **74** (2014) 3024 [arXiv:1404.5630] [INSPIRE].

JHEP12(2022)132

[35] M.L. Mangano, M. Moretti and R. Pittau, *Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b}+n$ jets as a case study*, Nucl. Phys. B **632** (2002) 343 [hep-ph/0108069] [INSPIRE].

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

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

[38] D. Krohn, J. Thaler and L.-T. Wang, *Jet Trimming*, JHEP **02** (2010) 084 [arXiv:0912.1342] [INSPIRE].

[39] M. Cacciari, G.P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063 [arXiv:0802.1189] [INSPIRE].

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

[41] S. Marzani, G. Soyez and M. Spannowsky, *Looking inside jets: an introduction to jet substructure and boosted-object phenomenology*, in *Lecture Notes in Physics* **958**, Springer (2019) [DOI:10.1007/978-3-030-15709-8] [arXiv:1901.10342] [INSPIRE].

[42] A. Larkoski, S. Marzani, J. Thaler, A. Tripathee and W. Xue, *Exposing the QCD Splitting Function with CMS Open Data*, Phys. Rev. Lett. **119** (2017) 132003 [arXiv:1704.05066] [INSPIRE].

[43] A.J. Larkoski, J. Thaler and W.J. Waalewijn, *Gaining (Mutual) Information about Quark/Gluon Discrimination*, JHEP **11** (2014) 129 [arXiv:1408.3122] [INSPIRE].

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

[45] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, in proceedings of the *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, San Francisco, CA, U.S.A., 13–17 August 2016, Association for Computing Machinery, New York, NY, U.S.A. (2016), pp. 785–794 [arXiv:1603.02754] [INSPIRE].

[46] D. Bourilkov, *Machine and Deep Learning Applications in Particle Physics*, Int. J. Mod. Phys. A **34** (2020) 1930019 [arXiv:1912.08245] [INSPIRE].

[47] A.S. Cornell, W. Doorsamy, B. Fuks, G. Harmsen and L. Mason, *Boosted decision trees in the era of new physics: a smuon analysis case study*, JHEP **2022** (2022) 15 [arXiv:2109.11815].

[48] G. Ke et al., *Lightgbm: A highly efficient gradient boosting decision tree*, in *Advances in Neural Information Processing Systems* **30**, I. Guyon et al. eds., Curran Associates, Inc. (2017) and online pdf vdersion at https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

[49] J. Cogan, M. Kagan, E. Strauss and A. Schwarztman, *Jet-Images: Computer Vision Inspired Techniques for Jet Tagging*, JHEP **02** (2015) 118 [arXiv:1407.5675] [INSPIRE].

[50] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *Jet-images — deep learning edition*, JHEP **07** (2016) 069 [arXiv:1511.05190] [INSPIRE].

[51] P.T. Komiske, E.M. Metodiev and J. Thaler, *Energy Flow Networks: Deep Sets for Particle Jets*, JHEP **01** (2019) 121 [arXiv:1810.05165] [INSPIRE].

[52] T. Buss et al., *What's Anomalous in LHC Jets?*, arXiv:2202.00686 [INSPIRE].

JHEP12(2022)132

[53] C. Shimmin et al., *Decorrelated Jet Substructure Tagging using Adversarial Neural Networks*, *Phys. Rev. D* **96** (2017) 074034 [arXiv:1703.03507] [INSPIRE].

[54] A. Ghosh, B. Nachman and D. Whiteson, *Uncertainty-aware machine learning for high energy physics*, *Phys. Rev. D* **104** (2021) 056026 [arXiv:2105.08742] [INSPIRE].

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

[56] F. Chollet et al., *Keras*, (2015) https://keras.io.

[57] D.P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv:1412.6980 [INSPIRE].

[58] A.M. Saxe, J.L. McClelland and S. Ganguli, *Exact solutions to the nonlinear dynamics of learning in deep linear neural networks*, arXiv:1312.6120.

[59] L. Hertel, J. Collado, P. Sadowski, J. Ott and P. Baldi, *Sherpa: Robust hyperparameter optimization for machine learning*, *SoftwareX* **12** (2020) 100591.

[60] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, *J. Mach. Learn. Res.* **12** (2011) 2825 [arXiv:1201.0490].

[61] V. Nair and G.E. Hinton, *Rectified linear units improve restricted boltzmann machines*, Omnipress (2010), pp. 807–814 and online pdf version at https://www.cs.toronto.edu/~fritz/absps/reluICML.pdf.

JHEP12(2022)132