Learning to identify electrons

Julian Collado,[1] Jessica N. Howard,[2] Taylor Faucett,[2] Tony Tong,[2] Pierre Baldi,[1,*] and Daniel Whiteson[2]
[1]*Department of Computer Science, University of California, Irvine, California 92627, USA*
[2]*Department of Physics and Astronomy, University of California, Irvine, California 92627, USA*

(Received 22 January 2021; accepted 21 May 2021; published 28 June 2021)

We investigate whether state-of-the-art classification features commonly used to distinguish electrons from jet backgrounds in collider experiments are overlooking valuable information. A deep convolutional neural network analysis of electromagnetic and hadronic calorimeter deposits is compared to the performance of typical features, revealing a ≈5% gap which indicates that these lower-level data do contain untapped classification power. To reveal the nature of this unused information, we use a recently developed technique to map the deep network into a space of physically interpretable observables. We identify two simple calorimeter observables which are not typically used for electron identification, but which mimic the decisions of the convolutional network and nearly close the performance gap.

DOI: 10.1103/PhysRevD.103.116028

I. INTRODUCTION

Production of electrons in high-energy collisions provides an essential handle on precision studies of the Standard Model [1,2] of particle physics as well as for searches for new physics [3,4]. The identification of electrons, and their separation from backgrounds which mimic their signature, is therefore a critical element in the data analysis toolkit, especially at lower transverse momentum, where the backgrounds rise rapidly [5].

In collider experiments, electrons are identified by an isolated track which aligns with a localized energy deposition, primarily in the electromagnetic calorimeter. The primary source of backgrounds is the production of hadronic jets, which typically feature multiple tracks and extended energy deposition in both electromagnetic and hadronic calorimeters, but can fluctuate to mimic electrons. The tracker and calorimeters, however, are very finely segmented, producing high-dimensional data which is difficult to analyze directly. A mature literature [6–8] contains higher-level features designed by physicists to highlight the distinct signature of the electron and suppress the backgrounds. The higher-level features define a lower-dimensional feature space.

Recent strides in machine learning for physics, particularly the advent of deep learning [9–15] and image-processing techniques [16–21], have demonstrated that high-level features designed by domain experts may not always fully capture the information available in the lower-level high-dimensional data. Specifically, the rich but subtle structure of the deposition of energy by jets provides a powerful potential handle for discrimination. Given their role as the dominant background, this suggests that additional classification power may be gained by applying image-based deep learning techniques to electrons.

In this study, we apply deep convolutional neural networks (CNNs) to the task of distinguishing between electrons and jets, using separate images from the electromagnetic and hadronic calorimeters. Due to the black-box nature of their operation, we do not propose to use CNNs in place of the high-level features. Instead, we apply CNNs to probe the information content of the low-level data in comparison to the high-level features. We show that the classification performance of the image-based CNNs exceeds the performance of the high-level features in common use by Large Hadron Collider (LHC) experiments, by a small, but significant, margin. We then identify the source of the untapped information and construct novel high-level features that capture it.

This paper is organized as follows. In Sec. II, we outline our approach. In Sec. III, we discuss the details of our image generation process and the corresponding dataset used for CNN experiments. In Sec. IV, we review the existing state-of-the-art ATLAS and CMS high-level features, which we combine to derive our benchmark performance. In Sec. V, we provide details of neural network architectures and training. In Sec. VI, we discuss the performance of these networks. In Sec. VII, we search for new high-level features to bridge the gap between CNNs and standard features. In Sec. VIII, we summarize and discuss the results, providing an intuitive understanding of the underlying landscape.

II. OVERVIEW

This study explores whether low-level, high-dimensional, $\mathcal{O}(10^3)$, calorimeter data contains information useful for

*pfbaldi@ics.uci.edu

distinguishing electrons from a major background not captured by the standard suite of high-level features designed by physicists. Similar studies in jets or flavor tags have revealed such gaps [10,17].

We probe this issue using a simulated dataset created with publicly available fast simulations tools [22]; while such samples do not typically match the fidelity of those generated with full simulations [23], we refine the calorimeter description for this study and find the modeling sufficiently realistic for a proof-of-principle analysis. Our focus is on comparing physically motivated, high-level features to low-level image techniques on equal footing. While we anticipate that the numerical results will be different when evaluated in a fully realistic scenario, the broad picture will likely remain the same. The technique described here is fairly general and applicable to more realistic experimental scenarios, so that valuable lessons can be learned in the present context.

We reproduce the standard suite of electron identification features, as described in Refs. [6,7], in the context of our simulated description. We then compare their combined performance to that of deep convolutional neural networks (CNNs) which have been trained to analyze the lower-level calorimeter cells using image recognition techniques [16–18]. We do not advocate for the use of CNNs to replace high-level features whose designs are grounded in physics; CNNs are difficult to interpret and the low level and large dimensionality of the input makes validation of the features and definition of systematic uncertainties nearly impossible. Instead, here we use the power of CNNs as a probe, to test whether further information is present in the low-level data. Having identified a gap, we then explore a complete space of novel high-level features, Energy flow polynomials (EFPs) [24] to interpret and bridge the gap.

III. DATASET GENERATION

In this section, we describe the process of generating simulated signal and background datasets, reproducing the standard suite of high-level features, and forming pixelated images from the electromagnetic and hadronic calorimeter deposits.

A. Processes and simulation

Simulated samples of isolated electrons are generated from the production and electronic decay of a Z' boson in hadronic collisions, $pp \to Z' \to e^+e^-$ at $\sqrt{s} = 13$ TeV. We set $m_{Z'}$ to 20 GeV in order to efficiently produce electrons in the range $p_{\mathrm{T}} = [10, 30]$ GeV, where hadronic backgrounds are significant. Simulated samples of background jets are generated via generic dijet production. Events were generated with MadGraph v2.6.5 [25], decayed and showered with PYTHIA v8.235 [26], with detector response described by DELPHES v3.4.1 [22] using ROOT version 6.0800 [27].

Our configuration of DELPHES approximates the ATLAS detector [28]. For this initial study, we model only the central layer of the calorimeters where most energy is deposited; future work will explore more detailed and realistic detector simulation. However, we maintain the critical separation between the electromagnetic and hadronic calorimeters and their distinct segmentation. Our simulated electromagnetic calorimeter (ECal) has segmentation of $(\Delta\phi, \Delta\eta) = (\frac{\pi}{126}, 0.025)$ while our simulated hadronic calorimeter (HCal) is coarser, $(\Delta\phi, \Delta\eta) = (\frac{\pi}{31}, 0.1)$. This approach allows us to investigate whether information about the structure of the many-particle jet is useful for suppressing their contribution. See Ref [19] for an analysis of the information contained in the shape of shower for individual particles.

No pile-up simulation was included in the generated data, as pileup subtraction techniques have been shown to be effective [29]. In total, we generated 107k signal and 107k background objects.

B. Electron candidate selection

We use DELPHES' standard electron identification procedures where loose electron candidates are selected from charged particle tracks which align with energy deposits in the ECal. We required the object to have track $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.125$, to avoid edge effects when forming calorimeter images, see Fig. 1. For later training, the background objects are reweighted to match the p_{T} distribution of the signal.

C. Image formation

The cells of the calorimeter can be naturally organized as pixels of an image, allowing for use of powerful image-processing techniques. Each pixel contains the energy (E) deposited in one cell. Alternatively, one may form images in which each cell represents $E_{\mathrm{T}} = E/\cosh\eta$, which folds in the location of the object relative to the collision point. For completeness we initially consider images in which pixels represent E and images where pixels represent E_{T}. Additionally, we create separate images for the ECal and HCal, in order to preserve the separate and powerful information they offer. In total, four images are created for each electron candidate: ECal E, ECal E_{T}, HCal E, HCal E_{T}. The pixels in the images were scaled to values between zero and one by dividing by the maximum pixel value in each image dataset. All standard classification and EFP variables were scaled by substracting the mean and dividing by the standard deviation.

The center of a calorimeter image is chosen to be the ECal cell with largest transverse energy deposit in the 9×9 cell region surrounding the track of the highest p_{T} electron in that event. This accounts for the curvature in the path of the electron as it propagates between the tracker and the calorimeter. The ECal image extends fifteen pixels in either

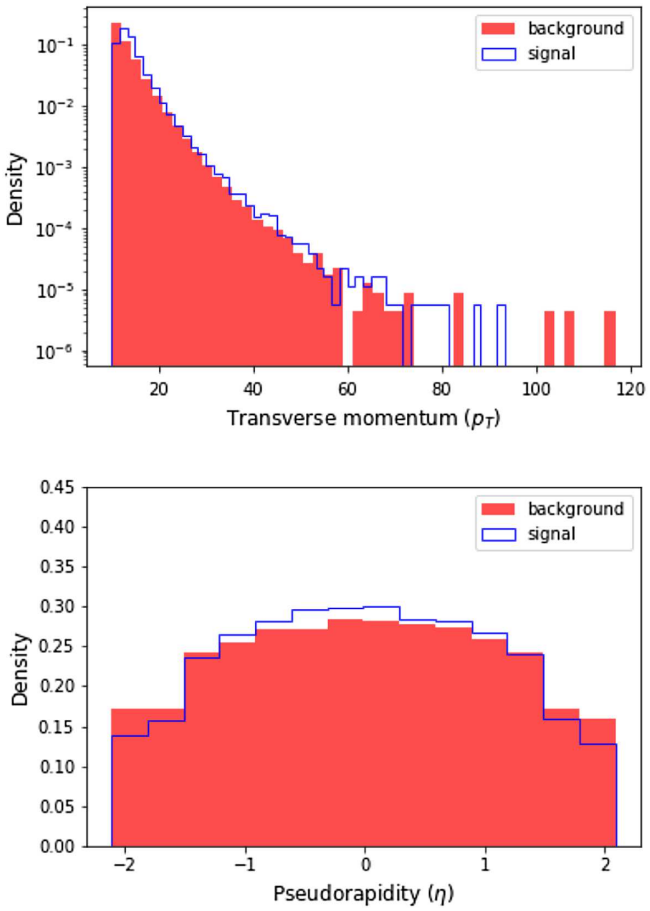

FIG. 1. Distribution of generated electron candidate p_T and η for simulated signal and background samples, before reweighting to match spectra.

direction, forming a 31×31 image. The HCal granularity is four times as course, and an 8×8 image covers the same physical region. Figures 2 and 3 show example and mean images for the ECal and HCal, respectively.

IV. STANDARD CLASSIFICATION FEATURES

To assess the performance of the high-level classifications features typically used by ATLAS [7] and CMS [6] which identify electrons and reject jet backgrounds, we reproduce their form here, where relevant.

Since electron candidates are confined to the longitudinal range $|\eta| < 2.125$, we only consider variables that are well-defined in this range. Additionally, we only consider variables which are based on information included in our simulation, to ensure the comparison uses information on equal footing. In addition, we do not perform clustering because it is unnecessary given the simplified nature of DELPHES' simulation; where a feature calls for the cluster energy, we replace it with the total energy of the candidate image. This is a reasonable proxy since the simplified simulation of the calorimeter response in

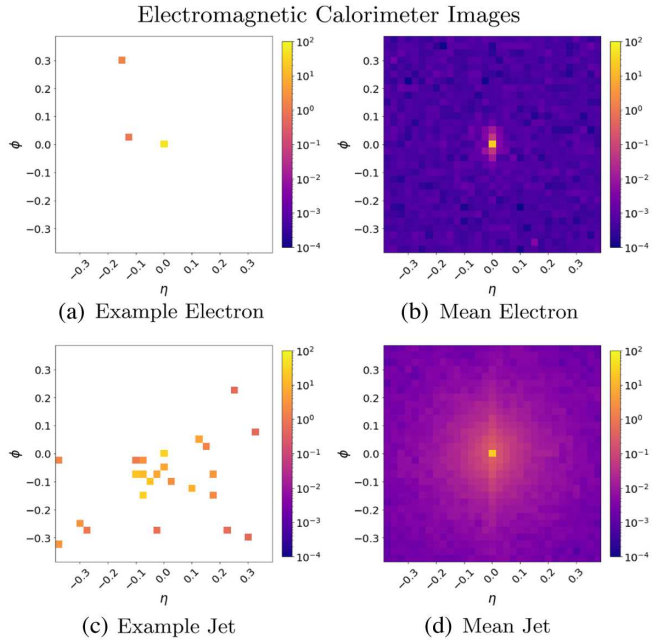

FIG. 2. Images in the electromagnetic calorimeter for signal electrons (top) and background jets (bottom). On the left are individual examples, on the right are mean images. See Fig. 3 for corresponding hadronic calorimeter images.

DELPHES is unlikely to deposit the electron's energy in multiple disconnected clusters. All high-level features are calculated from the ECal and HCal images, using E or E_T images where appropriate.

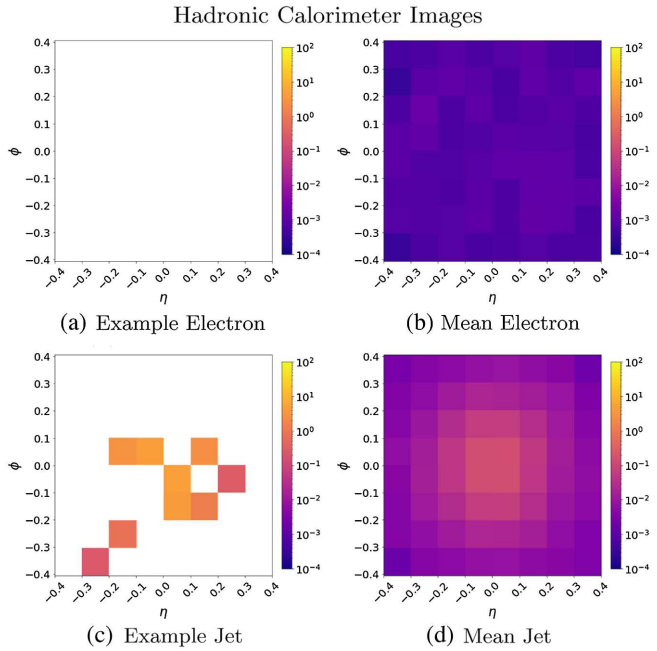

FIG. 3. Images in the hadronic calorimeter for signal electrons (top) and background jets (bottom). On the left are individual examples, on the right are mean images. See Fig. 2 for corresponding electromagnetic calorimeter images.

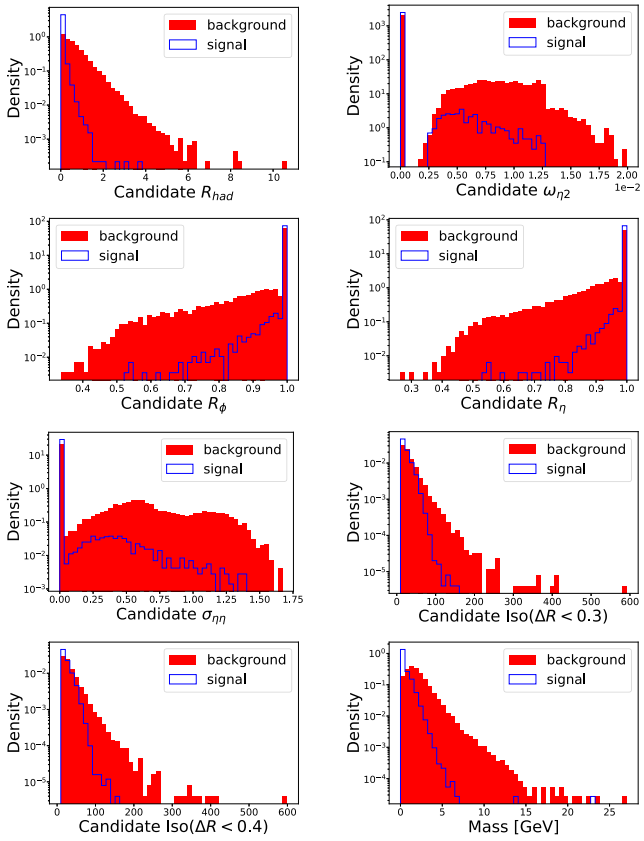

FIG. 4. Distribution of signal electron (red) and background jets (blue) for seven existing typically-used high-level features, as well as for mass.

We reproduce seven features: R_{had}, $\omega_{\eta 2}$, R_ϕ, R_η, $\sigma_{\eta\eta}$, and two isolation quantities. Together, these capture the typical strategies of suppressing objects with significant hadronic energy or extended energy deposits. Definitions of each feature are below, and distributions for signal and background samples are shown in Fig. 4.

A. Ratio of HCal and ECal energy: R_{had}

The feature R_{had} relates the transverse energy (E_{T}) in the electromagnetic calorimeter to that in the hadronic calorimeter. Specifically,

$$R_{\text{had}} = \frac{\Sigma_i E_{\text{T},i}^{\text{HCal}}}{\Sigma_j E_{\text{T},j}^{\text{ECal}}} \tag{1}$$

where i and j run over the pixels in the HCal and ECal images, respectively.

B. Lateral width of the ECal energy shower: $w_{\eta 2}$

The lateral width of the shower in the ECal, $w_{\eta 2}$, is calculated as

$$w_{\eta 2} = \sqrt{\frac{\Sigma_i E_i (\Delta\eta_i)^2}{\Sigma_i E_i} - \left(\frac{\Sigma_i E_i \Delta\eta_i}{\Sigma_i E_i}\right)^2} \tag{2}$$

where E_i is the energy of the ith pixel in the ECal image and $\Delta\eta_i$ is the pseudorapidity of the ith pixel in the ECal image measured relative to the image's center. The sum is calculated within an $(\Delta\eta \times \Delta\phi) = (3 \times 5)$ cell window centered on the image's center.

C. Azimuthal and longitudinal energy distributions: R_ϕ and R_η

To probe the distribution of energy in azimuthal (ϕ) and longitudinal (η) directions, we calculate two features R_ϕ and R_η. Qualitatively, these relate the total ECal energy in a subset of cells to the energy in a larger subset of cells extended in either ϕ or η, respectively. Specifically,

$$R_\phi = \frac{E_{3\times 3}}{E_{3\times 7}}, \qquad R_\eta = \frac{E_{3\times 7}}{E_{7\times 7}} \tag{3}$$

where the subscript indicates the number of cells included in the sum in η and ϕ respectively. For example, $(\eta \times \phi) = (3 \times 7)$ is a subset of cells which extends 3 cells in η and 7 in ϕ relative to the center of the image.

D. Lateral shower extension: $\sigma_{\eta\eta}$

An alternative probe of the distribution of energy in η is $\sigma_{\eta\eta}$ [6]. Specifically,

$$\sigma_{\eta\eta} = \sqrt{\frac{\Sigma_i w_i (i_\eta - \bar{i_\eta})^2}{\Sigma_i w_i}} \tag{4}$$

where w_i is the weighting factor $|\ln(E_i)|$ with E_i being the ECal energy of the ith pixel. The sum runs over the nonzero cells in the $(\eta \times \phi) = (5 \times 5)$ subset of cells centered on the highest energy cell in the ECal. Here, i_η is measured in units of cells away from center, $\bar{i_\eta}$, as $i_\eta \in 0, \pm 1$, or ± 2 if we choose $\bar{i_\eta} = 0$.

E. Isolation

Jets typically deposit significant energy surrounding the energetic core, while electrons from heavy boson decays are typically isolated in the calorimeter.[1] To assess the degree of isolation, we sum the ECal energy in cells within the angular range $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$ or 0.4, where $\Delta\eta$ and $\Delta\phi$ are measured from a given cell's center and the center of the image.

[1]Electrons may also appear inside jets in decays of B-mesons for example, but here we focus on decays from real W and Z bosons.

V. NEURAL NETWORK ARCHITECTURES AND TRAINING

We construct multilayer neural networks that accept low-level images, or high-level features, or both, with a sigmoid unit as their output to classify between signal and background.

Each image input is passed through a number of convolutional blocks, with each block consisting of two convolutional layers with 3×3 kernels, rectified linear units [30] as the activation function, and a final 2×2 maxpooling layer. Finally, the outputs of the maxpooling layer are flattened and concatenated with the high-level inputs to form a high-dimensional vector. This high-dimensional vector is then processed by a sequence of fully connected layers with rectified linear units, using dropout [31,32]. The final output is produced by a single logistic unit and it can be interpreted as the probability of the input belonging to the signal class. The entire architecture is trained by stochastic gradient descent to minimize the relative entropy between the targets and the outputs, across all training examples. For each combination of high-level variables, we also train and tune multilayer, fully connected, neural networks with a similar sigmoid unit at the top.

All models were implemented using KERAS [33] with TENSORFLOW [34] as the backend and trained with a batch size of 128 with the ADAM optimizer [35]. The weights for all the models were initialized using Glorot [36] uniform weights and each network was tuned using 150 iterations of bayesian optimizaton with the SHERPA hyperparameter optimization library [37]. Additional details about the hyperparameters and their optimization are given in Tables IV–VI.

VI. PERFORMANCE

Initial studies indicated that having images that reflect both E and E_{T} provided no performance boost, so only results with E_{T}-based images are shown here and used for further studies. A comparison of the performance of the image networks and the seven standard high-level features (R_{had}, $\omega_{\eta 2}$, R_ϕ, R_η, $\sigma_{\eta\eta}$, Iso($\Delta R < 0.3$), Iso($\Delta R < 0.4$)) is shown in Fig. 5 and described in Table I.

Networks combining the standard high-level features (AUC of 0.945) do not match the performance of a network which analyzes the lower-level data expressed as images (0.972), indicating that the images contain additional, untapped information relevant to the identification of electrons. This is not unexpected, and is in line with similar results for jet substructure or flavor tagging [10,17]. Networks which see only one of the ECal or HCal images but not both do not match this performance, supporting the intuition that both calorimeters contribute valuable information. Adding the HL features to the CNN, however, gives an almost negligible boost in performance, suggesting that the CNN has succeeded in capturing the power of the HL features.

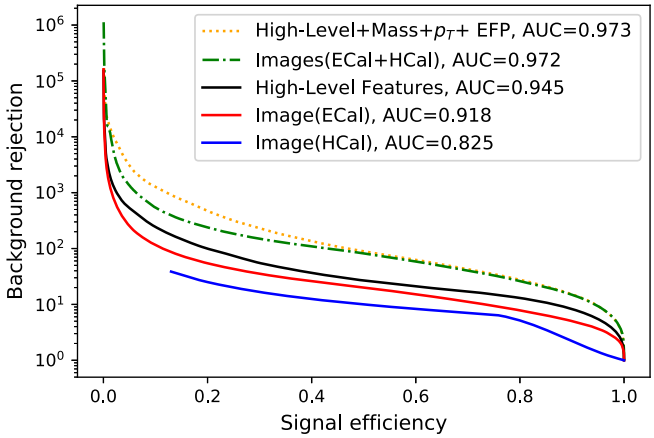

FIG. 5. Comparison of the performance in electron identification for networks with varying sets of input features. Shown is the signal efficiency versus background rejection, and the AUC, for networks which use a set of seven expert high-level features (see text for details), networks which use HCal and/or ECal images, and a network which combines the high-level features with jet mass, p_{T} and an energy-flow polynomial identified by a scan which aims to match the decisions of the image network.

VII. BRIDGING THE GAP

The performance of the deep CNN reveals that there is information in the low-level image that is not captured by the suite of existing high-level features. The goal, however, is not to replace the suite of features with an image-based network whose decisions are opaque to us and may not align with real physical principles. Instead, our aim is to identify new high-level features which bridge the gap between the existing performance and the superior performance of the CNN.

TABLE I. Electron classification power (AUC) for networks with various feature sets. Images refer to low-level pixel data. Standard features are the high-level (HL) features typically used [R_{had}, $\omega_{\eta 2}$, R_ϕ, R_η, $\sigma_{\eta\eta}$, Iso($\Delta R < 0.3$), Iso($\Delta R < 0.4$)], as described in the text. The uncertainty on AUC values was evaluated using bootstrapping to ± 0.001, unless otherwise specified.

Network Features				AUC
Images		7 Standard		
ECal	HCal	HL Features	M_{jet}	
	✓			0.82 ± 0.02
✓				0.918
✓	✓			0.972
✓	✓	✓		0.973
✓	✓	✓	✓	0.973
		✓		0.945
		✓	✓	0.956

We note that the design of the high-level features focuses on highlighting the characteristics of the signal electrons, localized energy depositions primarily in the ECal without significant structure. The background, however, is due to jets, which potentially can exhibit a rich structure and comprise a mixture of jets from gluons, light quarks, and heavy quarks. Each parton may produce jets with a distinct structure and varying probability to mimic electrons. We hypothesize that features which are sensitive to the structure of the jet, or subclasses of jets, may provide additional discrimination power.

We first consider the powerful feature of jet mass, M_{jet}, which is not often applied to electron identification, but has a distinct marginal distribution for electrons and jets, see Fig. 4. Including it in a network of HL features provides a small but distinct boost in performance, see Table I, indicating that it contains useful information for this classification task not duplicated by the standard seven HL features. This encourages us to explore further the space of jet observables as a way to understand the source of additional classification power of the CNN.

A. Set of observables

One could in principle consider an infinite number of jet observables. To organize our search, we use the energy flow polynomials (EFPs) [24], a large (formally infinite) set of parametrized engineered functions, inspired by previous work on energy correlation functions [38], which sum over the contents of the cells scaled by relative angular distances.

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

$$\text{each node} \Rightarrow \sum_{i=1}^{N} z_i, \tag{5}$$

$$\text{each edge} \Rightarrow (\theta_{ij})^k. \tag{6}$$

The observable corresponding to each graph can be modified with parameters (κ, β), where

$$(z_i)^\kappa = \left(\frac{p_{\text{T}i}}{\sum_j p_{\text{T}j}}\right)^\kappa, \tag{7}$$

$$\theta_{ij}^\beta = (\Delta\eta_{ij}^2 + \Delta\phi_{ij}^2)^{\beta/2}. \tag{8}$$

Here, $p_{\text{T}i}$ is the transverse momentum of cell i, and $\Delta\eta_{ij}$ ($\Delta\phi_{ij}$) is pseudorapidity (azimuth) difference between cells i and j. The original IRC-safe EFPs require $\kappa = 1$, however we consider examples with $\kappa \neq 1$ to explore a broader space of observables. Also, note that $\kappa > 0$ generically corresponds to IR-safe but C-unsafe observables.[2]

[2]For $\kappa < 0$, empty cells are omitted from the sum.

In principle, the space is complete, such that any jet observable can be described by one or more EFPs of some degree; in practice, the space is infinite and only a finite subset can be explored. We consider EFPs with up to seven edges and with β values of $[\frac{1}{2}, 1, 2]$ and κ values of $[-1, 0, 1, 2]$. We consider each graph as applied to the ECal or the HCal separately, effectively doubling the number of graphs, for a total of 12,072.[3]

B. Searching for observables

Rather than conduct a brute-force search of this large space, we aim to leverage the success of the CNN and find observables which mimic its decisions. We follow the black-box guided algorithm of Ref. [39], which isolates the portion of the input space where the CNN and existing HL features disagree and searches for a new observable that matches the decisions of the CNN algorithm in that subspace.

The subspace is defined as input pairs (x, x') that have a different relative ordering between the CNN and the network of n HL features (HLN_n). Mathematically, we express this using the *decision ordering* (DO)

$$\text{DO}[f, g](x, x') = \Theta((f(x) - f(x'))(g(x) - g(x'))), \tag{9}$$

where $f(x)$ and $g(x)$ are classification functions such as the CNN or the HLN_n, such that $\text{DO} = 0$ corresponds to inverted ordering and $\text{DO} = 1$ corresponds to the same ordering. The focus of our investigation are the set of pairs X_n where the two classifiers disagree, defined as

$$X_n = \{(x, x') | \text{DO}[\text{CNN}, \text{HL}_n](x, x') = 0\}. \tag{10}$$

As prescribed in Ref. [39], we scan a sub-space of EFPs to find the observable that has the highest average decision ordering (ADO) with the CNN when averaged over the disordered pairs X_n. The selected EFP is then incorporated into the new network of HL features, HLN_{n+1}, and the process is repeated until the ADO or AUC plateaus.

For all HLN_n used in this search, models were trained with KERAS [33] using TENSORFLOW [34] as the backend. Each model was built as a fully connected neural network of simple one dimensional input features and a single logistic unit output. The guided search requires training a new HLN_n after each new EFP selection. Performing a full Bayesian optimization with Sherpa and bootstrapping each network becomes computationally expensive. Instead, a simpler architecture was found to be provide consistent, stable, performance. These networks consisted of 3 hidden layers, each with 50 rectified linear units, separated by 2

[3]We also explored a version where ECal and HCal information were used simultaneously by each graph, but found no improvement.

dropout layers using a dropout value of 0.25 and trained with a batch size of 128.

The ADAM optimizer [35] was used with learning rate of 0.001 and initialized with Glorot [36] normal weights.

C. IRC safe observables

We begin our search by considering only the observables which are IRC safe, with $\kappa = 1$, a total of 3,018 graphs. Beginning with the seven HL features, the first graph selected is

$$(\kappa=1,\beta=\tfrac{1}{2}) = \sum_{a,b=1}^{N} z_a z_b \theta_{ab}^{\frac{1}{2}}$$

This graph has an ADO of 0.802 with the CNN over the input subspace where the CNN disagrees with the seven HL, suggesting that it is well aligned with the CNN's strategy. Adding it to the seven HL features achieves an AUC of 0.970 ± 0.001, very nearly closing the gap with the CNN performance of 0.972. This graph is very closely related to jet mass, a pairwise sum over cells which folds in angular separation, but more closely resembles the Les Houches Angularity variable [40], which similarly is sensitive to the distribution of energy away from the center, though with a smaller power of the angularity than jet mass, which suggests that it enhances small angles. Additional scans do not identify EFP observables with a useful ADO and do not contribute to the AUC, within our defined EFP subspace.

If instead, we begin with the seven HL features as well as the jet mass, the procedure selects two graphs:

$$(\kappa=1,\beta=1) = \sum_{a\cdots h=1}^{N} z_a \ldots z_h \theta_{ab}\theta_{ac}\theta_{ad}\theta_{ae}\theta_{af}\theta_{ag}\theta_{ah}$$

and

$$(\kappa=1,\beta=\tfrac{1}{2}) = \sum_{a,b,c=1}^{N} z_a z_b z_c \theta_{ab}^{\frac{1}{2}}\theta_{bc}^{\frac{1}{2}}\theta_{ac}^{\frac{1}{2}}$$

When combined with the seven HL features and M_{jet}, this set of ten observables achieves an AUC of 0.971 ± 0.001, almost matching the CNN performance. Distributions of these observables for signal and background samples are shown in Fig. 6.

As the EFPs are normalized, they are sensitive to relative distributions of energy rather than the overall scale. As suggested in Ref. [39], we add the jet p_{T} observable to provide this information, which when combined with the M_{jet} and seven HL features, achieves an AUC of 0.965. Searching the IRC safe EFPs, the guided search identifies the familiar graph ⟋ with $\kappa = 1, \beta = \frac{1}{2}$ and reaches an AUC of 0.973 ± 0.001, completely closing the gap.

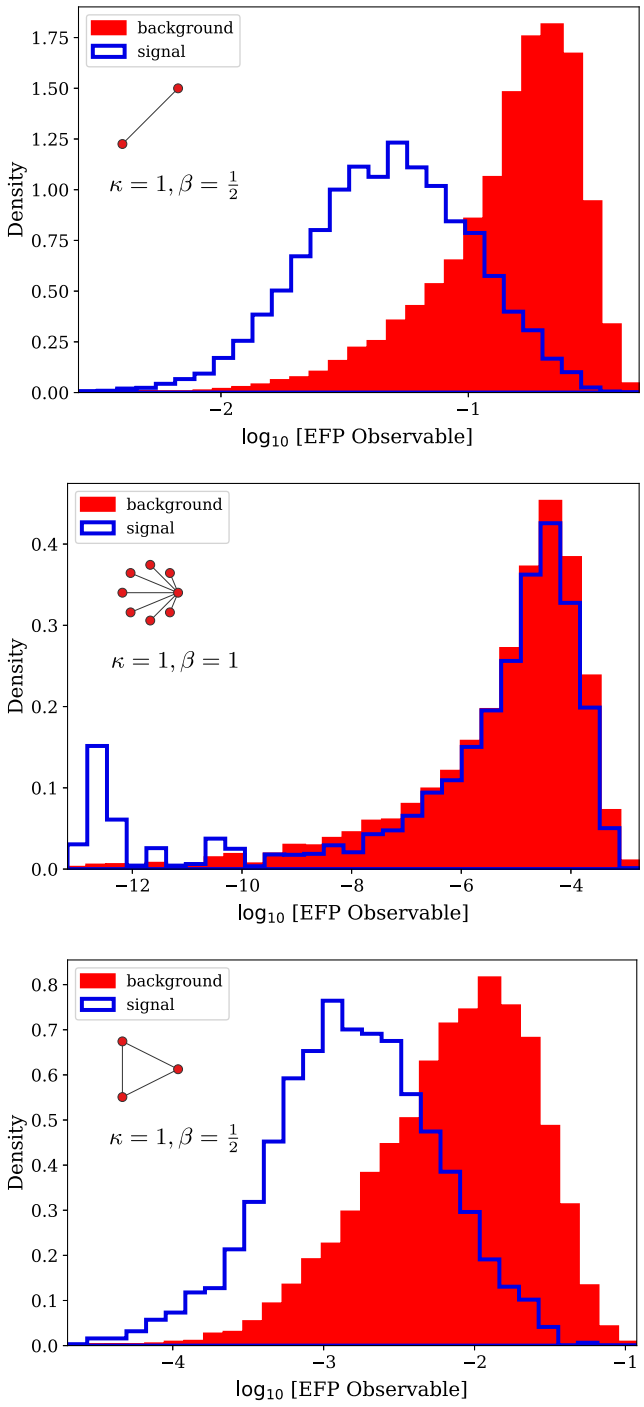

FIG. 6. $\log_{10}$ distributions of the selected IRC-safe EFPs as chosen by the black-box guided strategy, for signal electrons and background jets.

D. Broader scan

In this section, we present a scan of a larger set of EFPs, including values of κ which lead to IRC unsafe observables, $\kappa = [-1, 0, 1, 2]$.

Beginning from the seven standard HL features, the first pass selects a simple observable:

$$(\kappa=2) = \sum_{a=1}^{N} z_a^2$$

with no angular terms at all, but $\kappa = 2$. This is known in the jet substructure literature as p_{T}^{D} [41,42] and was originally developed to help distinguish between quark and gluon jets. When combined with the other seven HL features, this observable also reaches a performance of 0.970 ± 0.001. Further scans do not lead to statistically significant improvements in AUC.

If instead, we begin from the seven standard HL features and M_{jet}, we find , this time with $\kappa = 2$ as well as the simpler p_{T}^{D}. Distributions of these two IRC unsafe EFP observables for signal and background are shown in Fig. 7. Together with the seven HL and M_{jet}, these 10 observables reach a performance of 0.971 ± 0.001. Further scans do not lead to statistically significant improvements in AUC. However, beginning from the seven HL observables together with M_{jet} and p_{T}, the same EFP graphs are chosen by the guided search, reaching an AUC of 0.973 ± 0.001

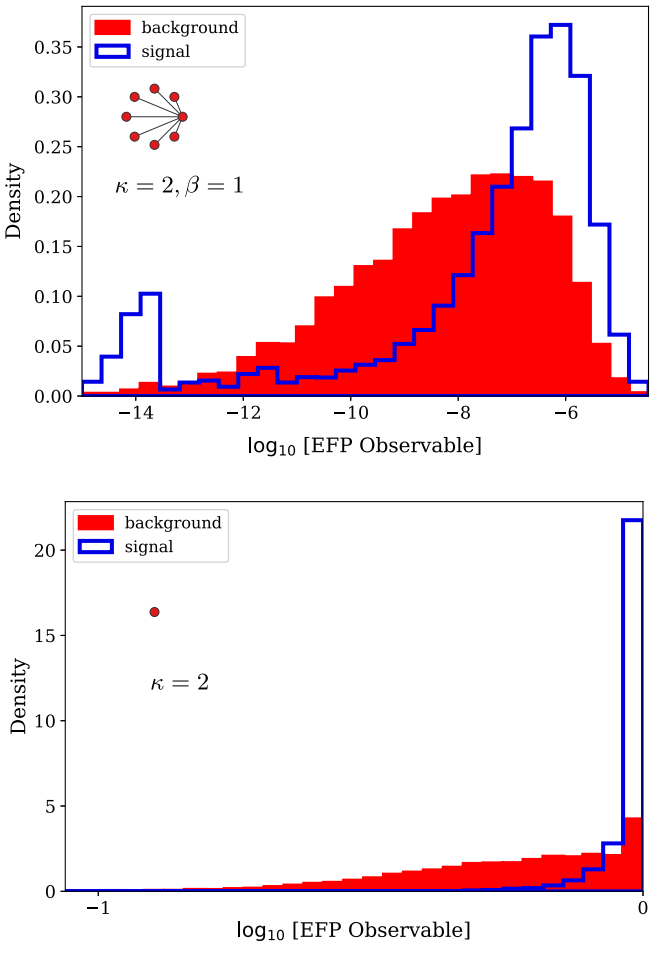

FIG. 7. $\log_{10}$ distributions of the selected EFPs as chosen by the black-box guided strategy, regardless of IRC safety, for signal electrons and background jets.

TABLE II. Summary of the performance of various networks considered. Uncertainty in the AUC value is ± 0.001, estimated using bootstrapping.

Base	Additions (κ, β)		(AUC)
7HL			0.945
7HL	$(1, \frac{1}{2})$		0.970
7HL	$(2, -)$		0.970
7HL + M_{jet}			0.956
7HL + M_{jet}	(1,1)	$(1, \frac{1}{2})$	0.971
7HL + M_{jet}	(2,1)	$(2, -)$	0.971
7HL + M_{jet} + p_{T}			0.965
7HL + M_{jet} + p_{T}	$(1, \frac{1}{2})$		0.973
7HL + M_{jet} + p_{T}	(2,1)	$(2, -)$	0.973
CNN			0.972
CNN + 7HL	$(1, \frac{1}{2})$		0.972
CNN + 7HL	$(2, -)$		0.973

TABLE III. Performance of selected networks, in terms of the AUC value as well as background rejection (R) at several choices of signal efficiency (ϵ).

Features	AUC	$R_{\epsilon=0.5}$	$R_{\epsilon=0.75}$	$R_{\epsilon=0.9}$
7HL	0.945	32.98	15.78	8.80
7HL + $(1, \frac{1}{2})$	0.970	88.63	34.73	15.07
CNN	0.973	94.07	36.89	15.93

See Table II for a summary of the additional observables needed to reach the performance of ≈0.97 in each case, and Table III for background rejection factors for several choices of signal efficiency.

VIII. DISCUSSION

Our deep neural networks indicate that low-level calorimeter data represented as images contains information useful for the task of electron identification that is not captured by the standard set of high-level features as implemented here.

A guided search [39] through the EFP space identified two EFP observables calculated on the ECal cells which mimic the CNN strategy and bridge the gap. Observables on the HCal information were not helpful to the classification task. The first,

$$(\kappa=1,\beta=\tfrac{1}{2}) = \sum_{a,b=1}^{N} z_a z_b \theta_{ab}^{\frac{1}{4}}$$

is closely related to the Les Houches angularity [40], and confirms our suspicion that the nontrivial structure of the background object provides a useful handle for classification. The second observable, p_{T}^{D} [41,42],

$$\bullet \quad (\kappa=2) = \sum_{a=1}^{N} z_a^2$$

is not IRC safe, and was originally developed to help distinguish between quark and gluon jets. It effectively counts the number of hard particles, which is sensitive to the amount of color charge, where electrons and jets are clearly distinct.

Both Les Houches angularity and p_{T}^{D} display power to separate electrons from the jet backgrounds, by exploiting the structure and nature of the jet energy deposits. The studies performed here use a simplified simulation of the detector, and notably lack an accurate description of the radiation of photons from electrons, which may result in an unrealistic pattern of energy deposition and secondary clusters. While the precise performance obtained here may depend at some level on the fidelity of the simulation used and the resulting limitations on the implementation of state-of-the-art high-level features, these results strongly suggest that these observables be directly studied in experimental contexts where more realistic simulation tools are available, or directly in data samples, using weakly supervised learning [43].

More broadly, the existence of a gap between the performance of state-of-the-art high-level features and CNN represents an opportunity to gather additional power in the battle to suppress lepton backgrounds. Rather than employing black-box CNNs directly, we have demonstrated the power of using them to identify the relevant observables from a large list of physically interpretable options. This allows the physicist to understand the nature of the information being used and to assess its systematic uncertainty.

Any boost in electron identification performance is extremely valuable to searches at the LHC, especially those with multiple leptons, where event-level efficiencies depend sensitively on object-level efficiencies.

All code and data used in this project is available at: https://github.com/TDHTTTT/EID, as well as through the UCI Machine Learning in Physics web portal at: http://mlphysics.ics.uci.edu/.

ACKNOWLEDGMENTS

The authors thank Jesse Thaler, Ian Moult, and Tim Tait for helpful discussions. We wish to acknowledge a hardware grant from NVIDIA. This material is based upon work supported by the National Science Foundation under Grant No. 1633631. The work of J. C. and P. B. is in part supported by Grants No. NSF 1839429 and No. NSF NRT 1633631 to P. B.. The work of J. N. H. is in part supported by Grants No. DE-SC0009920, No. DGE-1633631, and No. DGE-1839285. T. T. wants to thank Undergraduate Research Opportunities Program at UCI for Grant No. 02399s1. D. W. is funded by Department of Energy Office of Science Grant No. DE-SC0009920.

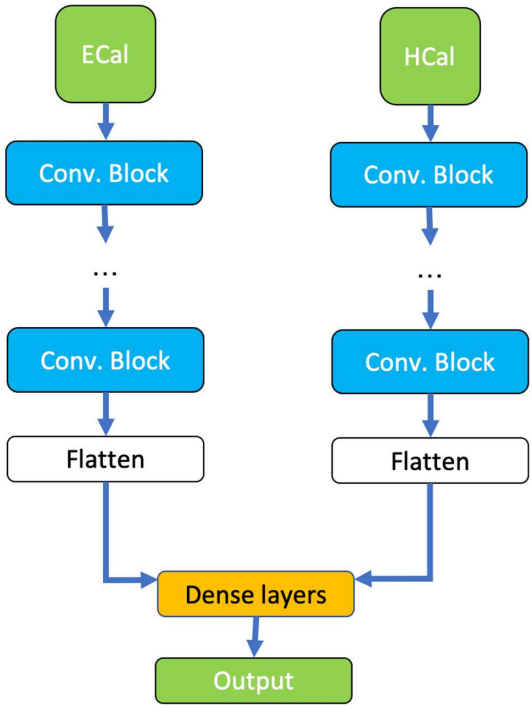

FIG. 8. Diagram of convolutional block appearing in network architecture, see Fig 9.

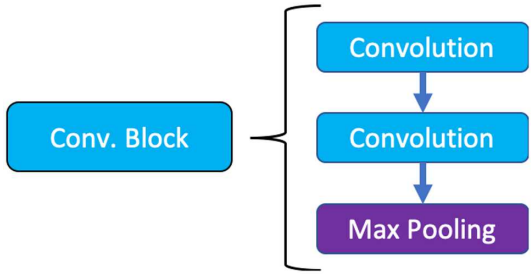

FIG. 9. Diagram of the architecture of the convolutional neural network.

APPENDIX: NEURAL NETWORK HYPERPARAMETERS AND ARCHITECTURE

Figures 8 and 9 show the architecture of the convolutional neural network.

TABLE IV. Hyperparameter ranges for Bayesian optimization of convolutional networks.

Parameter	Range
Num. of conv. blocks	[1, 4]
Num. of filters	[8, 128]
Num. of dense layers	[1, 3]
Num. of hidden units	[1, 200]
Learning rate	[0.0001, 0.01]
Dropout	[0.0, 0.5]

TABLE V. Hyperparameter ranges for Bayesian optimization of fully connected networks.

Parameter	Range
Num. of dense layers	[1, 8]
Num. of hidden units	[1, 200]
Learning rate	[0.0001, 0.01]
Dropout	[0.0, 0.5]

TABLE VI. Best hyperparameters found per model.

Features	Conv.	Filters	Dense	Hidden	LR	DP
ECal	3	117	2	160	0.0001	0.0
Hcal	2	27	2	84	0.01	0.5
Ecal + HCal	3	47	2	146	0.0001	0.0
HL	…	…	5	149	0.001	0.0019

[1] G. Aad *et al.* (ATLAS Collaboration), Measurement of $W^{\pm}$ and Z-boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV with the atlas detector, Phys. Lett. B **759**, 601 (2016).

[2] E. S. Almeida, A. Alves, N. Rosa-Agostinho, O. J. P. Eboli, and M. C. Gonzalez-Garcia, Electroweak sector under scrutiny: A combined analysis of LHC and electroweak precision data, Phys. Rev. D **99**, 033001 (2019).

[3] G. Aad *et al.* (ATLAS Collaboration), Search for supersymmetry in final states with jets, missing transverse momentum and one isolated lepton in $\sqrt{s} = 7$ TeV pp collisions using 1 fb^{-1} of ATLAS data, Phys. Rev. D **85**, 012006 (2012); , Erratum, Phys. Rev. D **87**, 099903 (2013).

[4] S. Chatrchyan *et al.* (CMS Collaboration), Search for supersymmetry in pp collisions at $\sqrt{s} = 8$ TeV in events with a single lepton, large jet multiplicity, and multiple b jets, Phys. Lett. B **733**, 328 (2014).

[5] I. Hoenig, G. Samach, and D. Tucker-Smith, Searching for dilepton resonances below the Z mass at the LHC, Phys. Rev. D **90**, 075016 (2014).

[6] V. Khachatryan *et al.* (CMS Collaboration), Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{}$ s = 8 TeV, J. Instrum. **10**, P06005 (2015).

[7] Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data, Technical Report No. ATLAS-CONF-2016-024, CERN, Geneva, 2016.

[8] M. Hushchyn and V. Chekalina (LHCb Collaboration), Particle-identification techniques and performance at LHCb in Run 2, Nucl. Instrum. Methods Phys. Res., Sect. A **936**, 568 (2019).

[9] P. Baldi, P. Sadowski, and D. Whiteson, Searching for exotic particles in high-energy physics with deep learning, Nat. Commun. **5**, 4308 (2014).

[10] 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).

[11] P. Baldi, *Deep Learning in Science: Theory, Algorithms, and Applications* (Cambridge University Press, Cambridge, England, to be published).

[12] R. Santos, M. Nguyen, J. Webster, S. Ryu, J. Adelman, S. Chekanov, and J. Zhou, Machine learning techniques in searches for $t\bar{t}h$ in the $h \to b\bar{b}$ decay channel, J. Instrum. **12**, P04014 (2016).

[13] A. Aurisano, A. Radovic, D. Rocco, A. Himmel, M. Messier, E. Niner, G. Pawloski, F. Psihas, A. Sousa, and P. Vahle, A convolutional neural network neutrino event classifier, J. Instrum. **11**, P09001 (2016).

[14] T. Cohen, M. Freytsis, and B. Ostdiek, (Machine) learning to do more with less, J. High Energy Phys. 02 (2018) 034.

[15] M. Andrews, M. Paulini, S. Gleyzer, and B. Poczos, End-to-end event classification of high-energy physics data, J. Phys. Conf. Ser. **1085**, 042022 (2018).

[16] J. Cogan, M. Kagan, E. Strauss, and A. Schwartzman, Jet-images: Computer vision inspired techniques for jet tagging, J. High Energy Phys. 02 (2015) 118.

[17] 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).

[18] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, Jet-images—deep learning edition, J. High Energy Phys. 07 (2016) 069.

[19] L. De Oliveira, B. Nachman, and M. Paganini, Electromagnetic showers beyond shower shapes, Nucl. Instrum. Methods Phys. Res., Sect. A **951**, 162879 (2020).

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

[21] 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).

[22] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lematre, A. Mertens, and M. Selvaggi (DELPHES 3 Collaboration), DELPHES 3, A modular framework for fast simulation of a generic collider experiment, J. High Energy Phys. 02 (2014) 057.

[23] S. Agostinelli *et al.* (GEANT4 Collaboration), GEANT4: A simulation toolkit, Nucl. Instrum. Methods Phys. Res., Sect. A **506**, 250 (2003).

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

[25] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching

to parton shower simulations, J. High Energy Phys. 07 (2014) 079.
[26] T. Sjostrand, S. Mrenna, and P. Z. Skands, PYTHIA 6.4 physics and manual, J. High Energy Phys. 05 (2006) 026.
[27] R. Brun and F. Rademakers, ROOT: An object oriented data analysis framework, Nucl. Instrum. Methods Phys. Res., Sect. A **389**, 81 (1997).
[28] G. Aad *et al.* (ATLAS Collaboration), The ATLAS experiment at the CERN large hadron collider, J. Instrum. **3**, P07007 (2008).
[29] P. Berta, M. Spousta, D. W. Miller, and R. Leitner, Particle-level pileup subtraction for jets and jet shapes, J. High Energy Phys. 06 (2014) 092.
[30] X. Glorot, A. Bordes, and Y. Bengio, in *Deep Sparse Rectifier Neural Networks* (PMLR, Fort Lauderdale, FL, USA, 2011), pp. 315–323.
[31] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, arXiv:1207.0580.
[32] P. Baldi and P. Sadowski, The dropout learning algorithm, Artif. Intell. **210C**, 78 (2014).
[33] F. Chollet *et al.*, Keras, https://keras.io, 2015.
[34] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems (2015), software available from tensorflow.org.
[35] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.
[36] X. Glorot and Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (JMLR Workshop and Conference Proceedings, Sardinia, Italy, 2010), pp. 249–256.
[37] L. Hertel, J. Collado, P. Sadowski, J. Ott, and P. Baldi, Sherpa: Robust hyperparameter optimization for machine learning, SoftwareX **12**, 100591 (2020).
[38] A. J. Larkoski, G. P. Salam, and J. Thaler, Energy correlation functions for jet substructure, J. High Energy Phys. 06 (2013) 108.
[39] T. Faucett, J. Thaler, and D. Whiteson, Mapping machine-learned physics into a human-readable space, Phys. Rev. D **103**, 036020 (2021).
[40] P. Gras, S. Höche, D. Kar, A. Larkoski, L. Lönnblad, S. Plätzer, A. Siódmok, P. Skands, G. Soyez, and J. Thaler, Systematics of quark/gluon tagging, J. High Energy Phys. 07 (2017) 091.
[41] F. Pandolfi, Search for the standard model Higgs boson in the $H \to ZZ \to l^+l^-q\bar{q}$ decay channel at CMS, Ph.D. thesis, Zurich, ETH, New York, 2012.
[42] S. Chatrchyan *et al.* (CMS Collaboration), Search for a Higgs boson in the decay channel H to ZZ^* to $q\bar{q}\,\ell^-\ell^+$ in pp collisions at $\sqrt{s} = 7$ TeV, J. High Energy Phys. 04 (2012) 036.
[43] L. M. Dery, B. Nachman, F. Rubbo, and A. Schwartzman, Weakly supervised classification in high energy physics, J. High Energy Phys. 05 (2017) 145.