EPJ Web of Conferences **251**, 04030 (2021) https://doi.org/10.1051/epjconf/202125104030
CHEP 2021

End-to-End Jet Classification of Boosted Top Quarks with CMS Open Data

Michael Andrews[1], *Bjorn* Burkle[2,*], *Shravan* Chaudhari[3], *Davide* DiCroce[4], *Sergei* Gleyzer[4], *Ulrich* Heintz[2], *Meenakshi* Narain[2], *Manfred* Paulini[1], and *Emanuele* Usai[2]

[1]Department of Physics, Carnegie Mellon University, Pittsburgh, USA
[2]Department of Physics, Brown University, Providence, USA
[3]Department of Electrical and Electronics Engineering, BITS Pilani, Goa, India
[4]Department of Physics and Astronomy, University of Alabama, Tuscaloosa, USA

Abstract. We describe a novel application of the end-to-end deep learning technique to the task of discriminating top quark-initiated jets from those originating from the hadronization of a light quark or a gluon. The end-to-end deep learning technique combines deep learning algorithms and low-level detector representation of the high-energy collision event. In this study, we use low-level detector information from the simulated CMS Open Data samples to construct the top jet classifiers. To optimize classifier performance we progressively add low-level information from the CMS tracking detector, including pixel detector reconstructed hits and impact parameters, and demonstrate the value of additional tracking information even when no new spatial structures are added. Relying only on calorimeter energy deposits and reconstructed pixel detector hits, the end-to-end classifier achieves a ROC-AUC score of 0.975±0.002 for the task of classifying boosted top quark jets. After adding derived track quantities, the classifier ROC-AUC score increases to 0.9824±0.0013, serving as the first performance benchmark for these CMS Open Data samples.

1 Introduction

The Large Hadron Collider (LHC) is a prolific top quark factory: since the beginning of data-taking in 2010, over 10^8 top quarks have been produced. The measurement of the top quark's properties and production rates at the LHC remains one of the main research priorities at experiments like the Compact Muon Solenoid (CMS) at the LHC. Moreover, investigating the resonant production of top quarks offers potential hints of the presence of new physics that may lie beyond the standard model (BSM).

Top quarks are unique in that they decay before they have time to hadronize, always decaying to a bottom quark and a W-boson. During the top decay chain, the W-boson will decay hadronically to quarks 66.5% or leptonically to a lepton and neutrino pair 33.5% of the time [1]. At hadron colliders like the LHC, the low production cross section of prompt electrons and muons can be exploited to boost tagging efficiency when identifying top quarks with a leptonically decaying W-boson in its decay chain. However, hadronic decays of top quarks can be much harder to identify, since the primary features used to identify them are

*e-mail: bjorn_burkle@brown.edu

the topology of its decay products and the track features of the bottom quark decay products. In particular, at high transverse momenta, the hadronic decay of highly a Lorentz-boosted top quark can lead to a single merged cluster of particles in the detector, hereby referred to as jets, offering a unique and challenging view into the study of the top quark's properties. Because of this, discriminating boosted top quark-jets from light flavour- or gluon-jets has become an important challenge for the LHC experiments, and a popular benchmark for data analysis techniques involving machine learning (ML) algorithms in high-energy physics (HEP).

Most jet identification techniques rely on inputs provided by the Particle Flow (PF) algorithm used to convert detector level information to physics objects [2]. The Particle Flow algorithm has many advantages due to its ability to greatly reduce the size and complexity of particle physics data while providing a physically intuitive and easy to use representation in physics analyses. Many of the modern machine learning approaches to jet discrimination are based on PF-based inputs [3–10]. However, there is some invariable loss of information from reducing the data set complexity. Despite the very high reconstruction efficiency of PF algorithms, some physics objects may fail to be reconstructed, are reconstructed imperfectly, or exist as fakes [11]. For that reason it is advantageous to consider end-to-end reconstruction that allows a direct application of machine learning algorithms to low-level data representation in the detector.

In this work, we extend the end-to-end deep learning approach for particle and event classification [12]. Specifically, we extend the use of end-to-end jet images introduced for quark- vs. gluon-jet discrimination [13] to the task of boosted top quark- vs. light quark- or gluon-jet discrimination. In previous work [13] we found that the track information was the leading contributor to the classifier's performance. Due to this insight and the importance of identifying displaced tracks associated with bottom quark decays, this new work introduces a number of key features from the CMS tracking detectors to exploit the full topology of hadronically decaying top quarks.

2 Open Data Simulated Samples

The end-to-end deep learning technique relies on high-fidelity simulated detector data, which in this work comes from the simulated Monte Carlo in the CMS Open Data Portal [14]. We use a sample of SM top-antitop ($t\bar{t}$) pair production where the W-boson from the top quark decay is required to decay to quarks as a source of boosted top quarks. Light flavour- and gluon-jets were obtained from three samples of QCD dijet production in different ranges of the hard-scatter transverse momentum. The full datasets used for this study can be found in [15–18]. For all samples, the detector response is simulated using Geant4 with the full CMS geometry and is processed through the CMS PF reconstruction algorithm using CMSSW release 5_3_32 [19]. An average of ten additional background collisions or pileup (PU) interactions are added to the simulated hard-scatter event, which are sampled from a realistic distribution of simulated minimum bias events. For this study, we additionally use a custom CMS data format which includes the low-level tracker detector information, specifically, the reconstructed clusters from the pixel and silicon strip detectors [20]. From the tracker clusters, we then do a parametric estimate of the position of the hit on the sensor surface.

For jet selection, we take reconstructed jets clustered using the anti-k$_t$ algorithm [21] with a radius parameter R of 0.8 (AK8 jets) and require $p_{\mathrm{T}} > 400\,\mathrm{GeV}$ and $|\eta| < 1.37$ for our event selection. Here, η is the pseudorapidity and equates to the polar angle of the CMS detector according to $\eta = -\ln(\tan\frac{\theta}{2})$. This η cut is to ensure that the jet image does not extend beyond the $|\eta| < 2.4$ acceptance limit of the current CMS tracker. Additionally, for the top jets we require the generator-level top quark, its bottom quark and W-boson daughters, and W-boson daughters to be within an angular separation of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.8$ from

the reconstructed AK8 jet axis, where ϕ is the azimuthal angle of the CMS detector. In order to avoid biases caused by the different p_T distributions of the two jet samples, we pseudo-randomly drop jets from the three QCD samples such that the total number of jets and p_T distribution of the $t\bar{t}$ sample is reproduced. The total number of jets used in the training, validation, and testing of our networks can be found in Table 1.

3 CMS Detector & Images

CMS is a multi-purpose detector composed of several cylindrical subdetector layers, with both barrel and endcap sections, encasing a primary interaction point. It features a large B = 3.8 T solenoid magnet to bend the trajectories of charged particles that aid in p_T measurement. At the innermost layers, close to the beamline, there is a silicon tracker used to reconstruct the trajectory of charged particles and find their interaction vertices. The tracker can be divided in two parts the silicon pixel detector and silicon strip detector. The first silicon pixel detector is the inner most part and composed of three layers in the barrel region (BPIX) and three disks in the endcap region (FPIX). Each layer is composed of pixel sensors that provide a very precise position of the passage of a charged particle. The pixel detector provides crucial information for vertexing and track seeding. The outer part of the tracking system is composed of several layers of silicon strip. These provide a precise position in the ϕ coordinate, but not in the η coordinate. This is followed by the electromagnetic calorimeter (ECAL), made of lead-tungstate crystals, to measure the energy of electromagnetically interacting particles, then the hardonic calorimeter (HCAL), made of brass towers, to measure the energy of hadrons. These are surrounded by the solenoid magnet which is finally encased by the muon chambers to detect the passage of muons [22].

We construct the jet images using low-level detector information where each subdetector is projected onto one or multiple image layers in a grid of 125 x 125 pixels with the image centered around the most energetic HCAL deposit of the jet. Each pixel corresponds to the span of an ECAL barrel crystal which covers a 0.0174×0.0174 in the $\eta - \phi$ plane, giving our images an effective ΔR of 2.175. For the ECAL and HCAL images, each crystal or tower is directly mapped to one or more image pixels containing the energy deposited in that crystal or tower, as described in [13]. Reconstructed particle tracks are weighted by their reconstructed p_T and their location is projected to an ECAL crystal. In order to better overlap with the calorimeter images, the $\eta - \phi$ position of the tracks are determined by assuming the track originated from the primary vertex, the location of the collision with the highest $\sum p_T^2$, before being propagated to the ECAL surface.

To improve the identification of tracks coming from the hadronization of b quarks, we added additional layers motivated by the long flight distance of b hadrons producing reconstructed tracks that do not converge to the primary vertex. To make the network aware of this information, we tried two approaches: a) additional two layers corresponding to the re-

Table 1. Number of jets used for training, validation, and testing in the top quark and non-top quark jet categories. Jets in the validation set were used during training to ensure that the network was not over-training, and jets in the testing set were used after training to quantify network performance. Numbers are reported after the p_T-resampling procedure.

Category	Top quark jets	QCD jets	Total Jets
Train	1280830	1279170	2560000
Validation	47859	48141	96000
Test	319819	320181	640000

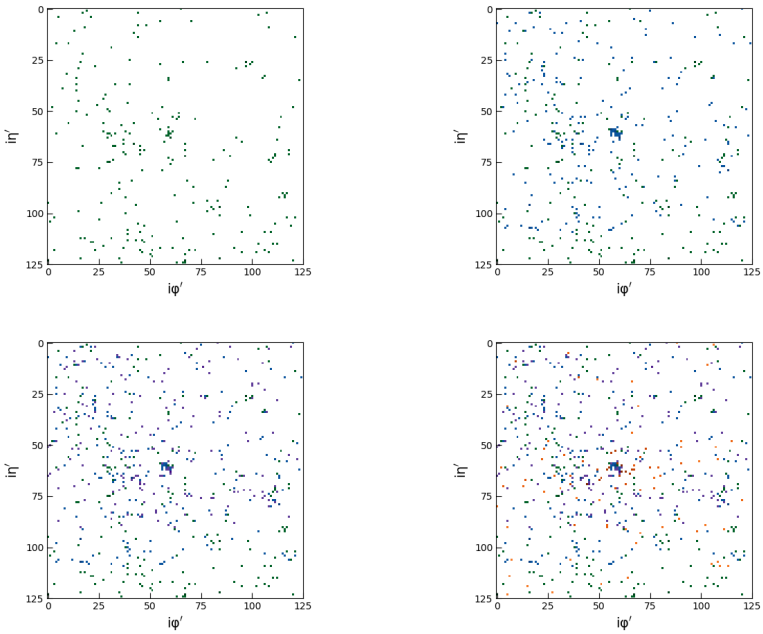

Figure 1. Jet images of the 1st layer of the BPIX (top left), 1st & 2nd layers of the BPIX (top right), 1st, 2nd, & 3rd layers of the BPIX (bottom left), and all layers of the BPIX with the reconstructed track p_{T} (bottom right) of a simulated 703 GeV p_{T} merged top jet. Images are created at the nominal ECAL resolution.

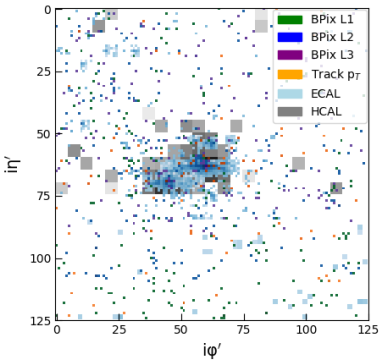

Figure 2. Composite image of a simulated boosted top quark jet at ECAL resolution

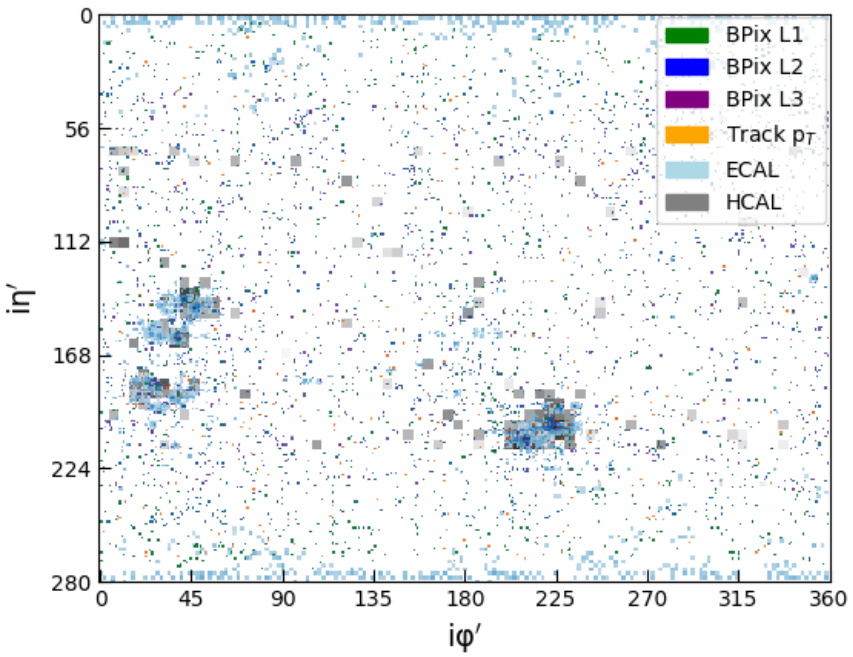

Figure 3. Full detector composite image of a $t\bar{t}$ simulated event. Image is constructed at the ECAL resolution.

constructed tracks weighted by their transverse ($d0$) and longitudinal (dZ) impact parameter significance and b) additional layers from the BPix detector that contain low-level representation of tracker RecHits. The impact parameter (IP) is defined as the distance vectors of minimum approach between the track helix and the primary vertex. To obtain the IP significance, the $d0$ and dZ values are divided by their respective uncertainties. These quantities are computed without using approximations making them accurate even for tracks relatively far from the primary vertex. Any $d0$ (dZ) values larger than 10 cm (20 cm) are suppressed to zero to prevent training degradation caused by the inclusion of tracks with superfluously large IP. Such tracks are expected to originate from photon conversions in the tracker or from poor track reconstruction, and these cuts are not expected to negatively impact network performance. Finally, each layer is independently normalized such that the value of the average cell, ignoring empty cells, is approximately unity to facilitate training convergence.

In an effort to extract as much information as possible from the tracking subdetector, we include the low-level detector information from this system: the tracking hits traditionally used for track reconstruction. There are multiple steps in the conversion from charge clusters produced via charged particles passing through the tracker to fully reconstructed tracks. In this study, we consider the reconstructed hit (RecHit) information from the three layers of the BPIX, but not from the FPIX or the silicon strip detector. The majority of tracks will pass through the BPIX, which allows us to simplify the geometry of the RecHit layers by omitting the FPIX hits while minimizing the amount of omitted information. RecHits are obtained by first clustering nearby pixels of a given sensor which pass an adjustable charge threshold. A

straight line fits the pixel cluster to center of the beam, and it's angle with the sensor surface is used to compute a hit location which is corrected for the Lorentz drift the charges experience before being read off the sensor. Given the hit location on the sensor and location of the sensor in the detector, the location of the RecHit in η and ϕ is obtained. The RecHits serve as a good intermediary between raw detector outputs and reconstructed track quantities, serving as a map between a module based coordinate system and the detector coordinates.

For this study, one additional step is performed on the RecHits prior to producing the image layers. The η and ϕ position of the RecHit is re-calculated with respect to the primary vertex of the collision rather than the geometric center of the detector. This is done so that the η and ϕ of the RecHits better match the η and ϕ of their corresponding tracks when reaching the ECAL, which would otherwise deviate due to the pixel detectors closeness to the beamline. After these computations are performed, image layers are produced where each pixel intensity is set to one if the image pixel contains a RecHit and zero otherwise. We generate three different image layers, one for each of the three concentric layers of the BPIX. Figure 1 shows the successive addition of the three pixel layers and the track p_{T}. The cluster of RecHits in the center of the images corresponds to a cluster of jet particles, while many of the outer RecHits originate from pile-up or detector noise. Figures 2 and 3 shows a end-to-end image featuring all the image layers considered in this work for a single jet and the full detector. The only layers that cannot be seen are the track $d0$ and dZ values because the perfectly overlap with the track p_{T} layer. A full list of all image layers along with their description can be found in Table 2.

4 Network, Training and Jet Identification Results

The network architecture and hyperparameters used in this work closely follow what was previously used in [12, 13], making use of a ResNet-15 CNN [23] trained with the ADAM optimizer [24]. The full network infrastructure is outlined in Table 3. The initial learning rate is 5×10^{-4} and is explicitly reduced by half every 10 epochs. The network was trained on a set of 2.56M jets, and we found that training for 20 epochs was sufficient for convergence. However, for our final network evaluations we trained for an additional 20 epochs. The network was developed and trained using the TensorFlow library v1.14 [25].

Table 4 shows the area under the receiver operator curve (AUC) for the different combinations of track and calorimeter layers at ECAL granularity. The network was evaluated on a separate sample of 200k jets giving an AUC statistical uncertainty of 0.002.

Our previous end-to-end deep learning results showed that the Track p_{T} layer gave the best single layer performance for jet discrimination [13]. Therefore, we choose track p_{T} layer performance as a baseline for our models' performance. We observe that the largest single-subdetector performance increase comes with the inclusion of the $d0$ and dZ track information, leading to an AUC score improvement of 0.014–0.017. Comparing rows 2 and 3 in Table 4 shows that the combination of track p_{T}, $d0$, and dZ outperforms the nominal

Table 2. Summary descriptions of the image layers considered.

Layer	Description
Track p_{T}	track momentum, position at ECAL surface
Track $d0$ & dZ	track impact parameter projections, w.r.t PV, at ECAL surface
BPIX1–3	Barrel pixel hits, position, binary value, PV-shifted
ECAL	EM energy deposit, per crystal
HCAL	Had energy deposit, per tower

Table 3. Network infrastructure of the ResNet-15 CNN used in this study. All convolutional layers used same padding and are followed by a ReLU activation function. The network was built using TensorFlow version 1.14, extra parameters refers to the arguments used when constructing layers.

Block	Layer Type	Extra Parameters
Input Node	Conv 2D	kernel size 7, stride 2, 16 channels
	Global Pool 2D	2×2 pool size
Resblock 1	Conv 2D	kernel size 3, stride 1, 16 channels
	Conv 2D	filter size 3, stride 1, 16 channels
Resblock 2	Conv 2D	kernel size 3, stride 2, 32 channels
	Conv 2D	filter size 3, stride 1, 32 channels
Resblock 3	Conv 2D	kernel size 3, stride 1, 32 channels
	Conv 2D	filter size 3, stride 1, 32 channels
Output Node	Max Pooling 2D	global pool size
	Dense	size 32×2
	Activation	Sigmoid

Table 4. Performance of the classifier trained up to 20 epochs and evaluated on a set of 200000 jets. AUC and signal-efficiency at 1% misidentification rates are shown for different combinations of tracking and calorimeter layers; AUC scores have a statistical uncertainty of ±0.002.

Layer Combinations	AUC	Sig-eff at 1%
Track p_{T} (baseline)	0.955±0.002	39.6%
Track p_{T} + ECAL + HCAL (nominal)	0.967±0.002	56.9%
Track p_{T} + $d0$ + dZ	0.972±0.002	57.2%
Track p_{T} + $d0$ + dZ + ECAL + HCAL	0.981±0.002	64.4%

combination layers despite the fact that the p_{T} + $d0$ + dZ images are agnostic to neutral particles, since they do not produce charge clusters in the tracker. What we observe is in agreement with [13] where the tracks were observed as the most important feature for jet discrimination, as well as more traditional jet tagging approaches which require the presence of a b-tagged subjet tagged using IP variables [26, 27].

Table 5 shows the network performance when including the BPIX RecHits in the jet images. On their own, the BPIX RecHits give a worse performance than the track p_{T}. However, we observe multiple improvements in network performance after combining the BPIX RecHit images with other layers. When training the network on images composed of BPIX1–3, ECAL, and HCAL layers we find that it outperforms the nominal combination of layers, shown in the second row of Table 4, and improves the AUC score by 0.008. Comparing row 4 of Table 5 with row 3 of Table 4 shows that adding the BPIX RecHits to the track p_{T} + $d0$ + dZ images improves the AUC by 0.005. To study the effect of BPIX RecHit resolution on network performance, we additionally trained the network on images produced at sub-ECAL granularity. However, we found that the higher granularity produced no significant changes in network performance.

The bottom row of Table 5 reports the performance of our network when trained on all 8 image channels. The network was trained for 40 epochs and used the training, validation, and testing dataset sizes listed in Table 1. When evaluating the network, we find an AUC score of 0.9824±0.0013 and a signal efficiency of 66.41% at 1% misidentification.

Table 5. AUC and signal efficiency at 1% misidentification rates of the classifier after including BPIX layers. When training on a subset of image layers, the network was trained for 20 epochs and evaluated on 200k jets. Training on the full images was performed for 40 epochs and evaluated on 640k jets.

Layer Combinations	**AUC**	**Sig-eff at 1%**
BPIX1–3	0.947±0.002	38.2%
BPIX1–3 + Track p_T	0.965±0.002	45.6%
BPIX1–3 + ECAL + HCAL (no reconstruction)	0.975±0.002	56.9%
BPIX1–3 + Track p_T + $d0$ + dZ	0.977±0.002	64.4%
BPIX1–3 + Track p_T + $d0$ + dZ + ECAL + HCAL	0.9824±0.0013	66.41%

5 Interpretation and Discussion

An in depth look at the networks' performance when trained on different image layer combinations provides an insight into the features that the network is learning. We first note that the strongest single subdetector performance comes from the reconstructed tracks weighted by their p_T and IP variables. This is in agreement with the expectation based on the current understanding of high momentum top jets. We expect a large number of high p_T tracks, due to the jet containing three merged subjets, and a small subset of tracks having large IP values, attributed to a decaying B-meson. What is particularly interesting is that the network is able to successfully extract this IP information from the addition of the $d0$ and dZ layers to the track p_T image layer. By design, these track-only images are composed of a set of sparse layers with the same distribution of activated pixels. Intuitively, extracting information from such images using 2D convolutions becomes much more difficult than the traditional computer vision tasks. However, in this difficult to parse regime, our network achieves an AUC of 0.972±0.002, outperforming the denser jet images used for our nominal layer combination.

The second insight comes from the performance of the BPIX RecHits. As mentioned in Section 4, the BPIX RecHits do not show strong standalone single-layer performance. However, this is to be expected for multiple reasons. The pixel detector has an η and ϕ resolution of 10 μm, giving the inner most layers a 1D spatial resolution that is almost eight times finer than the ECAL [22]; the ECAL resolution is too coarse to derive vertex information from pixel hits. Furthermore, we only considered the barrel region of the pixel detector, and do not include any RecHits from the forward region of the pixel detector. Any jets that border the η acceptance of this study will only have RecHit information for a portion of the jet image. Finally, our network is agnostic to each layer's distance from the beamline, giving the network incomplete information about the RecHits global positioning. For example, the RecHits will drift in ϕ as the charged particle bends in the CMS detector's magnetic field. But unless more layers are added to the image, the network does not have enough information to know the order of each hit nor the direction in ϕ the particle is moving. But despite the shortcomings of our current RecHit implementation, we find remarkable results. With the exception of the final layer combination, where BPIX RecHits are added to images composed of track p_T + $d0$ + dZ + ECAL + HCAL information, we note that adding the RecHits gives a significant increase in network performance. The most notable are cases where BPIX RecHits are added on top of the tracking variables (1), and the case where BPIX RecHits are used in lieu of the derived tracking information (2).

In the case of (1) we see that the network is able to use the BPIX RecHits to derive new jet features which were not present in the derived track quantities alone. One possible feature is the track charge, where motion through ϕ can be combined with the final location of the track to determine its direction of curvature, and thus the charge, of the track. However, more abstract features can also exist in these images. In the case of (2), the network does not use

any reconstructed variables for its inputs. We see that despite the lack of derived variables, the network outperforms the track $p_T + d0 + dZ$ images, and only performs marginally worse than the final performance on the full images. The overall success of our network's ability to learn from BPIX RecHits paves the foundation for future studies of an end-to-end top tagger where no derived variables are used. In addition to including the forward region of the pixel detector, future work can include RecHits from the silicon strip detector, which is used for track seeding and track momentum measurement. We also look to explore new types of architectures, such as graph neural networks [28], that can exploit the full spatial resolution of the CMS tracker and the 3D correlation of its layers to complement existing architecture in other layers.

6 Conclusions

In this work we have extended the end-to-end deep learning technique to top quark jet classification. To enhance the performance of the classifier we added additional layers containing information about track parameters and pixel detector reconstructed hits, marking the first top-tagging algorithm which uses tracking RecHits as input variables. The model was trained using CMS Open Data datasets containing low-level tracking information [15–18].

The end-to-end classifier trained on all input features achieves the performance of AUC of 0.9824±0.0013. We find that the addition of $d0$ and dZ variables gives the largest boost to network performance when compared to subdetector information used in previous end-to-end jet discrimination studies [13]. At ECAL image granularity, we observe that the BPIX RecHits do not provide the network with information that is not already present in the combination of track pT, $d0$, dZ, ECAL, and HCAL layers. However, we find that it still improves subgroups of these layers, and the network achieves an AUC score of 0.975±0.002 when training on images void of derived variables. These findings lay the ground work for future studies which look to incorporate RecHits from the full CMS tracker, higher-resolution training, and to explore new deep learning architectures that can fully exploit the tracker granularity. Furthermore, we believe that the improvements in classifier performance observed after the inclusion of BPIX RecHits signals that more jet tagging algorithms should incorporate these features into their algorithms.

7 Acknowledgments

We would like to thank the CMS Collaboration and the CERN Open Data group for releasing their simulated data under an open access policy. We strongly support initiatives to provide such high-quality simulated datasets that can encourage the development of novel but also realistic algorithms, especially in the area of machine learning. We believe their continued availability will be of great benefit to the high energy physics community in the long run. We would also like to thank Google for access to the Google Cloud which helped us speed up the training of the deep learning models used in this study.

References

[1] M. Tanabashi et al. (Particle Data Group), Phys. Rev. D **98**, 030001 (2018)
[2] CMS Collaboration, JINST **12**, P10003. 82 p (2017)
[3] CMS Collaboration, CMS-DP-2017-049 (2017)
[4] A. Butter et al., SciPost Phys. **7**, 014 (2019), 1902.09914
[5] G. Kasieczka, T. Plehn, M. Russell, T. Schell, JHEP **05**, 006 (2017), 1701.08784

[6] P.T. Komiske, E.M. Metodiev, J. Thaler, Journal of High Energy Physics **2019** (2019)
[7] A. Butter, G. Kasieczka, T. Plehn, M. Russell, SciPost Phys. **5**, 028 (2018), 1707.08966
[8] CMS Collaboration, CMS-DP-2017-027 (2017)
[9] M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeloos, D.K. Abhayasinghe, S.H. Abidi, O.S. AbouZeid, N.L. Abraham, et al. (ATLAS), The European Physical Journal C **79** (2019)
[10] A.M. Sirunyan et al. (CMS), JINST **15**, P06005 (2020), 2004.08262
[11] Tech. rep., Geneva (2006), https://cds.cern.ch/record/922757
[12] M. Andrews, M. Paulini, S. Gleyzer, B. Poczos (2018), 1807.11916
[13] M. Andrews, J. Alison, S. An, B. Burkle, S. Gleyzer, M. Narain, M. Paulini, B. Poczos, E. Usai, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **977**, 164304 (2020)
[14] CMS Collaboration, *CMS data preservation, re-use and open access policy* (2014), http://opendata.cern.ch/record/411
[15] CMS Collaboration (2019), http://opendata.cern.ch/record/12200
[16] CMS Collaboration (2019), http://opendata.cern.ch/record/12201
[17] CMS Collaboration (2019), http://opendata.cern.ch/record/12202
[18] CMS Collaboration (2019), http://opendata.cern.ch/record/12203
[19] *Cms software version 5_3_32 (cmssw_5_3_32)* (2016), http://opendata.cern.ch/record/221
[20] E. Usai, M. Andrews, B. Burkle, S. Gleyzer, M. Narain, CERN Open Data Portal (2019)
[21] M. Cacciari, G.P. Salam, G. Soyez, Journal of High Energy Physics **2008**, 063–063 (2008)
[22] Tech. rep., Geneva (2006), https://cds.cern.ch/record/922757
[23] K. He, X. Zhang, S. Ren, J. Sun (2015), 1512.03385
[24] D.P. Kingma, J. Ba (2014), 1412.6980
[25] M.A. et al., *TensorFlow: Large-scale machine learning on heterogeneous systems* (2015), software available from tensorflow.org, http://tensorflow.org/
[26] C. Weiser, Tech. Rep. CMS-NOTE-2006-014, CERN, Geneva (2006), http://cds.cern.ch/record/927399
[27] CMS Collaboration, CMS-PAS-JME-13-007 (2014)
[28] J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, M. Sun (2018), 1812.08434