



ATLAS PUB Note

ATL-PHYS-PUB-2017-017

17th July 2017



Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector

The ATLAS Collaboration

Distinguishing quark-initiated from gluon-initiated jets is useful for many measurements and searches at the LHC. This note presents a jet tagger for distinguishing quark-initiated from gluon-initiated jets, which uses the full radiation pattern inside a jet processed as an image in a deep neural network classifier. The study is conducted using simulated dijet events in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector. Across a wide range of quark jet identification efficiencies, the neural network tagger achieves a gluon jet rejection that is comparable to or better than the performance of the jet width and track multiplicity observables conventionally used for quark-versus-gluon jet tagging.

1 Introduction

Differentiating quark-initiated (quark) jets from gluon-initiated (gluon) jets has broad applicability to searches and measurements at the LHC. Identifying the nature of a jet through its internal structure has a long history, originating with the discovery of the gluon at PETRA [1–4]. Recent interest has resulted from an enhanced theoretical [5–7], phenomenological [8, 9], and experimental [10–14] understanding of quark-versus-gluon jet tagging as well as the development of powerful machine learning techniques that can utilize the entire jet internal radiation pattern [15, 16]. The key difference between quark and gluon jets is that the former carry only one quantum chromodynamic (QCD) color while the latter have both a color and anti-color. More precisely, the Altarelli-Parisi splitting functions [17] contain a factor of $C_A = 3$ for gluon radiation from a gluon and a factor of $C_F = 4/3$ for gluon radiation from a quark. Due to this difference, gluon jets tend to have more constituents and a broader radiation pattern than quark jets. All experimental studies so far have focused on combining a small number of key jet observables that capture these expected differences. The purpose of this note is to present a first full detector simulation study of quark versus gluon jet tagging using the entire radiation pattern inside a jet. The approach is based on state-of-the-art image classification techniques and is benchmarked against classical quark versus gluon jet tagging schemes. A complete comparison with a (physically-motivated) dimensionally reduced set of inputs is beyond the current scope.

This note is organized as follows. Section 2 introduces the jet constituent reconstruction with the ATLAS detector and how they are used to form an image of the radiation pattern inside a jet. Deep convolutional neural networks are briefly described in Sec. 3 and then their application to quark versus gluon jet tagging is documented in detail. The note concludes in Section 4.

2 Quark and gluon jet images

The ATLAS experiment [18] features a multi-purpose particle detector with nearly 4π coverage in solid angle.¹ Charged particle trajectories are reconstructed as tracks within a 2 T axial magnetic field inside the inner detector (ID) consisting of silicon pixels, silicon micro-strips, and a transition radiation tracking detector. Tracks are reconstructed from fits to hits in all three ID subsystems covering the kinematic range $\phi \in [0, 2\pi)$, $|\eta| < 2.5$, and $p_T > 400$ MeV. Tight quality criteria [19] on the track properties are used to mitigate the impact of multiple overlapping proton–proton collision (pile-up) as well as to reject spurious (fake) tracks that result from multiple charged particles or noise. Tracks used for this study are required to have $p_T > 0.5$ GeV and to originate from the hard-scatter primary vertex. Tracks are assigned to primary vertices based on the track-to-vertex matching resulting from the vertex reconstruction. Tracks not included in vertex reconstruction are assigned to the nearest vertex based on the distance $|\Delta z \times \sin \theta|$, up to a maximum distance of 3 mm.

Surrounding the ID are electromagnetic (EM) and hadronic calorimeters. Energy deposits in the calorimeters are the inputs to jet reconstruction and classification algorithms. Two types of calorimeter inputs are considered as inputs for constructing jet images.

¹ ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z -axis along the beam pipe. The x -axis points from the IP to the centre of the LHC ring, and the y -axis points upward. Polar coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the beam pipe. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln \tan(\theta/2)$. Transverse momentum and energy are defined in the x – y plane as $p_T = p \cdot \sin(\theta)$ and $E_T = E \cdot \sin(\theta)$.

The first energy deposit organization scheme makes use of topological calorimeter-cell clusters (topo-clusters) [20]. The algorithm uses as seeds calorimeter cells with energy significance² $|E_{\text{cell}}|/\sigma_{\text{noise}} > 4$, iteratively combines all neighbouring cells with $|E_{\text{cell}}|/\sigma_{\text{noise}} > 2$ and finally adds neighbouring cells without any significance requirement. Topo-clusters at the EM scale are used as input for jet reconstruction with the anti- k_t jet algorithm [22] with distance parameter $R = 0.4$, as provided by FastJet [23]. In order to remain fully inside the tracker acceptance, jets are required to have $|\eta| < 2.1$.

An alternative organization scheme, calorimeter towers, is used to associate calorimeter information with the jets reconstructed from topo-clusters. Calorimeter towers are fixed-size objects ($\Delta\eta \times \Delta\phi = 0.1 \times 0.1$) [24] that ensure a uniform segmentation of the calorimeter information. Instead of building clusters, the cells are projected onto a fixed grid in η and ϕ corresponding to 6400 towers for the full calorimeter coverage $|\eta| < 5$. Calorimeter cells which completely fit within a tower contribute their total energy to the single tower. Other cells extending beyond the tower boundary contribute to multiple towers, depending on the overlap fraction of the cell area with the towers. In the following, towers are matched geometrically to jets built from topo-clusters, by considering all the towers within the jet radius $R = 0.4$ ³.

In this note, we also consider jet images built from the tracks associated with the jet. Tracks are assigned to jets by adding them to the jet clustering process with infinitesimal p_T , a procedure known as ghost-association [25].

Truth-particle jets in Monte Carlo (MC) simulation are built from generated stable particles with a mean lifetime $\tau > 30$ ps, excluding muons and neutrinos. As with the detector-level jets, truth-particle jets are clustered with the anti- k_t , $R = 0.4$ algorithm. The p_T of the truth-particle jet is used to select the jets. Two p_T ranges are considered for this study: $150 \text{ GeV} < p_T < 200 \text{ GeV}$ and $400 \text{ GeV} < p_T < 500 \text{ GeV}$. As quarks and gluons carry color charge and jets are color neutral, there is some ambiguity in the labeling of jets in simulation as quark or gluon. In this note, jets in simulation are labeled based on the highest-energy parton emerging from the hard-scatter collision within the jet catchment area [25], just as was used and studied in previous studies [12]. Only jets labeled as gluon or light quark (i.e. excluding bottom and top quark) are considered.

The jets considered in this note are from generic dijet events generated with PYTHIA 8 [26, 27] using the A14 tune [28], the NNPDF2.3 [29] PDF set, and processed with a full simulation of the ATLAS detector [30, 31]. Additional samples generated with SHERPA 2.1.1 (CT10 PDF [32]) and HERWIG++ 2.7.1 (UE-EE5 tune [33] and CTEQ6L1 PDF [34]) are used to quantify the model dependence. Both PYTHIA 8 and HERWIG++ is interfaced with EvtGen v1.2.0 [35] for heavy flavor decays.

There is an extensive literature on state-of-the-art machine learning classification, regression, and generation of images. In line with this literature, the radiation pattern inside jets can be spatially discretized into pixels to form jet images [36]. Early classification studies showed promising results when adding non-linearities beyond the Fisher linear discriminant (FLD) [36, 37]. Since the first application of modern (deep) neural networks to jet images [38], there have been many studies with state-of-the-art techniques for classification [15, 39–42] on jet images.

This note documents the first study of jet images with full detector simulation, using quark versus gluon tagging as a benchmark. As a first step in constructing a jet image, all of the constituents inside a jet are rotated and Lorentz boosted (all treated as massless; equivalent to translating in η) so that $\phi_{\text{jet}} = \eta_{\text{jet}} = 0$.

² The cell noise σ_{noise} is the sum in quadrature of the readout electronic noise and the cell noise due to pile-up, estimated in simulation [20, 21].

³ The position of jets and towers are defined by the jet and tower axis, respectively.

Then, a fixed grid of size 16×16 in η and ϕ with pixel sizes 0.05×0.05 is centered on the origin. The intensity of each pixel is the total p_T within the pixel, using a particular set of inputs (calorimeter-cell clusters, towers, tracks, or truth particles). Each image is then normalized so that $\sum_i I_i = 1$, where I_i is the intensity of the i^{th} pixel. Normalization is known to remove useful discrimination information, but studies suggest the impact is small and it is useful for training. For an extensive description of the impact of image preprocessing on the information content of a jet image, see Ref. [38].

One representative gluon jet image is shown in Figure 1 and the average quark and gluon jet images and image differences are shown in Figure 2 and Figure 3. Jet images are significantly different than natural images - they are sparse and lack distinct features with well-defined edges. Generic quark and gluon jets have mostly one hard core. As expected, the radiation pattern around the core is broader for gluon jets relative to quark jets. The slightly reduced central activity in track images shown in Figure 2 is compatible with the lower track reconstruction efficiency at the core of the jet [43]. Topo-cluster images are found to be more collimated than tower images (Figure 3) as a result of the built-in noise suppression mechanism that removes some of the soft large-angle radiation.

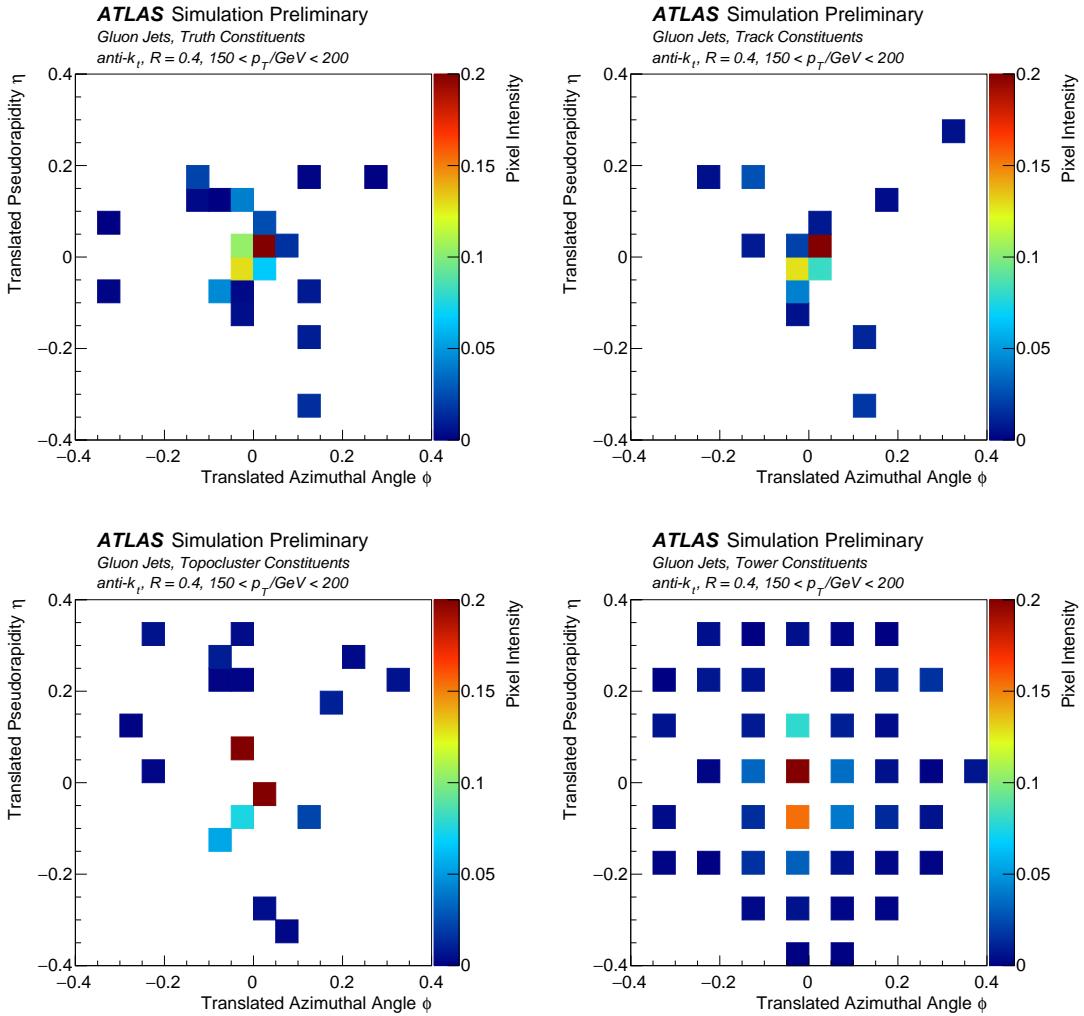


Figure 1: The stable particles (top left), track (top right), topocluster (bottom left), and tower (bottom right) images for an example gluon jet image. The tower image has gaps between hit pixels because the 0.1×0.1 towers are projected onto a 0.05×0.05 jet image.

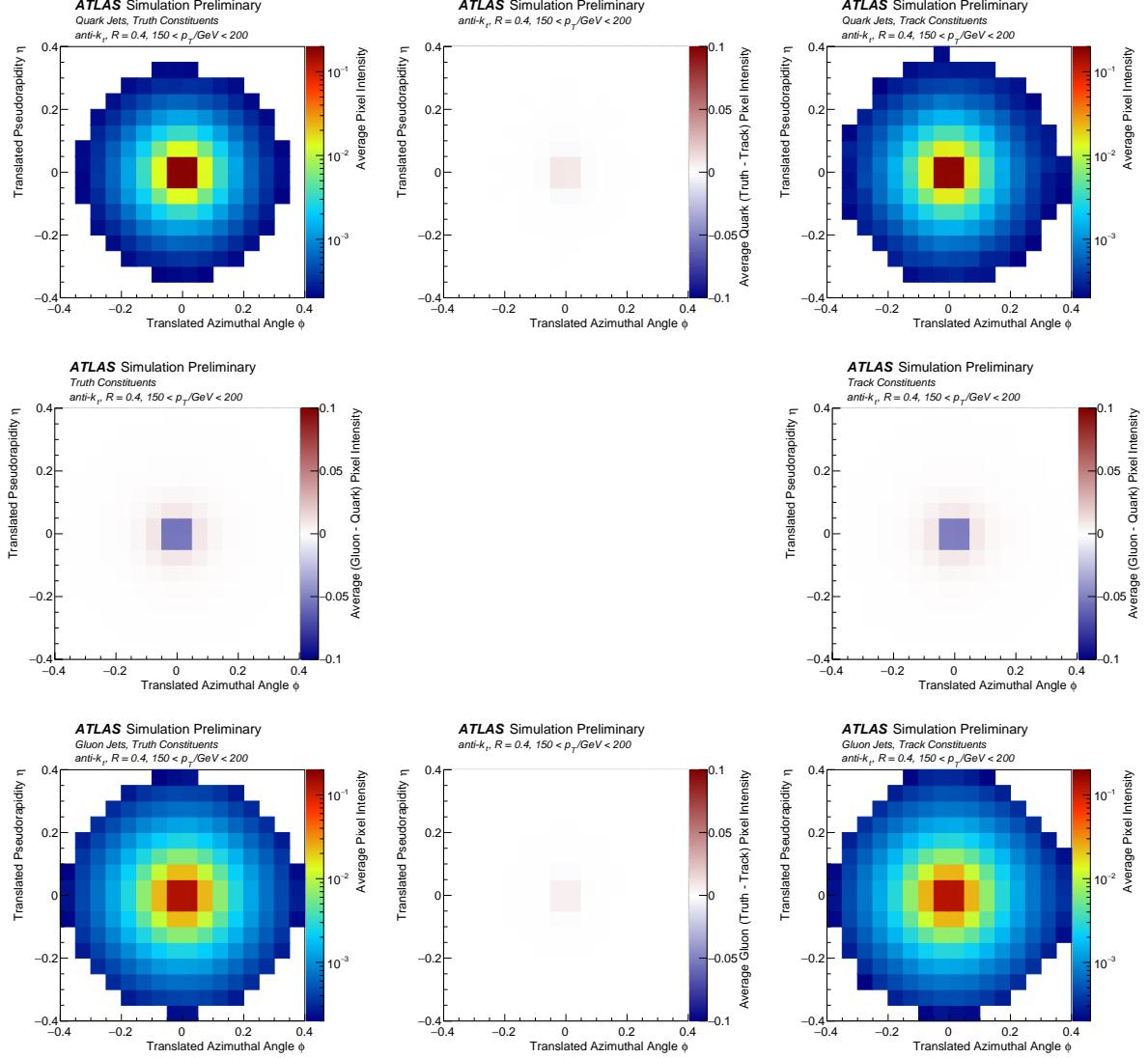


Figure 2: The four corners show the average quark (upper) and gluon (lower) jet images, from true constituents, both charged and neutral, (left) and reconstructed tracks (right); the four plots on the edges show the difference between the adjacent plots, for example the top plot shows the difference between the average quark jet for stable particles and reconstructed tracks. Quark-jets are more collimated than gluon ones, and track images show slightly less central activity than in the true jet.

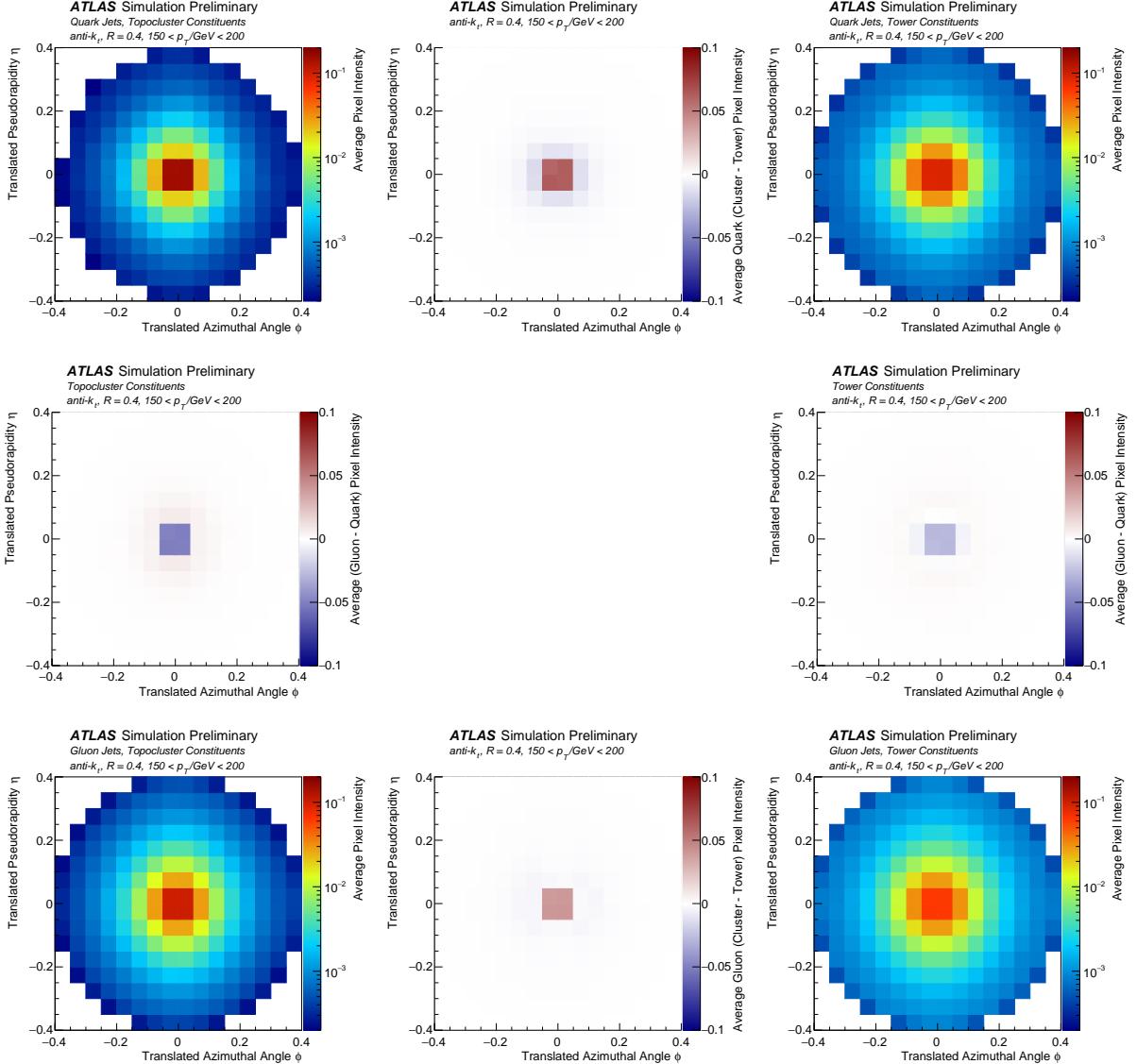


Figure 3: The four corners show the average quark (upper) and gluon (lower) jet images, from topo-clusters (left) and towers (right); the four plots on the edges show the difference between the adjacent plots, for example the top plot shows the difference between the average quark jet for topo-clusters and towers. Quark-jets are more collimated than gluon ones, and topo-cluster images are more collimated than tower images.

ATLAS Simulation Preliminary

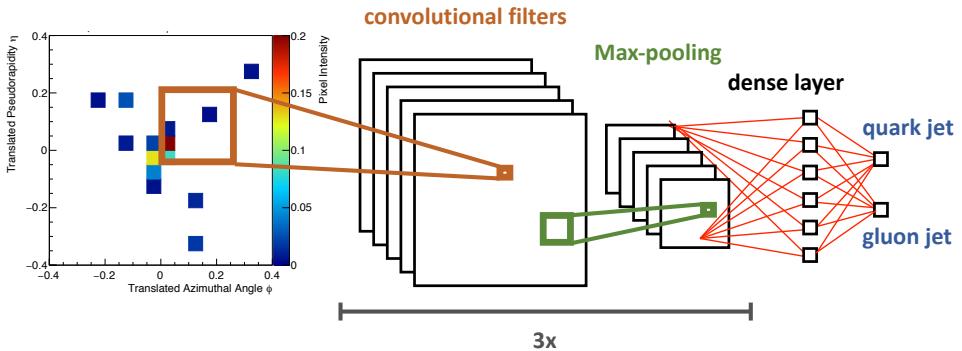


Figure 4: Illustration of the deep convolutional neural network architecture.

3 CNN tagger

The jet images are used as input for a deep neural network classifier. This section describes the implementation of the neural network and characterizes the properties and performance of the classifier.

3.1 Network architecture and training

In the studies presented in this note we make use of a deep convolutional neural network (CNN) to build a classifier of jet images. CNNs are a class of deep, feed-forward artificial neural networks based on modular sets of weights (filters) that operate linearly on a small $m \times n$ patch of the input image. The output of each filter is the dot-product between the weights and the pixel in the corresponding patch and it is typically assigned the position of the centre of the patch. Each filter is then *convolved* with the input image, by applying it to the whole image, a patch at a time, while moving the filter location across the image with a given stride. As a result of this operation, a new output image is obtained for each filter used. A non-linear activation function is typically applied to each pixel of the convolved image. A down-sampling procedure, referred to as Max-pooling [44], is then performed. This procedure takes non-overlapping patches of convolution outputs as input, and outputs the maximum value for each patch. The sets of filters, the activations and the Max-pooling constitute the fundamental building block for CNNs. The depth of the network is determined by the number of convolutions concatenated in the network.

The CNN architecture used in this note follows the example of Ref. [15] and consists of three iterations of a convolutional layer with a Rectified Linear Unit (ReLU) activation [45] and paired with a Max-pooling layer, all followed by a dense layer of 128 neurons with a ReLU activation. The output of the network is a softmax function [46] of size two, predicting the probability for the quark jet and the gluon jet class, respectively. The convolutional layers consist of 128, 128 and 64 filters, with filter sizes of 5×5 , 5×5 and 3×3 , respectively. The Max-pooling layers perform a 2×2 downsampling with a stride length of 2. In order to prevent overfitting, dropout [46] is applied to each convolution and the final fully connected layer with rate 0.3. In addition, a L2 regularization [46] with strength 10^{-8} is applied to all layers. A coarse scan of the various hyper-parameters was performed prior to settling on the architecture described above. An illustration of the architecture used is shown in Figure 4.

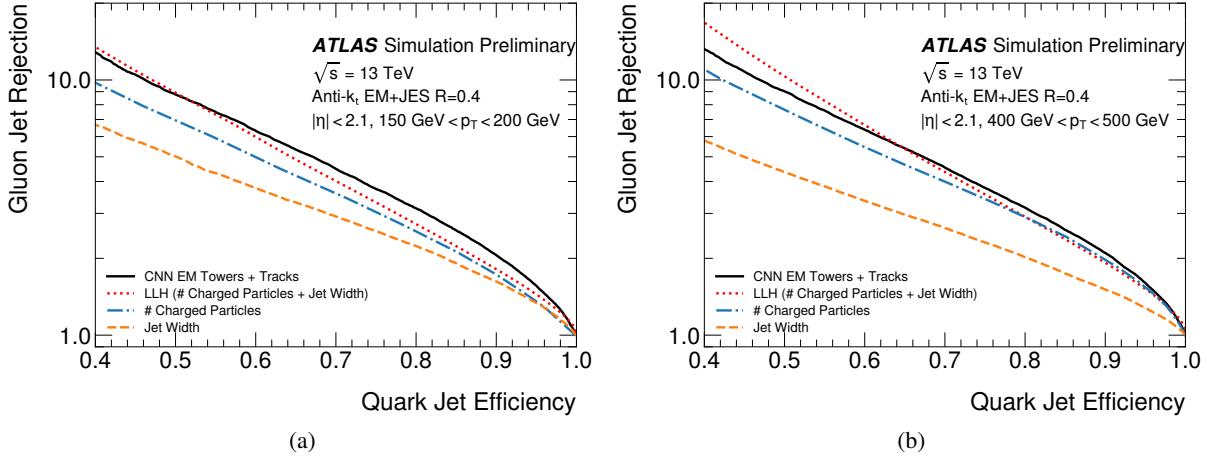


Figure 5: Gluon jet rejection as a function of the quark jet efficiency using physics motivated observables and jet image discriminants for jets with (a) $150 < p_T < 200$ GeV and (b) $400 < p_T < 500$ GeV. The LLH is a tagger constructed from the optimal (likelihood) combination of n_{track} and jet width.

Training is performed by minimizing the categorical crossentropy [46]. Minimization is performed with the Adam optimizer [47] as implemented in Keras [48] with a learning rate of 0.0001 over 50 iterations. Training is performed using a single NVidia Tesla K80 GPU with 224000 jet images, while 56000 jet images are used for testing. A typical training requires about 1 hour. The network is retrained for each of the two p_T ranges considered.

The output of the network corresponding to the quark jet class is used as a discriminant (*CNN tagger*). The discriminating power of the CNN tagger is compared with that of individual physically motivated observables, the calorimeter jet width w and the number of tracks n_{track} , and their combination with the 2D binned likelihood ratio (LLH) in Figure 5. Simple thresholds are applied to construct the n_{track} and jet width curves. Interestingly, the CNN tagger has a similar performance to the classic $n_{\text{track}}+jet$ width tagger that has been extensively studied in the past for quark versus gluon jet tagging [10, 11]. The overall performance improves with p_T as does the importance of n_{track} relative to jet width since the number of particles inside quark and gluon jets increases with p_T .

Given the promising performance suggested by Fig. 5, especially at lower p_T and high efficiency, the remainder of this note is dedicated to a preliminary probe of where and how the CNN is able to distinguish quark and gluon jets based on their radiation pattern.

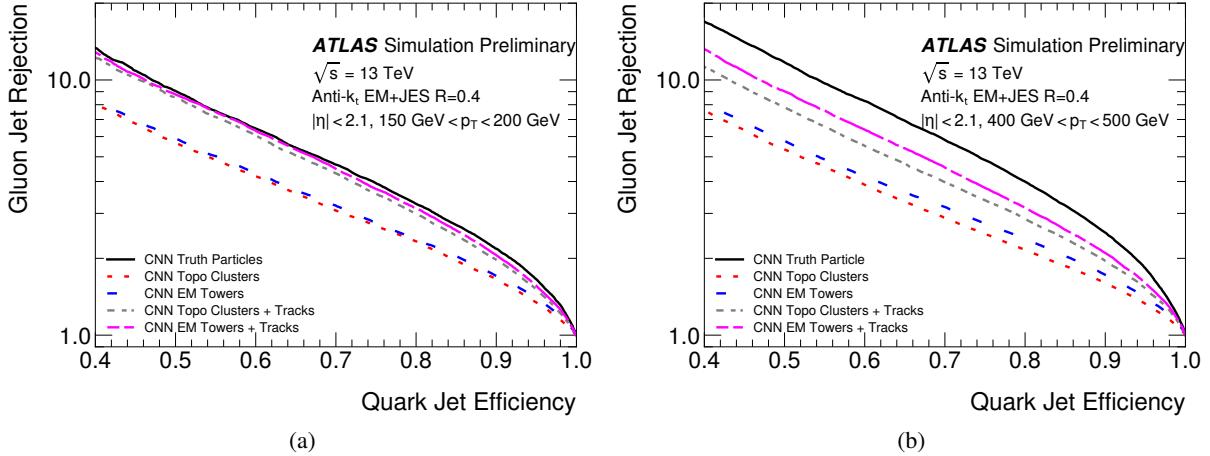


Figure 6: Gluon jet rejection as a function of the quark jet efficiency using the CNN tagger with different inputs for jets with (a) $150 < p_T < 200$ GeV and (b) $400 < p_T < 500$ GeV.

3.2 Inputs

The performance of the CNN tagger strongly depends on the choice of physics objects used to build the input images. Each input type is affected by experimental effects in different ways. Tracks only account for the charged particles in the jet; calorimeter towers include both charged and neutral component of the shower, but do not attempt to resolve individual particles. Calorimeter clusters aim at retaining the full granularity of the calorimeters while applying a zero suppression. They often merge close by particles and have a worse energy resolution than the track momentum resolution at low and moderate particle p_T . In addition, both calorimeter towers and clusters are more strongly affected by pile-up contributions. Figure 6 shows that the best discriminating performance is achieved by combining tower or topocluster images and track images as input for the CNN tagger. The two types of image are stacked so that the input to the CNN is a multi-dimensional array of size $16 \times 16 \times 2$. In the lower p_T range shown in Fig. 6, this combination actually saturates the achievable performance at low p_T bound set by using truth particles. There is a small gap in the higher p_T bin, suggesting further experimental improvements may help improve performance.

The difference between the full truth-particle line and the calorimeter + reconstructed tracks lines in Figure 6 depends on many details of the detector response, angular resolution, reconstruction efficiency, etc. However, it is possible to isolate one component of the difference by focusing only on the charged-particles. Figure 7(a) shows the performance of a track-only-image at both particle-level and detector-level. As might be expected, due to the precise measurement of charged-particle trajectories, there is little difference between the particle- and detector-levels. Degradation due to efficiency and resolution effects are only expected at much lower and higher transverse momenta.

3.3 Detector Regions

Both tower and topocluster inputs are projected onto a regular grid, but variations in the underlying geometry of the calorimeter could have an impact on performance. Focusing on towers, Figure 7(b)

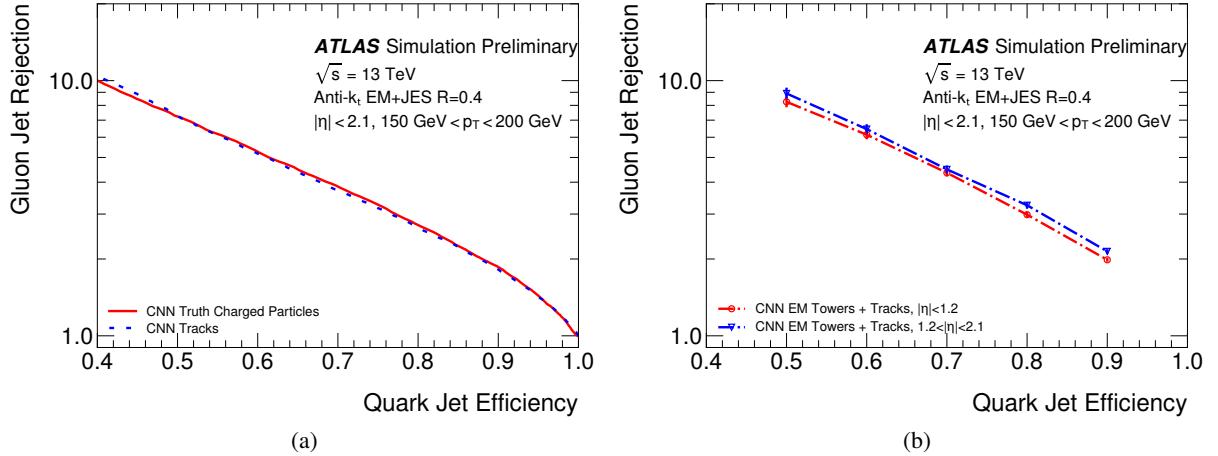


Figure 7: Gluon jet rejection as a function of the quark jet efficiency using the CNN tagger for jets with $150 < p_T < 200 \text{ GeV}$. **(a)** Comparison between using track images and truth charge particle images as input. **(b)** Comparison between different $|\eta|$ ranges. The full $|\eta|$ range ($|\eta| < 2.1$) is used for training.

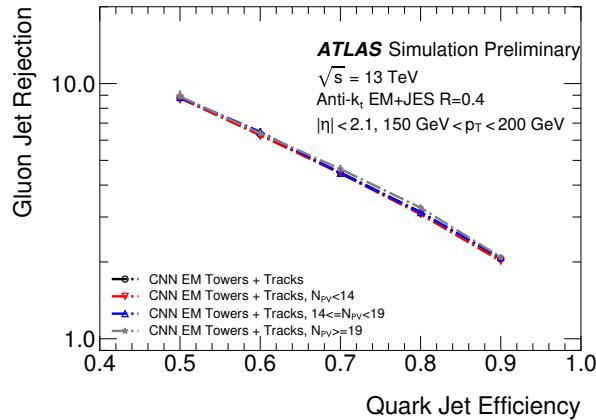


Figure 8: Gluon jet rejection as a function of the quark jet efficiency evaluated at different pileup conditions, quantified by the number of reconstructed primary vertices (N_{PV}).

shows that similar performance is achieved when testing the tagger on jets that are predominately in the barrel ($|\eta| < 1.2$) to those that are in the transition region between the barrel and endcap calorimeters ($1.2 < |\eta| < 2.1$). The full $|\eta|$ range ($|\eta| < 2.1$) is used for training.

3.4 Pileup dependence

The impact of pileup effects on the performance of the CNN tagger is evaluated by considering events with a different number of reconstructed primary vertices (N_{PV}). The test sample is divided into three bins: $N_{\text{PV}} < 13$, $13 < N_{\text{PV}} < 20$ and $N_{\text{PV}} > 20$. The distributions of the pixel intensities do vary with pileup, but the performance of the CNN tagger is found to be robust. This is demonstrated by Figure 8.

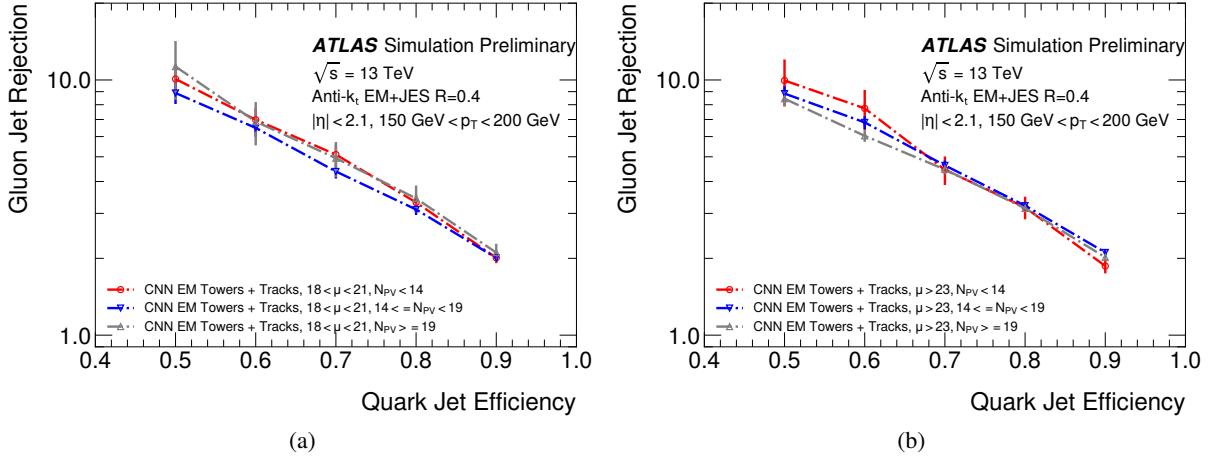


Figure 9: Gluon jet rejection as a function of the quark jet efficiency evaluated at different levels of N_{PV} for (a) $18 < \mu < 23$ and (b) $\mu > 23$. The two μ bins were chosen to have roughly the same number of events.

Calorimeter-based discrimination is not only effected by in-time pileup (quantified by N_{PV}) but also by out-of-time pileup. Figure 9 compares the tagger performance in two different regimes of the average number of collisions per bunch crossing (μ) regimes corresponding to the out-of-time pileup representative of LHC Run 2 conditions.

3.5 Fragmentation modeling

Quark versus gluon tagging is sensitive to the detailed modeling of perturbative and non-perturbative effects involved with fragmentation. We compare images and image classification performance with two very different fragmentation models. Figure 10 shows the average image (differences) between quark and gluon jets simulated with PYTHIA 8 and with HERWIG++. The radiation pattern inside gluon jets is similar between the two models, whereas there are larger differences for quark jets. In particular, quark and gluon jets are more different according to PYTHIA 8 relative to HERWIG++. This is observed with the CNN in Fig. 11. As gluon jets in HERWIG++ are more similar to quark jets than in PYTHIA 8, any observable (including the CNN tagger output) tends to have less discrimination power when tested in HERWIG++. The training sample determines the nature of the observable the CNN tagger learns. By comparing the CNN tagger trained on PYTHIA 8 events with the one trained on HERWIG++ events using the same test sample (PYTHIA 8), we see a smaller discrepancy, indicating that the learned feature is not strongly sensitive to the difference between PYTHIA 8 and HERWIG++. PYTHIA 8 and SHERPA produce relatively similar images and so unsurprisingly, the CNN performance is similar when trained/tested on either of these generators. As also noted by Ref. [39], when generators produce different images, the CNN returns a different performance when training and testing with images from various simulators. However, if the same network is used for testing and only the training sample is varied, the gap in performance is mostly removed (see also [15]). One explanation is that the network is learning robust features for quark versus gluon tagging, but the degree to which the features are expressed in the radiation patterns varies between generators.

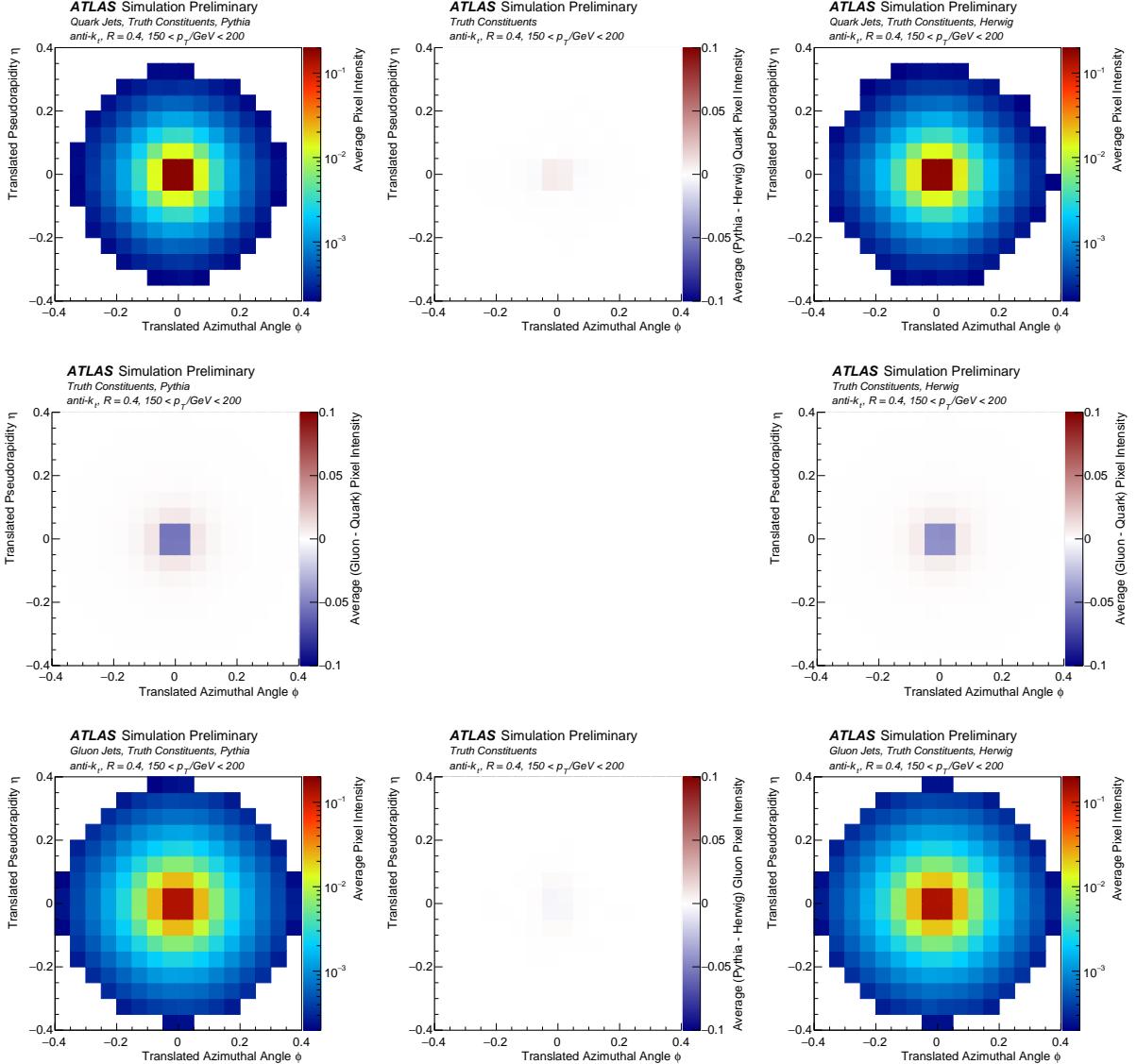


Figure 10: The four corners show the average quark (top) and gluon (bottom) jet images, for jets generated with PYTHIA (left) and and HERWIG (right); the four plots on the edges show the difference between the adjacent plots, for example the top plot shows the difference between the average quark jet in PYTHIA and and HERWIG.

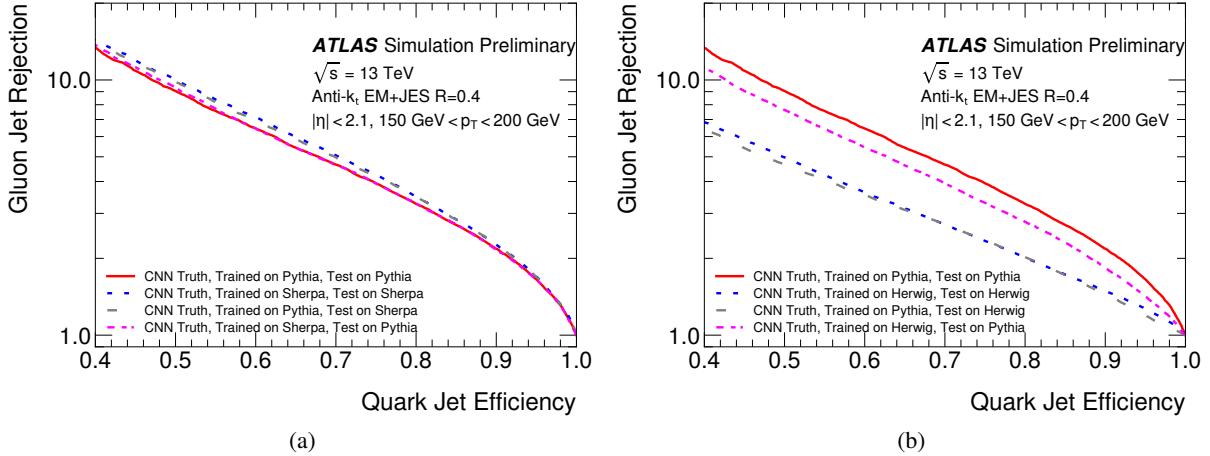


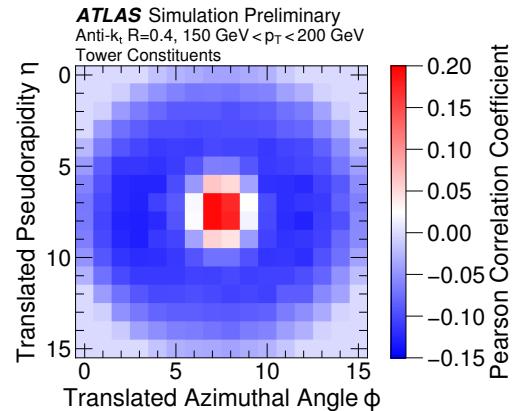
Figure 11: Gluon jet rejection as a function of the quark jet efficiency comparing PYTHIA to (a) SHERPA and (b) HERWIG for jets with $150 < p_T < 200$ GeV.

3.6 Visualizing learned information

In order to gain a deeper understanding of the physics learned by the CNN tagger, we examine how the internal structure of the network relates to the properties of quark and gluon jets, following the strategy outlined in Ref. [38]. Figure 12 shows, for each pixel of the jet image, the Pearson Correlation Coefficient of the pixel intensity with the final CNN tagger output. The intensity of the four pixels at the core of the jet are strongly correlated with CNN tagger output and are thus important to identify quark jets. The intensity of outer pixels in the jet image are instead anti-correlated with the CNN tagger output, in agreement with the intuition that gluon jets tend to feature a wider radiation pattern.

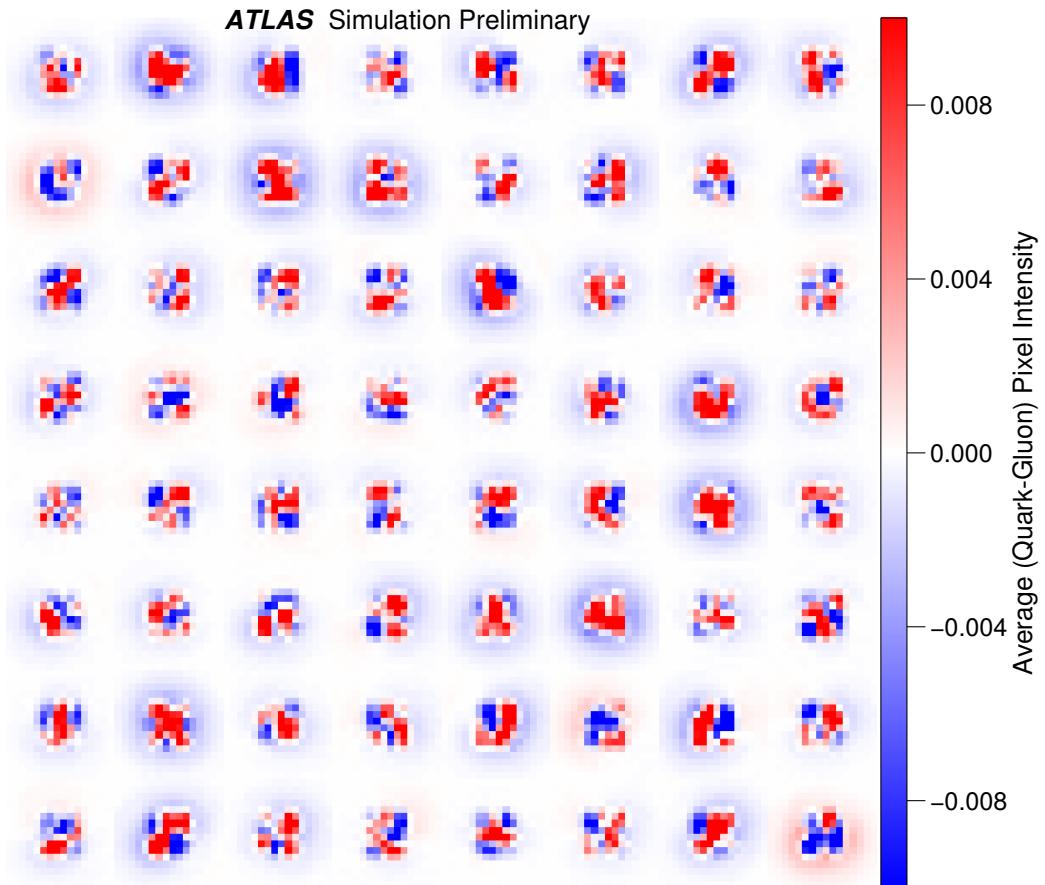
The discriminating information extracted by each convolutional filter can also be investigated. This is done by visualizing the average difference between the convolution of each filter with the input image for quark and gluon jets. More formally, let I_q and I_g be the average quark and gluon jet images, respectively. The average difference between the convolution of a filter w_i is obtained by computing $I_q * w_i - I_g * w_i$, where $*$ represents the convolution operator. The visualizations for the first convolutional layer of the CNN tagger are shown in Figure 13. As the images are not rotated prior to training, many of the filters are simply rotations of other filters. Some of the filters show features that are readily interpretable in terms of physical notions such as the soft haze around a jet core, but others are more complex and require further study to understand.

The visualizations in Figures 12 and 13 are an important start to probing what the CNN is learning about the jet radiation pattern.



(a)

Figure 12: Per-pixel linear correlation with CNN tagger output.



(a)

Figure 13: Average convolved filter differences for jet images (same color scheme as left plot; red is more quark-like). The filters of the first convolutional layer are considered.

4 Conclusion

This note presents the first full simulation study of deep learning with jet images. Convolutional neural networks are used to distinguish images of the radiation pattern inside quark jets from the pattern inside gluon jets. Across a wide range of quark jet efficiencies, the CNN tagger achieves a gluon jet rejection that is comparable to the performance of classical quark versus gluon tagging observables jet width and track multiplicity. For quark jet efficiencies above $\sim 60\%$, the CNN tagger outperforms track multiplicity and jet width. Combining or extending these taggers may improve the discrimination further. Furthermore, the origin of the CNN performance is investigated, including a variation of experimental conditions. The image-based approach described in this note is a promising avenue for future research to confront a variety of tagging challenges.

References

- [1] W. Bartel et al., JADE Collaboration, *Observation of planar three jet events in e^+e^- annihilation and evidence for gluon bremsstrahlung*, *Phys. Lett.* **B91** (1980) 142.
- [2] C. Berger et al., PLUTO Collaboration, *Evidence for Gluon Bremsstrahlung in e^+e^- Annihilations at High-Energies*, *Phys. Lett.* **B86** (1979) 418.
- [3] D.P Barber et al., MARK-J Collaboration, *Discovery of Three Jet Events and a Test of Quantum Chromodynamics at PETRA Energies*, *Phys. Rev. Lett.* **43** (1979) 830.
- [4] R. Brandelik et al., TASSO Collaboration, *Evidence for Planar Events in e^+e^- Annihilation at High-Energies*, *Phys. Lett.* **B86** (1979) 243.
- [5] A. J. Larkoski, J. Thaler and W. J. Waalewijn, *Gaining (Mutual) Information about Quark/Gluon Discrimination*, *JHEP* **11** (2014) 129, arXiv: [1408.3122 \[hep-ph\]](https://arxiv.org/abs/1408.3122).
- [6] P. Gras et al., *Systematics of quark/gluon tagging*, (2017), arXiv: [1704.03878 \[hep-ph\]](https://arxiv.org/abs/1704.03878).
- [7] C. Frye, A. J. Larkoski, J. Thaler and K. Zhou, *Casimir Meets Poisson: Improved Quark/Gluon Discrimination with Counting Observables*, (2017), arXiv: [1704.06266 \[hep-ph\]](https://arxiv.org/abs/1704.06266).
- [8] J. Gallicchio and M. D. Schwartz, *Quark and Gluon Tagging at the LHC*, *Phys. Rev. Lett.* **107** (2011) 172001, arXiv: [1106.3076 \[hep-ph\]](https://arxiv.org/abs/1106.3076).
- [9] J. Gallicchio and M. D. Schwartz, *Quark and Gluon Jet Substructure*, *JHEP* **04** (2013) 090, arXiv: [1211.7038 \[hep-ph\]](https://arxiv.org/abs/1211.7038).
- [10] ATLAS Collaboration, *Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Eur. Phys. J.* **C74** (2014) 3023, arXiv: [arXiv:1405.6583 \[hep-ex\]](https://arxiv.org/abs/1405.6583).
- [11] ATLAS Collaboration, *Discrimination of Light Quark and Gluon Jets in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector*, ATLAS-CONF-2016-034 (2016), URL: <http://cds.cern.ch/record/2200202>.
- [12] ATLAS Collaboration, *Quark versus Gluon Jet Tagging Using Charged Particle Multiplicity with the ATLAS Detector*, ATL-PHYS-PUB-2017-009 (2017), URL: <https://cds.cern.ch/record/2263679>.
- [13] CMS Collaboration, *Performance of quark/gluon discrimination in 8 TeV pp data*, CMS-PAS-JME-13-002 (2013), URL: <https://cds.cern.ch/record/1599732>.
- [14] CMS Collaboration, *Performance of quark/gluon discrimination in 13 TeV data*, CMS-DP-2016-070 (2016), URL: <https://cds.cern.ch/record/2234117>.
- [15] P. T. Komiske, E. M. Metodiev and M. D. Schwartz, *Deep learning in color: towards automated quark/gluon jet discrimination*, *JHEP* **01** (2017) 110, arXiv: [1612.01551 \[hep-ph\]](https://arxiv.org/abs/1612.01551).
- [16] L. M. Dery, B. Nachman, F. Rubbo and A. Schwartzman, *Weakly supervised classification in high energy physics*, *JHEP* **5** (2017) 145, arXiv: [1702.00414 \[hep-ph\]](https://arxiv.org/abs/1702.00414).

- [17] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, *Nucl. Phys.* **B126** (1977) 298.
- [18] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.
- [19] ATLAS Collaboration, *Early Inner Detector Tracking Performance in the 2015 data at $\sqrt{s} = 13 \text{ TeV}$* , ATL-PHYS-PUB-2015-051 (2015), URL: <http://cds.cern.ch/record/2110140>.
- [20] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, CERN-PH-EP-2015-304 (2016), arXiv: [arXiv:1603.02934 \[hep-ex\]](https://arxiv.org/abs/1603.02934).
- [21] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7 \text{ TeV}$* , *Eur. Phys. J. C* **73** (2013) 2304, arXiv: [1112.6426 \[hep-ex\]](https://arxiv.org/abs/1112.6426).
- [22] M. Cacciari, G. P. Salam and G. Soyez, *The Anti- $k(t)$ jet clustering algorithm*, *JHEP* **0804** (2008) 063, arXiv: [arXiv:0802.1189 \[hep-ph\]](https://arxiv.org/abs/0802.1189).
- [23] M. Cacciari, G. P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896, arXiv: [arXiv:1111.6097 \[hep-ph\]](https://arxiv.org/abs/1111.6097).
- [24] ATLAS Collaboration, *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, (2009), arXiv: [0901.0512 \[hep-ex\]](https://arxiv.org/abs/0901.0512).
- [25] M. Cacciari, G. P. Salam and G. Soyez, *The Catchment Area of Jets*, *JHEP* **0804** (2008) 005.
- [26] T. Sjostrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **0605** (2006) 026, arXiv: [arXiv:0603175 \[hep-ph\]](https://arxiv.org/abs/0603175).
- [27] T. Sjostrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852, arXiv: [arXiv:0710.3820 \[hep-ph\]](https://arxiv.org/abs/0710.3820).
- [28] ATLAS Collaboration, *ATLAS Run 1 Pythia 8 tunes*, ATL-PHYS-PUB-2014-021 (2014), URL: <https://cds.cern.ch/record/1966419>.
- [29] R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* **1504** (2015) 040, arXiv: [arXiv:1410.8849 \[hep-ph\]](https://arxiv.org/abs/1410.8849).
- [30] S. Agostinelli et al., GEANT4 Collaboration, *GEANT4: A Simulation toolkit*, *Nucl. Instrum. Meth.* **A506** (2003) 250.
- [31] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70** (2010) 823, arXiv: [arXiv:1005.4568 \[physics.ins-det\]](https://arxiv.org/abs/1005.4568).
- [32] J. Gao et al., *CT10 next-to-next-to-leading order global analysis of QCD*, *Phys. Rev. D* **89** (2014) 033009, arXiv: [arXiv:1302.6246 \[hep-ph\]](https://arxiv.org/abs/1302.6246).
- [33] M. H. Seymour and A. Siódmiak, *Constraining MPI models using σ_{eff} and recent Tevatron and LHC Underlying Event data*, *JHEP* **10** (2013) 113, arXiv: [arXiv:1307.5015 \[hep-ph\]](https://arxiv.org/abs/1307.5015).
- [34] D. Stump et al., *Inclusive jet production, parton distributions, and the search for new physics*, *JHEP* **10** (2003) 046, arXiv: [arXiv:hep-ph/0303013 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0303013).
- [35] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth.* **A462** (2001) 152.

- [36] J. Cogan et al., *Jet-Images: Computer Vision Inspired Techniques for Jet Tagging*, *JHEP* **02** (2015) 118, arXiv: [arXiv:1407.5675 \[hep-ph\]](https://arxiv.org/abs/1407.5675).
- [37] L. G. Almeida, M. Backovi, M. Cliche, S. J. Lee and M. Perelstein, *Playing Tag with ANN: Boosted Top Identification with Pattern Recognition*, *JHEP* **07** (2015) 086, arXiv: [1501.05968 \[hep-ph\]](https://arxiv.org/abs/1501.05968).
- [38] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *Jet-images ? deep learning edition*, *JHEP* **07** (2016) 069, arXiv: [1511.05190 \[hep-ph\]](https://arxiv.org/abs/1511.05190).
- [39] J. Barnard, E. N. Dawe, M. J. Dolan and N. Rajcic, *Parton Shower Uncertainties in Jet Substructure Analyses with Deep Neural Networks*, *Phys. Rev.* **D95** (2017) 014018, arXiv: [1609.00607 \[hep-ph\]](https://arxiv.org/abs/1609.00607).
- [40] G. Kasieczka, T. Plehn, M. Russell and T. Schell, *Deep-learning Top Taggers or The End of QCD?*, *JHEP* **05** (2017) 006, arXiv: [1701.08784 \[hep-ph\]](https://arxiv.org/abs/1701.08784).
- [41] P. Baldi, K. Bauer, C. Eng, P. Sadowski and D. Whiteson, *Jet Substructure Classification in High-Energy Physics with Deep Neural Networks*, *Phys. Rev.* **D93** (2016) 094034, arXiv: [1603.09349 \[hep-ex\]](https://arxiv.org/abs/1603.09349).
- [42] G. Louppe, K. Cho, C. Becot and K. Cranmer, *QCD-Aware Recursive Neural Networks for Jet Physics*, (2017), arXiv: [1702.00748 \[hep-ph\]](https://arxiv.org/abs/1702.00748).
- [43] ATLAS Collaboration, *Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC run 2*, (2017), arXiv: [1704.07983 \[hep-ex\]](https://arxiv.org/abs/1704.07983).
- [44] D. Scherer, A. Müller and S. Behnke, ‘Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition’, *Artificial Neural Networks – ICANN 2010: 20th International Conference, Thessaloniki, Greece, September 15-18, 2010, Proceedings, Part III*, ed. by K. Diamantaras, W. Duch and L. S. Iliadis, Springer Berlin Heidelberg, 2010 92, ISBN: 978-3-642-15825-4, URL: http://dx.doi.org/10.1007/978-3-642-15825-4_10.
- [45] X. Glorot, A. Bordes and Y. Bengio, *Deep Sparse Rectifier Neural Networks*, *Journal of Machine Learning Research* **15** (2011) 315, URL: <http://proceedings.mlr.press/v15/glorot11a/>.
- [46] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, <http://www.deeplearningbook.org>, MIT Press, 2016.
- [47] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, CoRR **abs/1412.6980** (2014), URL: <http://arxiv.org/abs/1412.6980>.
- [48] F. Chollet, ‘Keras’, <https://github.com/fchollet/keras>, 2015.