Quark jet versus gluon jet: deep neural networks with high-level features

Hui Luo\textsuperscript{a,b}, Ming-xing Luo\textsuperscript{a}, Kai Wang\textsuperscript{a}, Tao Xu\textsuperscript{a} and Guohuai Zhu\textsuperscript{a}

\textsuperscript{a} Zhejiang Institute of Modern Physics and Department of Physics, Zhejiang University, Hangzhou, Zhejiang 310027, China
\textsuperscript{b} II. Institut für Theoretische Physik, Universität Hamburg Luruper Chaussee 149, D- 22761 Hamburg, Germany

Abstract

Jet identification is one of the fields in high energy physics that deep learning has begun to make some impact. More often than not, convolutional neural networks are used to classify jet images with the benefit that essentially no physics input is required. Inspired by a recent paper by Datta and Larkoski, we study the separation of quark/gluon-initiated jets based on fully-connected neural networks (FNNs), where expert-designed physical variables are taken as input. FNNs are applied in two ways: trained separately on various narrow jet transverse momentum $p_{T,J}$ bins; trained on a wide region of $p_{T,J} \in [200, 1000] \text{ GeV}$. We find their performance are almost the same, and the larger $p_{T,J}$ the better. Comparing with results from deep convolutional neural networks, our results are comparable for low $p_{T,J}$, and even slightly better than those for high $p_{T,J}$. We also test the performance of FNNs with full set or different subsets of jet observables as input features. The FNN with one subset, consisting of fourteen observables, shows nearly no degradation of performance. This indicates that these fourteen expert-designed observables may have captured most of the information useful for the separation of quark/gluon jets.
I. INTRODUCTION

At the Large Hadron Collider (LHC), hadronic decay final states of bosons such as $W/Z$ or supersymmetric squarks are dominated by the light-quark-initiated jets while the corresponding standard model (SM) background often consists of gluon-initiated jets. Therefore, discrimination between quark jets and gluon jets is crucial in such searches. The quark-initiated jets and gluon-initiated jets were known to have different qualitative features for long time since the early measurements at PETRA and LEP colliders. For instance, radiation from color octet gluon will result in wider jet width in gluon jet comparing with the quark jet. However, the qualitative feature to discriminate quark from gluon was never as robust as the well-known b-jet tagging until a practical quark gluon jet tagging via charged particle multiplicity and jet width proposed in [1, 2] was finally employed by the ATLAS collaboration. This initiative effort was later joined by other quark gluon tagging observable proposals such as the N-subjettiness observables [3–5] and energy correlation functions [6, 7], etc. Detailed discussions can be found in recent reviews on jet substructure [8, 9].

On the other hand, the shallow neural networks has long been used in data analysis of high energy physics. Recently, the deep neural network technology has made breakthrough in pattern recognition. This new method has been employed in high energy physics as well, such as searches for new physics [10, 11], the identification of boosted top quarks and/or heavy bosons [12–22] and discussions on jet classification in general [23–31], neutrino physics [32–34], and other various topics [35–44].

For jet classification, the energy deposited in the calorimeter can be viewed as a grayscale image. Deep convolutional neural networks (DCNNs), which have shown to be very powerful in computer vision, were then applied to classify jet images [25]. But the information of charged particle multiplicity, which is one of the best variables for quark gluon tagging, is not included in grayscale jet images. As an attempt, pixel-level charged particle counts, transverse momentum $p_T$ of charged particles and $p_T$ of neutral particles are introduced in [26] as three “colors” of jet images. It is encouraging to observe that this DCNN outperforms traditional state-of-the-art methods in separating quark gluon jets. This is very impressive, especially considering that almost all of the expert-designed jet variables are not used in this procedure. But it is a pity that it is hard for DCNNs to tell us what physics they learn. Unlike daily life pictures, a jet image is often sparse, i.e. only few pixels are activated. Therefore it is also arguable whether DCNN is an optimal choice.
for jet classification [20].

In [29], a different method is developed: the physics-motivated N-subjettiness observables are taken as input to study the jet substructure using fully connected deep neural networks. The logic behind this method is that N-subjettiness observables can completely span the phase space of a jet substructure which contains all kinematic information in a jet. As a concrete example, [29] demonstrates that jet mass plus just eight N-subjettiness observables are enough for the separation of boosted Z jets from QCD jets. As noticed in [29], the choice of observable basis is not unique, one may use energy correlation functions as well. It was argued in [7] that a new series of energy correlation functions $U_i^{(\beta)}$ are powerful variables for the discrimination of light quark jets and gluon jets.

In this paper, inspired by [29], we study the quark/gluon tagging using fully connected deep neural networks. To capture as much as physical information, more jet substructure observables besides the jet mass and N-subjettiness are considered. Deep neural networks are very effective to analyze multidimensional problems, thus we implement this technique to classify signature/background. In the next section, we discuss the jet observables used as the input of the deep neural networks and the event generation procedure. We discuss the architecture and the results of the fully connected deep neural network in the third and fourth sections. The final section is devoted to a summary.

II. OBSERVABLES AND EVENT GENERATIONS

In the first part of this section, all observables as the input of neural networks are enumerated, which elaborate both entirety properties and substructures of a jet.

To generate a dataset for training and testing neural networks, we follow standard event generation procedures: generate all parton level events with MadGraph [46], and then do the showering with PYTHIA8.2 [47], the jets are clustered with FastJet [48] and observables are extracted using FastJet contrib as well as our private codes. Data generation processes are described detailedly in the second part of this section.

A. Observables as inputs of neural networks

The objects for quark/gluon discrimination considered in this work are the following:
• Particle multiplicity and charged particle multiplicity in a jet

• Jet mass $m_J$, transverse momentum of the jet $p_{TJ}$ and their ratio $m_J/p_{TJ}$

• Generalized angularities of two parameters, which were proposed in [49] and further discussed in [9]

\[ \lambda_\beta = \sum_{i \in \text{jet}} z_i^\kappa \theta_i^\beta. \] (1)

For proton-proton collision, using anti-$k_t$ algorithm [50] with E-scheme recombination [51], one has

\[ z_i \equiv \frac{p_{Ti}}{\sum_{j \in \text{jet}} p_{Tj}}, \quad \theta_i \equiv \frac{R_{in}}{R_0}, \] (2)

where $z_i$ is the momentum fraction of particle $i$, and $R_{in}$ is the rapidity/azimuth angle from a chosen axis\(^1\) of particle $i$, and $R_0$ is the radius of the jet under consideration. In particular, five values of $(\kappa, \beta)$, the same with [9, 49], are set as benchmarks,

\[(0, 0), \quad (2, 0), \quad (1, 0.5), \quad (1, 1), \quad (1, 2)\]

multiplicity $p_D^{pT}$ LHA width mass

Therein, variables with $\kappa = 1$ are all IRC safe while other two are IRC unsafe variants. Each of them corresponds to a specific physical quantity: (1) $(0, 0)$ duplicates the jet’s multiplicity; (2) $(2, 0)$ is known as $p_D^{pT}$ [6, 52]; (3) $(1, 0.5)$ is denoted as “LHA” (Les Houches Angularity) [9]; (4) width from $(1, 1)$ is related to broadening or girth [53–55]; (5) mass from $(1, 2)$ is related to thrust [56].

• N-subjettiness observables [4, 5] measure the radiation about N axes in a jet with a definition

\[ \tau_N^{(\beta)} = \frac{1}{p_{TJ}} \sum_{i \in \text{jet}} p_{Ti} \min \left\{ R_1^\beta, R_2^\beta, \ldots, R_N^\beta \right\}. \] (3)

As pointed out in [29], a basis from N-subjettiness objects can be constructed to span the phase space of appropriately identifying $M$ particles.\(^1\) This basis can then be taken as the

---

\(^1\) We use the axis directly from E-recombination in this paper. An alternative is with winner-take-all recombination scheme [60, 61].
input of deep neural networks to discriminate boosted hadronic Z decays from light parton intiated jets.

The N-subjettiness observables chosen in this work are the following
\[
\left\{ \tau_{1}^{(0.5)}, \tau_{1}^{(1)}, \tau_{2}^{(0.5)}, \tau_{2}^{(1)}, \tau_{2}^{(2)}, \tau_{3}^{(0.5)}, \tau_{3}^{(1)}, \tau_{3}^{(2)}, \tau_{4}^{(0.5)}, \tau_{4}^{(1)}, \tau_{4}^{(2)}, \tau_{5}^{(1)}, \tau_{5}^{(2)}, \tau_{5}^{(3)} \right\},
\]
which are identical to the ones used in [29, 30] for spanning the 6-body phase space in a jet. Two ratios \( \tau_{21}^{(1)} = \tau_{1}^{(1)}/\tau_{1}^{(2)} \) and \( \tau_{21}^{(2)} = \tau_{2}^{(1)}/\tau_{1}^{(2)} \) are also input in our work. So called “OnePass_WTA_KT_Axes” in FastJet package is chosen for above N-subjettiness objects.

- Generalized energy correlation functions [6, 7] can identify N-prong jet substructure without finding subjets first as done for N-subjettiness. In this work, we employ \( C_{N}^{(\beta)} \) (for quark gluon discrimination, \( N = 1 \ )) [6] and \( U_{i}^{(\beta)} \) [7]. Before moving on, let’s briefly review these variables.

The definition of \( C_{N}^{(\beta)} \) [6] is as follows,
\[
ECF(N, \beta) = \sum_{i_{1} < i_{2} < \ldots < i_{N} \in J} \left( \prod_{a=1}^{N} pT_{i_{a}} \right) \left( \prod_{b=1}^{N-1} \prod_{c=b+1}^{N} R_{i_{b} i_{c}} \right)^{\beta},
\]
\[
C_{N}^{(\beta)} = \frac{r_{N}^{(\beta)}}{r_{N-1}^{(\beta)}}, \quad r_{N}^{(\beta)} = \frac{ECF(N+1, \beta)}{ECF(N, \beta)},
\]
where \( r_{N} \) is much like the N-subjettiness \( \tau_{N} \) while \( C_{N}^{(\beta)} \) is similar to the N-subjettiness ratios \( \tau_{N,N-1}^{(\beta)} = \tau_{N}^{(\beta)}/\tau_{N-1}^{(\beta)} \). Both \( C_{N}^{(\beta)} \) and \( \tau_{N,N-1}^{(\beta)} \) are good to probe N-prong substructures in a jet. In particular, they are good measures for higher-order radiations from leading-order.

Observable \( U_{i} \) is proposed in [7] as
\[
U_{i}^{(\beta)} = 1^{e_{i+1}^{(\beta)}}, \quad \text{with} \quad \epsilon_{n}^{(\beta)} = \sum_{1 \leq i_{1} < i_{2} < \ldots < i_{n} \leq N} z_{i_{1}} z_{i_{2}} \ldots z_{i_{n}} \prod_{m=1}^{w} \min_{s < t \in \{i_{1}, i_{2}, \ldots, i_{n}\}} \left\{ \theta_{s t}^{\beta} \right\}.
\]

For proton proton collision, \( z_{i} \)’s share the same definition as (2) and \( \theta_{ij} \) denoting the opening angles between particle \( i \) and \( j \) in a jet. In this paper, we produce \( C_{1}^{(0.5)}, C_{1}^{(0.2)} \) and six different \( U_{i}^{(\beta)} \) objects with \( i = 1, 2, 3 \) and \( \beta = 0.5, 0.2 \) respectively. As a complementary, three normal energy correlators \( ECF(2, \beta) \) with \( \beta = 0.5, 1.0 \) and 2.0 are also produced.
B. Database generation

In this subsection, event generation from simulations is described. We generate pure quark and gluon dijet events at $\sqrt{s} = 13$ TeV in $pp$ collisions. The database is consist of observables from four different ranges of jet-transverse-momentum ($p_{TJ}$): $[200, 220]$ GeV, $[500, 550]$ GeV, $[1000, 1100]$ GeV and a wide one $[200, 1000]$ GeV. The reason to consider the range $[200, 1000]$ GeV is to train a single classifier for jets in broad kinematic regions, which is advantageous compared to the case of DCNN, as we shall discuss later.

With MadGraph5 v2.5.5 [46], hard processes $pp \rightarrow qq$ and $pp \rightarrow q\bar{q}$ for pure quark dijet are running, where $q$ represents for light-quarks ($u$, $d$, $s$). Similarly, $pp \rightarrow gg$ runs for pure gluon dijet productions at parton level. Mixed quark/gluon final state processes $pp \rightarrow qg$ has been shut down for simplicity. The $p_T$ cuts at parton level are $20\%$ broader than $p_{TJ}$ while all other cuts are turned off. In this way, we can partly prevent bias from cuts and ensure the efficiency of data generations.

Then, we pass the parton level results into PYTHIA 8.226 [47] and shower them with default parameters. Neutrinos are discarded and final state particles with $|\eta| < 2.5$ are kept. FastJet 3.3.0 [48] is used to cluster jets with the anti-$k_T$ algorithm [50] and E-scheme recombination [51] as well as the jet radius $R_0 = 0.4$. Jet mass and transverse momentum can be directly read out by using FastJet commands, and then it’s easy to obtain their ratio. Observables including charged multiplicity, five benchmark generalized angularities are generated with our private codes based on FastJet. N-subjettiness observables and the relevant energy correlation functions are all derived with FastJet contrib 1.027.

Finally we obtain one million events (evenly split into quark jets and gluon jets) for every narrow region of $p_{TJ}$: $[200, 220]$ GeV, $[500, 550]$ GeV, $[1000, 1100]$ GeV and 1.3 million events for the case of $[200, 1000]$ GeV. Notice that the generation of $[200, 1000]$ GeV data requires more careful treatment for two reasons. First, bias from kinematic cut becomes negligible in this situation. Secondly, the $p_T$ differential distribution drops quickly with larger $p_T$ at parton level. This suggests that we’d better generate events, not once-through, but in many sub-windows, with ap-

\footnote{It’s worth specifying definitions of quark/gluon jets. As already discussed in [9], there are ambiguities for quark/gluon tagging. However, since the processes considered here are simple enough and the radius of jet clustering is not that big, we take the most simplest definition of quark/gluon jets, namely, a one-to-one map between a jet and its initiating parton.}
proximately equal number of events in each sub-window. Therefore, we generate events every 25 GeV step within \([200, 600]\) GeV region and every 50 GeV step within \([600, 1000]\) GeV. The data gained from these 24 sub-windows formed the broadest kinematics region for our following study.

III. ARCHITECTURE OF FULLY CONNECTED NEURAL NETWORKS

We have described in the above section how to generate simulation data containing 36 expert-designed jet observables. In Fig. 1, the distributions of charged particle multiplicity, N-subjettiness \(\tau_1^{(0.5)}\) and energy correlation function \(U_1^{(0.5)}\) are plotted as examples. These observables are sensitive to the quark/gluon tagging. These curves are obtained based on the two millions of events showered with Pythia in the \(p_TJ\) bins of \([200, 220]\) GeV and \([1000, 1100]\) GeV, respectively.

Before feeding these observables into the neural networks, it is better to first standardize them, so that mean value and standard deviation of every feature are set to be 0 and 1, respectively. The one million events in the each given \(p_TJ\) range are split into three sets: \(8 \times 10^5\) events for training, \(10^5\) events for validation and another \(10^5\) events as test data. We use a fully connected deep neural network with six hidden layers. Each hidden layer has 300 nodes. For the activation function, we choose the rectified linear unit function (ReLU) for the hidden layers and the sigmoid function for the output layer. The binary cross-entropy loss function is minimized using the RMSprop optimization \([59]\).

The model contains more than 460000 unknown weights and biases as parameters and it is important to prevent its overfitting. For this purpose, the dropout regularization and validation-based early stop are adopted. The dropout ratio is taken to be 0.1 for all six hidden layers, and the training would be stopped if its performance on the validation data does not improve anymore for 20 epochs. The neural network model is implemented with Tensorflow \([57]\) and scikit-learn packages \([58]\). Running on a NVidia GTX 1080 GPU, it takes just several minutes to train the model with the input events in one given \(p_TJ\) region.

IV. RESULTS

Here, light quark jets are signals as they are presumably associated with the new physics, while the gluon jets are backgrounds. Therefore, the performance of neural networks can be measured by gluon rejection efficiency \((1 - \epsilon_g)\) as a function of quark acceptance efficiency \((\epsilon_q)\), known as
FIG. 1: Distributions of some observables sensitive for quark/gluon jets discrimination, measured on samples showered by Pythia in $p_T$ bins of $[200, 220]$ GeV (left) and $[1000, 1100]$ GeV (right). From top to bottom are the distributions of quark (red solid) and gluon jets (green dashed) of:

- charged particle multiplicity
- N-subjettiness $\tau_1^{(0.5)}$
- energy correlation function $U_1^{(0.5)}$

receiver operating characteristic (ROC) curve which is widely used in the field of machine learning. The area under the ROC curve (AUC) is also a useful quantity to measure the performance of models. Moreover, in collider physics, it may be better to plot the significance improvement char-
acteristic (SIC) curve $\epsilon_q/\sqrt{\epsilon_g}$ which is directly related to the statistical significance of the signal and background separation.

In Fig. 2, red solid lines represent ROC and SIC curves of neural networks with all jet observables discussed in Section II as input. The ROC AUCs are 0.877, 0.891 and 0.899 for the jets transverse momentum $p_{TJ}$ in the range of $[200, 220]$ GeV, $[500, 550]$ GeV and $[1000, 1100]$ GeV, respectively. Currently, deep convolutional neural network (DCNN) with color [26] has the best performance in quark/gluon discrimination. To compare with results from DCNN with color, we show in Table I the gluon jet efficiency at 50% quark jet acceptance. It turns out that our results from fully connected neural networks are basically as good as those of DCNN with color for $p_{TJ}$ around 200 GeV. For larger $p_{TJ}$, for example around 1000 GeV, our results are even slightly better than those from DCNN with color. This indicates that almost all information of quark/gluon jets has been included in jet observables used in this paper.

To get an idea of which subset of jet observables are important for the quark/gluon tagging, as an attempt, we also train neural networks by using different subsets of jet observables as follows:

• Case A. Input features include jet mass $m_J$, fourteen N-subjettiness observables $\tau_N^{(\beta)}$ listed in Eq. 4 and three ratios $m_J/p_{TJ}$, $\tau_1^{(1)}/\tau_1^{(2)}$ and $\tau_2^{(2)}/\tau_1^{(2)}$. This basis is essentially the same as the one proposed in [29], which is powerful for discrimination of boosted Z jets from QCD background. The results are shown in Fig. 2 as magenta dotted curves. One can directly compare these results with the red solid lines (with all of the jet observables). The difference appears very small using ROC measures, while the difference becomes more and more obvious in SIC curves with the jet $p_{TJ}$ increasing.

• Case B. Input features first include jet mass, particle multiplicity, generalized angularities $\lambda_{ij}^\kappa$ with $(\kappa, \beta)$ equals to $(2, 0)$, $(1, 0.5)$, $(1, 1)$ and $(1, 2)$, respectively. Then, energy correlation functions $ECF(2, \beta)$ with $\beta = 0.5, 1, 2$ and ratios $C_1^{(0.2)}$ and $C_1^{(0.5)}$ are involved. It was found in [7] that a new set of energy correlation functions $U_i^{(\beta)}$ are powerful for the quark/gluon tagging. So six $U_i^{(\beta)}$ observables ($i = 1, 2, 3$ and $\beta = 0.2$ or 0.5) are also taken into account here. The results are shown in Fig. 2 as green dot-dashed curves. However, this case has somehow inferior performance compared to other cases.

• Case C. Fourteen jet observables are considered, which includes particle multiplicity, charged particle multiplicity, LHA $\lambda_1^{(0.5)}$, jet mass, energy correlation function $U_1^{(0.5)}$ and
nine N-subjettiness observables \( (\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \tau_3^{(0.5)}, \tau_3^{(1)}, \tau_3^{(2)}) \). The results are shown in Fig. 2 as blue dashed curves. Interestingly, these results are basically as good as those using all jet observables. This implies that these fourteen observables may have captured most physical information which is useful for discriminating quark/gluon jets.

Notice that the jet images may look different for various transverse momenta, one has to train as many DCNNs as the number of transverse momentum bins. For example, three DCNNs are trained in [26] to classify quark/gluon jets with \( p_{TJ} \in [200, 220] \) GeV, [500, 550] GeV or [1000, 1100] GeV. Even so, there is no guarantee that jets with \( p_{TJ} \) falling outside the considered bins, say for example \( p_{TJ} \sim 700 \) GeV, could be classified efficiently using any of the three well-trained DCNNs. Certainly, this is not a very efficient way.

Up to now, we have strictly followed the way of DCNN such that different fully connected neural networks (FNNs) are trained for various jet transverse momentum regions. However, since transverse momentum is just one of the input features of FNN, due to the robust characteristic deep neural networks, nothing can stop us to train a single FNN for jets with very different transverse momentum.

As an attempt, a single FNN is trained for quark/gluon jets with \( p_{TJ} \) in the range of [200, 1000] GeV. We then test this FNN on jets with \( p_{TJ} \) in the bins of [200, 220] GeV, [500, 550] GeV and [1000, 1100] GeV, respectively. The corresponding ROC and SIC curves are shown in Fig. 3 with ROC AUCs being 0.876, 0.890 and 0.897 for \( p_{TJ} \in [200, 220] \) GeV, [500, 550] GeV and [1000, 1100] GeV, respectively. Comparing Figs. 2 and 3 one can see that a single FNN can discriminate quark/gluon jets in a wide range of transverse momentum without loosing any visible efficiency.

To compare our results from FNNs in a quantitative way with those from DCNNs with color [26], the gluon jet efficiencies at 50% quark jet acceptance are shown in Table I. For \( p_{TJ} \) around 200 GeV, the performance of FNN is comparable to that of DCNN with color. As \( p_{TJ} \) increases, the performance of FNN becomes even slightly better than that of DCNN with color.

V. SUMMARY

Deep learning approaches have developed many applications in high energy physics, among of which is jet identification, such as the separation of quark-initiated jets from gluon-initiated jets.
A natural way is to take the energy deposited in the calorimeter as a jet image. As a powerful tool, deep convolutional neural networks can then be used to classify jet images.
Motivated by [29], in this paper, we take 36 expert-designed jet observables as input features to discriminate quark/gluon jets using fully connected deep neural networks. One advantage of this method is that, the architecture of fully connected neural networks is much simpler than that of convolutional neural networks, and the former is also less GPU time-consuming. Since jet images do not lose any information, the approach of convolutional neural networks may be more powerful in jet classification if (nearly) all of the useful information could be extracted from the

\[ (b) \]

FIG. 3: The ROC (left) and SIC (right) curves of the fully connected neural network which trained on jets with the transverse momentum in the range of [200, 1000] GeV and tested on jets with the transverse momentum in the bins of [200, 220] GeV (magenta dotted lines), [500, 550] GeV (blue dashed lines) and [1000, 1100] GeV (red solid lines).

### TABLE I: Gluon jet efficiency at 50% quark jet acceptance for the transverse momentum of the jets in the range of [200, 220] GeV, [500, 550] GeV and [1000, 1100] GeV, respectively.

| Gluon jet efficiency at 50% quark jet acceptance | 1000 GeV (%) | 500 GeV (%) | 200 GeV (%) |
|-------------------------------------------------|--------------|--------------|--------------|
| FNN using all jet observables                    | 2.8          | 3.3          | 4.5          |
| FNN with Case A                                  | 3.2          | 3.7          | 5.0          |
| FNN with Case B                                  | 3.8          | 3.9          | 4.8          |
| FNN with Case C                                  | 2.8          | 3.3          | 4.6          |
| A single FNN trained using jets with \( p_T \) \( J \in [200, 1000] \) GeV | 2.8          | 3.4          | 4.6          |
| DCNN with color [26]                             | 3.4          | —            | 4.6          |
data. However, only few pixels are activated in a jet image, it remains an open question whether convolutional neural networks are the most efficient method in this situation.

We use a neural network with six hidden layers where 300 nodes are set in each hidden layer. Three neural networks are trained separately, based on one million events for each jet transverse momentum bins of [200, 220] GeV, [500, 550] GeV and [1000, 1100] GeV. As expected, the larger the jet transverse momentum, the better the performance of the neural networks. Specifically, the ROC AUCs are 0.877, 0.891 and 0.899 for the above three transverse momentum bins (larger AUC usually means better performance). The gluon jet efficiency at 50% quark jet acceptance is 4.5%, 3.3% and 2.8% for the three transverse momentum bins. These results are comparable to those from convolutional neural network for $p_{TJ} \in [200, 220]$ GeV, and even slightly better than those from convolutional neural network for $p_{TJ} \in [1000, 1100]$ GeV.

Many of these 36 jet observables should be complementary for the quark/gluon tagging, but some of them may be redundant. As an attempt, we test the performance of neural networks by choosing different subsets of the jet observables as input features. It is interesting to find that the neural network using only fourteen observables has as good performance as the neural network using all of thirty-six jet observables. These fourteen observables include particle multiplicity, charged particle multiplicity, LHA $\lambda_{1}^{(0.5)}$, jet mass, energy correlation function $U_{1}^{(0.5)}$ and nine $N$-subjettiness observables ($\tau_{1}^{(0.5)}, \tau_{1}^{(1)}, \tau_{1}^{(2)}, \tau_{2}^{(0.5)}, \tau_{2}^{(1)}, \tau_{2}^{(2)}, \tau_{3}^{(0.5)}, \tau_{3}^{(1)}, \tau_{3}^{(2)}$).

Since quark/gluon jet images may have different characteristics with various jet transverse momentum, it is customary to train different convolutional neural networks for different transverse momentum bins. But $p_{TJ}$ is just one of the input features of the fully connected neural networks (FNNs), it should be possible to train a single FNN for jet tagging with very different transverse momentum. Therefore we train such a FNN using 1.3 million data with the jet transverse momentum in the range of [200, 1000] GeV. Again, the performance of such a FNN is almost the same as those of FNNs trained separately for each transverse momentum bins.

In [30], a new method was proposed to construct novel jet observables with the help of fully connected neural networks. Such novel observables may have better discrimination power than widely-used observables. The novel observables may also deepen our understanding of the jet identification. It should be interesting to see in the future whether novel observables could also be constructed in the case of quark/gluon tagging.
Acknowledgement

We gratefully thank Andrew Larkoski for helpful discussions, and thank Gavin Salam and Torbjörn Sjostrand for clarifying technical details. HL is supported by the German Science Foundation (DFG) within the Collaborative Research Center 676 “Particles, Strings and the Early Universe”. Other authors are supported in part by the National Science Foundation of China (11135006, 11275168, 11422544, 11375151, 11535002) and the Zhejiang University Fundamental Research Funds for the Central Universities (2017QNA3007).

[1] J. Gallicchio and M. D. Schwartz, Phys. Rev. Lett. 107, 172001 (2011) doi:10.1103/PhysRevLett.107.172001 [arXiv:1106.3076 [hep-ph]].
[2] J. Gallicchio and M. D. Schwartz, JHEP 1304, 090 (2013) doi:10.1007/JHEP04(2013)090 [arXiv:1211.7038 [hep-ph]].
[3] I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, Phys. Rev. Lett. 105, 092002 (2010) doi:10.1103/PhysRevLett.105.092002 [arXiv:1004.2489 [hep-ph]].
[4] J. Thaler and K. Van Tilburg, JHEP 1103, 015 (2011) doi:10.1007/JHEP03(2011)015 [arXiv:1011.2268 [hep-ph]].
[5] J. Thaler and K. Van Tilburg, JHEP 1202, 093 (2012) doi:10.1007/JHEP02(2012)093 [arXiv:1108.2701 [hep-ph]].
[6] A. J. Larkoski, G. P. Salam and J. Thaler, JHEP 1306, 108 (2013) doi:10.1007/JHEP06(2013)108 [arXiv:1305.0007 [hep-ph]].
[7] I. Moult, L. Necib and J. Thaler, JHEP 1612, 153 (2016) doi:10.1007/JHEP12(2016)153 [arXiv:1609.07483 [hep-ph]].
[8] A. J. Larkoski, I. Moult and B. Nachman, arXiv:1709.04464 [hep-ph].
[9] P. Gras et al., JHEP 1707, 091 (2017) doi:10.1007/JHEP07(2017)091 [arXiv:1704.03878 [hep-ph]].
[10] P. Baldi, P. Sadowski and D. Whiteson, Nature Commun. 5, 4308 (2014) doi:10.1038/ncomms5308 [arXiv:1402.4735 [hep-ph]].
[11] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski and D. Whiteson, Eur. Phys. J. C 76, no. 5, 235 (2016) doi:10.1140/epjc/s10052-016-4099-4 [arXiv:1601.07913 [hep-ex]].
[12] P. Baldi, P. Sadowski and D. Whiteson, Phys. Rev. Lett. 114, no. 11, 111801 (2015).
[13] L. G. Almeida, M. Backović, M. Cliche, S. J. Lee and M. Perelstein, JHEP **1507**, 086 (2015) doi:10.1007/JHEP07(2015)086 [arXiv:1501.05968 [hep-ph]].

[14] J. Searcy, L. Huang, M. A. Pleier and J. Zhu, Phys. Rev. D **93**, no. 9, 094033 (2016) doi:10.1103/PhysRevD.93.094033 [arXiv:1510.01691 [hep-ph]].

[15] J. K. Behr, D. Bortoletto, J. A. Frost, N. P. Hartland, C. Issever and J. Rojo, Eur. Phys. J. C **76**, no. 7, 386 (2016) doi:10.1140/epjc/s10052-016-4215-5 [arXiv:1512.08928 [hep-ph]].

[16] J. S. Conway, R. Bhaskar, R. D. Erbacher and J. Pilot, Phys. Rev. D **94**, no. 9, 094027 (2016) doi:10.1103/PhysRevD.94.094027 [arXiv:1606.06859 [hep-ex]].

[17] R. Santos, M. Nguyen, J. Webster, S. Ryu, J. Adelman, S. Chekanov and J. Zhou, JINST **12**, no. 04, P04014 (2017) doi:10.1088/1748-0221/12/04/P04014 [arXiv:1610.03088 [hep-ex]].

[18] A. Alves, JINST **12**, no. 05, T05005 (2017) doi:10.1088/1748-0221/12/05/T05005 [arXiv:1612.07725 [hep-ph]].

[19] G. Kasieczka, T. Plehn, M. Russell and T. Schell, JHEP **1705**, 006 (2017) doi:10.1007/JHEP05(2017)006 [arXiv:1701.08784 [hep-ph]].

[20] J. Pearkes, W. Fedorko, A. Lister and C. Gay, arXiv:1704.02124 [hep-ex].

[21] M. Erdmann, B. Fischer and M. Rieger, JINST **12**, no. 08, P08020 (2017) doi:10.1088/1748-0221/12/08/P08020 [arXiv:1706.01117 [hep-ex]].

[22] A. Butter, G. Kasieczka, T. Plehn and M. Russell, arXiv:1707.08966 [hep-ph].

[23] P. Baldi, K. Bauer, C. Eng, P. Sadowski and D. Whiteson, Phys. Rev. D **93**, no. 9, 094034 (2016) doi:10.1103/PhysRevD.93.094034 [arXiv:1603.09349 [hep-ex]].

[24] D. Guest, J. Collado, P. Baldi, S. C. Hsu, G. Urban and D. Whiteson, Phys. Rev. D **94**, no. 11, 112002 (2016) doi:10.1103/PhysRevD.94.112002 [arXiv:1607.08633 [hep-ex]].

[25] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, JHEP **1607**, 069 (2016) doi:10.1007/JHEP07(2016)069 [arXiv:1511.05190 [hep-ph]].

[26] P. T. Komiske, E. M. Metodiev and M. D. Schwartz, JHEP **1701**, 110 (2017) doi:10.1007/JHEP01(2017)110 [arXiv:1612.01551 [hep-ph]].

[27] G. Louppe, K. Cho, C. Becot and K. Cranmer, arXiv:1702.00748 [hep-ph].

[28] C. Shimmin, P. Sadowski, P. Baldi, E. Weik, D. Whiteson, E. Goul and A. S?gaard, arXiv:1703.03507 [hep-ex].

[29] K. Datta and A. Larkoski, JHEP **1706**, 073 (2017) doi:10.1007/JHEP06(2017)073 [arXiv:1704.08249]
[hep-ph]].

[30] K. Datta and A. J. Larkoski, arXiv:1710.01305 [hep-ph].

[31] T. Cheng, arXiv:1711.02633 [hep-ph].

[32] E. Racah, S. Ko, P. Sadowski, W. Bhimji, C. Tull, S. Y. Oh, P. Baldi and Prabhat, arXiv:1601.07621 [stat.ML].

[33] A. Aurisano et al., JINST 11, no. 09, P09001 (2016) doi:10.1088/1748-0221/11/09/P09001 [arXiv:1604.01444 [hep-ex]].

[34] J. Renner et al. [NEXT Collaboration], JINST 12, no. 01, T01004 (2017) doi:10.1088/1748-0221/12/01/T01004 [arXiv:1609.06202 [physics.ins-det]].

[35] J. Barnard, E. N. Dawe, M. J. Dolan and N. Rajcic, Phys. Rev. D 95, no. 1, 014018 (2017) doi:10.1103/PhysRevD.95.014018 [arXiv:1609.00607 [hep-ph]].

[36] L. G. Pang, K. Zhou, N. Su, H. Petersen, H. Stocker and X. N. Wang, arXiv:1612.04262 [hep-ph].

[37] L. de Oliveira, M. Paganini and B. Nachman, arXiv:1701.05927 [stat.ML].

[38] P. Sadowski, B. Radics, Ananya, Y. Yamazaki and P. Baldi, arXiv:1706.01826 [physics.ins-det].

[39] T. Cohen, M. Freytsis and B. Ostdiek, arXiv:1706.09451 [hep-ph].

[40] P. T. Komiske, E. M. Metodiev, B. Nachman and M. D. Schwartz, arXiv:1707.08600 [hep-ph].

[41] M. Erdmann, J. Glombitza and D. Walz, arXiv:1708.00647 [astro-ph.IM].

[42] E. M. Metodiev, B. Nachman and J. Thaler, arXiv:1708.02949 [hep-ph].

[43] J. Ren, L. Wu, J. M. Yang and J. Zhao, arXiv:1708.06615 [hep-ph].

[44] S. Chang, T. Cohen and B. Ostdiek, arXiv:1709.10106 [hep-ph].

[45] D. Adams et al., Eur. Phys. J. C 75, no. 9, 409 (2015) doi:10.1140/epjc/s10052-015-3587-2 [arXiv:1504.00679 [hep-ph]].

[46] J. Alwall et al., JHEP 1407, 079 (2014) doi:10.1007/JHEP07(2014)079 [arXiv:1405.0301 [hep-ph]].

[47] T. Sjostrand et al., Comput. Phys. Commun. 191, 159 (2015) doi:10.1016/j.cpc.2015.01.024 [arXiv:1410.3012 [hep-ph]].

[48] M. Cacciari, G. P. Salam and G. Soyez, Eur. Phys. J. C 72, 1896 (2012) doi:10.1140/epjc/s10052-012-1896-2 [arXiv:1111.6097 [hep-ph]].

[49] A. J. Larkoski, J. Thaler and W. J. Waalewijn, JHEP 1411, 129 (2014) doi:10.1007/JHEP11(2014)129 [arXiv:1408.3122 [hep-ph]].

[50] M. Cacciari, G. P. Salam and G. Soyez, JHEP 0804, 063 (2008) doi:10.1088/1126-6708/2008/04/063 [arXiv:0802.1189 [hep-ph]].
[51] G. C. Blazey et al., hep-ex/0005012.

[52] S. Chatrchyan et al. [CMS Collaboration], JHEP 1204, 036 (2012) doi:10.1007/JHEP04(2012)036 [arXiv:1202.1416 [hep-ex]].

[53] S. Catani, G. Turnock and B. R. Webber, Phys. Lett. B 295, 269 (1992). doi:10.1016/0370-2693(92)91565-Q

[54] P. E. L. Rakow and B. R. Webber, Nucl. Phys. B 191, 63 (1981). doi:10.1016/0550-3213(81)90286-8

[55] R. K. Ellis and B. R. Webber, Conf. Proc. C 860623, 74 (1986).

[56] E. Farhi, Phys. Rev. Lett. 39, 1587 (1977). doi:10.1103/PhysRevLett.39.1587

[57] M. Abadi et al., TensorFlow: Large-scale machine learning on heterogeneous systems, software available from http://tensorflow.org/.

[58] F. Pedregosa et al., J. Mach. Learn. Res. 12, 2825 (2011).

[59] G. Hinton, http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf

[60] D. Bertolini, T. Chan and J. Thaler, JHEP 1404, 013 (2014) doi:10.1007/JHEP04(2014)013 [arXiv:1310.7584 [hep-ph]].

[61] A. J. Larkoski, D. Neill and J. Thaler, JHEP 1404, 017 (2014) doi:10.1007/JHEP04(2014)017 [arXiv:1401.2158 [hep-ph]].