Deep learning for *R*-parity violating supersymmetry searches at the LHC

Jun Guo,[1,2,*] Jinmian Li,[1,3,†] Tianjun Li,[2,4,‡] Fangzhou Xu,[2,5,§] and Wenxing Zhang[2,4,‖]

[1]*Center for Theoretical Physics, College of Physical Science and Technology, Sichuan University, Chengdu 610064, China*

[2]*CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China*

[3]*School of Physics, Korea Institute for Advanced Study, Seoul 130-722, Korea*

[4]*School of Physical Sciences, University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, China*

[5]*Institute of Modern Physics, Tsinghua University, Beijing 100084, China*

(Received 4 June 2018; published 30 October 2018)

Supersymmetry with hadronic *R*-parity violation in which the lightest neutralino decays into three quarks is still weakly constrained. This work aims to further improve the current search for this scenario by the boosted decision tree method with additional information from jet substructure. In particular, we find a deep neural network turns out to perform well in characterizing the neutralino jet substructure. We first construct a convolutional neutral network (CNN) which is capable of tagging the neutralino jet in any signal process by using the idea of jet image. When applied to pure jet samples, such a CNN outperforms the *N*-subjettiness variable by a factor of a few in tagging efficiency. Moreover, we find the method, which combines the CNN output and jet invariant mass, can perform better and is applicable to a wider range of neutralino mass than the CNN alone. Finally, the ATLAS search for the signal of gluino pair production with subsequent decay $\tilde{g} \to qq\tilde{\chi}_1^0(\to qqq)$ is recast as an application. In contrast to the pure sample, the heavy contamination among jets in this complex final state renders the discriminating powers of the CNN and *N* subjettiness similar. By analyzing the jets substructure in events which pass the ATLAS cuts with our CNN method, the exclusion limit on gluino mass can be pushed up by ∼200 GeV for neutralino mass ∼100 GeV.

DOI: 10.1103/PhysRevD.98.076017

I. INTRODUCTION

As one of the most promising new physics beyond the Standard Model (SM), supersymmetry (SUSY) [1,2] has been copiously searched for at the LHC [3,4]. With the Z_2 *R*-parity [5], the lightest supersymmetric particle (LSP) can be a weakly-interacting-massive-particle dark matter candidate with correct relic density [6]. Moreover, the *R*-parity conserving (RPC) SUSY at the hadron collider can be probed by looking for the particles with high transverse momenta and large missing energies in the final state. The gluino/squark masses have been excluded up to a couple of TeV [7,8] at the current stage of the LHC.

However, the *R* parity is not mandatory in SUSY models. In contrast to the RPC scenario where the yields of colored sparticles are constrained down to $\mathcal{O}(10)$ at the LHC run II with integrated luminosity of 36 fb^{-1}, some of the *R*-parity violating (RPV) scenarios are still weakly constrained. Thus, some improvements on the RPV searches are desired. In particular, the bounds on the RPV operators $U^cD^cD^c$, where U^c and D^c denote the right-handed up-type and down-type quark superfields, respectively, are quite weak due to the large hadronic activities expected at the LHC [9–14]. In our recent work [15], the status of LHC reaches on stop and sbottom masses with this kind of $U^cD^cD^c$ operators are studied. We found the stop and sbottom with mass ∼500 GeV are still not fully excluded. One of the important reasons is that the RPV scenarios were studied in the simplified model framework, such that the information of a specific signal was not fully explored.

In the hadronic RPV case, the decay products of the boosted heavy sparticle will be collimated, forming a single fat jet at the detector. The information from the fat jet substructure (see Refs. [16–21] for reviews) was found

*hustgj@itp.ac.cn
†jmli@kias.re.kr
‡tli@itp.ac.cn
§xfz14@mails.tsinghua.edu.cn
‖zhangwenxing@itp.ac.cn

to be useful in improving the search sensitivities, e.g., neutralino jet substructure [22] or stop jet substructure [23,24]. To characterize the jet substructure, traditionally, some high-level kinematic variables such as mass drop [25] and N subjettiness [26] are defined on the jet. On the other hand, all information of a jet can be inferred from the electromagnetic and hadronic calorimeters, with the basic observables being the position in the $\eta-\phi$ plane and energy deposit of each calorimeter cell. By identifying each cell as a pixel and the energy deposit in the cell as the intensity (or gray scale color) of that pixel, the jet can be naturally viewed as a digital image. The recent developments of computer vision can be applied as helpful tools for us to tag the jet nature with low-level inputs. There are a number of works that use the jet image to discriminate the hadronic W/Z jet [27–30] and top quark jet [31–33] from the QCD jet, and discriminate the quark jet from the gluon jet [34,35]. These studies show that the jet taggers based on computer vision perform comparably or even slightly better than those based on the high-level kinematic variables. Some improved algorithms have been proposed in Refs. [36–38]. It has been realized recently that the idea of jet image suffers from the disadvantage of low efficiency attributed to sparsity [28]. Machine learning techniques other than image recognition have been considered, such as using recursive neural networks [39,40], taking ordered sequence of jet constituents as inputs [41], and working on Lorentz vectors of jet constituents [42].

In this work, we will try to improve a realistic RPV SUSY search at the LHC by using the boosted decision tree (BDT) method [43] that takes into account the jet substructure information. In particular, a convolutional neutral network (CNN) (for pedagogical introductions, see Refs. [44,45]) is found to be efficient in tagging the substructure of the neutralino jet. The signal process under consideration is the gluino pair production, which decays into two quarks and a neutralino. The neutralino will subsequently decay into three quarks through the hadronic RPV operator $U^cD^cD^c$. The main task of the CNN is to discriminate the boosted neutralino jet in this signal process from the QCD jet in SM background processes. First, there is no prototype in the SM that is producing the same three-prong structure from three body decay as a neutralino jet. Also, the mass of the neutralino is an unknown parameter. We will show that the change of the CNN tagging efficiency when it is applied to the neutralino mass is different from the one that the CNN is trained on. In order to tag the neutralino jet irrespective of its production mechanism, our CNN is first trained on events of a simplified process with only a visible neutralino jet in the final state. Then it will be applied to each jet in both the signal and background events that pass all selections in the ATLAS search. Combining the discriminating power of the CNN scores and the jet invariant masses of the leading three jets with the BDT method, the signal and background can be separated further, leading to better search sensitivity.

The paper is organized as follows. In Sec. II, the architecture of the CNN that is adopted in this paper will be given. Section III discusses the training process and performance of the CNN on a simplified signal process. Its application to a realistic RPV gluino search is studied in Sec. IV. Our conclusion is provided in Sec. V.

II. THE CNN ARCHITECTURE

There exist many CNN architectures, such as the `VGGNet` [46] and `ResNet` [47]. They have been proved to be very successful in classifying images of either large size (in the PASCAL visual object classes data set [48]) or small size (in the `CIFAR-10` [49] data set). As for our

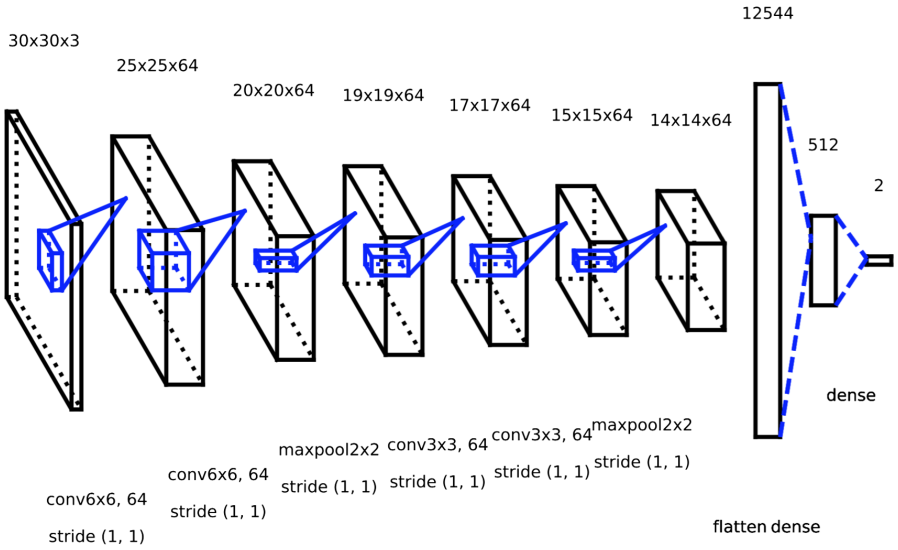

FIG. 1. The architecture of our CNN for one of the parameter choices.

case, due to the limited angular resolution of the detectors at the hadron collider, the jet image is usually smaller than 30×30 pixels. It has similar size as the images in the CIFAR-10 [49] data set. Inspired by the VGGNet architectures that were optimized for the CIFAR-10 data set, the sketch of our CNN architecture is shown in Fig. 1.

The input consists of three layers defined as the energy distribution of all particles, the energy distribution of charged particles, and the number of charged particles in calorimeter cells. The more detailed jet image preprocessing will be introduced later. These data are then passed through two iterations of two convolutional layers with rectified linear unit (ReLU) activation and a max-pooling layer. The size and the total number of convolution kernels (also called "filter") in each convolutional layer are free parameters. In practice, we need trial and error to figure out the best choice. In the figure, at the first step of iteration, the input is convoluted twice by 64 filters with same size of 6×6, followed by max pooling with a filter of size 2×2 and with stride of one. The second step of iteration has the same parameters except the size of the convolutional filters is reduced to 3×3. The total number of filters in each convolutional layer and the filter in the pooling layer remain the same as the first iteration. The feature map is flattened and read by the fully connected neural network (FCNN). There are 512 neural nodes defined in the hidden layer of the FCNN where the ReLU activation function has been adopted. The final output layer contains two nodes with a sigmoid activation function. With the output value of each node between [0,1], it can be used to characterize the probability of being either signal or background.

III. TRAINING AND TESTING OF THE CNN

Our goal is to employ a CNN that can recognize the jet image of a neutralino from jet images of a quark and gluon, so that the signal processes with a neutralino in the final state can be separated further from the backgrounds. To make our CNN a general neutralino jet recognizer which is not specific to any detailed production process, the training of the CNN is based on the signal event samples with only one visible neutralino in the final state, which subsequently decays into three quarks. Throughout the work, the hard-scattering signal and background events as well as the neutralino decay are simulated by the MADGRAPH5_AMC@NLO program [50]. The PYTHIA8 package [51] is used to perform the parton shower and hadronization. The detector effects are simulated by DELPHES3 [52] with the ATLAS configuration card, in which the jet reclustering algorithm is implemented via FASTJET [53] software. Our CNN is implemented in PYTHON using the deep learning library Keras [54].

The training and testing samples are generated and processed as follows. First, the signal events with a single visible neutralino jet are generated by the $pp \to \tilde{\chi}_1^0 \tilde{\chi}_2^0$ process in the SUSY model, with $\tilde{\chi}_2^0 \to bcs$ through the $U_2^c D_2^c D_3^c$ operator [15]. The $\tilde{\chi}_1^0$ is assumed to be stable here, which leaves nothing inside the detector.[1] As a benchmark, we choose the mass of $\tilde{\chi}_2^0$ to be 100 GeV. Its transverse momentum is required to be $p_T(\tilde{\chi}_2^0) > 200$ GeV, so its decay products are collimated and behave as a jet at the detector. Furthermore, it is obvious that the neutralino jet image will be varying if the polar angle (or pseudorapidity) of the neutralino is changed. To consider this effect, two classes of signal events sample are generated: one with the requirement of $|\eta(\tilde{\chi}_2^0)| < 0.1$ (central sample) and the other allows a much larger pseudorapidity $|\eta(\tilde{\chi}_2^0)| < 2.5$ (wide sample). Second, the background events in training and testing are generated by $pp \to j\tilde{\chi}_1^0 \tilde{\chi}_1^0$ in the SUSY model, where j can be either a quark or gluon and $\tilde{\chi}_1^0$ is stable at the detector. As in signal event generation, the transverse momentum of j is required to be $p_T(j) > 200$ GeV, and two classes of background samples with cuts of $|\eta(\tilde{\chi}_2^0)| < 0.1$ and $|\eta(\tilde{\chi}_2^0)| < 2.5$ are defined. It should be noted that during the training and testing stage, the initial state radiation and multiparticle interaction have been turned off in PYTHIA8 for both signal and background event generation.[2] Thus, their contaminations to the target jet image are suppressed, and the CNN can grab the important features of the target jets more efficiently. Third, in both signal and background events, jets are reconstructed by the anti-k_t algorithm [55] with cone size $R = 1.0$. The minimal transverse momentum of the target jet should be 100 GeV.[3] An event will be dropped if there is no jet with $p_T(j) > 100$ GeV. In the case of more than one jet with $p_T(j) > 100$ GeV in an event, the jet with the highest p_T is chosen. For signal events, we also require that the selected jet lie within a cone size of $R < 1.0$ of the parton level $\tilde{\chi}_2^0$.

At this stage, each event has been associated with a single jet, which is expected to be a neutralino jet (QCD jet) for the signal (background) event. Next, we need to convert the jet information into a grid image. Given a jet, its hardest constituent is located on the $\eta - \phi$ plane. Afterwards, a grid with a step of 0.1×0.1 and size of 30×30, which is centralized at the hardest constituent, is defined. Based on the grid and jet constituent information, we can define three different layers for the jet image: (1) The layer that shows the energy grid of all jet constituents, where the energies of the jet constituents belong to the same cell are added up; (2) as in the first layer, but only the energy of the charged jet constituents is taken into account; (3) the layer counts the

[1]This is a trick in generating process-independent neutralino jets for training and testing. In the next section, considering a complete model, $\tilde{\chi}_1^0$ is the LSP that decays into three quarks, i.e., bcs.

[2]These effects will be included when considering a realistic gluino search in the next section.

[3]This requirement is looser than that at parton level because we find the reconstructed jet can be softer than the parton level jet sometimes due to large angle splitting.

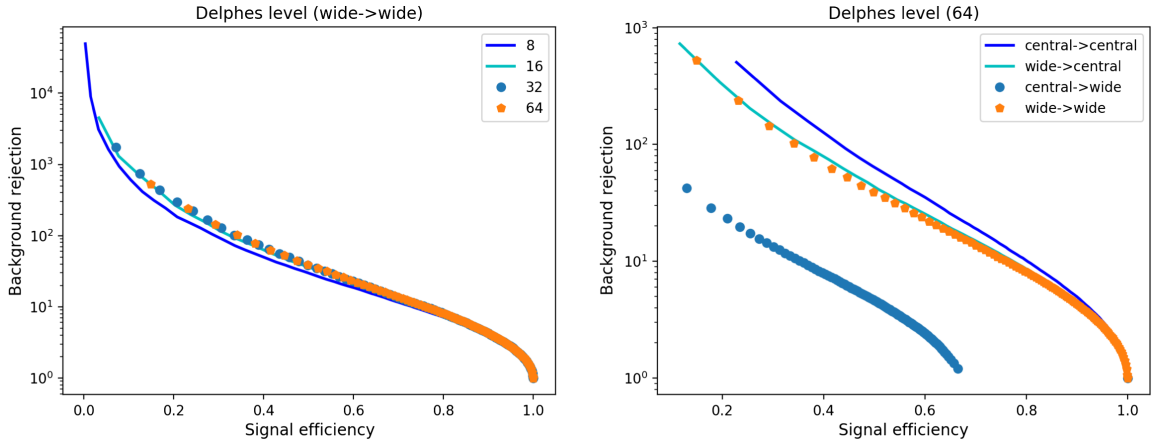

FIG. 2. Left panel: Performances of the CNNs with different number of convolutional filters in convolutional layers. Right panel: Performances of the CNNs that have been trained on the central sample (wide sample) and applied to either the central sample or wide sample. Details of other parameter choices are discussed in the text.

number of charged jet constituents in each cell. Since the CNN is found to be most efficient in dealing with numbers between [0,1], all numbers in each layer are divided by the maximum value in that layer, e.g., the maximum energy of the cell in the first layer. We will not apply anymore image preprocessing procedures, such as rotation and flipping, because they were found to decrease the performance of our CNN (same finding as in Ref. [32]).

Finally, to use our data set in a more efficient way (we have generated 10^6 signal and background events for training), 30 epochs are required during the training process, and to avoid the overtraining problem, an independent data set of 10^6 signal and background events is used for testing.

There are a number of free parameters in the CNN that can only be optimized through trial and error, including the sizes and numbers of the convolutional kernels in the convolutional layers, the dropout rates after two iterations and FCNN, the number of nodes in the hidden layer of the FCNN, and the learning rate in the NAdam algorithm [56]. We find the performance of the CNN only mildly depends on these parameters. In the left panel of Fig. 2, the performances of the CNNs with the number of convolutional filters in the convolutional layers being 8, 16, 32, and 64 are shown (the same number is adopted in all convolutional layers). The CNN with more than 16 convolutional kernels performs equally well—slightly better than the one with eight convolutional kernels. To obtain the results, we have taken the size of the convolutional kernel to be 6×6,[4] the dropout rate in two iterations as 0.25, while it is 0.5 for the FCNN. The number of nodes in the hidden layer is 512, and the learning rate is taken to be 0.001. This parameter choice will be used throughout this work. Even though the number of trainable parameters here ($\sim 6.5 \times 10^6$) is larger than the size of the training sample, our CNN is still working fine because of the following two reasons. First, we have tried the CNN with a much smaller parameter set (with eight filters in the convolutional layers and 64 nodes in the hidden layer of the FCNN, the parameter number is $\sim 0.1 \times 10^6$), and its performance is slightly worse than the one shown in Fig. 1. Second, the trained CNN has been tested on an independent event sample, which gives similar accuracy. So, the CNN is not overtrained on the training sample. Note that we have defined two CNNs that are trained and tested on a central sample and a wide sample of signal and background events, respectively. The results presented in the left panel correspond to the wide-sample-trained CNN applied to another independent wide sample. In the right panel, to characterize the dependence of the jet image feature on the jet pseudorapidity, we show the performance of these two sets of CNNs (both with 64 filters in all convolutional layers) on different samples. There is no doubt that the central jet ($|\eta| < 0.1$) is easier to tag than the jet within the wide pseudorapidity range ($|\eta| < 2.5$). The CNN trained and tested on the central sample is not working for tagging neutralino jets in the wide sample, mainly because the features captured by the CNN in the central sample are not useful for the wide sample. On the other hand, the CNN trained and tested on the wide sample performs well in tagging neutralino jets in the central sample, even though it is slightly worse than the CNN that is trained and tested directly on the central sample. This means we do not have to limit our analysis to the phase space with the target jet in the central region. It is especially useful in a realistic signal search at the LHC, so more signal events can be saved. In the following, we will keep using the CNN that is trained and tested on the wide sample with the filter number in each convolutional layer being 64.

We should compare the performance of our CNN with these high-level jet substructure variables. Among these,

[4]We find the CNNs with filter sizes of 2×2 and 4×4 perform worse.

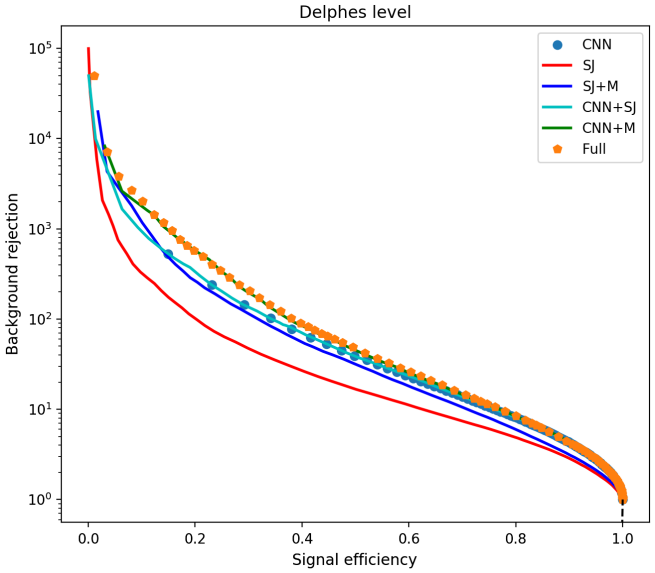

FIG. 3. Comparison among the performances of different methods of discriminating the neutralino jet from the QCD jet; SJ denotes the N-subjettiness variable, M is the jet invariant mass, and the combination of different variables is managed by the BDT method.

the N subjettiness is a general and effective discriminating variable that can characterize the multiprong structure of a jet. It is defined as [26]

$$\tau_N = \frac{\sum_k \min\{\Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k}\}}{\sum_k p_{T,k} R_0}, \qquad (3.1)$$

where k runs over all constituent particles in a given jet, $p_{T,k}$ is the transverse momentum of the kth constituent, $R_{J,k}$ is the distance between a candidate subjet J and the kth constituent in the $\eta - \phi$ plane, and R_0 is the characteristic jet radius that is used in the original jet clustering algorithm. A jet with an N prong will have $\tau_N \sim 0$ when all of its constituents are aligned with candidate subjets, while $\tau_I \gg 0$ for $I < N$ because there are constituents distributed away from the candidate subjet directions. As a result, the variable τ_N/τ_{N-1} is found to be efficient in tagging jets with N-prong structure. In our case, the neutralino jet substructure can be tagged by τ_3/τ_2. The performance of the N-subjettiness technique is shown by the red solid line in Fig. 3. We find that the performance of our CNN (represented by blue dots) is a few times better than that of the N subjettiness. Moreover, the jet invariant mass is a powerful discriminating variable that is independent of N subjettiness. To combine the discriminating power of both variables, the BDT method is adopted. Because the BDT only needs to learn two-dimensional information here, a relatively small size of forest should be enough. It uses a 100 tree ensemble that requires minimum training events in each leaf node of 2.5% and a maximum tree depth of three. The rest of the parameters are set to the default ones in the TMVA package [57]. It is trained on half of the reconstructed neutralino and QCD jets and is tested on the rest of the jets ($\sim 0.5 \times 10^6$ each). To avoid overtraining, the Kolmogorov-Smirnov test [58] in the BDT training and testing is required to be greater than 0.01.[5] The performance of the combination of N subjettiness and jet invariant mass is given by the blue solid line, which shows the similar tagging efficiency as the CNN alone.

Meanwhile, it is worth finding out whether our CNN is clever enough to learn both the N-prong structure and the jet invariant mass [30,59]. This can be seen through the tagging efficiencies of their combinations. In Fig. 3, the performances of the CNN + N subjettiness (SJ) and the CNN + jet invariant mass (M) are shown by cyan and green solid lines, respectively. The combination of their sensitivities is managed by the BDT method, with the same parameters as introduced above. The CNN + SJ does not show much more improvement than the CNN alone. while the tagging efficiency can be improved by a factor of a few after including the jet invariant mass. Thus, we can conclude that the full information of the prong structure in a jet can be learned by the CNN, but the jet invariant mass cannot be directly extracted from the jet image by the current method. One reason is that the image preprocessing procedures do not respect the Lorentz symmetry, so the jet invariant mass is broken down in the preprocessing [21,28,32].

In the above study, the neutralino mass has been taken to be 100 GeV in all event samples. In practice, for the purpose of signal discovery, the neutralino mass is an unknown parameter. It will be unrealistic to have the CNNs with the same neutralino mass as the signals that we want to probe. One way[6] in discovery is to train several CNNs, each at a chosen neutralino mass, and apply those CNNs to a wide range of neutralino mass. Then, for any given neutralino mass, the CNN, which was trained on the closest neutralino mass, is able to tag the signal efficiently. The generality of the CNN, which is trained on a fixed neutralino mass, can be seen in Fig. 4. In the left panel, we show the performances of the CNN on event samples with neutralino mass in the range of [70,150] GeV, where the CNN is trained with an $m_{\tilde{\chi}_2^0} = 100$ GeV event sample only. We find that the neutralino mass varying in the range of [90,125] GeV does not reduce the sensitivity much, and the CNN is more vulnerable to lower neutralino mass. On the other hand, the CNN can be more useful if it is used in combination with jet invariant mass (CNN + M). In the right panel, the performances of the combinational CNN + M on different neutralino masses are shown. The information from the jet invariant mass helps improve the tagging efficiency a lot, especially in the light neutralino mass

[5]In practice, we find the Kolmogorov-Smirnov tests are always greater than 0.1 for both neutralino and QCD jets.

[6]One can also train a neural network on the event sample that contains events of all neutralino masses as in Ref. [60].

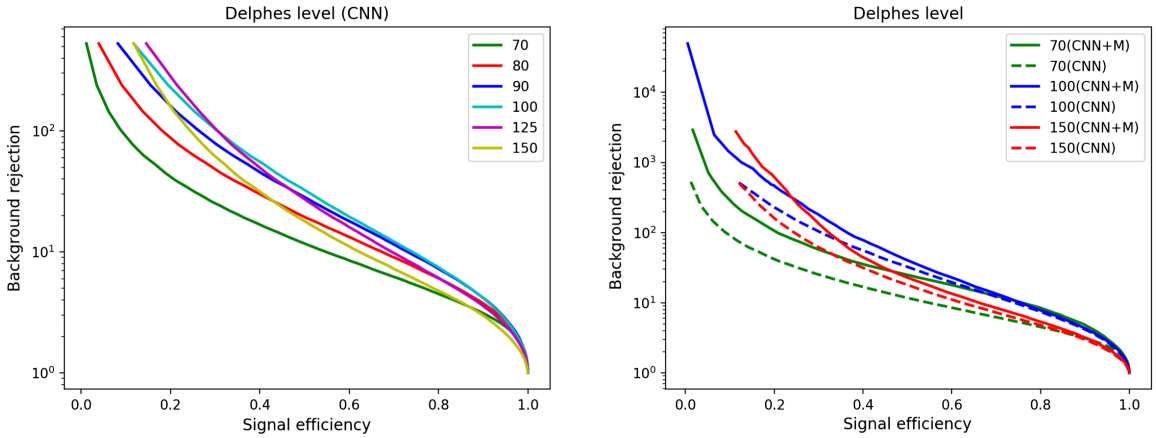

FIG. 4. Left panel: Performances of the CNN when applied to the samples with different neutralino masses. Right panel: Performances of the method that combines the information of the CNN and jet invariant mass. In both panels, the CNN is trained with the $m_{\tilde{\chi}_2^0} =$ 100 GeV event sample only. The numbers in the legend indicate the neutralino masses of the event samples.

region and compensating for the weakness of the CNN. The efficiency of the CNN + M method only mildly depends on the neutralino mass. To conclude, including jet invariant mass can not only improve the tagging efficiency but also extend the application of our CNN. They should be used together in realistic signal searches.

IV. APPLICATION TO THE LHC GLUINO SEARCH

Having shown the power and generality of our CNN method, we are ready to show its explicit application in a RPV gluino search.[7] The signal process is the gluino pair production, in which each gluino decays into two quarks and a neutralino. The neutralino decays through the hadronic RPV operator into three quarks. This signal has been searched for by the ATLAS Collaboration in Ref. [62]. For neutralino with mass ∼100 GeV, a gluino lighter than ∼1.1 TeV has been excluded. The dominant background process in the search is the QCD multijet background. In this section, we will show how the CNN helps improve the ATLAS gluino search. Before that, we need to recast the experimental analysis on both signal and background.

The QCD multijet process is simulated by the MADGRAPH5 framework at leading order.[8] According to the cuts adopted in the ATLAS analysis, we only consider the multijet processes with four or five jets at parton level, and each jet should have $p_T > 200$ GeV and $|\eta| < 2.0$. The matching of these processes is handled by the MLM method [63] in MADGRAPH5. Events with higher jet multiplicity are obtained after performing the initial state radiation and final state radiation in PYTHIA8. The signal events are generated at leading order as well, based on the benchmark points that have neutralino mass in the range of [50,200] GeV and gluino mass in the range of [1,2] TeV with step size 50 GeV.

We recast the ATLAS analysis [62] as follows. (1) For each event, large-R jets are reconstructed by the anti-k_T algorithm with radius parameter $R = 1.0$. A "trimming" process [64] with a subjet radius parameter of $R_{\text{subjet}} = 0.2$ and the minimal transverse momentum fraction of 5% is applied on each large-R jet. The resulting trimmed large-R jets are required to have $p_T > 200$ GeV and $|\eta| < 2.0$. The analysis only selects the events with at least four trimmed large-R jets ($N_{\text{jet}} \geq 4$) in which the leading one should have $p_T > 440$ GeV. (2) Meanwhile, the small-R jets of each event are reconstructed by the anti-k_T algorithm with radius parameter $R = 0.4$. They are required to have $p_T >$ 50 GeV and $|\eta| < 2.5$. These jets are used to count the number of b-tagged jets (N_b) in the final state. The b-tagging efficiency is taken to be 70% [65] with mistagging rates for the charm- and light-flavor jets of 0.15 and 0.008, respectively. (3) Two discriminative variables are defined for each event: the total jet mass variable (M_J^{Σ}) [66–69] which is the scalar sum of invariant masses of four leading trimmed large-R jets and the pseudorapidity difference between the two leading trimmed large-R jets ($|\Delta\eta_{12}|$). (4) Four signal regions are defined in Table I.

Because we are interested in the low neutralino mass region, the 4jSRb1 signal region provides the most sensitive probe. Only the signal and background events, which can pass all of the selections of the 4jSRb1 signal region, are kept for later analysis. In the simulation, the selected signal and background event numbers are guaranteed to be

[7] An attempt to improve the same search using the whole event image with the CNN was studied in Ref. [61].

[8] The higher order QCD corrections, which change the distributions of jet multiplicity, jet pseudorapidity, and jet transverse momentum, can only have indirect influences on the jet substructure, such as more contaminations between jets due to higher jet multiplicity, spread jet profile for larger pseudorapidity, and/or smaller transverse momentum. Our results are insensitive to these effects because the parton showing with PYTHIA8 includes all the leading logarithmic contributions, and our CNN is capable of tagging jets in a wide range of pseudorapidity and transverse momentum.

TABLE I. The definitions, the expected numbers of background events, and the observed event numbers of four signal regions in the ATLAS analysis [62]. Three components of background prediction uncertainty in the seventh column are statistical uncertainty, residual p_T-dependence uncertainty, and the Monte Carlo–based nonclosure uncertainty, respectively.

Signal region	N_{jet}	N_b	M_J^Σ	$\lvert\Delta\eta_{12}\rvert$	Observed	SM predicted
4jSR	≥4	⋯	>0.8 TeV	<1.4	122	$151 \pm 15 \pm 17 \pm 20$
4jSRb1	≥4	>0	>0.8 TeV	<1.4	46	$61 \pm 10 \pm 6 \pm 12$
5jSR	≥5	⋯	>0.6 TeV	<1.4	64	$51.4 \pm 7.7 \pm 7.2 \pm 6.5$
5jSRb1	≥5	>0	>0.6 TeV	<1.4	30	$18.2 \pm 4.2 \pm 2.5 \pm 3.0$

around 10 000 to suppress the statistical uncertainty. The cross section for the signal at this stage can be calculated as $\sigma^{13}(\tilde{g}\,\tilde{g}) \times \epsilon^{4\text{jSRb1}}$, where $\sigma^{13}(\tilde{g}\,\tilde{g})$ is the gluino pair production cross section at the 13 TeV LHC, which can be calculated at next-to-leading order by PROSPINO2 [70], and $\epsilon^{4\text{jSRb1}}$ is the selection efficiency of the 4jSRb1 signal region that is obtained from our recasted analysis. The background cross section (σ^{BG}) at this stage is simply estimated by the numbers in the "SM predicted" column of Table I divided by the integrated luminosity of the analysis $\mathcal{L} = 14.8\ \text{fb}^{-1}$.

Now, we can apply the CNN tag on the jets in the selected signal and background events. First, in each of the selected events, jets are reconstructed in the same way as the training sample, i.e., anti-k_t with radius parameter $R = 1.0$ and transverse momentum $p_T > 100\,\text{GeV}$. Since two neutralino jets can be either energetic or relatively soft in the signal process, all reconstructed jets are passed to our CNN for neutralino tagging. Each of them will be assigned a signal possibility (there are two outputs of the CNN: signal and background possibilities; the background possibility is correlated with the signal possibility). Then, the jets are ranked by the signal possibility. The distributions of the signal possibilities for the leading three jets are shown in Fig. 5, where the gluino mass and neutralino mass are set to 1.5 TeV and 100 GeV, respectively. We can see that the jets in the signal events obtain larger signal possibility than those in the background events. This information can help separate the signal and background further. On the other hand, it can be readily seen from the figure that even the background jets can obtain relatively high CNN scores of signal possibility. This indicates that the neutralino jet and QCD jet in the full signal and background events after the selections are much more difficult to discriminate than those in the training sample. The difficulty is mainly attributed to the severe contaminations among jets in the selected events. As will be demonstrated later, these contaminations tend to make a multiprong QCD jet, which also reduces the discriminating power of the N-subjettiness variable. The dashed lines in the same figure correspond to the CNN tagging efficiencies on jets after preforming the jet trimming, with trimming parameters the same as in the ATLAS analysis [62]. Because of the hardness of the contamination, the trimming fails to resolve the jets.

Because the CNN scores (signal possibility) for the leading three jets of the signal and background are correlated to some extent, we employ again the BDT method to study the discriminating power of the combination of this information (including jet invariant mass and N subjettiness). Compared to the BDT analysis in the previous section, we now have fewer available events (∼10 000) and more input variables (CNN scores and invariant masses of the three leading jets). However, the same BDT parameters turn out to perform quite well here. At each mass point, this BDT is trained on 5000 signal and 5000 background events, and it is tested on the rest of the

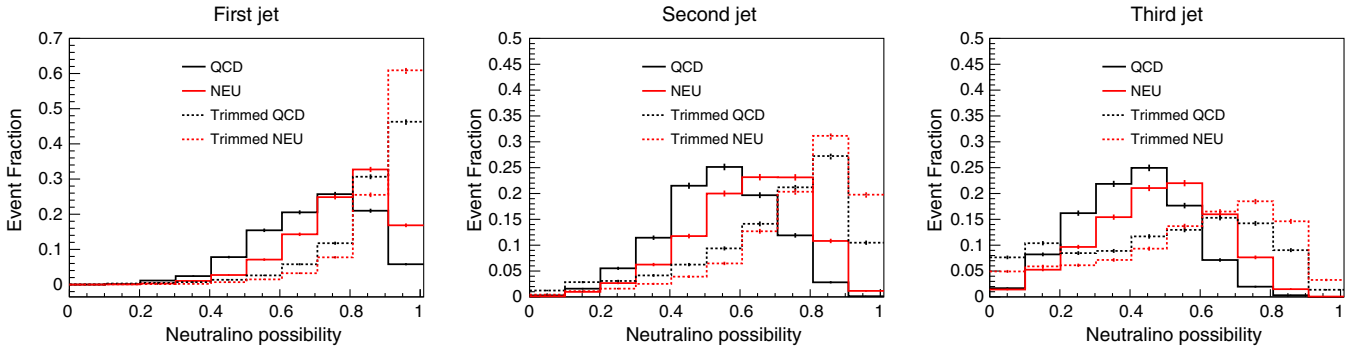

FIG. 5. The signal possibility of the leading three jets with (dashed line) and without (solid line) the trimming procedure in the selected signal (red line) and background (black line) events. The gluino mass and neutralino mass for the signal process are taken to be 1.5 TeV and 100 GeV for illustration.

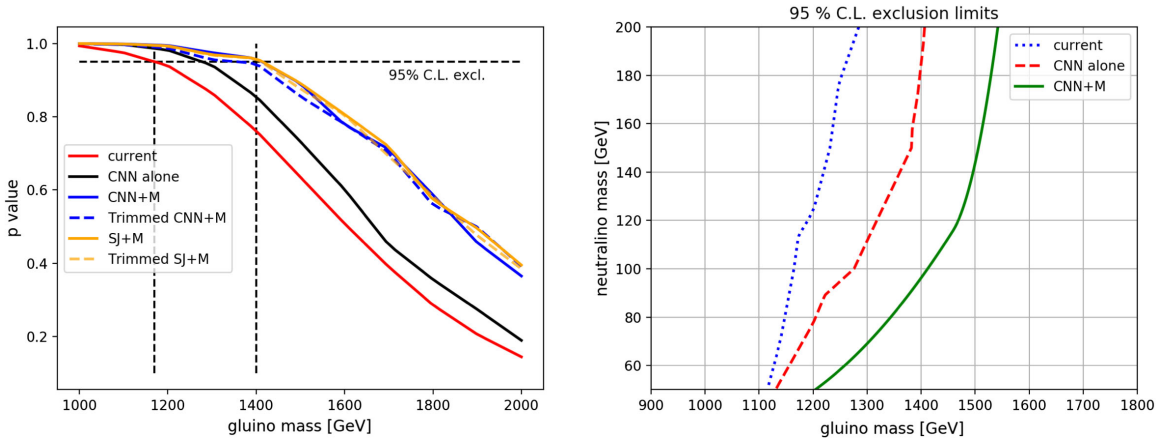

FIG. 6. Left: The p values for the original ATLAS analysis (red solid line) and BDT analyses with either CNN alone (dark solid line) or with various additional information as indicated in the legend. Right: The 95% C.L. exclusion limits for the original ATLAS analysis (blue dotted line), CNN alone analysis (red dashed line), and CNN + M analysis (green solid line).

independent events. The Kolmogorov-Smirnov tests are found to be always greater than ∼0.1 for both the signal and background, which indicates that the BDT is free from the overtraining problem. For each BDT trained and validated on the given gluino and neutralino masses, applying a cut on its BDT response will reduce the signal and background cross section further down to $\sigma^{13}(\tilde{g}\,\tilde{g}) \times \epsilon^{4\text{jSRb1}} \times \epsilon_S^{\text{BDT}}$ and $\sigma^{\text{BG}} \times \epsilon_B^{\text{BDT}}$, respectively. The $\epsilon_{S/B}^{\text{BDT}}$ corresponds to the selection efficiency of a BDT cut on signal/background events. We assume the observed event number is reduced by the same factor of ϵ_B^{BDT}. As for the background uncertainties, the statistical component is rescaled by a factor of $\sqrt{\epsilon_B^{\text{BDT}}}$, while others are rescaled by a factor of ϵ_B^{BDT}. We will adopt the p value [71] to characterize the probability of the signal exclusion, which is defined as

$$p = 1 - \frac{\mathcal{P}(\mathcal{H}_{S+B})}{\mathcal{P}(\mathcal{H}_B)}, \tag{4.1}$$

where $\mathcal{P}(\mathcal{H}_{S+B})$ and $\mathcal{P}(\mathcal{H}_B)$ are the probabilities of signal plus background hypothesis and background only hypothesis, respectively. So, the $\mathcal{H}_{S+B}$ hypothesis is excluded at 95% C.L. if the p value is greater than 0.95. The BDT cut that maximizes the p value will be taken at each gluino-neutralino mass point in each analysis.[9] The p values for the original ATLAS analysis and our BDT analyses with either the CNN output alone or with combined information are shown in the left panel of Fig. 6, where we have fixed $m_{\tilde{\chi}_1^0} = 100$ GeV. Our recasting of the ATLAS analysis shows that the benchmark points with gluino mass below ∼1.18 TeV can be excluded at 95% C.L., which coincides with the experimental result. Including the information of the CNN output alone will push the lower bound of the gluino mass to ∼1.3 TeV. By adding the jet invariant mass into the BDT, the non-observation of any excess will exclude the gluino mass lighter than ∼1.4 TeV. For comparison, we have also shown the p values for the analysis with the information of N subjettiness and jet invariant mass, which does not perform better than the CNN + M analysis. Furthermore, the dashed lines correspond to the p values of the analyses in which the jets are trimmed before performing the tagging. It turns out the trimming procedure does not help improve the signal and background discrimination.

We have demonstrated that the CNN + M method (without trimming) provides one of the most sensitive probes for the RPV gluino search. Finally, we show the application of the method (with the CNN trained on the $m_{\tilde{\chi}_1^0} = 100$ GeV events sample) to the two-dimensional $m_{\tilde{g}} - m_{\tilde{\chi}_1^0}$ plane. In the right panel of Fig. 6, the 95% C.L. exclusion limits for the original ATLAS analysis, the CNN alone analysis, and CNN + M analysis are given. Here, the CNN is trained on the event sample with $m_{\tilde{\chi}_1^0} = 100$ GeV. We can observe that such a CNN is vulnerable to lower neutralino mass; i.e., the improvement is dramatically decreased for the neutralino mass less than 100 GeV, while it is much less sensitive to the higher neutralino mass. This is mainly because the neutralino with mass $m_{\tilde{\chi}_1^0} \lesssim 200$ GeV from heavy gluino decay has transverse momentum larger than ∼400 GeV. All its decay products are captured by the jet reconstruction. So, the jet substructure is detectable except when the neutralino is so light that its subjets become overlapping. Including the jet invariant mass information can help push the gluino bounds by ∼100 GeV higher. In particular, the jet invariant mass has better discriminating power for heavier neutralino

[9]The optimized BDT cut efficiencies are found to be ∼50% for the signal and 10%–20% for the background. So, the signal and background event numbers in our simulation after the BDT cut are around 5000 and 1000–2000, respectively.

mass, compensating for the slight decrement of neutralino jet tagging efficiency, which can also be seen in Fig. 4.

V. CONCLUSION

In the paper, we study the possible improvement of the current hadronic RPV search by the BDT method with information from the jet substructure. In particular, the convolutional neutral network is adopted to tag the neutralino jet which decays into three quarks. The application of the CNN to an existing RPV gluino search by the ATLAS Collaboration in the final state with multiple energetic jets is investigated.

The information of a jet can be formatted into a jet image by identifying each calorimeter cell as a pixel. The energy distribution of all particles, the energy distribution of charged particles, and the number of charged particles in calorimeter cells are regarded as the RGB color of those pixels. The CNN is trained on events of a simplified process with only a visible neutralino jet in the final state. So, it is able to tag a neutralino jet by using the jet image, irrespective of its production mechanism. According to the small size and sparsity of the jet image, the `VGGNet` CNN architecture, which is optimized for the `CIFAR-10` data set, is adopted. It is able to tag the neutralino jet with efficiency of 50% while only accepting ∼1% of the QCD jet. These efficiencies are found to be insensitive to the CNN parameters in a wide range. Moreover, due to the cylinder shape of the detector, the jet image has strong dependence on the pseudorapidity of the jet. The CNN performs well for the jet either in the central region ($|\eta| \to 0$) or with relatively large pseudorapidity ($|\eta| \lesssim 2.5$). Our CNN can outperform the high-level jet substructure variable N subjettiness by a factor of a few in the neutralino jet and QCD jet discrimination. However, the jet invariant mass information is not fully learned by the CNN, partly because the image preprocessing does not respect the Lorentz symmetry. Combining the CNN output with the jet invariant mass can improve the signal efficiency further. More importantly, for the CNN being trained on a given neutralino mass, the CNN + M tagging method performs much better than the method with the CNN alone when applied to the neighbor of that neutralino mass.

To study the realistic application of the CNN, the ATLAS analysis is recast. Only the events (for both signal and background) which can pass all selection cuts of the 4jSRb1 signal region in the ATLAS analysis are kept. The CNN assigns "neutralino jet possibilities" to all jets in these events. The jets in the signal events are likely to obtain higher neutralino jet possibilities than those in the background events. Compared to the simplified processes (for generating a training sample) with a single target jet in the final state, the heavy contaminations due to multiple energetic jets in the final state greatly reduce the discriminating power of both the CNN and N subjettiness, but the BDT analysis with information from the CNN scores of three leading jets is still able to push the lower bound of the gluino mass by ∼100 GeV. The combined analyses of either CNN + M or N subjettiness + M have similar sensitivities, i.e., excluding the gluino mass lighter than ∼1.4 TeV for $m_{\tilde{\chi}_1^0} = 100$ GeV. By applying the CNN and CNN + M analyses to the two-dimensional $m_{\tilde{g}} - m_{\tilde{\chi}_1^0}$ plane, we find the CNN tagging efficiency is vulnerable to a lighter neutralino while it is insensitive to a heavier neutralino up to ∼200 GeV. The CNN + M method can help push the gluino bounds by 100–250 GeV higher depending on the neutralino mass.

ACKNOWLEDGMENTS

We thank the Korea Institute for Advanced Study for providing computing resources (KIAS Center for Advanced Computation Linux Cluster System) for this work. This research was supported in part by Projects No. 11475238, No. 11747601, and No. 11875062 supported by National Natural Science Foundation of China, and by the Key Research Program of Frontier Science, CAS (T. L.) and by National Research Foundation of Korea Research Grant No. NRF-2015R1A2A1A05001869 (J. L.).

[1] H. P. Nilles, Supersymmetry, supergravity and particle physics, Phys. Rep. **110**, 1 (1984).
[2] H. E. Haber and G. L. Kane, The search for supersymmetry: Probing physics beyond the standard model, Phys. Rep. **117**, 75 (1985).
[3] ATLAS Collaboration, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResults.
[4] CMS Collaboration, https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS.
[5] L. J. Hall and M. Suzuki, Explicit R-parity breaking in supersymmetric models, Nucl. Phys. **B231**, 419 (1984).
[6] G. Jungman, M. Kamionkowski, and K. Griest, Supersymmetric dark matter, Phys. Rep. **267**, 195 (1996).
[7] M. Aaboud *et al.* (ATLAS Collaboration), Search for squarks and gluinos in final states with jets and missing transverse momentum using 36 fb^{-1} of $\sqrt{s} = 13$ TeV pp collision data with the ATLAS detector, Phys. Rev. D **97**, 112001 (2018).
[8] A. M. Sirunyan *et al.* (CMS Collaboration), Search for natural and split supersymmetry in proton-proton collisions at $\sqrt{s} = 13$ TeV in final states with jets and missing transverse momentum, J. High Energy Phys. 05 (2018) 025.

[9] B. C. Allanach and B. Gripaios, Hide and seek with natural supersymmetry at the LHC, J. High Energy Phys. 05 (2012) 062.
[10] G. Durieux and C. Smith, The same-sign top signature of R-parity violation, J. High Energy Phys. 10 (2013) 068.
[11] B. Bhattacherjee, J. L. Evans, M. Ibe, S. Matsumoto, and T. T. Yanagida, Natural supersymmetry's last hope: R-parity violation via UDD operators, Phys. Rev. D **87**, 115002 (2013).
[12] S. Diglio, L. Feligioni, and G. Moultaka, Stashing the stops in multijet events at the LHC, Phys. Rev. D **96**, 055032 (2017).
[13] M. R. Buckley, D. Feld, S. Macaluso, A. Monteux, and D. Shih, Cornering natural SUSY at LHC run II and beyond, J. High Energy Phys. 08 (2017) 115.
[14] J. A. Evans and D. Mckeen, The light gluino gap, arXiv:1803.01880.
[15] J. Li, T. Li, and W. Zhang, The least constrained supersymmetry with R-parity violation, arXiv:1805.06172.
[16] G. P. Salam, Towards jetography, Eur. Phys. J. C **67**, 637 (2010).
[17] A. Abdesselam *et al.*, Boosted objects: A probe of beyond the Standard Model physics, Eur. Phys. J. C **71**, 1661 (2011).
[18] A. Altheimer *et al.*, Jet substructure at the Tevatron and LHC: New results, new tools, new benchmarks, J. Phys. G **39**, 063001 (2012).
[19] A. Altheimer *et al.*, Boosted objects and jet substructure at the LHC, Eur. Phys. J. C **74**, 2792 (2014).
[20] D. Adams *et al.*, Towards an understanding of the correlations in jet substructure, Eur. Phys. J. C **75**, 409 (2015).
[21] A. J. Larkoski, I. Moult, and B. Nachman, Jet substructure at the large hadron collider: A review of recent advances in theory and machine learning, arXiv:1709.04464.
[22] J. M. Butterworth, J. R. Ellis, A. R. Raklev, and G. P. Salam, Discovering Baryon-Number Violating Neutralino Decays at the LHC, Phys. Rev. Lett. **103**, 241803 (2009).
[23] Y. Bai, A. Katz, and B. Tweedie, Pulling out all the stops: Searching for RPV SUSY with stop-jets, J. High Energy Phys. 01 (2014) 040.
[24] B. Bhattacherjee and A. Chakraborty, Study of the baryonic R-parity violating MSSM using the jet substructure technique at the 14 TeV LHC, Phys. Rev. D **89**, 115016 (2014).
[25] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, Jet Substructure as a New Higgs Search Channel at the LHC, Phys. Rev. Lett. **100**, 242001 (2008).
[26] J. Thaler and K. Van Tilburg, Identifying boosted objects with N-subjettiness, J. High Energy Phys. 03 (2011) 015.
[27] J. Cogan, M. Kagan, E. Strauss, and A. Schwartzman, Jet-images: Computer vision inspired techniques for jet tagging, J. High Energy Phys. 02 (2015) 118.
[28] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, Jet-images–Deep learning edition, J. High Energy Phys. 07 (2016) 069.
[29] P. Baldi, K. Bauer, C. Eng, P. Sadowski, and D. Whiteson, Jet substructure classification in high-energy physics with deep neural networks, Phys. Rev. D **93**, 094034 (2016).
[30] K. Datta and A. Larkoski, How much information is in a jet?, J. High Energy Phys. 06 (2017) 073.
[31] L. G. Almeida, M. Backović, M. Cliche, S. J. Lee, and M. Perelstein, Playing tag with ANN: Boosted top identification with pattern recognition, J. High Energy Phys. 07 (2015) 086.
[32] G. Kasieczka, T. Plehn, M. Russell, and T. Schell, Deep-learning top taggers or the end of QCD?, J. High Energy Phys. 05 (2017) 006.
[33] S. Macaluso and D. Shih, Pulling out all the tops with computer vision and deep learning, arXiv:1803.00107.
[34] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz, Deep learning in color: Towards automated quark/gluon jet discrimination, J. High Energy Phys. 01 (2017) 110.
[35] H. Luo, M.-x. Luo, K. Wang, T. Xu, and G. Zhu, Quark jet versus gluon jet: Deep neural networks with high-level features, arXiv:1712.03634.
[36] T. Cohen, M. Freytsis, and B. Ostdiek, (Machine) learning to do more with less, J. High Energy Phys. 02 (2018) 034.
[37] E. M. Metodiev, B. Nachman, and J. Thaler, Classification without labels: Learning from mixed samples in high energy physics, J. High Energy Phys. 10 (2017) 174.
[38] K. Datta and A. J. Larkoski, Novel jet observables from machine learning, J. High Energy Phys. 03 (2018) 086.
[39] G. Louppe, K. Cho, C. Becot, and K. Cranmer, QCD-aware recursive neural networks for jet physics, arXiv:1702.00748.
[40] T. Cheng, Recursive neural networks in quark/gluon tagging, Comput. Software Big Sci. **2**, 3 (2018).
[41] J. Pearkes, W. Fedorko, A. Lister, and C. Gay, Jet constituents for deep neural network based top quark tagging, arXiv:1704.02124.
[42] A. Butter, G. Kasieczka, T. Plehn, and M. Russell, Deep-learned top tagging with a Lorentz layer, SciPost Phys. **5**, 028 (2018).
[43] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, Boosted decision trees, an alternative to artificial neural networks, Nucl. Instrum. Methods Phys. Res., Sect. A **543**, 577 (2005).
[44] M. Nielsen, Neural networks and deep learning, http://neuralnetworksanddeeplearning.com/.
[45] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016).
[46] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv:1409.1556.
[47] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, arXiv:1512.03385.
[48] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, The Pascal visual object classes (VOC) challenge, Int. J. Comput. Vis. **88**, 303 (2010).
[49] A. Krizhevsky, Learning multiple layers of features from tiny images, University of Toronto Technical report, 2009.
[50] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, J. High Energy Phys. 07 (2014) 079.
[51] T. Sjostrand, S. Mrenna, and P. Z. Skands, PYTHIA 6.4 physics and manual, J. High Energy Phys. 05 (2006) 026.
[52] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi (DELPHES 3 Collaboration), DELPHES 3, A modular framework for fast

simulation of a generic collider experiment, J. High Energy Phys. 02 (2014) 057.
[53] M. Cacciari, G. P. Salam, and G. Soyez, FASTJET user manual, Eur. Phys. J. C **72**, 1896 (2012).
[54] F. Chollet, https://github.com/fchollet/keras.
[55] M. Cacciari, G. P. Salam, and G. Soyez, The anti-$k(t)$ jet clustering algorithm, J. High Energy Phys. 04 (2008) 063.
[56] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.
[57] A. Hocker *et al.*, TMVA—Toolkit for multivariate data analysis, Proc. Sci. ACAT2007 (**2007**) 040 [arXiv:physics/0703039].
[58] I. M. Chakravarty, R. G. Laha, and J. D. Roy, *Handbook of Methods of Applied Statistics* (McGraw-Hill, New York, 1967).
[59] S. Chang, T. Cohen, and B. Ostdiek, What is the machine learning?, Phys. Rev. D **97**, 056009 (2018).
[60] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski, and D. Whiteson, Parameterized neural networks for high-energy physics, Eur. Phys. J. C **76**, 235 (2016).
[61] W. Bhimji, S. A. Farrell, T. Kurth, M. Paganini, Prabhat, and E. Racah, Deep neural networks for physics analysis on low-level whole-detector data at the LHC, in *Proceedings of 18th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT 2017) Seattle, WA, USA, 2017*, http://stacks.iop.org/1742-6596/1085/i=4/a=042034.
[62] ATLAS Collaboration, Search for massive supersymmetric particles in multi-jet final states produced in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC, CERN Technical Report No. ATLAS-CONF-2016-057, 2016.
[63] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, Matching matrix elements and shower evolution for top-quark production in hadronic collisions, J. High Energy Phys. 01 (2007) 013.
[64] D. Krohn, J. Thaler, and L.-T. Wang, Jet trimming, J. High Energy Phys. 02 (2010) 084.
[65] ATLAS Collaboration, Optimisation of the ATLAS b-tagging performance for the 2016 LHC run, CERN Technical Report No. ATL-PHYS-PUB-2016-012, 2016.
[66] A. Hook, E. Izaguirre, M. Lisanti, and J. G. Wacker, High multiplicity searches at the LHC using jet masses, Phys. Rev. D **85**, 055029 (2012).
[67] S. El Hedri, A. Hook, M. Jankowiak, and J. G. Wacker, Learning how to count: A high multiplicity search for the LHC, J. High Energy Phys. 08 (2013) 136.
[68] T. Cohen, M. Jankowiak, M. Lisanti, H. K. Lou, and J. G. Wacker, Jet substructure templates: Data-driven QCD backgrounds for fat jet searches, J. High Energy Phys. 05 (2014) 005.
[69] G. Aad *et al.* (ATLAS Collaboration), Search for massive supersymmetric particles decaying to many jets using the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV, Phys. Rev. D **91**, 112016 (2015).
[70] W. Beenakker, R. Hopker, and M. Spira, PROSPINO: A program for the production of supersymmetric particles in next-to-leading order QCD, arXiv:hep-ph/9611232.
[71] A. L. Read, Presentation of search results: The CL(s) technique, J. Phys. G **28**, 2693 (2002).