

Boosted Higgs boson jet reconstruction via a graph neural network

Jun Guo,¹ Jinmian Li^{2,*} Tianjun Li,^{3,4} and Rao Zhang²

¹Department of Physics, Jiangxi Normal University, Nanchang 330022, China

²College of Physics, Sichuan University, Chengdu 610065, China

³CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China

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



(Received 16 October 2020; accepted 8 June 2021; published 28 June 2021)

By representing each collider event as a point cloud, we adopt the graphic convolutional network (GCN) with focal loss to reconstruct the Higgs jet in it. This method provides higher Higgs tagging efficiency and better reconstruction accuracy than the traditional methods, which use jet substructure information. The GCN, which is trained on events of the $H + \text{jets}$ process, is capable of detecting a Higgs jet in events of several different processes, even though the performance degrades when there are boosted heavy particles other than the Higgs boson in the event. We also demonstrate the signal and background discrimination capacity of the GCN by applying it to the $t\bar{t}$ process. Taking the outputs of the network as new features to complement the traditional jet substructure variables, the $t\bar{t}$ events can be separated further from the $H + \text{jets}$ events.

DOI: 10.1103/PhysRevD.103.116025

I. INTRODUCTION

The precision measurements [1,2] of the Higgs boson couplings at the LHC are compatible with the Standard Model (SM) predictions. However, there are several evidences, such as dark matter observation and neutrino oscillation, indicating new physics beyond the SM (BSM). In many of the SM extensions, the Higgs boson is a portal to BSM sector. Measuring the Higgs boson precisely at colliders is crucial for probing new physics.

The decay channel of $H \rightarrow b\bar{b}$ has not been probed with promising precision in spite of its large branching ratio, mainly due to the large QCD backgrounds. The jet substructure technique has been found to be efficient in tagging this channel when the Higgs boson is Lorentz boosted. Various jet substructure tools and techniques [3–8] have been developed since the first proposal in Ref. [9]. In those methods, jets are reconstructed from sequential recombination algorithms, such as the anti- k_T algorithm [10], the Cambridge/Aachen algorithm [11], and so on. Based on the distribution of jet constituents, several jet substructure variables are built to characterize the jet. This procedure

involves a cone size parameter R in jet clustering, which should be adjusted for different searches, depending on the detector parameters and the target jet properties. As a result, hadrons with different origins may be clustered in a same jet if R is too large, or those with the same origin may be misclustered into different jets if R is too small. Moreover, in most jet substructure analyses, a large R is chosen in order to fully capture the jet constituents during the jet clustering. It means that the jet is heavily contaminated by pileup events. A jet grooming method [9,12,13] should be applied to mitigate the effects of pileup events.

Machine learning techniques have proven to be efficient in jet tagging and jet reconstruction (see Refs. [14–16] for reviews). They can outperform the jet substructure method in tagging the Higgs jet [17–21]. In Ref. [22], we take the event information as a digital image and adopt the mask R-CNN framework to detect/reconstruct the Higgs jet in the event image. The method achieves higher efficiency of Higgs jet detection and higher accuracy of Higgs momentum reconstruction than the traditional jet clustering and substructure tagging methods. In this work, we choose the graph neural network (GNN) for Higgs jet reconstruction. Comparing to the imaged based methods, the GNN can be implemented in a more efficient way. The GNN methods classify final state hadrons with particular relations to form jets while the image based methods, such as the mask R-CNN, consist of two stages: image classification and region proposal. Despite its simpleness, the output of the GNN is more informative, i.e., the output of GNN is particlewise. On the other hand, the Mask R-CNN only

*jmli@scu.edu.cn

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

predicts continuous regions on the event image. All particles in a masked region are supposed to have the same origin. There are two drawbacks: (1) For charged particles, the tracker has better angular resolution than the electromagnetic/hadronic calorimeter. In experiments, a charged particle can be distinguished from other charged/neutral particles even when they are flying along a similar direction. This feature can be easily encoded in GNN. However, the granulation in the mask R-CNN method is usually based on the cell size of calorimeters. All particles hitting the same calorimeter cell will be treated as a whole. (2) In the mask R-CNN method, a calorimeter cell that contains only the particles from the pileup events but is surrounded by the Higgs constituents will also be assigned to the Higgs jet. The GNN does not have this tendency so that the pileup subtraction is less urgent.

The GNNs have been applied to a growing variety of scenarios including social networks, knowledge graphs, recommender systems, 3D shape analysis as well as the data analysis in high energy physics [23]. They have been used for pileup mitigation at hadron colliders [24], tracker reconstruction [25,26], jet identification [19,27–32], and event classification [33,34]. In particular, studies in Refs. [26,35] demonstrate that the GNN is capable of labeling constituents of a specific jet after supervised training. However, the effects from pileup events are not considered. They only need to label $\sim 10\%$ of total particles. In practice, with an average number of $\langle n \rangle = 50$ pileup interactions, the total number of particles for an event can reach $\mathcal{O}(10^4)$, whereas only $\lesssim \mathcal{O}(10^2)$ particles are labeled. This renders very low efficiency in training the GNN.

In this work, we adopt the dynamic graph convolutional network (GCN) [36] to reconstruct a boosted Higgs jet in any events, taking into account the pileup effects. The GCN is trained on the events of the $H + \text{jets}$ process [with $p_T(H) > 200$ GeV] overlaid with $\langle n \rangle = 50$ pileup events. It is designed with focal loss [37], in order to train with the imbalanced dataset; *i.e.*, the number of unlabeled particles is around 2 orders of magnitude larger than that of the labeled ones. We will demonstrate that the GCN is capable of detecting Higgs jet in events of several other processes efficiently, provided that there are no boosted heavy particles other than the Higgs boson in the event. Moreover, we will show that the GCN method has improved background discrimination power compared with the traditional jet substructure method.

The paper is organized as follows. In Sec. II, we introduce how the event samples are generated and preprocessed. Some concepts related to the network are briefly introduced in Sec. III. The performance of the network is discussed in Sec. IV. We provide a conclusion and outlook in Sec. V.

II. EVENT SIMULATION AND DATA PREPARATION

Events in the analysis are simulated within the MG5_aMC@NLO [38] framework. PYTHIA8 [39] is used

to perform the parton shower, hadronization, and hadrons decay. The events of $H + \text{jets}$ with the Higgs boson decaying into a pair of b quark are used for training and validating our GCN. The Higgs boson is required to have transverse momentum $p_T > 200$ GeV. Moreover, there are multiple proton-proton collisions (referred to as pileup), during a bunch crossing period at the LHC. We adopt the A3 tune of PYTHIA8 with phenomenological parameters provided in Refs. [40,41] to simulate pileup events. In preparation of event samples, the final states of each event of a hard process are overlaid with final states of $\langle \mu \rangle = 50$ (average number with poisson distribution) pileup events. This leads to dramatically increased difficulty in detecting the Higgs jet constituents with the GCN.

With the contamination of pileup events, the total number of final state particles for each Higgs production event can reach $\mathcal{O}(10^4)$. However, we find that most of the Higgs constituents have relatively large transverse momenta. In order to reduce the complexity of the algorithm, we only consider the leading 1024 (or 2048) particles with the highest p_T for each event. For the $H + \text{jets}$ process with $p_T(H) > 200$ GeV, the lowest p_T of a selected particle is ~ 0.9 GeV (~ 0.6 GeV for selecting 2048 particles) for most of the events. So the selection can also help to alleviate the infrared effects of events. In the left panel of Fig. 1, we show the invariant mass distributions for the vector sum of the selected Higgs constituents momenta ($m_{\text{inv}}^{\text{sum}}$). We find a certain distortion (deduction) in the invariant mass distribution due to our selection. The distortion is slightly less severe in the case with 2048 particles. We will show later that such an amount of distortion is acceptable in the sense that the GCN can perform better than the traditional jet substructure method. After the selection, the Higgs constituents still take only a small fraction of total particles. In the right panel of Fig. 1, the distributions of the ratio between the number of the unlabeled particles (particles not belonging to a Higgs jet) and that of the labeled ones (Higgs constituents) are given. Because the typical number of Higgs constituents is 30–50, the ratio is ~ 25 (50) in the case of selecting 1024 (2048) particles.

The GCN takes the pseudorapidity (η), azimuthal angle (ϕ), and transverse momentum (p_T) of all particles in an event as the input features. We have also tried to feed the particle type (lepton, photon, or hadron) and particle electric charge to the GCN without finding much improvements.¹

¹The GCN is sensitive to the collinear splittings of input momenta. However, infrared effects of the hadronic level events are already alleviated in Monte Carlo simulation, since the very soft and collinear gluons in parton emission are not resolved in an infrared safe fragmentation framework, such as the string model [42] in PYTHIA8. For example, the MC simulations of the charged-particle multiplicities at the LHC are found to match the data well [43]. (The charged particles are required to have $p_T \gtrsim 500$ MeV to improve the infrared stability.) And the number of charged particles in a jet is also used to distinguish quark and gluon jet at experiments [44].

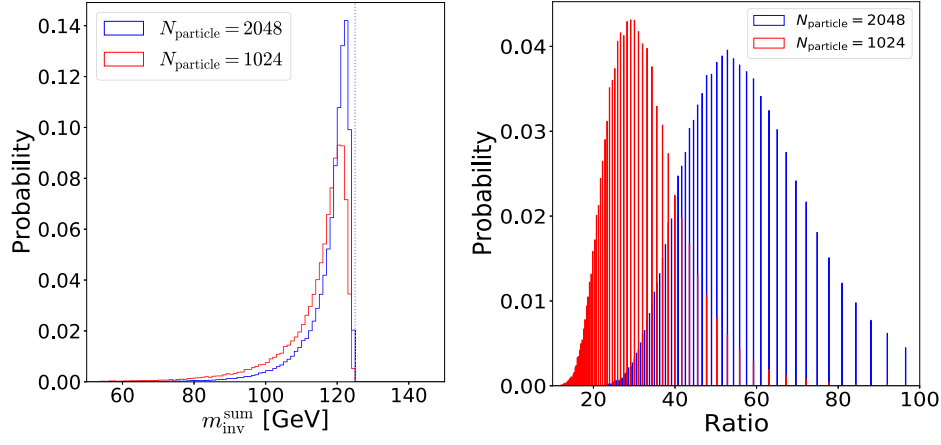


FIG. 1. Left: the invariant mass distributions for the vector sum of the selected Higgs constituents in two cases. Right: the ratio between the number of the unlabeled particles and that of the labeled ones. Events used here correspond to the $H + \text{jets}$ process with $p_T(H) > 200$ GeV overlaid with pileup contaminations.

Because the detector is a cylinder, the azimuthal angle $[\phi, \text{ in the range of } [0, 2\pi]]$ is periodic. We find the GCN gains the best detection efficiency with the definition of ϕ such that the Higgs constituents do not distribute across the line of $\phi = 0$. In the training sample, this condition is fulfilled by applying two preprocessing procedures: (1) The Higgs boson, whose azimuthal angle is outside the range of $[\pi/2, 3\pi/2]$, is shifted by $\pm\pi$ in the ϕ coordinate to keep $\phi(H) \in [\pi/2, 3\pi/2]$. (2) The Higgs constituents, whose angular distances ($\Delta R = \sqrt{\Delta y^2 + \Delta \phi^2}$, y is the rapidity) to the Higgs boson are larger than $\pi/2$, are not labeled. While during the testing, one can not obtain the angular position of the Higgs boson in advance. This condition can be fulfilled with the ϕ definition such that those high p_T particles of an event are away from the line of the $\phi = 0$. In addition, the training sample is purified by dropping events with $m_{\text{inv}}^{\text{sum}}$ smaller than 115 GeV, in order to not confuse the network with events that do not include any featured Higgs jets. (All events are used in testing sample.)

Events of several processes different from the $H + \text{jets}$ are also simulated to test the generality of the network: 1) two Higgs bosons plus three QCD jets; 2) one Higgs boson plus top quark pair; 3) a hypothetical supersymmetric (SUSY) model process, $pp \rightarrow \tilde{t}_1^* \tilde{t}_1 \rightarrow \tilde{t}_1^* \tilde{t}_1 \tilde{\chi}_1^0 \tilde{\chi}_2^0 \rightarrow \tilde{t}_1^* \tilde{t}_1 \tilde{\chi}_1^0 \tilde{\chi}_1^0$, with $m_{\tilde{\chi}_1^0} = 100$ GeV, $m_{\tilde{\chi}_2^0} = 800$ GeV, and $m_{\tilde{t}_1} = 1$ TeV. The Higgs bosons in the first two processes are forced to be boosted ($p_T > 200$ GeV) and decayed into two b quarks. We note that each event of all those processes is overlaid with average number of 50 pileup events.

III. GRAPH NEURAL NETWORK AND FOCAL LOSS

The performance of the network in jet detection depends on both the data representation and the network architecture. In this section, we will give a brief introduction to the setups of our GCN and explain how focal loss works.

A. Event representation and graph neural network

An event at the LHC corresponds to a collection of detected particles, with each particle assigned with a four momentum (η, ϕ, p_T, E) and an identity.² The point cloud, which provides a flexible representation for collection of points in two- or three-dimensional space, can be naturally used for representing an event. Comparing to the image representation of collider events, which requires a uniform granulation, the different angular resolutions for different types of particles are easier to implement in the point cloud representation.

Treating the point cloud as a graph, GCN is an efficient network to analyze it. A graph consists of nodes and edges. For collider events, nodes correspond to particles, and edges are the connections between particles. Our GCN uses the connections of the K -nearest neighbors (KNN) for each node in the feature space, with the squared distance defined as the sum of squared differences in all coordinates. In the first edge-convolution (EdgeConv) block, it is

$$D(x_i, x_j) = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}. \quad (3.1)$$

As for the second and third EdgeConv blocks, the feature vectors learned by previous block are viewed as new coordinates of the nodes in the latent space. The neighbors of each node will be dynamically updated with EdgeConv operations.

²Only long-lived particles are detected. Electrons, muons, and photons can be identified directly. Long-lived charged hadrons (pion, kaon, and proton) can be distinguished through Cherenkov effect, ionization energy loss, and so on, while neutral hadrons are not distinguished from each other. However, the particle identities do not provide improved performance in our study. They will not be used as input to the GCN.

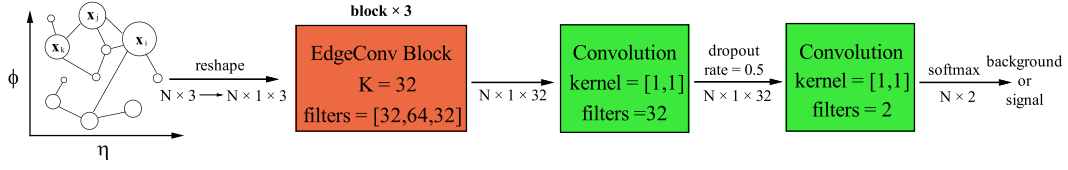


FIG. 2. Architecture of our GCN, the edge-convolution block is the same as the one proposed in Ref. [28].

In the edge-convolution blocks, the edge feature is calculated by [28,36]

$$e_{ijm} = \text{ReLU}(\theta_m \cdot ((\mathbf{x}_j - \mathbf{x}_i) \oplus \mathbf{x}_i)), \quad (3.2)$$

where the index i runs over all particles in an event (which is 1024 in our case), the index j runs over the K neighbors of the particle i , and the index m corresponds to the number of convolutional filter. The direct sum indicates that the feature vector of the i th particle \mathbf{x}_i (with the dimension C) being concatenated with the vector difference between the feature vectors of particle i and its j th neighbor. θ_m is a $2C$ -dimensional vector, which should be adjusted during the training. The inner product between θ_m and the concatenated feature vector is delivered to the rectified linear unit (ReLU) activation function to give the final expression. An aggregation function is applied after obtaining the edge features,

$$x'_{im} = \max_j e_{ijm}. \quad (3.3)$$

The new feature x'_{im} will be passed to the next edge-convolution block.

The architecture of our GCN is illustrated in Fig. 2. The input coordinates are η and ϕ , and the input features are η , ϕ , and p_T . It contains three EdgeConv blocks, with a universal $K = 32$. Each EdgeConv block is implemented as three multilayer perceptrons, which is the same as the one proposed in Ref. [28]. The number of channels (filters) of these perceptrons are 32, 64, and 32, respectively. Two convolutional layers with kernel size $[1, 1]$ are used for classification task. The first convolution layer contains 32 kernels, and the second contains 2 kernels. To prevent overfitting, a dropout layer with drop rate of 0.25 is inserted. The final classification scores for input particles are assigned by the softmax function.

Note that usually a spatial transformer network (STN) [45] is inserted in front of edge-convolution block to make the GCN invariant under translation, scale, and rotation of input data. However, in our case, the coordinates η and ϕ are not fully interchangeable; *i.e.*, they have different distributions for given processes. We find including the STN module in the GCN does not improve the Higgs jet detection efficiency, as long as the training sample is sufficient. So the STN module is not adopted in our GCN. Moreover, by taking the η , ϕ , and p_T as input features for each particle (node), our GCN is sensitive to the

position and the orientation of the graph, and each event can be associated with a graph without ambiguity.

B. Focal loss

Focal loss was firstly introduced in Ref. [46], which is especially useful in an object detection task. It is used for addressing the large class (signal and background) imbalance issue and makes the model focus on signals that have much smaller number than backgrounds. It is a modification of cross entropy, given as

$$FL(p, y) = -\alpha y(1-p)^\gamma \log(p) - (1-\alpha)(1-y)p^\gamma \log(1-p), \quad (3.4)$$

where p is the probability of being positive class, and y is the label which will be given beforehand in the supervised learning. Compared with cross entropy, the hyper parameter α is used to keep a balance between positive and negative samples, and $(1-p)^\gamma$ and p^γ terms suppress the loss contribution of well classified samples so that loss function is dominated by the difficult ones. For our case, given the ratio between the number of unlabeled particles and number of labeled ones around 30–50, we find setting $\alpha = 0.65(0.75)$, $\gamma = 1$ for $N_{\text{particle}} = 1024(2048)$ samples works well.

IV. RESULT

Our GCN model is implemented with TensorFlow [47]. The model is trained on one million $H + \text{jets}$ events with $p_T(H) > 200$ GeV. The learning rate is set to 10^{-3} . In this section, we present some results obtained by the well trained GCN.

A. Number of input particles

As have been discussed in Sec. II, we consider input datasets with a number of particles $N_{\text{particle}} = 1024$ and

TABLE I. The recall, accuracy and precision of the GCN that is trained on the $N_{\text{particle}} = N_{\text{train}}$ event sample and tested on the $N_{\text{particle}} = N_{\text{test}}$ event sample.

	N_{train}	N_{test}	Recall	Accuracy	Precision
I	1024	1024	0.699	0.979	0.724
II	1024	2048	0.635	0.984	0.604
III	2048	1024	0.645	0.976	0.673
IV	2048	2048	0.709	0.985	0.599

$N_{\text{particle}} = 2048$ for each event, respectively. We note that the GCN does not require a fixed size of input. So the GCN which is trained on either $N_{\text{particle}} = 1024$ or $N_{\text{particle}} = 2048$ event sample can be tested on both samples. There will be in total four different applications. We provide the recall (number of correct signal/number of predicted signal), accuracy (number of correct predictions/total number of predictions), and precision (number of correct signal/number of actual signal) for those cases in Table I. It is not surprising to find out that the recalls are the highest for the cases where the training and testing samples have the same input size. The accuracies are always high because most of the particles are belong to the background, and they are easy to classify. Even though the event sample with larger N_{particle} provides a more complete list of Higgs constituents, the fraction of signal is significantly smaller (as shown in Fig. 1). As a result, the precision is higher (the detected signal is purer) when the GCN is tested on the event sample with $N_{\text{particle}} = 1024$. In the following, we will only consider the first case.

B. Reconstructed kinematic variables

The aim of our GCN is to mark the Higgs constituents out (with classification score greater than 0.5) from the particle list with heavy pileup contamination. The four momentum of the reconstructed Higgs boson candidate is given by the sum of the four-momenta of the marked hadrons. In most cases, our GCN works efficiently, as shown in the left panel of Fig. 3. While in other cases, particles with relatively large transverse momenta and far away from the Higgs in the $\eta - \phi$ plane will also be misassigned to the Higgs, as shown in the right panel of Fig. 3. So after the detection by the GCN, we further require that the Higgs constituents should lie within

$$\Delta R = \sqrt{(\eta_i - \eta_H)^2 + (\phi_i - \phi_H)^2} < 1.5 \quad (4.1)$$

of the reconstructed Higgs jet; otherwise, the predicted Higgs constituents are unlabeled.

To obtain an intuitive impression of the GCN performance, we compare it to a popular Higgs tagging method, which is composed of Cambridge-Aachen (C/A) jet clustering [11], mass-drop tagger [9], and trimming [12], denoted by MDT in the following. In this method, the final state particles are clustered with the C/A jet algorithm with an appropriate cone size parameter R_0 in order to capture most of the Higgs decay products. They are referred to as fat jets. The mass-drop tagger uses two criteria to characterize the substructure of the fat jet: by undoing the jet clustering, there is at least one step which breaks the jet j into two subjets j_1 and j_2 ($m_{j_1} > m_{j_2}$) such that the mass drop is significant ($m_{j_1} < 0.67m_j$) and the splitting is not asymmetric $\frac{\min(p_T^2(j_1), p_T^2(j_2))}{m_j^2} \Delta R_{j_1, j_2}^2 > 0.09$. The trimming method selects the hard subjets inside a fat jet, in order to mitigate the pileup contamination. The cone size parameter of jet clustering and trimming parameters in the MDT analysis are optimized for each Higgs production process to achieve the highest reconstruction efficiency of Higgs jet within 125 ± 5 GeV mass window.

In the Fig. 4, we plot the invariant mass distributions of the Higgs jets from the ground truth selection, MDT analyses, and GCN analyses, respectively. The ground truth selection corresponds to selecting the Higgs constituents that belong to the 1024 highest- p_T particles for each event. The Higgs jet momentum is given by the vector sum of momenta of all selected Higgs constituents. In the figure, the height of histogram for the ground truth Higgs invariant mass is scaled by a factor of 0.4 for

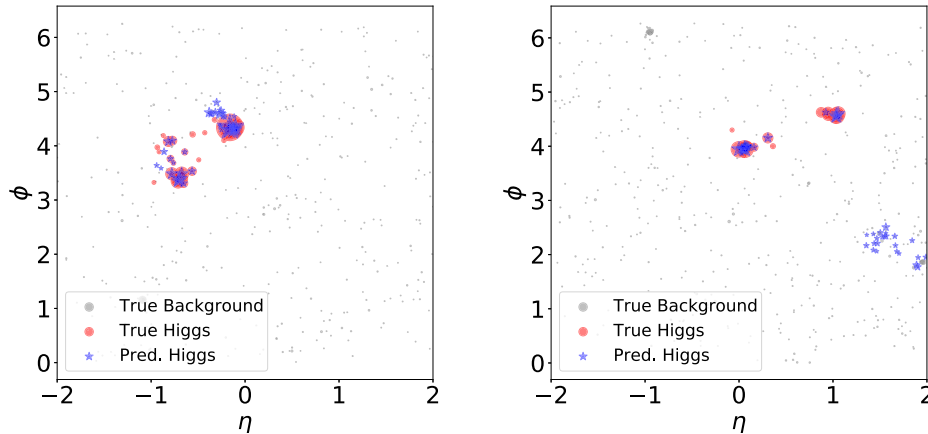


FIG. 3. Each panel is showing a $H + \text{jets}$ event in the $\eta - \phi$ plane, including the pileup contribution. The truth level Higgs constituents are marked by red disks and the predicted Higgs constituents are marked by blue stars. The size of disk and star is indicating the transverse momentum for each particle. For visibility, the sizes of the Higgs constituents (both truth level and GCN prediction) are magnified by a factor of 20.

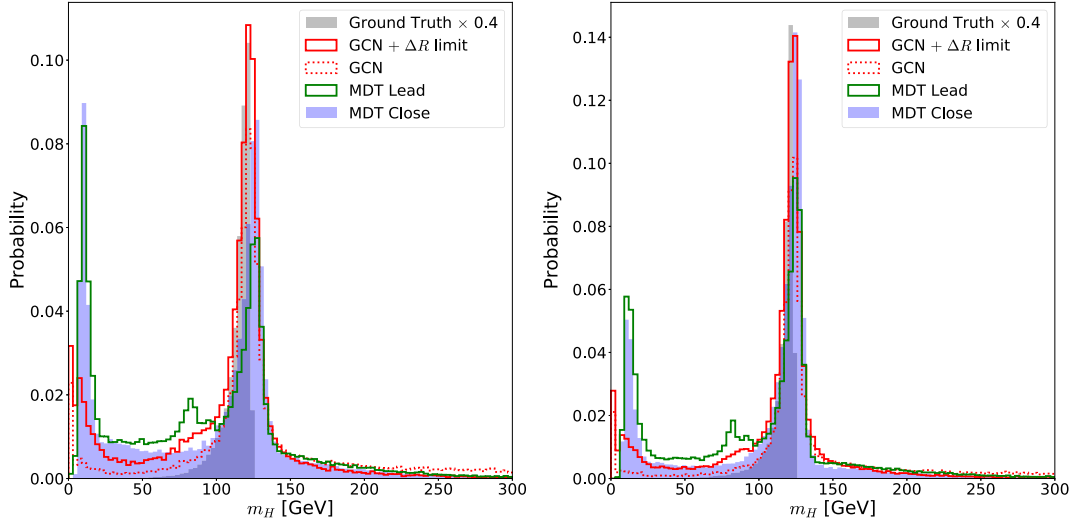


FIG. 4. The invariant mass distribution of the selected Higgs constituents at ground truth level (the height of which is scaled by a factor of 0.4 for visibility), the reconstructed Higgs jet from the MDT method (by choosing the jet with highest p_T and the jet closest to the true Higgs), and the GCN method (with and without the angular separation cut $\Delta R < 1.5$). Methods applied to the event samples of $H + \text{jets}$ process with cut $p_T(H) > 200$ GeV (left panel) and $p_T(H) > 300$ GeV (right panel), respectively.

visibility. In the MDT analyses, $R_0 = 1.5$ (1.3) is chosen for fat jet clustering in an event sample with a Higgs boson $p_T > 200$ (300) GeV. The fat jets are trimmed by reclustering the constituents into $R_{\text{sub}} = 0.20(0.19)k_t$ -subjets and discarding those with $p_T^{\text{subjet}} < f_{\text{cut}} p_T^{\text{jet}}$, where $f_{\text{cut}} = 0.05$. Among those jets with substructure, the Higgs jet candidates are selected either as the leading p_T trimmed fat jet (denoted by MDT lead, which is oversimplified but realistic) or the one closest (in term of angular separation $\sqrt{\Delta\eta^2 + \Delta\phi^2}$) to the true Higgs boson (denoted by MDT close, corresponds to the ideal case, which can be approached by using sophisticated Higgs tagging method). The performances of the GCN both with and without angular separation cut [Eq. (4.1)] are shown. We can find that the GCN method with an angular separation cut performs the best, in the sense that it has the largest number of events with reconstructed Higgs jet invariant mass close to 125 GeV. In particular, when the Higgs decay products are well separated, the Higgs constituents can not be fully captured by the CA jet algorithm in the MDT lead and MDT close analyses. This leads to the peaks in the distributions of the reconstructed Higgs invariant mass at $\lesssim 10$ GeV in these two methods. The problem is less severe in the GCN methods. Comparing left and right panels of Fig. 4, because the higher transverse momentum of the Higgs lead to more remarkable Higgs jet features, the efficiency and the precision of the Higgs reconstructions are improved in all methods,³ and the fake peak at

³Although the parameters in the MDT method are adjusted accordingly for each event sample, the GCN is always the one trained on the event sample with $p_T > 200$ GeV.

low invariant mass is greatly suppressed in the events sample with a Higgs boson $p_T > 300$ GeV.

The accuracies of the Higgs momentum reconstructions from the MDT close method and the GCN method with ΔR limit are shown in Fig. 5, in terms of two-dimensional distributions on the $\frac{\Delta m}{m} - \frac{\Delta p_T}{p_T}$ plane and $\Delta\eta - \Delta\phi$ plane, where the $m = 125$ GeV and p_T is the Higgs boson transverse momentum at parton level; Δm , Δp_T , $\Delta\eta$, and $\Delta\phi$ are the deviations between the reconstructed Higgs momenta and the truth-level Higgs boson momentum. Two event samples of the $H + \text{jets}$ process with $p_T(H) > 200$ GeV and $p_T(H) > 300$ GeV are analyzed for illustration. The different shades of gray regions and black contours indicate 20%, 40%, and 60% of events, respectively. The closer they are to the center, the higher accuracy they stand for. The GCN method outperforms the MDT method in kinematic variables reconstruction except the transverse momentum, where a systematic excess is prominent. Because the pileup effects are mitigated by trimming in the MDT method, no such pileup mitigation procedure is applied in the GCN method. We have used the vector sum of Higgs constituents momenta as the reconstructed Higgs momentum, the transverse momentum of the Higgs jet can be very sensitive to misassigned particles from pileup events, leading to an over estimated Higgs boson p_T . Using the GCN method to reconstruct the Higgs jet with $p_T > 200$ GeV, more than half of the events can be reconstructed with $\frac{\Delta m}{m} \in [-0.4, 0.2]$, $\frac{\Delta p_T}{p_T} \in [-0.1, 0.2]$, $\Delta\eta \in [-6, 6] \times 10^{-2}$, and $\Delta\phi \in [-6, 6] \times 10^{-2}$. The resolutions are even better for reconstructing a Higgs jet with $p_T > 300$ GeV.

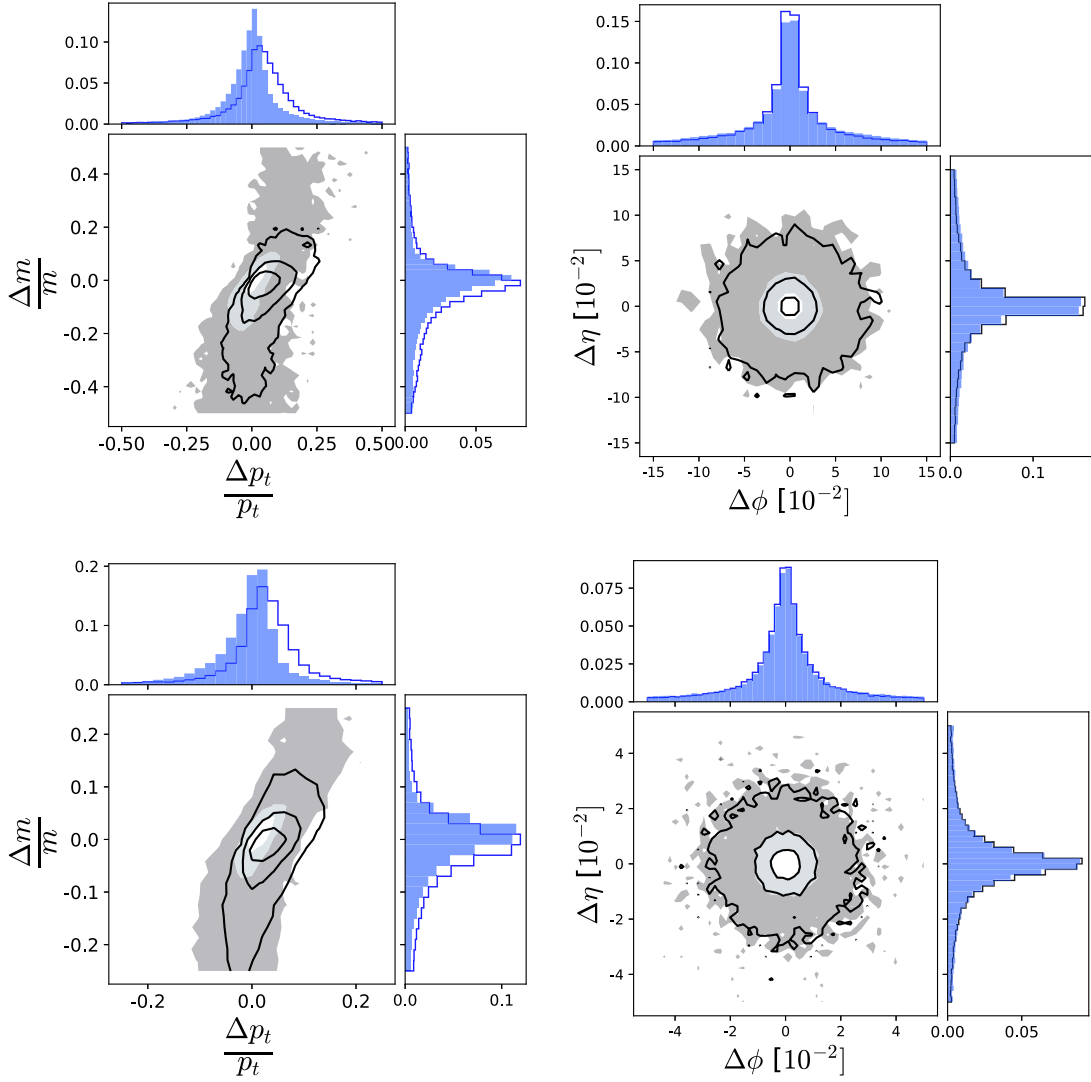


FIG. 5. $\frac{\Delta m}{m} - \frac{\Delta p_t}{p_t}$ and $\Delta\eta - \Delta\phi$ distributions for the reconstructed Higgs boson in the event samples of the $H + \text{jets}$ process with cut $p_T(H) > 200$ GeV (upper panels) and $p_T(H) > 300$ GeV (lower panels). The grey (blue) regions indicate the distributions of reconstructed variables from the MDT close method. And the black contours (blue lines) correspond to the distributions of reconstructed variables from the GCN method with ΔR cut.

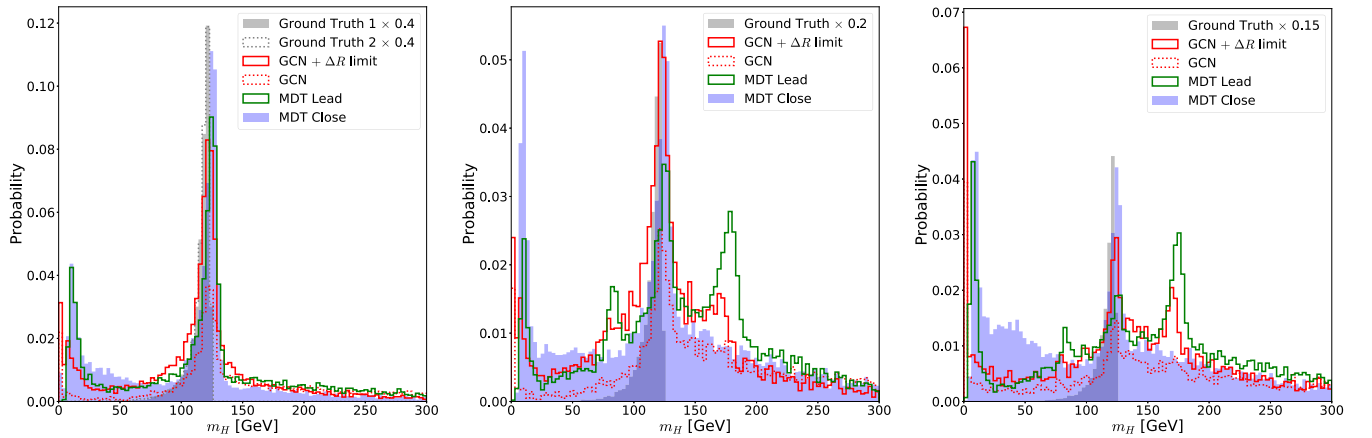


FIG. 6. Same as Fig. 4. Methods applied to the events samples of two Higgs bosons plus jets process (left), Higgs boson plus a top quark pair process (middle), and a SUSY process (right).

TABLE II. Jet clustering and trimming parameters in the MDT methods. The MDT lead and MDT close methods take the same values.

Process	R_0	R_{sub}	f_{cut}
$HHjjj$	1.4	0.19	0.05
$H\bar{t}t$	1.4	0.19	0.05
$\tilde{t}_1\tilde{t}_1^*$	0.9	0.18	0.05

C. Applications to different processes

Even though the GCN is trained on the events of the $H + \text{jets}$ process, it may serve as a general Higgs tagger for many other processes. For illustration, we perform the tests of the GCN on three processes different from the one used for training. Each process contains at least one Higgs jet in the final state. They are two Higgs bosons plus jets process, a Higgs boson plus a top quark pair process, and a SUSY process. The details of their simulations have been discussed in Sec. II.

In Fig. 6, we present the Higgs jet invariant mass distributions obtained by the MDT methods and the GCN methods for these three processes. As before, the ground truth distributions are superposed. For the two Higgs boson plus jets process, the Higgs jet candidate can be selected without ambiguity in the GCN method and MDT lead method. Unlike the image based method as proposed in Ref. [22], which can detect multiple Higgs jets in an event, the GCN is only capable of detecting one Higgs jet for each event. According to the left panel of Fig. 6, the GCN tends to find out the Higgs boson with a more remarkable substructure in the Higgs pair production process. Thus, the peak of the invariant mass distribution at ~ 125 GeV is sharper than the one showing in the left panel of Fig. 4. While for the MDT close method, the trimmed fat jet that is closest to either of the true Higgs bosons is chosen. The jet clustering and trimming parameters in the MDT analyses (same for MDT lead and MDT close) for different processes are optimized in the ranges of $R_0 \in [0.5, 2.0]$, $R_{\text{sub}} \in [0.1, 0.4]$, and $f_{\text{cut}} \in [0.02, 0.1]$ with step sizes 0.1, 0.01, and 0.01, respectively. The values with the highest reconstruction efficiency of Higgs jet within a mass window 125 ± 5 GeV⁴ are chosen, as shown in Table II. Due to the stringent trimming parameters that we choose, part of the Higgs constituents are also removed in some events. It leads to the peaks of Higgs invariant mass distributions at around zero. The effect is more remarkable when there are larger number of energetic hadrons (which are not originated from the Higgs boson) flying around the Higgs jet.

The performance of the GCN degrades in the second process due to the existence of energetic top quark decay

⁴Same values are obtained when optimizing the efficiency within the 125 ± 20 GeV mass window.

TABLE III. The standard derivatives for the Gaussian smearing of azimuth, pseudorapidity, and energy of different particles.

	$\sigma(\phi) = \sigma(\eta)$	$\sigma(E)$
Electron	0.002	$\sqrt{0.03^2 + (0.0013E)^2}$
Muon	0.001	$\sqrt{0.01^2 + (0.0001E)^2}$
Tracker	0.002	$\sqrt{0.06^2 + (0.0013E)^2}$
Photon	0.02	$\sqrt{0.101^2 E + (0.0017E)^2}$
Neutral hadron	0.1	$\sqrt{1.59^2 + 0.5205^2 E + (0.0302E)^2}$

products. While in the third process, one of the top quark is boosted, producing clusters of particles with high transverse momenta. It can be easily misassigned as Higgs constituents by the GCN (similar as in the right panel of Fig. 3, GCN tends to assign high classification scores to the energetic particles within clusters), and they are too energetic to be thrown away by the angular separation cut $\Delta R < 1.5$. As a result, the Higgs detection of the GCN method is not successful in most of the cases for this process.

In practice, if the target process has been set and for which the performance of our GCN is not satisfying, one can improve the performance by transfer learning: initialize the model with our well trained weights, train the model by using events of the target process and with lower learning rate.

D. Signal and background discrimination

To study the background resisting capacity of the GCN, we carry out a comparison study on the performances of our GCN on the $H + \text{jets}$ process with $p_T(H) > 200$ (300) GeV and on the top quark pair production process. In simulation of the $t\bar{t}$ events, the parton level $t\bar{t}$ production is matched to parton shower up to two additional jets, and each event is overlaid with an average number of 50 pileup events.

The finite momentum resolutions are considered by smearing each component of a momentum with a Gaussian distribution. The standard derivatives (using the parameters of the ATLAS detector [48,49]) for different particles are listed in Table III.

We adopt the boosted decision tree (BDT) technique to obtain a combined discriminating powers of several variables. The variables are classified into two categories:

- (i) Jet substructure variables for the Higgs jet candidate.⁵
- (ii) Variables reconstructed from the scores that the GCN assigns to Higgs constituents.

The first category includes the N-subjettiness variables [50] $\tau_{21} = \tau_2/\tau_1$ and $\tau_{32} = \tau_3/\tau_2$, with τ_N being defined as

⁵The variables in this class can be calculated for the Higgs jet candidate in both the MDT analysis and the GCN analysis.

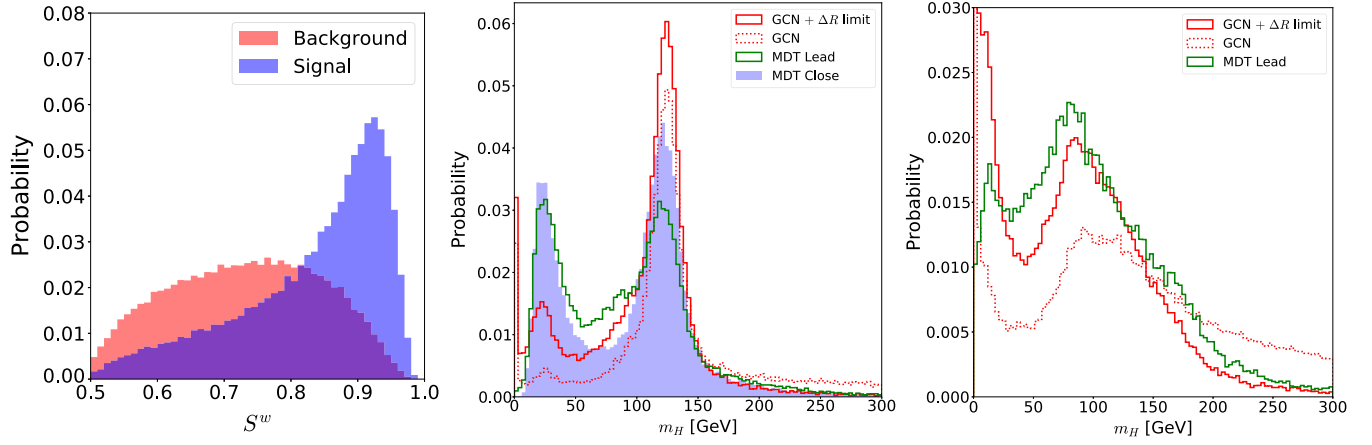


FIG. 7. The distributions of the p_T weighted score S^w (left panel) and Higgs invariant mass for signal [H + jets with $p_T(H) > 200$ GeV, middle panel] and $t\bar{t}$ background (right panel) processes. The detector effects have been taken into account.

$$\tau_N \equiv \frac{\sum_k \min\{\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}\}}{\sum_k p_{T,k} R_0}, \quad (4.2)$$

where the summations run over all Higgs jet constituents, R_0 is the cone size parameter in the original jet clustering algorithm, $\Delta R_{I,k}$ denotes the angular distance ($\sqrt{\Delta\eta^2 + \Delta\phi^2}$) between the subjet I and jet constituent k . There are also variables that characterize the jet profile: $f_{\text{cent}} = E^{0.1}/E^{0.2}$ with $E^{0.1}/E^{0.2}$ is the summed energy of particles inside a cone of $R = 0.1/0.2$ centered on the jet direction, the transverse momentum of the most energetic track inside a cone of radius $R = 0.2$ centered on the jet direction, and p_T weighted sum of the angular distances of all particles inside the jet.

The second category includes the average $\langle S \rangle$ and the standard deviation σ_S of the scores that are assigned to the constituents of the Higgs jet by the GCN. Moreover, we define a p_T weighted score S^w for the Higgs jet candidate,

$$S^w = \sum (s_i \cdot p_{T,i}) / \sum p_{T,i}, \quad (4.3)$$

where i runs over all Higgs constituents, s_i and $p_{T,i}$ are the GCN score and the transverse momentum of the i th constituent. The distribution of the S^w is presented in the left panel of Fig. 7. Only events which contain GCN tagged Higgs jet are used. It includes 96% of simulated signal events and 87% of simulated background events. The Higgs jet candidates in the events of the signal process tend to gain higher S^w than that in the events of background process, which makes this variable useful for signal and background discrimination.

Beside the variables in those two categories, the transverse momentum $p_T(H)$ and invariant mass m_H of the reconstructed Higgs are also powerful discriminating variables. In the middle and right panels of Fig. 7, we plot the distributions of reconstructed Higgs invariant mass for signal [H + jets with $p_T(H) > 200$ GeV] and background ($t\bar{t}$) processes, taking into account the detector effects. For the signal process, compared with the left panel of Fig. 4, we can find that the detector effects broaden the peaks at 125 GeV in all methods. The GCN method still outperforms the MDT method after taking into account the

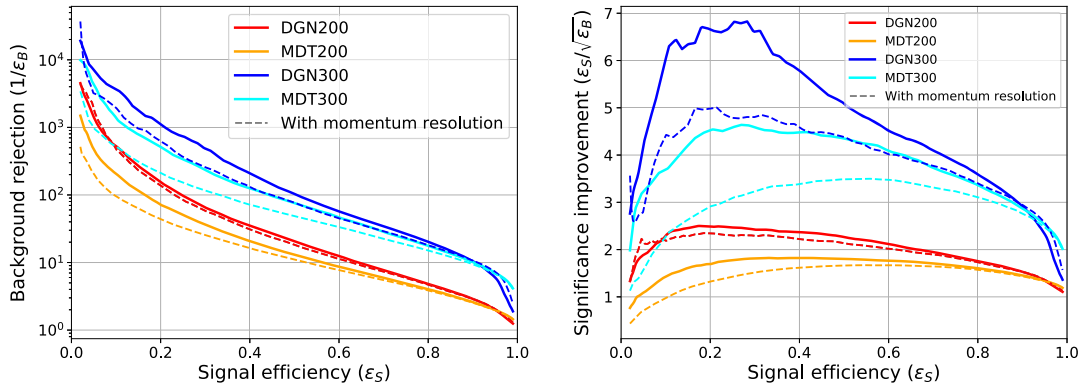


FIG. 8. The receiver operating characteristic curves (left panel) and the significance improvement curves (right panel) of the GCN analyses and MDT analyses. Two different signal samples are considered: H + jet with $p_T(H) > 200$ GeV and $p_T(H) > 300$ GeV. Results which take into account the finite momentum resolution are shown by the dashed lines.

detector effects. The distributions of the fake Higgs jet invariant mass for background are typically around 50–100 GeV in MDT and GCN analyses, and the cut $\Delta R < 1.5$ in the GCN method is found to be effective in suppressing the distribution in large m_H region.

In the following, the BDT analysis with input of variables in the first category as well as the $p_T(H)$ and the m_H calculated on the MDT lead tagged Higgs jet will be referred to as MDT analysis. And the BDT analysis with an input of variables in both categories as well as the $p_T(H)$ and the m_H calculated on the GCN tagged Higgs jet will be referred to as GCN analysis. The receiver operating characteristic (ROC) curves for these two BDT analyses applying to two signal samples [with $p_T(H) > 200$ GeV and $p_T(H) > 300$ GeV, respectively] are plotted in the left panel of Fig. 8. The solid and dashed lines correspond to the results with and without considering the finite momentum resolutions, respectively. We can find the GCN analysis has improved signal and background discriminating power than the MDT analysis for both signal samples. Moreover, the results obtained by the GCN analyses are less sensitive to the momentum resolution than the results of the MDT analyses. The significance improvement curves are presented in the right panel of Fig. 8. It shows that the signal significances improve most when the signal selection efficiency $\epsilon_S \sim 0.2$. And the GCN analysis outperforms the MDT analysis by a factor of 1.5 on the signal significance for both signals.

V. CONCLUSION AND OUTLOOK

Representing each collider event as a point cloud, we adopt the GCN with focal loss to reconstruct the Higgs jet in the event. The $H + \text{jets}$ with a cut $p_T(H) > 200$ GeV is taken as the benchmark process, on events of which the GCN is trained. We find the GCN analysis outperforms the traditional jet substructure analysis in both Higgs tagging efficiency and momentum reconstruction accuracy. And the GCN method is less sensitive to pileup contamination. In the GCN method without dedicated pileup mitigation, 60%

of $H + \text{jets}$ events with cut $p_T > 200$ (300) GeV can be reconstructed with $|\Delta\phi| \sim |\Delta\eta| < 6 \times 10^{-2}$ (2.5×10^{-2}), $\Delta m/m \in [-0.4, 0.2]$ ($[-0.25, 0.12]$), and $\Delta p_T/p_T \in [-0.1, 0.2]$ ($[-0.05, 0.1]$).

The GCN which is trained on the events of the $H + \text{jets}$ process is also capable of reconstructing the Higgs jet in events of other processes, even though its performance is degraded when there are boosted particles other than the Higgs boson in an event. Finally, we show that the features learned by the GCN are complementary to jet substructure variables in separating the events of the $H + \text{jets}$ process and the $t\bar{t}$ process.

There are several limitations of the current GCN method: (1) The method can only reconstruct a single Higgs jet for each event (even in the event with multiple Higgs bosons). (2) Although we try to propose a general Higgs jet detection method which does not depend on the Higgs production process, it turns out that the current method is not efficient in detecting the Higgs jet in processes where the Higgs is accompanied by other energetic particles. Since events of those new processes are not used in training, modifying the network and loss function can not help improving the performance. On the other hand, if the target process has been set, we can always improve the performance by transfer learning (probably with modified loss). (3) The b-tagging is not considered in our study, which is a very important step towards the Higgs tagging. These subjects will be studied in future works.

ACKNOWLEDGMENTS

This work was supported in part by the Fundamental Research Funds for the Central Universities, by the National Science Foundation of China under Grant No. 11905149, by Projects No. 11847612 and No. 11875062 supported by the National Natural Science Foundation of China, and by the Key Research Program of Frontier Science, Chinese Academy of Sciences.

-
- [1] ATLAS Collaboration, Combined measurements of Higgs boson production and decay using up to 80 fb⁻¹ of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment, *Phys. Rev. D* **101**, 012002 (2020).
 - [2] CMS Collaboration, Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13$ TeV, *Eur. Phys. J. C* **79**, 421 (2019).
 - [3] A. Abdesselam *et al.*, Boosted objects: A probe of beyond the standard model physics, *Eur. Phys. J. C* **71**, 1661 (2011).
 - [4] A. Altheimer *et al.*, Jet substructure at the tevatron and LHC: New results, new tools, new benchmarks, *J. Phys. G* **39**, 063001 (2012).
 - [5] A. Altheimer *et al.*, Boosted objects and jet substructure at the LHC. Report of BOOST2012, held at IFIC Valencia, 23rd-27th of July 2012, *Eur. Phys. J. C* **74**, 2792 (2014).
 - [6] D. Adams *et al.*, Towards an understanding of the correlations in jet substructure, *Eur. Phys. J. C* **75**, 409 (2015).
 - [7] A. J. Larkoski, I. Moult, and B. Nachman, Jet substructure at the large hadron collider: A review of recent advances in theory and machine learning, *Phys. Rep.* **841**, 1 (2020).

- [8] R. Kogler *et al.*, Jet substructure at the large hadron collider: Experimental review, *Rev. Mod. Phys.* **91**, 045003 (2019).
- [9] 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).
- [10] M. Cacciari, G. P. Salam, and G. Soyez, The anti- k_t jet clustering algorithm, *J. High Energy Phys.* **04** (2008) 063.
- [11] Y. L. Dokshitzer, G. Leder, S. Moretti, and B. Webber, Better jet clustering algorithms, *J. High Energy Phys.* **08** (1997) 001.
- [12] D. Krohn, J. Thaler, and L.-T. Wang, Jet trimming, *J. High Energy Phys.* **02** (2010) 084.
- [13] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, Techniques for improved heavy particle searches with jet substructure, *Phys. Rev. D* **80**, 051501 (2009).
- [14] D. Guest, K. Cranmer, and D. Whiteson, Deep learning and its application to LHC physics, *Annu. Rev. Nucl. Part. Sci.* **68**, 161 (2018).
- [15] K. Albertsson *et al.*, Machine learning in high energy physics community white paper, *J. Phys. Conf. Ser.* **1085**, 022008 (2018).
- [16] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, Machine learning at the energy and intensity frontiers of particle physics, *Nature (London)* **560**, 41 (2018).
- [17] K. Datta, A. Larkoski, and B. Nachman, Automating the construction of jet observables with machine learning, *Phys. Rev. D* **100**, 095016 (2019).
- [18] J. Lin, M. Freytsis, I. Moulton, and B. Nachman, Boosting $H \rightarrow b\bar{b}$ with machine learning, *J. High Energy Phys.* **10** (2018) 101.
- [19] E. A. Moreno, T. Q. Nguyen, J.-R. Vlimant, O. Cerri, H. B. Newman, A. Perival, M. Spiropulu, J. M. Duarte, and M. Pierini, Interaction networks for the identification of boosted $H \rightarrow b\bar{b}$ decays, *Phys. Rev. D* **102**, 012010 (2020).
- [20] A. Chakraborty, S. H. Lim, and M. M. Nojiri, Interpretable deep learning for two-prong jet classification with jet spectra, *J. High Energy Phys.* **07** (2019) 135.
- [21] CMS Collaboration, Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques, *J. Instrum.* **15**, P06005 (2020).
- [22] J. Li, T. Li, and F.-Z. Xu, Reconstructing boosted Higgs jets from event image segmentation, *arXiv:2008.13529*.
- [23] J. Shlomi, P. Battaglia, and J.-R. Vlimant, Graph neural networks in particle physics, *arXiv:2007.13681*.
- [24] J. Arjona Martínez, O. Cerri, M. Pierini, M. Spiropulu, and J.-R. Vlimant, Pileup mitigation at the Large Hadron Collider with graph neural networks, *Eur. Phys. J. Plus* **134**, 333 (2019).
- [25] S. Farrell *et al.*, Novel deep learning methods for track reconstruction, in *4th International Workshop Connecting The Dots 2018* (2018), p. 10.
- [26] X. Ju *et al.*, Graph neural networks for particle reconstruction in high energy physics detectors, in *33rd Annual Conference on Neural Information Processing Systems* (2020).
- [27] P. T. Komiske, E. M. Metodiev, and J. Thaler, Energy flow networks: Deep sets for particle jets, *J. High Energy Phys.* **01** (2019) 121.
- [28] H. Qu and L. Gouskos, ParticleNet: Jet tagging via particle clouds, *Phys. Rev. D* **101**, 056019 (2020).
- [29] E. A. Moreno, O. Cerri, J. M. Duarte, H. B. Newman, T. Q. Nguyen, A. Perival *et al.*, JEDI-net: A jet identification algorithm based on interaction networks, *Eur. Phys. J. C* **80**, 58 (2020).
- [30] V. Mikuni and F. Canelli, ABCNet: An attention-based method for particle tagging, *Eur. Phys. J. Plus* **135**, 463 (2020).
- [31] A. Chakraborty, S. H. Lim, M. M. Nojiri, and M. Takeuchi, Neural network-based top tagger with two-point energy correlations and geometry of soft emissions, *J. High Energy Phys.* **07** (2020) 111.
- [32] E. Bernreuther, T. Finke, F. Kahlhoefer, M. Krämer, and A. Mück, Casting a graph net to catch dark showers, *arXiv:2006.08639*.
- [33] IceCube Collaboration, Graph neural networks for Icecube signal classification, *arXiv:1809.06166*.
- [34] M. Abdughani, J. Ren, L. Wu, and J. M. Yang, Probing stop pair production at the LHC with graph neural networks, *J. High Energy Phys.* **08** (2019) 055.
- [35] X. Ju and B. Nachman, Supervised jet clustering with graph neural networks for Lorentz boosted bosons, *arXiv:2008.06064*.
- [36] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, Dynamic graph CNN for learning on point clouds, *arXiv:1801.07829*.
- [37] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, Focal loss for dense object detection, *arXiv:1708.02002*.
- [38] 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.
- [39] T. Sjostrand, S. Mrenna, and P. Z. Skands, A brief introduction to PYTHIA 8.1, *Comput. Phys. Commun.* **178**, 852 (2008).
- [40] P. Skands, S. Carrazza, and J. Rojo, Tuning PYTHIA 8.1: The Monash 2013 tune, *Eur. Phys. J. C* **74**, 3024 (2014).
- [41] ATLAS Collaboration, The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model (2016).
- [42] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, Parton fragmentation and string dynamics, *Phys. Rep.* **97**, 31 (1983).
- [43] ATLAS Collaboration, Charged-particle distributions in $\sqrt{s} = 13$ TeV pp interactions measured with the ATLAS detector at the LHC, *Phys. Lett. B* **758**, 67 (2016).
- [44] CMS Collaboration, Quark-gluon jet discrimination at CMS, in *2nd Large Hadron Collider Physics Conference* (2014), p. 9.
- [45] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, Spatial transformer networks, *arXiv:1506.02025*.
- [46] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, Focal loss for dense object detection, *arXiv:1708.02002*.

- [47] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean *et al.*, TensorFlow: A system for large-scale machine learning, in *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, OSDI'16, USA* (USENIX Association, Savannah, 2016), pp. 265–283.
- [48] ATLAS Electromagnetic Barrel Calorimeter Collaboration, Energy linearity and resolution of the ATLAS electromagnetic barrel calorimeter in an electron test-beam, [Nucl. Instrum. Methods Phys. Res., Sect. A **568**, 601 \(2006\)](#).
- [49] Y. A. Kulchitsky, M. V. Kuzmin, J. A. Budagov, V. B. Vinogradov, and M. Nessi, Hadron energy reconstruction for the ATLAS barrel prototype combined calorimeter in the framework of the nonparametrical method, [arXiv:hep-ex/0004009](#).
- [50] J. Thaler and K. Van Tilburg, Identifying boosted objects with N-subjettiness, [J. High Energy Phys. **03** \(2011\) 015](#).