Automatic detection of boosted Higgs boson and top quark jets in an event image

Sang Kwan Choi,[†] Jinmian Li,[*] Cong Zhang,[‡] and Rao Zhang[§]
College of Physics, Sichuan University, Chengdu 610065, China

(Received 17 July 2023; accepted 9 November 2023; published 4 December 2023)

We build a deep neural network based on the Mask region-based convolutional neural network framework to detect the Higgs jets and top quark jets in any event image. We propose an algorithm to assign the top quark final states at the ground truth level so that the network can be trained in a supervised manner. A new jet branch is added to the network, which uses constituent information to predict the four-momenta of the original parton, thus intrinsically implementing the pileup mitigation. The network can predict both the shapes and the momenta of target jets. We show that the network surpasses the LorentzNet in top and Higgs tagging and the PELICAN network in momentum regression for certain cases, in terms of reconstruction efficiency and accuracy. We also show that the performance of the network does not degrade much when applied to events from a different process from the trained one and to events with overlapping jets.

DOI: 10.1103/PhysRevD.108.116002

I. INTRODUCTION

Because of the color confinement, the quarks and gluons produced in a hard process cannot be detected individually. Instead, they will go through parton showering and hadronization after production, giving rise to a collimated spray of energetic detectable color singlet hadrons, which are referred to as jets. Establishing the correspondence between jets and partons is essential for understanding the underlying physics of collider events. It requires an infrared-safe algorithm that can attribute the related final state hadrons to their partonic ancestors and predict the four-momentum of the parton.

At the LHC and other collider experiments, the jets are reconstructed by using sequential recombination algorithms [1], in which the final state hadrons are pairwise recombined according to some distance measures, such as the Cambridge/Aachen algorithm [2], anti-k_T algorithm [3], and so on [4–6]. In those jet clustering algorithms, an appropriate cone-size parameter needs to be taken according to the configuration of the detector and the properties of the target jet. At the LHC, the anti-k_T jet algorithm with cone size $R = 0.4$ works efficiently in finding quark and gluon jets in the ATLAS and CMS detectors. Another origin of a jet is a boosted heavy particle decaying into hadronic final states, for example, the top quark and the Higgs boson. As the typical jet cone size of a heavy resonance is given by $R \sim 2m/p_T$, a much larger value is adopted to fully capture the jet constituents. In contrast to the jet originating from the light quark and gluon, those fat jets are characterized by a remarkable substructure. A variety of jet substructure techniques [7–12] have been proposed for tagging heavy resonant jets, such as mass-drop tagging [13] for the Higgs boson, the HEPTopTagger algorithm [14] for the top quark, and N-subjettiness [15,16] for general fat jets. One practical issue for the fat-jet reconstruction at hadron colliders would be the heavy contamination from pileup events as well as underlying events. The average number of pileup events reaches $\langle\mu\rangle = 35$ for the LHC Run 2 and $\langle\mu\rangle = 140$ for the High-Luminosity LHC. The large cone size of the fat jet renders many of those background particles to be included as the jet constituents. Although the jet grooming methods [17,18] have been found to be very helpful in mitigating the pileup effects, there are still relatively large errors in obtaining the original parton momentum when it is calculated by the vector sum of momenta of the groomed jet constituents.

In terms of fat-jet tagging efficiency, machine-learning techniques have proven to substantially outperform those jet substructure techniques [11,19–22]. According to the jet formation, a jet can be either viewed as sequences/trees formed through sequential parton showering and hadronization or viewed as graphs/point clouds with the information encoded in the adjacency nodes and edges. Moreover, the calorimeters inside the detector measure the angular position and energy of particles on fine-grained spatial

[*]Corresponding author: jmli@scu.edu.cn
[†]hermit1231@sogang.ac.kr
[‡]zhangcong.phy@gmail.com
[§]zhangrao@stu.scu.edu.cn

cells. Considering each calorimeter cell as a pixel and the energy deposition as the intensity, a jet can be naturally viewed as a digital image. All three representations of the jet are common objects in machine learning. They can be proceeded by different kinds of neural network architectures, i.e., recurrent neural networks [23] and a transformer network [24] for sequence, graph neural networks (GNNs) [25–27] for point cloud, recursive neural networks [28,29] for tree networks, and two-dimensional convolutional neural networks (CNNs) [30–33] for jet images. Those delicate deep-learning approaches can better leverage the fine resolution of detectors and automatically figure out the complex pattern of a jet from the low-level inputs. However, those methods rely on traditional jet clustering algorithms to reconstruct the jet at the first stage. As a result, a predefined cone-size parameter is required, and the jet representation could suffer from distortion due to an inappropriate cone-size parameter or contaminations from pileup events. ParticleNet [27], Particle Transformer [34], LorentzNet [35], and PELICAN [36] are among the state-of-the-art methods for Higgs boson and top tagging in this field. They achieve typical area under curve (AUC) values of over 0.98 for top tagging, without considering the pileup effects. In addition, the momentum reconstruction component of the PELICAN network can predict the p_T and mass of a W boson with standard deviations of a few percent.

Several studies [37–40] attempt to propose jet definition alternatives to the clustering methods, so that the presumed cone-size parameter is no longer mandatory. Meanwhile, the techniques of object detection and semantic segmentation in computer vision provide new ways to label the jet constituents. In the Monte Carlo simulation of collider events, the final state hadrons can be attributed to their ancestor parton without ambiguity, so it will be possible to build a neural network to label the jet constituents among final state particles based on supervised learning. Reference [41] studies the construction of a W boson jet from final state particles with the supervised GNN. In Ref. [42], we improve the GNN with a focal loss function, such that the method can remain efficient when heavy pileup contaminations are taken into account. Moreover, we demonstrate that the GNN, which is trained on events of the $H+$ jets process, is capable of detecting a Higgs jet in events of several different processes. The image segmentation with the convolutional network also works well in detecting the Higgs jets in event images. In Ref. [43], we take the event information as a digital image and adopt the Mask region-based convolutional neural network (Mask R-CNN) framework [44] to reconstruct the Higgs jet in the event image. Those deep-learning methods reach higher efficiency of Higgs jet detection and higher accuracy of Higgs momentum reconstruction than the traditional jet clustering and substructure tagging methods. However, those methods have not been tested for detecting different kinds of multiple jets in an event. In this work, based on the event image representation, we adopt a modified version of Mask R-CNN to detect/reconstruct all Higgs boson and top quark jets in an event. Since the top quark is carrying a color charge, its energy flow is interconnected with other colored particles in the production process and with the beam remnants. There is no unique way to associate the hadronic final states with it. Based on the rule that the vector sum of top jet constituents can reproduce the top quark momentum well, we propose a pattern of attribution for the top quark final states in the training sample. As a result, the Mask R-CNN can be trained to detect the top quark jet in a supervised way, similar to the Higgs jet. Moreover, the Mask R-CNN can only predict regions with masks, which will also include a large number of particles from pileup events. The Higgs boson and top quark momenta cannot be simply obtained by the vector sum of momenta of masked pixels (calorimeter cells) in the jet image. We add a new fully connected network component to the Mask R-CNN, which takes the input of bounding box information to predict the Higgs boson and top quark momenta. Given the Higgs boson and top quark momenta at the ground truth level, this network component is capable of pileup mitigation in an automatic way after supervised learning. It turns out that the modified Mask R-CNN can not only provide the jet regions (masks) of the Higgs boson and top jets in an event but also predict the four-momenta of the Higgs boson and top parton precisely. We compare the Higgs boson and top jet tagging performance with LorentzNet and the momentum regression performance with PELICAN.

This paper is organized as follows. In Sec. II, we describe the event generation and the event preprocessing. In Sec. III, we briefly introduce the Mask R-CNN framework and illustrate how events proceed. The changes to the Mask R-CNN are discussed in detail. In Sec. IV, the performance of the network being applied to the $Ht\bar{t}$ process as well as other processes that have not been used for training is presented. We summarize our work and conclude in Sec. V.

II. EVENT PREPARATION AND PREPROCESSING

The proton-proton collision events are simulated by the MG5_aMC@NLO framework [45] with center-of-mass energy $\sqrt{s} = 13$ TeV. PYTHIA8 [46] is used for the quark parton showering, hadronization, and hadron decay. The detector effects are not considered except for the angular granularity of calorimeters.[1] The angular size of the calorimeter cell is assumed to be 0.02×0.02 on the $\eta \times \phi$ plane. This is an idealized setup since the hadron calorimeter at the LHC usually has η/ϕ resolution larger than ~ 0.15. Although the precision of momentum reconstruction is limited by the cell size, we find that

[1]The effects of energy smearing will be discussed separately later.

the network performance, i.e., the Higgs boson/top detection efficiency, is barely changing with the cell size. The event image is built by presenting each calorimeter cell on the $\eta \times \phi$ plane as a pixel of an image, and the transverse momentum of the cell as the intensity (or gray scale color) of that pixel.

The network is trained on 1×10^6 events of the $Ht\bar{t}$ process, where both the Higgs boson and top quark jets are marked. It should be noted that training our network on other processes is certainly possible, and the performance of the network should be similar. To show the generality of the network, the performance of various test samples are studied, including $Ht\bar{t}$, $t\bar{t}t\bar{t}$, $HHt\bar{t}$ production in the Standard Model (SM) as well as the neutralino pair production and top squark pair production with subsequent decay $\tilde{\chi}_2^0 \to H\tilde{\chi}_1^0$ and $\tilde{t} \to t\tilde{\chi}_1^0$ in the supersymmetric (SUSY) model. The transverse momenta of the Higgs boson and top quark in the SM processes are required to be greater than 200 and 300 GeV, respectively. As for the SUSY case, we set the masses $m_{\tilde{\chi}_2^0} = 450\,\text{GeV}$, $m_{\tilde{t}} = 650\,\text{GeV}$, and $m_{\tilde{\chi}_1^0} = 100\,\text{GeV}$. We do not specify the decay modes of the Higgs boson and top quark in the training sample. However, we find the network exhibits better performance on events with a hadronically decaying Higgs boson ($H \to b\bar{b}$) and top quark ($t \to bW, W \to qq$). Therefore the Higgs boson and top quark in the test sample are forced to decay through those modes.

Moreover, there are multiple proton-proton collisions (referred to as pileup) in each bunch crossing at the LHC. Those collisions are dominated by nondiffractive events with small transverse momentum transfer. Simulation of the pileup events requires a perturbative parton shower, Lund-string hadronization, multiple parton interaction, and color reconnection, which are usually described by phenomenological models. The parameters in the models are not unique and need to be inferred from experimental data. The set of appropriately chosen parameters is dubbed PYTHIA tunes [47]. We adopt the A3 tune of PYTHIA8 with phenomenological parameters provided in Refs. [48,49] to simulate pileup events. The number of pileup events per bunch crossing at the LHC follows the Poisson distribution with an average value around $\langle\mu\rangle = 35$ at the LHC Run 2 and $\langle\mu\rangle = 200$ at the High-Luminosity LHC. We took the average number of pileup events of $\langle\mu\rangle = 50$ in our simulation. A detailed study of the effects of different pileup levels will be given later. Finally, we note that particles flying into the same calorimeter cell can only be identified with the summation of their momenta, so the pileup will increase the momenta of the target jet constituents, and the pileup mitigation procedure is essential to obtain a precise parton momentum.

A. Attributions of final state particles

In the training sample, the constituents of the Higgs jets and top/antitop quark jets need to be assigned beforehand. However, due to color confinement, some of the top quark final states could have multiple ancestors other than the top quark according to the Monte Carlo simulation. In constructing the top jet, we hope to only include the constituents whose momenta are mostly inherited from the top quark.

The final states of an $Ht\bar{t}$ event fall into four categories. Those who only have a unique ancestor should be assigned to the H, t, and $\bar{t}$ categories without ambiguity. The rest of the final states have multiple ancestors (dubbed the MA category) and should be assigned as top/antitop with some criteria. Hadrons in the MA category (one of the ancestors is the top quark or antitop quark) are ranked according to their angular distances ($\Delta R = \sqrt{\Delta\phi^2 + \Delta y^2}$) to the top/antitop quark, where ϕ is the azimuth angle, and y is rapidity. They are assigned to the top/antitop categories in order until the reconstructed top/antitop jet invariant mass exceeds $1.05m_t$ with m_t being the top quark invariant mass (which could be off shell) in the event.[2] In the left panel of Fig. 1, we show the distributions of the invariant masses for the Higgs jet, top quark jet, and antitop quark jet before and after the assignment. In terms of the invariant mass, we observe that the selection of hadrons in the MA category is necessary, and our assignment criterion is appropriate. In the right panel of the same figure, we illustrate the event image on the pseudorapidity (η) versus the azimuth angle (ϕ) plane after applying the assignment. Both the top and antitop jets have focused shapes since their constituents are selected according to the angular separation.

B. Data preparation for the network

The images fed to the network represent the transverse momentum deposited in the $\eta \times \phi$ plane. Since the pixel values of typical images are integers ranging from 0 to 255 in each channel, one may consider mapping from p_T to the integer values in red-green-blue (RGB). However, we instead use p_T itself as we find no disadvantages in the network performance. When the range of ϕ in the images is from 0 to 2π, the constituents of a jet with ϕ near the boundary ($\phi = 0$) may locate in two regions far apart from each other, which are susceptible to being considered as two different objects. In order to incorporate the periodicity in the network, one may introduce a periodic padding in convolutional layers, where an input image is padded in the periodic manner so that convolution kernels can read the values on the opposite side of the boundary. Another way to handle the periodicity is to have the images with an augmented range, for instance, $\phi \in [0, 3\pi]$, such that a

[2]Although this criterion cannot guarantee the total match between the momenta of the top quark and the top jet, this method already helps to capture most of the high energy constituents. Modifying the criterion will change the predicted mask (or jet shape), but will not have significant effects on the predicted momenta.

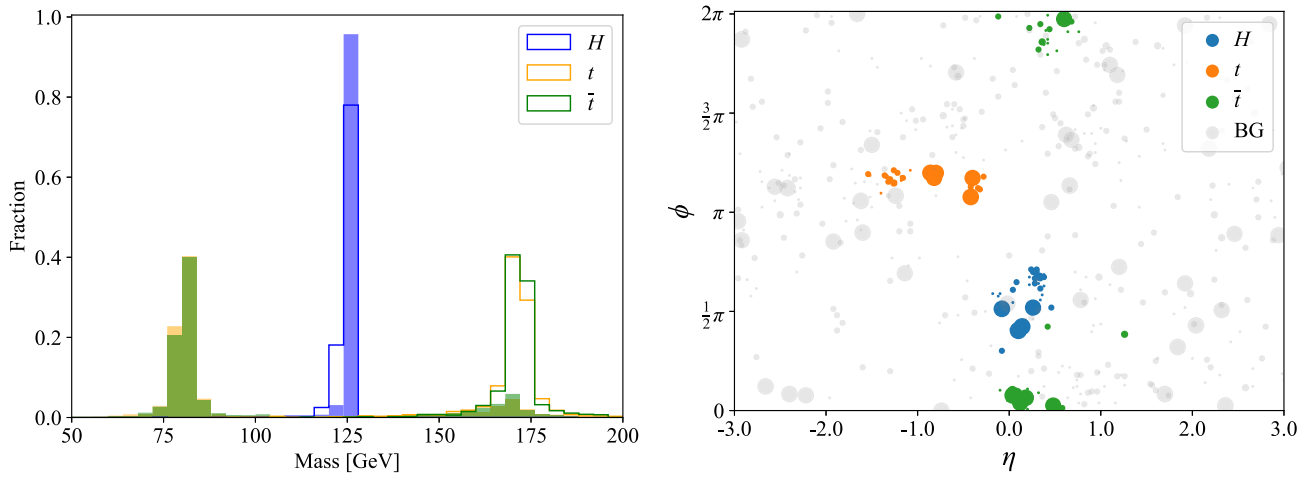

FIG. 1. Left: distributions of the invariant masses for the Higgs jet, top quark jet, and antitop quark jet in events of $Ht\bar{t}$ production at the LHC. The shaded histograms and the solid lines correspond to the distributions before and after including the selected hadrons in the MA category. Right: the event image after applying the assignment, where the size of each dot indicates the energy of each final state hadron.

continuous picture of a jet appears on the image at least once. In our brief implementation of the two schemes, we observed better performance from the augmented images. Therefore, we use the augmentation scheme in this paper. Considering together network requirements for image dimensions, the ϕ range is chosen to be from 0 to 2.85π. The spatial size of the input images is then 448×448 pixels corresponding to the $\eta \times \phi$ plane across $[-4.48, 4.48] \times [0, 2.85\pi]$ where the resolution is given as $\Delta\eta \times \Delta\phi = 0.02 \times 0.02$. Having three copies along the channel, the dimension of the input images is $448 \times 448 \times 3$. Having three copies is not just redundant due to the network requirement, but it has a nontrivial effect on the network. It implies that the kernel at the first convolutional stage should also have three channels, tripling the number of learnable parameters.

Given an input image, the original Mask R-CNN has three outputs for each candidate object: a class label, a bounding box, and a mask. Therefore, one needs to provide the ground truth of them to train the network, but there is an issue to address here. The mask is a binary image with the same spatial size as the input image, in which object pixels have the value 1 and background pixels have 0, and the smallest rectangle enclosing all the object pixels is the bounding box. In the jet detection task, it is reasonable to define the pixels where individual jet constituents are located as the object pixels. Note that, however, Mask R-CNN will crop candidate regions in the mask and resize them into a fixed size to compare to corresponding outputs of the network. The sparsity of the constituent pixels is not robust to resizing, in particular, when scattered in a broad region. Instead, we define a jet area that gives a mask of connected pixels. First, we preselect constituents which will compose the jet area. The preselection process is as follows. Boost along the beam direction to the frame where p_z of the parton is 0, and discard the constituents with energy lower than 0.1 GeV or with angular separation to the parton greater than $\pi/2$. Then, among the remaining constituents, select those having at least three others in the 20×20 pixels neighborhood or with p_T greater than 5 GeV. The neighborhood condition is imposed since we want to construct jet areas that do not change drastically depending on a couple of constituents unless p_T is significantly large, otherwise the bounding box predictions will fluctuate widely according to whether or not the network precisely detects a few constituents far from some clusters. The size of the neighborhood and the number of neighbors are empirically determined after monitoring the network performance of several cases. With the selected constituents, we define two types of jet areas: the convex hull and the enlargement. The first one is the area bounded by the convex hull covering all the selected ones, and the second one is obtained by expanding each selected pixel into the area of 9×9 pixels with the selected one at the center. Now we define each pixel in the jet area as the object pixel. The convex hull mask is a simply connected region, as is usual for object masks in natural images, whereas the enlargement mask may consist of several regions useful to identify clusters of constituents more precisely (Fig. 2).

In natural images, the bounding box is an intuitive and also robust concept since objects have their boundary, and the deviation of the bounding box depending on specific choices of boundary pixels is not large. On the contrary, there is no such thing as a boundary for jets, and instead, we introduced the jet area in order to alleviate the issues of the sparse object pixels and the large fluctuation of the bounding box. Nevertheless, the size and shape of the bounding box will rely on our preselection rules. It may be helpful to adjust the bounding box by involving a ground truth value irrelevant to our definition of jet areas.

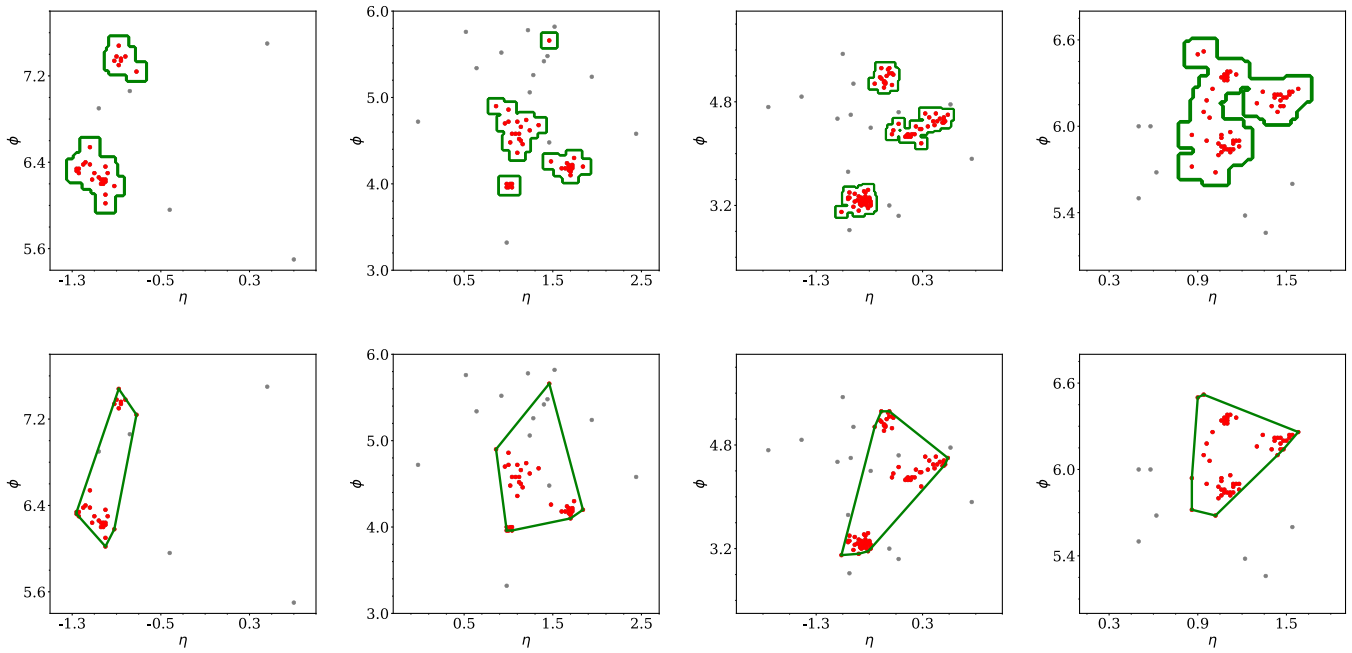

FIG. 2. Top: enlargement areas. Bottom: convex hull areas. Red dots are the constituents constructing the jet areas represented by the green lines. Gray dots are the ones ruled out by the neighborhood condition. One may see that, without the condition, how drastically few gray dots could change jet areas. Nevertheless, we spare a constituent with high p_T (>5 GeV) as shown in the second example (the red dot on the top).

Therefore, we employ the ground truth coordinates of partons, i.e., the rapidity y and azimuth ϕ, to be the center of the bounding box. The bounding box is now the smallest rectangle enclosing the jet area, while its center is fixed at the ground truth coordinates of the parton.[3] We expect that the adjustment tells the network consistent information about the keypoint of a jet around which its features should be extracted and gathered through convolutions regardless of our choice of preselection rules. Indeed, we find a significant improvement in the network performance compared to using the default bounding box.

III. MASK R-CNN AND ITS MODIFICATIONS

Mask R-CNN [50] is a state-of-the-art framework for object detection and instance segmentation. It was progressively developed from R-CNNs that first take candidate regions from separate region proposal methods, and then use a convolutional network to extract features for classifications and bounding box regressions. The original R-CNN [51] was computationally expensive as it performs a CNN for each region proposal. By extracting a feature map from the entire input image to share across region proposals, SPPnet [52] greatly reduced the computational cost, and Fast R-CNN [53] streamlined a multistage pipeline of the predecessors, as the classifier and box regressor are jointly trained with the feature extraction network. While the previous models take the region proposals from separate region proposal methods, Faster R-CNN [54] has brought them into one unified network by introducing region proposal networks (RPNs) that can share the convolutional feature map, showing remarkable gains in speed and accuracy. Finally, Mask R-CNN extends Faster R-CNN by adding a mask branch, for instance, segmentation in parallel with the classifier and box regressor.

Let us briefly review the structure of Mask R-CNN (Fig. 3). It can be mainly divided into three modules: a backbone architecture for feature extraction, an RPN for region proposal generation, and a detection head for classification, box regression, and mask segmentation. The RPN and detection head share the backbone such that the features are used for both regional proposals and detections. Although the backbone can be any convolutional architecture, Feature Pyramid Network (FPN) [55] is commonly used to take advantage of multiscale features in addition to the base architecture such as residual neural networks (ResNets) [56]. The ResNet extracts features from an input image by successively scaling down its spatial size through convolutions, which in turn produces a pyramid of outputs at several scales. The FPN then uses a top-down pathway with lateral connections to the outputs at each scale in order to build a feature pyramid. The feature map at each level of the pyramid is fed into the RPN to propose candidate bounding boxes referred to as regions of interest (RoIs). An anchor is a reference box whose center

[3]Since the approximation $y \simeq \eta$ is valid for ultrarelativistic particles, the coordinates of a parton on the $\eta \times \phi$ plane should take its y and ϕ.

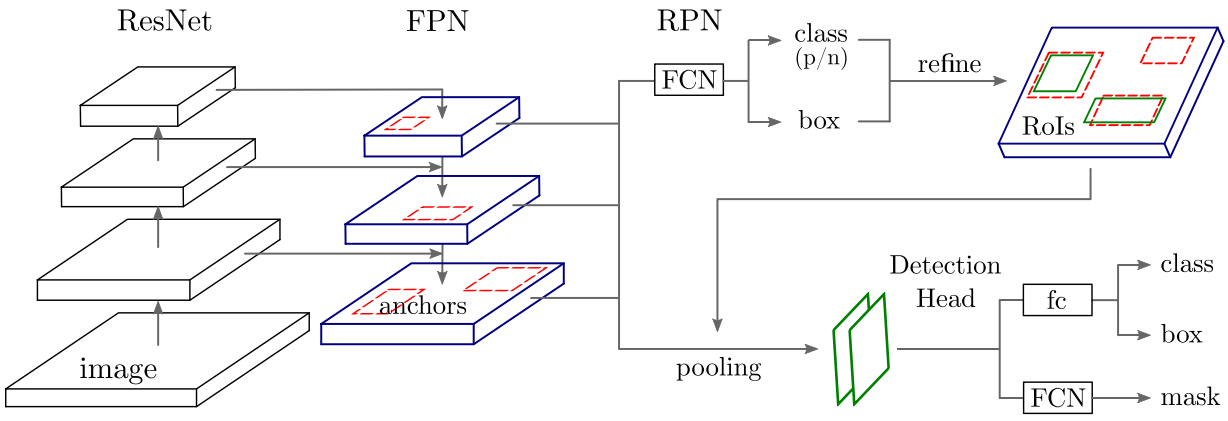

FIG. 3. Schematic diagram of Mask R-CNN, adapted from [43]. The backbone shows only three convolutional stages of ResNet instead of five for simplicity. FCN denotes a fully convolutional network, and fc represents fully connected layers.

is at a pixel of the feature pyramid. By default, every pixel has three anchors with aspect ratios of 1∶1, 1:2, and 2∶1, where the anchor scale is different according to the level of the feature pyramid. Through a small convolutional network, the RPN outputs an objectness score and box regression for each anchor. It is trained so that an anchor has a high score if its intersection over union (IoU) with a ground truth box is greater than a threshold, and the box regression refines the size and location of positive anchors to fit their ground truth boxes better. The refined positive boxes are the RoIs that are passed on to RoI pooling. While the RoIs have variable sizes due to the refinement in the RPN, the detection head requires a fixed-size input. Therefore, we pool RoI features of a fixed size by cropping the RoIs from the feature pyramid and resizing them using bilinear interpolation.[4] The RoI features are then fed to the detection head that has two branches. The classification branch consists of two fully connected layers followed by classification and box regression outputs. The mask branch is a small fully convolutional network to predict binary masks.

In the jet reconstruction task, the original Mask R-CNN can be used to tag jets and predict the jet areas. On the other hand, the jet areas, especially the convex hull areas, are susceptible to the pileup contamination since they include background particles as well. Furthermore, the preselection process to define jet areas may exclude some constituents that significantly contribute to the four-momentum of their parton. Therefore, in order to accurately obtain four-momenta of partons from jet area predictions, one needs separate methods carrying out the pileup mitigation as well as compensating for the preselection. On the contrary, instead of using separate methods, we bring a pileup mitigation and compensation network into Mask R-CNN to facilitate end-to-end jet reconstructions. In other words, we extend Mask R-CNN by adding an additional branch for predicting the mass p_T and coordinates of partons,[5] which we call the jet branch. The jet branch has the same architecture as the classification branch, i.e., two fully connected layers, and also shares the RoI inputs with the classification branch (Fig. 4). It is worth noting that we predict two boxes and two masks for each RoI (denoted by "×2" in the figure), although it is usual to predict one box and mask per class so that the default number of predictions for each RoI is 3 corresponding to the three classes (background/Higgs/top). If objects in each class have their typical shape or ratio, the class-specific prediction may be more effective. However, the Higgs jets are indistinguishable from top jets by their bounding box or mask. Therefore, we predict for two classes (background/jet) considering Higgs and top as one class. On the other hand, the jet branch outputs a single four-momentum (coordinates, p_T, and mass) for each RoI regardless of class. This reflects the fact that it is always possible to calculate the four-momentum even for a background region in a consistent way. Furthermore, this approach also can tell the network that the mass of an RoI is essential to determine its class. Consequently, the mass and class prediction tasks will be closely intertwined and enhance each other.

In our implementation of Mask R-CNN, which is based on the open-source code in [57], we also employed the Composite Backbone Network (CBNet) [58] as our backbone and Cascade R-CNN [59] as an extension of the detection head, so as to further improve accuracy. We present a brief introduction to the two architectures while

[4]We use the crop and resize operation following the source code [57] for simplicity. The original paper [50] proposed a more elaborate method called RoIAlign to reduce misalignment between the RoIs and the extracted features.

[5]In practice, the vector sum of the constituents momenta is used as ground truth instead of the momentum of the original parton in training our network, because the vector sum is more directly related to the masked constituents. However, in the training sample, the momentum difference between the vector sum and the parton is less than 2% for the Higgs class and 5% for the top class, for more than 90% of events.

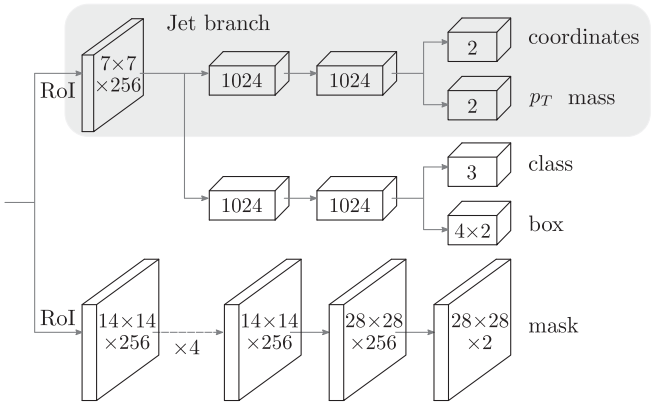

FIG. 4. Detection head together with the jet branch. Numbers inside boxes denote dimensions. ×4 represents four consecutive convolutions. The figure of classification and mask branches are adapted from [50].

referring readers to the original papers [58,59] for details. The CBNet groups multiple identical backbones by connecting them in parallel. We use CB-ResNet50, which consists of two ResNet50s, an assisting one and a lead one, connected in such a way that the features of higher-level stages in the assisting backbone flow to the lower-level stages in the lead backbone. Therefore, the lead backbone can integrate the high-level features into its low-level convolutional stages for more effective feature extractions. Cascade R-CNN extends the detection head in order to have more accurate bounding box predictions. The detector requires an IoU threshold to decide whether an RoI is positive or negative, and the commonly used threshold value is 0.5, which will be robust to poor proposals but also can be loose, leading to noisy box predictions. To address the problem, Cascade R-CNN adds detectors and constructs a sequence of detectors with increasing thresholds at each stage. We use three detectors where the first two have only the classification branch to refine RoIs, while the last detector has all three branches.

IV. NETWORK PERFORMANCE

In object detection, it is usual to evaluate the network performance using average precision (AP). Given an IoU threshold, one can obtain the precision-recall curve by varying the score threshold as shown in the Fig. 5, and AP is the area under the curve. In jet detection, however, the value itself is not directly comparable to that of other models since AP also heavily depends on the mask scheme. Nonetheless, it is a useful metric within one mask scheme, and hence we report the mask AP and box AP in Table I to show the performance in jet area detection (see also Fig. 6 for an example).

On the other hand, the main goal of the jet reconstruction task is to calculate the four-momentum of partons, which are independent of the mask scheme. Therefore, we want to measure how accurately our extended Mask R-CNN can predict the mass, p_T, and coordinates. In experimental analyses, we have prior knowledge about the type and number of jets in selecting the signal events. Let us assume we have three ground truth Higgs jets on an image for illustrative purposes. First recall that, since we use the augmented images, there may be two predictions for the same jet duplicated along the augmented ϕ coordinate. To sort this out, we postprocess the output by taking it back to the ϕ range of one period and eliminating one with a lower score if two predictions are of the same class and their mask overlap is large. After the postprocessing, we choose three predictions of the Higgs boson with the highest scores if there are more than three Higgs predictions. Otherwise, we take all available Higgs predictions. A prediction is a true positive if the distance between the predicted and ground truth coordinates is less than a threshold (we use 30 pixels $\simeq 0.6$). We measure the coordinates, mass, and

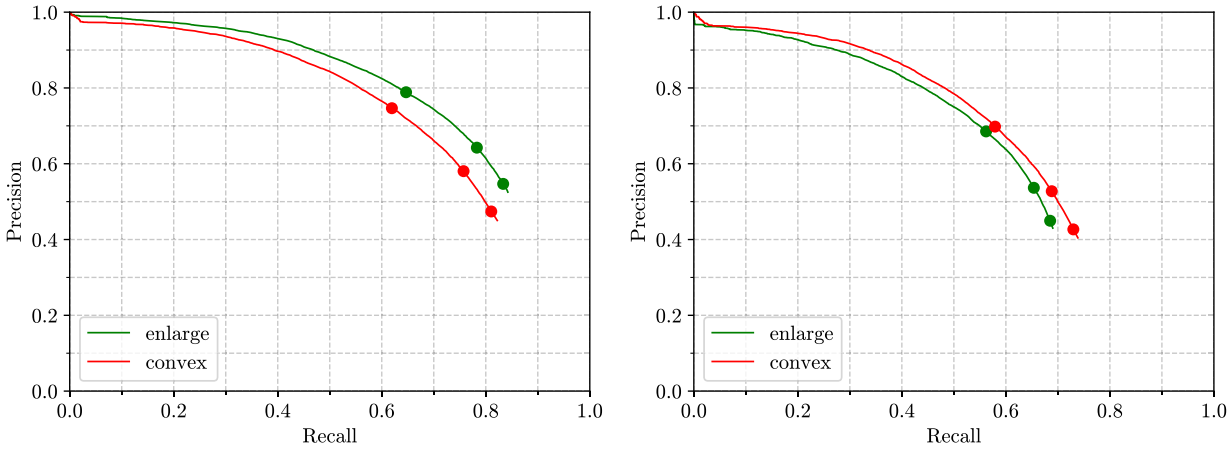

FIG. 5. Precision-recall curve at IoU $= 0.5$ for the bounding box (left) and mask (right). Three dots on each curve indicate the points where the score threshold is 0.9, 0.7, and 0.5 from left to right.

TABLE I. AP (%) on the $Ht\bar{t}$ test set. AP_{50} denotes the AP at IoU = 0.5, and AP denotes the AP averaged across the IoU from 0.5 to 0.95 with a step size of 0.05.

	AP^{mask}	AP^{mask}_{50}	AP^{box}	AP^{box}_{50}
Convex hull	24.7	60.7	37.1	68.9
Enlargement	16.0	56.6	43.0	73.5

p_T differences between the true positive prediction and its ground truth (obtained by the vector sum of the constituents momenta).

To gain an intuitive understanding of the network performance, we contrast our results with those from existing state-of-the-art object tagging architecture. We adopt LorentzNet for classification and PELICAN for momentum regression as an illustration. The validations for the application of these two networks are provided in the Appendix. These networks require a jet clustering algorithm to localize jets for a given event image, as they use a jet as the input. We retrain these networks on the Higgs boson and top quark jets from our 0.3×10^6 $Ht\bar{t}$ event sample (with an average number of 50 pileup events superposed on each signal event), where the Higgs boson ($p_{T_H} > 200$ GeV) and top quark ($p_{T_t} > 300$ GeV) jets are reconstructed by the anti-k_T algorithm with cone-size parameter R varying from 0.8 to 1.8 in steps of 0.2. The detector effects have been ignored in this comparison study.

It should be noted that LorentzNet and PELICAN follow a fixed input format: a set of four momenta of the jet constituents. To simplify the training, we construct a training file for the reconstructed anti-k_T jets mentioned above, mainly containing the following information.

(i) Nobj: the count of jet constituents.
(ii) Pmu: the four-momentum of jet constituents (E, p_x, p_y, p_z) sorted in descending order of p_T. This part serves as the network input, with a shape of [N × 4], where N(= 200) is the maximum number of jet constituents for a single input, and the insufficient parts are padded with zeros.
(iii) label: indicates whether Pmu at a certain position is a constituent(1) or a padding value(0), with a shape of [1 × 200].
(iv) truth_Pmu: the four-momentum of the parton to which the jet belongs, used for momentum prediction.
(v) is_signal: the type of jet determined by the distance from jet to parton. For example, if it is closest to the Higgs jet (i.e., when $\Delta R = \sqrt{\Delta\phi^2 + \Delta y^2}$ is the smallest), it is classified as a Higgs jet(0), and top jet(1) is in a similar situation. This value is used for jet classification.
(vi) mass: the invariant mass of the parton to which the jet belongs.

The above file format refers to that found in Ref. [60], and there are some other configurations mentioned in this reference, which are irrelevant to the training. For more

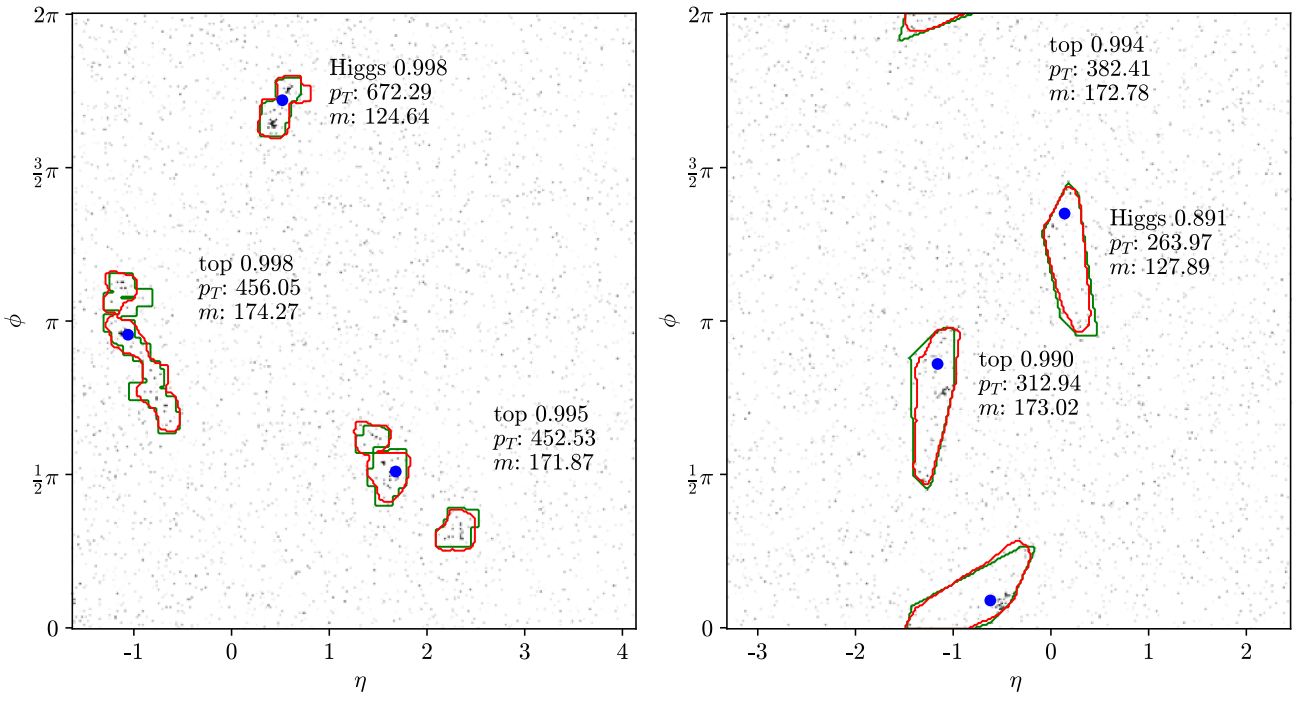

FIG. 6. Prediction examples. Left: enlargement mask. Right: convex hull mask. Green and red lines represent ground truth and predicted masks, respectively. Labels include score, p_T, and mass predictions. Blue dots denote predicted coordinates.

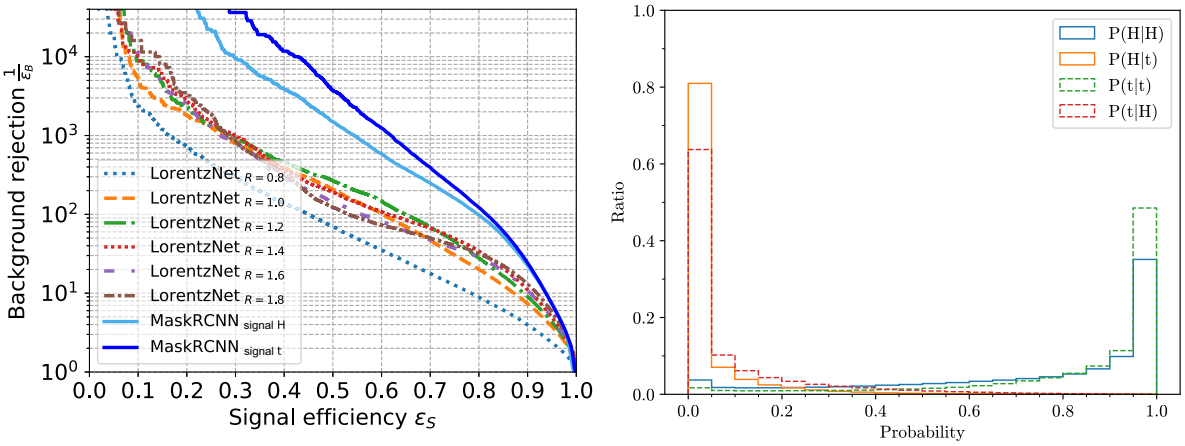

FIG. 7. Left: ROC curves for LorentzNet with different cone-size parameters (the top quark jet is taken as the signal and the Higgs jet is the background) and for Mask R-CNN (the signal Higgs boson and background top in cyan color; the signal top and background Higgs boson in blue color). Right: the distributions for the classification scores in Mask R-CNN.

details, please refer to Ref. [60]. The hyperparameters for both networks are set to the default values as shown in Refs [35,36].

The receiver operating characteristic (ROC) curves of LorentzNet for the top quark (signal) and Higgs (background) jet discrimination with different cone sizes are shown in the left panel of Fig. 7. It can be observed that the cone-size parameter has non-negligible effects on the performance of LorentzNet. In particular, the performance degrades dramatically when the cone size becomes too small to fully capture the jet constituents. On the other hand, a larger cone size does not impair jet classification much, despite distorting the jet shape. The performance of those classifiers is not directly comparable to that of Mask R-CNN. However, we may consider only the classification part of Mask R-CNN to obtain the ROC curve. In Mask R-CNN, the RPN does a job similar to jet clustering algorithms. The RPN in our hyperparameter setting proposes up to 500 RoIs for each image,[6] which will be passed to the detection head. To calculate the ROC, we first get all the RoIs of an event from the RPN and compute the IoUs between the RoIs and ground truth bounding box. The RoIs with IoUs greater than 0.5 are selected and fed into the classifier branch. The RoI with the highest classification score is chosen for calculating the ROC curve of the Mask R-CNN. As our network predicts three classes, we provide two ROC curves, one of which considers the Higgs jet as the positive class and the top quark jet as the negative class, while the other considers the opposite. The ROC curves are presented in Fig. 7. Mask R-CNN achieves higher performance than LorentzNet in general, due to its more accurate jet boundaries. Those features can be quantified in terms of AUC values, as given in Table II. In addition, Mask R-CNN can tag the top quark jet more accurately than the Higgs jet in its trinary classification. The right panel of Fig. 7 illustrates this fact with the classification scores distributions, where $P(H|H)$, $P(H|t)$, $P(t|t)$, and $P(t|H)$ are the probabilities of tagging a true Higgs boson as a Higgs boson, a true top as a Higgs boson, a true top as a top, and a true Higgs boson as a top, respectively.

A. The reconstruction accuracy for the test sample

The reconstruction accuracies of the Higgs boson and top four-momenta obtained from applying the PELICAN and the Mask R-CNN methods to the test sample are illustrated in Fig. 8 in terms of two-dimensional distributions on the $\Delta y \times \Delta\phi$ plane and $\frac{\Delta m}{m} \times \frac{\Delta p_T}{p_T}$ plane. Here, $\Delta m, \Delta p_T, \Delta y$, and $\Delta\phi$ correspond to the differences of the invariant mass, transverse momentum, rapidity, and azimuth angle between the reconstructed jet and the ground truth jet. It should be noted that the four-momentum of the ground truth jet is obtained by the vector sum of its constituent momenta. The m and p_T in the denominator correspond to the values of the ground truth jet. The distributions are normalized such that the sum of all simulated events for each process is equal to 1. In the figure, the red and pink contours correspond to the distributions of the PELICAN method with cone size $R = 1.0$ and $R = 1.4$, respectively. The blue contours and gray shades correspond to the distributions of the Mask R-CNN methods with the convex hull mask and the enlargement mask, respectively. Different shades of gray regions and colored contours from inside out indicate 20%, 40%, and 60% of events, respectively. The closer they are to the center, the higher accuracy they stand for. It can be observed that the Mask R-CNN methods with different definitions of mask can achieve similar accuracies in both

[6]Overlaps between RoIs are allowed as long as the IoU is less than a threshold (we use 0.7) such that several RoIs indicate one jet in general.

TABLE II. AUC values for LorentzNet with different jet cone sizes and for Mask R-CNN with either the Higgs boson or top being the signal.

	LorentzNet						Mask R-CNN	
	$R = 0.8$	$R = 1.0$	$R = 1.2$	$R = 1.4$	$R = 1.6$	$R = 1.8$	Signal Higgs boson	Signal top
AUC	0.920	0.953	0.960	0.964	0.962	0.966	0.9723	0.9754

Higgs boson and top jet momenta reconstruction. For the test sample of the $H\bar{t}t$ process, about 60% of Higgs jets can be reconstructed with $\Delta y \sim \Delta\phi \in [-0.06, 0.06]$, $\frac{\Delta m}{m} \in [-0.08, 0.05]$, and $\frac{\Delta p_T}{p_T} \in [-0.14, 0.1]$, and about 60% of top jets can be reconstructed with $\Delta y \sim \Delta\phi \in [-0.05, 0.05]$, $\frac{\Delta m}{m} \in [-0.07, 0.05]$, and $\frac{\Delta p_T}{p_T} \in [-0.11, 0.08]$. The Mask R-CNN method surpasses the PELICAN method in regressing both Higgs boson and top

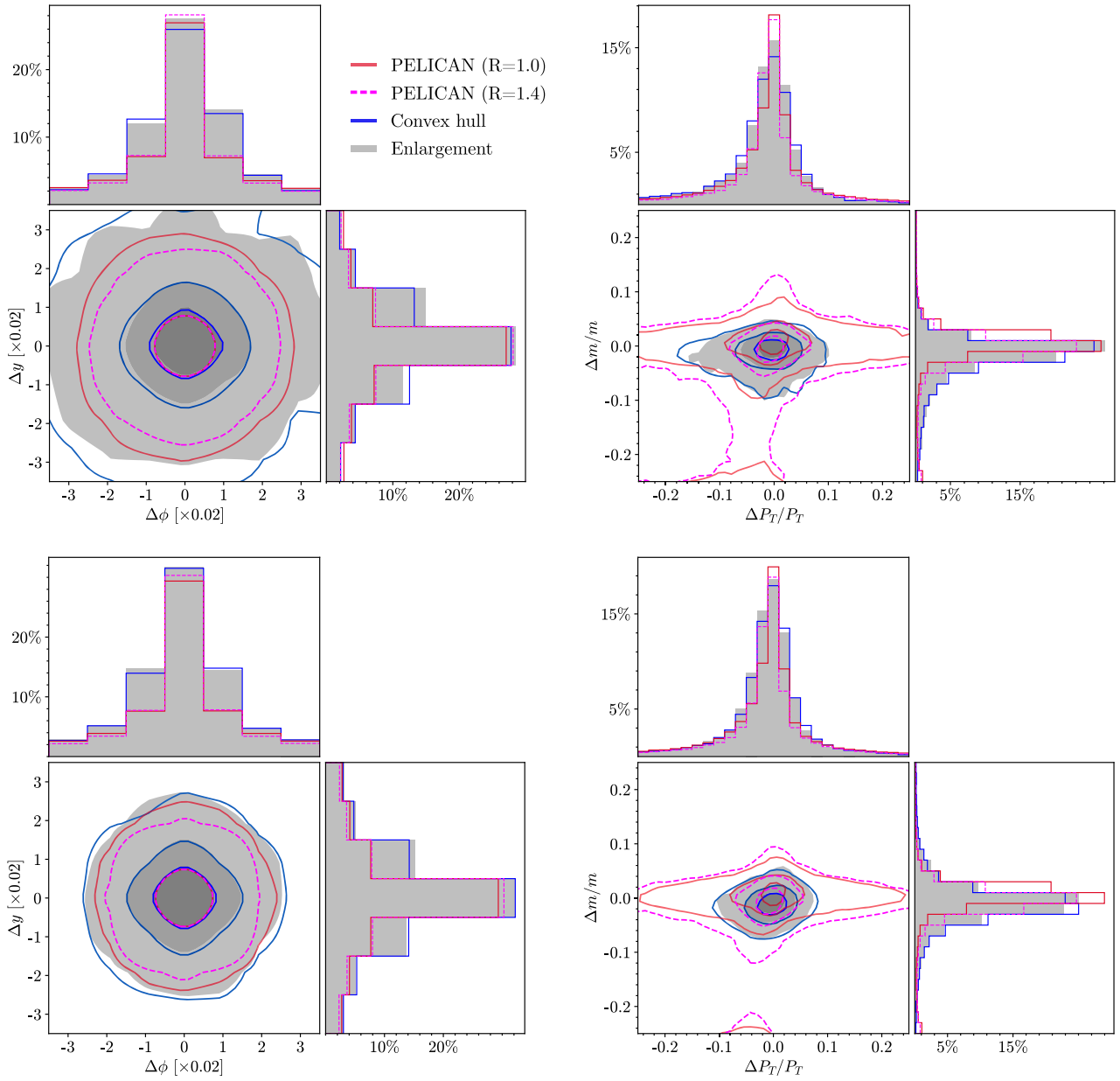

FIG. 8. The two-dimensional event distributions of the $H\bar{t}\bar{t}$ event sample. Contours/shade from the inside out correspond to the event fraction of 20%, 40%, and 60%, respectively. Upper panels: reconstructed Higgs jets. Lower panels: reconstructed top quark jets.

TABLE III. The values of RMSD for given percentages (from 10% to 50%) of the most accurate events of the $Ht\bar{t}$ process. The results for the Higgs boson and top obtained from the enlarged Mask R-CNN method and PELICAN method are shown.

	$Ht\bar{t}$			
	Higgs		Top	
	Enlarge	PELICAN	Enlarge	PELICAN
10%	0.0098	0.0082	0.0100	0.0077
20%	0.0150	0.0132	0.0147	0.0122
30%	0.0198	0.0189	0.0192	0.0173
40%	0.0251	0.0272	0.0240	0.0240
50%	0.0322	0.0424	0.0295	0.0353

momenta for any choice of the cone-size parameter, when 60% of most accurate events are considered. Note that the PELICAN efficiencies of successful Higgs boson and top tagging on $Ht\bar{t}$ events are around 78% for $R = 0.8$ and around 85% for other cone sizes.

To give a quantitative comparison between the performance of Mask R-CNN and PELICAN, we calculated the root-mean-square distance (RMSD) using the following method. We note that using RMS in the usual way is not available since many of the predictions are false positives, which are outliers. From the prediction of two networks, we first sort the events by the distance of $(\Delta p_T/p_T, \Delta m/m)$ from the origin (0, 0), i.e., sort the events according to the p_T and m prediction accuracies. Then, we calculate the RMSD for 10%, 20%, 30%, 40%, and 50% of the most accurate events. The results of the test sample are given in Table III. In the sense of RMSD, the fraction of events when Mask RCNN surpasses PELICAN is 35% for Higgs momentum reconstruction and 40% for top momentum reconstruction, respectively. In other words, those two methods have the same RMSD on Higgs (top) momentum reconstruction when the first 35% (40%) of events with the highest accuracy are considered.

Compared to the results in our earlier works [42,43] where the reconstructed jet momentum is calculated by the vector sum of momenta of all marked particles so that the accuracy is highly affected by the pileup contamination, the method in this work adopts an independent jet branch to predict the ground truth jet momentum. It turns out that the jet branch has been trained to implement pileup mitigation in an automatic and efficient way. No further pileup mitigation procedures are required.

B. The detector and pileup effects

Although the network is trained on Monte Carlo events without detector effects, it is supposed to also work well on events where detector effects are included. For illustration, we apply our network (which is trained on events without detector effects) to events where the energy of final state particles is Gaussian smeared with the standard deviation varying from 1% to 20% of the total energy. In terms of RMSD as discussed above, the results of the Higgs boson and top reconstruction accuracy for events with different standard deviations are given in Tables IV and V. We find that the reduction of reconstruction accuracy due to the detector effect is mild, especially for the top. Given a specific detector configuration, it is also possible for the network to learn the dedicated detector effects (which render the assignment of Higgs boson/top constituents

TABLE IV. The values of RMSD for given percentages (from 10% to 50%) of the most accurately predicted Higgs boson in the $Ht\bar{t}$ process. Different columns correspond to the different standard deviations that are taken in the Gaussian smearing of the jet constituent energy E. The second column gives the RMSD values without detector smearing (DS). Results are obtained by the enlarged Mask R-CNN method.

	Without DS	$0.01E$	$0.02E$	$0.04E$	$0.08E$	$0.1E$	$0.12E$	$0.14E$	$0.18E$	$0.2E$
10%	0.0098	0.0098	0.0101	0.0110	0.0129	0.0133	0.0143	0.0152	0.0167	0.0173
20%	0.0150	0.0150	0.0153	0.0165	0.0189	0.0197	0.0214	0.0225	0.0250	0.0260
30%	0.0198	0.0203	0.0202	0.0219	0.0242	0.0258	0.0277	0.0291	0.0325	0.0341
40%	0.0251	0.0261	0.0259	0.0279	0.0302	0.0322	0.0343	0.0362	0.0404	0.0424
50%	0.0322	0.0337	0.0333	0.0354	0.0372	0.0400	0.0421	0.0444	0.0496	0.0520

TABLE V. Same as Table IV, for the predicted top in the $Ht\bar{t}$ process.

	Without DS	$0.01E$	$0.02E$	$0.04E$	$0.08E$	$0.1E$	$0.12E$	$0.14E$	$0.18E$	$0.2E$
10%	0.0099	0.0095	0.0100	0.0101	0.0105	0.0101	0.0108	0.0107	0.0112	0.0113
20%	0.0147	0.0144	0.0149	0.0151	0.0154	0.0152	0.0160	0.0162	0.0167	0.0170
30%	0.0192	0.0188	0.0194	0.0196	0.0201	0.0200	0.0210	0.0212	0.0219	0.0224
40%	0.0239	0.0235	0.0243	0.0246	0.0252	0.0251	0.0261	0.0265	0.0275	0.0282
50%	0.0295	0.0292	0.0302	0.0306	0.0312	0.0310	0.0324	0.0330	0.0342	0.0351

ambiguous) by training the network on particle-gun MC events (which contain only a single top quark or Higgs boson in the final state) where the detector effects are included. This can help to mitigate the detector smearing on the momentum precision to some extent. However, the possible drawbacks of such a procedure are that the method could become detector dependent and may not be able to learn the feature for the overlapped case as will be discussed later.

Another important practical issue is the pileup effects during the collision. At the High-Luminosity LHC, the average number of pileup interactions per bunch crossing can reach $\langle\mu\rangle \sim 140$–$200$. On the other hand, there are pileup mitigation algorithms based on vertex and calorimeter information, which help to suppress the pileup effects in the final data. Irrespective of a specific pileup mitigation method, we apply our network (which is trained on events with 50 pileups denoted by network@PU50) to events with pileup levels varying from 5 to 200 (denoted by PU5 to PU200). The results are given in Table VI. The AUC of Mask R-CNN is barely changing for events with pileups smaller than 50, and it decreases steadily with increasing the pileup for $\langle\mu\rangle \gtrsim 50$. Moreover, we further fine-tune the network@PU50 on 300 000 events with 200 pileups and obtain the network@PU200 version of Mask R-CNN. In terms of AUC values, the upgraded network has stable performance on events with pileup level up to 200. Meanwhile, the performance is comparable to that of the network@PU50 for low pileup events.

C. Tests on different processes

Although the networks have been trained with event samples of the $Ht\bar{t}$ process, they can be used as general Higgs boson and top jets taggers for events of many other processes. For demonstration purposes, we showcase their capabilities in other processes at the LHC that produce Higgs bosons and top jets: (1) $pp \to HHt\bar{t}$ in the SM, (2) neutral Higgsino pair production with subsequent decay $\tilde{\chi}_2^0 \to H\tilde{\chi}_1^0$ in the SUSY model, (3) top squark pair production with subsequent decay $\tilde{t} \to t\tilde{\chi}_1^0$, and (4) $pp \to t\bar{t}t\bar{t}$ in the SM. The corresponding accuracy contours are shown in Figs. 9 and 10. In all cases, we find that the network performance only slightly depends on the definition of the mask. The accuracies for all of the variables (m, p_T, y, ϕ) are always slightly worse than those of the $Ht\bar{t}$ process. Generally speaking, about 40% of the Higgs bosons/top jets in those test samples can be reconstructed with $\Delta y \sim \Delta\phi \in [-0.04, 0.04]$, $\frac{\Delta m}{m} \in [-0.08, 0.05]$, and $\frac{\Delta p_T}{p_T} \in [-0.1, 0.07]$. The Higgsino pair $(\tilde{\chi}_2^0\tilde{\chi}_2^0)$ events contain the fewest detectable particles in the final state so that higher accuracies can be obtained for the momentum variables. For the stop pair process, only the masses of the stop and neutralino are set. There are $\mathcal{O}(10\%)$ of events containing a top quark with momentum less than 300 GeV, leading to lower reconstruction efficiencies and decreased accuracies of the momentum variables. The mass predictions for the Higgs boson/top are generally lower for all cases, and the situation is more severe for the stop pair and neutralino pair processes because the network tends to predict smaller jet masks, which may drop some jet constituents with non-negligible energy. In the stop pair and neutralino pair processes, such an operation happens more frequently, because there are a certain fraction of events with $p_T(H) < 200$ GeV and $p_T(t) < 300$ GeV. The quantitative results in terms of RMSD for those processes are provided in Table VII. The fraction of events when Mask RCNN

TABLE VI. AUC values of two versions of the enlarged Mask R-CNN being tested on an event sample with different pileup levels. Besides the version that is used throughout the paper (denoted by network@PU50, because it is trained on events with 50 pileups), the network@PU200 version of Mask R-CNN is obtained by further training the network@PU50 on 300 000 events with 200 pileups.

		PU5	PU10	PU20	PU30	PU50	PU80	PU100	PU120	PU150	PU180	PU200
network@PU50	Signal Higgs boson	0.9724	0.9728	0.9729	0.9732	0.9723	0.9670	0.9589	0.9446	0.9016	0.8051	0.7037
	Signal top	0.9743	0.9746	0.9751	0.9756	0.9754	0.9718	0.9659	0.9547	0.9228	0.8732	0.8323
network@PU200	Signal Higgs boson	0.9609	0.9620	0.9636	0.9656	0.9678	0.9691	0.9696	0.9702	0.9705	0.9697	0.9691
	Signal top	0.9684	0.9690	0.9701	0.9710	0.9723	0.9737	0.9741	0.9744	0.9744	0.9743	0.9740

TABLE VII. Same as Table III, for the $HHt\bar{t}$, $\tilde{\chi}_2^0\tilde{\chi}_2^0$, $\tilde{t}\bar{\tilde{t}}$, and $t\bar{t}t\bar{t}$ processes, respectively.

	Higgs boson of $HHt\bar{t}$		Top of $HHt\bar{t}$		$\tilde{\chi}_2^0\tilde{\chi}_2^0$		$\tilde{t}\bar{\tilde{t}}$		$t\bar{t}t\bar{t}$	
	Enlarge	PELICAN	Enlarge	PELICAN	Enlarge	PELICAN	Enlarge	PELICAN	Enlarge	PELICAN
10%	0.0133	0.0119	0.0127	0.0110	0.0133	0.0080	0.0184	0.0160	0.0143	0.0152
20%	0.0210	0.0221	0.0197	0.0198	0.0207	0.0139	0.0271	0.0261	0.0221	0.0315
30%	0.0292	0.0416	0.0268	0.0342	0.0284	0.0236	0.0364	0.0399	0.0306	0.0938
40%	0.0405	0.0804	0.0359	0.0616	0.0376	0.0504	0.0489	0.0698	0.0413	0.1712
50%	0.0593	0.1497	0.0485	0.1108	0.0508	0.1227	0.0731	0.1254	0.0561	0.2438

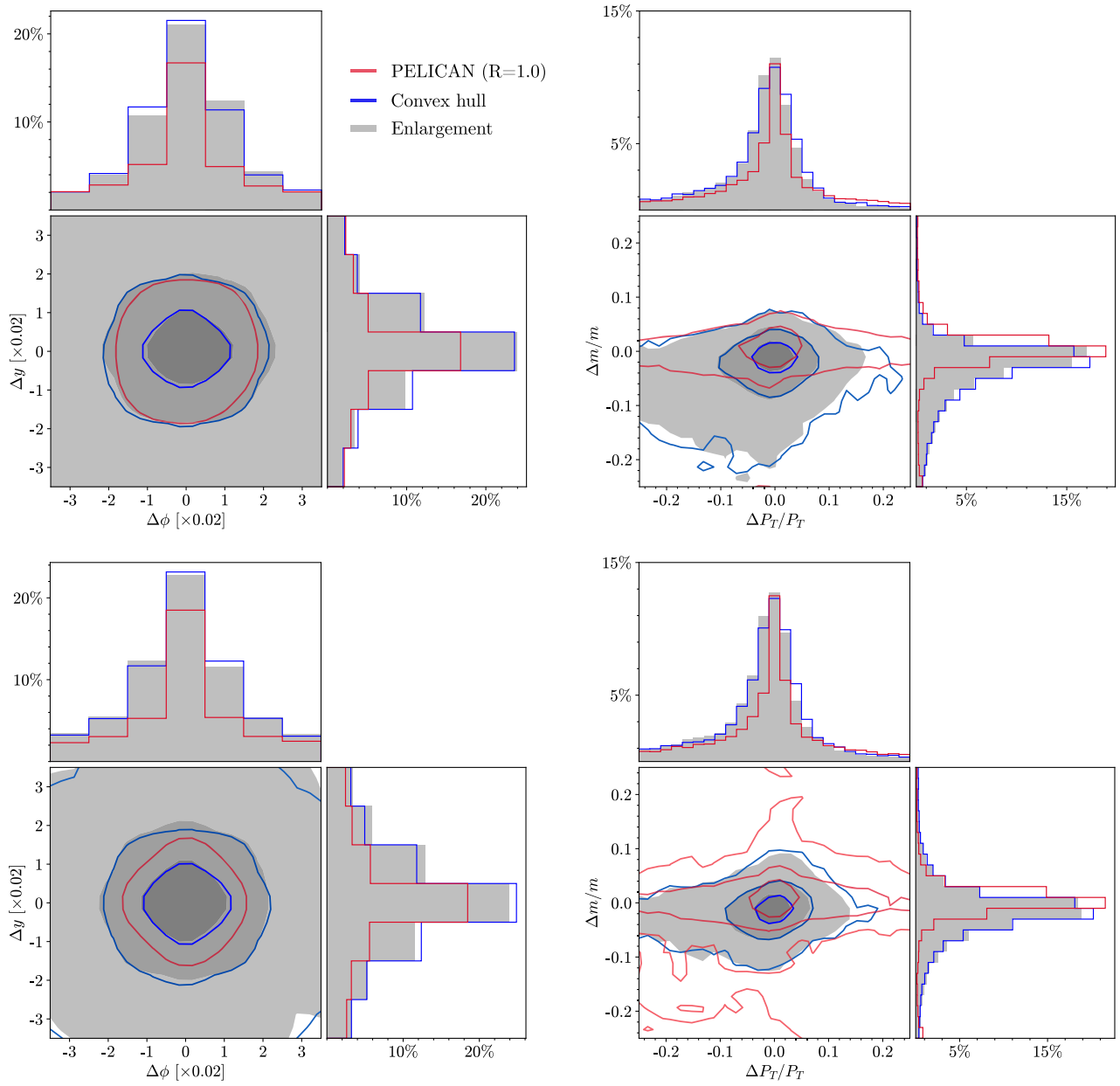

FIG. 9. Similar to Fig. 8 for the reconstructed Higgs jet (upper panels) and top jet (lower panels) in the $HHt\bar{t}$ event sample.

surpasses PELICAN is 18% for Higgs reconstruction in the $HHt\bar{t}$ process, 20% for top reconstruction in the $HHt\bar{t}$ process, 35% for Higgs reconstruction in the $\tilde{\chi}_2^0\tilde{\chi}_2^0$ process, 24% for top reconstruction in the $\tilde{t}\,\bar{\tilde{t}}$ process, and 7% for top reconstruction in the $t\bar{t}t\bar{t}$ process, respectively.

Since there are multiple jets on an image, some of the jet areas can be very close to each other or even overlap each other. In such cases, the sequential recombination algorithms with a large cone size will end up with merging nearby or overlapping jets into one jet. In fact, when two final state particles are close to each other, it is practically impossible to tell if they originate from different ancestors. Therefore, the overlapping jets will also greatly impede the network accuracy. To see how much it affects the accuracy, the network is tested on the $t\bar{t}t\bar{t}$ event sample with a separation condition in which the ground truth coordinates are at least 50 pixels ($\Delta R \simeq 1$) away from each other such that a large overlap almost never occurs. The momenta reconstruction result is shown as dashed lines in the lower panels of Fig. 10 for comparison. On the other hand, unlike the PELICAN method, which requires the jet clustering, the network is still mostly capable of finding the correct number of jets. It seems that the network can identify if there are constituents of multiple jets in an RoI and outputs the most plausible overlap configuration based on what it has seen in the training sample, although the prediction is less accurate as shown in Fig. 11, for instance.

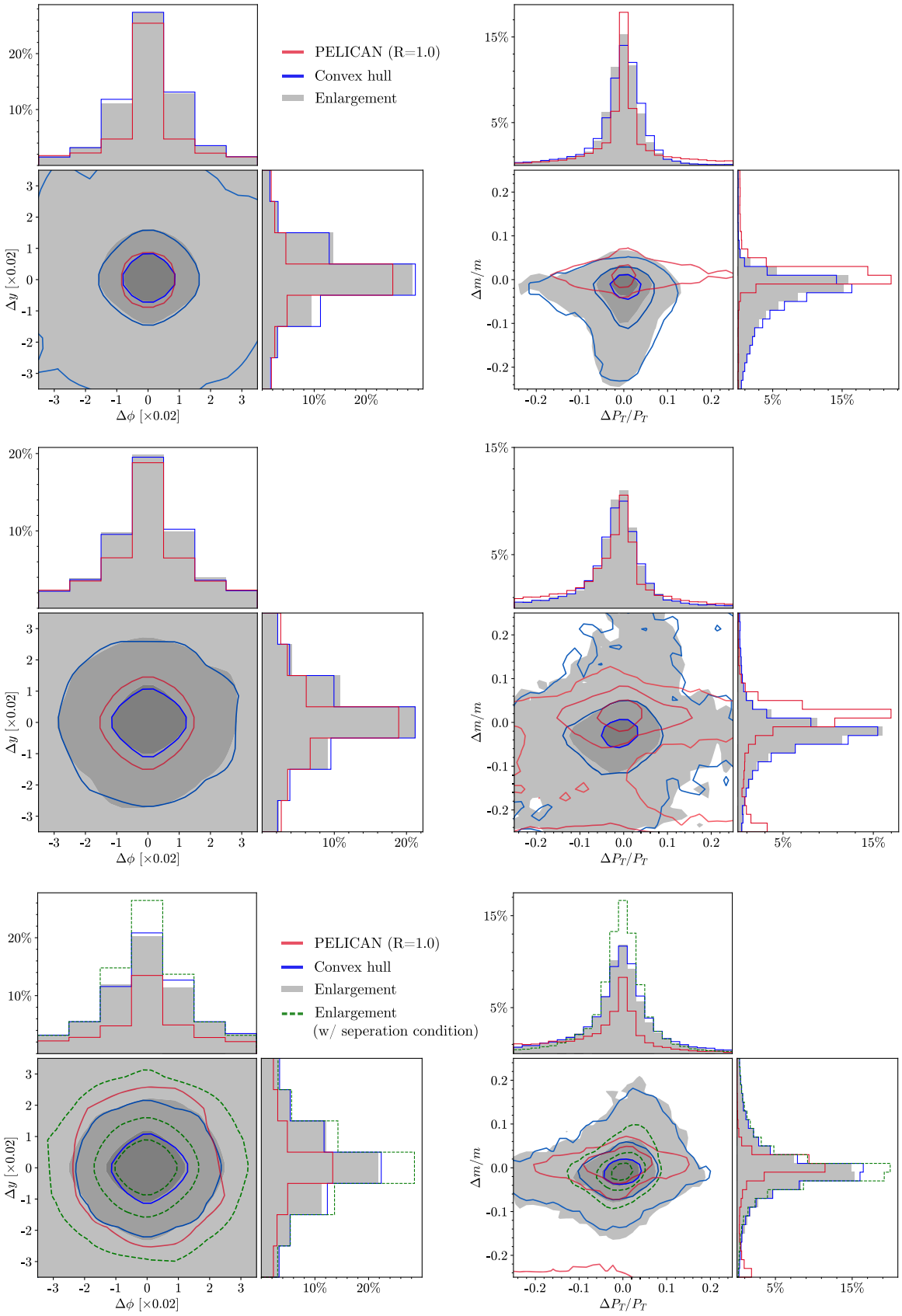

FIG. 10. Similar to Fig. 8 for reconstructed Higgs jets in the $\tilde{\chi}_2^0\tilde{\chi}_2^0$ event sample (upper panels), reconstructed top jets in the $\tilde{t}\,\bar{\tilde{t}}$ event sample (middle panels), and reconstructed top jets in the $t\bar{t}t\bar{t}$ event sample (lower panels).

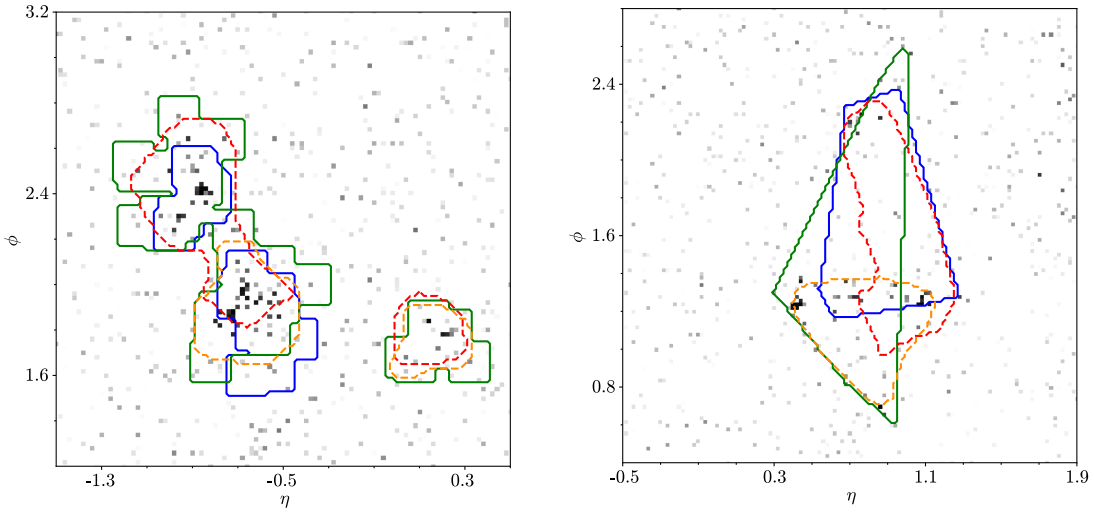

FIG. 11. Solid lines indicate ground truth masks and dashed lines represent predicted masks. Overlapping jet areas make the network less accurate.

Finally, we apply our network to events of the QCD multijet process $pp \to jjj$ where each jet in the final state is required to have $p_T > 200$ GeV. Mask R-CNN predicts only 14 Higgs jets and zero top quark jets out of 10 000 images proving its high rejection rate for pure QCD jets. For comparison, the same event sample is tested by the BDRS Higgs tagging algorithm and the HEPTopTagger top tagging algorithm. With the default setting of the jet substructure parameter as given in Refs. [13,14], the BDRS method predicts ~600 Higgs jets with mass in $[120\ \text{GeV}, 130\ \text{GeV}]$, and the HepTopTagger predicts ~550 top jets with mass in $[150\ \text{GeV}, 200\ \text{GeV}]$, among the 10 000 event sample.

V. CONCLUSION

This work aims to build a deep neural network to label the constituents of target jets among hadronic final state particles based on supervised learning. In particular, the hadronic final state of the top quark cannot be identified unambiguously according to the Monte Carlo simulation due to the color interconnection in hadronization. We propose an algorithm based on the ground truth information as well as the angular distance measure to determine the top quark jet constituents.

The Mask R-CNN framework is adopted to detect the Higgs jets and the top quark jets in collision events at the LHC. The network can predict the shape (or mask) of target jets of different kinds on event images. The definition of jet shape/mask is not unique, even from a theoretical perspective. Two schemes of mask definition are proposed in this work: enlargement mask and convex hull mask. More importantly, an additional jet branch is built for predicting the four-momenta of the original partons, in which the pileup mitigation is intrinsically implemented.

The network is trained on events of the $Ht\bar{t}$ process at the LHC, where the transverse momenta of the Higgs boson and top quarks are required to be $p_T(H) > 200$ GeV and $p_T(t/\bar{t}) > 300$ GeV. Each event is overlaid with an average number of $\langle\mu\rangle = 50$ pileup events. Compared with LorentzNet for jet classification and the PELICAN network for jet momentum regression, Mask R-CNN can detect and reconstruct both the Higgs boson and top jets in a more efficient and accurate way, mainly because it predicts jets with more accurate boundaries. The networks with two different definitions of the mask have similar performance. In terms of two-dimensional distributions on the $\Delta y \times \Delta\phi$ plane and $\frac{\Delta m}{m} \times \frac{\Delta p_T}{p_T}$ plane, about 60% of Higgs jets can be reconstructed with $\Delta y \sim \Delta\phi \in [-0.06, 0.06]$, $\frac{\Delta m}{m} \in [-0.08, 0.05]$, and $\frac{\Delta p_T}{p_T} \in [-0.14, 0.1]$, and about 60% of top jets can be reconstructed with $\Delta y \sim \Delta\phi \in [-0.05, 0.05]$, $\frac{\Delta m}{m} \in [-0.07, 0.5]$, and $\frac{\Delta p_T}{p_T} \in [-0.11, 0.08]$.

The generality of the method is demonstrated by applying the Mask R-CNN method to processes different from the trained one, including (1) $pp \to HHt\bar{t}$ in the SM, (2) $\tilde{\chi}_2^0\tilde{\chi}_2^0$ production with decay $\tilde{\chi}_2^0 \to H\tilde{\chi}_1^0$ in thet SUSY model, (3) $\tilde{t}\bar{\tilde{t}}$ production with decay $\tilde{t} \to t\tilde{\chi}_1^0$, and (4) $pp \to t\bar{t}t\bar{t}$ in the SM. In all cases, we find the dependence of the network performance on the mask definition is little, and the network outperforms the PELICAN method in the accuracy of momenta reconstruction, especially for processes with higher visible final state multiplicity. In general, the performance is slightly worse than that for the $Ht\bar{t}$ process. About 40% of the Higgs boson/top jets in those test samples can be reconstructed with $\Delta y \sim \Delta\phi \in [-0.04, 0.04]$, $\frac{\Delta m}{m} \in [-0.08, 0.05]$, and $\frac{\Delta p_T}{p_T} \in [-0.1, 0.07]$.

Moreover, we show that the network is capable of detecting the target jets even when they overlap each other on the event image, although the accuracy of the reconstructed momentum is degraded. The network exhibits high background jet rejection power when applied to events of the QCD multijet process.

Although we focus on the generalization capability to other processes in this work, conversely one may have the network to specialize in a particular process through the transfer learning. By training only the detection head with a small dataset of a certain process, the accuracy on the process increases while the network becomes rapidly insensitive to other processes. The Mask R-CNN method proposed in this work can be used to detect the boosted Higgs boson/top at the hardware trigger when being loaded to FPGA. Meanwhile, this method can also supplement the conventional analysis by detecting the Higgs boson/top and removing the Higgs boson/top constituent before applying a usual jet clustering algorithm. In the future, we will try to generalize this method to detect all kinds of jets in collider events, then it can simply replace the jet clustering and identification algorithms in the conventional data analysis.

ACKNOWLEDGMENTS

We are grateful to Alexander Bogatskiy for providing the updated version of PELICAN. This work was supported by the Natural Science Foundation of Sichuan Province under Grant No. 2023NSFSC1329 and the National Natural Science Foundation of China under Grant No. 11905149. S. C. was supported by the Fundamental Research Funds for the Central Universities, Sichuan University Full-Time Postdoctoral Research and Development Fund (Grant No. 2022SCU12118).

APPENDIX: VALIDATION OF LORENTZNET AND PELICAN

To validate the applications of LorentzNet and PELICAN, we try to reproduce the results of Refs. [35,36]. LorentzNet and the classification component of PELICAN are trained and tested on datasets provided in Ref. [61], and both models were trained by the commands given on their github.[7,8]

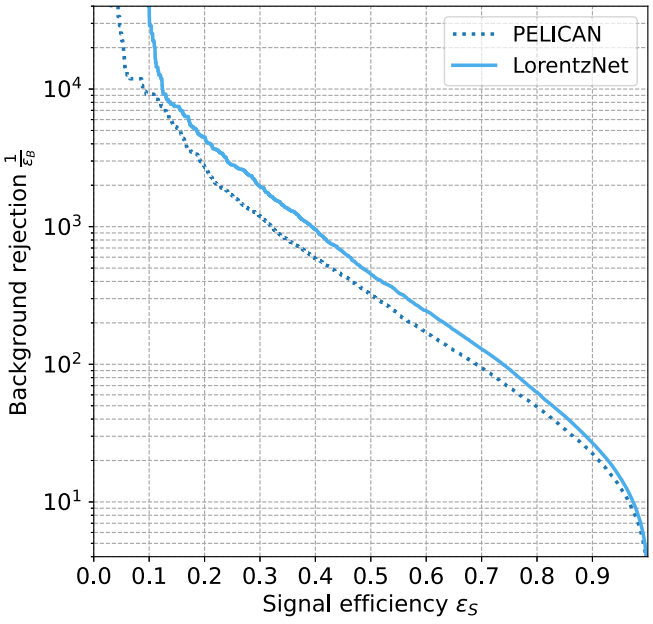

FIG. 12. ROC curves for top quark jet tagging with LorentzNet and PELICAN.

The ROC curves of both methods are shown in Fig. 12, for which the signal is the top quark jet and the background is the gluon/light-flavor jets. The [accuracy, AUC] are [0.9417, 0.9865] for LorentzNet and [0.9362, 0.9837] for PELICAN, respectively. The slight degradation of performance for PELICAN is attributed to the limited number of training samples (60 000 events are used).

We also train PELICAN to predict the four-momentum of a W boson for the case without detector effects, as described in Ref. [36]. The regression component of PELICAN is trained and tested on the dataset provided in Ref. [60]. The resolutions that we have obtained are $\sigma_{p_T} = 1.12\%$, $\sigma_m = 1.5\%$, $\sigma_\Psi = 0.55$. Those numbers are close to the values in Ref. [36].

[7]LorentzNet: https://github.com/sdogsq/LorentzNet-release.
[8]PELICAN: https://github.com/abogatskiy/PELICAN.

[1] G. P. Salam, Towards jetography, Eur. Phys. J. C **67**, 637 (2010).
[2] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, Better jet clustering algorithms, J. High Energy Phys. 08 (1997) 001.
[3] M. Cacciari, G. P. Salam, and G. Soyez, The anti-k_t jet clustering algorithm, J. High Energy Phys. 04 (2008) 063.
[4] M. Czakon, A. Mitov, and R. Poncelet, Infrared-safe flavoured anti-k_T jets, J. High Energy Phys. 04 (2023) 138.

[5] R. Gauld, A. Huss, and G. Stagnitto, Flavor identification of reconstructed hadronic jets, Phys. Rev. Lett. **130**, 161901 (2023).
[6] F. Caola, R. Grabarczyk, M. L. Hutt, G. P. Salam, L. Scyboz, and J. Thaler, Flavoured jets with exact anti-k_t kinematics and tests of infrared and collinear safety, Phys. Rev. D **108**, 094010 (2023).
[7] A. Abdesselam *et al.*, Boosted objects: A probe of beyond the standard model physics, Eur. Phys. J. C **71**, 1661 (2011).
[8] A. Altheimer *et al.*, Jet substructure at the tevatron and LHC: New results, new tools, new benchmarks, J. Phys. G **39**, 063001 (2012).
[9] 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).
[10] D. Adams *et al.*, Towards an understanding of the correlations in jet substructure, Eur. Phys. J. C **75**, 409 (2015).
[11] 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).
[12] R. Kogler *et al.*, Jet substructure at the Large Hadron Collider: Experimental review, Rev. Mod. Phys. **91**, 045003 (2019).
[13] 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).
[14] T. Plehn, M. Spannowsky, M. Takeuchi, and D. Zerwas, Stop reconstruction with tagged tops, J. High Energy Phys. 10 (2010) 078.
[15] J. Thaler and K. Van Tilburg, Identifying boosted objects with N-subjettiness, J. High Energy Phys. 03 (2011) 015.
[16] J. Thaler and K. Van Tilburg, Maximizing boosted top identification by minimizing N-subjettiness, J. High Energy Phys. 02 (2012) 093.
[17] D. Krohn, J. Thaler, and L.-T. Wang, Jet trimming, J. High Energy Phys. 02 (2010) 084.
[18] 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).
[19] D. Guest, K. Cranmer, and D. Whiteson, Deep learning and its application to LHC physics, Annu. Rev. Nucl. Part. Sci. **68**, 161 (2018).
[20] K. Albertsson *et al.*, Machine learning in high energy physics community white paper, J. Phys. Conf. Ser. **1085**, 022008 (2018).
[21] 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).
[22] M. Feickert and B. Nachman, A living review of machine learning for particle physics, arXiv:2102.02770.
[23] A. Andreassen, I. Feige, C. Frye, and M. D. Schwartz, JUNIPR: A framework for unsupervised machine learning in particle physics, Eur. Phys. J. C **79**, 102 (2019).
[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need, arXiv:1706.03762.
[25] E. A. Moreno, O. Cerri, J. M. Duarte, H. B. Newman, T. Q. Nguyen, A. Periwal, M. Pierini, A. Serikova, M. Spiropulu, and J.-R. Vlimant, JEDI-net: A jet identification algorithm based on interaction networks, Eur. Phys. J. C **80**, 58 (2020).
[26] P. T. Komiske, E. M. Metodiev, and J. Thaler, Energy flow networks: Deep sets for particle jets, J. High Energy Phys. 01 (2019) 121.
[27] H. Qu and L. Gouskos, ParticleNet: Jet tagging via particle clouds, Phys. Rev. D **101**, 056019 (2020).
[28] G. Louppe, K. Cho, C. Becot, and K. Cranmer, QCD-aware recursive neural networks for jet physics, J. High Energy Phys. 01 (2019) 057.
[29] T. Cheng, Recursive neural networks in quark/gluon tagging, Comput. Softw. Big Sci. **2**, 3 (2018).
[30] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, Jet-images—deep learning edition, J. High Energy Phys. 07 (2016) 069.
[31] 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.
[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, J. High Energy Phys. 10 (2018) 121.
[34] H. Qu, C. Li, and S. Qian, Particle transformer for jet tagging, arXiv:2202.03772.
[35] S. Gong, Q. Meng, J. Zhang, H. Qu, C. Li, S. Qian, W. Du, Z.-M. Ma, and T.-Y. Liu, An efficient Lorentz equivariant graph neural network for jet tagging, J. High Energy Phys. 07 (2022) 030.
[36] A. Bogatskiy, T. Hoffman, D. W. Miller, and J. T. Offermann, PELICAN: Permutation equivariant and Lorentz invariant or covariant aggregator network for particle physics, arXiv:2211.00454.
[37] D. Y. Grigoriev, E. Jankowski, and F. V. Tkachov, Optimal jet finder, Comput. Phys. Commun. **155**, 42 (2003).
[38] L. Mackey, B. Nachman, A. Schwartzman, and C. Stansbury, Fuzzy jets, J. High Energy Phys. 06 (2016) 010.
[39] G. Cerro, S. Dasmahapatra, H. A. Day-Hall, B. Ford, S. Moretti, and C. H. Shepherd-Themistocleous, Spectral clustering for jet physics, J. High Energy Phys. 02 (2022) 165.
[40] B. Mukhopadhyaya, T. Samui, and R. K. Singh, Dynamic radius jet clustering algorithm, J. High Energy Phys. 04 (2023) 019.
[41] X. Ju and B. Nachman, Supervised jet clustering with graph neural networks for Lorentz boosted bosons, Phys. Rev. D **102**, 075014 (2020).
[42] J. Guo, J. Li, T. Li, and R. Zhang, Boosted Higgs boson jet reconstruction via a graph neural network, Phys. Rev. D **103**, 116025 (2021).
[43] J. Li, T. Li, and F.-Z. Xu, Reconstructing boosted Higgs jets from event image segmentation, J. High Energy Phys. 04 (2021) 156.
[44] K. He, G. Gkioxari, P. Dollár, and R. Girshick, Mask R-CNN, arXiv:1703.06870.
[45] 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.
[46] T. Sjöstrand, S. Mrenna, and P. Skands, A brief introduction to PYTHIA 8.1, Comput. Phys. Commun. **178**, 852 (2008).
[47] ATLAS Collaboration, Summary of ATLAS PYTHIA 8 tunes, Report No. ATL-PHYS-PUB-2012-003, 2012.
[48] P. Skands, S. Carrazza, and J. Rojo, Tuning PYTHIA 8.1: The Monash 2013 tune, Eur. Phys. J. C **74**, 3024 (2014).
[49] ATLAS Collaboration, The PYTHIA 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model, Report No. ATL-PHYS-PUB-2016-017, 2016.
[50] K. He, G. Gkioxari, P. Dollár, and R. Girshick, Mask R-CNN, in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)* (IEEE, New York, 2017), pp. 2980–2988, 10.1109/ICCV.2017.322.
[51] R. Girshick, J. Donahue, T. Darrell, and J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, New York, 2014), pp. 580–587, 10.1109/CVPR.2014.81.
[52] K. He, X. Zhang, S. Ren, and J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, in *Proceedings of Computer Vision—ECCV 2014*, edited by D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars (Springer International Publishing, Cham, 2014), pp. 346–361.
[53] R. Girshick, Fast R-CNN, in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)* (IEEE, New York, 2015), pp. 1440–1448, 10.1109/ICCV.2015.169.
[54] S. Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, in *Advances in Neural Information Processing Systems*, edited by C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (Curran Associates, Red Hook, New York, 2015), Vol. 28 [https://proceedings.neurips.cc/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf].
[55] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, Feature pyramid networks for object detection, in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, New York, 2017), pp. 936–944, 10.1109/CVPR.2017.106.
[56] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR '16* (IEEE, New York, 2016), pp. 770–778, 10.1109/CVPR.2016.90.
[57] W. Abdulla, Mask R-CNN for object detection and instance segmentation on KERAS and TensorFlow, https://github.com/matterport/Mask_RCNN, 2017.
[58] T. Liang, X. Chu, Y. Liu, Y. Wang, Z. Tang, W. Chu, J. Chen, and H. Ling, CBNet: A composite backbone network architecture for object detection, IEEE Trans. Image Process. **31**, 6893 (2022).
[59] Z. Cai and N. Vasconcelos, Cascade R-CNN: Delving into high quality object detection, in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, New York, 2018), pp. 6154–6162, 10.1109/CVPR.2018.00644.
[60] J. T. Offermann, A. Bogatskiy, and T. Hoffman, Top Quark Momentum Reconstruction Dataset, Zenodo, 10.5281/zenodo.7126443 (2022).
[61] S. Gong, Converted Top Tagging Datasets (2022), https://osf.io/7u3fk/?view_only=8c42f1b112ab4a43bcf208012f9db2df.