Reconstructing boosted Higgs jets from event image segmentation

Jinmian Li, Tianjun Li, and Fang-Zhou Xu

College of Physics, Sichuan University, Chengdu 610065, China
CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China
School of Physical Sciences, University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, China
Center for High Energy Physics, Tsinghua University, Beijing 100084, China

E-mail: jmli@scu.edu.cn, tli@mail.itp.ac.cn, xfz14@mails.tsinghua.edu.cn

Abstract: The Mask R-CNN framework is adopted to reconstruct Higgs jets in Higgs boson production events, with the effects of pileup contamination taken into account. This automatic reconstruction method achieves higher efficiency of Higgs jet detection and higher accuracy of Higgs boson four-momentum reconstruction than traditional jet clustering and tagging methods using substructure information. Moreover, the Mask R-CNN trained on events containing a single Higgs jet is capable of detecting one or more Higgs jets in events of several different processes, without apparent degradation in reconstruction efficiency and accuracy. Taking the outputs of the network as new features to complement traditional jet substructure variables, the signal events can be discriminated from background events even further.
1 Introduction

Jet is one of the most prominent objects in the event reconstruction at the LHC. Due to the color confinement, quarks and gluons cannot be detected freely. Almost immediately after being produced, a quark/gluon goes through parton showering and hadronization, leading to a collimated spray of energetic detectable particles, which is referred to as jet [1–4]. At a high energy collider such as the LHC, a boosted hadronically decaying heavy particle can also give rise to a jet, e.g., W/Z boson, Higgs boson and top quark in the Standard Model (SM).

The aims of jet reconstruction are to obtain the original parton momentum and identity by using the information of final state hadrons. Since the first proposal for jet reconstruction [5], many jet clustering algorithms have been developed and adopted in experiments. At lepton colliders and hadron colliders, jets are usually reconstructed through a sequential recombination algorithm, such as the $k_T$ algorithm [6, 7], the Cambridge/Aachen algorithm [8], the anti-$k_t$ algorithm [9] and so on. Those algorithms involve a cone size parameter ($R$) that should be adjusted according to the detector architecture and the property of the target jet. The primary task after jet clustering is to identify the jet origin (jet tagging), for which the jet substructure is found to be very useful. Many dedicated variables/methods (see Refs. [10–15] for reviews) have been proposed to distinguish top quark jet from light flavor jet [16–19], W/Z/H jet from QCD jet [20], as well as quark jet from gluon jet [21, 22]. Despite the great success of the jet clustering algorithm, there are several possible issues remain: (1) The jet clustering can be easily infected by another close hard parton which could from multiple minimum-bias interactions (pileup) or from

---

1The parton here also refers to a boosted hadronically decaying W/Z boson or Higgs boson.
underlying events \(^2\); (2) The cone-size \(R\) is a priori parameter in jet clustering. Hadrons that have different origins may be clustered in the same jet if the \(R\) is too large, or those have the same origin may be miss-clustered into different jets if the \(R\) is too small. In both cases, the jet substructure is distorted and the jet tagging is no longer successful.

Even though the jet constituents are clustered sequentially in jet clustering algorithm, they have no intrinsic order from the theory perspective. The calorimeters measure the positions and energy depositions of jet constituents on fine-grained spatial cells. Treating each cell as a pixel, and the energy deposit in the cell as the intensity (or grayscale color) of that pixel, the jet can be naturally viewed as a digital image. The recent developments of computer vision, especially the application of deep learning \([25–27]\), can be used to reconstruct and tag the jet nature with low-level inputs (four-momentum vectors of final state hadrons). There are many works that use the image-based approach for various jet tagging tasks, e.g., W/Z jet tagging \([28–30]\), top quark jet tagging \([31–35]\), Higgs jet tagging \([36]\), quark-gluon jet discrimination \([37, 38]\) and new heavy particle jet tagging \([39]\). In these methods, the traditional jet clustering algorithm is used to reconstruct the jet in an event and a deep neural network (DNN) is applied subsequently for jet identification. Apparently those methods suffer from similar issues as in jet clustering, i.e., contaminations from soft-radiation and pileup events, jet image distortion due to inappropriate cone-size parameter. Moreover, in our previous study \([39]\), we find the jet tagging efficiency of the neural network which is trained on the event sample of a given process degrades when applying to another process. This implies that the network learns some process dependent features of the jet. The problem is even more severe as the final state multiplicity is higher.

A natural extension of the DNN application to jet tagging is implementing the jet detection within the DNN, so that the manual jet clustering is not needed. The techniques of object detection and image segmentation in computer vision just meet the need. There are mainly two types of object detection methods: (1) the region proposal based framework such as Mask R-CNN \([40]\); (2) the regression/classification based framework such as Yolo \([41, 42]\). In this work, we adopt the Mask R-CNN architecture to detect (reconstruct) the Higgs jet in Higgs events which is overlaid with abundant pileup events. The loss function of the network is designed to achieve the highest precision of the Higgs jet reconstruction\(^3\). We find this automatic Higgs detection method outperforms the traditional jet clustering and jet substructure tagging method, in the sense that the Higgs jet momentum is more precise and the background rejection rate is higher. Moreover, we will show that the Mask R-CNN trained on single Higgs events is capable of detecting one or more Higgs jets in different processes.

The paper is organized as follows. In Sec. 2 and Sec. 3, we introduce the data preparation and the architecture of the network. The performances of the CNN method are presented in Sec. 4. We provide conclusion and outlook in Sec. 5.

---

\(^2\)The jet grooming techniques \([20, 23, 24]\) can mitigate those effects.

\(^3\)The four-momentum of the reconstructed Higgs jet should be close to that of the original Higgs particle.
2 Event generation and data preparation

Two sets of data were generated by MG5\_aMC@NLO [43] for training and testing, either of which consists of 150k events of $H$+jets with the Higgs boson decaying to a pair of $b$-quark jets. Then 50 pileup events are superposed on each one of the hard events. In the first (second) set of events, the Higgs boson is forced to be boosted such that its transverse momentum satisfies $p_T > 200(300)$ GeV. $p_T$ is not restricted to a particularly narrow range, because the purpose here is to demonstrate the flexibility and universality of this algorithm. The Higgs momenta are altered a little bit by initial-state showers, whose simulations are handled by Pythia8 [44]. So the transverse momenta of some Higgs bosons fall below 200 (300) GeV despite the forced threshold in the MG5. We decide to leave it that way. The $p_T$ distributions of Higgs bosons are shown on the left panel of Fig. 1.

The data fed to the network are in the form of 320×320 grayscale images, representing the transpose momenta deposited in the $\eta - \phi$ plane across $[-\pi, \pi] \times [0, 2\pi]$. Each pixel corresponds to $\Delta\eta \times \Delta\phi \simeq 0.020 \times 0.020$. In order for the images to exploit as many $p_T$ scales as possible, the intensities of the pixels are normalized such that the grayscale spectrum of Higgs jet constituents is flat across the range 0 – 255, i.e., each one of the 256 color grades contains the same number of final state particles. Note that the statement holds true only for particles coming from Higgs bosons but not whole events. The procedure can be thought of as an image enhancement method, analogous to processing high dynamic range (HDR) images with the tone mapping technique to increase local contrast. The particle momenta of interest span over 4 orders of magnitudes, one simple way to represent them is a logarithmic transformation, but the consequence is lack of sensitivity to soft radiation and waste of intensity gradient in regions above 10 GeV. The middle and right panels of Fig. 1 show the $p_T$ distributions of the Higgs decay products and the grayscale mappings. Numerically, curves on the right panel equal the ones on the middle panel being integrated and multiplied by 256, which is just 8 bit color depth. An image represented in this way is expected to maximize the discerning ability of the network. Note that there is only a slight difference between two normalizations, so in principle, a unified normalization scheme is acceptable. Moreover, although it is optimized based on Higgs jets of certain energy, this procedure should be applicable to other objects to some extent, e.g., hadronically decaying

---

**Figure 1**: Left: Higgs boson $p_T$ distributions for training samples. Middle: $p_T$ distributions of Higgs decay products for training samples. Right: pixel intensity normalization schemes for 2 sets of data.
vector bosons and top quarks. Since $p_T$ distributions of jet constituents depend largely on the energy scales of parton shower and hadronization rather than identities of the particles which initiate the jets.

To tell the network the location of the Higgs jet on a detector image, each image is complemented by a mask in the form of $320 \times 320$ binary image (in the case of multiple Higgs jets, several masks) where signal pixels are labeled 1 and the rest 0. This is where things get complicated. Technically, only individual pixels representing particles coming from the Higgs boson are supposed to light up, no matter their $p_T$, which is obviously easy to implement. The issue is that our network, heavily relying on convolution, could not handle masks composed of sparse pixels very well. On the contrary, a connected mask is preferred, i.e., a jet area. Though this kind of labeling renders the algorithms more susceptible to pile-up, we will see that, due to its ability to keep the jet area noticeably smaller than conventional clustering algorithms, the influences are well under control. And a pileupmitigation procedure based on jet areas will make the four-momenta of reconstructed jets more accurate. We present two schemes of defining masks in this work. They share common pre-selection rules. Firstly, boost along the beam direction to the frame where $p_z$ of the Higgs boson becomes zero, then discard all constituent particles with energy lower that 1 GeV or outside the hemisphere whose axis coincide with the Higgs boson momentum. A simple alternative is requiring $p_T > 1$ GeV and $\Delta R < \pi/2$, but it dose not treat particles with different orientations on an equal footing. The inequalities are more manifest when decay products are scattered more widely on the $\eta - \phi$ plane due to higher rapidities or lower $p_T$ of Higgs bosons. The coordinates representing a Higgs boson on the $\eta - \phi$ plane are taken to be its rapidity $y$ and azimuth $\phi$. Only differences in rapidity $\Delta y$ are invariant under boosts along the beam direction, thus the proper measure of relative locations of particles on an image should be $y - \phi$ instead of $\eta - \phi$. Approximating $y$ by $\eta$ is fit only for ultrarelativistic particles which do not include Higgs bosons of our interest. The first mask is simply a convex hull covering all the constituents, while the second mask is an irregular polygon with serrated edge. We construct it as follows. Put the constituents in order according to their polar angle with respect to the initial Higgs boson, then connect them sequentially to form a closed loop, whose inside forms a radial mask. This mask is bound to be equal to or smaller than the convex hull in size. A mask is supposed to meet three requirements: 1) The $y - \phi$ coordinates of the Higgs boson lie within the mask, which serves as the criterion of a correct tag when tested; 2) The mask be simply connected, so as to suit a convolutional network; 3) The area of a mask be as small as possible to reduce the affect of pile-up. Other ways to construct a mask are possible and could potentially boost the performance. We show some instances in Fig. 2. Note that the pixelation of edges due to finite spacial resolution is not considered at this stage.

Considering that some jet constituents (in the case of radial masks, all jet constituents) are located right on edges of masks and that it is impossible for a CNN to be accurate down to pixel level, we will enlarge a convex hull/radial mask by one/two pixel(s) around its boundary during test.
Figure 2: Top: convex hull masks. Bottom: radial masks. Green dots represent $y - \phi$ coordinates of Higgs bosons, while red dots represent $\eta - \phi$ ($\approx y - \phi$) coordinates of their decay products.

3 Network Architecture

Neural networks have a common multilayer structure. Each layer consists of a certain number of nodes representing neurons. Each node is assigned a value computed from the previous layer with trainable weights and a bias, then operated on by an activation function, normally a rectified linear unit (ReLU), $f(x) = \max\{0, x\}$ to model the neuron’s nonlinear response to stimuli.

CNNs are well suited for image recognition tasks. Basically, convolution layers compute feature maps at different levels, pooling layers perform downsampling on them, and fully-connected layers are used for regression or classification. These are basic building blocks of a ConvNet. Different network architectures can be constructed by stacking multiple layers in various combinations to accomplish all kinds of recognition tasks, such as classification, detection and segmentation. Certain networks also incorporate deconvolution layers for upsampling, so that the spatial dimension of data can be increased.

Convolution operations work in a translation-invariant manner. The nature of images and CNNs correspond exactly to the finite space resolution of a detector and the Lorentz boost invariance of $y - \phi$ coordinates, making them perfectly capable of the jet detection task.

Mask R-CNN is a framework extensively adopted in the computer vision industry for object detection and semantic segmentation. It was developed progressively from the region-based CNN (R-CNN) framework and first put forth in 2017. Mask R-CNN is created by intricately combining three major functional modules, a region proposal network (RPN), a Region-based CNN (R-CNN) and a fully convolutional network (FCN). Both RPN and R-CNN are adapted by replacing the single-scale feature map with a feature pyramid
Figure 3: A sketch illustrating the components of Mask R-CNN and its processing pipeline, inspired by figures in Refs. [40, 45–47]. The orange square represents an input image and blue squares represent feature maps of multiple scales. Each map in FPN is upsampled and merged with the corresponding map in ResNet to generate a lower-level one. The sketch shows only three levels instead of five in the actual implementation. Anchors (dashed rectangles) are a set of reference boxes with fixed scales and aspect ratios that spread across the feature pyramid. RPN consists two sibling layers, a classification layer that assign positive/negative (red/green) labels to anchors and a regression layer that refines the boxes. Taken together, they define RoIs (solid red rectangles). The design of FPN and anchors enable Mask R-CNN to deal with multi-scale objects. It might be overkill for our purpose, but since the framework yields decent results only tweaked a little, we do not bother to change the main structure. The feature map pooled from each RoI is then fed to fully-connected layers to determine whether it is a Higgs jet and to fine-tune its bounding-box, and to a small FCN to predict the final mask.

network (FPN) to detect multi-scale objects. In the original paper, different convolutional backbone architectures used for feature extraction are compared. We choose the residual neural network of 50 layers (ResNet-50). We also test a deeper network, ResNet-101, and observe no significant improvements.

First, RPN outputs a set of rectangular proposals, referred to as regions of interest (RoIs), exploiting the feature hierarchy computed by the backbone network. Then, a pooling layer extracts features from each RoI. Finally, fully-connected layers perform classification and bounding-box regression. In our case, there are only two classes, Higgs jet and background. The features are shared between the proposal and detection networks. In parallel to predicting the class and bounding-box, a small FCN outputs a binary mask determining whether a pixel belongs to the jet. Though seemingly simple in concept, Mask R-CNN is quite a sophisticated framework to implement. Fortunately, source codes of at least three different implementations have been made publicly available. We use the one at https://github.com/matterport/Mask_RCNN. For elaboration of concepts and implementation details, see Ref. [40] and references therein.

A detector image has slightly different topology than an ordinary one, i.e., $\phi = 0$ and $\phi = 2\pi$ represent the same line. The decay products of a Higgs boson with $\phi \gtrsim 0$ or $\phi \lessapprox 2\pi$ will locate in two separate regions on an image, making the jet different to detect and cluster. In order to accommodate the cylindrical topology in a CNN, one need
to incorporate a periodic padding feature into the underlying framework (in our case, TensorFlow), which is beyond our ability and interest. We opt for another approach to bypass this challenge. If the $\phi$-coordinate of a Higgs boson lies outside of $[\pi/2, 3\pi/2]$, then a shift of $\pm \pi$ is applied to $\phi$ accordingly to move the Higgs boson into $\phi \in [\pi/2, 3\pi/2]$ . Combined with the pre-selection rule that the angles between a Higgs boson and its jet constituents be smaller than $\pi/2$, this procedure ensures the unity of a mask. This trick works for cases of at most two Higgs bosons in one event.

4 Network Performance

Fig. 4 and Fig. 5 show ground truth labels and test results of two typical events. We compare the performance of our algorithm, denoted as CNN, to a conventional one composed of mass-drop tagger and trimming, denoted as MDT. We optimize the parameters to achieve best reconstruction efficiency within 5 GeV of $m_H$. For dataset 1 (2), $R = 1.5$ (1.3) fat jets are trimmed by re-clustering the constituents into $R_{\text{sub}} = 0.20$ (0.19) $k_T$ subjets and discarding those with $p_T^{\text{subjet}} < 0.05 p_T^\text{jet}$. The goal is to reconstruct the four-momentum of a jet as accurately as possible, so we use reconstructed mass, transverse momentum, rapidity and azimuth as the criteria to measure the qualities of jet finding and clustering algorithms. Among them, the distribution of reconstructed mass is the most obvious, as it has a unique ground truth value, 125 GeV.

4.1 Detection efficiency

Fig. 6 shows the mass distributions of reconstructed Higgs bosons through CNN and MDT. Both of CNN and MDT may find multiple Higgs jets in one event, in that case, we keep the one with the highest score $p_T$ for CNN/MDT. We conclude from the plots that: 1) All four CNNs outperform MDT+trimming; 2) CNNs perform better if the test sample and train sample belong to the same dataset. When tested on the sample with $p_T > 300$ GeV, the network trained on the sample with $p_T > 200$ GeV is inferior to the network trained on the sample with $p_T > 300$ GeV, although in a sense, the sample with $p_T > 300$ GeV is the subset of the one with $p_T > 200$ GeV. This is probably due to insufficiency of samples or preference of the network caused by $p_T$ distribution; 3) CNNs trained using radial masks outperform those using convex hull masks. 4) The improvements of CNNs over MDT+trimming are more significant for Higgs jets with lower $p_T$. The right panel of Fig. 6 shows the efficiency boost, defined as the ratio of the number of jets with $m_{\text{rec}} \in [m_H - \Delta m, m_H + \Delta m]$. Here, only cases where the test sample and the train sample belong to the same dataset are shown. In dataset 1, radial masks are noticeably better than convex hull masks, while in dataset 2, the advantage is less evident. If we keep jets with $m_{\text{rec}} \in [105, 145]$ GeV, then CNN keeps 50%-80% more signal events than MDT+trimming.

Although the networks are trained on a single process, i.e., one Higgs boson plus three QCD jets, it may serve as a general Higgs tagger in all kinds of events. For demonstration purposes, we showcase its capability in three other processes: 1) two Higgs boson plus three QCD jets; 2) one Higgs boson plus two top quarks; 3) a hypothetical SUSY process,
Figure 4: A typical event containing one Higgs and three QCD jets (overlaid by pileup events). Each pixel corresponds to $\Delta \eta \times \Delta \phi = 0.020 \times 0.020$. The grayscale of the image are inverted for better perception. Particles of the Higgs jet are highlighted in blue. The yellow circle has $R = 0.8$, which is the minimum radius required for a C/A jet to enclose the Higgs decay products completely. This region is magnified three times for clearer visibility. The Higgs boson in this event has $p_T = 327$ GeV. This fat jet has $p_T = 417$ GeV and $m = 209$ GeV prior to any glooming procedure. The green contour indicates the input mask we constructed in radial shape, and the red one indicates the output. Note that we mandatorily enlarge the output by two pixels around its boundary as the final detection result. In this example, particles inside the enlarged contour produce a reconstruction of $m = 118$ GeV and $p_T = 329$ GeV, as compared to a trimmed jet of $m = 133$ GeV and $p_T = 336$ GeV.

$$pp \rightarrow \tilde{t}_1 \tilde{t}_1 \rightarrow \tilde{t}_1^0 \tilde{t}_1^0 \rightarrow \tilde{t}_1^0 \chi_1^0 H \chi_1^0,$$
where $m_{\tilde{t}_1^0} = 100$ GeV, $m_{\chi_1^0} = 800$ GeV and $m_{\tilde{t}_1} = 1$ TeV. The Higgs bosons in all processes are forced to be boosted, $p_T > 200$ (300) GeV and decay to two $b$-quarks. We apply the exact network trained using radial masks directly to these scenarios with no further alteration and see quite acceptable results. Higgs mass distributions are shown in Fig. 7. As a comparison, distributions of trimmed jets are also plotted. The parameters for trimming are shown in Table. 1. Again, we optimize the parameters to achieve best reconstruction efficiency within 5 GeV of $m_H$. Solid lines in the plots represent doubly $b$-tagged jets. Without $b$-tag, MDT is much more likely to mistake a $W$ boson or top quark for a Higgs boson. Here for simplicity, we assume a 100% efficiency of $b$-tag.
Figure 5: Another grayscale-inverted event image. The yellow circle has \( R = 0.4 \), which is magnified six times for clearer visibility. The Higgs boson has \( p_T = 327 \) GeV. The green contour indicates the input mask we constructed in convex hull shape, and the red one indicates the output enlarged by one pixel around its boundary. Reconstructed jets inside the red contour has \( m = 125 \) GeV and \( p_T = 927 \) GeV, as compared to the trimmed one of \( m = 127 \) GeV and \( p_T = 934 \) GeV.

Figure 6: Left and middle: reconstructed mass distribution of different methods on 2 test samples. Right: reconstruction efficiency improvement of CNN over MDT, in terms of the number of reconstructed Higgs within \( \Delta m \) of \( m_H \). At least 50% more Higgs are identified in a narrow mass range, say, \( m \subset [120, 130] \) GeV.

| process  | \( R_0 \) | \( R_{sub} \) | \( f_{cut} \) |
|----------|----------|------------|-------------|
| HHjjj    | 1.4      | 0.19       | 0.5         |
| Htt      | 1.4      | 0.19       | 0.5         |
| \( \tilde{t}_1^* \tilde{t}_1 \) | 0.9      | 0.18       | 0.5         |

Table 1: Trimming parameters in the MDT method.
4.2 Reconstruction accuracy

To show that detection and segmentation produce better accuracy than clustering, tagging and trimming, we present the deviations of reconstructed variables in Fig. 8 and Fig. 9. Only correctly tagged Higgs bosons are taken into account. We consider it a correct tag if the true Higgs boson falls within $R_0$ of the reconstructed one or falls within the mask. Note that these criteria can not be applied to actual event selections, hence they are only suitable for demonstration of reconstruction accuracy. Different shades of gray regions and colored contours indicate 20%, 40%, 60% and 80% of events, respectively. The closer they are to the center, the higher accuracy they stand for.

The performance of CNN is worse than MDT in terms of $p_T$ reconstruction, where a systematic excess is prominent, because we do not carry out any pileup reduction procedure, other than try to contain its contamination via small jet areas. One average, CNN reconstructs mass more accurately than trimming. Convex hull mask and lower $p_T$ suffers more from pile-up, because they produce a larger jet area. The diameters (defined as the largest angular distance $\Delta R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$ between any two marked particles) and areas of Higgs jets predicted by CNN are shown in Fig. 10.

We adopt a method similar to [48] for pileupmitigation. For as many as 50 pileup events, their $p_T$ distribute almost uniformly on the $\eta - \phi$ plane. Let the density of $p_T$ be $\rho$. According to the mask shape of each jet, subtract pileup from the original jet, then we obtain the corrected four-momentum. The corrections of mass and $p_T$ are roughly,

$$
\delta p_T \simeq -\rho A \tag{4.1}
$$

$$
\delta m^2 \simeq -\rho p_T A (\Delta R^2) \gtrsim -\rho p_T A \frac{\Delta R_{\text{max}}^2}{2} \tag{4.2}
$$
where \( A \) is the area of the mask, \( \Delta R \) represent the distance between a point on the mask and the Higgs boson, and \( \langle \Delta R^2 \rangle \) is the area-averaged \( \Delta R^2 \). Note that \( \Delta R_{\text{max}} \) does not equal half of the diameter. We do not use above approximations for pileup subtraction.

In simulated events, \( \rho \simeq 55 - 60 \text{ GeV} \). But in practice, we find setting \( \rho = 35 \text{ GeV} \) and imposing \( m = 0 \text{ GeV} \) on subtracted four-momentum yields best results. One of the reasons is that masks predicted by CNN do not cover all jet constituents. Another reason is that the mass and \( p_T \) of pileup are overestimated when substituting a continuous distribution for a discrete one. This method is by no means flawless, but it proves effective. The results are shown in Fig. 11 and Fig. 12.
Figure 9: The same as Fig. 8, except using convex hull masks.

Figure 10: Left: jet diameter distribution. Right: jet area distribution, vertical lines indicate jet areas with $R = 0.1, 0.2, \ldots, 0.7$. 
4.3 Signal and background discrimination

Given its high efficiency, one may be concerned that this network is more likely to be fooled by false Higgs. So we also examine the discrimination power of our trained networks on a dataset composed of two top quarks plus two QCD jets events. Ideally, the score returned by the network alone should suffice as a discriminant between real and fake Higgs. With a varying number between 0 and 1 as the threshold, one would get a receiver operating characteristic curve (ROC) displaying how well the signals and backgrounds can be differentiated. Unfortunately, our network is not powerful enough yet to accomplish this mission in one step, we will probably get there with a more sophisticated representation of the event and a network with more input branches. For now, we complement the outputs of our network with a few substructure variables and implement a BDT to fulfill this task. Variables in consideration are listed in Table. 2, inspired by [49]. The ROC curves and significance improvement curves (SIC) are shown in Fig. 13, where solid lines indicate

Figure 11: The same as Fig. 8 except pileupsubtraction.
doubly $b$-tagged events. Here, we consider it a $b$-tag if the jet or subjet with a certain radius cover the $\eta - \phi$ coordinates of a $b$ quark with $p_T > 20$ GeV. On top of that, a $b$-tag efficiency of 70%, $c$-quark mis-tagging rate of 15% and the other light quark and gluon mis-tagging rates of 0.8% are set. The overall efficiency of double $b$-tag is about 35%, so the ROC curve does not exceed $\varepsilon \simeq 0.35$. From the ROC, we can find that the background rejection for a given signal selection efficiency can be improved by an order of magnitude once the Mask R-CNN score is included in the BDT analysis. Equipped with the Mask R-CNN score and $b$-tagging, the signal significance factor can achieve $\sim \mathcal{O}(10)$ with signal efficiency $\varepsilon \lesssim 0.25$.

5 Conclusion and outlook

The Mask R-CNN framework is adopted to reconstruct the Higgs jets in Higgs production events which is overlaid with abundant pileup events. Because the network is capable to

Figure 12: The same as Fig. 9 except pileup subtraction.
detect object in a translation-invariant manner, it is suitable for reconstructing Higgs jet in wide ranges of \( p_T \) and \( \eta \).

We train the Mask R-CNN on two datasets of \( H+\)jets production overlaid by pileup events, with the requirements of \( p_T(H) > 200 \) GeV and \( p_T(H) > 300 \) GeV, respectively. The Higgs jet in an event is defined by a mask which is built according to the truth level information from Monte Carlo simulation. Two schemes of mask definition are proposed in this work: convex hull mask and radial mask. We find that the CNN method outperform the traditional jet substructure method in the Higgs jet reconstruction, i.e., provide more accurate Higgs jet momentum.

Even though the Mask R-CNN is trained on the events of the \( H+\)jets process, it is totally applicable to detect Higgs jets in events of different processes. For illustration, three different processes are considered in the work: 1) two Higgs bosons plus three QCD jets; 2) one Higgs boson plus two top quarks; 3) a hypothetical SUSY process, \( pp \rightarrow \tilde{t}_1 \tilde{t}_1 \rightarrow \tilde{\chi}_1^0 \tilde{\chi}_2^0 \rightarrow \tilde{\chi}_1^0 t H \chi_1^0 \). Promising Higgs detection efficiencies are obtained for all processes. In particular, it is surprised to find that the Mask R-CNN which is trained on single Higgs
events is capable to reconstruct two Higgs jets in the first process. Finally, the background resistance of the method is demonstrated by applying it to the $t\bar{t}+\text{jets}$ process, which is typically the dominant background in Higgs-related searches. In addition to the traditional jet substructure variables, the information provided by the output of the Mask R-CNN can help to further improve the background rejection rate by an order of magnitude.

The method proposed in this work can be generalized to detect multiple different objects (such as W/Z boson jets) simultaneously. Moreover, the $\eta$, $\phi$ and $p_T$ of the Higgs jet are obtained by the vector sum of momenta of all marked particles (pixels) at current stage. One could integrate the momentum estimation into the network, which could be even more accurate (more close to the true Higgs boson momentum). It should be noted that two schemes of mask definition given in this work are by no means unique and the best. They may be improved in realistic experiments. As for jet induced by a colored particle, the assignment of the final state particles is ambiguous. It will be more difficult to define an appropriate mask for them. We leave those points to future works.

Note Added

While we were preparing this manuscript, a paper with similar purpose [50] appears on the arXiv. Different neural network was adopted in their study. The effects from pileup events as well as the application to processes different from the one used in training are not studied though.

Acknowledgments

We thank Jun Guo for valuable comments and suggestions. 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.

References

[1] G. Hanson et al., Evidence for jet structure in hadron production by $e^+e^-$ annihilation, Phys. Rev. Lett. 35 (1975) 1609.

[2] G.C. Blazey et al., Run II jet physics, in Physics at Run II: QCD and Weak Boson Physics Workshop: Final General Meeting, pp. 47–77, 2000 [hep-ex/0005012].

[3] S. Ellis, J. Huston, K. Hatakeyama, P. Loch and M. Tonnesmann, Jets in hadron-hadron collisions, Prog. Part. Nucl. Phys. 60 (2008) 484 [0712.2447].

[4] G.P. Salam, Towards Jetography, Eur. Phys. J. C 67 (2010) 637 [0906.1833].

[5] G.F. Sterman and S. Weinberg, Jets from Quantum Chromodynamics, Phys. Rev. Lett. 39 (1977) 1436.

[6] S.D. Ellis and D.E. Soper, Successive combination jet algorithm for hadron collisions, Phys. Rev. D 48 (1993) 3160 [hep-ph/9305266].
[7] S. Catani, Y.L. Dokshitzer, M. Seymour and B. Webber, Longitudinally invariant $k_t$ clustering algorithms for hadron hadron collisions, *Nucl. Phys. B* **406** (1993) 187.

[8] Y.L. Dokshitzer, G. Leder, S. Moretti and B. Webber, Better jet clustering algorithms, *JHEP* **08** (1997) 001 [hep-ph/9707323].

[9] M. Cacciari, G.P. Salam and G. Soyez, The anti-$k_t$ jet clustering algorithm, *JHEP* **04** (2008) 063 [0802.1189].

[10] A. Abdesselam et al., Boosted Objects: A Probe of Beyond the Standard Model Physics, *Eur. Phys. J. C* **71** (2011) 1661 [1012.5412].

[11] A. Altheimer et al., Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks, *J. Phys. G* **39** (2012) 063001 [1201.0008].

[12] 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** (2014) 2792 [1311.2708].

[13] D. Adams et al., Towards an Understanding of the Correlations in Jet Substructure, *Eur. Phys. J. C* **75** (2015) 409 [1504.00679].

[14] 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. Rept.* **841** (2020) 1 [1709.04464].

[15] R. Kogler et al., Jet Substructure at the Large Hadron Collider: Experimental Review, *Rev. Mod. Phys.* **91** (2019) 045003 [1803.06991].

[16] D.E. Kaplan, K. Rehermann, M.D. Schwartz and B. Tweedie, Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks, *Phys. Rev. Lett.* **101** (2008) 142001 [0806.0848].

[17] T. Plehn, G.P. Salam and M. Spannowsky, Fat Jets for a Light Higgs, *Phys. Rev. Lett.* **104** (2010) 111801 [0910.5472].

[18] T. Plehn, M. Spannowsky, M. Takeuchi and D. Zerwas, Stop Reconstruction with Tagged Tops, *JHEP* **10** (2010) 078 [1006.2833].

[19] D.E. Soper and M. Spannowsky, Finding top quarks with shower deconstruction, *Phys. Rev. D* **87** (2013) 054012 [1211.3140].

[20] 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** (2008) 242001 [0802.2470].

[21] J. Gallicchio and M.D. Schwartz, Quark and Gluon Tagging at the LHC, *Phys. Rev. Lett.* **107** (2011) 172001 [1106.3076].

[22] J. Gallicchio and M.D. Schwartz, Quark and Gluon Jet Substructure, *JHEP* **04** (2013) 090 [1211.7038].

[23] D. Krohn, J. Thaler and L.-T. Wang, Jet Trimming, *JHEP* **02** (2010) 084 [0912.1342].

[24] S.D. Ellis, C.K. Vermilion and J.R. Walsh, Techniques for improved heavy particle searches with jet substructure, *Phys. Rev. D* **80** (2009) 051501 [0903.5081].

[25] D. Guest, K. Cranmer and D. Whiteson, Deep Learning and its Application to LHC Physics, *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 161 [1806.11484].
[26] K. Albertsson et al., *Machine Learning in High Energy Physics Community White Paper*, *J. Phys. Conf. Ser.* **1085** (2018) 022008 [1807.02876].

[27] A. Radovic et al., *Machine learning at the energy and intensity frontiers of particle physics*, *Nature* **560** (2018) 41.

[28] J. Cogan, M. Kagan, E. Strauss and A. Schwartzman, *Jet-Images: Computer Vision Inspired Techniques for Jet Tagging*, *JHEP* **02** (2015) 118 [1407.5675].

[29] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *Jet-images — deep learning edition*, *JHEP* **07** (2016) 069 [1511.05190].

[30] P. Baldi, K. Bauer, C. Eng, P. Sadowski and D. Whiteson, *Jet Substructure Classification in High-Energy Physics with Deep Neural Networks*, *Phys. Rev. D* **93** (2016) 094034 [1603.09349].

[31] L.G. Almeida, M. Backović, M. Cliche, S.J. Lee and M. Perelstein, *Playing Tag with ANN: Boosted Top Identification with Pattern Recognition*, *JHEP* **07** (2015) 086 [1501.05968].

[32] G. Kasieczka, T. Plehn, M. Russell and T. Schell, *Deep-learning Top Taggers or The End of QCD?*, *JHEP* **05** (2017) 006 [1701.08784].

[33] S. Macaluso and D. Shih, *Pulling Out All the Tops with Computer Vision and Deep Learning*, *JHEP* **10** (2018) 121 [1803.00107].

[34] A. Butter et al., *The Machine Learning Landscape of Top Taggers*, *SciPost Phys.* **7** (2019) 014 [1902.09914].

[35] CMS Collaboration collaboration, *Machine learning-based identification of highly Lorentz-boosted hadronically decaying particles at the CMS experiment*, Tech. Rep. CMS-PAS-JME-18-002, CERN, Geneva (2019).

[36] J. Lin, M. Freytsis, I. Moult and B. Nachman, *Boosting $H \rightarrow b \bar{b}$ with Machine Learning*, *JHEP* **10** (2018) 101 [1807.10768].

[37] P.T. Komiske, E.M. Metodiev and M.D. Schwartz, *Deep learning in color: towards automated quark/gluon jet discrimination*, *JHEP* **01** (2017) 110 [1612.01551].

[38] ATLAS Collaboration collaboration, *Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector*, Tech. Rep. ATL-PHYS-PUB-2017-017, CERN, Geneva (Jul, 2017).

[39] J. Guo, J. Li, T. Li, F. Xu and W. Zhang, *Deep learning for R-parity violating supersymmetry searches at the LHC*, *Phys. Rev. D* **98** (2018) 076017 [1805.10730].

[40] K. He, G. Gkioxari, P. Dollár and R. Girshick, *Mask r-cnn*, in *ICCV*, pp. 2980–2988, 2017 [1703.06870].

[41] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, *You Only Look Once: Unified, Real-Time Object Detection*, pp. 779–788, 2016.

[42] J. Redmon and A. Farhadi, *YOLOv3: An Incremental Improvement*, arXiv (2018) [1804.02767].

[43] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [1405.0301].

[44] T. Sjostrand, S. Mrenna and P.Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852 [0710.3820].
[45] R. Girshick, J. Donahue, T. Darrell and J. Malik, *Rich feature hierarchies for accurate object detection and semantic segmentation*, pp. 580–587, 2014.

[46] R. Girshick, *Fast r-cnn*, in *ICCV*, pp. 1440–1448, 2015.

[47] S. Ren, K. He, R. Girshick and J. Sun, *Faster r-cnn: Towards real-time object detection with region proposal networks*, in *NIPS*, pp. 91–99, 2015 [1506.01497].

[48] M. Cacciari and G.P. Salam, *Pileup subtraction using jet areas*, *Phys. Lett. B* 659 (2008) 119 [0707.1378].

[49] ATLAS collaboration, *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at √s=8 TeV*, *Eur. Phys. J. C* 75 (2015) 303 [1412.7086].

[50] X. Ju and B. Nachman, *Supervised Jet Clustering with Graph Neural Networks for Lorentz Boosted Bosons*, 2008.06064.