Improving the measurement of the Higgs boson-gluon coupling using convolutional neural networks at e^+e^- colliders

Gexing Li,[1,2,*] Zhao Li,[1,2,†] Yan Wang,[3,1,‡] and Yefan Wang[1,2,§]

[1]*Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China*
[2]*School of Physics Sciences, University of Chinese Academy of Sciences, Beijing 100039, China*
[3]*College of Physics and Electronic Information, Inner Mongolia Normal University, Hohhot 010022, China*

(Received 27 August 2019; published 17 December 2019)

In this paper we propose to use convolutional neural networks (CNNs) to improve the precision measurement of the Higgs boson-gluon effective coupling at lepton colliders. The CNN is employed to recognize the Higgs boson and a Z boson associated production process, with the Higgs boson decaying to a gluon pair and the Z boson decaying to a lepton pair at the center-of-mass energy 250 GeV and integrated luminosity 5 ab^{-1}. By using CNNs, the uncertainty of the effective coupling measurement can be decreased from 1.94% to about 1.28% using the PYTHIA data and from 1.82% to about 1.22% using the HERWIG data in the Monte Carlo simulation. Moreover, the performance of CNNs using different final state constituents shows that the energy distributions of the leading and subleading jets constituents play a major role in the identification and the optimal uncertainty of effective coupling using CNNs is reduced by about 35% compared to that using conventional method.

DOI: 10.1103/PhysRevD.100.116013

I. INTRODUCTION

The Higgs boson occupies a distinct place in the Standard Model (SM) of particle physics. Many lingering physics problems are linked to the Higgs boson, for instance, the stability of the vacuum, electroweak hierarchy problem, and dark matter. These problems imply the existence of new physics beyond the SM and require a good understanding of the Higgs properties. The effective coupling of the Higgs boson to a gluon pair is one of the most important parameters. Many theories beyond the SM predict that the Higgs boson-gluon coupling may have deviation from the SM prediction by direct or indirect effects, for example, the stop in supersymmetry or the T quark in little Higgs models can contribute to the coupling through the loop effects [1–10]. Therefore, the precision measurement of the Higgs boson-gluon coupling will be a touchstone of the SM and may lead to a breakthrough for new physics.

Although the gluon fusion is the most important process of the Higgs boson production at the CERN Large Hadron Collider, the Higgs boson-gluon coupling is still difficult to be determined accurately due to the overwhelming large QCD radiation [11,12]. The better candidates for the precision measurement of Higgs boson-gluon coupling can be electron positron colliders, which have the clean environment and the high luminosity. The possible future electron positron colliders, which are usually called the Higgs factory at 250 GeV center-of-mass energy, include the Circular Electron-Positron Collider [13–15], Future Circular Collider-electron-positron [16–18], and International Linear Collider [19–23]. At the Higgs factory, the measurement on most of the Higgs properties can reach percent level accuracy [11,12,24]. For the Higgs boson-gluon effective coupling the κ_g [5,14] is always used to parametrize its deviation from the SM prediction, where $\kappa_g^{\mathrm{SM}} = 1$. With the conventional method (only using the kinematic cuts and b tagging) [25] the uncertainty of the κ_g will reach about 2.2% for the channel of a Z boson decaying to a lepton pair including the detector effect at the Circular Electron-Positron Collider.

The measurement accuracy of the Higgs boson-gluon coupling can be further improved through an effective identification of jet types. In the last few decades, many different observables motivated by color charge, color connections, electrical charge, or spin have been proposed and achieved good performance [26–28]. For example,

*ligx@ihep.ac.cn
†zhaoli@ihep.ac.cn
‡wangyan728@ihep.ac.cn
§wangyefan@ihep.ac.cn

the jet energy profile is one of the useful jet substructure observables to distinguish quark and gluon jets by the energy distribution of jet constituents. By using the jet energy profile, the uncertainty of the Higgs boson-gluon coupling can be further reduced to about 1.6% for the channel of a Z boson decaying to a lepton pair [29].

However, an observable usually only describes a certain aspect of the jets or some special processes. Although it is better to choose a set of complementary observables to extract more comprehensive characteristics to identify different types of jets or events, the applicable scope of different observables and the degree of association between them will also be difficult problems. Moreover, the deeper correlations between the jet or event constituents may be difficult to be extracted by the artificial observables.

Deep learning has been applied to solve many complicated problems in particle physics. In particular, deep neural networks have been employed to distinguish different types of jets, including Higgs boson tagging [30], boosted W boson tagging [31,32], boosted top tagging [33,34], single merged jet tagging [35], heavy-light quark discrimination [36], and quark-gluon discrimination [37–40]. They all get an exciting recognition capability and superior to the conventional method. A convolutional neural network (CNN) is one of the most popular and powerful algorithms. Its powerful ability of image recognition makes it easy to extract more comprehensive and deeper features to analyze the jet substructure. It is very suitable for jet tagging and also for testing different shower and hadronization schemes by comparing different Monte Carlo (MC) generators.

In this paper, we propose to use the CNN for the precision measurement of Higgs boson-gluon effective coupling by distinguishing the background processes from the process of a Z boson decaying to a lepton pair and a Higgs boson decaying to a gluon pair ($2\ell 2g$) at lepton colliders. The global information in an event is used for the training of the CNN instead of the jet information. We will use events from different event generators for neural network training and testing to illuminate the difference between the different shower and hadronization schemes.

The content is organized as follows. In the next section, the CNN is briefly reviewed. In the third section, the MC events are generated by PYTHIA and HERWIG. The production of images and CNN architecture are introduced in the fourth section. In the fifth section, we show the results using the CNN. The conclusion is made in the last section.

II. CONVOLUTIONAL NEURAL NETWORKS

A neural network is one of the most popular algorithms in machine learning. Generally, a neural network consists of an input layer, hidden layer, and output layer. A layer is dense if each of its units connects to all of the units in the previous layer. If a neural network consists of a dense layer completely, it will tune a large number of parameters and waste a lot of computing resources. Actually, each neuron only needs to perceive the local image instead of the global image for image recognition, and then the global information can be obtained by integrating the local information at a higher level. This motivates the design of the CNN [41]. In the last few years, based on the development of computer technology, the CNN has been a mainstay of many major breakthroughs in various fields.

In the image identification, the images in the CNN will pass a convolutional layer, pooling layer, and dense layer. The function of the convolutional layer is extracting features of the image. This can be implemented by the convolution of the filter and the image. A filter is a $n \times n$ grid of weights, where n is the filter size. The convolution is that each weight in a filter multiplies the corresponding pixel intensity in a patch the same size as an image. Then, we sum the convolutional values, add a bias, and feed it to an activation function. Activation functions introduce the nonlinear properties into neural networks, which enable the neural networks to learn the deeper information. The most used activation function in CNNs are rectified linear units (ReLU), which are defined as $f(x) = \max\{0, x\}$. Each convolutional layer usually has many different filters to extract different features of a image. For the multichannel images, there are different colors and convolutional filters in each channel. Each color or channel will be solved by a corresponding filter, like the single color image, and will be accumulated in the final step.

Then, a pooling layer, following the convolutional layer, is used to reduce the number of parameters. The filter of the pooling layer is the $m \times m$ grid, where m is the pooling size. The max pooling and average pooling are the most common pooling functions. Max pooling takes the largest value while average pooling takes the average of all values in a filter region. A dropout usually is added to avoid the overfitting. It refers to the random discarding of some neural network units at certain probability in each training [42]. Finally, the dense layers are added to integrate the features in the feature maps extracted by the convolution layers and pooling layers to obtain the high-level meanings of the features and then use them for image recognition.

The error of the model can be quantified by the binary cross entropy loss function [43]

$$f_{\text{loss}} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \ln Y_i + (1 - y_i)\ln(1 - Y_i)], \quad (1)$$

where N is the number of training events. The y_i and Y_i are the real value and the predicted value by the CNN of the ith event. The training process is tuning the parameters in the model to minimize the loss function.

III. PREPROCESSING

The main process of the Higgs boson production is $e^+e^- \to Z^*/\gamma^* \to Zh$ at the future e^+e^- colliders. We choose the process of the Z boson decaying to a lepton

pair and the Higgs boson decaying to a gluon pair ($2\ell 2g$) as the signal process since the Z boson can be reconstructed very well by the lepton pair. The process of different Z boson decay modes $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ are discussed first. Then the two lepton channels are combined as $Z \to \ell^+\ell^-$. The backgrounds are divided into two-fermion leptonic (final states are a lepton pair from the Z or γ^* intermediate states), two-fermion hadronic (final states are two quarks), four-fermion leptonic (final states are four leptons from the vector boson pair intermediate states), four-fermion semileptonic (final states are a pair of charged leptons and a pair of quarks from the vector boson pair intermediate states), four-fermion hadronic (final states are four quarks), and the Higgs boson production with the final states, which are different from the signal [mainly the Higgs boson and a Z boson associated production process with the Z boson decaying to a lepton pair and the Higgs boson decaying to a b/c quark pair (hbb/hcc) or W/Z boson pair (hWW/hZZ)] [15,44]. Both the signal and background events are simulated at future e^+e^- colliders [13–23] for the center-of-mass energy 250 GeV and integrated luminosity 5 ab^{-1}. The parton level MC events are generated by WHIZARD 1.95 [45,46] and transferred to hadron level by PYTHIA 6 [47] and HERWIG 7 [48], respectively. For clarity, we call them PYTHIA data and HERWIG data, respectively.

We select a pair of isolated leptons to reconstruct the Z boson. The rest of the final state constituents are clustered into jets via FASTJET 3.3.0 [49] using the anti-k_t algorithm with a large jet cone of $R = 1.5$, and the energy of each jet is required to be more than 5 GeV. To suppress the two-fermion leptonic and four-fermion leptonic backgrounds [15], we add two cuts at first. One is the number of the stable charge particles in the final state $N_{\text{charge}} \geq 10$, and another is the electromagnetic energy ratio in the final state $R_{\text{EM}} < 0.99$. Then, the kinematic cuts, i.e., invariant mass, recoil mass, and other constraints of the lepton pair and jet pair, are used to ensure that the lepton pair and jet pair, respectively, come from the Z boson and the Higgs boson to reject the two-fermion hadronic and four-fermion hadronic backgrounds. More details of the analysis can be found in Ref. [29]. The reference also shows that the c tagging cannot decrease the κ_g uncertainty effectively since its mistag rate for the gluon jet will exclude some gluon jets. Therefore, we only use the b tagging in this paper.

The kinematic cuts and b tagging can remove a large number of the distinct backgrounds, which will greatly improve the efficiency of the neural network. The remaining backgrounds contain the hbb, hcc, hWW, hZZ, and four-fermion semileptonic. The jets in the backgrounds hbb/hcc and four-fermion semileptonic are mainly heavy quark jets and light quark jets, respectively. But the jets in the backgrounds hWW/hZZ are W/Z jets and light quark jets since quite a few of the light quark jets are merged into the W/Z jets with a large jet cone of $R = 1.5$. It is the complex jet types in the backgrounds that make the signal identification be a challenge.

TABLE I. The uncertainties of κ_g in the different Z boson decay modes with the conventional method using PYTHIA data and HERWIG data.

	$Z \to e^+e^-$	$Z \to \mu^+\mu^-$	$Z \to \ell^+\ell^-$
PYTHIA	2.93%	2.53%	1.94%
HERWIG	2.67%	2.47%	1.82%

After all the cuts, the uncertainties of κ_g should be evaluated. The evaluation of systematic uncertainties requires a detailed detector study and is unknown yet for the Higgs factory. But the statistical uncertainty of κ_g around the SM prediction can be explicitly expressed as

$$\delta\kappa_g = \frac{\sqrt{N}}{2N_g}, \tag{2}$$

where N_g and N are the numbers of the Higgs boson decaying to gluon pair events and total events, respectively.

In Table I, the second and the third lines are the uncertainties of κ_g with the conventional method using PYTHIA data and HERWIG data, respectively. The difference between the results using PYTHIA data and HERWIG data may come from the different shower and hadronization schemes. The k_T-ordered and the angular-ordered schemes are used for shower effect, and the Lund string and the cluster models are used for the hadronization effect in PYTHIA6 and HERWIG7, respectively.

IV. ARCHITECTURE OF THE CNN

For the training of the CNN, we use the combined lepton channel $Z \to \ell^+\ell^-$. The entire spherical surface, where the azimuthal angle $\phi \in [-\pi, \pi]$ and the polar angle $\theta \in [0, \pi]$, is treated as a two dimensional plane image. Each image is designed to have a 66-pixel length in the ϕ direction and a 34-pixel length in the θ direction. The energy of all the final state stable particles is discretized into pixels as our pixel intensity at lepton colliders. The images of the signal process $2\ell 2g$ are given the sign one and the other images as the background process are given the sign zero. All the images are divided into the training, validation, and test sets in proportion to 8:1:1.

The neural network is implemented by using Keras [43] with TensorFlow backend. Our CNN architecture is inspired by the VGGNet [50] architectures and consisted of four iterations of convolutional layers and maxpooling layers shown in Fig. 1. Then the feature map is flattened and fed to a dense layer with 128 units. Finally, a dense layer with one unit and a sigmoid activation is added to classify the signal and background processes. Each convolutional layer consists of 64 or 128 filters with filter size 3×3 and a ReLU activation. The uniform distribution is used to initialize the filters. The stride length of the

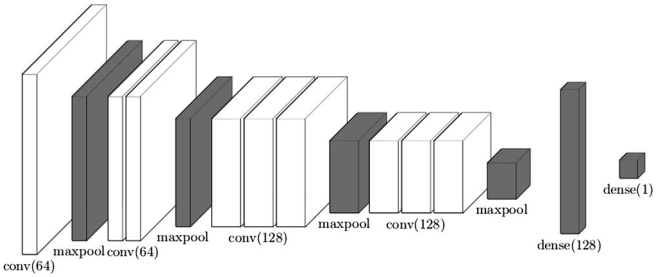

FIG. 1. The architecture of our CNN.

convolution is 1. The first convolutional layer is set without padding to weaken the influence of the edge information of the image at the beginning while the others are set with padding to keep all the information of the feature map. Each maxpooling layer performs a 2×2 down-sampling with a stride length of 2. A dropout layer follows each maxpooling layer and the dense layer to avoid overfitting. All the dropout rates of dropout layers are 0.5 except that the first one is 0.25.

The binary cross entropy is used as the loss function. The optimization of training uses the Adam algorithm [51] and the learning rate is 0.0005. The training is set with batch sizes 128 and 100 epochs and an early stopping patience of 5. Thus, the training will stop early if the value of the validation loss does not go down 5 times.[1]

The receiver operator characteristic (ROC) curve is usually used to quantify the performance of neural networks. A ROC curve is generated by plotting the true positive rate against the false positive rate. The area under the curve (AUC) is defined to compare the overall performance of the neural networks. In this paper, the true positive rate is the signal process ($2\ell 2g$) acceptance efficiency R_g and the false positive rate is the mistag efficiency R_B of the background processes.

Then we test the performance of our neural network and compare it to several different neural networks. Figure 2 shows the background rejection rate $1 - R_B$ as a function of the signal acceptance efficiency R_g for the CNN with different architectures. The lines marked as "3-conv," "Alex," and "MiniVGG" represent the performance of the CNN architectures in Refs. [40,52,53], respectively. The green dotted line is the result using the neural network, which contains three iterations of a convolutional layer and a maxpooling layer. The blue dashed line is the result using the famous AlexNet, which uses a stack of convolutional layers to increase the nonlinearity of the neural network and bigger filter size to increase the receptive field. So, the performance of the AlexNet has a significant improvement compared to that of the 3-conv. The red dash-dotted line is the result using the neural network, which is inspired by the

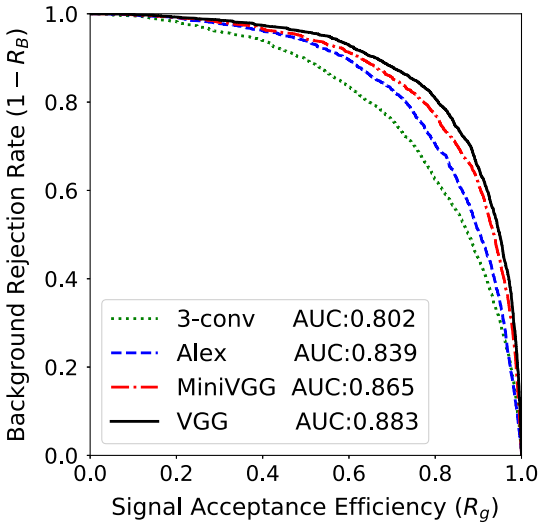

FIG. 2. The background rejection rate $1 - R_B$ as a function of the signal acceptance efficiency R_g for the CNN with different architectures.

MiniVGGNet architecture but with a bigger filter size in the first two convolutional layers. More iterations of the convolution layer stack further enhance the nonlinearity of the neural network and lead to improved performance. According to the advantages of the VGGNet, our neural network uses a stack of convolutional layers with 3×3 filter size instead of a single convolutional layer with a big filter size, which can increase the nonlinearity of the neural network and reduce the number of parameters. The black solid line is the result using our neural network architecture, which is better than other three neural network structures for the identification of our signal and background processes.

V. RESULTS

In this section, we will present the improvement on the κ_g uncertainty archived by using the CNN.

Figure 3 shows the background rejection rate $1 - R_B$ as a function of the signal acceptance efficiency R_g for our CNN. The area under these curves are the AUC values of the different cases. Both training and testing have been applied to the PYTHIA and HERWIG data. For convenience, The symbol P(H) + P(H) is used to represent training with the PYTHIA (HERWIG) data and testing with the PYTHIA (HERWIG) data. It can be found that at around $R_g = 80\%$ the background rejection rate can reach about 80%; meanwhile, the signal acceptance efficiency could still be acceptable. Furthermore, it can be seen that the AUC value of the "H + H" is slightly better than that of the "P + P." More specifically, the curves of the P + P and H + H are very similar at the low signal acceptance efficiency region $R_g < 70\%$, but the curve of the H + H is higher than that of the P + P at the high signal acceptance efficiency region $R_g > 70\%$. In general, the performances of the P + P and

[1]Example code is provided at https://github.com/zhaoli-IHEP/Higgs-ML.

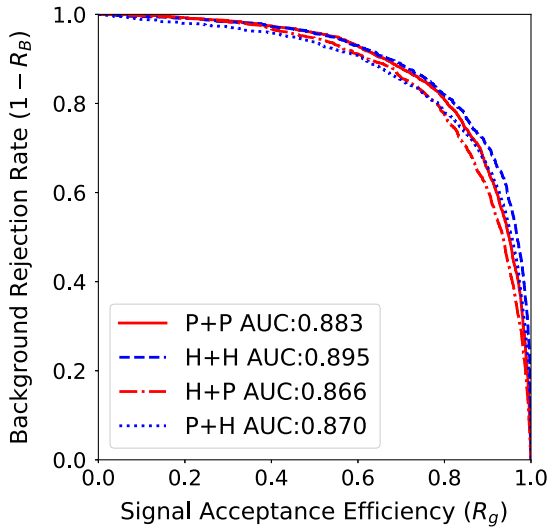

FIG. 3. The background rejection rate $1-R_B$ as a function of the signal acceptance efficiency R_g for our CNN. The symbol "P(H) + P(H)" means training with the PYTHIA (HERWIG) data and testing with the PYTHIA (HERWIG) data.

H + H are similar, which indicates the similar performance of the shower and hadronization schemes in PYTHIA and HERWIG.

The "H + P" and "P + H" are training and testing with different data as a cross-check to illustrate the universality of the CNN model. It makes sense to compare the performance of the CNN models, which are trained with the different data but tested with the same data. The CNN models are universal if their performances are similar. By comparing the "P + P(H)" to the "H + P(H)" in Fig. 3, the performance of the CNN model tested with different data is just slightly worse than that tested with same data in all the signal acceptance efficiency regions. It means that our CNN models do not have too much overfitting since they are not overly dependent on the certain data.

The different ratios of the remaining signal and backgrounds can be obtained on the ROC curve in Fig. 3. The uncertainty of κ_g after using the CNN at each point (R_g, R_B) can be expressed as

$$\delta\kappa_g^{\mathrm{CNN}}(R_g, R_B) = \frac{\sqrt{N_g R_g + N_B R_B}}{2N_g R_g}. \tag{3}$$

Figure 4 presents the uncertainty of κ_g after the CNN as a function of the signal acceptance efficiency R_g using the PYTHIA and HERWIG data. At the optimal point $R_g = 70\%$, $\delta\kappa_g^{\mathrm{CNN}}$ can reach about 1.28% by using the P + P and 1.22% by using the H + H. Compared to Table I, it shows that $\delta\kappa_g^{\mathrm{CNN}}$ can be further reduced by 34% for the P + P and 33% for the H + H. The results using the H + H is about 5% smaller than that using the P + P. The small difference of the results may come from the different shower and hadronization schemes in PYTHIA and HERWIG. The results

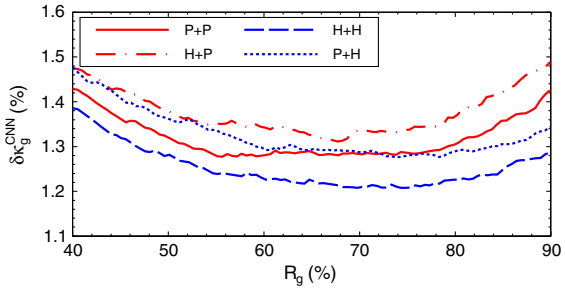

FIG. 4. The uncertainty of κ_g after the CNN as a function of the signal acceptance efficiency R_g. Both training and testing use the PYTHIA and HERWIG data.

of the cross check are slightly worse than that of the training and testing with the same data. Comparing the P + P to the H + P, the uncertainties of κ_g using the H + P is slightly worse than that using the P + P. But the difference between the P + P and the H + P is less than 0.1%, which far exceeds the measurement accuracy of the future electron positron colliders. The H + H and the P + H are in the same situation. The similar results mean that our CNN models do not have too much overfitting and the results are reliable.

In the previous part, one image is constructed with the information of all the final state stable particles in an event. To gain insight into the improvement by the CNN and find the most important features of the signal and background, different images are constructed with different final state constituents. The following analysis only uses the PYTHIA data. Figure 5 shows the uncertainty of κ_g after the CNN as a function of the signal acceptance efficiency R_g using the different images. The line marked as "all" is the result using the images constructed with the information of all the final state stable particles, and the line marked as "multijet" is the result using the images constructed with the information of all the jets clustered by anti-k_T algorithm in an event. The multijet result is slightly better than the all result in the region $R_g \in [60\%, 70\%]$. However, the difference of the all and multijet results is less than 0.2% at the optimal points

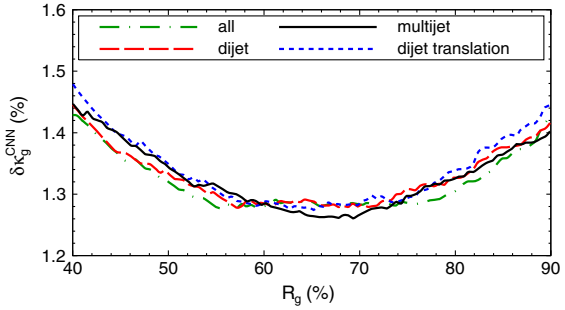

FIG. 5. The uncertainty of κ_g after the CNN as a function of the signal acceptance efficiency R_g using the different images, which are constructed with the information of all the final state stable particles, all the jets, or the first two jets sorted by their energy.

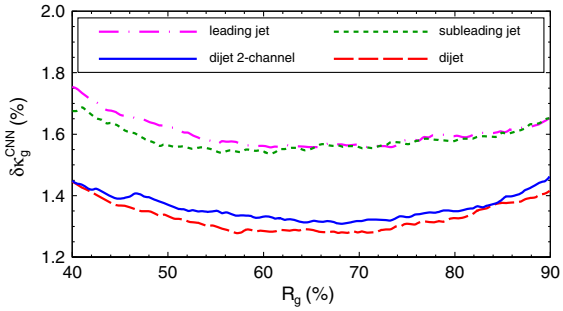

FIG. 6. The uncertainty of κ_g after the CNN varies with the signal acceptance efficiency R_g using different single-jet images, which are constructed only with the leading jet or subleading jet.

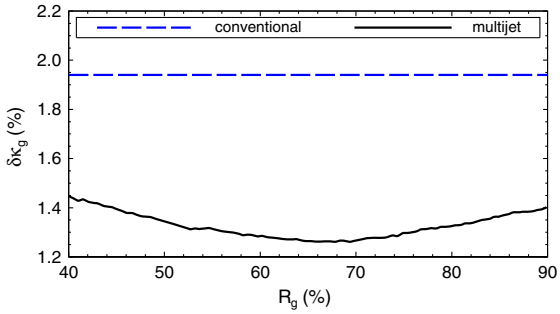

FIG. 7. The best result using the CNN is compared to the result using the conventional method for the PYTHIA data.

and can be ignored. This indicates that the information of jets makes a major contribution to the identification of the signal and background processes. The reason is that most of the information except the jets in an event is the lepton pair, which are very similar in the signal and background processes after using the kinematic cuts. The line marked as "dijet" is the result using the images only constructed with the information of the leading and subleading jets. The all and dijet results are very similar, which shows that the leading and subleading jets nearly contribute all the features for the CNN. The multijet and dijet results are also very similar since most of the events only have two jets with a large jet cone of $R = 1.5$. If the images are constructed only with the leading and subleading jets, the center of the two jets can be chosen as the image center. Then the constituents of the two jets are discretized into pixels to obtain the "dijet translation" images. By this operation, the jets will not be split into two parts at the margins of the image. It can be seen that the dijet translation and the dijet results are also very similar, which indicates that the symmetry property in the ϕ direction has been recognized by the CNN.

After showing that the information of the leading and subleading jets makes a major contribution to the identification of the signal and background processes, we further analyze the contribution of each jet. Figure 6 shows the uncertainty of κ_g after the CNN as a function of the signal acceptance efficiency R_g using the different single-jet images. Each single-jet image has the size $2R \times 2R$ with the jet cone $R = 1.5$ and is designed to have 34×34 pixels. The jet axis is chosen at the image center so that there is a complete jet on the single-jet image. The lines marked as "leading jet" and "subleading jet" represent the results using the leading jet images and the subleading jet images, respectively. We can see that the leading and subleading jets are equally important for the identification. Then the leading and subleading jet images as two different channels are combined as the "dijet two-channel" by analogy with the recognition of color images, with red, green, and blue intensities treated as separate input layers. Compared to the dijet, which puts the leading and subleading jets in one image, the dijet two-channel removes the relative location information of the two jets. It can be seen that the dijet two-channel result is just slightly worse than the dijet result, so the relative location information of the jets is not important for this discrimination. From the above analysis, we can conclude that the leading and subleading jets make a major contribution to the identification of the signal and background processes.

In the third section, the analysis shows that the jets in the signal process are mainly gluon jets and the jets in the background processes can be mainly divided into quark jets and W/Z jets. The three types of jets have different energy distributions of their constituents. Each jet image using energy as pixel intensity records the energy distribution of the jet constituents. This information can be extracted from the jet images by the CNN to identify the signal and background processes. Therefore, the energy distributions of the leading and subleading jets constituents make a major contribution to the identification of the signal and background processes.

Figure 7 shows the best result using the CNN (the line marked as multijet) and the result using the conventional method (the line marked as "conventional") for the PYTHIA data. Comparing to the result using the conventional method, the CNN has a significant improvement in a wide signal acceptance efficiency region. At the optimal point $R_g = 70\%$, the uncertainty of κ_g can be decreased from 1.94% to about 1.26% by using the CNN and reduced by about 35% compared to that using the conventional method for the PYTHIA data. Moreover, the result using the HERWIG data is similar to that using the PYTHIA data.

VI. CONCLUSIONS

In this paper, the CNN is used to improve the precision measurement of the Higgs boson-gluon effective coupling at lepton colliders. By using the CNN the uncertainty of κ_g can be decreased from 1.94% to about 1.28% using the PYTHIA data and from 1.82% to about 1.22% using the HERWIG data in the channel of a Z boson decaying to a lepton pair in the MC simulation for the center-of-mass energy 250 GeV and integrated luminosity 5 ab^{-1}.

The difference between the expected κ_g uncertainties using the PYTHIA and the HERWIG data is less than 0.1%. Moreover, the performance of the CNN using different final state constituents is proof that the energy distributions of the leading and subleading jets constituents play a major role on the identification and the optimal uncertainty of κ_g using the CNN is reduced by about 35% compared to that using the conventional method.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under Grant No. 11675185. Y. W. is supported by the China Postdoctoral Science Foundation under Grant No. 2016M601134. The authors want to thank Yu Bai, Gang Li, Qiang Li, and Manqi Ruan for helpful discussions.

[1] K. Blum, R. T. D'Agnolo, and J. Fan, J. High Energy Phys. 01 (2013) 057.
[2] X.-F. Han, L. Wang, J. M. Yang, and J. Zhu, Phys. Rev. D **87**, 055004 (2013).
[3] M. B. Einhorn, in Conference on Unified Symmetry in the Small and in the Large Coral Gables, Florida, 1993 (1993), pp. 407–420, http://inspirehep.net/record/353863.
[4] S. Kanemura, Y. Okada, E. Senaha, and C. P. Yuan, Phys. Rev. D **70**, 115002 (2004).
[5] X.-G. He, Y. Tang, and G. Valencia, Phys. Rev. D **88**, 033005 (2013).
[6] A. Moyotl, S. Chamorro, H. Castilla-Valdez, and M. A. Prez, arXiv:1610.06299.
[7] S. Baek and X.-B. Yuan, Phys. Lett. B **774**, 662 (2017).
[8] W.-S. Hou and M. Kikuchi, Phys. Rev. D **96**, 015033 (2017).
[9] S. Kanemura, M. Kikuchi, K. Sakurai, and K. Yagyu, Phys. Rev. D **96**, 035014 (2017).
[10] S. Paehr and G. Weiglein, Eur. Phys. J. C **78**, 222 (2018).
[11] M. E. Peskin, arXiv:1207.2516.
[12] M. E. Peskin, in Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, 2013 (2013), http://inspirehep.net/record/1272697.
[13] CEPC Study Group, arXiv:1809.00285.
[14] CEPC Study Group, arXiv:1811.10545.
[15] X. Mo, G. Li, M.-Q. Ruan, and X.-C. Lou, Chin. Phys. C **40**, 033001 (2016).
[16] M. Bicer *et al.* (TLEP Design Study Working Group), J. High Energy Phys. 01 (2014) 164.
[17] W. Barletta, M. Battaglia, M. Klute, M. Mangano, S. Prestemon, L. Rossi, and P. Skands, Nucl. Instrum. Methods Phys. Res., Sect. A **764**, 352 (2014).
[18] M. Benedikt and F. Zimmermann, Proc. Sci., LeptonPhoton2015 (**2016**) 052.
[19] T. Behnke, J. E. Brau, B. Foster, J. Fuster, M. Harrison, J. M. Paterson, M. Peskin, M. Stanitzki, N. Walker, and H. Yamamoto, arXiv:1306.6327.
[20] H. Baer, T. Barklow, K. Fujii, Y. Gao, A. Hoang, S. Kanemura, J. List, H. E. Logan, A. Nomerotski, M. Perelstein *et al.*, arXiv:1306.6352.
[21] C. Adolphsen, M. Barone, B. Barish, K. Buesser, P. Burrows, J. Carwardine, J. Clark, H. M. Durand, G. Dugan, E. Elsen *et al.*, arXiv:1306.6353.
[22] C. Adolphsen, M. Barone, B. Barish, K. Buesser, P. Burrows, J. Carwardine, J. Clark, H. M. Durand, G. Dugan, E. Elsen *et al.*, arXiv:1306.6328.
[23] H. Abramowicz *et al.*, arXiv:1306.6329.
[24] J. Gao, J. High Energy Phys. 01 (2018) 038.
[25] Y. Bai, C. Chen, Y. Fang, G. Li, M. Ruan, S. Jingyuan, W. Bo, K. Panyu, L. Boyang, and L. Zhanfeng,, Chin. Phys. C **44**, 013001 (2020).
[26] J. Gallicchio and M. D. Schwartz, J. High Energy Phys. 04 (2013) 090.
[27] J. Shelton, in Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales (TASI 2012): Boulder, Colorado, 2012 (2013), pp. 303–340, http://inspirehep.net/record/1217434.
[28] A. J. Larkoski, I. Moult, and B. Nachman, arXiv:1709.04464.
[29] G. Li, Z. Li, Y. Liu, Y. Wang, and X. Zhao, Phys. Rev. D **98**, 076010 (2018).
[30] P. Chiappetta, P. Colangelo, P. De Felice, G. Nardulli, and G. Pasquariello, Phys. Lett. B **322**, 219 (1994).
[31] J. Cogan, M. Kagan, E. Strauss, and A. Schwartzman, J. High Energy Phys. 02 (2015) 118.
[32] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, J. High Energy Phys. 07 (2016) 069.
[33] L. G. Almeida, M. Backovi, M. Cliche, S. J. Lee, and M. Perelstein, J. High Energy Phys. 07 (2015) 086.
[34] J. Pearkes, W. Fedorko, A. Lister, and C. Gay, arXiv:1704.02124.
[35] P. Baldi, K. Bauer, C. Eng, P. Sadowski, and D. Whiteson, Phys. Rev. D **93**, 094034 (2016).
[36] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban, and D. Whiteson, Phys. Rev. D **94**, 112002 (2016).
[37] L. Lonnblad, C. Peterson, and T. Rognvaldsson, Phys. Rev. Lett. **65**, 1321 (1990).
[38] L. Lonnblad, C. Peterson, and T. Rognvaldsson, Nucl. Phys. **B349**, 675 (1991).
[39] C. Peterson, T. Rognvaldsson, and L. Lonnblad, Comput. Phys. Commun. **81**, 185 (1994).
[40] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz, J. High Energy Phys. 01 (2017) 110.
[41] Y. Lecun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, Neural Comput. **1**, 541 (1989).
[42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, J. Mach. Learn. Res. **15**, 1929 (2014).

[43] F. Chollet *et al.*, Keras, https://github.com/keras-team/keras (2015).
[44] J. Yan, S. Watanuki, K. Fujii, A. Ishikawa, D. Jeans, J. Strube, J. Tian, and H. Yamamoto, Phys. Rev. D **94**, 113002 (2016).
[45] W. Kilian, T. Ohl, and J. Reuter, Eur. Phys. J. C **71**, 1742 (2011).
[46] M. Moretti, T. Ohl, and J. Reuter, arXiv:hep-ph/0102195.
[47] T. Sjostrand, S. Mrenna, and P. Z. Skands, J. High Energy Phys. 05 (2006) 026.
[48] J. Bellm *et al.*, Eur. Phys. J. C **76**, 196 (2016).
[49] M. Cacciari, G. P. Salam, and G. Soyez, Eur. Phys. J. C **72**, 1896 (2012).
[50] K. Simonyan and A. Zisserman, arXiv:1409.1556.
[51] D. P. Kingma and J. Ba, arXiv:1412.6980.
[52] A. Krizhevsky, I. Sutskever, and G. E. Hinton, in *Advances in Neural Information Processing Systems 25*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Curran Associates, Inc., Red Hook, NY, 2012), pp. 1097–1105.
[53] J. Guo, J. Li, T. Li, F. Xu, and W. Zhang, Phys. Rev. D **98**, 076017 (2018).