Determination of the impact parameter in high-energy heavy-ion collisions via deep learning*

Pei Xiang(相培)[1] Yuan-Sheng Zhao(赵渊晟)[1] Xu-Guang Huang(黄旭光)[1,2†]

[1]Department of Physics and Center for Field Theory and Particle Physics, Fudan University, Shanghai 200433, China
[2]Key Laboratory of Nuclear Physics and Ion-beam Application (MOE), Fudan University, Shanghai 200433, China

Abstract: In this study, Au+Au collisions with an impact parameter of $0 \le b \le 12.5$ fm at $\sqrt{s_{NN}} = 200$ GeV are simulated using the AMPT model to provide preliminary final-state information. After transforming this information into appropriate input data (the energy spectra of final-state charged hadrons), we construct a multi-layer perceptron (MLP) and convolutional neural network (CNN) to connect final-state observables with the impact parameters. The results show that both the MLP and CNN can reconstruct the impact parameters with a mean absolute error approximately 0.4 fm, although the CNN behaves slightly better. Subsequently, we test the neural networks at different beam energies and pseudorapidity ranges in this task. These two models work well at both low and high energies. However, when conducting a test for a larger pseudorapidity window, the CNN exhibits a higher prediction accuracy than the MLP. Using the Grad-CAM method, we shed light on the 'attention' mechanism of the CNN model.

Keywords: heavy ion collisions, impact parameter, deep learning

DOI: 10.1088/1674-1137/ac6490

I. INTRODUCTION

As a unique means of generating quark-gluon plasma (QGP) on earth, high-energy heavy-ion collision experiments provide opportunities to study this type of extremely hot and dense matter. With further research on the collective behaviour of quarks and gluons, the deep structure of a nucleus and the state of the universe a few microseconds after the Big Bang have been discovered. For heavy-ion collisions, besides collision energy, the impact parameter (denoted by b) is a crucial quantity that determines the initial geometry of a collision. Numerous quantities have essential correlations with the impact parameter. For example, the elliptic flow of hadrons is sensitive to the impact parameter [1]. Electromagnetic (EM) fields in heavy-ion collisions roughly satisfy $e|\text{Field}| \propto \sqrt{s} f(b/R_A)$, where $f(b/R_A)$ is a function of b/R_A with R_A as the nucleus radius [2, 3]. When studying the EM properties of QGP, dilepton production is a significant probe. For lepton pair production, both the cross section and azimuthal asymmetry have a strong dependence on the impact parameter [4]. The recently observed hyperon spin polarization increases with increasing impact parameter of collisions [5, 6]. However, the impact parameter of a single collision cannot be measured directly in heavy-ion experiments. Usually, it is estimated using particular final-state observables sensitive to it, such as charged-particle multiplicity. By introducing the concept of centrality, which is defined as classes categorized by b, and comparing experimental data with simulation results using, for example, the Glauber model, one can determine the rough interval of the impact parameter of an event [7].

Owing to their powerful ability to establish a reliable map between input data and the target value without significant prior knowledge, deep learning (DL) methods are widely used in both science research and our daily life. When applying DL algorithms to face recognition tasks, a machine can identify a person's ID with his/her facial information [8]. For heavy ion collisions, the impact parameter can be viewed as one of the IDs of an event. Several studies have proved the effectiveness of DL methods on impact parameter 'recognition' [9–17]. From a simple neural network [9] to a PointNet model [13] and boosted decision trees [17], with the development of machine learning algorithms, more appropriate learning methods have been proposed to improve the performance of 'recognition' and satisfy experimental requests. However,

Received 17 March 2022; Accepted 7 April 2022; Published online 27 May 2022

* Supported by the National Natural Science Foundation of China (12075061) and Shanghai NSF (20ZR1404100)

† E-mail: huangxuguang@fudan.edu.cn

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Article funded by SCOAP[3] and published under licence by Chinese Physical Society and the Institute of High Energy Physics of the Chinese Academy of Sciences and the Institute of Modern Physics of the Chinese Academy of Sciences and IOP Publishing Ltd

most of this research only involves collisions at low or intermediate energies [9–16]. Although a recent study [17] considered LHC energies, the adopted machine learning model is not a deep neural network (DNN).

Here, we investigate RHIC Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV and choose final-state charged hadrons as probes. After transforming the momentum information of these particles into energy spectra as input data for the learning models, we use a multi-layer perceptron (MLP) model and convolutional neural network (CNN) model to obtain a map between the energy spectrum and impact parameter, respectively. Then, we analyze the influences of beam energy and the range of pseudorapidity on the accuracy of predictions. Furthermore, we examine the interpretability of the CNN serving as a regression machine. An 'attention' map of the CNN model is obtained using the Grad-Cam algorithm, and all of the collisions are simulated by a multi-phase transport (AMPT) model [18].

II. DEEP LEARNING ALGORITHMS

The discovery of the Higgs Boson in 2012 completed the jigsaw of elementary particles according to the current Standard Model. Soon after, CERN announced the Higgs boson machine learning challenge [19] to the public. The goal of this challenge is to help experimentalists better distinguish the signal of Higgs boson decay from background noise using machine learning (ML) methods. ML was found to be extremely effective in this task.

As a branch of the field of artificial intelligence, ML technology has promoted intelligentization in many areas of industry, such as autonomous driving, smart mobile electronic devices, and the internet. Because scientists often deal with large quantities of data or require information that is difficult to obtain using traditional methods, they have applied ML techniques to fundamental science research, and physics is no exception. On the whole, the applications of ML in physics can be divided into two categories. The first is the replacement of physical models with ML models if the latter are more effective for specific problems. By training ML models with a large quantity of data, a map can be constructed between two or more physical quantities. For instance, by training neural networks to approximately imitate wave functions, we can construct a map between the potential and one particle's energy without solving the Schrödinger equation [20] or quantum many-body problem [21]. The second is to recognize the target signal (such as a physical phenomenon) from the background with noise. For example, deep neural networks can help distinguish the Higgs boson or other exotic particles of interest (signal) from other particles (background) [22]. In short, ML algorithms can be used to deal with regression and classification problems in physics. As a branch of ML, DL has become the most popular AI method in physics research. In the field of relativistic heavy ion collisions, DL has been applied to the problems of QCD phase transition [23– 25], relativistic hydrodynamics [26], jet structure [27, 28], the search for the chiral magnetic effect [29], and recognition of the initial clustering structure in nuclei [30]. Among various DL algorithms, the DNN, especially the convolutional neural network, are commonly used.

The DNN is one of the early DL models. Owing to its remarkable ability to realize nonlinear mapping with a comparatively simple structure, the DNN is still a preferred tentative model for most regression tasks. A DNN (Fig. 1), also known as a multi-layer perceptron (MLP), is composed of an input layer, sufficient hidden layers to be considered 'deep', and an output layer.

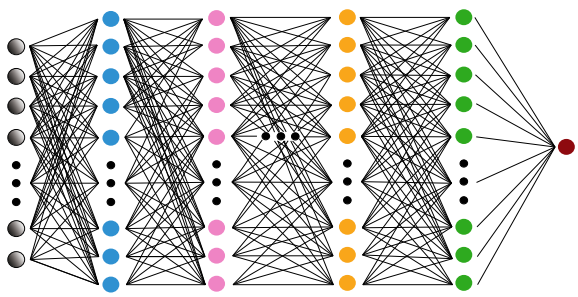

Fig. 1. (color online) Architecture of the multi-layer perceptron (MLP) we use in this study. It contains four hidden layers. The first layer is composed of 512 neurons, the second layer has 256 neurons, the third layer has 128 neurons, and the final hidden layer has 64 neurons.

The CNN commonly appears in 2D image-related tasks. A typical CNN consists of the input layer, convolutional layers, pooling layers, fully connected layers, and the output layer. Our CNN architecture is shown in Fig. 2.

As a supervised learning regression task, impact parameter determination aims at constructing a map between the input observables and a single value, that is, the impact parameter of an event. Thus, it is appropriate to choose the mean squared error (MSE) as the loss function to examine the performance of the learning models. This is defined as

$$\text{Loss} = \frac{1}{N_{\text{batch}}} \sum_{i=1}^{N_{\text{batch}}} (y_i^{\text{pred}} - \hat{y}_i^{\text{true}})^2, \tag{1}$$

where $\hat{y}_i^{\text{true}}$ is the true value of the impact parameter of an event among a batch of events with a size of N_{batch}, and y_i^{pred} is the output of the MLP/CNN model, corresponding to the prediction value.

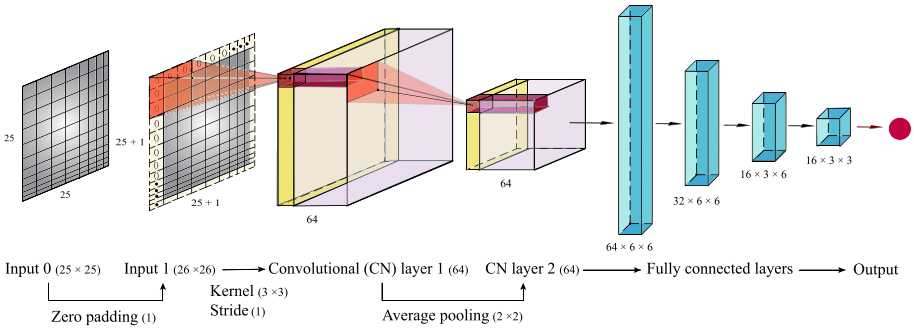

Fig. 2. (color online) Structure of the convolutional neural network (CNN). It contains two convolutional layers and four fully-connected layers.

III. SIMULATION OF EVENTS AND SELECTION OF OBSERVABLES

We consider Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV and use a multi-phase transport (AMPT) model to perform the simulation. The AMPT model [18] is a hybrid transport model that contains four basic stages: the initial condition, partonic scattering, hadronization, and hadronic interaction. The initial condition is generated using the HIJING model [31]. Scatterings among partons are modeled by Zhang's Parton Cascade (ZPC) model [32]. Two different versions of the AMPT model are available for the process of hadronization. The first is the default model, in which partons are recombined with their parent strings and the Lund string fragmentation model is used to turn partons into hadrons. The second contains a string melting model. This combines partons with hadrons via a quark coalescence model. Finally, the rescatterings of hadronic matter are performed by a relativistic transport (ART) [33]. Here, we choose the model with string melting as the simulator.

The selection of features serving as input information is the first step in a deep learning process. Considering the experimental observability, we choose the momenta of the freezed-out charged hadrons with pseudorapidity η satisfying $-1 \le \eta \le 1$ as observables.

As shown in Sec. II, the input data of an MLP model and that of a CNN model have different forms. Here, for the CNN model, the observables above are transformed into two-dimensional (2D) energy spectra in $(p_x,\ p_y)$ space. Fig. 3 illustrates the case of an event with $b = 1.65$ fm. The x, y components of momenta are cut by the interval $[-1.5, 1.5]$, and the $(p_x,\ p_y)$ space is cut by a 25×25 grid. Each cell in this grid contains the sum of the energies of charged particles whose p_x and p_y are within corresponding intervals. Note that the energy contains not only the momentum but also the mass of the particle. Thus, an energy spectrum carries more physical information than a multiplicity spectrum. By arranging the pixels of a 2D energy spectrum into a 1D chain, data is generated as an input for the MLP model.

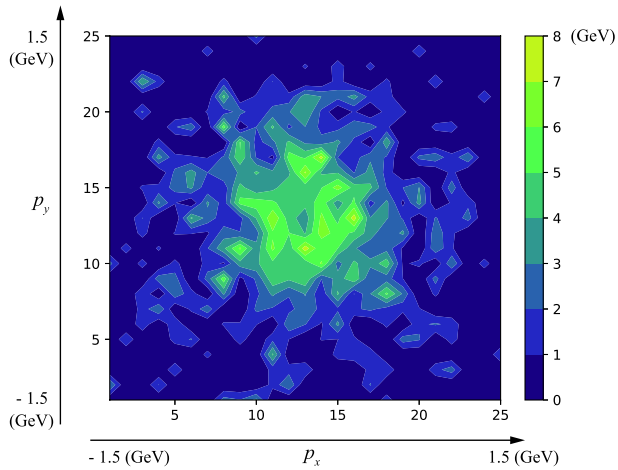

Fig. 3. (color online) Energy spectrum in $(p_x,\ p_y)$ space of final-state charged particles of an event with $b = 1.65$ fm. This is cut by a 25×25 grid. Each unit in the grid represents the sum of the energies of the enclosed particles. This type of energy spectrum is used as the input data for DL models.

IV. RESULTS AND DISCUSSION

We generate 28000 events per centrality and divide them into two parts: 20000 events for training and 8000 events for validation. The interval of the impact parameter $b \in [0, 12.5]$ fm is divided into nine centrality classes [34]. Thus, the total training dataset is composed of 180000 events, and the total validation dataset contains 72000 events. After appropriate training of the model, 20000 events (satisfying a differential distribution $\propto bdb$ in the impact parameter) are input into the model to test its prediction accuracy.

In Fig. 4, we show the errors of the predicted impact parameters compared to the true values. Both the MLP and CNN models exhibit high prediction accuracy for semi-central and semi-peripheral events. The mean error of the CNN model for events with the impact parameter satisfying $2 \le b \le 11$ fm ranges from -0.06 to 0.05 fm, and that of the MLP model ranges from -0.08 to 0.08 fm. The mean absolute prediction error is 0.40 fm for the CNN model and 0.41 fm for the MLP model, indicating

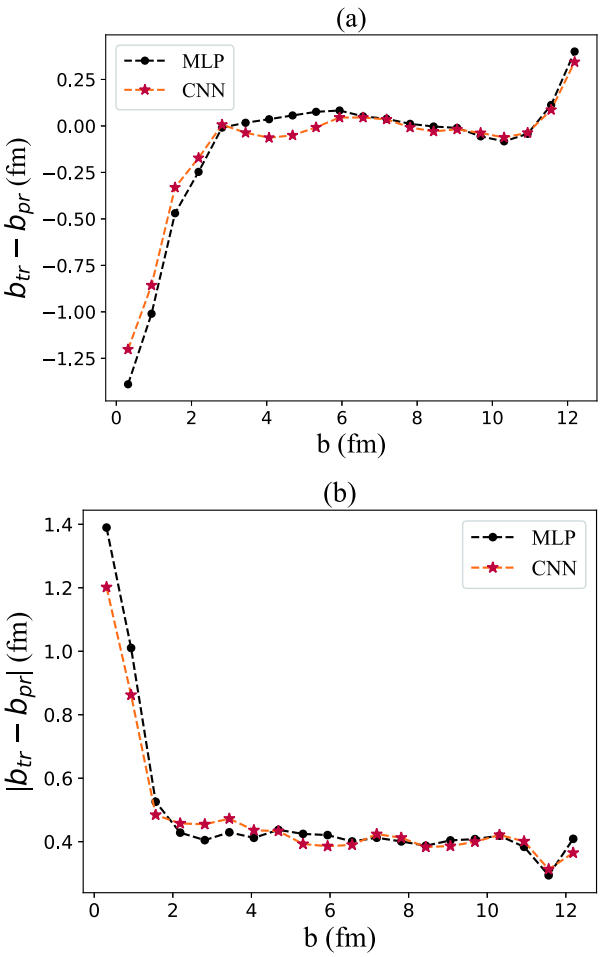

Fig. 4. (color online) Errors between the true values of the impact parameter and those predicted by the MLP/CNN model for Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV: (a) The mean errors between the true values (denoted by $b_{\rm tr}$) and predicted values (denoted by $b_{\rm pr}$); (b) Corresponding mean absolute errors.

that the CNN model performs slightly better than the MLP model. However, the accuracy decreases in central regions. The prediction values are greater than the true values for central collisions and vice versa for peripheral events.

A. Robustness over collision energies

In heavy ion collisions, the beam energy is another crucial quantity that largely determines the bulk properties of the matter created in an event. Low-energy collisions differ from high-energy collisions in many aspects. With a higher beam energy, colliding nuclei generate more partons, and eventually more varieties of hadrons freeze out. To illustrate this, we generate a dataset for events with collision energies $\sqrt{s_{NN}} = 7.7$, 39.0, 62.4, 130.0, 200.0 GeV, corresponding to the beam energy scan program performed at the RHIC. Then, we analyze their differences in the multiplicity and composition of final

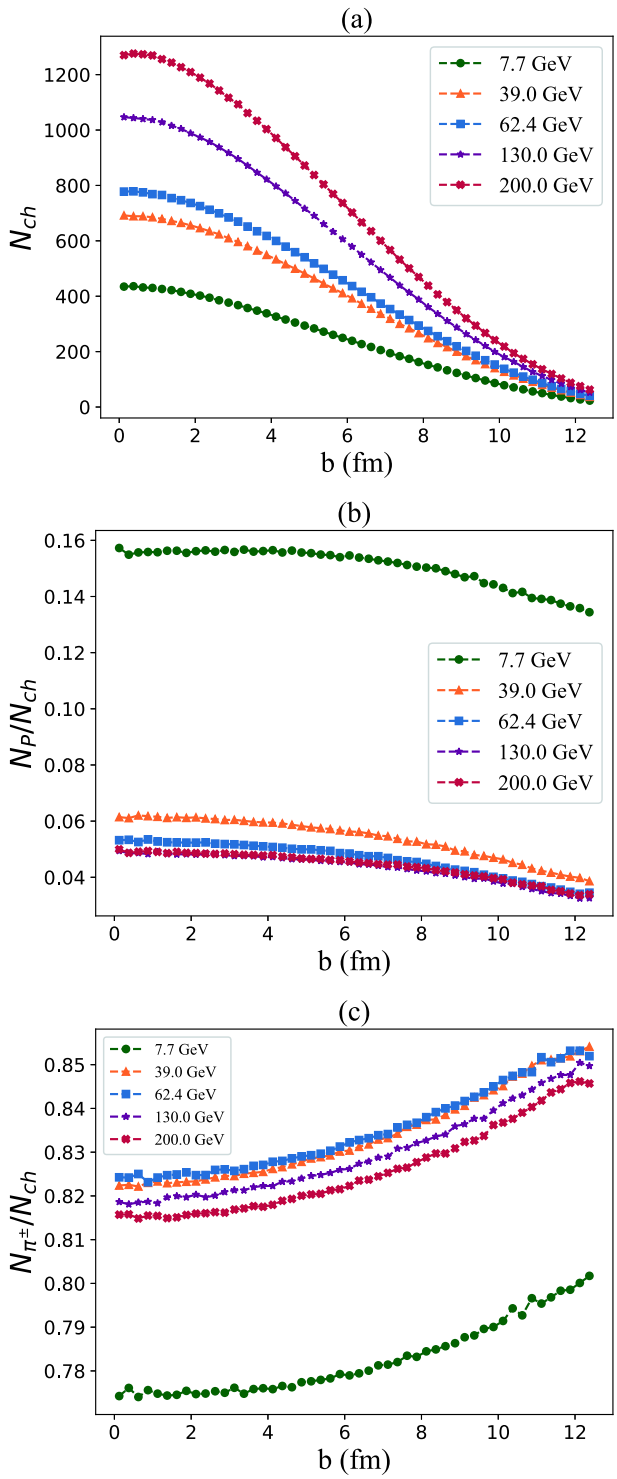

Fig. 5. (color online) Multiplicity and composition of charged particles produced in collisions with different beam energies, that is, $\sqrt{s_{NN}}$ = 7.7, 39.0, 62.4, 130.0, 200.0 GeV: (a) The multiplicity of charged particles; (b) The fraction of protons in charged particles; (c) The fraction of charged pions.

state charged particles. As shown in Fig. 5, the multiplicity of charged particles increases with beam energy. In addition, the fraction of protons in charged hadrons is

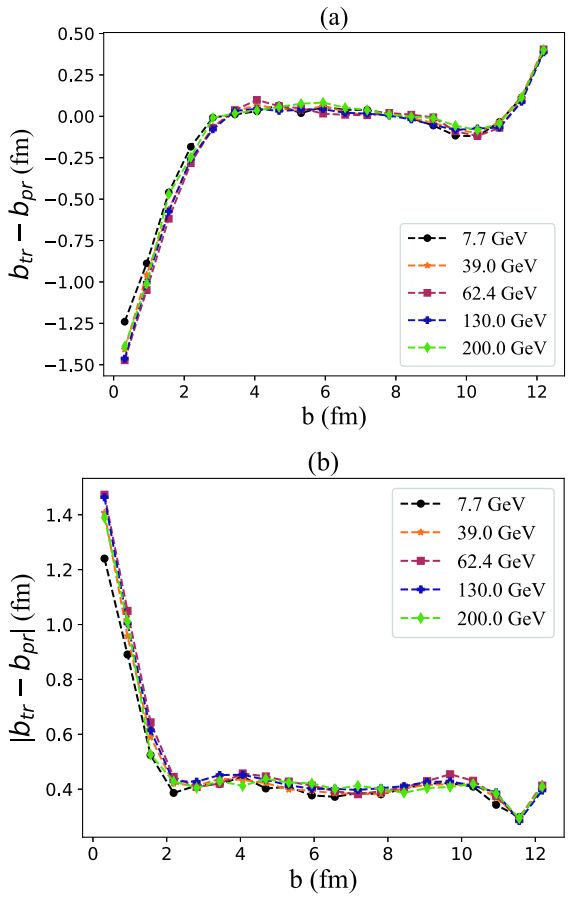

Fig. 6. (color online) Performance of the MLP model in dealing with collisions with different beam energies, that is, $\sqrt{s_{NN}}$ = 7.7, 39.0, 62.4, 130.0, 200.0 GeV: (a) The mean errors between the true values (denoted by $b_{\rm tr}$) and predicted values (denoted by $b_{\rm pr}$); (b) Corresponding mean absolute errors.

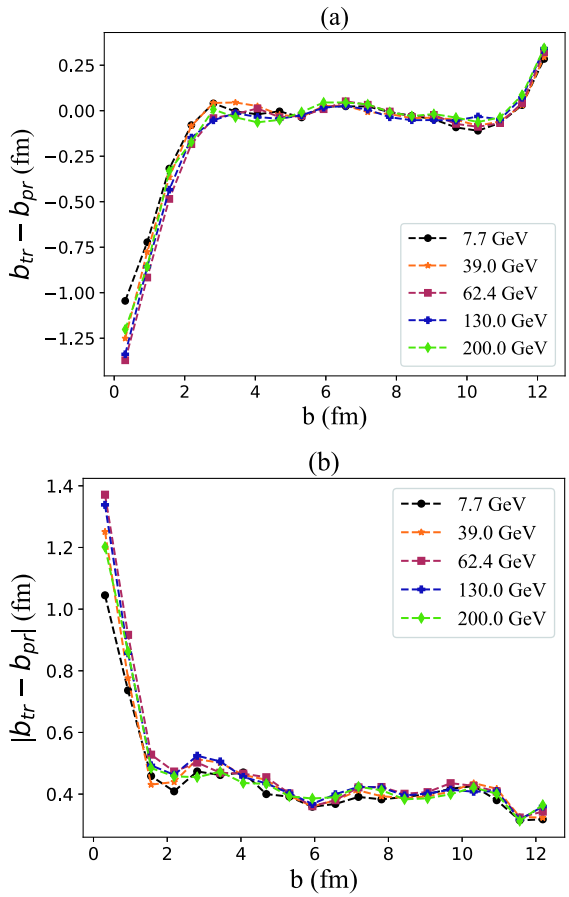

Fig. 7. (color online) Performance of the CNN model in dealing with collisions with different beam energies, that is, $\sqrt{s_{NN}}$ = 7.7, 39.0, 62.4, 130.0, 200.0 GeV: (a) The mean errors between the true values (denoted by $b_{\rm tr}$) and predicted values (denoted by $b_{\rm pr}$); (b) Corresponding mean absolute errors.

higher at lower collision energies; however, the fraction of charged pions is not monotonic in collision energy. Hence, we test whether the differences in multiplicity and composition of the produced hadrons at lower collision energies affect the ability of the DL models.

Therefore, we train and test the DL models to the cases of $\sqrt{s_{NN}}$ = 7.7, 39.0, 62.4, 130.0, 200.0 GeV. The results are shown in Fig. 6 and Fig. 7 for the MLP and CNN models, respectively. The DL models perform well for low and intermediate energy cases. Thus, we find that the DL models are robust against different collision energies.

B. Influence of pseudorapidity cut

In most relativistic heavy ion collision experiments, attention is mainly focused on the mid-rapidity region, that is, $-1 \le \eta \le 1$, owing to the coverage limit of most detectors. However, the mid-rapidity region only covers a small part of final-state charged particles [35] (see Fig. 8). Therefore, more information can be obtained from the

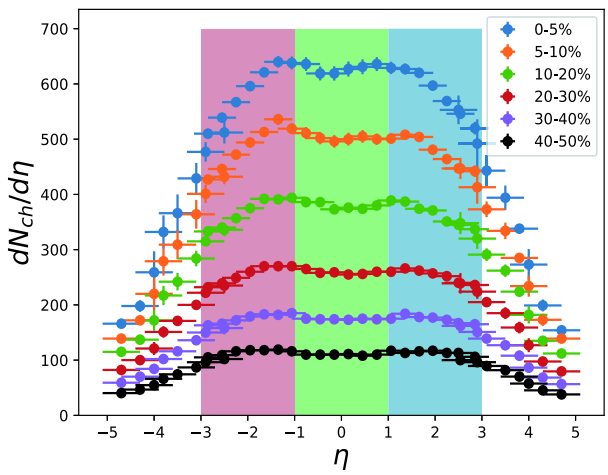

Fig. 8. (color online) Distributions of $dN_{\rm ch}/d\eta$ for centrality ranges of, top to bottom, 0%–5%, 5%–10%, 10%–20%, 20%–30%, 30%–40%, and 40%–50% [35]. The particles in the pseudorapidity window of $-3 \le \eta \le 3$ are divided into three parts, [-3, -1], [-1, 1], and [1, 3], corresponding to three channels, 'R', 'G', and 'B', of the input layer.

energy spectrum if we can observe a larger region in pseudorapidity. The octagon detector in the PHOBOS experiment can accept charged particles with $|\eta| < 3.2$ [36], and the recent upgrade of the inner Time Projection Chamber (iTPC) detector at the RHIC can extend the rapidity acceptance to $-1.5 \le \eta \le 1.5$ [37]. Thus, we take the regions of $-3 \le \eta \le -1$ and $1 \le \eta \le 3$ into consideration.

Now, we expand the training and detection region to $-3 \le \eta \le 3$. Because particles with positive pseudorapidity move in the opposite direction to those with negative pseudorapidity, we add two extra channels into the input layer of the CNN. Similar to inputting a colored picture with three channels 'RGB' into the CNN, we choose the region $-3 \le \eta \le -1$ as channel 'R,' $-1 \le \eta \le 1$ as channel 'G,' and $1 \le \eta \le 3$ as channel 'B' (Fig. 8). Then, we feed the energy spectra with the three channels above to the CNN we trained with only data in the 'G' region. The CNN exhibits a higher prediction accuracy than that with one channel (Fig. 9). The mean absolute prediction error of the CNN for $2 \le b \le 12.5$ fm is 0.35 fm. This value is 0.40 fm in the case of the CNN with only one input channel. Thus, extending the pseudorapidity window can truly improve the performance of regression.

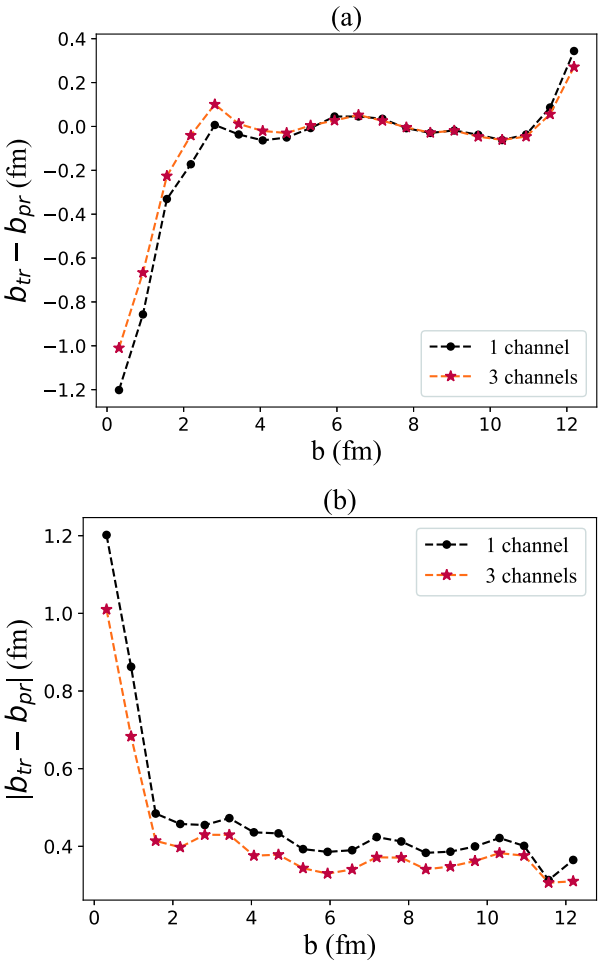

Fig. 9. (color online) Errors between the true values of the impact parameter and those predicted by the CNN model with $\sqrt{s_{NN}} = 200$ GeV. The number of input channels is one and three, respectively: (a) The mean errors between the true values (denoted by b_{tr}) and predicted values (denoted by b_{pr}); (b) Corresponding mean absolute errors.

C. Comparison with the multiplicity method

As mentioned previously, the impact parameter of a single event cannot be directly measured experimentally. This also applies to geometric quantities, such as the participant number N_{part} and binary collision number N_{coll}. Instead, one can introduce the quantity of 'centrality,' which is usually expressed as a percentage of the total nuclear interaction cross section [7] and is strongly correlated with the impact parameter b, to estimate an event's b range. In heavy ion collision experiments, the multiplicity of final state charged particles (N_{ch}) is usually chosen as the main observable to classify event centrality. It is assumed that the average multiplicity of charged particles decreases monotonically when the impact parameter increases. With the Glauber model and Monte Carlo method, we can establish the centrality classes for heavy ion collisions [38]. By comparing the charged-particle multiplicity of an event measured by experiments with the results given by the MC-Glauber model, one can determine this event's centrality and further estimate its impact parameter.

Here, we propose a scheme to compare the uncertainty on b determined using the multiplicity method with the CNN method. Based on the above assumption that the impact parameter is monotonically related to charged-particle multiplicity, if we select a batch of events with the same multiplicity (N_{ch}), their impact parameters should satisfy a Gaussian-like distribution. We select the standard deviation (σ_b^{mult}) of these impact parameters as the quantity to characterize the uncertainty of mapping N_{ch} to the corresponding impact parameter, which can be taken as the center of the Gaussian-like distribution, that is, μ_b^{mult}. Then, a batch of events with the same impact parameter μ_b^{mult} are generated by the AMPT model. With a well-trained CNN model, we can obtain the corresponding predicted impact parameters and calculate their standard deviation, that is, σ_b^{CNN}. By comparing σ_b^{mult} with σ_b^{CNN}, we can evaluate the uncertainty on impact parameter determination using the multiplicity or CNN method.

Owing to event-by-event fluctuation in the nucleon distribution of nuclei before collision and the complex evolution process after the collision, events with the same impact parameter should have different charged-particle multiplicities and energy spectra. Thus, evaluating the uncertainty of these two methods is equivalent to measuring the degree of single-valued correspondence between the impact parameter and multiplicity or energy spec-

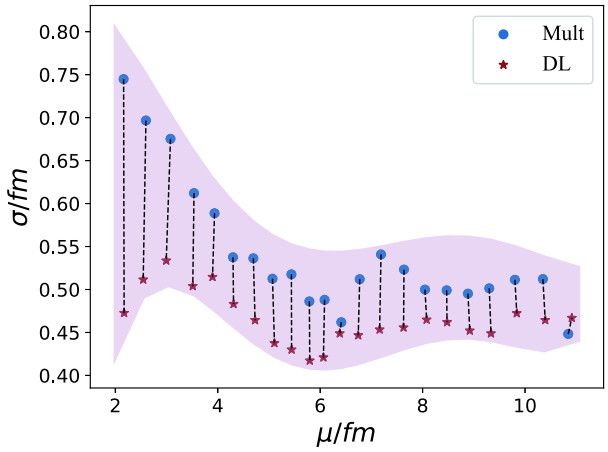

Fig. 10. (color online) Comparing the standard deviation of the b distribution obtained using the multiplicity method with that obtained using the CNN. μ is the mean value of the b distribution. 'Mult' denotes the multiplicity method and 'DL' refers to the deep learning method, which is the CNN model here.

trum. From the results shown in Fig. 10, it can be concluded that the charged-particle energy spectrum with CNN as the identifier of the impact parameter is better than that with the multiplicity method.

V. OBTAINING PHYSICAL INFORMATION FROM THE CNN

DL algorithms have revealed a strong capability to construct a map between the input data and target. As a result, we can succeed in various regression or classification tasks without prior knowledge. However, most widely used DL models are not interpretable. Because of their 'complexity' and 'dimensionality', it is difficult to understand how these models work and obtain instructive information from them [39]. Therefore, these DL models are viewed as 'black boxes' in most cases. More effort has been made to open the 'black boxes' of DL algorithms. By analyzing the DNNs on the information plane [40], one can understand the training and learning processes and hence the internal representations of the DNNs [41]. In addition, for CNNs, which are usually used in visual recognition tasks, there have been a number of studies aimed at visualizing them. By operating global average pooling and defining a quantity measuring the importance of neurons in a CNN, one can generate a class activation map (CAM) [42], with which we can localize the crucial regions of a 2D matrix for a CNN to succeed in classification tasks. Based on the CAM method, R. R. Selvaraju *et al*. proposed a new CNN interpretation method known as gradient-weighted class activation mapping (Grad-CAM) [43]. Compared with the former method, Grad-CAM can be applied to more types of CNNs and can provide more information about what is learned by the neuron network.

In the CAM method, the class activation map for class c (in our case, c is trivial and can be supressed because we have a regression problem instead of a classification problem) is defined as M^c.

$$M^c_{x,y}(A) = \sum_k \omega^c_k f_k(A;x,y). \tag{2}$$

Here, $f_k(A;x,y)$ represents the activation of unit k in the final convolutional layer at spatial location (x,y) of input data A, and ω^c_k is the weight measuring the importance of unit k for class c. In Grad-CAM, ω^c_k is defined as the result of performing global average pooling on the gradient of the score for class c with respect to activations f_k:

$$\omega^c_k = \frac{1}{Z}\sum_x\sum_y \frac{\partial y^c}{\partial f_k(A;x,y)}, \tag{3}$$

where $(1/Z)\sum_x\sum_y$ represents the operation of global average pooling. Subsequently, the gradient-weighted class activation map is given as

$$L^c = \mathrm{ReLU}\left(\sum_k \omega^c_k f^k\right). \tag{4}$$

CAM and Grad-CAM have succeeded in classification problems. However, several adjustments are required in the definitions of the maps to apply them to our CNN for regression. When Grad-CAM meets classification tasks, only the regions positively correlated with the class of interest should be preserved. Thus, a ReLU operation is performed on the linear combination of maps. However, for regression problems, both the positively-correlated and negatively-correlated features should be considered. Consequently, we redefine an activation map as the absolute value of the linear combination of maps.

$$L^c = Abs\left(\sum_k \omega^c_k f^k\right). \tag{5}$$

We obtain average 'attention' maps for impact parameters in the interval $[2,11]$ fm, where our CNN behaves well in prediction (see Fig. 11). It turns out that compared with the cases of central collision, the CNN turns its 'attention' to regions of larger transverse momenta for peripheral collisions. The CNN gives high marks to charged particles with small transverse momenta when it

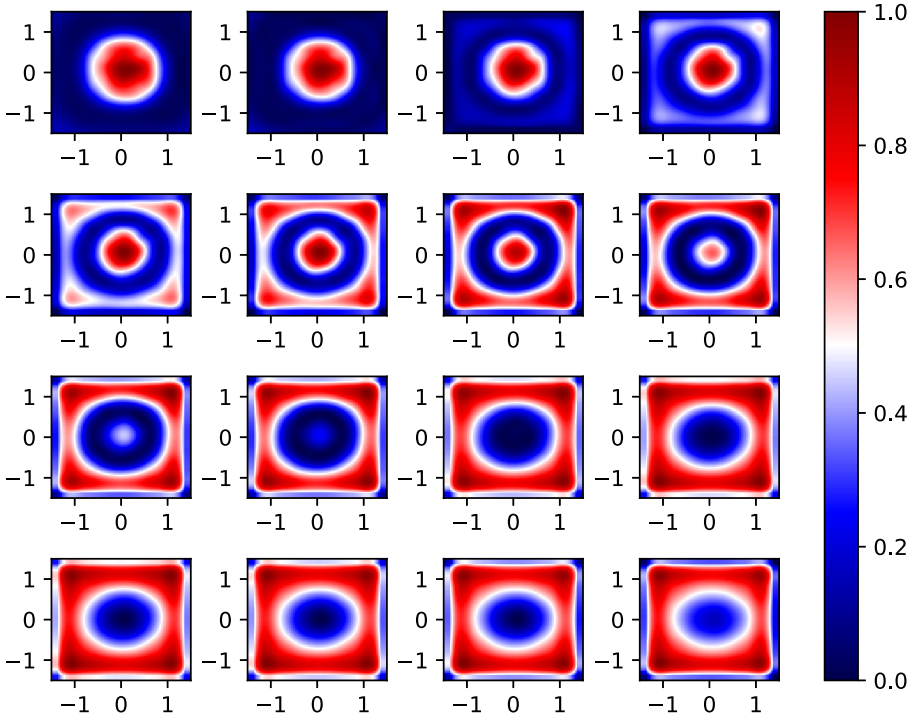

Fig. 11. (color online) Average 'attention' maps for the cases of b = 1.65, 3.54, 4.69, 5.07, 5.44, 5.79, 6.09, 6.41, 6.77, 7.19, 7.64, 8.47, 8.89, 9.30, 9.80, 10.35, and 11.1 fm.

tries to 'recognize' a central event's impact parameter. With an increase in b, the peripheral area of the energy spectrum begins to attract the CNN's attention and becomes more important for distinguishing peripheral events from central ones. This suggests that the CNN focuses on particles with larger transverse momenta for peripheral events.

VI. SUMMARY

In this paper, we investigate the feasibility of applying the DL method to determine a single event's impact parameter in relativistic heavy-ion collisions. By constructing DNN (a MLP) and CNN models, we establish a map between the energy spectrum of final-state charged particles and the impact parameter for a single collision event.

Simulated using the AMPT model, 180000 events of an Au + Au collision at $\sqrt{s_{NN}} = 200$ GeV are generated to train the DL models. In addition, the validation dataset contains 72000 events, which aims at measuring the performance of the DL models. After selecting the best of the trained models, 20000 events are used to test its prediction accuracy. The MLP and CNN models both exhibit high accuracy in the cases of semi-central and semi-peripheral collisions, that is, with an impact parameter of $2-10$ fm. The mean prediction error of the CNN model for events with $2 \le b \le 11$ fm ranges from -0.06 to 0.05 fm, and that of the MLP model ranges from -0.08 to 0.08 fm. The mean absolute prediction error on the b interval $[2, 12.5]$ fm is 0.40 fm for the CNN model. However, these two DL models operate worse for central and peripheral collisions.

To investigate the influence of beam energy in this task, we also apply these two DL models with the same architectures to the cases of $\sqrt{s_{NN}} = 7.7$, 39.0, 62.4, 130.0. The DL models turn out to be effective for both low-energy and high-energy collisions. By extending the pseudorapidity cut, the performance of the CNN is improved, and it reconstructs the impact parameter satisfying $2 \le b \le 12.5$ fm with a mean absolute error of 0.35 fm. Subsequently, we propose a scheme to compare the uncertainty on b determination using the multiplicity method with the CNN method. We find that the charged-particle energy spectrum is a better observable than the multiplicity to realize this type of impact parameter determination.

By modifying the Grad-CAM algorithm for classification tasks to that available for regression tasks, we obtain 'attention' maps for the CNN model. When meeting central collisions, the CNN has a tendency to focus on particles with small transverse momenta and vice versa for peripheral events.

ACKNOWLEDGEMENT

We acknowledge useful discussions with W. B. He, L. G. Pang, X. N. Wang, and K. Zhou.

References

[1] U. Heinz and R. Snellings, Annual Review of Nuclear and Particle Science **63**, 123 (2013)
[2] X.-G. Huang, Reports on Progress in Physics **79**, 076302 (2016)
[3] K. Hattori and X.-G. Huang, Nuclear Science and Techniques **28**, 26 (2017)
[4] C. Li, J. Zhou, and Y.-J. Zhou, Phys. Rev. D **101**, 034015 (2020)
[5] Y.-C. Liu and X.-G. Huang, Nuclear Science and Techniques **31**, 56 (2020)
[6] J.-H. Gao, G.-L. Ma, S. Pu *et al.*, Nuclear Science and Techniques **31**, 90 (2020)
[7] B. Abelev, J. Adam, D. Adamova *et al.*, Phys. Rev. C **88**, 044909 (2013)
[8] S. Balaban, in *Biometric and Surveillance Technology for Human and Activity Identification XII*, Vol. 9457 (International Society for Optics and Photonics, 2015) p. 94570B
[9] C. David, M. Freslier, and J. Aichelin, Phys. Rev. C **51**, 1453 (1995)
[10] S. A. Bass, A. Bischoff, J. Maruhn *et al.*, Phys. Rev. C **53**, 2358 (1996)
[11] F. Haddad, K. Hagel, J. Li *et al.*, Phys. Rev. C **55**, 1371 (1997)
[12] F. Li, Y. Wang, H. Lü *et al.*, Journal of Physics G: Nuclear and Particle Physics **47**, 115104 (2020)
[13] M. O. Kuttan, J. Steinheimer, K. Zhou *et al.*, Phys. Lett. B **811**, 135872 (2020)
[14] F. Li, Y. Wang, Z. Gao *et al.*, Phys. Rev. C **104**, 034608 (2021)
[15] C. Y. Tsang *et al.*, (2021), arXiv: 2107.13985[physics.ins-det]
[16] X. Zhang *et al.*, (2021), arXiv: 2111.06597[nucl-th]
[17] N. Mallick, S. Tripathy, A. N. Mishra *et al.*, Phys. Rev. D **103**, 094031 (2021)
[18] Z.-W. Lin, C. M. Ko, B.-A. Li *et al.*, Phys. Rev. C **72**, 064901 (2005)
[19] C. Adam-Bourdarios, G. Cowan, C. Germain *et al.*, *in NIPS 2014 Workshop on Highenergy Physics and Machine Learning* (PMLR, 2015) pp. 19– 55
[20] K. Mills, M. Spanner, and I. Tamblyn, Phys. Rev. A **96**, 042113 (2017)
[21] G. Carleo and M. Troyer, Science **355**, 602 (2017)
[22] P. Baldi, P. Sadowski, and D. Whiteson, Nature communications **5**, 1 (2014)
[23] Y.-L. Du, K. Zhou, J. Steinheimer *et al.*, European Physical Journal C **80**, 516 (2020)
[24] L.-G. Pang, K. Zhou, N. Su *et al.*, Nature communications **9**, 1 (2018)
[25] R. Wang, Y.-G. Ma, R. Wada *et al.*, Physical Review Research **2**, 043202 (2020)
[26] H. Huang, B. Xiao, Z. Liu *et al.*, Physical Review Research **3**, 023256 (2021)
[27] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz, Journal of High Energy Physics **2017**, 1 (2017)
[28] Y.-L. Du, D. Pablos, and K. Tywoniuk, arXiv: 2106.11271 [hep-ph]
[29] Y.-S. Zhao, L. Wang, K. Zhou *et al.*, arXiv: 2105.13761[hep-ph]
[30] J. He, W.-B. He, Y.-G. Ma *et al.*, Phys. Rev. C **104**, 044902 (2021)
[31] X.-N. Wang and M. Gyulassy, Phys. Rev. D **44**, 3501 (1991)
[32] B. Zhang, Computer Physics Communications. **109**, 193 (1998)
[33] B.-A. Li and C. M. Ko, Phys. Rev. C **52**, 2037 (1995)
[34] Z. Qiu and U. Heinz, Phys. Rev. C **84**, 024911 (2011)
[35] I. Bearden, D. Beavis, C. Besliu *et al.*, Phys. Rev. Lett. **88**, 202301 (2002)
[36] P. Sarin, *Measurement of charged particle multiplicity distribution in Au+ Au collisions up to 200 GeV*, Ph.D. thesis, Massachusetts Institute of Technology (2003)
[37] X. Luo and N. Xu, Nuclear Science and Techniques **28**, 112 (2017)
[38] M. L. Miller, K. Reygers, S. J. Sanders *et al.*, Annual Review of Nuclear and Particle Science **57**, 205 (2007)
[39] Y. Bathaee, Harvard Journal of Law & Technology **31**, 889 (2017)
[40] N. Tishby and N. Zaslavsky, in *2015 IEEE Information Theory Workshop (ITW)* (IEEE, 2015) pp. 1–5
[41] R. Shwartz-Ziv and N. Tishby, arXiv: 1703.00810[cs.LG]
[42] B. Zhou, A. Khosla, A. Lapedriza *et al.*, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016) pp. 2921–2929
[43] R. R. Selvaraju, M. Cogswell, A. Das *et al.*, in *Proceedings of the IEEE international conference on computer vision* (2017) pp. 618–626