Deep learning stochastic processes with QCD phase transition

Lijia Jiang,[1,2] Lingxiao Wang,[2] and Kai Zhou[2,*]
[1]*Institute of Modern Physics, Northwest University, 710069 Xi'an, China*
[2]*Frankfurt Institute for Advanced Studies,*
Ruth Moufang Strasse 1, D-60438, Frankfurt am Main, Germany

(Received 22 March 2021; accepted 22 May 2021; published 25 June 2021)

It is nontrivial to recognize phase transitions and track dynamics inside a stochastic process because of its intrinsic stochasticity. In this paper, we employ the deep learning method to classify the phase orders and predict the damping coefficient of fluctuating systems under Langevin description. As a concrete setup, we demonstrate this paradigm for the scalar condensation in QCD matter near the critical point, in which the order parameter of the chiral phase transition can be characterized in a $1+1$-dimensional Langevin equation for the σ field. In a supervised learning manner, convolutional neural networks accurately classify the first-order phase transition and crossover based on σ field configurations with fluctuations. Noise in the stochastic process does not significantly hinder the performance of the well-trained neural network for phase order recognition. For mixed dynamics with diverse dynamical parameters, we further devise and train the machine to predict the damping coefficients η in a broad range. The results show that it is robust to extract the dynamics from the bumpy field configurations.

DOI: 10.1103/PhysRevD.103.116023

I. INTRODUCTION

The phenomena of phase transition are extensively observed in various many-body systems. Through measuring thermodynamic quantities such as the susceptibility or heat capacity, the information about the phase transition (e.g., the order of the phase transition, the critical exponent, etc.) could be extracted no matter in classical or quantum systems. In addition to its own intricacy for quantum systems, the situation becomes more complicated in a nonequilibrium dynamical evolution, especially for a stochastic process. The intrinsic randomness breaks the deterministic description for such stochastic dynamics, which hinders our further understanding to those exotic nonequilibrium systems, e.g., cold atoms in a moving optical lattice [1] or heavy-quark diffusion in the quark-gluon plasma (QGP) [2,3].

Deep learning with a hierarchical structure of artificial neural networks is emerging as a novel tool to deal with high-level representations of intricate data [4]. With the advancement of hardware and computational power, there is significant progress of applications of deep learning in an increasing number of fields, such as audio recognition, medical image analysis, computer vision, and board game programs, etc. In these cases, the machine has produced results comparable or even superior to human experts. Recently, the deep learning method has also been utilized in the field of physics research [5], such as in nuclear physics [6–12], particle physics [13–17], and condensed matter physics [18–24]. The advantage of the deep learning method is that it could help us extract hidden correlations from complex nonlinear physical systems, which might be difficult to tackle in conventional computation.

Model-free prediction with machine learning on state evolution has been discussed in recent years for nonlinear and chaotic dynamical systems [25–27]. In the present paper, we explore application of deep learning to detect phase transition and dynamical information in stochastic processes, which would be of great potential application in a variety of fields. In such problems, unlike the case with thermal equilibrium, the raw observations (configurations) for the system are nothing but the stochastic time series data. To uncover the phase transition or the dynamical information from limited raw data with stochasticity is, in principle, challenging but crucial for studying the properties of the dynamical system. A related paradigm that has been developed for a deterministic (with minor stochasticity in some stages) dynamical system [10,11] is to train a deep neural network with supervision to identify the phase order of the QCD equation of state in heavy-ion collisions. We generalize the idea further to recognize the phase order and extract the dynamical parameters in a stochastic dynamical process with a phase transition. With regard to the effective inputs to the deep neural networks, while the final-state particle spectra could be a proper choice for deterministic systems [10,11], we feed the event-by-event temporal-spatial scalar field configurations in the final stage to the neural network to identify the dynamics,

*zhou@fias.uni-frankfurt.de

including the phase order and dynamical parameter. Specifically, we design a deep convolutional neural network (CNN) to track the dynamical process with a phase transition. It is worth noting that the magnitude of the noise in the dynamical evolution does not hinder the classification ability of the well-trained CNNs noticeably. The neural network have a consistent performance on predicting the phase order of the configuration with unknown fluctuations. Moreover, we utilize the machine to predict the damping coefficient from configurations in the final stage of the stochastic process. In this part, we evolve the system within the crossover scenario with diverse damping coefficients. As the dominant dynamical parameter, the damping coefficient is set inside a limited range for the training datasets. The test on a broader range of the damping achieves a good performance, which suggests that deep CNNs could help extract crucial dynamical information in such stochastic Langevin dynamics given just the limited raw states.

The paper is organized as following: In Sec. II, we introduce the paradigm that apply deep learning to Langevin dynamics of QCD matter, in which the parameter setups and details of the event-by-event simulation in $1+1$D Langevin equation are described. In Sec. III, the CNNs are adopted to classify the first-order transition and crossover from σ field configurations with fluctuations. The stability of the performance is evaluated for datasets with various noise. In Sec. IV, we prepare mixed configurations with a damping coefficient in the range of (1.0–2.5) and (4.6–5.5) fm^{-1} for training the machine to recognize the damping coefficient; however, the predictions to the dynamics are made for a broader range beyond training. It is found that the dynamics recognition from such a stochastic process with a neural network is robust. In Sec. V, we summarize the main findings in this work and conclude the potentials of our paradigm.

II. DEEP LEARNING LANGEVIN DYNAMICS

In this section, we introduce a deep learning approach to track the stochastic process driven by a Langevin equation, in which the scalar σ field evolves as the order parameter of QCD phase transition [28–31] near the critical point [32,33]. Although the dynamical process is discussed in the context of high-energy heavy-ion collision systems, it can be naturally extended into general nonequilibrium stochastic systems [34] in a broader areas.

In particular, we implement deep CNNs to explore the dynamical process as the flowchart in Fig. 1 shows. To prepare the training data for the CNNs, we numerically solve the event-by-event Langevin equation for the scalar σ field. The temporal-spatial information of the field configurations $\sigma(x,t)$ is recorded when the system temperature drops far below the critical temperature T_c. To perform supervised learning, the configurations labeled with crossover and first-order phase transition are prepared to be our

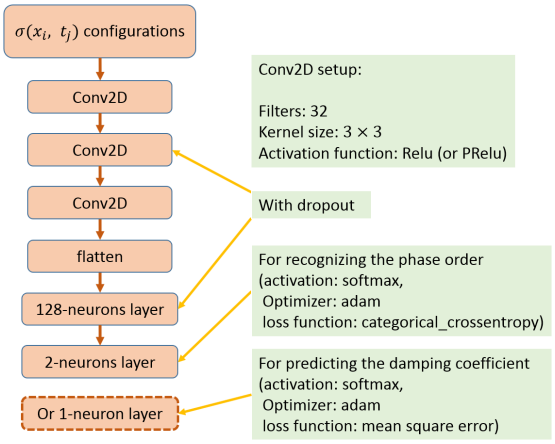

FIG. 1. The architecture of the deep CNNs for recognizing the phase order and predicting the damping coefficient from the σ field configurations. Details of the network structure can be found in the text, especially for the damping regression task. The loss function "categorical cross entropy" is defined as $\mathrm{loss} = -\sum_{i=1}^{C} y_i \log f_i(x)$, where x is input, $f_i(x)$ is output, y_i is the ground truth, and C is the number of the categories.

training set, with which we train the CNNs to classify the order of the phase transition given configurations only in a later stage of the dynamical evolution. As a demonstration of robustness, we test the trained CNNs in various cases with different magnitudes of noise. Furthermore, we also show that, as the key dynamical information, the damping coefficient η can be well recognized via deep CNNs.

With increasing temperature T and/or baryon chemical potential μ, the QCD matter will undergo a phase transition from hadronic matter to the QGP phase. Furthermore, the QCD transition is conjectured to be a crossover at small chemical potential μ (and moderately high temperature) and first order at moderate values of μ (and lower temperature), with a critical point separating the two. Although the QCD phase transition is complicated [35,36], its general thermodynamics and phase behaviors could be describe effectively by models such as the Nambu–Jona-Lasinio model [37–39], quark-meson model [40,41], or linear sigma model (LSM) [42]. As a practical example, the effective potential from the LSM presents a scenario in which the crossover locates at the small chemical potential region and first-order phase transition occurs at the large chemical potential region [43]. In heavy-ion collision (HIC) experiments, the hot and dense fireball created sets an extreme dynamical environment where the QCD phase transition can happen [44–46]. To model the phase transition processes in HICs, the Langevin equation is adopted to describe the semiclassical evolution for the long-wavelength mode of the σ field (for more context and details, see [47,48]):

$$\partial^{\mu}\partial_{\mu}\sigma(t,x) + \eta\partial_t\sigma(t,x) + \frac{\delta V_{\mathrm{eff}}(\sigma)}{\delta\sigma} = \xi(t,x), \quad (1)$$

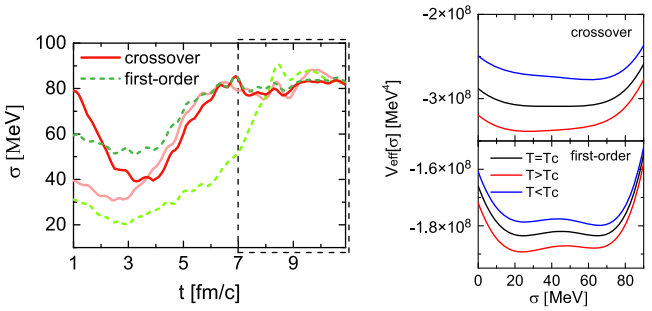

FIG. 2. Dynamical evolution of a scalar σ field in different phase transition scenarios with solid lines standing for crossover and dashed lines for first-order phase transition, with the small box in the upper right corner masking the input range we chose for the CNNs. The right figures are the corresponding effective potential which induces the different phase transitions, in which the upper one shows the crossover with temperature across the critical temperature T_c and the bottom one shows the first-order transition.

where η is the damping coefficient, $\xi(t,x)$ is the noise term, and the effective potential V_{eff} decides the type of phase transition in the stochastic process as shown in Fig. 2 (see more details in Ref. [28]). The terms η and ξ are both from the interaction between the σ field and the thermal background, which follow the fluctuation-dissipation theorem. In the zero-momentum mode limit, the correlation has the form $\langle\xi(t)\xi(t')\rangle = \frac{1}{V} m_\sigma \eta \coth(\frac{m_\sigma}{2T})\delta(t-t')$ [47].

In our calculation, η is taken as a free parameter, while the noise is set as white noise. Note that the above zero-momentum approximation suits only the critical point, where the correlation length is infinite. To be realistic in the crossover and first-order phase transition scenario, we set the spatial noise as $\xi(x) = B\sqrt{\frac{1}{V} m_\sigma \eta \coth(\frac{m_\sigma}{2T})\frac{1}{dt}}G(x)$, where B is a free parameter being encoded in the noise term to control the strength of the white noise and $G(x)$ is a random number generator of the standard normal distribution. Since the correlation length in the crossover and first-order phase transition scenario is finite, the difference of the σ field at different spatial points is large; thus, the magnitude of B reflects the strength of the σ correlation at different spatial points: A larger B corresponds to a weaker σ correlation. Besides, B also eliminates the unreasonable reliance of the noise on the spatial grid size.

In a practical implementation, we adopt the method proposed in Ref. [30] to construct the initial profiles of the σ field, according to the probability distribution function $P(\sigma) \sim \exp(-E(\sigma)/T)$ with the energy function $E[\sigma(\mathbf{x})] = \int d^3x[\frac{1}{2}(\nabla\sigma(x))^2 + V_{\text{eff}}(\sigma(x))]$. To mimic the realistic dynamical process in HICs where the created fireball expands and cools rapidly, in principle, we need to embed the local temperature $T(t,x,y,z)$ and baryon chemical potential $\mu(t,x,y,z)$ into the effective potential. For simplicity but without loss of generality, we assume the heat bath evolves along trajectories with constant baryon chemical potential, and the temperature drops down in a Hubble-like way:

$$\frac{T(t)}{T_0} = \left(\frac{t}{t_0}\right)^{-0.45}, \tag{2}$$

where $T_0(> T_c)$ is the initial temperature and $t_0 = 1$ fm is the initial time for the evolution. With regard to the dynamical evolution of the σ field, we set the damping coefficient η to be constant across the evolution with values ranging from 1.0 to 5.5 fm^{-1}. As for the details of the numerical setup, we simulate the evolution of the σ field in one-dimensional space with range $L = 6.0$ fm, and the spatial grid size $dx = 0.2$ fm. The duration of the evolution is 16 fm at most with the temporal step size $dt = 0.1$ fm/c. With the above setups and initial profiles, the σ field was evolved according to Eq. (1) on an event-by-event basis. As a practical choice, the configurations from later episodes well after phase transition with 4 fm duration of the σ field are censored as the input dataset. Thus, the input is in $t \in [7, 11]$ fm^{-1}, or in the last 40 time steps from the evolution, where the ambient temperature is already much lower than T_c, ensuring that the potential phase transition already happened. Therefore, the prepared input configuration contains $N = 40 \times 30 = 1200$ pixels in each event.

III. RECOGNIZING PHASE TRANSITION IN STOCHASTIC PROCESS

In this section, we first demonstrate that QCD phase order could be recognized from the stochastic process by a deep CNN. Despite that the evolution is simulated in $1+1$-dimensional space for our Langevin systems, its stochastic nature induces elusiveness because of the randomness from interactions. For time series data analysis, long short-term memory (LSTM) neural networks are routinely adopted. However, it is laborious to capture the dynamics in such simplified stochastic processes [49] with LSTM networks. As a succinct alternative, we adopt CNNs to classify the QCD phase order from noisy configurations, since the deep CNN could unearth sufficient correlations in high-dimensional data [50]. The main architecture of the deep CNNs is shown in Fig. 1. The configurations of the σ field with a 40×30 temporal-spatial "resolution" are fed into the deep CNN as images, followed by three convolutional layers with rectified linear unit (ReLU) activation functions as the core structure. For each layer, there are 32 filters with size 3×3. To avoid overfitting, dropout is applied after the second convolutional layer. After all CNN layers, the outputs are flattened and further fed into a 128-neuron fully connected layer, with ReLU activation function and dropout follows. To tackle this binary classification task, the final output layer of the model is another fully connected layer with softmax activation and two neurons to indicate the two phase transition classes.

To perform supervised learning for the binary classification task here, configurations labeled with typical QCD phase transition types (i.e., crossover and first order) serve as the nuts and bolts. To prepare the two categories of configurations, two trajectories in the $T-\mu$ phase diagram are adopted for the dynamical cooling process: The temperature decreases in a Hubble-like way shown in Eq. (2) with a fixed baryon chemical potential as $\mu = 180$ MeV and $\mu = 240$ MeV, separately. Both of the two trajectories would experience the phase transition, which provides a homogeneous description to the dynamical evolution. Since the initial configurations of the σ field are different in the two trajectories, the one with $\mu = 180$ MeV mimics the crossover transition type and is labeled as (0,1) in the training set, while the other one at $\mu = 240$ MeV mimics the first-order phase transition and is labeled as (1,0). The σ field evolution processes are recorded as images with size 40×30, where $N_t = 40$ is time grids and $N_s = 30$ is spatial grids. It is worth noting that the white noise introduced in Eq. (1) leads to intrinsic differences in the event-by-event-generated configurations. The corresponding damping coefficient is set to be $\eta = 1$ fm^{-1} in this section.

For preparing datasets, we simulated 10000 events at each parameter setup we considered in our following tasks. Typical σ field time evolutions are demonstrated in Fig. 2, with two different phase transition scenarios shown. Note that the phase order could not be naively recognized by eye from the spatial-averaged time evolution shown in the figure and much less provided only the final episode evolution after the phase transition happening. As a matter of fact, the effective inputs we adopt for the deep CNNs are demonstrated in Fig. 3, i.e., typical σ field configurations with the final episode masked to be the input. They are two-dimensional images containing the spatiotemporal information for the evolution, which thus embed more correlations from both the time and space of the dynamics. Different magnitudes of noise B are presented in the typical configurations shown in Fig. 3. As the magnitude of the noise is enlarged from $B = 1$ to $B = 3$, the configurations become jouncy, since the fluctuations are increasing for the dynamical evolution.

In the following, for applying the CNN to the phase transition identification task, from the generated events, 20% of the events are randomly chosen as the test set, and the left part of the data is used for training the neural network. Thus, the training dataset consists of 20000 events with configurations at $B = 0.5$ and 20000 at $B = 1$ (half with first-order and half with crossover phase transition), which are fed to the neural network with batch size 16. The training runs 2000 epochs, after which the validation accuracy reaches to 99.9%; see the details in Fig. 4. The validation loss tends to decrease with the training epoch increasing; in other words, it behaves similar to training loss, in which there is no distinct overfitting. The loss is an

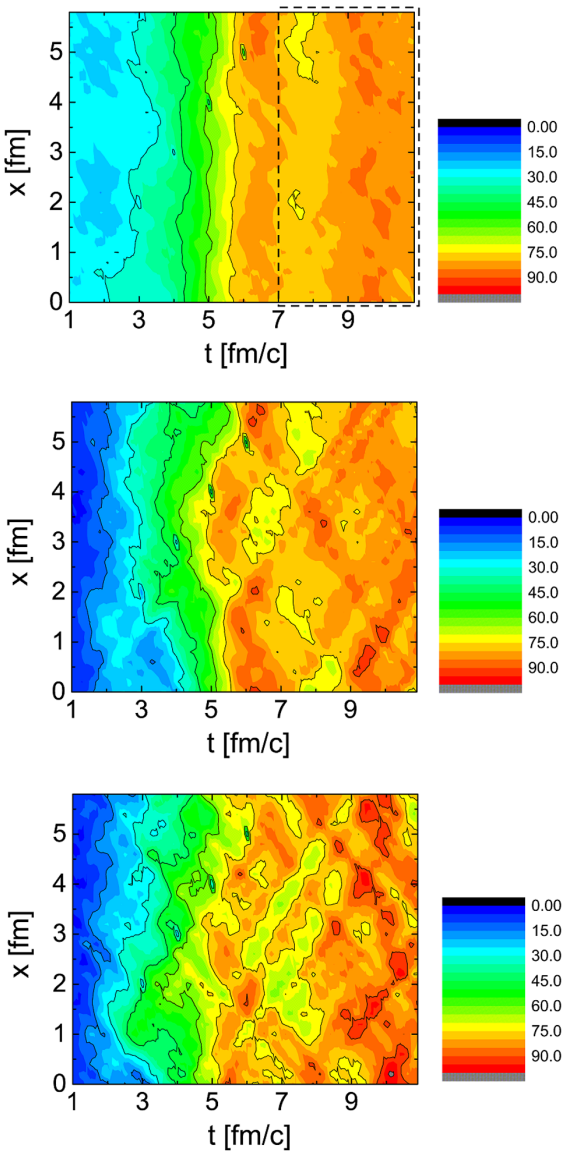

FIG. 3. Dynamical evolution of the σ field in space-time coordinate with noise term $B = 1$, 2, 3, separately (from top to bottom). The color bar listed on the right-hand side represents the strength of the field. The inputs to the CNNs are marked by the dashed box shown in the top figure, which applies to every configuration.

abbreviation of the loss function values in training the deep neural network model. Minimizing the loss leads us to a higher accuracy of prediction on the order of the phase transition in our case. The definition of the loss is shown in the caption of Fig. 1, and the accuracy is the number of correct predictions divided by the total number of predictions. With the trained CNN, we further make the test on a previously unseen dataset. As shown in Fig. 5, the deep CNN has an extraordinary performance on recognizing the order of the phase transition. It is worth noting that the input to the network is solely the final episode of the σ field configuration after phase transition. The test accuracy is

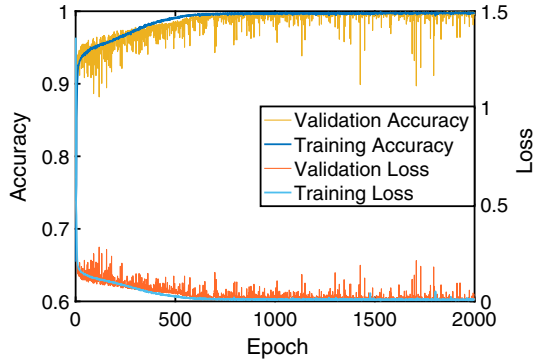

FIG. 4. The accuracy and loss on training and validation datasets with different noise parameters $B = 0.5$, 1, 1.5, 2, and 2.5.

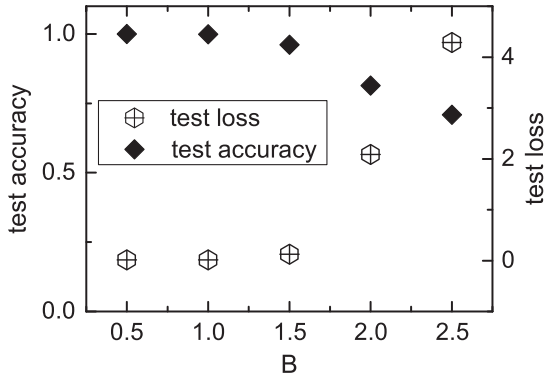

FIG. 5. The accuracy and loss on test datasets with different noise parameters $B = 0.5$, 1, 1.5, 2, and 2.5. The performance is snapped at training epoch 1000.

99% for configurations with $B = 0.5$ and $B = 1$, implying the existence of phase order information projected onto the final episode configurations from the dynamical evolution. We also trained a different neural network with LSTM layers to the task based on the same datasets, but the test accuracy reached only 68.7%. Furthermore, as shown in Fig. 5, the trained CNN could recognize with above 70% accuracy the phase order information from the configurations which have quite different noise magnitude than included in the training dataset. Notably, for $B = 1.5$, although the fluctuations are already so large that they can eventually break the deterministic evolution, the test accuracy achieves 95%. It means that the neural network has learned the underlying evolutionary patterns of the phase transition. Even though the σ configurations are totally different from the learned ones due to the intense noise, the machine could keep an accurate prediction on the phase order in each event. In the case of larger noise $B = 2$ or $B = 2.5$, the machine's performance decreased, which could be ameliorated by introducing more diverse configurations inside the training set to the deep CNN.

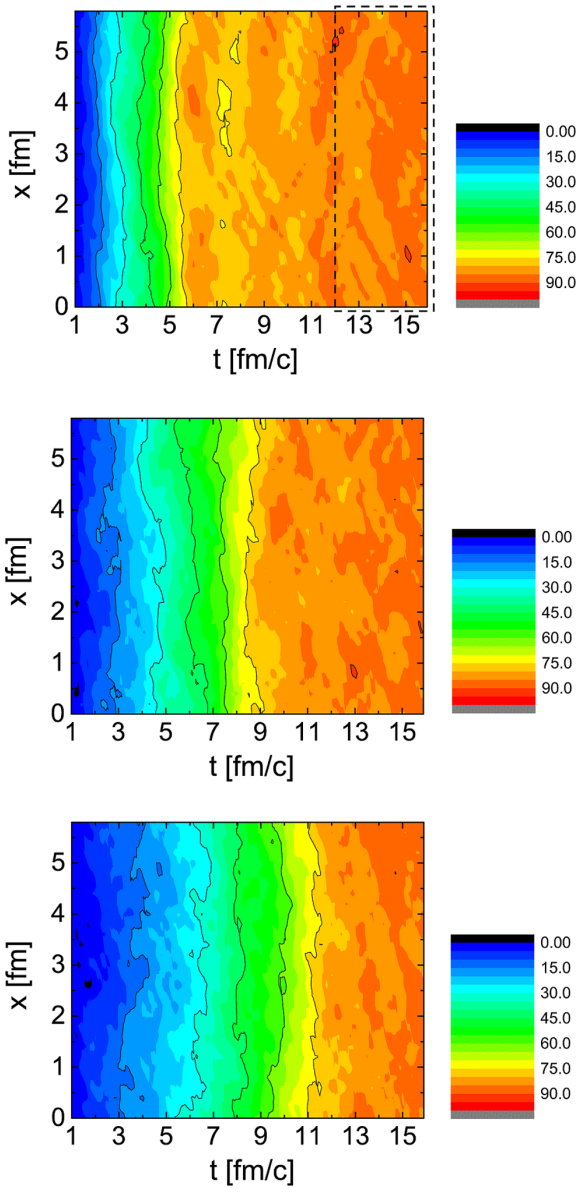

FIG. 6. Dynamical evolution of the σ field in space-time coordinates with the damping coefficients $\eta = 1$, 3, and 5 fm^{-1}, separately (from top to bottom). The color bar listed on the right-hand side represents the strength of the field. The inputs to the CNNs are marked by the dashed box shown in the top figure, which applies to every configurations.

IV. TRACKING DYNAMICAL PARAMETERS FROM EVOLUTION

In Eq. (1), the damping coefficient η drives the diffusion process for the σ field as a main dynamical parameter. In general, the σ field eventually reaches its vacuum expectation value faster in a smaller damping environment. However, the damping coefficient can influence not only the thermal diffusion speed of the system, but also the fluctuations due to the Einstein relation. Because of its ambiguity, it is hard to measure the η in quantum dynamical

systems, no matter whether it is the cold atoms or hot dense QGP. In this section, with a well-trained deep CNN, the good testing results indicate that the damping coefficient could be extracted from the bumpy field configurations collected as the "final episode states" of the stochastic process. To deploy supervised learning on the damping coefficient regression task, we employ a similar training strategy as used above for the phase transition binary classification. Specifically, for preparing the training datasets, we used events generated under the Langevin equation on both the crossover and the first-order side with the damping coefficient η in the range of (1.0–2.5) and (4.6–5.5) fm^{-1}, where η varies with size $d\eta = 0.1$. Within each η bin, 1000 events are generated. Three representative examples of the evolution with different η values are shown in Fig. 6, in which there are no distinctive features to recognize different η. To reduce the bias in datasets, we mix and shuffle the configurations with different η values and randomly divide them into two parts: 20% for validation and the remaining 80% configurations for training. To investigate the generalization ability of the machine, configurations with different damping coefficient values beyond the training set will be utilized to test the network prediction performance.

A network with similar architecture as shown in Fig. 1 is deployed, to which slight changes are performed to target this regression task. Specifically, the dropout layers are removed, and one more fully connected layer with 32 neurons is inserted before the final output layer. The latter is with one neuron representing the values of η. The activation functions are all changed to parameterized ReLU, and the loss function is set to be the mean squared error between the true values and network predictions. With that, the training process is depicted in Fig. 7, where the coefficient of determination R^2 are calculated to indicate the regression quality:

$$R^2 = 1 - \frac{SS_{\mathrm{res}}}{SS_{\mathrm{tot}}}, \tag{3}$$

with $SS_{\mathrm{res}} = \sum_i (\eta_{i,\mathrm{truth}} - \bar{\eta}_{\mathrm{truth}})$ and $SS_{\mathrm{tot}} = \sum_i (\eta_{i,\mathrm{truth}} - \eta_{i,\mathrm{pred}})$ the residual sum of squares and the total sum of squares, respectively. The R^2 is implemented to quantitatively evaluate the correlation between the ground truth and the prediction from the neural networks. As training goes on, R^2 grows quickly from 87% to 93% and from 91% to 97% for the crossover and the first-order scenario, respectively, and tends to be stable with the increasing epoch, which reveals that the deep CNNs have advantages to capture the hidden correlation in image-type inputs [51]. With regard to the inset in Fig. 7, the trained regression CNN is tested on previously unseen σ field configurations with different values of the damping coefficients. The results demonstrate that the predictions retain high consistency with the value of the damping coefficient for

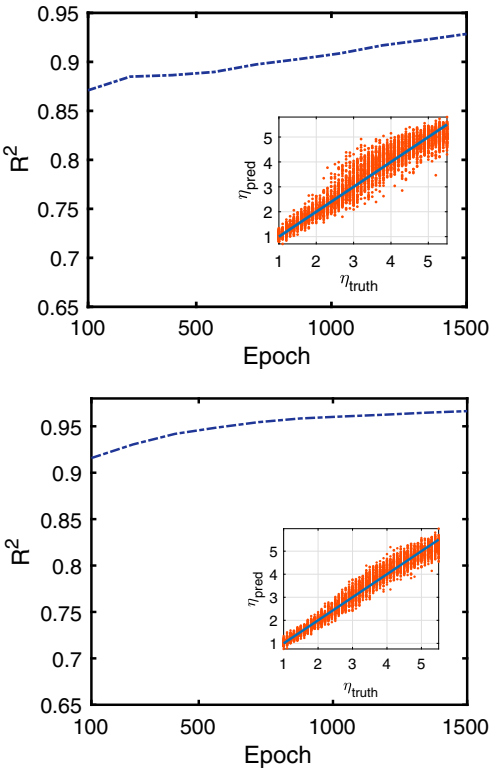

FIG. 7. The training process in predicting the damping coefficient from configurations. The upper panel is prediction results for the crossover scenario; the lower panel is prediction results for the first-order scenario. For both panels, the blue dashed line in the main figure is the training curve, which contains the coefficient of determination R^2 increasing with epochs. The inset part consists of 4600 orange dots, which are labeled by the damping coefficients of the ground truth and predictions from the trained CNNs.

diverse configurations with stochasticity. The predicted η value versus the corresponding ground truth lies around the diagonal line, with a band indicating the deviation. Remarkably, it is found that, although inside the training set there is no supervision in the region of $\eta \sim (2.6\text{–}4.5)$ fm^{-1}, the predictions of the trained network still keep a reasonable performance in this range. In such an interpolation region, the large damping will induce fluctuations definitely non-negligible for the evolution, which makes it hard to decode the dynamical parameters via any conventional analysis. The deep learning approach in our work offers an alternative way to track the dynamics of the stochastic process from intricate configurations driven by the damping coefficient.

V. SUMMARY AND OUTLOOK

In this paper, we introduced a method applying deep learning to identify phase transition information and also track the dynamics for a stochastic dynamical models near the QCD critical point. We numerically simulate

the time evolution of the fluctuating $1+1$-dimensional σ field within the framework of Langevin dynamics in an event-by-event manner and collect its spatial-temporal field configurations to form the datasets for the deep learning study on our tasks.

Based on the generated data, the machine can be trained to identify the nature of the phase transition: first-order or crossover type, encoded inside the stochastic field dynamical evolution. Although the field configurations are totally different from each other due to the noise terms and also the random initial conditions, the machine successfully learns to make an accurate prediction on the phase order for previously unseen evolution events in the testing stage. This is related to the powerful capability of deep CNNs for extracting hidden correlations in an image-type dataset, which facilitates the presented phase order identification from the field configurations. We further designed a regressive CNN to decode the dynamical parameter—damping coefficient—inside the Langevin dynamics from the evolution. It is found that extracting dynamics from such a stochastic process with the trained CNN shows robustness, which also reveals an acceptable generalization ability when tested on configurations containing damping beyond the training set. In summary, we demonstrate that the framework is effective in extracting the Langevin dynamics from complicated configurations associated with intrinsic stochasticity.

It should be noted that the critical point shows unique critical behavior in the thermal equilibrium context. However, when we focus on the dynamic evolution of the system based on the Langevin equation, the phase transition point is just an instantaneous point, so the dynamical process in the critical point scenario is similar to that in the crossover scenario. On the contrary, due to the minimum value coexisting in the phase transition region, the first-order phase transition shows different dynamic evolution of the σ field. Therefore, in this work, we choose the dynamic evolution of the σ field in the conditions of crossover and the first-order phase transition to analyze. The present method can be helpful for a broader field; for example, there is a potential application in a topological-dependent stochastic process [52], in which the topological charge could be extracted by deep CNNs in a similar manner. Moreover, transfer learning could also help us to understand the stochastic process through introducing well-trained deep CNNs into real physical observations [53].

ACKNOWLEDGMENTS

The authors acknowledge inspiring discussions with Horst Stoecker. The work is supported by the Artificial Intelligence (AI) grant at FIAS of SAMSON AG, Frankfurt (L. J., L. W., and K. Z.), by the BMBF under the ErUM-Data project (K. Z.), and by the NVIDIA Corporation with the generous donation of NVIDIA GPU cards for the research (K. Z.).

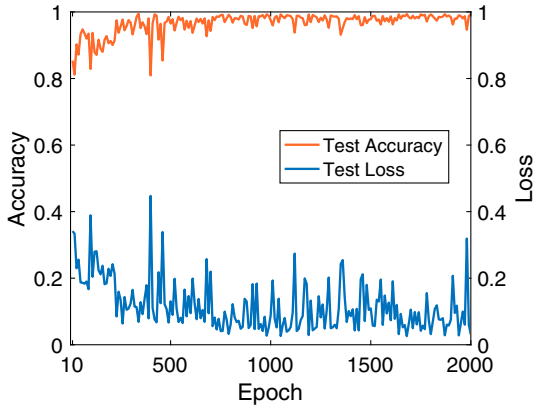

FIG. 8. The accuracy and loss on test datasets with the noise parameter $B = 1.5$.

APPENDIX: RECOGNIZING PHASE TRANSITION IN 2+1-DIMENSIONAL EVENTS

To show the robustness of our deep learning approach, in this Appendix, we present the investigation of recognizing a phase transition in a higher spatial dimension. We prepared 20000 events with σ field configurations at $B = 0.5$ and 20000 at $B = 1$ (half with first-order and half with crossover phase transition) in $2+1$-dimensional space. The configurations of the σ field with a $40 \times 10 \times 10$ "resolution" are input to the neural network for training and testing, where $N_t = 40$ is time grids and $N_x = 10$, $N_y = 10$ is spatial grids.

The main architecture of the deep CNNs is kept similar to that in Sec. III. There are again three convolutional layers with ReLU activation functions as the core structure. For the first two convolutional layers, there are 16 filters with size $3 \times 3 \times 3$. The dropout between the second and the third convolutional layers is omitted, and we add a maxpooling with the pool size $2 \times 2 \times 2$. A third convolutional layer follows the maxpooling with 64 filters and size $3 \times 3 \times 3$. The remaining setup is unchanged from that in Sec. III.

Similar to the training process in Sec. III, 20% of the events are randomly chosen as the test set, and the remaining part of the data is used for training the neural network. With the trained CNN, we further make the test on a previously unseen dataset with $B = 1.5$. As shown in Fig. 8, the deep CNN has an extraordinary performance on recognizing the order of the phase transition in the unknown events.

[1] K. O. Chong, J.-R. Kim, J. Kim, S. Yoon, S. Kang, and K. An, Commun. Phys. **1**, 25 (2018).
[2] H. van Hees, M. Mannarelli, V. Greco, and R. Rapp, Phys. Rev. Lett. **100**, 192301 (2008).
[3] J. Zhao, K. Zhou, S. Chen, and P. Zhuang, Prog. Part. Nucl. Phys. **114**, 103801 (2020).
[4] Y. LeCun, Y. Bengio, and G. Hinton, Nature (London) **521**, 436 (2015).
[5] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Rev. Mod. Phys. **91**, 045002 (2019).
[6] L.-G. Pang, K. Zhou, and X.-N. Wang, arXiv:1906.06429.
[7] H. Huang, B. Xiao, H. Xiong, Z. Wu, Y. Mu, and H. Song, Nucl. Phys. **A982**, 927 (2019).
[8] Z. Liu, W. Zhao, and H. Song, Eur. Phys. J. C **79**, 870 (2019).
[9] M. O. Kuttan, J. Steinheimer, K. Zhou, A. Redelbach, and H. Stoecker, Phys. Lett. B **811**, 135872 (2020).
[10] L.-G. Pang, K. Zhou, N. Su, H. Petersen, H. Stöcker, and X.-N. Wang, Nat. Commun. **9**, 210 (2018).
[11] Y.-L. Du, K. Zhou, J. Steinheimer, L.-G. Pang, A. Motornenko, H.-S. Zong, X.-N. Wang, and H. Stöcker, Eur. Phys. J. C **80**, 516 (2020).
[12] K. Zhou, G. Endrődi, L.-G. Pang, and H. Stöcker, Phys. Rev. D **100**, 011501 (2019).
[13] P. Baldi, P. Sadowski, and D. Whiteson, Nat. Commun. **5**, 4308 (2014).
[14] P. Baldi, P. Sadowski, and D. Whiteson, Phys. Rev. Lett. **114**, 111801 (2015).
[15] J. Barnard, E. N. Dawe, M. J. Dolan, and N. Rajcic, Phys. Rev. D **95**, 014018 (2017).
[16] I. Moult, L. Necib, and J. Thaler, J. High Energy Phys. 12 (2016) 153.
[17] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, Nature (London) **560**, 41 (2018).
[18] L. Wang, Phys. Rev. B **94**, 195105 (2016).
[19] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, Nat. Phys. **13**, 435 (2017).
[20] P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, Sci. Rep. **7**, 8823 (2017).
[21] J. Carrasquilla and R. G. Melko, Nat. Phys. **13**, 431 (2017).
[22] G. Torlai and R. G. Melko, Phys. Rev. B **94**, 165134 (2016).
[23] Z.-Y. Han, J. Wang, H. Fan, L. Wang, and P. Zhang, Phys. Rev. X **8**, 031012 (2018).
[24] L. Wang, Y. Jiang, L. He, and K. Zhou, arXiv:2005.04857.
[25] J. Pathak, Z. Lu, B. R. Hunt, M. Girvan, and E. Ott, Chaos **27**, 121102 (2017).
[26] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, Phys. Rev. Lett. **120**, 024102 (2018).
[27] H. Fan, J. Jiang, C. Zhang, X. Wang, and Y.-C. Lai, Phys. Rev. Research **2**, 012080 (2020).
[28] K. Paech, H. Stöcker, and A. Dumitru, Phys. Rev. C **68**, 044907 (2003).
[29] C. Herold, M. Nahrgang, I. Mishustin, and M. Bleicher, Phys. Rev. C **87**, 014907 (2013).
[30] L. Jiang, S. Wu, and H. Song, Nucl. Phys. **A967**, 441 (2017).
[31] L. Jiang, S. Wu, and H. Song, EPJ Web Conf. **171**, 16003 (2018).
[32] B. J. Schaefer and J. Wambach, Phys. Part. Nuclei **39**, 1025 (2008).
[33] S.-x. Qin, L. Chang, H. Chen, Y.-x. Liu, and C. D. Roberts, Phys. Rev. Lett. **106**, 172301 (2011).
[34] G. A. Pavliotis, *Stochastic Processes and Applications* (Springer, New York, 2014).
[35] K. Fukushima and T. Hatsuda, Rep. Prog. Phys. **74**, 014001 (2011).
[36] K. Fukushima and C. Sasaki, Prog. Part. Nucl. Phys. **72**, 99 (2013).
[37] T. Hatsuda and T. Kunihiro, Phys. Rep. **247**, 221 (1994).
[38] L.-j. Jiang, X.-Y. Xin, K.-L. Wang, S.-X. Qin, and Y.-x. Liu, Phys. Rev. D **88**, 016008 (2013).
[39] L. Wang, G. Cao, X.-G. Huang, and P. Zhuang, Phys. Lett. B **780**, 273 (2018).
[40] D.-U. Jungnickel and C. Wetterich, Phys. Rev. D **53**, 5142 (1996).
[41] B. J. Schaefer and J. Wambach, Nucl. Phys. **A757**, 479 (2005).
[42] N. Petropoulos, J. Phys. G **25**, 2225 (1999).
[43] H.-S. Roh and T. Matsui, Eur. Phys. J. A **1**, 205 (1998).
[44] Y. Aoki, G. Endrődi, Z. Fodor, S. D. Katz, and K. K. Szabó, Nature (London) **443**, 675 (2006).
[45] A. Bazavov *et al.* (HotQCD Collaboration), Phys. Rev. D **85**, 054503 (2012).
[46] A. Bzdak, S. Esumi, V. Koch, J. Liao, M. Stephanov, and N. Xu, Phys. Rep. **853**, 1 (2020).
[47] M. Nahrgang, S. Leupold, C. Herold, and M. Bleicher, Phys. Rev. C **84**, 024912 (2011).
[48] M. Nahrgang, C. Herold, S. Leupold, I. Mishustin, and M. Bleicher, J. Phys. G **40**, 055108 (2013).
[49] K. Yeo and I. Melnyk, J. Comput. Phys. **376**, 1212 (2019).
[50] S. Das Sarma, D.-L. Deng, and L.-M. Duan, Phys. Today **72**, No. 3, 48 (2019).
[51] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, Phys. Rev. Lett. **122**, 065301 (2019).
[52] L. Zhao, Z. Wang, X. Zhang, X. Liang, J. Xia, K. Wu, H.-A. Zhou, Y. Dong, G. Yu, K. L. Wang, X. Liu, Y. Zhou, and W. Jiang, Phys. Rev. Lett. **125**, 027206 (2020).
[53] H. Y. Kwon, H. G. Yoon, C. Lee, G. Chen, K. Liu, A. K. Schmid, Y. Z. Wu, J. W. Choi, and C. Won, Sci. Adv. **6**, eabb0872 (2020).