




# Machine learning method for the supplement, correction, and prediction of the nonlinear dynamics in pattern formation

Cite as: Phys. Fluids **33**, 024110 (2021); <https://doi.org/10.1063/5.0036762>

Submitted: 09 November 2020 • Accepted: 12 January 2021 • Published Online: 26 February 2021

 Yi Chen (陈一),  Di Wu (吴笛),  Li Duan (段俐), et al.



View Online



Export Citation



CrossMark

## ARTICLES YOU MAY BE INTERESTED IN

[A point-cloud deep learning framework for prediction of fluid flow fields on irregular geometries](#)

Physics of Fluids **33**, 027104 (2021); <https://doi.org/10.1063/5.0033376>

[Deep neural network-based strategy for optimal sensor placement in data assimilation of turbulent flow](#)

Physics of Fluids **33**, 025119 (2021); <https://doi.org/10.1063/5.0035230>

[Physics guided machine learning using simplified theories](#)

Physics of Fluids **33**, 011701 (2021); <https://doi.org/10.1063/5.0038929>

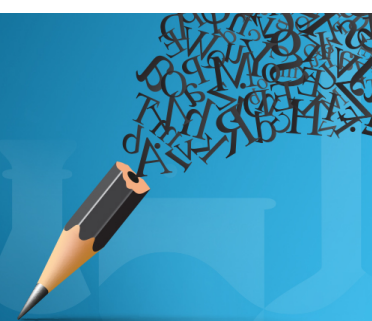


Author Services

**English Language Editing**

High-quality assistance from subject specialists

LEARN MORE



# Machine learning method for the supplement, correction, and prediction of the nonlinear dynamics in pattern formation

Cite as: Phys. Fluids **33**, 024110 (2021); doi: 10.1063/5.0036762

Submitted: 9 November 2020 · Accepted: 12 January 2021 ·

Published Online: 26 February 2021



View Online



Export Citation



CrossMark

Yi Chen (陈一),<sup>1,2</sup>  Di Wu (吴笛),<sup>1,2</sup>  Li Duan (段俐),<sup>1,2,a)</sup>  and Qi Kang (康琦)<sup>1,2,a)</sup> 

## AFFILIATIONS

<sup>1</sup> National Microgravity Laboratory, Institute of Mechanics, Chinese Academy of Sciences, Beijing 100190, China

<sup>2</sup> School of Engineering Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

<sup>a)</sup> Authors to whom correspondence should be addressed: [duanli@imech.ac.cn](mailto:duanli@imech.ac.cn) and [kq@imech.ac.cn](mailto:kq@imech.ac.cn)

## ABSTRACT

The pattern formation and spatial-temporal chaos are interesting issues in nonlinear dynamics. A novel model based on machine learning methods is designed to learn and imitate the pattern evolution in Bénard–Marangoni convection (BM convection). There is a supercritical process, which is an inevitable and unique experimental phenomenon, on the way to chaos in BM convection. A single layer of fluid uniformly heated at the bottom is used as the experimental system. During the experiment, the temperature difference between top and bottom of the liquid layer is increased first to make the system enter the supercritical convection state and then decreased after a while; surface temperature distribution of the liquid layer is measured in real time with an infrared thermal imager, which visualized the formation and re-organization of cellular convection during the supercritical state. The temperature data are used as the material that meets the conditions of machine learning and then the machine learning method in charge of predicting the picture of temperature distribution that it never has seen before in two steps. The experimental data are used to train an auto-encoder model based on convolutional neural networks and an RNN–CNN joint model, in which the former is used for extracting low-dimensional features of the temperature field, and the latter is used for predicting evolution results of the low-dimensional features and recovering them back to the temperature field. The models have finally achieved the objectives of supplementing the missing experimental data and correcting actual experimental data by comparing the actual experimental results with the prediction results of the machine learning approach and theoretical analysis results. On the other hand, active exploration has been undertaken in predicting physical experimental results that have never happened before.

Published under license by AIP Publishing. <https://doi.org/10.1063/5.0036762>

## I. INTRODUCTION

Studying the evolution of patterns is an interesting and significant subject. The self-organization phenomenon appeared in the Bénard–Marangoni (BM) convection system is a typical example. However, traditional research methods are not suitable to this kind of system because theoretical solutions are difficult to obtain, the experiment process is easily influenced, and numerical methods are time-consuming. Compared with actual physical experiments, our machine experiment can obtain results closer to theoretical results because its calculation characteristics are hardly influenced by the environment. In addition, the method of machine learning is to observe problems from a statistical perspective, which is helpful in analyzing structured information in the data and obtaining the distribution law of the data.

Convection is a very common phenomenon in nature. Different forces can drive different types of convection. BM convection is driven by the surface tension of liquid. In 1958, Pearson *et al.*<sup>1</sup> theoretically proved the instability mechanism of BM convection through linear stability analysis and determined the critical Marangoni number and the critical wavenumber of the convection.

Self-organization is an essential feature of Bénard convection. Bénard–Marangoni convection appears in the form of Bénard cells, which array themselves in a hexagonal honeycomb structure automatically. Golovin *et al.*<sup>2</sup> explained the generation of this structure using the weakly nonlinear analysis, and they indicated that it was because of the superimposition of disturbance waves that simultaneously existed in the flow field in three directions with an angle of 120° between each other. Koschmieder and Biggerstaff<sup>3</sup> first experimentally

discovered the subcritical bifurcation phenomenon in the critical convection stage of BM convection. By observing the pattern of the system, they found that the subcritical bifurcation behaved as follows: when the critical Marangoni number was reached, the pattern of convection was replaced by the hexagonal Bénard cells; as the Marangoni number started to drop after passing the critical point, the hexagonal Bénard cells did not disappear immediately when the Marangoni number was at the critical value but disappeared when it was less than the critical value. Scanlon and Segel<sup>4</sup> conducted the weakly nonlinear analysis on BM convection and determined the stable region around the critical value, coming up with a better understanding of this process from the point of view of the mechanism. Schatz *et al.*<sup>5</sup> obtained experimental results in agreement with theoretical analysis, providing a reliable basis for the existence of subcritical bifurcations.

As the temperature gradient increases, BM convection will gradually lose stability and enter into the supercritical convection state on the way from the steady state to chaos. Chaos is a phenomenon that appears in a nonlinear system. It is a seemingly irregular movement and sensitive to the initial values of parameters. That is to say, when the initial value is slightly disturbed, the disturbance will increase exponentially with time, resulting in unpredictable behavior of the system. Chaos is a common form of movement in nature. In 1963, Lorenz<sup>6</sup> extracted the Lorenz equations as a simplified atmospheric convection model (a Bénard convection model) in his study on the long-term weather forecast, and he also studied the chaotic mechanism involved. Lorenz found that under certain parameter conditions, the orbit starting from a finite set would be drawn to a specified final state. This ultimate state is referred to as the Lorenz attractor, which can be categorized as a strange attractor due to its sensitive dependence on the initial conditions. Tucker<sup>7</sup> eliminated the conditions of specific parameters in the subsequent research and concluded that the Lorenz equations had strange attractors under any conditions. With the increase in the Marangoni number, BM convection gradually enters into chaos through the critical state, the moderate supercritical state, and the highly supercritical state, and the arrangement of Bénard cells will turn into the two-dimensional spatiotemporal chaotic state after a long-term nonlinear evolution.

Nowadays, there are various methods of exploring the flow stability. In theoretical analysis, the linear stability theory represented by the normal mode method and the weakly nonlinear theory represented by the Landau equation are often used; however, due to the nonlinear term in the equation, it is difficult to obtain the exact solution of the differential equation. In the experimental study, due to the influence of gravity on the ground, experiments under many working conditions cannot be carried out smoothly. The method of numerical simulation exists the contradiction between calculation accuracy and calculation efficiency. Because the process of minimizing the calculation error always means too much calculating load. In addition, in the process of solving differential equations of the flow field by the numerical method, the calculation is based on the state change of each point on the grid, which will result in the lack of correlation between adjacent points. In recent years, the method of machine learning has been widely applied in various research fields,<sup>8</sup> and the development in computer vision has attracted particular attention. The method of machine learning is not based on the state change of each grid point in the flow field but treating the temperature distribution in the entire field as a picture and using the convolution kernel to extract characteristics of each picture, so the numerical values of adjacent points are

correlated, and structural information in the flow field is better expressed. Since Bénard cells formed by BM convection are typical temperature structure information, it has more advantage to employ the machine learning methods to study the evolution of patterns formed by BM convection.

Machine learning can be simply divided into supervised learning and unsupervised learning. The key idea of supervised learning is to revise the fitting (calculating) results of the model every time with the help of a teacher. During the transition process of the fluid system from the steady state to chaotic state, the law of motion of each point in the system is changing from describable to indescribable (spectral analysis), and intuitively, the system shows a developing trend into random motion eventually. However, in fact, the motion of each point has its own changing law, for example, strange attractors will appear when the system is fully developed. The nature of the machine learning methods is to use statistical methods to fit the data to distributions. Because the model uses many activation functions and there are a lot of weights inside, compared with traditional methods, the machine learning methods can determine the distribution characteristics of each data and the entire dataset better. This is also intended to solve the problems of fluid mechanics from another perspective.

This article attempts to associate the machine learning methods with the experimental study on the supercritical convection stage of BM convection. The work of the machine is to learn a certain number of practical experimental pictures of the patterns in the supercritical process; by summarizing the law of pattern evolution, it is to gain the ability to supplement the data missing in the experiment and predict the subsequent experimental results to a certain extent. In addition, by comparing with theoretical analysis results and prediction results of machine learning methods, the actual physical experiment results will be corrected.

The structure of this article is as follows: Section II introduces the fluid experiment system and the machine learning model; Sec. III introduces the dimensionality reduction model and evaluates the results it gained; Sec. IV introduces the model for supplement, correction, and prediction and evaluates the results it gained; and Sec. V makes a summary of the work discussed in this paper and identifies the prospects for future work.

## II. FLUID EXPERIMENT SYSTEM AND MACHINE LEARNING MODEL

### A. Fluid experiment system

Wu *et al.*<sup>9,10</sup> studied the characteristics of the pattern at various stages of the BM convection system with the change in the Marangoni number, obtained the experimental results of the pattern evolution in the supercritical process, and provided training data for the machine learning research discussed in this paper. The forming mechanism of the pattern can be simply expressed as follows: in a liquid layer with cold top and hot bottom, the disturbance causes the hot fluid in the lower part of the layer to migrate to the surface of the liquid layer, forming hotspots on the surface; since the surface tension of the fluid at the hotspot is less than that of the surrounding fluid, the fluid at the hotspot is pulled away by the surrounding fluid, and at the same time, the hotter fluid in the interior is driven upward to replenish; as a result, the disturbance increases and convection forms in the fluid. The model used in the experiment is a square liquid pool heated at the

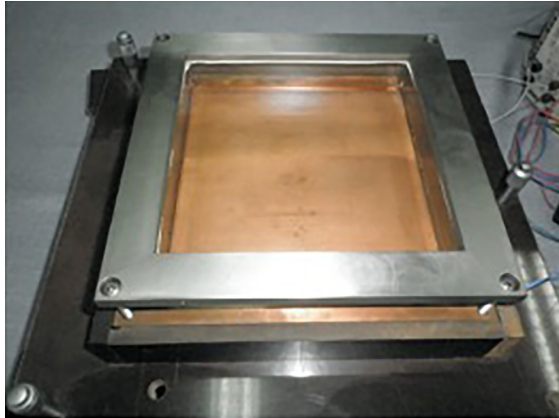


FIG. 1. Experimental system.

bottom. As shown in Fig. 1, the size of the horizontal cross section of the liquid pool is  $160 \times 160 \text{ mm}^2$ . The bottom is composed of copper plates, thin engineering plastic plates, and aluminum plates to ensure the uniformity of temperature at the bottom during heating. 50 cSt silicone oil is selected as the experimental medium because it has stable chemical nature and the interface is not easily polluted. In order to obtain the spatial distribution of temperature of the flow field, an infrared thermal imager (FLIR E60, resolution:  $640 \times 480$ ) is used to observe the temperature field on the upper surface of the liquid layer. (The time interval between the recordings is 10 s. For example, if the first picture is taken at 0 s, the ninth picture will be taken at 90 s.) The Marangoni number ( $Ma$ ) is a dimensionless number, and it is used to describe the ratio of heat transport caused by thermocapillary convection to heat transport caused by heat conduction. When it exceeds the critical number, convection actually starts,

$$Ma = \frac{(-\partial\sigma/\partial T)\Delta T d}{\mu\alpha}, \quad (1)$$

where  $\sigma$  represents the surface tension coefficient,  $\Delta T$  represents the temperature difference,  $d$  represents the thickness of the liquid layer,  $\mu$  represents the dynamic viscosity coefficient, and  $\alpha$  represents the thermal conductivity coefficient. In the rest of this paper, we will abbreviate the Marangoni number as  $Ma$  and express the critical Marangoni number corresponding to the onset of oscillation of the single-layer fluid as  $Ma_c$ .

This experiment is divided into two stages: heating and cooling. The heating rate is  $0.3^\circ\text{C}/\text{min}$ . The temperature at the hot end is gradually increased from the room temperature to  $70^\circ\text{C}$  and then decreased to the room temperature. Because the temperature difference increases first and then decreases, the liquid layer undergoes a process including static state, critical convection, supercritical convection, and then critical convection again. BM convection is a type of stable quasi-steady flow, but in the supercritical state, the appearance and arrangement of Bénard cells show a new unsteady evolution process. The infrared thermal imager records the change in the pattern during this process at a rate of 0.1 fps. The experimental results are shown in Fig. 2. The formation of Bénard cells in BM convection is essentially caused by the surface tension gradient generated from the uneven temperature distribution on the liquid surface. The polygonal hot region

in Fig. 2 represents a convective Bénard cell. The generation of new Bénard cells must be accompanied by the generation of hotspots, and the generation mechanism of a Bénard cell is influenced by the wave-number, arrangement, and migration of surrounding Bénard cells. Therefore, the motion of a Bénard cell shows itself in many forms at the macrolevel, including splitting, merging, and independent generation. The region in the orange square shown in Fig. 2 presents one of these generation processes.

As can be seen above, the generation and evolution of Bénard cells are complex nonlinear processes, belonging to problems of pattern dynamics. The formation and spatiotemporal chaos of the pattern are interesting issues for most scientists. The location where Bénard cells are generated is closely related to the local temperature change in the liquid pool. In addition, Bénard cells array themselves precisely, and a slight change in a Bénard cell will have a significant impact on the overall pattern. Therefore, except for practical fluid experiments, traditional research methods can hardly describe the overall pattern accurately and comprehensively. The machine learning methods combine statistics and algorithms. Statistics can be used to analyze the law of data distribution and deduce the changing law of data distribution after combined with specific algorithms. It is very helpful in the study of the evolution law of the pattern.

In the process of attempting to simulate actual experiments with the machine learning methods, the materials for machine learning are pictures captured by using the infrared camera in the heating and cooling processes after the BM convection enters the supercritical convection state. These pictures are used to train models for supplementing, correcting, and predicting the actual physical processes in the single-layer fluid system. In this process, the machine has learned 740 sets of continuous experimental data (7400 s in total) contained in the infrared images so that it has gained the ability to deduce the subsequent development of the experiment independently and the ability to judge the state of the system at a certain moment during the experiment. The division method of the datasets is shown in Table I. The experimental pictures are counted from 0.

## B. Machine learning model

In recent years, machine learning has been paid much attention as a field of computer science. Using machine learning methods to predict the evolution process is gradually applied to various disciplines. Shi *et al.*<sup>11</sup> used a convolutional long short-term memory (LSTM) network in the now-casting of precipitation; Huang *et al.*<sup>12</sup> successfully predicted the evolution of 3D flame using the neural network and VT technology; and Han *et al.*<sup>13</sup> used a complex neural network structure to predict the unstable flow field in the case of the flow around a circular cylinder.

In this experiment, the machine attempts to predict the pattern at a certain time point based on the pattern pictures at three previous moments in a specific order. In order to train the machine with the law of formation of Bénard cells in more detail and make the machine show it more clearly in the testing, in all the following cases, the machine should be allowed to observe a pattern with a relatively large change in the training process. Therefore, frame skipping is employed in the selection of pictures for the training set and testing set, and such a prediction method can be called a frame skipping prediction. In order to ensure the prediction effect, the number of skipped frames

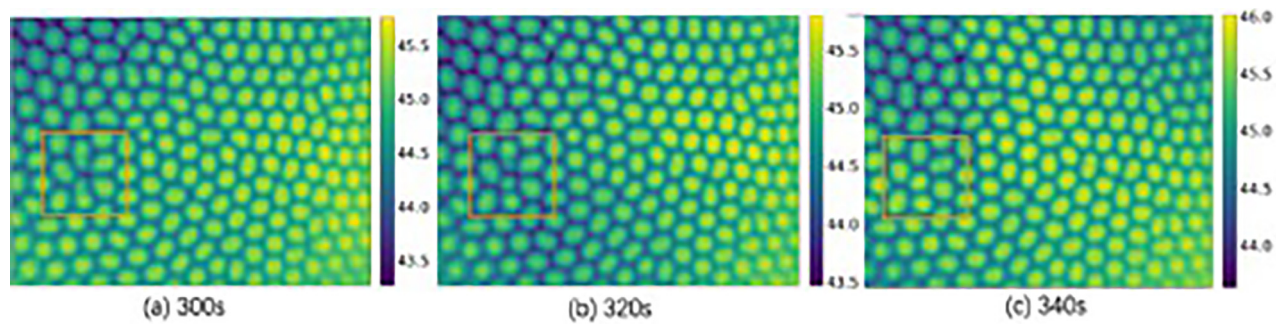


FIG. 2. The example of experimental results: (a) corresponding to the pattern at 300 s, (b) corresponding to the pattern at 320 s, and (c) corresponding to the pattern at 340 s.

should be the same when selecting the training set and the testing set. Based on such experimental objectives, the RNN structure that is good at time-series analysis and the CNN structure that is good at image processing are associated in our work to obtain a joint model. Since each pattern image is composed of 76 800 data points, the operation speed is too slow, and the amount of information is too large to memorize. Therefore, the entire experimental process is divided into two steps. In the first step, a CAE model (auto-encoder model based on the convolutional neural network) is utilized to reduce the dimensionality of the 76 800 data points so that the picture information can be compressed into feature vectors with a dimensionality as low as possible; in the second step, the RNN part of the RNN–CNN joint model is used to predict the change in the feature vector with a lower dimensionality, and the improved CNN part of the RNN–CNN joint model is used to restore the predicted feature vectors to the predicted picture.

Different from the usual numerical methods for simulation and prediction, machine learning methods are inclined to use statistical methods to summarize experimental information, rather than using

numerical methods to iterate differential equations. This is a preliminary attempt to use machine learning methods to fit results of real physical experiments.

### III. DIMENSIONALITY REDUCTION MODEL AND RESULTS ANALYSIS

#### A. Dimensionality reduction model

The CAE model is employed to reduce the dimensionality of the original image data in this machine learning experiment. The CAE model is an auto-encoder (AE) model based on the convolutional neural network. The AE<sup>14</sup> consists of an encoder and a decoder, as shown in Fig. 3. First, it extracts critical information from the sample through the encoder, a set of neural networks, which is referred to as encoding. Then, it reconstructs the data through the decoder, the mirror of the encoder, or a set of inversely distributed neural networks, which is referred to as reconstruction. In general, the reconstruction error is calculated from the input data and output data of the auto-encoder. The weight of each node in the network structure is adjusted at each

TABLE I. Division of datasets.

Case	Training data (numbering of pictures)	Testing data (numbering of pictures)	Dataset description	Prediction method
Supplement	710	30	From the continuously changing 740 pictures, 30 pictures are randomly selected as the testing set.	Based on actual experiment pictures
Correction	710	30	From the continuously changing 740 pictures, 15 consecutive pictures during the heating/cooling process (30 pictures in total) were selected as the testing set.	Based on actual experiment pictures
Short-term prediction	740	15	740 continuously changing pictures are used as the training set. After the training, based on the first 27 pictures of the 740 pictures, the results of subsequent experiments are predicted autonomously.	Based on pictures calculated by the machine learning methods
Long-term prediction	740	15	Continue to autonomously predict 15 pictures after short-term prediction.	Based on pictures calculated by the machine learning methods

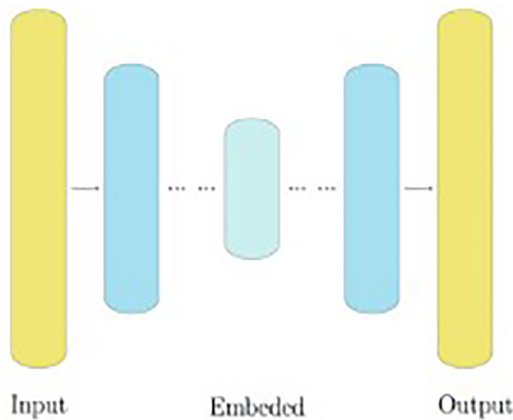


FIG. 3. The structure of an auto-encoder.

round of learning of the machine learning approach, and the reconstruction error will be minimized after several rounds of training. Thus, the data in the middle layer are considered to represent all the information of the original data, and dimensionality reduction is achieved.

In the first step of dimensionality reduction, the picture is encoded and compressed as much as possible by using the encoder. The encoder compresses the image data from a  $240 \times 320$  matrix to a feature vector consisting of 2048 feature values; the decoder recovers the feature vector back to a  $240 \times 320$  matrix through upsampling and convolution-and-pooling operations. If the picture can be recovered through the above process successfully (the reconstruction error is minimized), the auto-encoder is considered to have gained the ability to encode each picture through the encoder into a feature vector containing 2048 feature values that carries all important information of the picture. In this machine learning experiment, the mean square error between original and reconstructed images is used as the reconstruction error [Eq. (2)] to optimize the network structure. In Eq. (2), LF represents the loss function,  $y_m$  is the value of each pixel in the original image,  $\bar{y}_m$  is the value of the corresponding pixel in the reconstructed image, and  $M$  is the total number of pixels in the picture. That the function is multiplied by  $\frac{1}{2}$  makes the subsequent backpropagation algorithm easy to be accomplished. Due to the large amount of information contained in the experimental data, in order to predict successfully later on, the extracted feature vector should include the information of the temperature state, the position of Bénard cells, and other important information that is difficult to explicitly describe. Therefore, in this experiment, a complex neural network structure transformed from AlexNet<sup>15</sup> is selected as the training model. In the construction of the neural networks, the process of feature extraction mainly depends on convolutional kernels. If the number of convolution kernels is too small, the model can only recover the large-scale structure information (such as the temperature field distribution) of the experimental pictures due to the lack of information extracted from the model, but the detailed information (specific cell structure) cannot be mastered. As we can see in NIPS-2012, AlexNet is designed as a neural network structure for classification and shows an

excellent classification effect on Cifar-10, which shows that the network structure of the model has an excellent ability to extract features. AlexNet analyzed three-channel color images on the Cifar-10 dataset, while our images only have a single-channel, and the depth of the convolution kernel should be adjusted in order to design the neural network structure to meet the requirements. In the feature extraction (coding) stage of the auto-encoder, we use the convolution kernel of the same scale as that of AlexNet in feature extraction (encoding) and adopt the network structure with the same structure but reverse distribution as the coding stage in the feature recovery (decoding) stage of the auto-encoder. The model structure is illustrated in Fig. 4, and the specific structures of each layer are shown in Table II,

$$LF = \frac{1}{2M} \sum_{m=1}^M (y_m - \bar{y}_m)^2. \tag{2}$$

**B. Analysis of dimensionality reduction results**

Machine learning methods need to use the back-propagation algorithm in the process of parameter optimization, in which the learning rate can control the optimization degree of parameters. In the initial stage of optimization, the parameter value is generally far away from the optimal value, so we need to greatly optimize the parameters. At this time, we need a larger learning rate. In the later stage of training, the parameter value has been adjusted around the optimal value; we only need to fine-tune, so we need to use a smaller learning rate. Therefore, we developed a learning rate optimization method suitable for this experiment.

In order to enable the machine to grasp the important information of the experimental pictures better and help the machine to make a more reasonable calculation, we need to preprocess the pictures before putting them into the model for calculation. There are two kinds of normalization operations: (1) z-score normalization: the pictures provided to the machine experiment are infrared thermal imager pictures, so the average value of the pictures is closely related to the average temperature of the flow field. In the process of convolution operation, if the mean value difference between the same batch of images is too large, the machine cannot summarize the unified information about the specific structure in the feature extraction process, resulting in the misunderstanding of the original experimental results. (2) The layer normalization: in a deep neural network, the input of the middle layer is the output of the previous layer. Therefore, the previous calculation in the neural layer will lead to a large difference in the distribution of input in the current layer. When the gradient descent is used to update the parameters, each parameter update will cause the input distribution of each layer in the middle of the network to change.

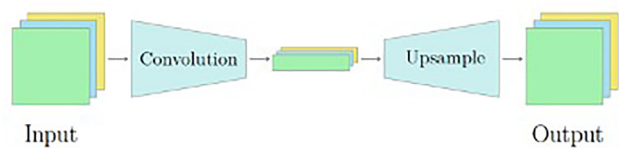


FIG. 4. The structure of the auto-encoder in the experiment.

TABLE II. Auto-encoder structure.

Numbering of neural network layers		Layer type	Convolution kernel size/full connection size	Layer normalization or not
Coding	Layer 1	Conv.-pooling	3, 3, 1, 96	Y
	Layer 2	Conv.-pooling	3, 3, 96, 256	Y
	Layer 3	Conv.-pooling	3, 3, 256, 384	Y
	Layer 4	Conv.-pooling	3, 3, 384, 256	N
	Layer 5	FC	4096	N
	Layer 6	FC	2048	N
Decoding	Layer 7	FC	2048	N
	Layer 8	FC	4096	N
	Layer 9	Conv.-pooling	3, 3, 384, 256	N
	Layer 10	Conv.-pooling	3, 3, 256, 384	N
	Layer 11	Conv.-pooling	3, 3, 96, 256	N
	Layer 12	Conv.-pooling	3, 3, 1, 96	N

The deeper the layer is, the more obvious the input distribution will be changed. In the matrix operation, the smaller data will be swallowed. We choose the layer normalization method to solve this problem.

1. Dimensionality reduction results

In this experiment, we use the reconstruction error of the model to describe the dimensionality reduction effect. The backpropagation algorithm is used to update the parameters of the model. The optimization curve of the loss function is shown in Fig. 5, where the abscissa represents the number of times the model is optimized and the ordinate represents the average of reconstruction errors of all pictures after each optimization, which is the value of the model's loss function. It can be seen that with the increase in training rounds, the value of the loss function decreases from the maximum value 44.46 to the minimum value around 1.04, with an obvious optimization effect.

Figure 6 shows the comparison between the reconstructed picture (a) at 500 s and its corresponding original picture (b) at 500 s. It

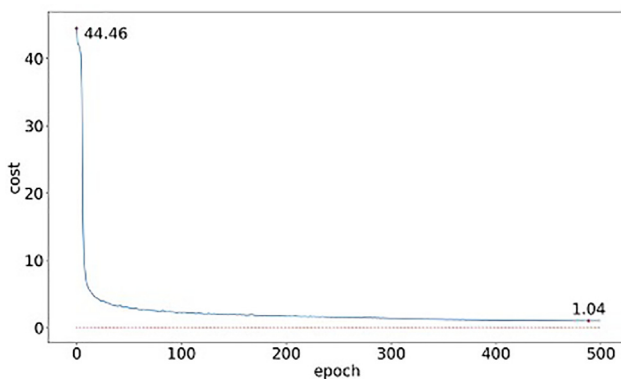


FIG. 5. The optimization curve of the loss function of the CAE.

can be seen intuitively that the two pictures are very similar in terms of the shape and position of the Bénard cells, as well as the overall temperature distribution.

2. Mean square error analysis

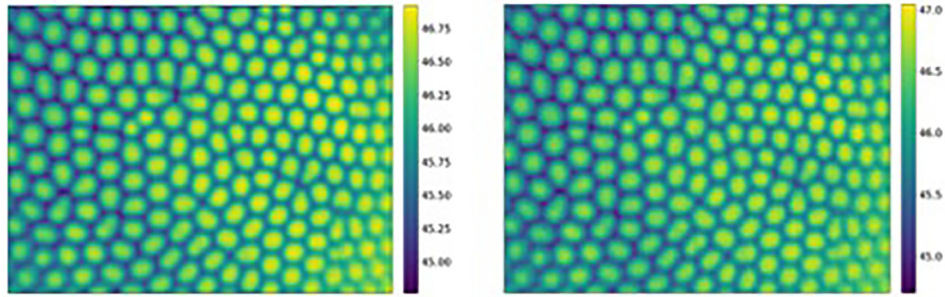
After multiple rounds of training on the model, each connection weight is optimized through the backpropagation algorithm, and effective training results have been achieved. According to Eq. (2), the mean square error between the reconstructed and original pictures at 500 s is only 0.002. The average mean square error of all pictures is 0.0023, and it can reach the order of  $10^{-3}$ . This indicates that each reconstructed picture and original picture have a very small difference in the total pixel value.

3. R<sup>2</sup> analysis

If the mean square error of a model is close to the mean square error of the average model,  $h_{avg}$ , the fitting effect of the model is not ideal, here  $R^2 \approx 0$ . The farther the mean square error of the model is from the mean square error of  $h_{avg}$ , the closer the value of  $R^2$  is to 1 and the better the model's fitting effect.  $R^2$  can be expressed as Eq. (3), where  $h(x^{(i)})$  is the mapping set corresponding to the original data point  $x^{(i)}$  after the calculation by model  $h$ ,  $y^{(i)}$  represents the actual result set corresponding to the original data point  $x^{(i)}$ , and  $\bar{y}$  represents the average of the actual results corresponding to all original data points in one picture,

$$R^2 = 1 - \frac{\sum_{i=1}^m (h(x^{(i)}) - y^{(i)})^2}{\sum_{i=1}^m (\bar{y} - y^{(i)})^2}. \tag{3}$$

After the first step of training, the final experimental results of the model can reach  $R^2 = 0.9695$ . This means that the CAE model has mapped the differences between different pixels in the original picture matrix well in the resulting picture matrix. The reconstructed picture and the original picture have similar distribution of pixel values, rather



**FIG. 6.** Comparison between the reconstructed picture and its corresponding original picture: (a) corresponding to the reconstructed result at 500 s and (b) corresponding to the experimental result at the same time (500 s).

than the elements in the original picture that are simply mapped to the average pixel value of the original picture to reduce the reconstruction error. The model has high credibility. Thus, the CAE model has been ready to provide feature vectors for subsequent models; then, we can use these low-dimensional data to supplement, predict, and correct the practical experimental data with RNN–CNN in the follow-up work.

#### IV. MODEL FOR SUPPLEMENT, CORRECTION, AND PREDICTION AND RESULTS AND DISCUSSION

##### A. Structure of the model for supplement, correction, and prediction

The RNN–CNN joint model is used to supplement, correct, and predict the actual experimental data in this paper. The RNN<sup>16</sup> part of the joint model is referred to as a recurrent neural network, and it is widely used in forecasting scenarios involving time-series information. In the machine learning experiment, the changing trend of a physical quantity in the past should be considered for the prediction of its value in the future. The same logic is used in the prediction of pattern evolution, that is to say, the overall appearance of the pattern at some previous moments is needed for the machine to deduce the pattern at present. In our machine learning experiment, the machine needs to grasp the preferred positions of new generated Bénard cells from the training set to reasonably deduce the evolution of Bénard cells in the testing set. In addition, although there is a big difference between two pictures when one of them has a new generated Bénard cell and the other has none, it is a gradual process when a Bénard cell is generated. If the machine wants to be able to describe this process coherently, it must learn to predict the small difference between adjacent pictures. Therefore, the RNN is needed in the entire network structure to predict the changing trend of the feature vectors.

CNN is referred to as a convolutional neural network. It is widely applied in the scenario where image processing is involved.<sup>17</sup> Because the CNN contains a lot of activation functions, it is frequently used in solving complex nonlinear problems. The two types of activation functions used in this paper are shown in Fig. 7, and the two kinds of activation operations are expressed by Eq. (4). In this kind of problem, the weight of the node is updated through the backpropagation algorithm for fitting the nonlinear relationship. In addition, a large number of convolutional kernels can be set in the convolutional neural network,

and the ability of feature extraction can be controlled by adjusting the number of kernels. This is a convenient way to regulate the efficiency of machine learning. The low-dimensional information extracted by using the encoder is too complicated to understand. Figure 8 shows the low-dimensional information (feature vectors) of two random figures, where *rdm* represents the sequence number of the extracted low-dimensional features and *max*, *min*, *means*, and *std* represent the maximum, minimum, mean, and variance values of the feature vector, respectively; the values in the feature vectors of each picture are represented by red dots; the blue curve is obtained by sorting these values represented by the red dots. The blue curves of both pictures have the same changing trend, and the statistical quantities such as the maximum, minimum, mean, and variance values of different feature vectors are very close. This indicates that the machine may have grasped the general law of temperature distribution after learning in step 1. Because the pictures can be reconstructed precisely, it is reasonable to believe that the complicated spatial–temporal information such as the position of Bénard cells and temperature is contained in the feature vectors. This information is coupled together to constitute the feature vectors finally. Therefore, it is an effective way to use the CNN to recover the picture of the temperature field from the feature vectors predicted by the RNN,

$$\begin{aligned} \text{Sigmoid}(x) &= \frac{1}{1 + e^{-x}}, \\ \text{ReLU}(x) &= \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0. \end{cases} \end{aligned} \tag{4}$$

In step 2, the joint model of RNN and CNN is utilized to predict the evolution of the pattern. LSTM is used to help the model extract and remember important information effectively. LSTM serves as a memory cell with a logic gate system. Its forgetting module helps the memory cell select information that should be remembered. The structure of LSTM is shown in Fig. 9. The operation in the memory cell is expressed by Eq. (5), where *C<sub>t</sub>* represents the state of the cell, *f<sub>t</sub>* represents the forget gate, *i<sub>t</sub>* represents the input gate, and *h<sub>t</sub>* represents the hidden cell. In order to recover the picture more effectively, in the joint model, the CNN is rebuilt by adding upsampling structures based on the nearest neighbor algorithm between convolutional layers instead of using traditional convolution and pooling operation. The model structure is shown in Fig. 10, and the specific structures of each layer are shown in Table III,



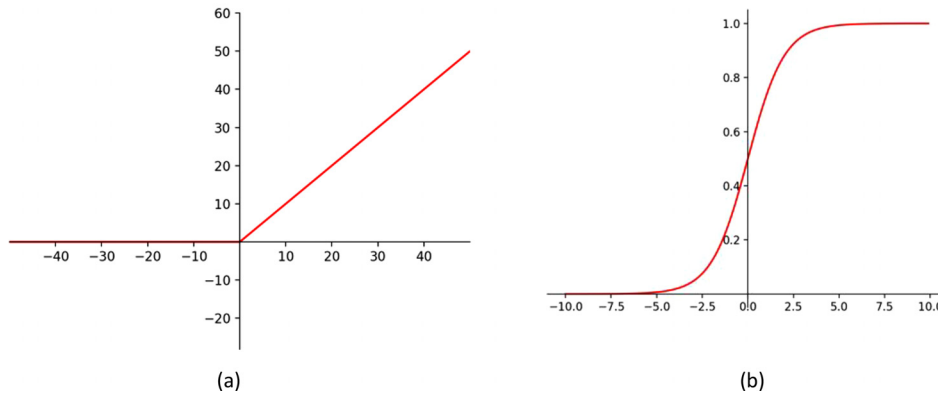


FIG. 7. ReLU (a) and sigmoid (b).

$$\begin{aligned}
 f_t &= \sigma(W_f \cdot [h_t, x_t]) + b_f, \\
 i_t &= \sigma(W_i \cdot [h_{t-1}, x_t]) + b_i, \\
 \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t]) + b_C, \\
 C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t, \\
 o_t &= \sigma(W_o \cdot [h_{t-1}, x_t]) + b_o, \\
 h_t &= o_t * \tanh(C_t).
 \end{aligned}
 \tag{5}$$

In this experiment, the mean square error between the original and predicted pictures is used as the loss function [Eq (6)] to optimize the network structure. In order to help the machine to learn the data better, the input low-dimensional feature vectors should also be normalized before calculating. In the normalization process before RNN\_CNN training. If we want to help the machine understand the data better, we need to ensure the reduced image matrix (740 × 2048) is normalized by column, otherwise, the data of a certain dimension will become so small that be ignored in the training.

$$LF = \frac{1}{2M} \sum_{m=1}^M (y_m - \bar{y}_m)^2,
 \tag{6}$$

where LF represents the loss function,  $y_m$  is the value of each pixel in the original picture,  $\bar{y}_m$  is the value of the corresponding pixel in the predicted picture, and M is the total number of pixels in the picture.

In order to help the model to calculate better, we added the lasso regression term after the loss function. Lasso regression is actually adding an L1 norm about model weight after the loss function, compared with other regularization methods, and the L1 norm can effectively generate sparse coefficients in the subsequent gradient descent calculation. Therefore, it can be said that lasso regression makes a “parameter selection,” and the unselected weights are returned to 0. The regression term can help the model avoid over-fitting and enhance the generalization effect of the model.

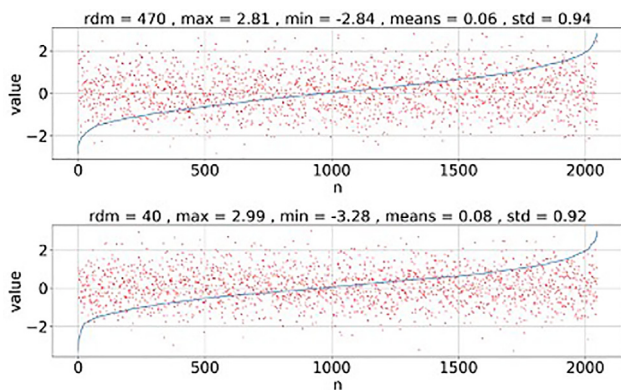


FIG. 8. Low-dimensional feature vectors.

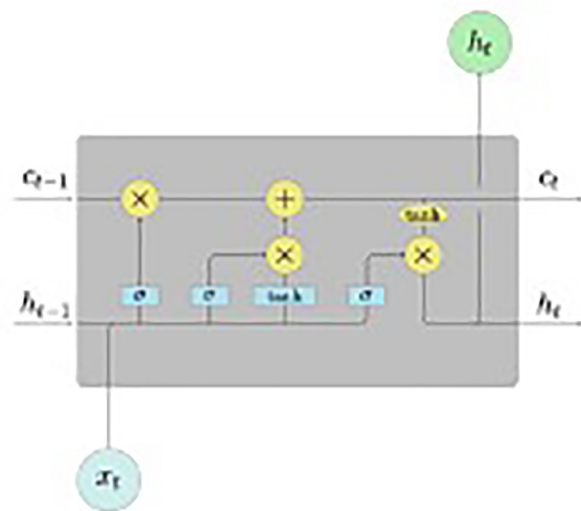


FIG. 9. The structure of LSTM.

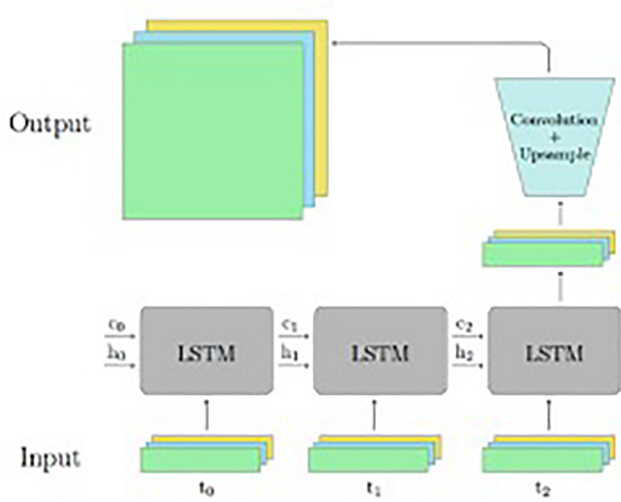


FIG. 10. The structure of the RNN-CNN joint model.

**B. Supplement and correction to the data**

The frame-by-frame prediction method is employed in the supplement and correction to the actual experimental data. The frame-by-frame prediction focuses on making the machine deduce the experimental result at present according to the experimental results at three previous moments. In the frame-by-frame prediction, two different ways of dividing the datasets are adopted. The first one is

to select 30 pictures randomly from 740 original pictures as the testing dataset and use the other 710 pictures as the training dataset in order to judge the machine's degree of mastery of the overall law and its ability to supplement actual data; the second one is to select 15 consecutive pictures in the process of heating and cooling, respectively (30 in total), as the testing dataset in order to judge whether the machine has the ability to identify the details of the data and understand the rules of generation and evolution of Bénard cells and whether it is possible to correct real experimental data.

**1. The data supplement**

In the process of data supplement, pictures in the testing set need to be selected randomly. After several rounds of training, weights in the network are optimized by machine, and good results have been achieved finally. Figure 11 shows the comparison between the result (d) on the testing set, which is predicted from results (a)–(c), and the corresponding practical experimental picture (e) after multiple rounds of learning. There is a Bénard cell generated obviously in the area framed by the red square. It shows that the prediction results of the model are highly similar to the actual results.

The following two methods are used to describe the quantitative relationships between the predicted image and the original image in the total pixel value and the distribution of pixel values. The weights in the model are optimized by using the backpropagation algorithm. According to Eq. (6), the average mean square error is 0.0024 on the training set and 0.02 on the testing set. This means there is little difference in the total pixel value between original and predicted pictures.

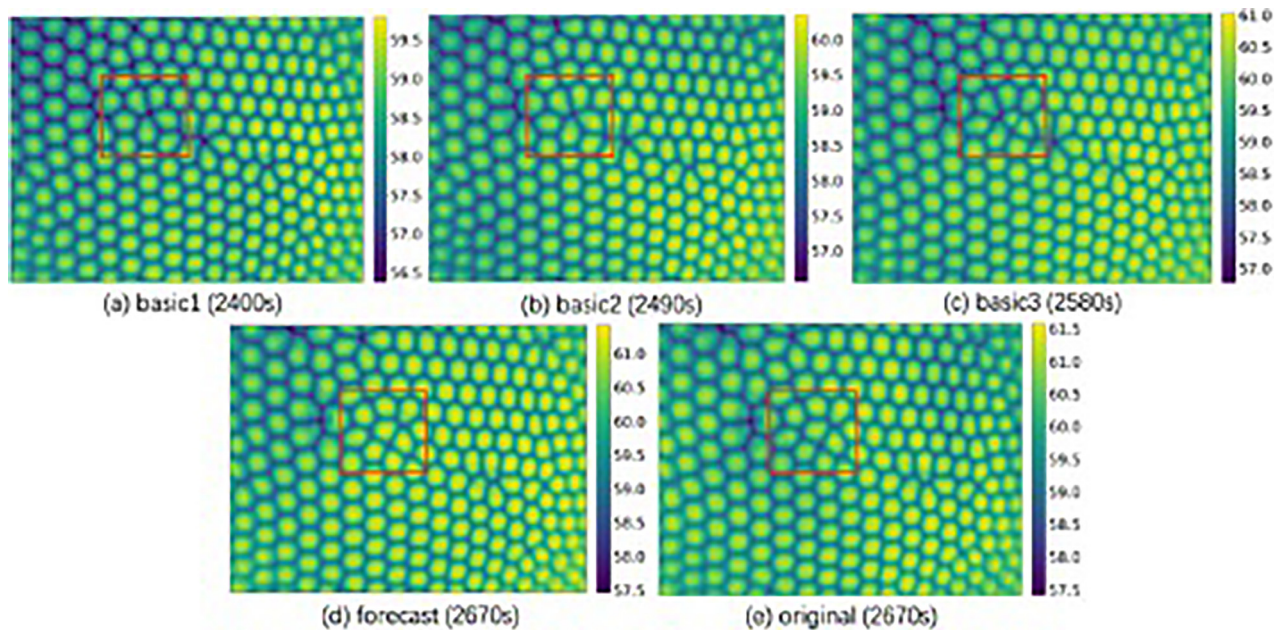
According to Eq. (3),  $R^2 = 0.9990$  on the testing set. This indicates that the joint model has represented the differences between

TABLE III. RNN-CNN structure.

Number of neural network layers	Neural network parameters	Activation function	Layer normalization or not
RNN	2048, 2048	ReLu	N
FC1	4 096	None	N
FC2	76 800	ReLu	N
Conv.1	3, 3, 256, 256	ReLu	N
Conv.2	3, 3, 256, 384	ReLu	N
Conv.3	3, 3, 384, 256	ReLu	N
Conv.4	3, 3, 256, 96	ReLu	N
Conv.5	3, 3, 96, 1	Sigmoid	N

TABLE IV. Cross validation.

	Val.1 (calculated)	Val.2 (calculated)	Val.3 (calculated)	Test (calculated)
Val.1 (experimental)	0.101	0.747	0.814	0.549
Val.2 (experimental)	0.806	0.059	0.854	0.831
Val.3 (experimental)	0.851	0.839	0.073	0.865
Test (experimental)	0.684	0.866	0.901	0.096



**FIG. 11.** Data supplement results: (d) corresponding to the supplement result on the testing set, which is predicted from experimental results (a)–(c), and the corresponding practical experimental picture (e).

different elements in each feature vector well in the resulting picture. The predicted pictures have similar distribution of pixel values with the original pictures, rather than that the elements in the feature vector are simply mapped to the average pixel value of the corresponding original picture to reduce the prediction error. Hence, the model has high credibility.

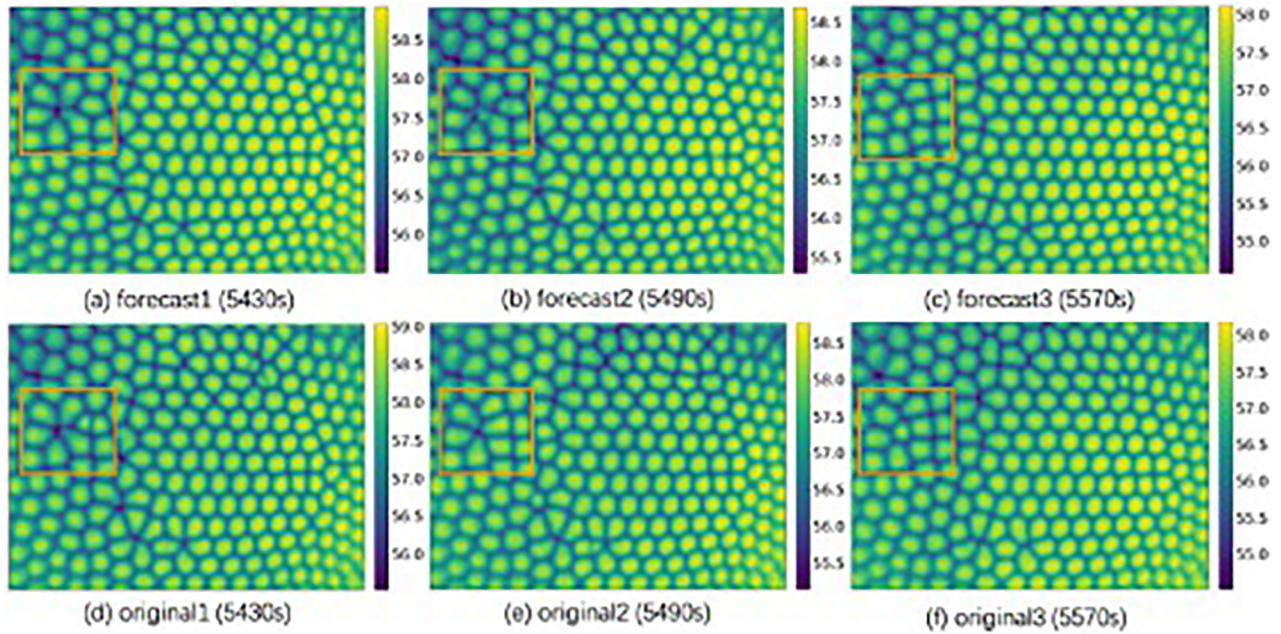
In this part, the predicted results by using the machine learning approach show remarkable agreement with the practical experimental results. This indicates that machine learning methods have strong ability to learn from data. A few randomly selected pictures are used as the testing set. Since the machine has mastered the evolution law of the pattern before and after the selected picture during the training, it can perform predictions on the selected picture in the testing set. It will be really helpful for data supplement in the future work. We can train the model by selecting pictures before and after the missing frame as the training set and treat the picture at the missing frame as the testing set. The predicted results in the testing set can be used as the supplement to the missing data of the experiment.

## 2. The correction of data

In the process of data correction, consecutive pictures need to be selected as the testing set. There are big differences in training methods between the experiment with the consecutive extraction method and the experiment with the random extraction method. In order to gain the ability to generate continuously changing pictures, the machine needs to have (gain) creativity when being trained to learn the law of evolution of the pattern. Therefore, the loss function should not be decreased too low, the regularization term should be added to the loss function, and the training rounds should be reduced as well.

After many rounds of training, every weight in the model has been optimized, and good results have been achieved finally. The comparisons between the forecast results (a)–(c) from the testing set and the corresponding actual results (d)–(f) are shown in Figs. 12 and 13. It can be seen that the machine learning approach has successfully predicted the generation of Bénard cells at different positions with good continuity in the Bénard cells generation process. This demonstrates that after learning through the training set, the machine has inferred the law of evolution of the pattern that is suitable for the supercritical convection, and then, it has predicted the evolution results in the testing set accurately.

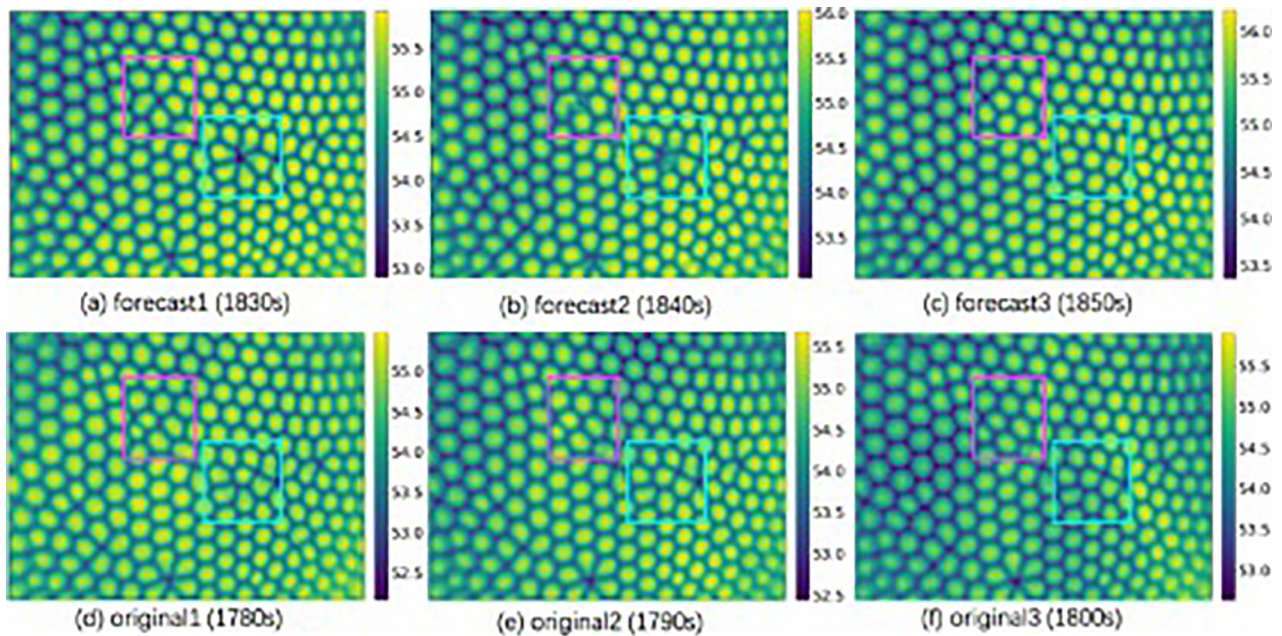
From the pattern evolution process, as shown in the square areas of Fig. 12 [the prediction images (a)–(c) correspond to the experimental images (d)–(f), respectively, at the same time points], it is clear that the new generated Bénard cell at the center of the square area is in the hotspot region; the prediction results have the same temperature distribution as the actual experimental results, and the prediction results are synchronized with actual experimental results. However, sometimes the generation time point is delayed, as shown by the square areas with blue frames in Fig. 13 [the prediction images (a)–(c) correspond to the experimental images (d)–(f), respectively, with the time points of the experimental pictures ahead of the predicted pictures], and the generation time points of Bénard cells predicted by machine learning methods are later than that shown in the actual experimental results. Sometimes, a completely different way of pattern formation is produced, as shown by the square areas with purple frames in Fig. 13. It is predicted by the machine learning methods that the Bénard cell is generated at the center because a new hotspot has appeared at the center; however, in the actual experiment, the hotspot is obviously closer to the Bénard cell on the left, and the generation process is manifested



**FIG. 12.** Synchronized results [forecast images on the top and original images on the bottom: the prediction images (a)–(c) correspond to the experimental images (d)–(f), respectively, at the same time points].

as the split of the Bénard cell on the left. During the practical experiment, the evolution of the pattern is closely related to the disturbance of temperature field; therefore, whether the machine learning approach can predict the results the same as experimental results

mostly depends on whether it can predict the distribution of the temperature field accurately. Comparing forecast pictures and experimental pictures of Fig. 13, we can see that the actual experimental results show a big change in the temperature distribution, but the prediction



**FIG. 13.** Delayed results [forecast images on the top and original images on the bottom: the prediction images (a)–(c) correspond to the experimental images (d)–(f), respectively, with the time points of the experimental pictures ahead of the predicted pictures].

results show little change in the temperature distribution. The hotspots appear on the surface before the generation of Bénard cells, so the two locations where Bénard cells are generated in the square areas in the predicted pictures are both in the higher temperature regions; well, in the actual experimental pictures, it does not have a higher temperature at the same locations. This can explain the difference between predicted results and actual results to a certain extent.

The following two methods are used to describe quantitative relationships between the predicted image and the original image in the total pixel value and the distribution of pixel values. According to Eq. (6), the average mean square error is 0.03 on the testing set. This shows that each forecast picture has little difference in the total pixel value from the original picture. According to Eq. (3),  $R^2 = 0.9416$  on the testing set. Although the differences between different elements in each feature vector have been well represented on the testing set, the prediction effect is not as good as the method with the training pictures being selected randomly. This is normal in fact.

The continuity in the results above shows that the machine has mastered the law of evolution of the pattern through the training set, and the model it gained can be generalized to a dataset that has not been seen before, which, to some extent, reflects that the machine has gained the ability to identify slight differences between adjacent pictures and the ability to imitate real experiments. As the number of consecutive pictures extracted for the testing set increases (1 and 15), although the overall effect (the value of  $R^2$ ) of the model is damaged, the predicted pictures still have good continuity, as shown in Figs. 12 and 13. This reflects that as the number of consecutive pictures extracted from the picture pool increases, the influence from outside becomes more and more non-negligible. The same as performing weather forecasts with machine learning methods, in the short-term forecast, if the disturbing factors are similar to those have been learned, the predicted results are similar to the actual results; however, in the long-term forecast, due to the complex and diversified influencing factors and the addition of unknown interference factors, the prediction results tend to have larger and larger deviations from the actual results. Therefore, the decrease in the value of  $R^2$  is quite reasonable.

The above results also show that the machine has learned the rule that Bénard cells will be generated at hotspots; in addition, it has formed its own opinion about temperature distribution and has obtained prediction results that are not consistent with the actual experimental results of the generation process of Bénard cells. At the same time, by observing the computing results of machine learning methods, it is found that the temperature drift has a great impact on the generation process of Bénard cells, but this kind of influence will be eliminated after the Bénard cell is generated. Since the real experimental process is accompanied by the interference of external environmental factors and the Marangoni number of the experiment has exceeded the critical value, even a small disturbance at any moment will affect the experimental results; as a result, the experimental results (pattern images) are uncertain. However, the calculation performed by machine learning methods is a matrix operation essentially. Except for the laws of the physical process itself, the machine simply summarizes the disturbances that have been observed before without any further disturbances being added. Therefore, there is no uncertainty in the results of the machine learning experiment, so it is quite normal that computing results do not match the actual results. From this perspective, it can be considered that the prediction results of the machine

learning approach are closer to theoretical facts compared with the actual experimental results, and they have a certain correction effect on the actual experimental results to some extent. Because of the influence of gravity, many working conditions of BM convection cannot be achieved on the ground. The machine can be fed with the actual experimental data under the conditions of adjacent parameters to supplement the experimental data under the condition of the current parameter that is missing or difficult to obtain. In addition, if the machine learns data from a more idealized experiment, it should be able to predict experimental results closer to theoretical facts and help scientists to correct actual experimental results.

Due to the small size of the data, the testing set accounts for a small proportion in the process of dividing the dataset. In order to prove that there is no over-fitting phenomenon in the model, we conduct cross validation. The dataset is divided into one testing set and three validation sets finally. If over-fitting exists in the model, the calculation results will be very similar in the data correction process, and there will be no continuous cell generation process in the results. The continuity of the final generated data qualitatively shows that the model maintains good creativity and has no over-fitting phenomenon. In the process of cross-comparison, we calculate the mean square error of the testing set and the validation set. The results show that the mean square error of each dataset reaches the minimum only when compared with the corresponding experimental results. In the mean square error calculation with other experimental datasets, we can see that the calculation result is one order of magnitude larger. It also shows that there is no over-fitting phenomenon in the data. The specific results of cross-validation are shown in Table IV.

### C. Short-term forecast

The continuous prediction method is adopted in the machine learning experiment for testing the forecast capability of the model. Continuous prediction allows the machine to use the learned model to predict subsequent continuous experimental results autonomously based on a given number of initial actual experimental pictures. In continuous prediction, all the 740 pictures are used as training data for the frame skipping learning, and then, in the test, only the preliminary 27 real experimental pictures are saved in the trained model for autonomous continuous prediction, and a total of 770 pictures are obtained in the end. Because this kind of forecast will cause error accumulation after every picture is predicted by the machine learning methods, the 27th–739th predicted pictures can reflect the self-correction ability of the machine learning approach. The 740st–754th pictures can demonstrate the short-term prediction ability of the machine learning approach, and the 755th–769th pictures can demonstrate the long-term prediction ability of the machine learning approach.

The backpropagation algorithm is used to update the model's parameters. The optimization curve of the loss function is shown in Fig. 14, where the abscissa represents the number of times the model is optimized and the ordinate represents the average of the forecast errors of all pictures after each optimization, i.e., the value of the model's loss function. It can be observed that with the increase in training rounds, the value of the loss function decreases from the maximum value 119.49 to the minimum value around 2.33, so the optimization effect is obvious.

After multiple rounds of training, each weight of the model has been optimized by the machine learning methods autonomously, and

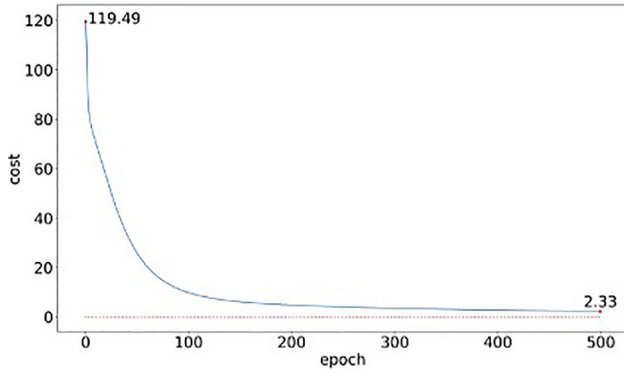


FIG. 14. The optimization curve of the loss function of RNN and CNN.

very good training results have been achieved finally. Figure 15(d) shows the predicted picture based on the consecutive forecast pictures in Figs. 15(a)–15(c) according to the law learned by the machine learning methods before. It can be seen intuitively from Fig. 15(d) that a Bénard cell is generated obviously in the square area. The actual experimental picture is shown in Fig. 15(e) for the convenience of comparison. The forecast picture matches the actual experimental picture well not only in the overall pattern but also in the distribution of the temperature field. It indicates that the machine learning approach has a great capability of self-correction because the error accumulation caused by the continuous prediction method does not affect the judgment of the machine learning approach on the Bénard cell's evolution and temperature distribution. Figure 16(d) shows the prediction effect

of the 749th (7500 s) picture by the machine learning methods. It can be observed that a Bénard cell has split obviously in the square area of the picture. The 749th (7500 s) picture of the practical experiment is shown in Fig. 16(e) for the convenience of comparison. Noticeably, Figs. 16(a)–16(d) show the calculation results by the machine learning methods, and in addition to that, Figs. 16(c) and 16(d) show pictures that the machine has never seen during the training process. From Figs. 16(c) and 16(d) as well as from Figs. 16(c) and 16(e), the central Bénard cell in the square area has increased evidently, indicating that the machine has realized that the wavenumber of the disturbance wave will decrease at this spot. The prediction result is also very similar to the actual experimental result, but the prediction effect is not as good as the 99th (1000 s) picture. In Fig. 16, the pattern shape is predicted correctly by the machine learning methods, but the temperature distribution is not predicted accurately. This is probably caused by unpredictable experimental noise and the addition of the law that has not been learned by the machine or because it is hard for the machine to make long-term predictions to the randomly changing temperature distribution.

The following two methods are used to describe quantitative relationships between the predicted image and the original image in the total pixel value and the distribution of pixel values. According to Eq. (6), the average mean square error is 0.004 on the testing set and the training set, which can reach the order of  $10^{-3}$ . This shows that each predicted picture has little difference in the total pixel value from the original picture. According to Eq. (3),  $R^2 = 0.9995$ . This shows that the joint model has represented the differences between different elements in each feature vector well on the resulting set. The forecast picture and the original picture are similar in the distribution of pixel values. It is not just that the elements

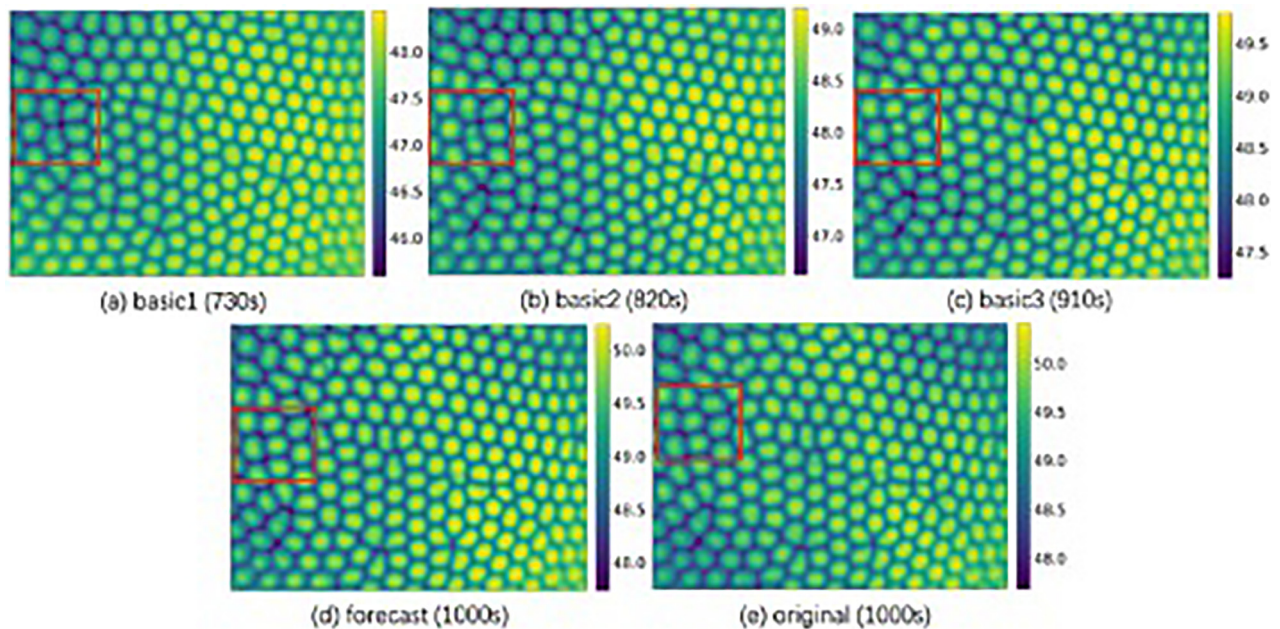
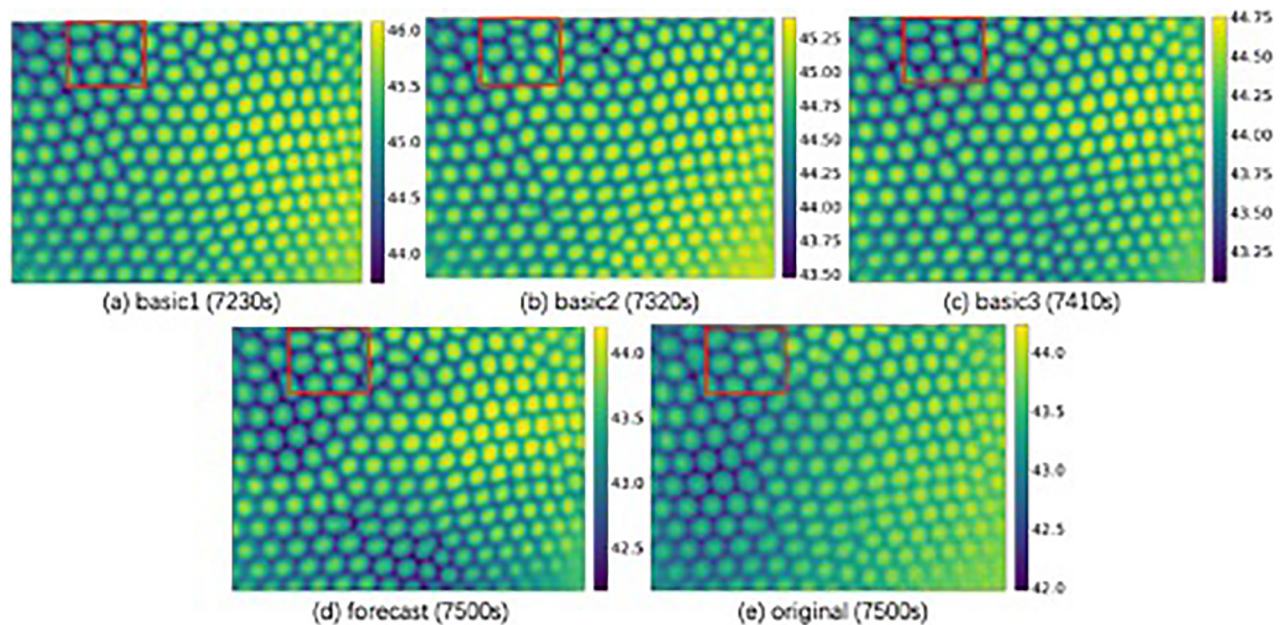


FIG. 15. The prediction effect of the 100th (1000 s) picture: (d) corresponding to the forecast result on the testing set, which is predicted from calculated results (a)–(c), and the corresponding practical experimental picture (e).



**FIG. 16.** The prediction effect of the 750th (7500 s) picture: (d) corresponding to the forecast result on the testing set, which is predicted from calculated results (a)–(c), and the corresponding practical experimental picture (e).

in each feature vector are simply mapped to the average pixel value of the original picture to reduce the prediction error. The model has high credibility.

From the above analysis, it can be seen that the machine has grasped the law of evolution of the pattern to some extent when the Ma number is increased first and then decreased within a certain range in this experimental environment, and it has gained a certain ability to perform experiments independently, that is to say, the experiment can be carried out accurately and autonomously by itself after the initial conditions are given, and the results similar to the practical experimental results can be obtained. At the same time, in order to obtain continuous experimental results, the machine must have the ability to distinguish adjacent frames of pictures; otherwise, there will be mutations caused by the generation of Bénard cells and their changes in position. The slight difference between adjacent frames of pictures is sometimes difficult to recognize by the human eyes, but the picture is just a two-dimensional matrix for the machine learning approach. When the value of an element in the matrix changes, it can be found by the machine in time. Therefore, compared with humans, it is not a complicated task for the machine learning approach to predict the tiny difference between adjacent frames of pictures. The success of the continuous prediction method also shows that the machine does have gained the ability to distinguish slight differences between neighboring pictures, reflecting the superiority of machine learning in this work. The results of experiments under unachievable experimental conditions can be predicted through this method. The prediction results demonstrate that the model has good ability in the short-term forecast in this process. During the short-term prediction of this machine learning experiment, uncontrollable interference of external

environmental factors does not affect the experimental results significantly, so this method can be well applied in actual predictions.

#### D. Long-term forecast

After the model discussed in 4.3 has been trained and the short-term prediction has been completed, long-term forecast is performed by the machine learning methods for predicting the experimental results after the 755th picture. Figure 17 shows the mean square errors of each picture calculated by Eq. (2), where the abscissa represents the sequence number of the picture and the ordinate represents the mean square error of the picture. It can be seen from Fig. 17 that the values of MSE are lower than 0.04 before the 755th picture (7560 s) but higher than 0.04 after that. This indicates that the model can only predict the experimental results of the first 755 pictures (before 7560 s) accurately. After the 755th picture (7560 s), the mean square error of the model's prediction result will rise significantly. In order to explain this phenomenon, we plot the changing curve of average values of 798 original experimental pictures, as shown in Fig. 18, where the abscissa represents the sequence number of the picture and the ordinate represents the average pixel value of the picture. It can be seen that the average pixel value of the first picture and that of the 755th picture are both around 42.98 on the same horizontal line. In terms of the experimental state, the average temperature of the fluid system is the same in the two pictures. In other words, after the 755th picture, the model is called upon to make predictions in a temperature range that has never been learned. It is more complicated that the model learns a hydrodynamic process that includes subcritical bifurcations and leads to chaos, which means that the evolution of the pattern during the decreasing process of temperature difference of the BM convection system is not

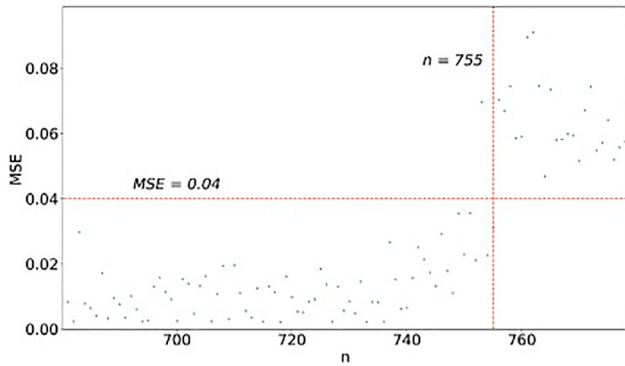


FIG. 17. Scatter plot of MSE.

a simple inverse of that during the increasing process of temperature difference. Therefore, sudden deterioration of the prediction effect is understandable when the temperature range and the amount of learning data are limited.

The Ma number is gradually decreasing after the 755th (7560 s) picture. There are two forms of abnormal prediction effect shown in Fig. 19 [the predicted pictures (a) and (b) correspond to the experimental pictures (d) and (e), respectively, at the same time points]. One of the abnormal results is that the pattern is incorrect and disappearing, as shown by the pictures on the left of Fig. 19. However, in some clear regions, as we can see in the square areas with orange frames, the calculated results are relatively correct in many locations but have a little displacement compared with the actual results. Noticeably, the pattern near the red line in the square area is having a new Bénard cell being generated, so the pattern here is more representative. To some extent, it reflects that although the evolution of the pattern during the decreasing process of temperature difference is not a simple inverse of that during the increasing process of temperature difference due to the existence of subcritical bifurcations, the formation mechanism of the Bénard cell has not changed. At the same time, it can be seen that there are differences between the pattern in the areas circled by blue curves and surrounding areas of the predicted picture and that in the corresponding areas of the original picture. The pattern has more Bénard

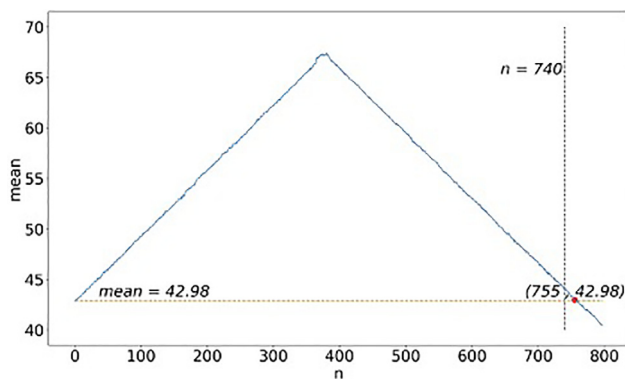


FIG. 18. Average (temperature) changing curve.

cells with a more complicated structure in these areas of the predicted picture. Since the temperature difference between the high-temperature zone and the low-temperature zone at these locations in the predicted picture is larger than that in the original picture, this effect is expected. The other abnormal result is the blurred pattern, as shown by the pictures on the right of Fig. 19. However, in some clear regions, as shown in the square area with the purple frame, we can see that the correct pattern is given by the machine. In this way, we can use machine learning methods to predict the results of some experiments that are not completed practically. The abnormal prediction results still have some value. On the one hand, the normal part of the predicted picture may reflect the characteristics of the pattern without external interference. On the other hand, the anomalous part of the predicted picture may reflect the influence of uncontrollable factors (such as changes in temperature and physical background) on the actual experimental results. They all have certain enlightenment values for the final result analysis.

The learning material prepared for the machine learning approach does not contain knowledge about how the Bénard cell will annihilate during the cooling process. Therefore, it is normal that the machine cannot determine how the Bénard cell should develop in the region where the Bénard cell is about to disappear during the prediction process.

The following three points are also the reasons for the abnormal results in the long-term forecast.

### 1. Unpredictable temperature field

During training, the machine has learned that the location where Bénard cells will be generated is closely related to the temperature distribution. During the actual experiment, the temperature distribution will be affected by various disturbances. In the short-term prediction, the temperature drift will not have too much influence on the generation location and moving direction of the Bénard cell, so the machine can obtain conclusions that match the experimental results; however, in the long-term prediction, such temperature changes are unpredictable, so the machine will be confused, leading to abnormal results.

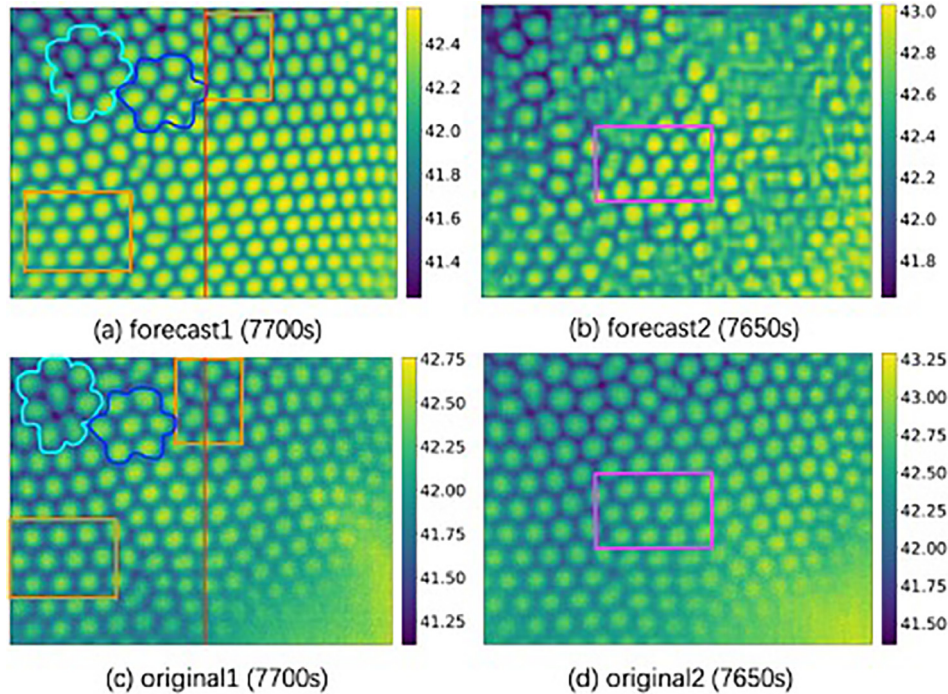
### 2. Chaos effect

When the Ma number exceeds the critical value in the actual experiment and BM convection gradually evolves into chaos, the differential equation will be more and more easily affected by the quadratic term (nonlinear term), so the influence of small disturbances on the convective system will be more and more significant. Due to error accumulation during the calculation process and the fact that the actual experiment may be subject to unpredictable interference of external factors in the following part, the machine learning approach will lose the basis of prediction in the relatively long-term prediction, resulting in abnormal prediction results. To a certain extent, this can also prove that the machine learning approach has discovered that the nonlinear equation satisfying the fluid system is sensitive to the initial value.

### 3. The complexity of the physical process

In the task of fitting the actual physical model with the machine learning methods, we select a section of the supercritical process with





**FIG. 19.** Two error modes (incorrect pattern on the left and blurred pattern on the right): the predicted pictures (a) and (b) correspond to the experimental pictures (d) and (e), respectively, at the same time points.

the flow system being heated up first and then cooled down. Because the selected experimental process is incomplete, the machine does not have chance to learn the rule of pattern disappearance and the knowledge about unpredictable interference of external factors in the following part of the experiment. Therefore, it is difficult for the machine learning methods to predict long-term results accurately based on the existing dataset.

In addition, when the temperature difference of the fluid system increases first and then decreases, the characteristics of the pattern at different stages (such as the proportion of different pattern shapes) will also change. Our experiment is a preliminary study on the characteristics of the overall change in the pattern under supercritical convection conditions. Because the heating and cooling processes are not subdivided into smaller sections, the knowledge about pattern evolution grasped by the machine is insufficient for long-term predictions. However, to some extent, this also reflects the authenticity and reliability of machine learning methods.

## V. CONCLUSIONS

This machine experiment is only a preliminary attempt to combine the machine learning method with the hydrodynamics experiment. The focus of the machine experiment is to judge whether the machine can master the laws of fluid mechanics and grasp the law of experimental data change from the perspective of algorithm and statistics. Through the above analysis, we know that the complexity of the physical process is the main reason for the deviation of the calculation

results from the experimental results. In future research, if we can further subdivide the physical process and add physical quantities such as boundary conditions to constrain the calculation system, then the prediction results of the machine will be more accurate.

In this experiment, 740 infrared thermal images of the one-layer fluid in the supercritical process of BM convection have been prepared for the machine learning. We have obtained a powerful model suitable for predicting the state of the BM convection system. It combines dimensionality reduction algorithms with prediction algorithms and has obtained better prediction results with less data. In this paper, the CAE model based on the convolutional neural network has excellently performed dimensionality reduction on the dataset and has successfully completed the task of extracting the feature vector of the image. 76 800 pixels in each figure are replaced with 2048 data points, the amount of data is compressed to 3% of the original, and the computational efficiency of the subsequent RNN-CNN model is improved; the RNN-CNN joint model has predicted the change in the feature vector extracted by using the CAE and has successfully recovered the final predicted feature vectors from predicted pictures. The three models of data supplement, data correction, and short-term data prediction have all achieved excellent results with  $R^2$  greater than 0.94 on the testing set, which shows that the model has grasped the rules of the distribution of data points and the rules of the evolution of data point distribution in the picture. The model has achieved the objective of supplementing the missing experimental data and the objective of correcting actual experimental data through a comparison between actual

experimental results, prediction results by the machine learning methods, and theoretical analysis results. These two achievements will have good application prospects in practical research work. At the same time, active exploration has been conducted in predicting physical experimental results. The model we have trained can be generalized to get more theoretical results. The prediction results that match or unmatch actual experimental results are all valuable in understanding the law of evolution of the supercritical process in BM convection (chaos system).

This is a novel attempt to model the supercritical process of Bénard–Marangoni convection in a one-layer fluid system. The machine learning methods no longer provide discrete differential equations and build a computational mesh to gradually deduce the state of the system at a certain moment but use statistical methods to fit the experimental data to the statistical distribution to infer unknown experimental results. In addition, a lot of convolution kernels are used to extract structured information of the pattern. All these are very helpful for the research on BM convection. The entire machine learning experiment shows obvious advantages of the new method in computational efficiency and the authenticity of results.

In the above machine learning experiment, after mastering the law of evolution of the pattern in the learning stage, the machine can autonomously deduce the pattern evolution in the testing stage and predict subsequent experimental results continuously from a few initial experimental pictures. The trained machine (network) has the ability to imitate actual physical experiments to some extent. In addition, the image data we have used in the machine learning experiment contain complex spatial–temporal information, so the structure of the model can be generalized to other application scenarios where image data are also used.

The trained machine has learned part of the law of evolution of the pattern in the fluid system under certain working conditions when the  $Ma$  number changes within a limited range. For the experimental study on BM convection, in the future, the pattern learning data that are more detailed, in a larger range, and under more working conditions of the fluid system will be provided to the machine for targeted learning in order to develop autonomous experiments completely conducted by the trained machine. At the same time, we will try to describe the strange attractor's appearance after the system enters into chaos under specified working conditions, which will help us to understand the turbulent structure better.

#### AUTHORS' CONTRIBUTIONS

Y.C. and D.W. contributed equally to this work.

#### ACKNOWLEDGMENTS

This project was financially supported by the National Natural Science Foundation of China (Grant Nos. 12032020 and 12072354),

the Manned Space Program of China, and the Strategic Priority Research Program on Space Science of the Chinese Academy of Sciences.

#### DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

#### REFERENCES

1. R. A. Pearson, "On convection cells induced by surface tension," *J. Fluid Mech.* **4**(05), 489–500 (1958).
2. A. A. Golovin, A. A. Nepomnyashchy, and L. M. Pismen, "Pattern formation in large-scale Marangoni convection with deformable interface," *Physica D* **81** (1–2), 117–147 (1995).
3. E. L. Koschmieder and M. I. Biggerstaff, "Onset of surface-tension-driven Bénard convection," *J. Fluid Mech.* **167**, 49 (1986).
4. J. W. Scanlon and L. A. Segel, "Finite amplitude cellular convection induced by surface tension," *J. Fluid Mech.* **30**(1), 149–162 (1967).
5. M. F. Schatz, S. J. VanHook, W. D. McCormick, J. B. Swift, and H. L. Swinney, "Onset of surface-tension-driven Bénard convection," *Phys. Rev. Lett.* **75**(75), 1938–1941 (1967).
6. E. N. Lorenz, "Deterministic nonperiodic flow," in *The Theory of Chaotic Attractors*, edited by B. R. Hunt, T. Y. Li, J. A. Kennedy, and H. E. Nusse (Springer, New York, NY, 2004).
7. W. Tucker, "The Lorenz attractor exists an auto-validated proof," *C. R. Seances Acad. Sci.* **328**(12), 1197–1202 (1999).
8. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016).
9. D. Wu, L. Duan, and Q. Kang, "Defects of Bénard cell on a propagating front," *Phys. Fluids* **32**(2), 024107 (2020).
10. D. Wu, L. Duan, and Q. Kang, "Wavenumber selection by Bénard–Marangoni convection at high supercritical number," *Chin. Phys. Lett.* **34**(5), 054702 (2017).
11. X. Shi, Z. Chen, H. Wang *et al.*, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems* (ACM, 2015) **Vol. 1** pp. 802–810.
12. J. Huang, H. Liu, and W. Cai, "Online *in situ* prediction of 3-D flame evolution from its history 2-D projections via deep learning," *J. Fluid Mech.* **875**, R2 (2019).
13. R. Han, Y. Wang, Y. Zhang, and G. Chen, "A novel spatial-temporal prediction method for unsteady wake flows based on hybrid deep neural network," *Phys. Fluids* **31**, 127101 (2019).
14. P. Vincent, H. Larochelle, Y. Bengio, and P. A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Machine Learning, Proceedings of the Twenty-Fifth International Conference (ICML 2008), June 5–9, 2008* (ACM, Helsinki, Finland, 2008).
15. A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *NIPS*, 2012.
16. T. Mikolov, M. Karafiát, L. Burget, and K. S. Jan Cernocký, "Recurrent neural network based language model," in *INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26–30, 2010* (DBLP, 2010).
17. Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks* (MIT Press, 1998).