

Machine learning based image processing technology application in bunch longitudinal phase information extraction

Xingyi Xu¹ and Yimei Zhou

*Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201800, China
and University of Chinese Academy of Sciences, Beijing 100049, China*

Yongbin Leng^{1*}

*Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201204, China;
Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201800, China;
and University of Chinese Academy of Sciences, Beijing 100049, China*



(Received 17 December 2019; accepted 16 March 2020; published 25 March 2020)

We report on the application of machine learning (ML) methods to extract longitudinal phase information such as parameters of the synchrotron damping oscillation. Parameters of the synchrotron damping oscillation are important for the evaluation of machine status and bunch stability. It is of concern to extract these parameters with high-speed and high-precision. The previous methods, such as multi-parameter nonlinear fitting and table look-up, are slower and easily fall into local optimal solutions. Our approach based on ML-image processing consists of training a virtual diagnostic to predict parameters using the beam position monitor (BPM) electrical signal data as inputs. We find that when the noise of data is large, our ML-model can still get better results than other methods, an important step toward on-line multiparameter extraction from multidimensional raw data.

DOI: [10.1103/PhysRevAccelBeams.23.032805](https://doi.org/10.1103/PhysRevAccelBeams.23.032805)

I. INTRODUCTION

The injection transient process is a good starting point to study the beam instabilities since it occurs frequently in top-off mode. Oscillations will be invoked due to the energy mismatch between the storage ring and the injector. The study of the injection transient is helpful for optimizing the state parameters of the injector and understanding the physical process of the merging of the stored charge and the refilled charge during the injection process [1–4]. Starting from 2012, the BI (beam instrument) group of SSRF (Shanghai Synchrotron Radiation Facility) performs bunch-by-bunch phase measurement and study the injection transient [5].

The refilled bunch contains the stored charge, refilled charge, and the crosstalks from other stored bunches. The stored bunch contains the stored charge and the crosstalks. If we get the turn-by-turn longitudinal phase of the refilled charge, the synchrotron damping oscillation can be monitored. The data collected by beam position monitors (BPMs) is a combination of the stored charge, refilled

charge, and the crosstalks from other stored bunches. The BI group of SSRF developed a series of algorithms to separate the refilled charge part from the refilled bunch data. The longitudinal phase of the refilled charge was obtained by using the look-up table method to process the separated refilled charge data. These phase data are fitted by gradient descent method or the fitting function library in commercial data software such as MATLAB to get the synchrotron damping oscillation parameter [6].

Although the existing method can obtain the synchrotron damping oscillation parameters from the BPM data, many problems have been found in practical applications. First, the method is very complicated and takes a long time to get the synchronous oscillation damping parameters. This method needs to construct the corresponding functions and look-up tables of the BPM in advance. Afterwards, this method also needs to eliminate the signal of stored charge and crosstalk between bunches. At last, phase of refilled charge needs to be fitted by gradient descent method or the fitting function library in commercial data software. This cannot meet the requirements of online measurement of the synchrotron damping oscillation at each injection transient. Second, this method wastes a lot of information in the BPM signal. In the existing methods, we believe that the longitudinal phase of the stored charge in each bunch is consistent. Therefore, we use the average of all the stored bunch phase as the stored charge phase in the refilled bunch, or simply choose the phase of several stored

*lengyongbin@sinap.ac.cn

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

bunches as the stored charge phase. This means that the phase relationship between the bunches is ignored. Underutilizing the data means more noise. Furthermore, linear fitting is easy to encounter the problem of getting into a local optimal solution. When the noise of signal is large, different initial parameter settings will result in different fitting results. When the initial parameters are too different from the fitted line, the time required for convergence cannot be estimated. Therefore, recently we developed a method based on machine learning image processing technology to solve these problems. With this method, we can fully extract information from large matrix data that is not easily quantified by physics theory, and get more stable conclusions quickly.

With the enhancement of computer computing performance, machine learning technology has been greatly developed in recent years. In particular, the concept of deep learning allows us to use algorithms to deal with the complex problem with a large amount of data. Accelerator engineers have also begun to gradually apply machine learning technology to beam measurement and diagnosis, and have achieved some remarkable results [7].

The development of machine learning in image processing is obvious to all, such as face recognition, automatic driving and so on. Using image processing techniques, we can give machines the ability to process image data [8,9]. In our field of beam measurement, we collect huge amounts of data every day. These data are often stored in two-dimensional or even multi-dimensional arrays. If the two-dimensional array is restored according to the storage method of the grey value image, it can be drawn into a greyscale image. Therefore, can we use cutting-edge image processing techniques to process such data in the form of multidimensional arrays and extract the information we need from it? In this regard, we made a preliminary attempt. The object we are dealing with is the electrical signal data

of the BPM when injection transients. We hope to use these data to extract the relevant parameters of synchrotron damping oscillation.

A digital acquisition board is used to collect the BPM data. These electrical signals are stored in the form of multi-dimensional arrays. A two-dimensional array can be treated as an image. We can use a large amount of data at the same time by using machine learning image processing technology. Even if a part of data has noise due to instability, we can get more accurate result than other methods, which can improve the robustness of the algorithm.

II. PREPARATION

Before building the model, we need to choose a suitable machine learning model. This model requires a fast running speed for online data processing. In addition, we need enough data for model building and testing. After consideration, we chose the convolutional neural networks (CNN) as our predictive model and established a data acquisition system to obtain training data. The true value of the training data is obtained by the bunch-by-bunch phase acquisition system that has been implemented by the SSRF BI Group. This method compares the electrical signal stimulated by the bunch through BPM with the corresponding function of the BPM electrode established by the oscilloscope to obtain the longitudinal phase. After that, the relevant parameters of the synchrotron damping oscillation during the injection transients process are obtained by the complex multi-parameter nonlinear fitting of refilled charge phase [5,10].

A. Dataset for training and testing

Figure 1 shows the framework of the bunch-by-bunch measurement system, which is used to obtain the dataset. The dataset we got consists of 482 sets of BPM electrical signals captured at the instant of 482 accelerator injection

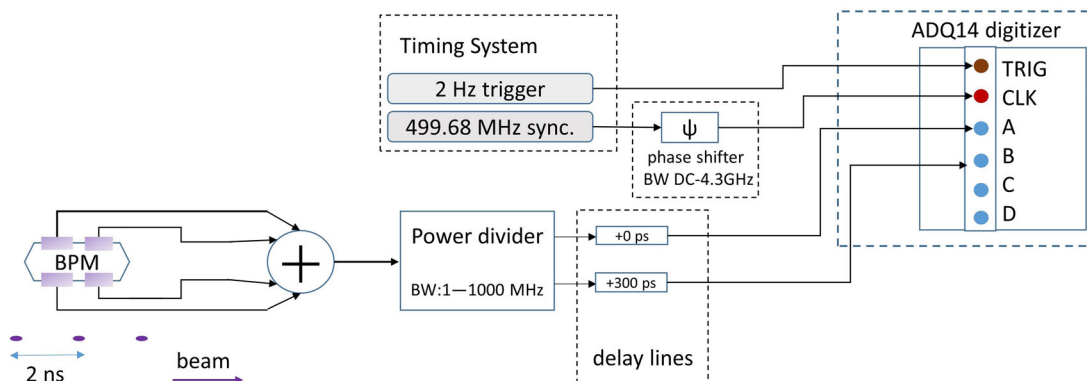


FIG. 1. The system framework of the bunch-by-bunch monitor. The BPM sum signal from four electrodes was split into two parts with 300 ps delay lines by a power divider (BW 1 MHz-1000 MHz). Two sampling points on the rising edge of the BPM signal were captured by a high-speed ADQ digitizer (14 bits, 1 GHz sampling rate, 1:2 GHz bandwidth). The ADC external clock signal of 499.654 MHz, directly come from the timing system of the machine, aiming at the synchrotron sampling of signals. The clock was followed by a tunable shifter with maximum phase excursion of 2 ns to ensure the capture of zero-crossing point of the rising edge of the signal. The system can achieve 3 million points (2083 turns) of data, which is enough for us to obtain the whole oscillation process.

TABLE I. Dataset for CNN.

Parameter	Original data	Preprocessed data
Shape	482*512*2080*2	482*800*1800*2
Set	482	482
Bunch	512	800
Turn	2080	1800
Channel	2	2
Range	-15000-20000	0-1

processes. We use a threshold trigger algorithm to automatically capture the injection process. In order to realize the acquisition of BPM electrical signals, we choose a high-speed data acquisition board. With the fixed delay lines, we realize that each single bunch collects two sampling points each turn when it goes through the BPM. The two data points are both sum of four electrodes of a BPM, separated by 300 ps delay. We use channel A and channel B to refer to these two data channels. The stored data format is $512 \times 2080 \times 2$. Among them, 512 means more than 500 bunches in the storage ring of SSRF. We collect more than 2000 turns data in each injection process, in which the injection process starts from the first hundred turns. 2 represents the two data channels A and B. In order to facilitate and speed up the training process, we cut and normalize the data. What is more, we expanded the refilled bunch by 300 times to highlight the refilled bunch. It means that there is only one refilled bunch, but we copy it into three hundred bunches to increase its visual size. The data finally introduced into the algorithm is $482 \times 800 \times 1800 \times 2$ in the range of 0-1. The dataset is shown in Table I.

B. Convolutional neural network

Convolutional neural networks (CNN) is a type of feedforward neural networks with convolutional computation and deep structure. It is one of the representative algorithms of deep learning [11–13]. CNN has the ability to represent learning, and can shift-invariant classification of input information according to their hierarchical structure. Therefore, it is also called “shift-invariant artificial neural networks (SIANN)” [14]. CNN is inspired by the visual organization of living things. Visual cortical cells receive signals from photoreceptors on the retina. A single visual cortical cell does not receive all of the signals from the photoreceptor, but only accepts the signal from the stimulus region it dictates. Only by feeling the stimulation in the field can the neuron be activated. Multiple visual cortical cells systematically superimpose the receptive field, completely receiving signals transmitted by the retina and establishing a visual space.

It is beneficial to use CNN to obtain synchrotron oscillation parameters from BPM multidimensional array data. First of all, the data we need to process is data which contains 500 bunches in more than thousands of turns. If we use common back propagation neural networks instead of CNN, it will require an extremely large number of weights [15]. If we want to process $M \times N$ two-dimensional array, the number of model’s weights will be

$$\text{weight number} \geq M^2 \times N^2. \quad (1)$$

According to Eq. (1), processing data with 500 bunches and 1800 turns requires more than 80 billion weights, which is unacceptable for online processing. However, CNN has the feature of weight sharing, which greatly

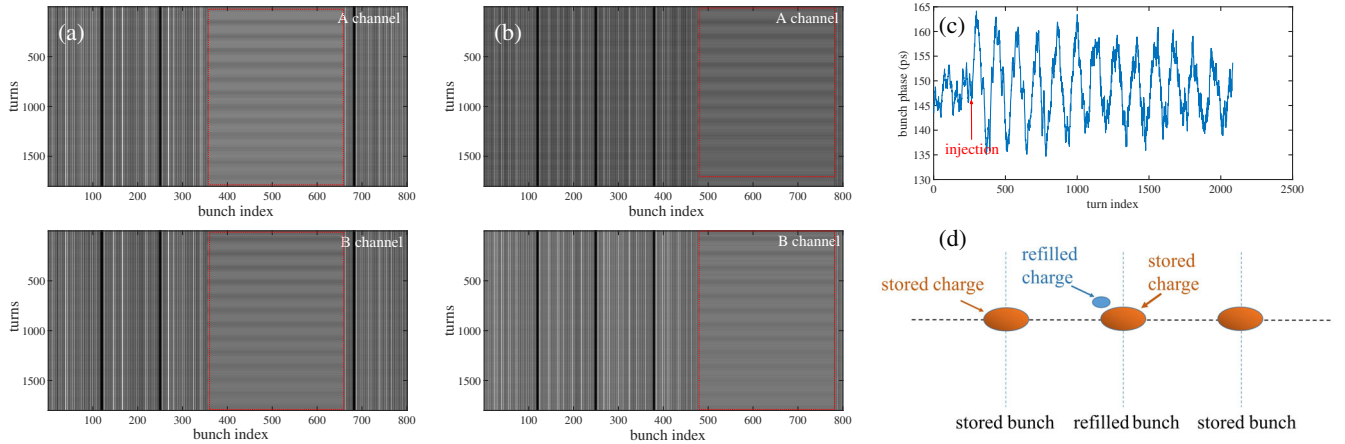


FIG. 2. Data is drawn into images. (a) and (b) show two sets data with two channels. B channel is similar to A channel. The difference between them is caused by the delay(300 ps). With both of them, we can remove the influence of different charge quantities to get the phase oscillation parameters. The area in the red dotted box is refilled bunch. There is only one refilled bunch, but we copy it into three hundred bunches to increase its visual size. Outside is stored bunches. (c) shows refilled bunch longitudinal phase. (d) shows injection process. The refilled bunch is consisted of refilled charge and stored charge. The difficulty of our work is how to remove the oscillation of refilled charge from the refilled bunch.

reduces the complexity of the network when the size of input data is large. Second, SSRF injects selectively during each injection process. Only a few bunches will be refilled. It means that which bunch will be refilled each time is not determined. Identification of refilled bunches can be done by directly comparing BPMs sum signal before and after injection. But we need using all the data as model's input because the synchrotron damping oscillation of refilled charge is stripped from refilled bunch (refilled charge and stored charge) by using stored bunch (stored charge). It is not sufficient with the refilled bunch data alone. As a result, the position of different kind of bunches in the data is not determined. The refilled bunch is in red dotted box which shown in Fig. 2. The position of red dotted boxes is different in different datasets. We use the whole image [Fig. 2(a) and (b)] as models input. It means that our model needs to find which area of images is the refilled bunch. This is a difficulty for ordinary machine learning models. Ordinary machine learning models such as BP neural networks will give different outputs when the location of the information is different. This does not satisfy our physical theory. But CNN has the characteristics of translation invariance, no matter where the refilled bunch is, the model can easily find the information we need. Last but not least, the image data processed by CNN retains the spatial and time relationship, that is, the laterally adjacent pixels are spatially adjacent bunches, and the vertically adjacent pixels are continuous time, which is beneficial to mining the intrinsic relationship of the data.

III. APPLICATION OF CONVOLUTIONAL NEURAL NETWORK IN DATASET

After data preprocessing is completed, the data of two channels A and B in each dataset can be drawn into two image which shown in Fig. 2. Figure 2(a) and (b) is two different sets of data with A and B channel. The area in the red dotted box is refilled bunch. There is only one refilled bunch, but we copy it into three hundred bunches to increase its visual size. Outside of the red dotted box is stored bunches. Our predictive model needs to distinguish where is refilled bunch area and stored bunch area which are different in different datasets. You can see ripples (oscillation) in the refilled bunch area (red dotted box). The refilled bunch longitudinal phase we obtained by look-up table method is shown in Fig. 2(c). The synchrotron damping oscillation cannot be distinguished because the refilled bunch is composed of refilled charge and stored charge [shown in Fig. 2(d)]. The synchrotron damping oscillation is carried by refilled charge. Therefore, what we must do is separate the refilled charge phase from refilled bunch by using stored bunches which are outside the red dotted box and get the oscillation amplitude and the synchrotron damping time. We let the machine learn how to deal with this problem through convolutional neural networks.

We divided the dataset of 482 injection processes into the training sets and the test sets according to 3:1, which means that there are 321 sets of data used to train the model. The remaining 121 sets of data are not known by the model. Therefore, we use this data to test the performance of the model. Our goal is to establish a model that can extract the amplitude and the damping time of the synchrotron damping oscillation of the refilled charge directly from the BPM sum signal image by a convolutional neural networks.

For prediction, we have two images including A and B channels. The corresponding position pixel of two images is the BPM data of the same bunch on the same turn. There are 300 ps delay between two channels. If two-channel sampling with delay lines is not used, it is difficult to determine whether the difference of data value is due to phase oscillations or due to differences of bunches' charge. Using this two-channel data, we can remove the effects of different charge between bunches and extract the phase oscillation parameters. Therefore, we want our CNN model to process two channels of data at once, that means using two images as input. Considering the relationship between two images, we refer to the three-channel simultaneous processing method of color image with three primary colors (cyan, magenta, and yellow). We treat each training image as a color image with two primary colors. It means that the input data is a 1800×800 image with a thickness of 2.

We establish a convolutional neural networks consisting of one input layer, five convolutional layers, six pooling layers, two fully connected layers, and one output layer to deal with this problem (shown in Fig 3). The convolution kernel in the convolutional layer can respond to converged data. The convolutional layers extract the different characteristics of the input by the movement of the convolution kernel on input layer. By pooling the data, the pooling layers downsamplings a large matrix into a small matrix, reducing the amount of computation and preventing overfitting. The fully-connected layers are similar to common BP artificial neural networks. Each neuron is connected by a weight parameter with a Rectified Linear Unit as the activation function for each neuron. In detail, when the data enters the model, it is in the shape of $1800 \times 800 \times 2$. The first layer is a convolutional layer. Considering that the data can better reflect the characteristics of the oscillation on the time scale, the convolutional kernel is set to a rectangular structure of 18×9 , so that the oscillation characteristics can be better extracted. The kernel's moving stride is 9 in horizontal and 4 in vertical. The data processed by the first convolutional layer passes through a pooling layer with a filter size of 2×2 and the stride of 2. After that, the data format becomes $100 \times 100 \times 16$. The convolutional kernels of the second to fifth convolutional layers all have a 3×3 square structure, and the convolution kernel's moving stride is 2. The second to sixth pooling layers have the same structure as the first pooling layer. These five layers of convolutional layers and pooling layers are alternately

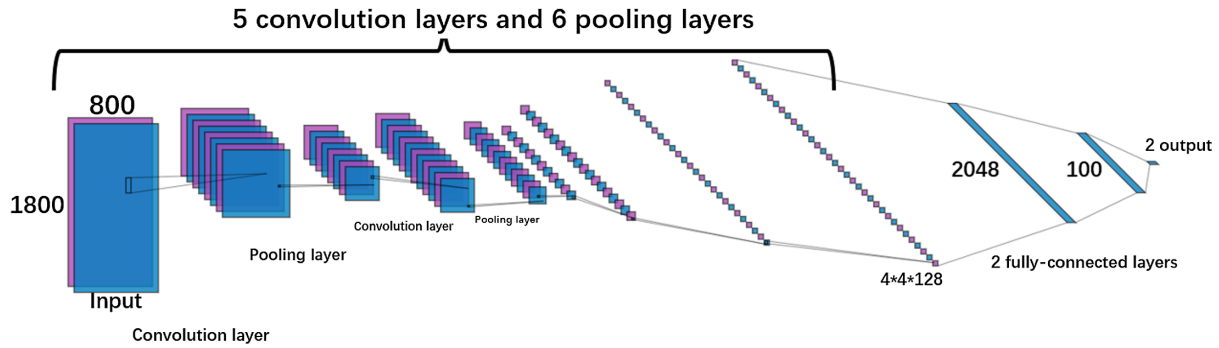


FIG. 3. The structure of the convolutional neural network built in this experiment which consisting of one input layer, five convolutional layers, six pooling layers, two fully connected layers, and one output layer.

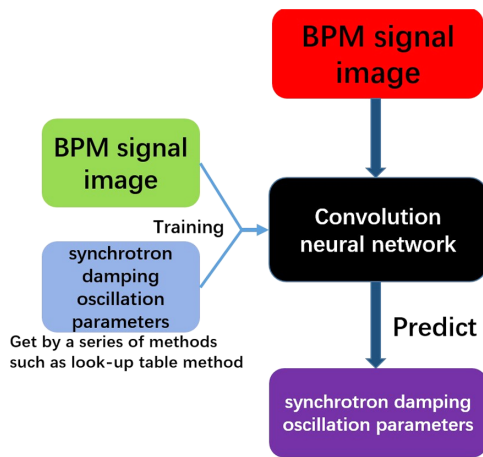


FIG. 4. Machine learning schematic to predict synchrotron damped oscillation parameters.

connected. After all this, the data is in the shape of $4 \times 4 \times 128$, and feature extraction at the image level has been completed. Therefore, the subsequent fully connected layer extends and compresses the data of $4 \times 4 \times 128$ to a long chain of 100, and the second fully connected layer further reduces and compress the data to 2 outputs. These two output values will represent synchrotron damping time and synchrotron oscillation amplitude. For all the examples presented, we use Google’s open source machine learning framework tensorflow to build models. Two NVIDIA RTX2080 graphics cards are used to speed up our training process. The schematic of our work is shown in Fig 4.

IV. PERFORMANCE ANALYSIS AND COMPARISON WITH THE TRADITIONAL METHOD

Training the convolutional neural networks with 361 sets of training data, our virtual diagnostic ML-models cost function (standard deviation between true and predictive values) quickly converges. After that, we analyze the

performance of the model and compare it with the previous traditional methods.

We apply the trained model to the 121 sets of test data and obtained the predictive value for each set of test data. The time required to process each set of data is less than 10 ms. The fast processing speed meets the real-time processing of each injection transient process data. In contrast, in the past traditional methods, it takes several minutes to remove noise, separate the stored charge data from the refilled bunch data, complete table lookup, and perform multiparameter nonlinear fitting.

We show these 121 sets of predictive value and test value in Fig. 5 (oscillation amplitude) and Fig. 6 (synchrotron damping time). We use standard deviation to estimate the difference. The standard deviation of oscillation amplitude is 5.3 ps and the standard deviation of damping time is 0.08 ms.

The difference between the predictive value and the test value is analyzed. On the one hand, because the capture time period of the injection process is long, there are only hundreds of sets of data in this training, which leads to the model not being adequately trained. On the other hand, the so-called true value (test value) is obtained by previous traditional methods. Therefore, the error introduced by this process is large. According to the error analysis about previous traditional methods, the error of the oscillation

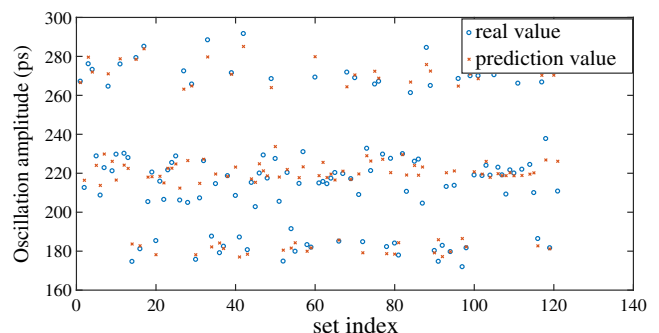


FIG. 5. Predictive value versus test value about oscillation amplitude. Blue mark: test value; orange mark: predictive value.

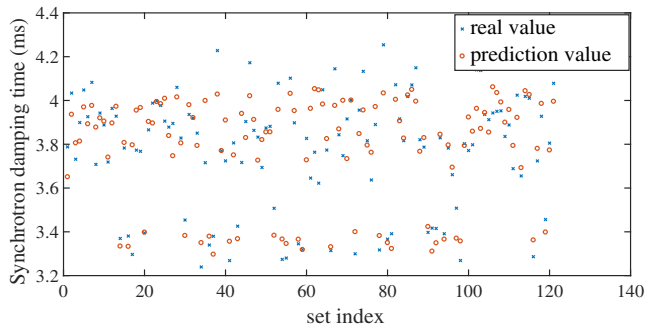


FIG. 6. Predictive value versus test value about synchrotron damping time. Blue mark: test value; orange mark: predictive value.

amplitude introduced by these algorithms will be around 2 ps. The error of the damping time is 0.1 ms. The main measurement error in the refilled charge phase extraction is derived from the random measurement error of the ADC and the look-up table error. These systematic errors will cause the calculated values to be evenly distributed above and below the actual values. In a machine learning model with a large amount of data as the training basis, errors that are evenly distributed above and below the zero value are averaged out by the learning process. This means that this ML-based virtual diagnostic method does not have these systematic errors. What is more, because of the CNN model taking into account more than 500 bunches of the whole ring rather than a few of them, it is convincing that the result got by CNN model will be closer to the real situation of the accelerator.

In order to compare the stability of the ML-based method and the traditional method, we present the results calculated by the two methods in the form of a distribution chart in Fig. 7. It can be seen that the results calculated by both methods are divided into three clusters. This is due to the fact that the data comes from the three days of SSRF injection in 15th March 2019, 22nd March 2019, and 17th July 2019. The results calculated by the ML-based method are more concentrated in the cluster center. The parameters of synchrotron damped oscillation depend on the state of the injector and the storage ring. As a result, they will oscillate in a very small range in one day. Because the ML-based method eliminates evenly distributed error through a large amount of training data, the results calculated by ML models will be more concentrated, which is consistent with the real situation.

Erroneous outliers caused by algorithm failures appear in the results of traditional method and machine learning algorithms do not have this problem. It can be seen that an outlier (red arrow shown in first image of Fig. 7) appears in the test values, which is unreasonable in a stable machine environment. So we reprocess this set of data and get the longitudinal phase of the refilled charge (shown in Fig. 8). Large noise appears at the end of the data, which will

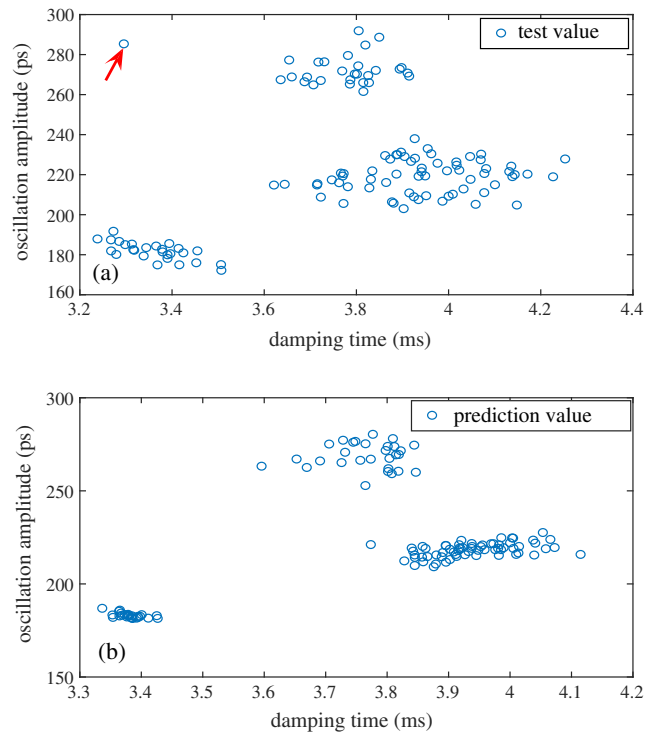


FIG. 7. The distribution of results. (a) results obtained by traditional method; (b) results obtained by CNN model.

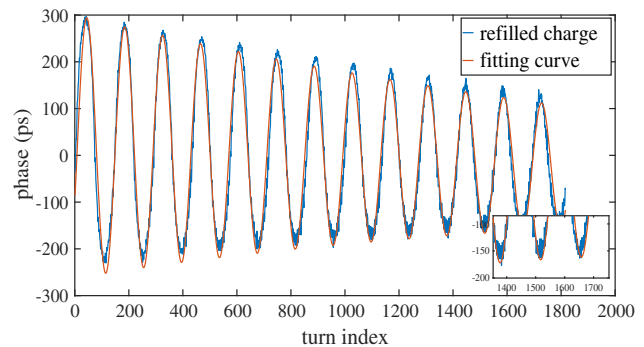


FIG. 8. The refilled charge phase and the fitting curve. Blue line: the refilled charge phase; orange line: the fitting curve of the refilled charge phase.

undoubtedly affect the accuracy of the fitting. We believe that the source of this noise may be that the refilled bunch is in a position where the crosstalk is large, and the signal of the stored charge of the refilled bunch is too different from the signal of the stored charge in other stored bunches. Therefore, the traditional method does not completely remove the stored charge signal and the crosstalk signal. We choose two methods to fit, including MATLAB Curve Fitting (we use this method to get test values) and gradient descent algorithm (realized by ourselves). We find that different fitting results will be obtained by different methods or different initial parameters. It means that the

result falls into a local optimal solution because of the noisy and different initial parameters. The synchrotron damping time we got by fitting is in the range of 3.29 ms to 3.62 ms. We plot a fitted result with a 3.6 ms damping time in the same figure. The fitting result and the actual value still have large residuals. Therefore, when the noise is large, it is difficult to obtain accurate results by fitting. The CNN model can consider the state of all bunches in a coordinated manner. After reprocessing, we find that the results obtained by the CNN prediction model are similar to the global optimal solution (3.54 ms). This reflects the stability of the CNN algorithm. Although the data used to train the model is from traditional methods, the number of outliers less than 1% of the total do not affect the model structure in a model which is not overfitting. The regularization term used in training the model causes a large number of correct values to eliminate the impact of small-scale abnormal training data.

V. CONCLUSION

In order to better understand the state of the beam, there will be more and more beam measuring instruments and higher and higher sampling rate. As a result, an accelerator will produce an extremely large amount of data every moment. These data are often time related or spatially related. It is a very important task to extract the beam parameters with high speed and high efficiency.

In this paper, the machine learning image processing technology, convolution neural network (CNN), is used to process the BPM electrical signal data stored in the form of multidimensional array. The innovation of this work is processing two-dimensional data with two channel as an image of two primary colors. Convolutional neural networks has two characteristics that is translation invariance and weight sharing. They solve the problem that the refilled bunch appear randomly in more than 500 bunches and realizes high-speed processing of large scale data. Using machine learning technology, we have realized extraction of synchrotron damping oscillation parameters directly from the BPM sum data which is important for the evaluation of machine state and bunch stability.

Compared with traditional methods, machine learning methods have many advantages and are more suitable for online systems. A trained virtual diagnostic model completely replaces the previously complicated algorithms. The time required for one parameter extraction was reduced from minutes to 10 milliseconds. Large amount of training data is eliminated the random measurement error of the ADC and the look-up table error which is unavoidable in traditional method. Because of the trapped local optimal solution and the effects of stored charge signals and crosstalk are not completely eliminated, the traditional method sometimes get wrong results. The ML-based model is trained on most of the correct data and takes the complete data of all bunches as input. We found that the results

obtained using machine learning models did not appear to be wrong. The oscillation parameter distribution is more concentrated, which is consistent with the state of the machine.

We still have a lot of work to do to optimize our model. For some reason, we only got more than 300 sets of data to train the model. This causes our model to be undertrained. In addition, we only used the data collected over three days to train the model. This means that the ML-based model might fail to predict other day data if it is very different from these three days. Therefore, we need to collect more data in different machine states to continue training the model. An online training data collection system is needed. In subsequent experiments, we intend to capture more training data to improve prediction accuracy. The BPM signal actually contains almost all the information of the bunch, we need to mine the data efficiently. Our future goal is to extract all the parameters we need from a multi-dimensional BPM data array. With this model, we can give the operating status of different monitors. What is more, we hope use this method to reduce unnecessary monitors in the future.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (No. 11575282) and Ten Thousand Talent Program and Chinese Academy of Sciences Key Technology Talent Program.

-
- [1] Z. Chen, Y. Yang, Y. Leng, and R. Yuan, Wakefield measurement using principal component analysis on bunch-by-bunch information during transient state of injection in a storage ring, *Phys. Rev. Accel. Beams* **17**, 112803 (2014).
 - [2] L. Duan, Y. Leng, R. Yuan, and Z. Chen, Injection transient study using a two-frequency bunch length measurement system at the SSRF, *Nucl. Sci. Techn.* **28**, 93 (2017).
 - [3] Y. Leng, Y. Yang, N. Zhang, Y. Yan, and R. Yuan, Bunch by bunch transverse beam position observation and analyze during injection at SSRF, in *Proceedings of the 2nd International Beam Instrumentation Conference (JACOW, Oxford, UK, 2013)* pp. 746–748.
 - [4] H. Chen, J. Chen, B. Gao, and Y. Leng, Bunch-by-bunch beam size measurement during injection at Shanghai synchrotron radiation facility, *Nucl. Sci. Techn.* **29**, 79 (2018).
 - [5] Y. Zhou, H. Chen, S. Cao, and Y. Leng, Bunch-by-bunch longitudinal phase monitor at SSRF, *Nucl. Sci. Techn.* **29**, 113 (2018).
 - [6] Y. Zhou, Z. Chen, B. Gao, N. Zhang, and Y. Leng, Bunch-by-bunch phase study of the transient state during injection, *Nucl. Instrum. Methods Phys. Res., Sect. A* **955**, 163273 (2020).
 - [7] E. Fol, J. C. de Portugal, and R. Tomas, Application of machine learning to beam diagnostics, in *Proceedings of the 7th International Beam Instrumentation Conference*,

- Shanghai, China* (JACoW Publishing, Geneva, Switzerland, 2018), <https://doi.org/10.18429/JACoW-IBIC2018-TUOA02>.
- [8] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature (London)* **521**, 436 (2015).
- [9] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogue, J. Yao, D. Mollura, and R. M. Summers, Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning, *IEEE Trans. Med. Imaging* **35**, 1285 (2016).
- [10] X. Xu, Y. Leng, and Y. Zhou, Machine learning application in bunch longitudinal phase measurement, in *Proceedings of the 10th International Particle Accelerator Conference (JACOW, Melbourne, Australia, 2019)*, pp. 2625–2628.
- [11] Y. Bengio, Learning deep architectures for AI, *Found. Trends Mach. Learn.* **2**, 1 (2009).
- [12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning* (MIT press, Cambridge, MA, 2016).
- [13] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai *et al.*, Recent advances in convolutional neural networks, *Pattern Recognit.* **77**, 354 (2018).
- [14] W. Zhang *et al.*, Shift-invariant pattern recognition neural network and its optical architecture, in *Proceedings of annual conference of the Japan Society of Applied Physics*, 1988 (unpublished).
- [15] S. Han, J. Pool, J. Tran, and W. Dally, Learning both weights and connections for efficient neural network, in *Advances in Neural Information Processing Systems 28*, edited by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett (Curran Associates, Inc., Montreal, Quebec, 2015), pp. 1135–1143.