Convolutional neural networks application in virtual diagnostics

Xingyi Xu, Yongbin Leng, Bo Gao, Ruitao Jiang

BI(beam instrument) Group, SSRF

LEAPS WG2 workshop

11^{th} May

Outline

- Introduction & Background
- Application in virtual beam diagnostics
  - deep learning image processing technology and beam parameter extraction from electrical signal
  - design accelerator parameter optimizer based on convolution kernel
- Summary & Future

Introduction & Background

- What is a convolutional neural network?
- Why should it be a Convolutional neural networks?

3

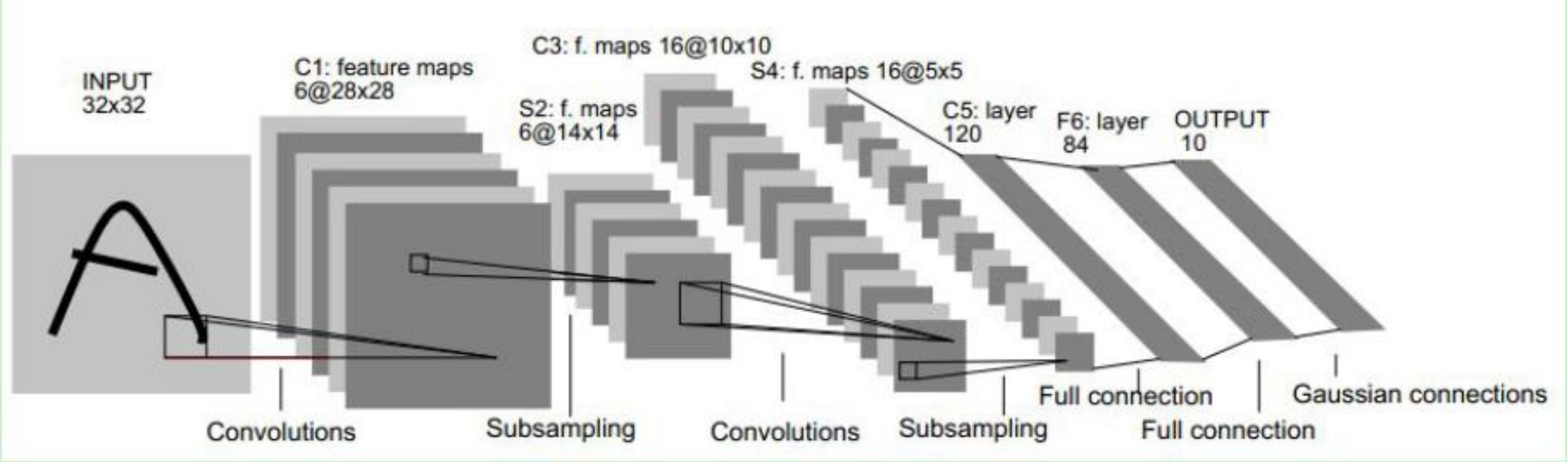
Introduction & Background

- **What is a convolutional neural network?**

What is a convolutional neural network?

In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analyzing visual imagery. —— From Wikipedia, the free encyclopedia

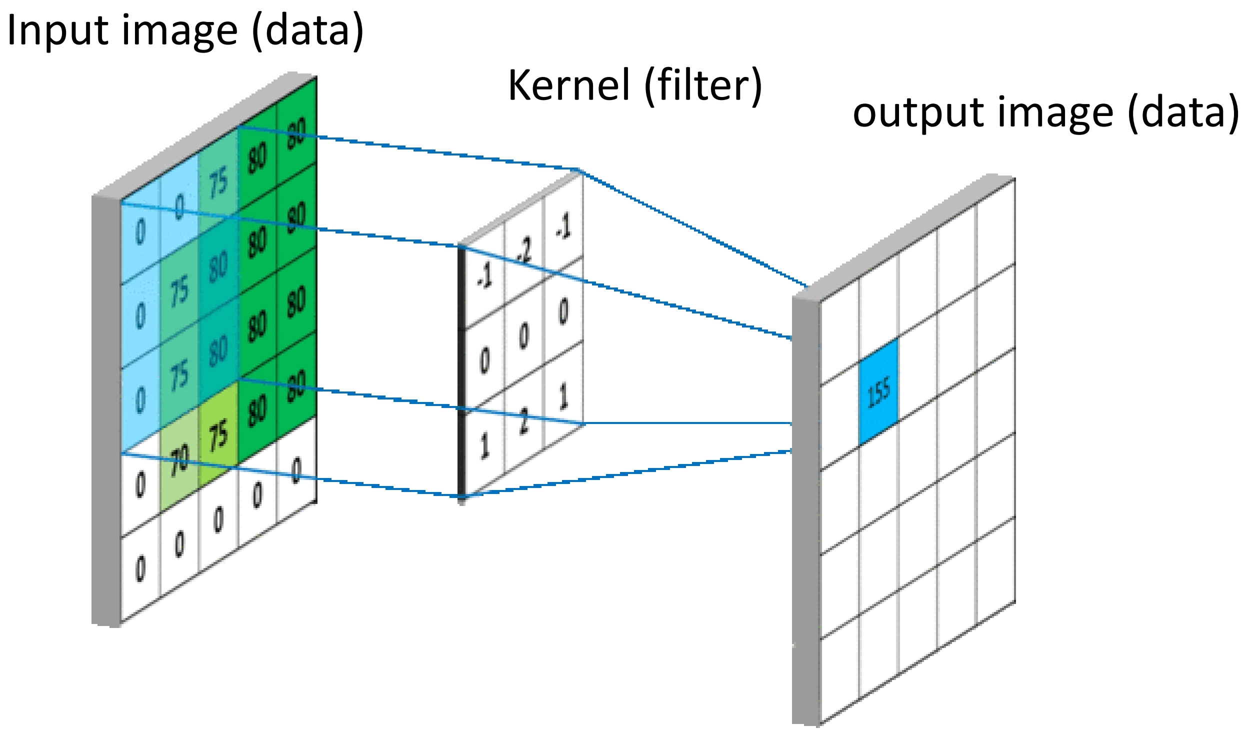
- ✓ Convolution neural network is a kind of feedforward neural network, which has the characteristics of simple structure, less training parameters and strong adaptability.
- ✓ CNN avoids the complex pre-processing of image(etc.extract the artificial features), we can directly input the original image.



schematic diagram of a classic deep convolutional neural network

Core part of CNN ___---> convolution

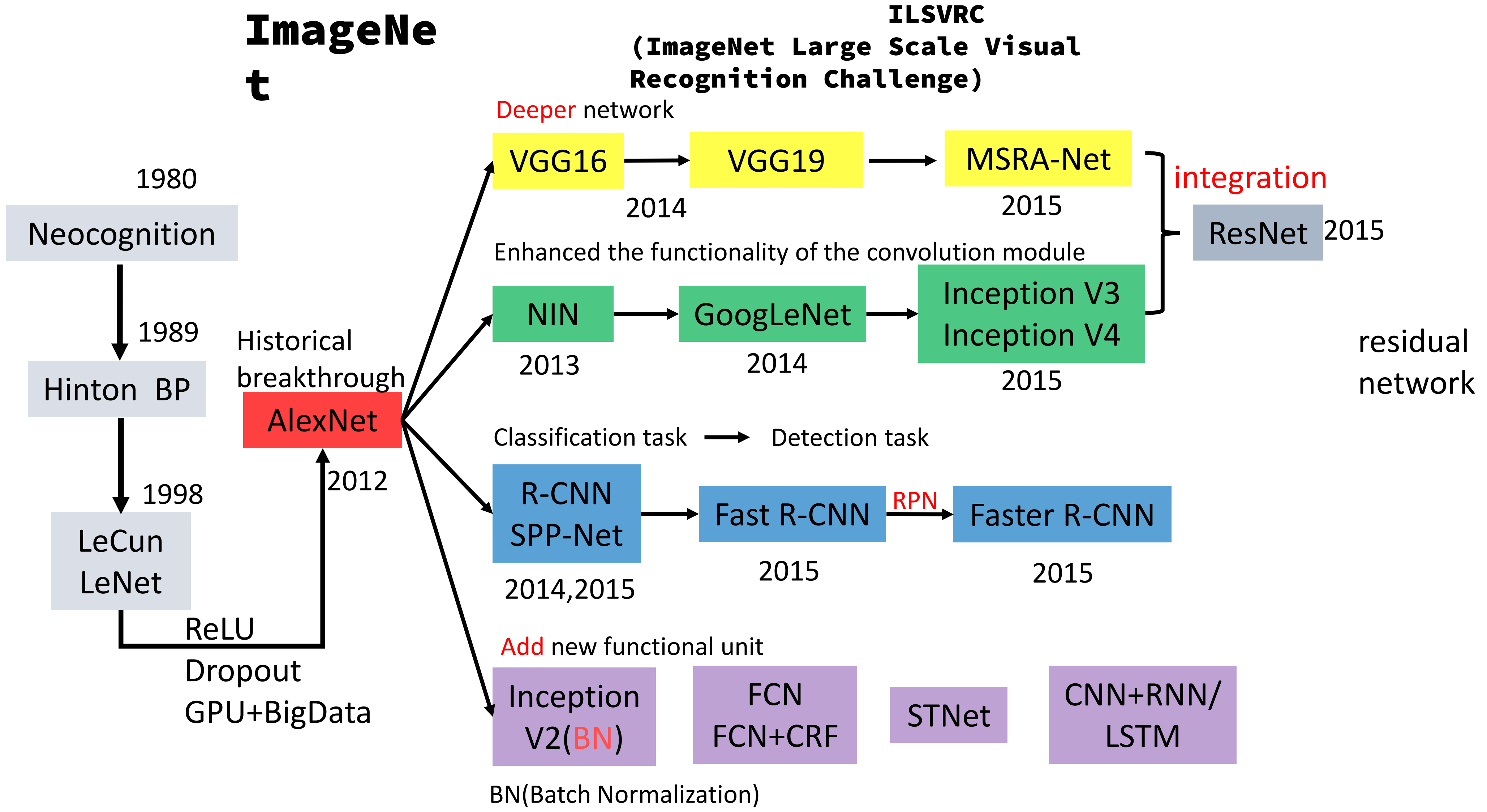
- ✓ The convolution kernel translates on a 2-dimensional plane, and each element of the convolution kernel is multiplied by the element at the corresponding position of the convolution image and then sum all the product.
- ✓ By moving the convolution kernel, we have a new image, which consists of the sum of the product of the convolution kernel at each position.



local receptive field
weight sharing

Reduced the number of parameters

CNN Structure Evolution



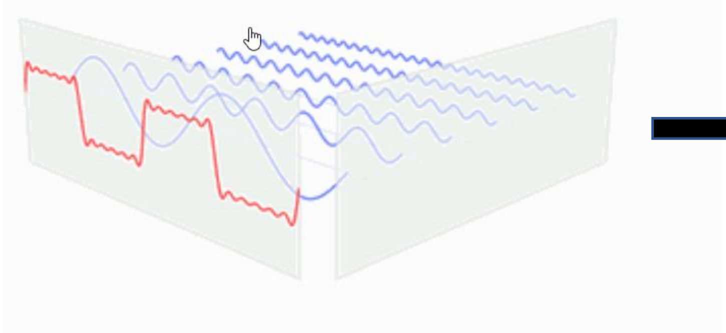
Introduction & Background

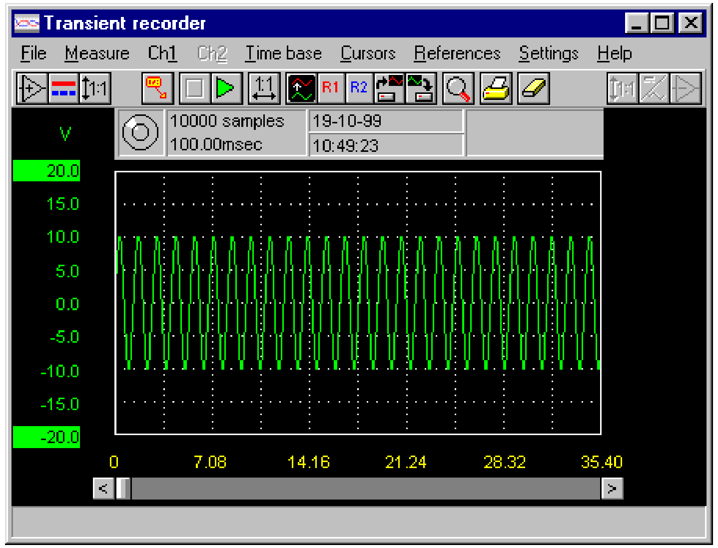
- Why should it be a Convolutional neural networks?

8

Why it should be convolutional neural networks?

In the field of beam diagnostics, a large amount of information extraction work is done in the frequency domain. Information at different frequencies contains different physical meanings.

convolutional kernel →  → filter

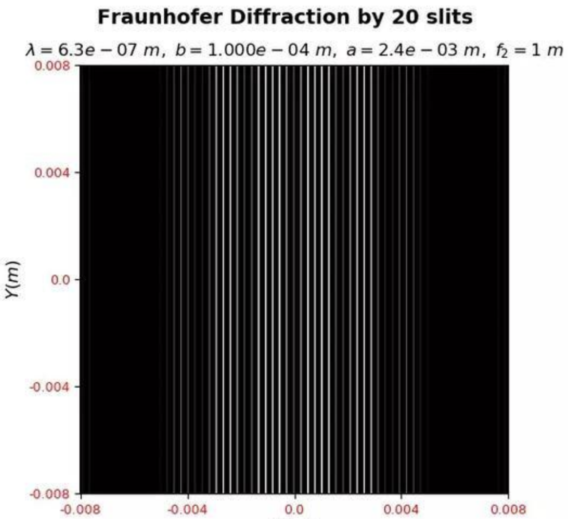


electric signal

a convolutional neural network

Just like

An intelligent filter bank



Optical signal

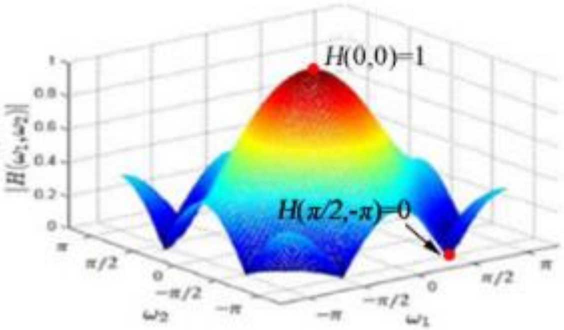
Example for CNNs (kernel) in the Frequency Domain

Low pass
convolutional kernel

0	1/6	0
1/6	1/3	1/6
0	1/6	0

Time domain

Fourier transform



Convolution(filter)

low pass

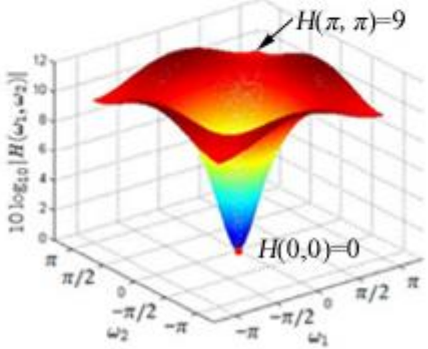
Time domain convolution = frequency domain multiplication

High pass
convolutional kernel

-1	-1	-1
-1	9	-1
-1	-1	-1

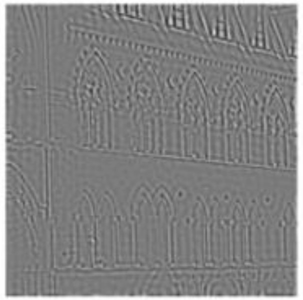
Time domain

Fourier transform



Convolution(filter)

high pass



By Training, CNNs **intelligently** learn to analyze data from different frequency domains and predict physical parameters.

What we want to obtain
Beam diagnostics

Application in virtual beam diagnostics

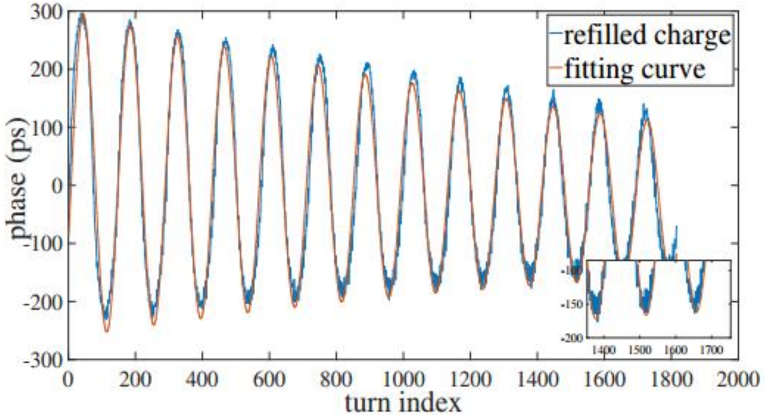
- deep learning image processing technology and beam parameter extraction from electrical signal
- design accelerator parameter optimizer based on convolution kernel

11

Application in virtual beam diagnostics

- deep learning image processing technology and beam parameter extraction from electrical signal

Q: What we want?



A: parameters of the synchrotron damping oscillation

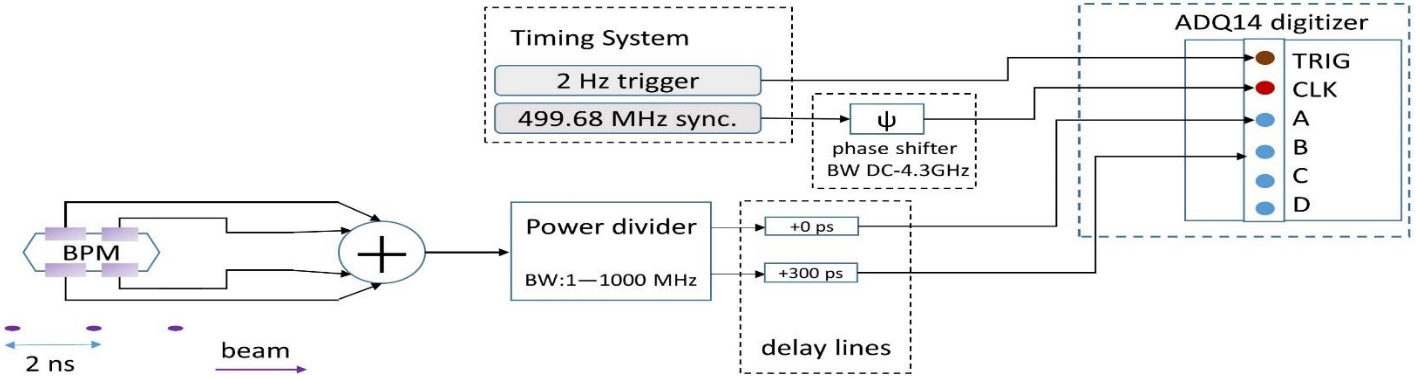
$$z_d = z_m \sin\left(\sqrt{\Omega^2 - \alpha_s^2}\, t + \varphi_0\right) e^{-\alpha_s t}.$$

z_m : *Oscillation* amplitude

αs : damping parameter

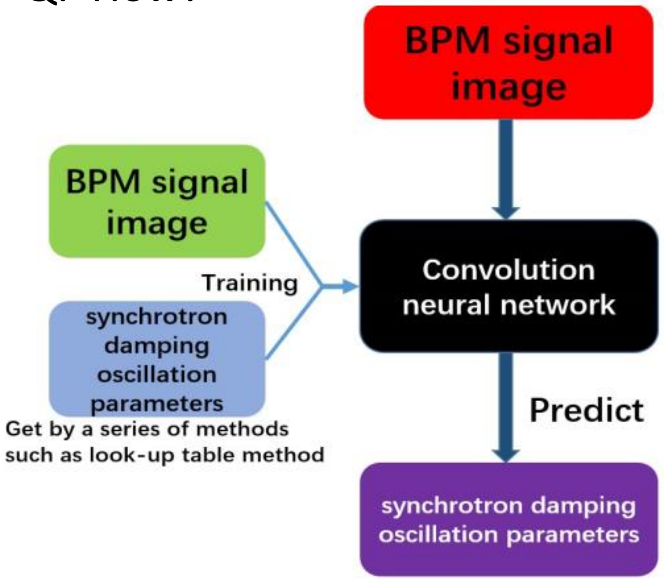
Q: what data is used to calculate?

A: raw electric signal from BPM pickup

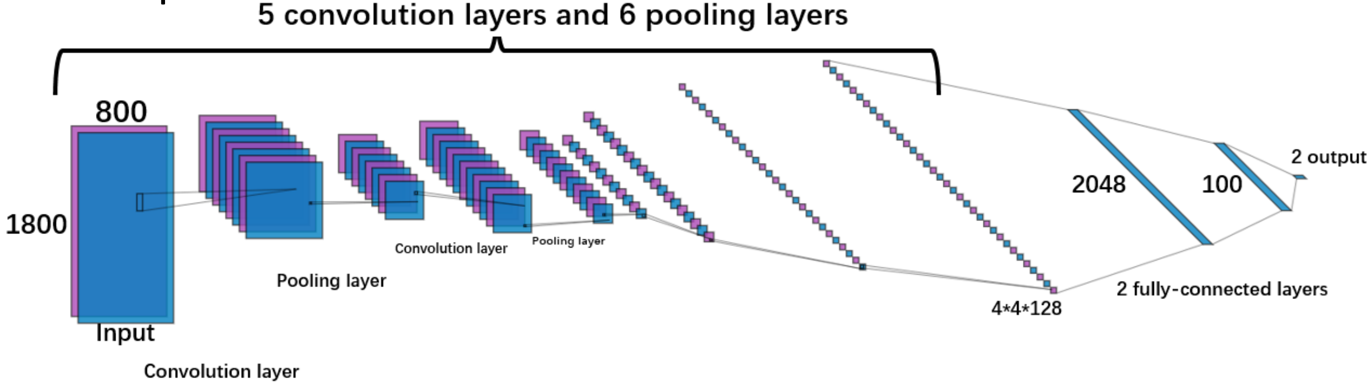


deep learning image processing technology and beam parameter extraction from electrical signal

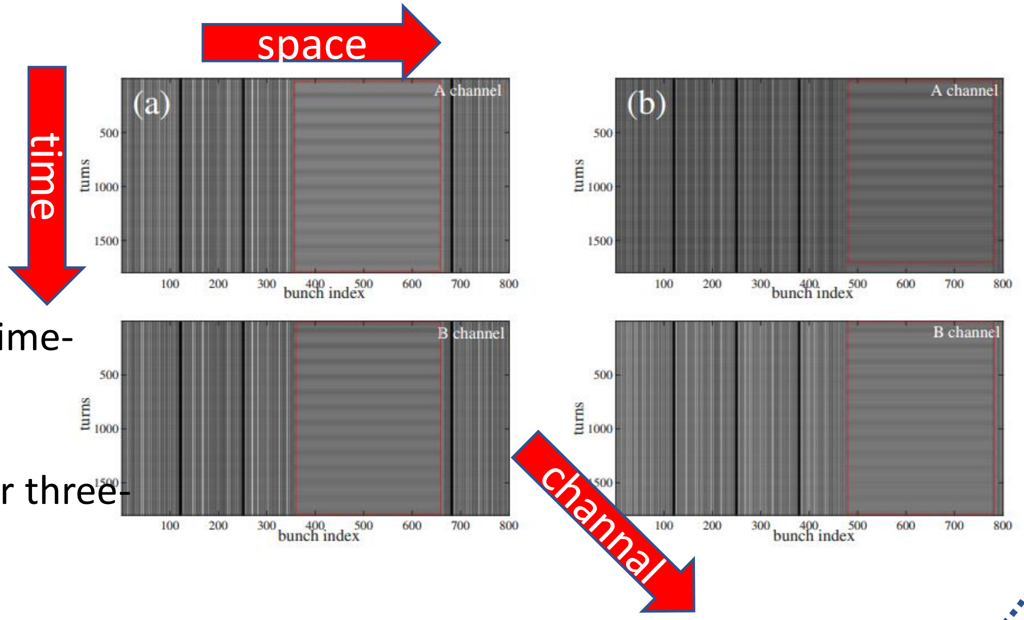
Q: How?



A: a deep convolutional neural network



3-D input data block



3-D convolution kernel

Input data : raw electric signal from BPM pickup

The one-dimensional time series signal is reshaped into a two-dimensional time-space signal

The two-channel electrical signal is processed in a way that imitates the color three-channel picture.

The high information density area is enhanced to increase the visual size

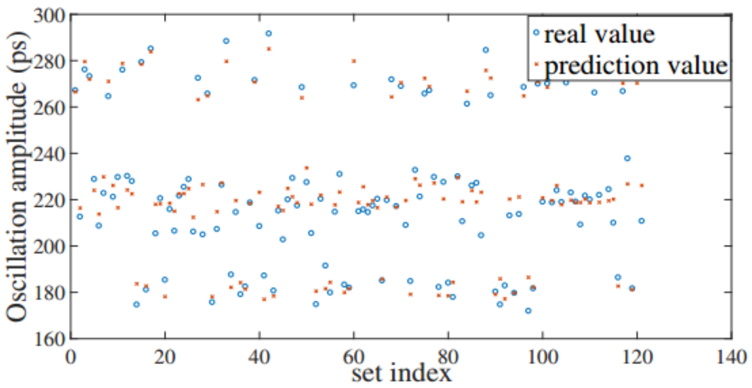
Q: performance? A: fine

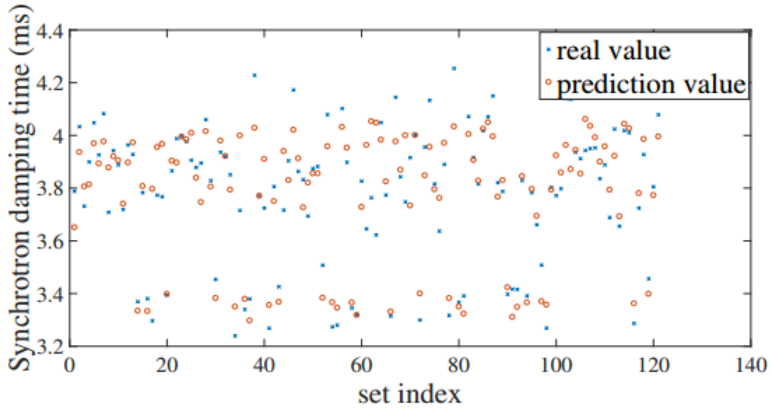
- **Faster** : each calculation takes less than 0.1 seconds

- **Easier**(for FPGA) : After the training is completed, the calculation consists entirely of basic operations. High-speed online processing can be completed by FPGA.

- **more robust** : The machine learning model has the characteristics of smoothness and average. There will be no abnormal results caused by algorithm errors when using CNN, which occasionally occurs in traditional algorithms.

- **Acceptable accuracy**: 2%-4%





Application in virtual beam diagnostics

- design accelerator parameter optimizer based on convolution kernel

16

The contradiction between machine learning and diagnosis

Some people will doubt whether such a black box is reliable.

Supervised learning

How to get train data?

Why not calculate directly?

1. Faster
2. Destructive to Non-destructive
3. Manual to Automation

- Supervised learning algorithms build a mathematical model of a set of data that contains both the inputs and the desired outputs. ——training data

Is it reliable?

What calculations did ML do?

1. ML model is a black box
2. difficult to visually and clearly understand what calculations are done in the black box.

The main machine learning algorithm

The most suitable algorithm for diagnosis (prediction and decision-making)

What is wakefield ?

impedance & wakefield

- The electromagnetic fields generated by a particle moving through a vacuum chamber are usually described as wakefields.

- Wakefields generated by the head particles can act back on following particles modifying their dynamics and (potentially) driving **instabilities**.

- A long range wakefield causes coupling **multi-bunch instability.**

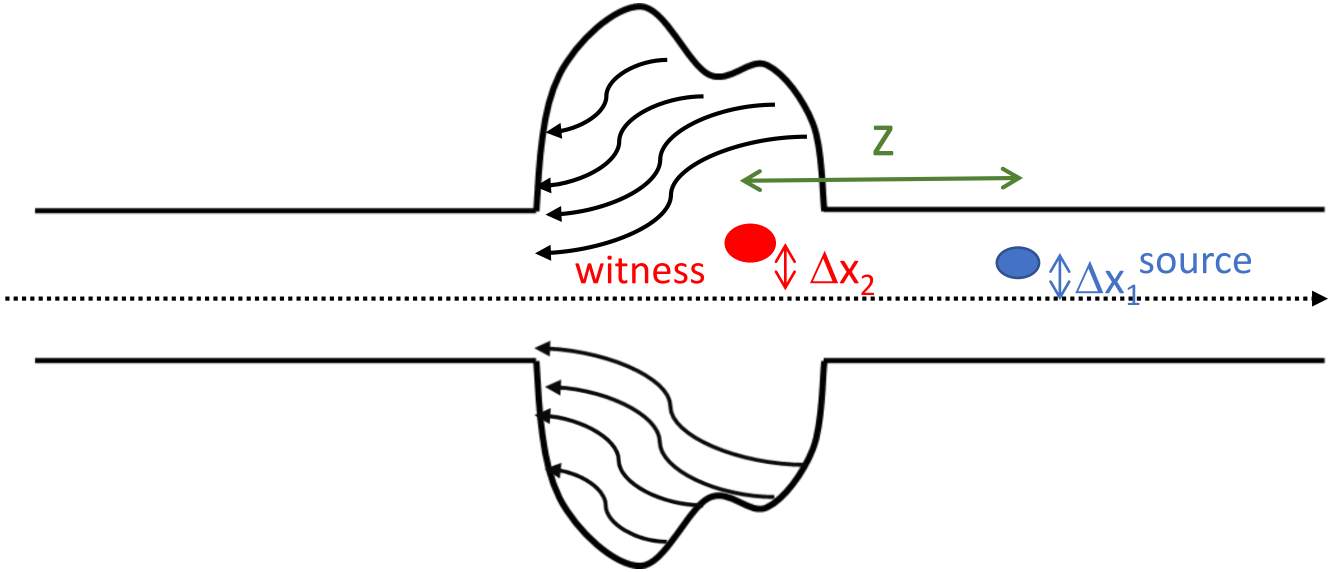
- A short range wakefield causes **single bunch instability**.

$$W_x(z) = -\frac{E_0}{q_1 q_2}\frac{\Delta x_2'}{\Delta x_1}$$

The transverse wakefield

$$W_{\|}(z) = -\frac{\Delta E_2}{q_1 q_2}$$

The longitudinal wakefield

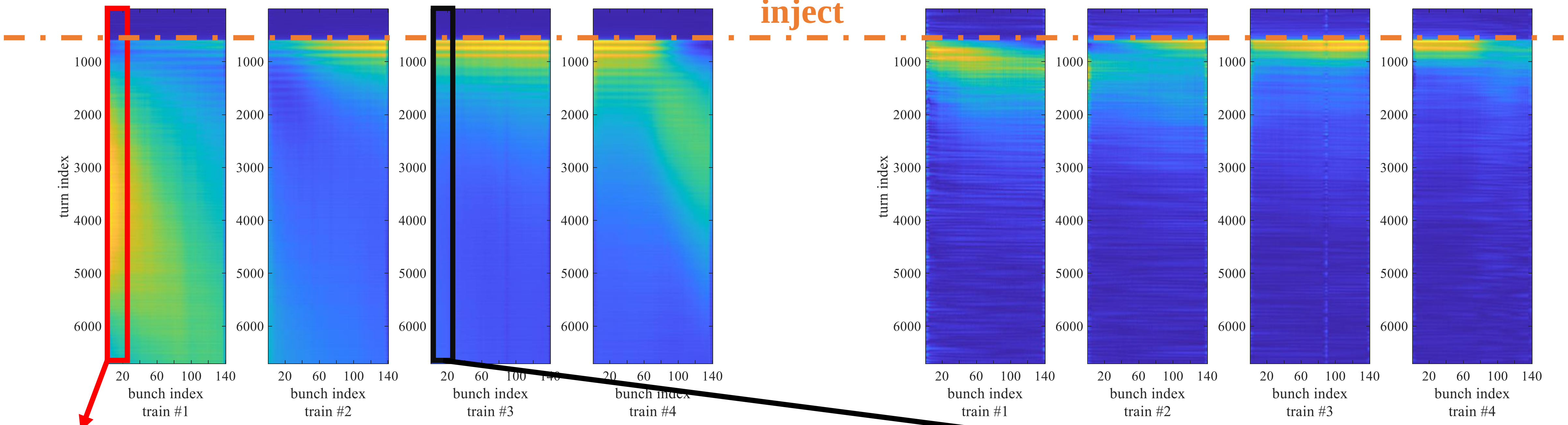


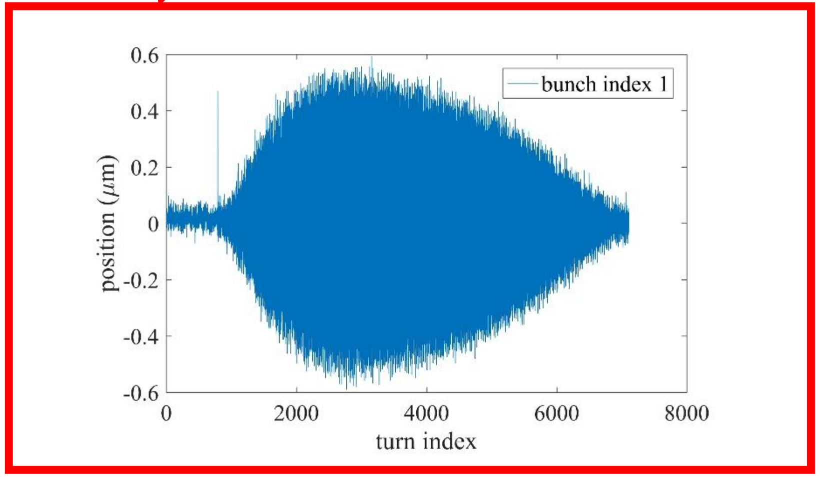
q_1, q_2: *charge*

Inject transient

Transverse position x-axis

Transverse position y-axis



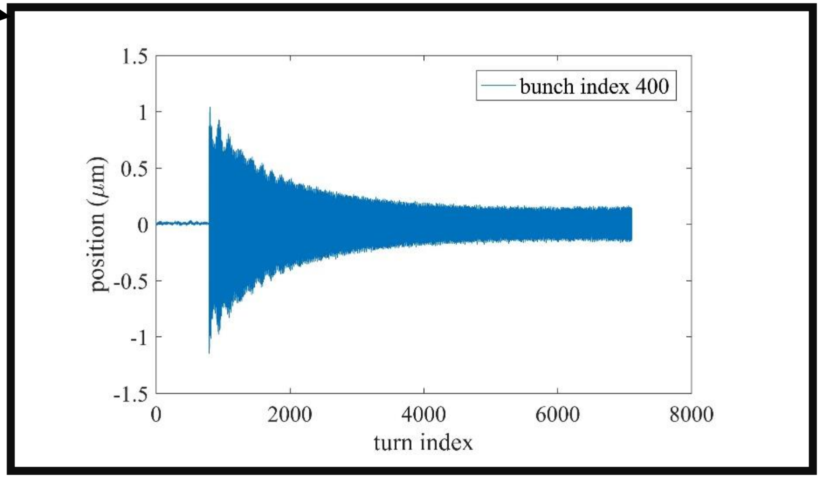


kichers field

- deviate from orbit (very short action time)
- influence on the bunches were not constant

Transverse coupled wakefield

- Transmission of oscillation between bunches
- Influence on oscillation evolution



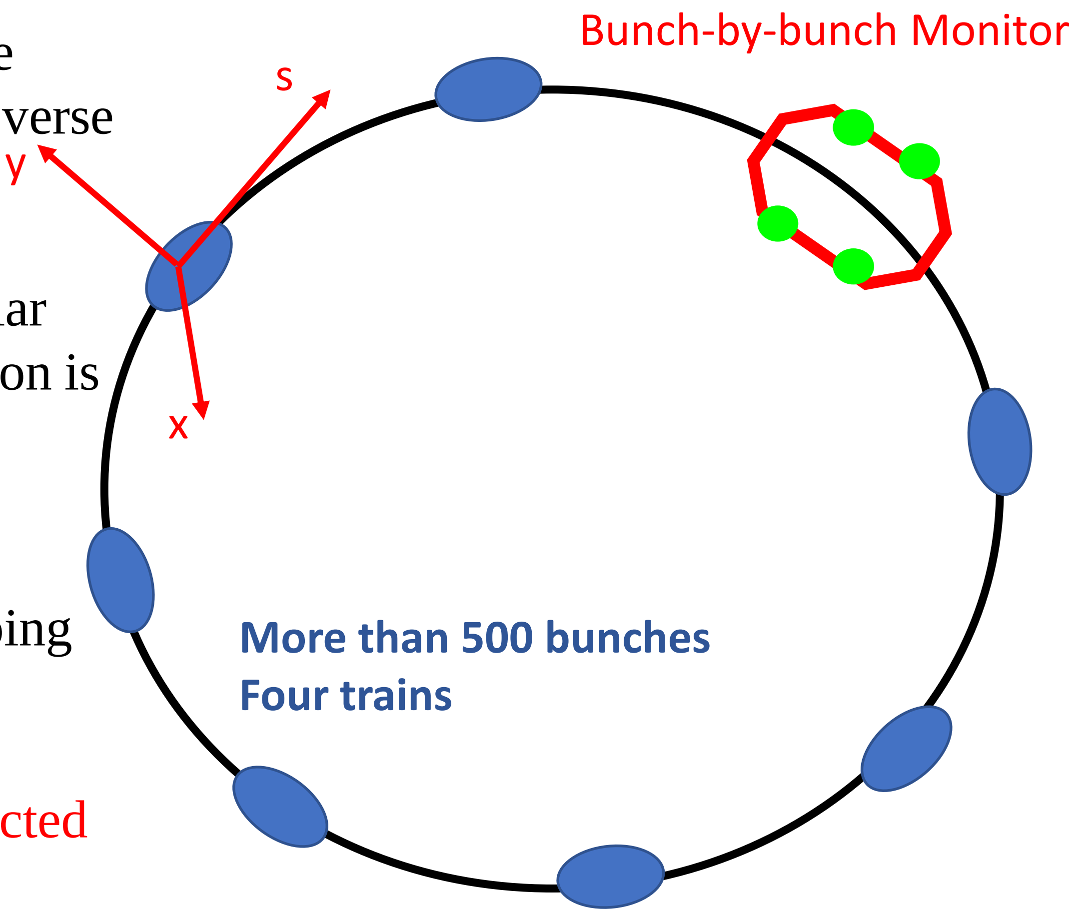
Transverse Equivalent Coupled Wakefield for the Whole Storage Ring

- **Transverse equivalent dipolar wakefield**: The influence of the transverse oscillation amplitude of the source bunch on the transverse oscillation amplitude of the witness bunch.

- **Simplified model**: The betatron oscillation phase and quadrupolar wakefield is not considered. Equivalent dipolar wakefield function is consistent for every source.

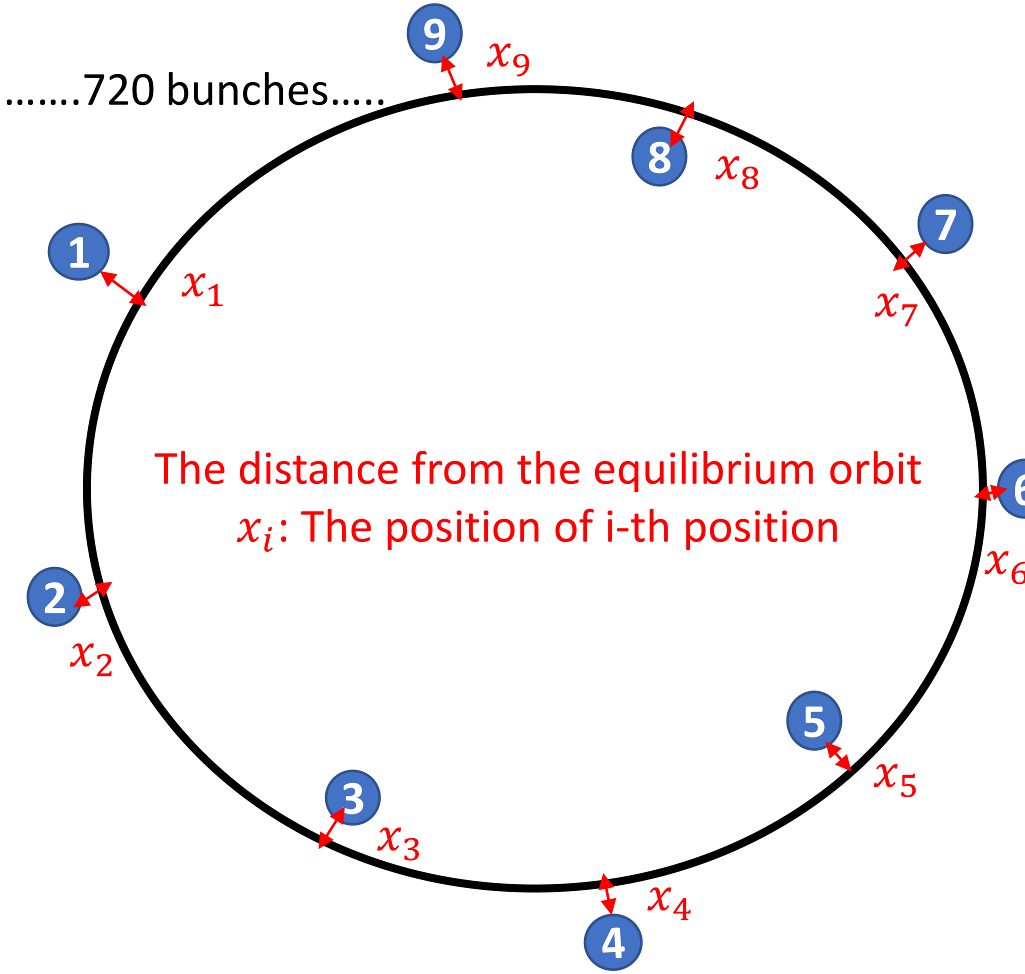
- **Target:** Predict the evolution trend of the transverse oscillation amplitude of each bunch under the combined action of the damping term and the wakefield.

- **Verification:** costfunction ——the difference between the predicted trend and the measured trend.



Modeling

turn \ bunch index	1	2	3	……	719	720
1	$x_1(1)$	$x_2(1)$	$x_3(1)$	……	$x_{719}(1)$	$x_{720}(1)$
2	$x_1(2)$	$x_2(2)$	$x_3(2)$	……	$x_{719}(2)$	$x_{720}(2)$
3	$x_1(3)$	$x_2(3)$	$x_3(3)$	……	$x_{719}(3)$	$x_{720}(3)$
……	……	……	……	……	……	……
N	$x_1(N)$	$x_2(N)$	$x_3(N)$	……	$x_{719}(N)$	$x_{720}(N)$



- **Coupled wakefield drive**: All bunches will produce a kick for the transverse oscillations of all bunches (including themselves) in each turn.

$$W_x(\mathrm{i}-j) = \frac{\Delta x_j}{q_i q_j} \cdot x_i$$

- **Damping term:** Landau Damping, Synchrotron Radiation Damping , Transverse feedback damping, etc.
The damping coefficient is considered to be a constant value.

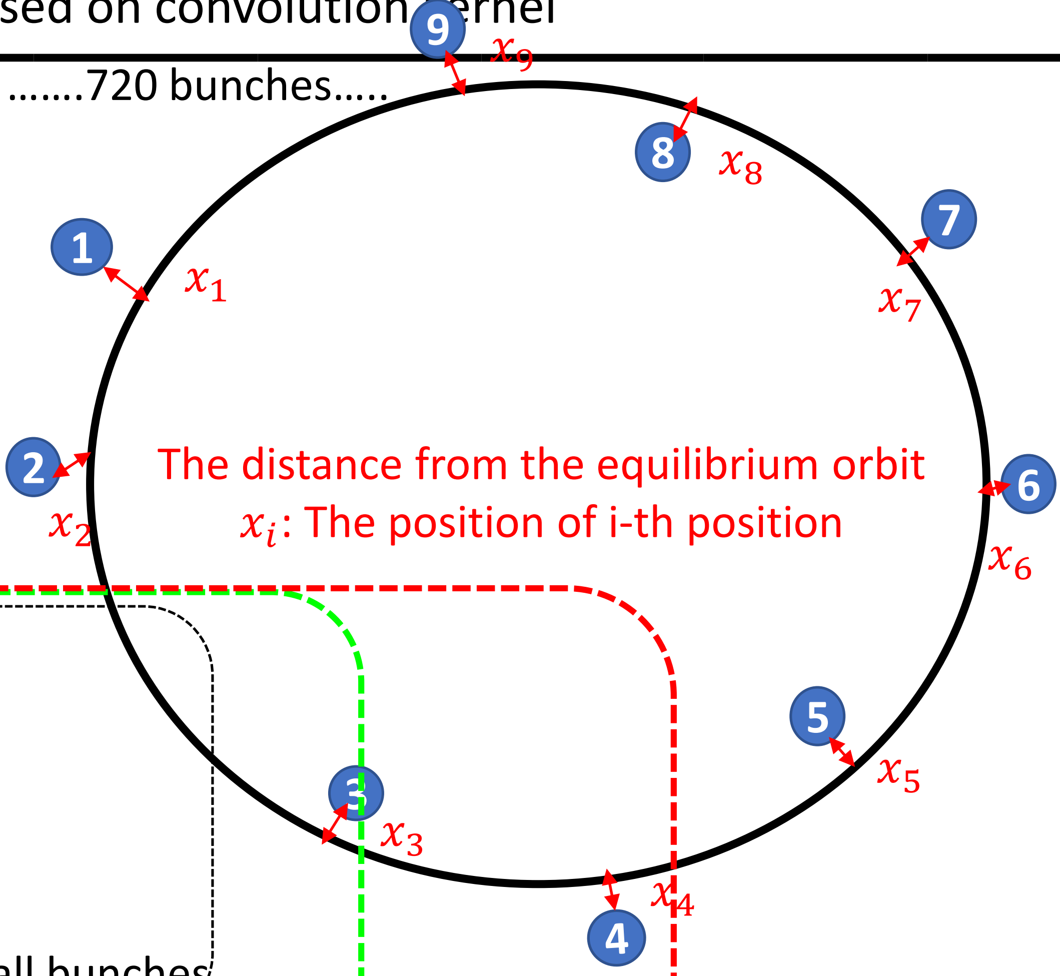
x(n+1) = x(n)*(1-η) + ΔX_{wf}

η : damping parameter
ΔX_{wf} : coupled wakefield drive

The iterative multi-turns transverse oscillation amplitude evolution trend speculate

design accelerator parameter optimizer based on convolution kernel

turn \ bunch index	1	2	3	……	719	720
1	$x_1(1)$	$x_2(1)$	$x_3(1)$	……	$x_{719}(1)$	$x_{720}(1)$
2	$x_1(2)$	$x_2(2)$	$x_3(2)$	……	$x_{719}(2)$	$x_{720}(2)$
3	$x_1(3)$	$x_2(3)$	$x_3(3)$	……	$x_{719}(3)$	$x_{720}(3)$
……	……	……	……	……	……	……
N	$x_1(N)$	$x_2(N)$	$x_3(N)$	……	$x_{719}(N)$	$x_{720}(N)$



position change of bunch j-th in (N+1)th turn:

$$\Delta X_j(N+1) = X_i(N+1) - X_i(N)$$

$$X(n+1) = X(n)*(1-\eta) + \Delta X_{wf}$$

$$\Delta X_{wf\_j}(N) = \sum_i W(j-i) \cdot X_i(N) \longleftarrow$$ Perform circular convolution on all bunches.

Repeat the above(black box) process for each bunch to get the position of all bunches in the next turn.

Repeat the above(green box) process for each turn to get transverse oscillation amplitude evolution trend.

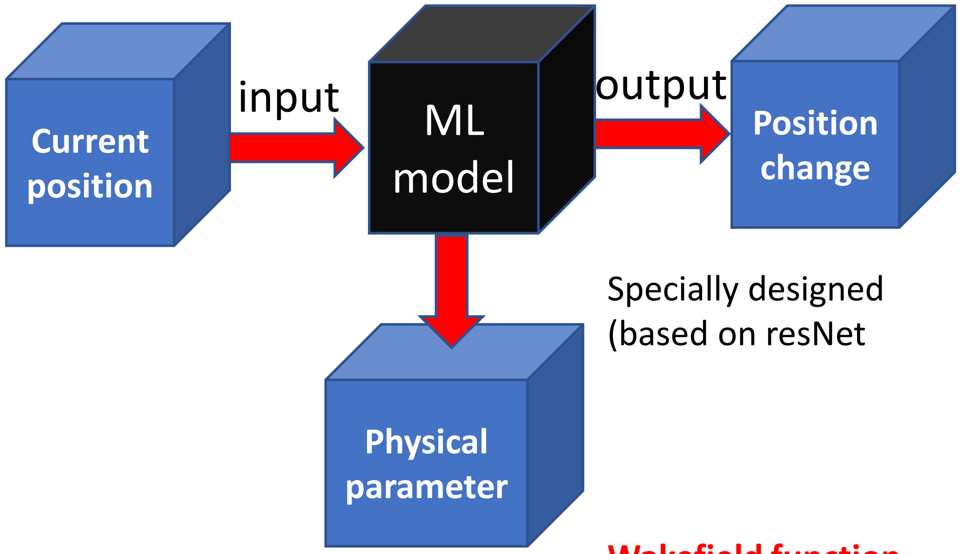
By designing a suitable exclusive network to restore the wakefield mechanism.

design accelerator parameter optimizer based on convolution kernel

$\Delta X_{wf\_j}(\mathrm{N}) = \sum_{\mathrm{i}} W(j-\mathrm{i}) \cdot X_{\mathrm{i}}(N)$ Perform circular convolution on all bunches.

$\Delta X_i(j) \Rightarrow$ just like residual

Design A exclusive ResNet to "simulate" this evolution trend calculation;



Specially designed (based on resNet

Wakefield function (for me)

- What is actually needed is the physical parameters hidden in the black box
- Hybrid neural networks in this project are **no longer black boxes**.
- The wakefield drive value is actually the parameter of the convolution kernel.

Cost function: the difference between the predicted trend and the measured trend.

The tensor data stream inside this residual convolutional neural network

turn \ bunch index	1	2	3	......	719	720
N	x_1(N)	x_2(N)	x_3(N)	......	x_{719}(N)	x_{720}(N)

Circular shift

1	2	3	......	719	720
x_{720}(N)	x_{719}(N)	x_{718}(N)	......	x_2(N)	x_1(N)
x_1(N)	x_{720}(N)	x_{719}(N)	......	x_3(N)	x_2(N)
x_2(N)	x_1(N)	x_{720}(N)	......	x_4(N)	x_3(N)
......	......	......	......	......	......
x_3(N)	x_2(N)	x_1(N)	......	x_5(N)	x_4(N)

Sum (multiply damping parameter)

Convolution

N	Δx_1(N)	Δx_2(N)	Δx_3(N)	......	Δx_{719}(N)	Δx_{720}(N)

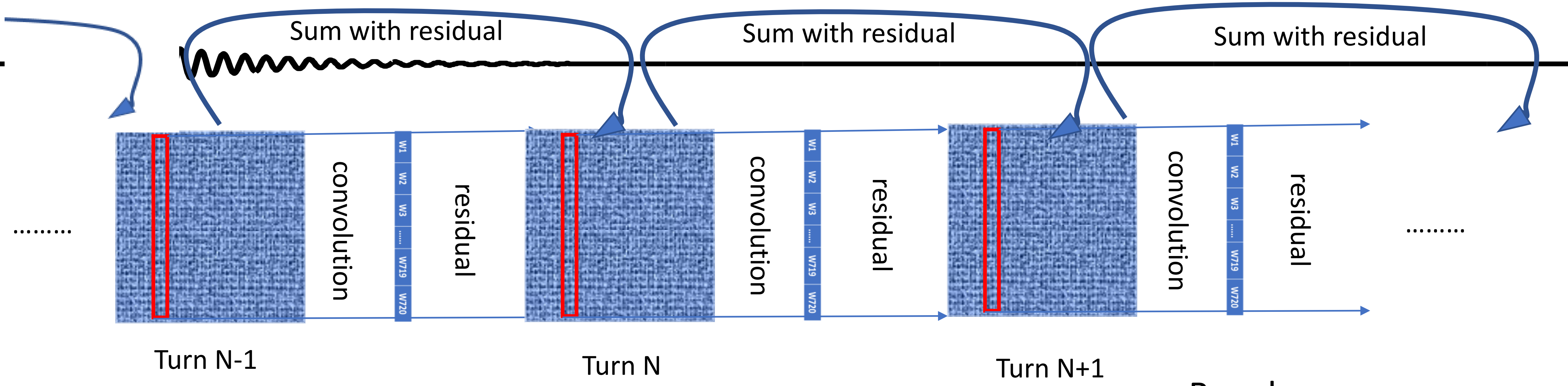
$$X_i(N+1) = X_i(N) + \Delta X_j(N+1)$$

turn \ bunch index	1	2	3	......	719	720
N+1	x_1(N+1)	x_2(N+1)	x_3(N+1)	......	x_{719}(N+1)	x_{720}(N+1)

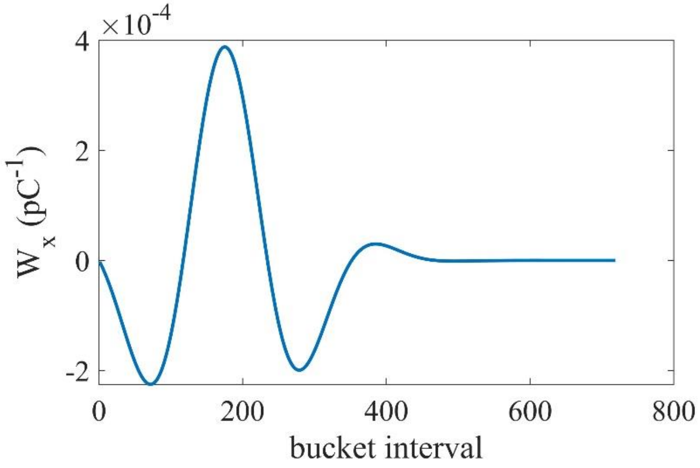
Set initial value and training

W1	W2	W3	......	W719	W720

Convolution kernel(**Coupled wakefield drive)**

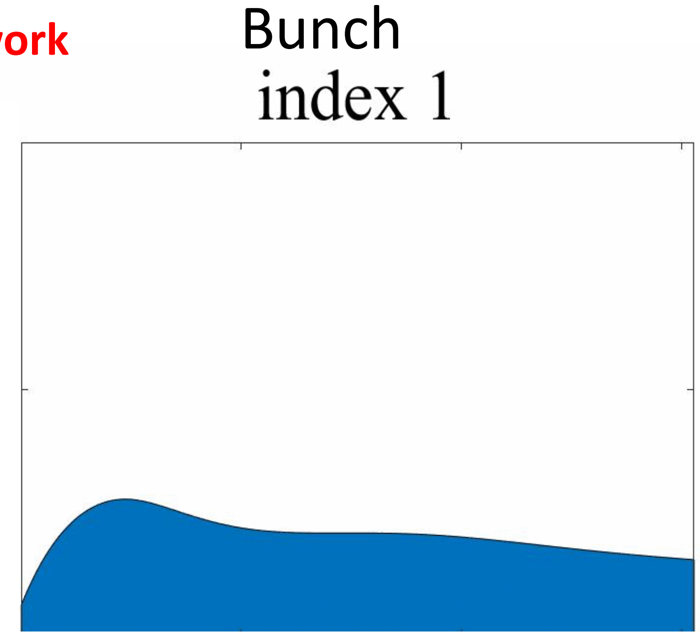


Schematic diagram of this special convolutional neural network



Coupled wakefield drive obtained by training

- Bunches that are 180 buckets away have the most obvious influence.
- the transmission of oscillating energy



The predicted trend of transverse oscillation amplitude(x).

Summary & Future

26

Summary & Future

1. Convolutional neural networks is a powerful multi-dimensional data processing tool.

2. In addition to the field of computer vision, convolutional neural networks may also be suitable for beam virtual diagnostic.

3. The SSRF Beam Instrument Group has done some simple tentative work on the application of convolutional neural network in virtual diagnostic.

4. In the future, Deep learning technology will play an increasing role in virtual beam diagnostics.

Thanks for your attention

Contact: xuxingyi@sinap.ac.cn
lengyongbin@sinap.ac.cn (my professor)

28