

# Detection of Keypoint in Press-Fit Curve Based on Convolutional Neural Network

Shoujia Fang, Guoqing Ding, Xin Chen

**Abstract**—The quality of press-fit assembly is closely related to reliability and safety of product. The paper proposed a keypoint detection method based on convolutional neural network to improve the accuracy of keypoint detection in press-fit curve. It would provide an auxiliary basis for judging quality of press-fit assembly. The press-fit curve is a curve of press-fit force and displacement. Both force data and distance data are time-series data. Therefore, one-dimensional convolutional neural network is used to process the press-fit curve. After the obtained press-fit data is filtered, the multi-layer one-dimensional convolutional neural network is used to perform the automatic learning of press-fit curve features, and then sent to the multi-layer perceptron to finally output keypoint of the curve. We used the data of press-fit assembly equipment in the actual production process to train CNN model, and we used different data from the same equipment to evaluate the performance of detection. Compared with the existing research result, the performance of detection was significantly improved. This method can provide a reliable basis for the judgment of press-fit quality.

**Keywords**—Keypoint detection, curve feature, convolutional neural network, press-fit assembly.

## I. INTRODUCTION

**P**RESS-FIT is widely used in automation and manufacturing. The quality of press-fit is important, especially in the automotive and aerospace fields, as it relates to the quality of the relevant products, which can cause safety problems and even more serious problems if not reliable enough. However, it has been very difficult to predict the pressing quality of the press-fit assembly by press-fitting process analysis [1].

There are currently two common methods for judging the quality of press-fit. They are finite element analysis (FEA) and judgment based on the press-fit curve. A new analytical method was established based on the TCT and resistant force calculation method [2]. Wang et al. Use this method to establish a simplified model, and derive the resistance calculation method, and finally establish an analysis method based on TCT and resistance calculation method. The judgment based on the press-fit curve generally uses three

Shoujia Fang is with Department of Instrument Science and Engineering, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, P. R. China (e-mail: fsj168@sjtu.edu.cn).

Guoqing Ding is an Associate Professor in Department of Instrument Science and Engineering, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, P. R. China (corresponding author, phone: +86-21-34204624, e-mail: gqding@sjtu.edu.cn).

Xin Chen is an Associate Professor in Department of Instrument Science and Engineering, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, P. R. China (e-mail: xchen.ie@sjtu.edu.cn).

methods, including key points, entry and exit frames, and envelopes.

The key point is the most important because it is related to the stopping conditions of the press-fit equipment. The searching of key points is based on the method of deriving the press-fit curve. Tan et al. [3] used mathematical formulas to verify the rationality of using the second derivative to search for key points. He interpolated the data using cubic spline interpolation and then used the least squares method for derivation. From the perspective of the image as a whole, Tan [4] associated with the use of a corner detection algorithm based on grayscale changes. He first drew the data and generates the image, applying the Harris corner detection algorithm to find the location of the key points. Considering the computational efficiency, he compressed the image. However, the whole process was still a bit cumbersome. Therefore, we consider the method of deep learning using original data.

Due to the complexity of the situation, such as the differences in individual parts, there is a lot of interference in the factory, and it is impossible to get good results with a unified method or detection algorithm. The main problem of the traditional method is that the feature extraction requires manual participation. The feature extraction is a model-based process, and there is no unified method guidance for the process. Therefore, the features obtained after a lot of work are not achieved ideally. If data mining and deep learning can be performed based on samples, adaptive feature extraction directly on the press-fit data will inevitably reduce the design difficulty of the perturbation recognition algorithm. There are significant features and differences in the time domain that can be expressed in natural language. The waveform features described by natural language are in fact the complex abstraction and understanding of the waveform image of the human brain. If it can imitate the human brain, the mechanism and process of identifying the characteristics of the press-fit data can realize the adaptive extraction of the feature to a certain extent. The current convolutional neural network in deep learning theory is an algorithm that can directly learn from original images, and is widely used in face recognition, text recognition and target tracking. This paper attempts to learn the press-fit data samples using a convolutional neural network.

CNN has many applications. Time-related applications are commonly used in one-dimensional convolutional neural networks. 1D CNN has many applications in the detection feature points and classification of electrocardiogram, and has achieved good results [5]. CNN also has many applications in

face key points detection, and the result is also very good [6]. There are also many applications for 1D CNN in industrial production. The vibration data of the bearing housings is preprocessed through fast fourier transformation (FFT) and inputted into CNN to detect faults [7].

Considering that using the press-fit data to generate a picture which as the input of convolutional neural network requires a long time of training and judgment, it does not meet the real-time requirements in industrial production. Both force data and distance data are time-series data. Therefore, one-dimensional convolutional neural network is used to process the press-fit data.

## II. PRESS-FIT CURVE AND CNN

### A. Press-Fit Curve

Press-fit equipment is one of the most important production equipments in today's machinery manufacturing industry [8], widely used in the machinery manufacturing industry represented by the automotive industry. Since press-fit is the final process of product manufacturing, it can reflect many quality indicators of the product, including accuracy, reliability, life and performance. Therefore, the automated assembly in modern machinery manufacturing requires not only high press-fit accuracy, Moreover, it is required to be able to perform automatic quality inspection on the press-fit products, that is, to monitor the press-fit curve force in real time, and to judge the quality of the press-fit by curve analysis.

In the past, bearing press-fit was mainly based on empirical judgment of the press-fit result. After the appearance of the press-fit force-displacement measuring device, it was possible to achieve a more accurate press-fit result determination using the press-fit force-displacement curve. It is an important function of the press-fit system to analyze the data obtained by the data acquisition system to determine whether the press-fit is suitable. On the other hand, the automatic start and stop process has become a hot issue in the current stage of press automation. After the press-fit process exceeds the critical point, it will not continue to press in, but will maintain press-fit force for a period of time. For this reason, the judgment of the key points becomes the key to solve the problem of termination process [9]. There are some of press-fit curve shown as Fig. 1

The key point is the point at which the first derivative of the press-fit-displacement curve changes rapidly, which is a critical position. The human visual system recognizes that the pressing force changes slowly before this point, and the pressing force increases sharply after this point [10].

Humans can easily identify the press-fit points of the press-fit curve, but it is difficult to find a robust and highly accurate deterministic algorithm for the press-fit point detection. Therefore, we think of using deep learning.

### B. CNN

The convolutional neural network is a feedforward neural network. It is usually composed of alternating convolutional layers and pooling layers. The convolutional layer can capture the regional connection features in the input information,

and the principle of weight sharing is applied to make the model to be trained. The amount of parameters is greatly reduced; the pooling layer combines adjacent nodes into one to merge similar features, further reducing the amount of data trained. Reflects the characteristics of its local connections, weight sharing, and subsampling. These characteristics make the convolutional neural network have a certain degree of translation, scaling and distortion invariance, and fewer parameters, which makes the training efficiency faster. The back propagation algorithm is used to update the weight in the training process [5].

The CNN is composed of convolutional layer neurons and pooled layer neurons, wherein the convolution layer extracts local features by convolving all the parts of the input signal, and the neurons of the same feature map share weights; the pooling layer, also called The downsampling layer performs independent operations on each feature map, such as average pooling or maximum pooling. The pooling layer can effectively reduce the feature resolution and reduce the number of network parameters that need to be optimized. The basic structure of the convolutional neural network is stacked, and the output of the previous layer is used as the input of the latter layer, which constitutes a convolutional neural network and has the ability of deep learning [11]

1) *1D Convolutional Layer*: The press-fit force data to be processed in this paper are discrete time series. Therefore, one-dimensional convolution is used as the convolution layer to construct a one-dimensional convolutional neural network suitable for pressure data feature extraction. Given input press-fit force sequence  $p_t, t = 1, \dots, n$  and displacement sequence  $d_t, t = 1, \dots, n$ , and filters  $w_t, t = 1, \dots, n$ , the filter sequentially performs a local convolution operation on the input features of the previous layer. In general, the length  $m$  of the filter is much smaller than the length  $n$  of the signal sequence. The output of the convolution is:

$$y_{t1} = \sum_{k=1}^m w_k * p_{t-k+1}$$

and

$$y_{t2} = \sum_{k=1}^m w_k * d_{t-k+1}$$

In the convolutional layer, each neuron in the  $l_{th}$  layer is connected only to neurons in a partial window of the  $l-1_{th}$  layer to form a local connection network. The convolutional layer requires an activation function  $f(x)$  for nonlinear feature mapping, then the input of the  $i$ -th neuron in the first layer is defined as:

$$a_i^l = f\left(\sum_{j=1}^m w_j^l * a_{i-j+m}^{l-1} + b^l\right) = f\left(w^l * a_{(i+m-1):i}^{l-1} + b^i\right)$$

where  $w^l \in R^m$  is the  $m$ -dimensional filter,  $w^l$  to all neurons in the convolutional layer are identical.  $a_{(i+m-1):i}^l = [a_{i+m-1}^l, \dots, a_i^l]^T$ ,  $b^i$  is the offset parameter,  $i = 1, \dots, n$ .

2) *1D Pooling Layer*: The operation of the pooling layer is also a feature obtained by some way from a region. The common pooling method is to take the maximum or

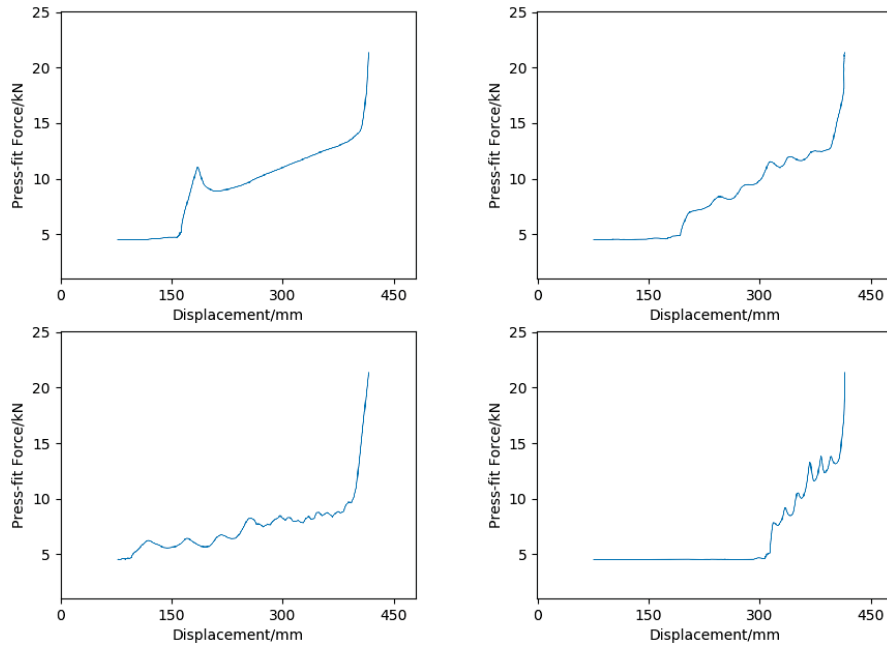


Fig. 1 Press-fit curves (X axis represents displacement, and Y axis represents press-fit force)

average value of all neurons in the region. For a feature map  $X_l$  obtained by the convolutional layer, it is divided into a plurality of regions  $R_k, k = 1, \dots, K$ .

$$pool_{max}(R_k) = \max_{i \in R_k} a_j$$

$$pool_{avg}(R_k) = \frac{1}{|R_k|} \sum_{i \in R_k} a_i$$

Pooling is a self-sampling process that greatly reduces the number of features, avoids overfitting, and allows the next layer of neurons to remain invariant to small morphological changes, providing strong robustness.

3) *Activation Function*: The feature map output to the pooling layer uses a nonlinear function as an activation function to avoid the problem of insufficient expression ability of the neural network model. In traditional convolutional neural networks, saturated nonlinear functions such as sigmoid are usually used [12], but due to the slow convergence of saturated nonlinear functions and even gradient disappearance in the reverse propagation phase, the unsaturated nonlinear activation function Relu (rectified linear units) is chosen in this paper [13] as the activation function of CNN, as shown below:

$$f(x) = \max(0, x)$$

4) *Training Method*: This paper uses the error back-pass law to train convolutional neural networks. The training sample is continuously convolved, pooled, and activated from the input layer to output a sampled feature map of the sample at the output layer. Pass the error layer by layer from the output layer to the input layer according to the differential chain rule, and Correspondingly update the convolution kernel and offset coefficients of each layer by using a random gradient descent

algorithm [14]. Detailed derivation process can be found in the literature [15].

### III. CNN ARCHITECTURE

Since the press-fit force and displacement time-series data are temporally continuous, the 1D convolutional neural network method was adopted to build CNN model. By applying 1D convolution, some temporal features can be extracted from the data. The base model consists of two parts. The first part includes three 1D convolution layers with different kernel sizes, and a dropout layer. In between is a reshape layer, which has similar effect as a flatten layer. The second part includes a dense layer, a dropout layer and the classification layer.

Specifically, based on the structure of the convolutional neural network LeNet-5 [16], through a lot of attempts, the network structure shown in Fig. 2 is finally determined. The convolutional neural network contains a total of 3 convolutional layers, 3 pooling layers, and 2 fully connected layers. Each layer except the input layer needs to be trained for weights. The input layer is a 2500 feature map, Conv1, Conv2 and Conv3 represent convolutional layers, respectively containing 3, 6, 12 convolution kernels, and the convolution kernel sizes are 5 by 5, 5 by 5, 5 by 5 respectively. Pool1, Pool2, and Pool3 represent the one-dimensional pooling layer, which adopts the maximum pooling mode. The pooling window size is 2. In CNN, the resolution between layers and layers is decremented, but the number of feature planes contained in each layer is incremented, which helps to identify more abstract features. The final output of the convolutional neural network is a 2 by 1 vector, which represents the ratio of the predicted value of the key point to the maximum value of the sequence.

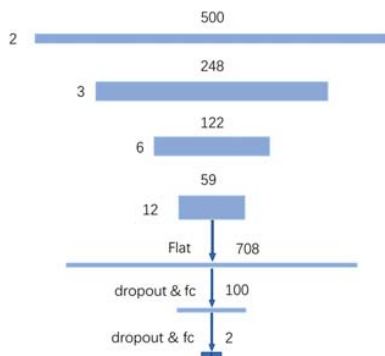


Fig. 2 Structure of 1D CNN for press-fit data

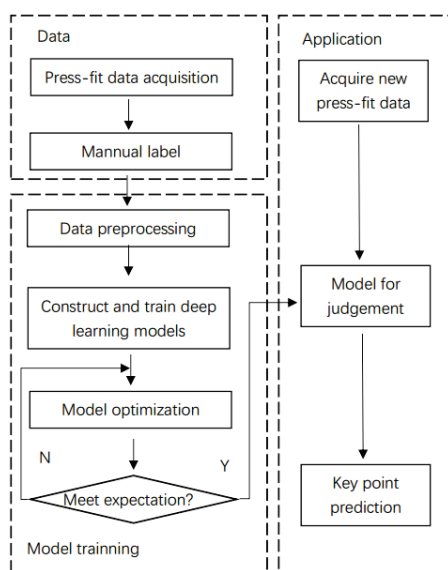


Fig. 3 System structure

The model is implemented on the Keras neural network framework with backend of TensorFlow.

#### IV. EXPERIMENT AND RESULT

The data comes from the automotive rear axle press-fit equipment in the actual factory.

From the factory's press-fit equipment, we obtained approximately 700 sets of press-fit data, each set of press-fit data including press-fit force data and displacement data. In order to make the features easier to learn, the press-fit data is pre-processed, including filtering and truncation. Because there is noise from different sources in the recorded press-fit data, we perform Gaussian filtering on the press-fit data. On the other hand, since the press-fit process generally includes three stages of approaching the workpiece, pressing into the workpiece, and retracting to the initial position, the most important data for judging the press-fit result is the stage of pressing into the workpiece. Therefore, we cut this part out to avoid the interference of useless information. At the same time, due to the small amount of data, we conducted data augmentation. Finally, 2000 sets of data were obtained.

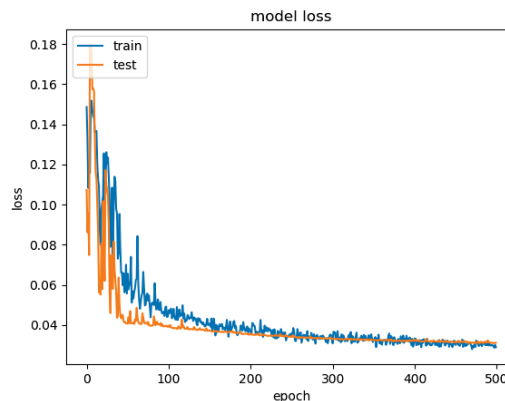


Fig. 4 Learning curve on training set and test set

Among them, 1600 groups were used as training sets, and the remaining 400 groups are used as test sets.

After the pre-processing is finished, each set of data is labeled, that is, the key points of each set of data are marked. Then the model is trained. The system structure can be seen as Fig. 3.

As we can see in Fig. 4, the error rate was higher at the beginning of training. After about 30 epochs of training, the error rate quickly decreased to less than 0.09. After 200 epochs, the error rate of the training set tended to be stable, about 0.03%, and the error rate of the training set is stable at around 0.04%. It can be seen from the error rate curve on the training set and the verification set that the trend and value of the two are basically the same, and the network has no over-fitting and under-fitting problems.

Some keypoint detection results can be seen in Fig. 5.

As for the computational efficiency, we tested 10,000 sets of data on the selected model and got results for 98.3 seconds. Therefore, the average time of per inference is 100ms. In actual production, we can get the press-fit data after the second phase of the press-fit process. The third phase takes about 3 seconds, which is enough to perform preprocessing and detect the key point. Overall, this result can meet the actual production time on the production line.

#### V. CONCLUSION

In this paper, a keypoint detection method for press-fit curves based on one-dimensional convolutional neural network is proposed. The morphological features of the press-fit curve can be automatically extracted through the convolutional layer, and the key points of the press-fit curve can be detected through the learning of the fully-connected layer. The experimental results show that the proposed method can achieve good detection performance and can be used as an auxiliary basis for the press-fit equipment to judge the quality of press-fit. Due to weight sharing of the convolutional neural network, the calculation speed is fast, making the model easy to transplant into the embedded motion controller. However, in practical applications, a certain type of press-fit data is selected as a specific data set, and the key points of the press-fit curve need to be manually labeled. This is a problem that needs to

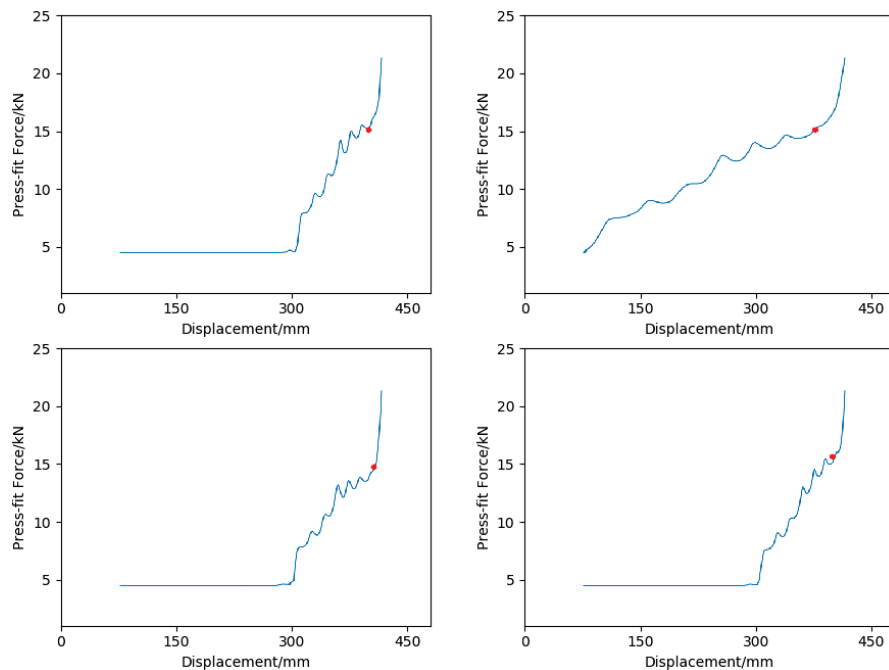


Fig. 5 Some results of key point detection based on CNN

be further studied and solved to realize the automatic detection of the key points of the press-fit curve.

#### REFERENCES

- [1] Bo You, Zhifeng Lou, Yi Luo, Yang Xu, and Xiaodong Wang. Prediction of Pressing Quality for Press-Fit Assembly Based on Press-Fit Curve and Maximum Press-Mounting Force, 2015.
- [2] Xingyuan Wang, Zhifeng Lou, Xiaodong Wang, and Chonglin Xu. A new analytical method for press-fit curve prediction of interference fitting parts. *Journal of Materials Processing Technology*, 250:16–24, December 2017.
- [3] Cong Tan, Guo-qing Ding, and Xin Chen. Algorithm in searching for the critical point of press-fit curve. In *ELECTRICAL ENGINEERING AND AUTOMATION: Proceedings of the International Conference on Electrical Engineering and Automation (EEA2016)*, pages 673–680. World Scientific, 2017.
- [4] Tan Cong. Research of ethercat-based auto parts distributed servo press system. Master's thesis, Shanghai JiaoTong University, 2016.
- [5] HUANG Jiao, BIN Guangyu, and WU Shuicai. Patient-specific ecg classification based on one-dimensional convolution neural network. *China Medical Devices*, (3):11–14, 2018.
- [6] Amarjot Singh, Devendra Patil, G Meghana Reddy, and SN Omkar. Disguised face identification (dfi) with facial keypoints using spatial fusion convolutional network. In *Computer Vision Workshop (ICCVW), 2017 IEEE International Conference on*, pages 1648–1655. IEEE, 2017.
- [7] Rui Zhao, Ruqiang Yan, Jinjiang Wang, and Kezhi Mao. Learning to monitor machine health with convolutional bi-directional lstm networks. *Sensors*, 17(2):273, 2017.
- [8] Daniel E Whitney. The potential for assembly modeling in product development and manufacturing. Technical report, In Proceedings of the 1995 IEEE International Symposium on Assembly and Task Planning, 1996.
- [9] Hu Hong-wei. A feasibility study of pressing force-offset curve as the judgment basis for acceptance of bearing press-fit. *LOCOMOTIVE & ROLLING STOCK TECHNOLOGY*, (5):7–10, 2010.
- [10] Yi Shanbo. *Research on the Pressing Force and Pressing Force Curve of Rolling Bearings for Railway Freight Cars*. PhD thesis, Central South University, 2005.
- [11] David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106–154, 1962.
- [12] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010.
- [13] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323, 2011.
- [14] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer, 2010.
- [15] Daniel Strigl, Klaus Kofler, and Stefan Podlipnig. Performance and scalability of gpu-based convolutional neural networks. In *Parallel, Distributed and Network-Based Processing (PDP), 2010 18th Euromicro International Conference on*, pages 317–324. IEEE, 2010.
- [16] R. Anil, K. Manjusha, S. Sachin Kumar, and K. P. Soman. Convolutional neural networks for the recognition of malayalam characters. In Suresh Chandra Satapathy, Bhabendra Narayan Biswal, Siba K. Udgata, and J. K. Mandal, editors, *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014*, pages 493–500. Cham, 2015. Springer International Publishing.

**Shoujia Fang** received his B.Sc. degree in 2016 from Hefei University of Technology. Now he is a Master in Shanghai Jiao Tong University. His main research interests include intelligent instrument.