Multi-input Convolutional Neural Network Fault Diagnosis Algorithm Based on the Hydraulic Pump

Lishan Zhang1,*, Lei Han1, Yuzhen Meng1 and Wenkui Zhao2
1 Beijing University of Posts and Telecommunications, Beijing, China
2 Aviation Key Laboratory of Science and Technology on Aero Electromechanical System Integration, Nanjing, China

*Corresponding author email: zlsgongzuo@126.com

Abstract. Convolutional neural network used in fault diagnosis can effectively extract fault features in vibration signals. However, in the feature extraction of mechanical fault diagnosis, usually more than two feature signals including at least axial and radial vibration signals can be extracted. This paper proposes two multi-input convolutional neural network models based on the fault data of the aircraft hydraulic pump including axial and radial vibration. The first is the Independent Input Multi-input Convolutional Neural Network model. The two inputs are respectively used for convolution pooling operation with CNN, and are combined through the concatenate function before the fully connected layer, and then all frames are integrated and flattened by the flatten function. A one-dimensional array, finally enters the fully connected layer and outputs the result through the softmax function. The second is the Combined Input Multi-input Convolutional Neural Network, that is, combine two one-dimensional signals into a two-dimensional signal in the input layer of the convolutional neural network and then perform convolution pooling, and finally output the result through the softmax function. The results show that the two models have good accuracy and stability, and the second one has a higher convergence and fitting efficiency than the first one.

Keywords: Fault diagnosis; Convolutional neural network; Multi-input.

1. Introduction
The aircraft hydraulic pump plays an important role in the operation of the aircraft. In the past, the traditional fault diagnosis of hydraulic pumps usually involves fault setting, signal collection, and comparison of the collected signals in time domain and frequency domain to find out the characteristic difference between normal and multiple faults, and pass More samples to judge the effectiveness and robustness of the method. This is a purely manual diagnosis method, which is inefficient. Since 2010, Deep learning has become the focus of machine learning research[1]. In recent years, great success in the field of pattern recognition have been done by convolutional neural networks. It can automatically mine features from signals and images. So convolutional neural networks began to be used in machine health detection. Since 2015, many researchers have begun to use deep learning models for bearing fault diagnosis. Taking the bearing data published by Case Western Reserve University in the United States as a sample, many researchers have proposed a variety of different convolutional neural network models for fault diagnosis experiments, verifying that the convolutional neural network has excellent performance in bearing fault diagnosis.

At present, the research on fault diagnosis of convolutional neural network is almost always a single input, that is, researchers only use the collected original time-domain signal as the input of the deep learning model for training. In 2018, Chenhua Ni et al. used the four parameters of the hydraulic...
data (hydraulic pressure, hydraulic flow, motor speed, motor torque) in the ocean power generation system as inputs to predict the amount of ocean power generation, and proposed to superimpose the four data to form a square matrix is then input into a multi-input model in a two-dimensional convolutional neural network [2]. And in 2019, Tao Zan et al. proposed a multi-input deep neural network model for bearing fault diagnosis. They converted the collected radial vibration time domain signals into frequency domain signals through Fourier transform, and then converted the frequency domain signals is the envelope spectrum, and then use the above time domain signal, frequency domain signal, and envelope spectrum as the three inputs of the convolutional neural network model for training [3]. This paper proposes two 2-input convolutional neural network models for fault diagnosis of aircraft hydraulic pumps. The two inputs are the radial and axial vibration time domain signals when the hydraulic pump is running. The sample data used for training are all self-collected data, and the radial signal and the axial signal are collected at the same time. We aims to discuss the advantages of the multi-input convolutional neural network model and other single-input models, and compare the similarities and differences between the two different input methods in terms of model performance.

2. Brief Introduction to CNN
This part will briefly introduce the structure of Convolutional Neural Networks (CNN). More details about Convolutional Neural Networks can be found in [4].

2.1. Convolutional Layer
The convolution layer convolutes the input local region through the kernel (convolution kernel), and then the activation unit generates the output feature. Each convolution kernel of the same convolution layer has the same size and is used to extract the features of the input local area. This is usually called weight sharing. A convolution kernel shall be applied to one frame of the next layer, and the number of frames represents the depth of the layer. We use $K_i^l$ and $b_i^l$ to respectively represent the weights and biases of the convolution kernel of the $i$-th layer, and use $X_j^l$ to represent the local area of the $j$-th layer, then the convolution process can be expressed as

$$y_j^{l+1}(j) = K_i^l \times X_j^l(j) + b_j^l \tag{1}$$

After the convolution process, the output of the convolution is "activated" through the function $f(?) = \max \{0, ?\}$ to accelerate the convergence of the convolutional neural network. The activation function usually used is ReLU.

2.2. Pooling Layer
The pooling layer is usually followed by the convolutional layer. It is a down-sampling operation that combines part of the features into one to reduce network parameters. The pooling layer includes maximum pooling, average pooling, overlapping pooling and so on. The most commonly used is max-pooling, which can reduce network parameters and remove redundant information while obtaining features that remain unchanged. The process of max-pooling can be expressed numerically as:

$$P_i^{l+1}(j) = \max_{(j-1)H+i, (j-1)H+i+1, \ldots, (j-1)H+i+W} \{Q_j^l(t)\} \tag{2}$$

Where $Q_j^l(t)$ is the value of the $t$-th neuron in the $i$-th frame of the $l$-th layer, $W$ and $H$ are the width and height of the pooling kernel, respectively, and $P_i^{l+1}(j)$ represents the corresponding value of the neuron in the next layer $l+1$ of the pooling layer.

2.3. Fully Connected Layer
After passing through multiple convolution layers and pooling layers, one or more full connection layers are connected, each neuron in the fully connected layer and all neurons in the previous layer fully connected. The fully connected layer can integrate the local information that is distinguished by
categories in the convolutional layer or the pooling layer. In order to increase the efficiency of the convolutional neural network, the activation function of each neuron in the fully connected layer generally adopts the ReLU function. The output value of the last layer from the fully connected layer is passed, and it would be classified by softmax logistic regression (softmax regression). This layer can also be called softmax layer [5]. The Softmax function is:

\[ (z_j) = \frac{e^{z_j}}{\sum_k e^{z_k}} \]  

Where \( z_j \) represents the logistic regression of the \( j \)-th output neuron.

3. Experiment and Analysis

3.1. Data Description
The original experimental data is the vibration signal of a certain type of aircraft hydraulic pump. By installing accelerometers in the axial and radial directions of the hydraulic pump housing, the time-domain signals of the axial and radial vibration acceleration of the hydraulic pump under various working conditions are collected, and the sampling frequency is 2kHz. Aircraft hydraulic pump failures can be basically divided into two types, namely, hydraball clearance increased, the swashplate misalign, and normal and no failure conditions. A total of three types of data have been collected.

In this experiment, because the collected data includes axial data and radial data, the axial data and the radial data have the same number of training samples and test samples. The set sample contains 2048 data points. The training sample data is extracted from the original data through the sample expansion method. Each sample offset is 256 data points. The number of training samples for each working condition is 460, and the total number of training samples is extracted. Is 1380. The test sample data extraction does not use sample expansion, there is no overlap between the test samples, 67 test sample data of each working condition are extracted, and the total number of test samples is 201. The experimental training test samples are shown in Table 1.

3.2. Experiment
The convolutional neural network models in this paper refer to the WDCNN model proposed by Zhang Wei in 2017 [6]. WDCNN uses the first layer of large convolution kernel to increase the receptive field of the convolutional neural network, and the tail adopts the design of multi-layer small convolution kernel, which can extract signal features in more detail. The batch normalization algorithm is also used in the WDCNN model to speed up the training speed.

This article intends to construct the convolutional neural network model through two different data input methods. The first is an Independent Input Multi-input Convolutional Neural Network (IIMI-CNN), and the second is a Combined Input Multi-input Convolutional Neural Network (CIMI-CNN).

3.2.1. Independent Input Multi-input Convolutional Neural Network. The so-called independent input is the axial and radial vibration data collected by the aircraft hydraulic pump, which are independently input into the convolutional neural network at the same time for convolution pooling, and then combined in the model, and finally output the result. The schematic diagram of the model structure is shown in Figure 1.
3.2.1.1. First generation model. In the past, the research on bearing fault diagnosis by researchers usually refers to collecting time-domain data using radial vibration. Since the experimental data contains axial data and radial data, we plan to design a 2-input convolutional neural network model that uses axial data and radial data as inputs for training at the same time. The CNN model built for the first time in the experiment is shown in Table 2. Taking WDCNN as the prototype, the two inputs are separately convolutional pooling operation in WDCNN, and the concatenate function is combined before the fully connected layer, that is, the axial and radial data samples After the final pooling layer, the corresponding frames are connected. Then use the flatten function to integrate and flatten all frames into a one-dimensional array. Finally, it enters the fully connected layer and outputs the result through the softmax function.

Table 2. Multi-input convolutional neural network first-generation model structure parameters.

| No. | Layer     | Kernel size/Step | Kernels number | Output size   | Padding |
|-----|-----------|------------------|----------------|---------------|---------|
| 1   | Conv 1(2) | 64*1/16*1        | 16             | 128*16        | same    |
| 2   | Pool 1(2) | 2*1/2*1          | 16             | 64*16         | valid   |
| 3   | Conv 3(4) | 3*1/1*1          | 32             | 64*32         | same    |
| 4   | Pool 3(4) | 2*1/2*1          | 32             | 32*32         | valid   |
| 5   | Conv 5(6) | 3*1/1*1          | 64             | 32*64         | same    |
| 6   | Pool5(6)  | 2*1/2*1          | 64             | 16*64         | valid   |
| 7   | Conv7(8)  | 3*1/1*1          | 64             | 16*64         | same    |
| 8   | Pool7(8)  | 2*1/2*1          | 64             | 8*64          | valid   |
| 9   | Conv9(10) | 3*1/1*1          | 64             | 6*64          | valid   |
| 10  | Pool9(10) | 2*1/2*1          | 64             | 3*64          | valid   |
| 11  | Concatenate|                 |                | 6*64          |         |
| 12  | Fully Connect | 100              | 1              | 100           |         |
| 13  | Softmax   | 3                | 1              | 3             |         |

In the model, the batch-normalization layer is added before the first convolutional layer, before the second convolutional layer, and after the fifth pooling. To prevent overfitting, a Dropout layer is added after the flatten layer after the combination of axial and radial data. During the training process, Adam stochastic optimization algorithm is used to train our model. Use the sparse_categorical_crossentropy loss function to calculate the loss value, each training 40 epoch, Batch is 32.

3.2.1.2. Training results and analysis of the first generation model. After many trials, the results of the experiment are shown in Figure 2. The three figures represent 3 trials respectively. The blue line is the
self-correct rate of the training set, the yellow line is the correct rate of the test set, and the x-axis is the epochs.

Figure 2. The training accuracy of the first generation model.

It can be seen from the figure that the model can usually reach a peak at the 12th cycle of training, and sometimes has 100% accuracy at the peak, but as the number of training cycles increases, the model will appear very severe over-fitting phenomenon causes the model's training effect to be very unstable. After analysis, the model may have three over-fitting phenomena. One is that the previous axial and radial data are convolved too many times, and the second is that the width of the first layer of convolution is not enough, which leads to insufficient receptive fields and signals. Feature extraction is unstable. Third, the step size of the pooling layer and the width of the relative pooling kernel are large, which causes the pooling to lose valuable features.

It can be seen from the peak accuracy rate that can reach 100% that the 2-input convolutional neural network combined by the axial and radial directions can improve the accuracy of fault diagnosis. Therefore, modify the structure of the first-generation convolutional neural network model, change the width of the first-layer convolution kernel from 64 to 128, remove the convolutional pooling layer of the last three layers, remove the middle layer of the three BN layers, and then pool. The step length of the chemical layer is shortened. The modified model structure parameters are shown in Table 3.

| No. | Layer   | Kernel size/Step | Kernels number | Output size | Padding |
|-----|---------|------------------|----------------|-------------|---------|
| 1   | Conv1(2)| 128*1/16*1       | 16             | 128*16      | same    |
| 2   | Pool1(2)| 2*1/1*1          | 16             | 64*16       | valid   |
| 3   | Conv3(4)| 3*1/1*1          | 32             | 64*32       | same    |
| 4   | Pool3(4)| 2*1/1*1          | 32             | 32*32       | valid   |
| 5   | Concatenate |               |                | 64*32       |         |
| 6   | Fully Connect | 100             | 1              | 100         |         |
| 7   | Softmax | 3                | 1              | 3           |         |

Figure 3. The training accuracy of the second generation model.

3.2.1.3. Results and analysis of second-generation model training. The experimental results are shown in Figure 3. It can be clearly seen that the over-fitting phenomenon has basically disappeared, and the accuracy rate can basically be stabilized above 98%, and the highest accuracy rate can reach 100%. However, in several experiments, there will be slight instabilities each time, and the accuracy rate cannot be stabilized at 100%. So a conjecture is put forward: the axial and radial data are combined after...
convolution and directly enter the fully connected layer, so that the axial and radial data are not perfectly integrated.

Therefore, the second-generation model is modified to perform two-layer convolution pooling on the fused data after the fusion layer, so that the axial and radial data are more closely integrated. The modified third-generation model structure model is shown in Table 4.

| No. | Layer  | Kernel size/Step | Kernels number | Output size | Padding |
|-----|--------|------------------|----------------|-------------|---------|
| 1   | Conv1(2) | 128*1/16*1        | 16             | 128*16      | same    |
| 2   | Pool1(2) | 2*1/1*1           | 16             | 64*16       | valid   |
| 3   | Conv3(4) | 3*1/1*1           | 32             | 64*32       | same    |
| 4   | Pool3(4) | 2*1/1*1           | 32             | 32*32       | valid   |
| 5   | Concatenate |                 |                | 64*32       |         |
| 6   | Conv 5  | 3*1/1*1           | 64             | 64*64       | same    |
| 7   | Pool 5  | 2*1/1*1           | 64             | 32*64       | valid   |
| 8   | Conv 6  | 3*1/1*1           | 64             | 32*64       | same    |
| 9   | Pool 6  | 2*1/1*1           | 64             | 16*64       | valid   |
| 10  | Fully Connect | 100             | 1              | 100         |         |
| 11  | Softmax | 3                | 1              | 3           |         |

3.2.1.4. Results and analysis of third-generation model training. The experimental results are shown in Figure 4. It can be seen that the accuracy of the test set of the third-generation model can reach nearly 100% and is quite stable. This model verifies the above conjecture: the 2 input convolutional neural network needs to be combined through the concatenate function to correspond to the frame, and the two inputs need to be tightly combined through the convolutional layer and the pooling layer to further improve the stability of the model.

![Figure 4](image1.png)

**Figure 4.** The training accuracy of the third-generation model.

3.2.2. Combined Input Multi-input Convolutional Neural Network. The combined input means that the axial and radial data are fused before the input layer. The combine process is shown in Figure 5, that is,
the size of each data of the axial and radial data in this experiment is 1*2048. After fusion and superposition, the data the size becomes 2*2048, and then the combined data is input into the convolutional neural network. The schematic diagram of the model structure is shown in Figure 4. Since the data is combined before the input layer, this model can be designed with reference to the second-generation model of the independent input convolutional neural network. The model structure parameters are shown in Table 5.

Table 5. Combined Input Multi-input Convolutional Neural Network model structure parameters.

| No. | Layer     | Kernel size/Step | Kernels number | Output size | Padding |
|-----|-----------|------------------|----------------|-------------|---------|
| 1   | Combine   | 2048*2           |                |             |         |
| 2   | Conv1     | 128*1/16*1       | 16            | 128*16      | same    |
| 3   | Pool1     | 2*1/1*1          | 16            | 64*16       | valid   |
| 4   | Conv2     | 3*1/1*1          | 32            | 64*32       | same    |
| 5   | Pool2     | 2*1/1*1          | 32            | 32*32       | valid   |
| 6   | Fully Connect | 100       | 1             | 100         |         |
| 7   | Softmax   | 3                | 1             | 3           |         |

3.2.2.1. Results and analysis. The experimental results are shown in Figure 6. It can be seen that the training accuracy of the CIMI-CNN model after training can reach 100%. And after many experiments to compare with the above-mentioned independent input convolutional neural network third-generation model, the independent input convolutional neural network three-generation model generally needs about 13-14 training epochs to reach the highest accuracy rate, and for the combined input model, it only takes 10-12 training epochs to reach the highest accuracy rate. It can be seen that the CIMI-CNN has higher convergence and fitting efficiency than the IIMI-CNN. However, the combined input model does not perform well in training stability. It can be seen from the figure that unstable fluctuations sometimes occur during the training process.

Figure 6. The training accuracy of combined input model.

4. Comparative Analysis

In this section, we will use the convolutional neural network model proposed by the predecessors in the bearing fault diagnosis, to train in the aircraft hydraulic pump data. Through the WDCNN model, only the radial data is input for training, and the training accuracy result is shown in Figure 7.

Figure 7. WDCNN model training hydraulic pump radial data.  
Figure 8. Time domain and frequency domain 2 input model training.
It can be seen that the training results are relatively stable, the accuracy rate is stable at about 92%, and the highest accuracy rate is nearly 97%. Compared with the multi-input convolutional neural network, it shows that 2 inputs add more features to the fault signal, and the features are more obvious, which is of great significance to fault diagnosis. Tao Zan et al. proposed multi-input training for time-domain data, frequency-domain data, and envelope spectrum data. Here, in order to compare the multi-input proposed in this article, we use time-domain data to obtain frequency-domain data through Fourier transform. The time domain and frequency domain are taken as two inputs into the IIMI-CNN third-generation model proposed above for training. The result of training accuracy is shown in Figure8. 

It can be seen that the training result is relatively stable, but the accuracy rate is low, and the accuracy rate is stable at about 74.5%. The 2-input combination of time-domain signal and frequency-domain signal is less effective than training with time-domain signal alone. This result shows that the time domain signal and the frequency domain signal cannot be simply combined through the above model and achieve better results.

5. Conclusion
This paper proposes two multi-input convolutional neural network models for fault diagnosis training of aircraft hydraulic pumps. Found through experiments:
(1) Using two data of the same type that respectively show the characteristics of the machine in two different directions as the two inputs of the convolutional neural network will greatly improve the accuracy of the convolutional neural network for machine fault diagnosis.
(2) The two inputs of the CIMICNN need to be combined after separate convolution, and then convolution pooling and fusion are performed to make the model performance better.
(3) Compared with the independent input model, the combined input model has higher convergence and fitting efficiency in the model training process, but the stability of the former is Slightly worse.
(4) The multi-input CNN has strong adaptability to the two inputs described in (1).

References
[1] Duy-Tang Hoang, Hee-Jun Kang. Rolling element bearing fault diagnosis using convolutional neural network and vibration image[J]. Cognitive Systems Research, 2018, 53.
[2] Chenhua Ni, Xiangdong Ma. Prediction of Wave Power Generation Using a Convolutional Neural Network with Multiple Inputs[J]. Energies, 2018, 11(8).
[3] Tao Zan, Hui Wang, Min Wang, Zhihao Liu, Xiangsheng Gao. Application of Multi-Dimension Input Convolutional Neural Network in Fault Diagnosis of Rolling Bearings[J]. Applied Sciences, 2019, 9(13).
[4] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).
[5] Yandong Li, Zongbo Li, Hang Lei. Summary of Convolutional Neural Network Research[J]. Computer Application, 2016, 36(09):2508-2515+2565.(in Chinese)
[6] Wei Zhang. Study on bearing fault diagnosis algorithm based on convolutional neural network[D]. Harbin Institute of Technology, 2017.(in Chinese)
[7] Nikolaos Tziolas et al. Employing a Multi-Input Deep Convolutional Neural Network to Derive Soil Clay Content from a Synergy of Multi-Temporal Optical and Radar Imagery Data[J]. Remote Sensing, 2020, 12(9).
[8] Yang Yinghua, Li Doliang, Liu Xiaozhi. Fault Diagnosis Based On One-Dimensional Deep Convolution Neural Network[A]. Northeastern University, Chinese Society of Automation Cyber-Physics System Control and Decision Professional Committee. Proceedings of the 32nd China Control and Decision Conference (3) [C]. Northeastern University, Chinese Society of Automation Cyber-Physical System Control and Decision Professional Committee: "Control and Decisions" Editorial Department, 2020:6.(in Chinese)