Thesis for the Degree of Master of Engineering

A study on the fault diagnosis based on improved deep convolutional neural network

by

SHAO XIAORUI

Department of Information Systems
(Interdisciplinary Program)

The Graduate School

Pukyong National University

May 2019

A Study on the fault diagnosis based on improved deep convolutional neural network

개선된 심층 신경망 기반 오류진단 연구

Advisor: Prof. Chang-Soo Kim

by

SHAO XIAORUI

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Engineering

in the Department of Information Systems, the Graduate School, Pukyong National University

May 2019

A study on the fault diagnosis based on improved deep convolutional neural network

A dissertation

by

SHAO XIAORUI

Approved by:

(Chairman) ***Chang-Soo Kim***

(Member) ***Kyung-Hyune Rhee*** (Member) ***Ha-Kyung Kim***

May 16, 2019

Contents

List of Figures

List of Tables

List of Abbreviations

FD : Fault Diagnosis

CNN : Convolutional Neural Network

DNN : Deep Neural Network

APS : Air Pressure System

SVM : Support Vector Machine

AI : Artificial Intelligence

PCA : Principal Component Analysis

ICA : Independent Component Analysis

DBN : Deep Brief Network

RNN : Recurrent Neural Network

DPCA : Deep Principal Component Analysis

DICA : Deep Independent Component Analysis

LSTM : Long Short-Term Memory

RLEU : Rectifier Linear Unit

FPR : Fault Positive Rate

NA : Not available values (Missing Values)

NO : Normal status (No fault)

OF1 : Outer ring Fault of bearing Diameter 1

IF1	: Inner ring Fault of bearing Diameter 1
BF1	: Ball ring Fault of bearing Diameter 1
OF2	: Outer ring Fault of bearing Diameter 2
IF2	: Inner ring Fault of bearing Diameter 2
BF2	: Ball ring Fault of bearing Diameter 2
OF3	: Outer ring Fault of bearing Diameter 3
IF3	: Inner ring Fault of bearing Diameter 3
TP	: True Positive
FP	: False Positive
TN	: True Negative
FN	: False Negative

개선된 심층 신경망 기반 오류진단 연구

SHAO XIAORUI

부경대학교 대학원 정보시스템협동과정

요약

오류 진단은 제품 생산의 중요한 문제이다. 이는 대부분의 불안정한 요소를 발견할 수 있는 장점이 있고, 정확한 조치를 통해 손실을 줄이고 비용을 절약할 수 있는 방법이다. 최근에는 기계 학습(machine learning)의 연구들이 발전함에 따라 데이터 구동에 기반한 오류 진단기술이 이슈화 되고 있다. 딥 러닝기반의 컨볼루션 뉴런 네트워크(CNN: Convolutional Neural Network)의 강력한 추출 특성 능력은 오류 진단을 진행할 수 있는 방법으로 인식되고 있다. 그러나 CNN 의 레벨들이 깊어지면서 숨겨진 유용한 정보들이 사라지는 문제점이 있어, 진단 성능을 떨어뜨리는 단점이 있다.

본 논문에서는 이러한 문제점을 극복하기 위해 CNN 에서 이전 계층에서 계산한 결과가 다음 계층에 더해지는 새로운 입력 계층을 추가하여 정확성을 높이고자 하였다. 따라서 제안된 방법을 적용하기 위해 본 연구는 3 개의 공공 데이터를 사용하여 비교하였다. 테스트 데이터로는 베어링(bearing) 데이터, 반도체 제조 프로세스 데이터, APS 고장 데이터를 사용하였다. 3 개 공공 데이터에 대한 실험 결과는 베어링의 정밀도가 100%, 반도체 제조 공예의 정밀도가 100%, APS 고장 데이터는 97.7%였다. 본 연구는 기존의 딥 러닝 기반인 Deep Belief Network(DBN)과 Deep Neural Network(DNN), 그리고 Normal CNN 모델과 비교하여 제안된 방법은 입력 데이터를 간단하게 사용하고, 다른 예비 처리가 필요 없이 좋은 성능을 얻을 수 있음을 제시하고 있다.

Keyword: 오류 진단, 심도 학습, CNN.

Chapter 1.

Introduction

1.1 Study Background

Since the German Industrial 4.0 Group published its report on industrial 4.0 in 2013[1], the fourth industrial revolution has attracted worldwide attention. For example, the United States put forward the Industrial Internet in 2014[2], China proposed the Internet + in 2015[3], and the Korean government put forward the implementation plan of manufacturing innovation 3.0 strategy in 2015[4]. The essence of industry 4.0 is the integration of physical information and interconnected information[5]. In order to realize the intelligent factory under the industrial 4.0 manufacturing system, cloud computing technology, big data technology, and AI technology are needed deeply. In the process of building a smart factory, the damage of some parts will lead to equipment stop operation, reduce production, even stop production and pose a serious threat to personal safety. Finding an effective way to detect the fault and take the right action to stop the fault is the key factor to implement smart manufacturing.

1.2 Motivation

At the background of industrial 4.0, higher requirements are put forward for the accurate mastery of faults. But the currently proposed methods spend more time for fault diagnosis, however, the performance is low. It does not satisfy the need for human beings. In this study, combined fault diagnosis with deep learning technology, adopt an image processing method to handle vibration signals for fault diagnosis. Hope to give a new view for fault diagnosis and improve the accuracy of fault diagnosis by using the proposed method. Different from the traditional methods, do not need to select the features of the original data set to fault diagnosis and do not need to do more features engineering. Just by inputting the original data set, the proposed method will do fault diagnosis automatically and intelligently. Thus, can save much more time and offer the fault status for the manager in time to reduce the loss and ensure the production security and safety.

1.3 Thesis Research Objective

The objective of this thesis research is as follows.

(1) Looking for a new method for fault diagnosis with higher accuracy, lower cost, lower time.

(2) Giving the framework of proposed improved CNN based method.

(3) Using improve CNN based method to improve the accuracy of fault diagnosis.

1.4 Structure of the Thesis

The structure of this thesis has been divided into five chapters as follows:

Chapter 1 Introduction: introducing the background of fault diagnosis and motivation for this thesis, also giving the research objective of this thesis, the whole structure.

Chapter 2 Related Research: This chapter describes the related research for fault diagnosis in the world. Summary of the advantages and disadvantages of pass fault diagnosis.

Chapter 3 Related methodologies: This chapter summarizes the typical CNN framework and explains some mathematical knowledge about CNN, Improved CNN based method for fault diagnosis was described.

Chapter 4 Case study: This chapter uses 3 different public data set for testing the effectiveness of improved CNN based method for diagnosis.

Chapter 5 Conclusion: This chapter gives the conclusion of improved CNN based method using for fault diagnosis. And discussed the future work.

Chapter 2.

Related research

2.1 Fault Diagnosis

Fault diagnosis mainly compresses two parts, one is fault detection, another one is fault diagnosis. Fault detection is only needed to identify whether the fault exists or not. And if there is a fault state has been detected, can take some actions to prevent it. Especially, it can be treated as special fault diagnosis. For fault diagnosis, it is not only necessary to detect whether a fault exists or not, but also what kind of fault it is. Simply, can treat fault detection as a binary classification problem, and fault diagnosis can be treated as a multi-classified problem.

2.2 Traditional Fault Diagnosis Method

In order to reduce the harm caused by a failure. Some methods have been conducted. Traditional fault diagnosis methods were first applied in the field of fault diagnosis. one of traditional diagnosis is mechanism-based method[6]. The mechanism-based method needs operators to have a clear understanding of the whole operation mechanism, including the principle, structure, and function of each equipment. According to the operation mechanism of equipment, a mathematical model can be established to describe the whole system. However, the modern industry presents a large and complex direction of development. Operational data and states are multivariate and heterogeneous, which makes the mechanism-based method very limited[7]. Another fault diagnosis method is knowledge rules-based method. It also needs experts who are engaged in equipment management or have rich knowledge about the whole system[8]. However, with the rapid development of manufacturing industry, the formalization of equipment is becoming more and more diverse, and the speed of replacement is fast. Specialized equipment needs to train special operators to control equipment. Knowledge rules-based method requires enterprises to spend too much time and money to train special operators. In addition, more and more people are reluctant to accept the profession of equipment operators for certain reasons, which has a bad impact on the operation and management of equipment. To overcome the limitations of traditional methods. Some researchers have turned a new fault diagnosis area data-driven based method.

2.3 Data-Driven Fault Diagnosis Method

2.3.1 Machine Learning Based method for Diagnosis

With the development of big data technology and artificial intelligence technology, data-driven based fault diagnosis method has been proposed[8]. Firstly, statistical analysis methods have been successfully applied in an industrial system for fault diagnosis, such as principal component analysis (PCA)[9][11][12][13], independent component analysis (ICA)[14][15][16]. PCA and ICA only get the features by utilizing the linear operations, it will be lost the detailed information, and they are not suitable for complex, massive data[17]. To overcome those shortcomings, some methods have been proposed and applied widely in industrial production. In particular, machine learning-based methods have been widely used in fault diagnosis. In Si-Yu Shao et al. Also called machine learning based method as artificial intelligence (AI)-based fault diagnosis techniques[18]. To avoid ambiguity, in this study called this method as a machine learning based fault diagnosis. Support vector machine (SVM)[19][20] is the main machine learning method used for fault diagnosis because of its outstanding features extraction abilities by mapping the multi-dimensional features into a higher dimension. It can be used for multi-classification directly. Recently, some researchers combined the SVM with other algorithms for improving the performance of fault diagnosis. In reference[21] uses Gaussian kernel and SVM select nonlinear feature for fault diagnosis. Reference[22] combine the SVM and model-based method for bearing faults diagnosis. However, SVM requires the huge data set for training the model, and if data is not balanced, it will not get the high performance. Moreover, SVM needs a long time to train the model and consume most of the computing power when data is massive. Another powerful algorithm for fault diagnosis is random forest, it is a tree-based method and insensitive to missing values. Thus, it is adopted by many industry systems. As a promotion of decision tree, random forest usually performs better than SVM and decision tree. Reference[23][24] [25] applied random forest for fault diagnosis and comparing with wavelet and SVM, it shows the better per

formance can be got by Random Forest. However, it cannot give a full play to the superiority of the baggings when only a few samples are variable for model establishment[17]. In other words, random forest and SVM cannot extract the full useful and hidden features from the original data set. Common machine learning based fault diagnosis method cannot do fault diagnosis accurately due to it lacks some critical and hidden feature information and has some limitations such as memory consumption and time consumption. Feature representation is a key factor for fault diagnosis, how to overcome the limitations of common machine learning based method is becoming a hot topic recently. One effective method has been successfully applied. It is a deep learning-based method.

2.3.2 Deep Learning Based Fault Diagnosis Method

As a new method in the field of machine, deep learning has achieved a brilliant achievement in the fields of handwritten digit recognition[26], computer vision[27][28][29][30], speech recognition[31][32] etc... because of its powerful automatic feature extraction capability[9]. Deep learning originates from neural networks. Multiple-layers perceptron and multiple hidden layers are the prominent features of deep learning models. Unlike ordinary light learning algorithms, deep learning algorithms use multi-layer hidden layers to realize the layer-by-layer transformation of data features, in order to ensure the most effective information extraction and feature expression.

The deep learning methods we know well include deep belief networks (DBN), convolution neural networks (CNN), recurrent neural networks (RNN). Among them, in the fields of fault diagnosis, DBN was adopted firstly. Reference[33] uses the DBN to classify the faults of induction motors. [18][34] all use DBN to realize the fault diagnosis of rolling bearings and gearboxes. [35] implement a global two-layer back-propagation network based on DBN for complex chemical process fault diagnosis. Compared with some mainstream fault diagnosis methods, the robustness and accuracy of the proposed method are verified. However, deep belief network uses unsupervised way to pre-train RBM and uses a specific classification algorithm to fine-tune the global, when data is large, will cost too much time, and improper parameter selection is easy to fall into a local optimum. For the application of fault diagnosis using RNN, because of RNN performs well in sequential data set, so, most of RNN based fault diagnosis method is long-short term memory (LSTM). [36] utilized LSTM for sequential fault diagnosis and tested on the Tennessee Eastman benchmark process data set. Get the mean accuracy is about 0.821. And comparing it with DPCA, DPCA+SVM, DLDA +SVM, it shows the DBN-based method performs much better than others. However, [37] using the CNN method for diagnosis in the same data set to [36], get the mean accuracy is 0.882 in the test data set. It shows the LSTM cannot get the full hidden information in the original data set. So many evidences show that CNN is the most powerful algorithm for fault diagnosis when the data set has many columns because of CNN has an excellent feature extraction ability. So, this study is focusing on CNN based fault diagnosis. Following, a summary of the recent research related fault diagnosis based on CNN method.

CNN needs at least two-dimensional data, it can extract the hidden features by convolution operations, and reduce the dimension of the data by polling operation. [38] adopted two softmax classifiers to solve the classification of two non-independent problems of fault diagnosis and fault phase selection inside and outside the area, and realized the weight sharing to a greater extent. [39] used CNN to classify gear faults and compared it with traditional classification algorithm SVM. [40] proposed a new CNN based on LeNet-5 and tested on the different data sets, which adopts two fully connected layers to fault diagnosis. In fact, it is the different type of CNN. How to define the depths of the network still need to think more. The other research is based on normal CNN. Specifically [41] proposed a multiscale convolutional neural network for turbine gearbox fault diagnosis. It adopts three different inputs for full features extractions. Comparing it with traditional CNN, the proposed method gets the highest performance for all of faults types. However, for [41] proposed multiscale CNN, how to determine multiscale is confusing us sometimes. So, in this study, proposed a new improved CNN model for fault diagnosis. And tested on different data sets.

Chapter 3.

Methodologies

3.1 Overview of Convolutional Neural Network

CNN is proposed by LeCun et al[31] which is a typical feedforward neural network, essentially constructs a number of filters that can extract the characteristics of input data. Through these filters, the input data is convoluted and pooled, and the topology features hidden in the data are extracted step by step. With the deep entry of the network layer, the extracted features are abstracted and finally obtained. The final obtained data has translation, scaling and rotation invariance. Its main features are a sparse connection, weight sharing, and downsampling. The sparse connection can reduce the number of training parameters and speed up the convergence; weight sharing can effectively avoid algorithm overfitting; downsampling makes full use of the features of the data, reduces the data dimension, optimizes the network structure[42][43][44]. CNN can deal with one-dimensional (1-D) signals and sequences, two-dimensional (2-D) images, and three-dimensional (3-D) videos, due to their powerful ability of automatic feature extraction and classification[41].

3.2 Typical Convolutional Neural Network

The typical convolution neural network contains the input layer, the convolution layer, pooling layer, fully connected layer and the output layer as shown in Figure 3.1. Each sample is input in a two-dimensional array and then mapped to the hidden layer by the convolution kernel (the weight matrix of neurons) to obtain the abstract features. Because CNN is working in two-dimension data. After convolution operations get the features, which was called the feature map. The hidden layer is made up of the convolutional layer (C layer) and pooling layer (P layer). The C layer and the P layer alternately appear, and the output of each layer is the inpu1 t of the next layer. The feature of the original data is abstracted to the upper level through the hidden layer, and finally, the output is obtained through a fully connected layer in which transform the two-dimension data into one-dimension data and determinate which class it is belonging according to the maximum probability of each class.

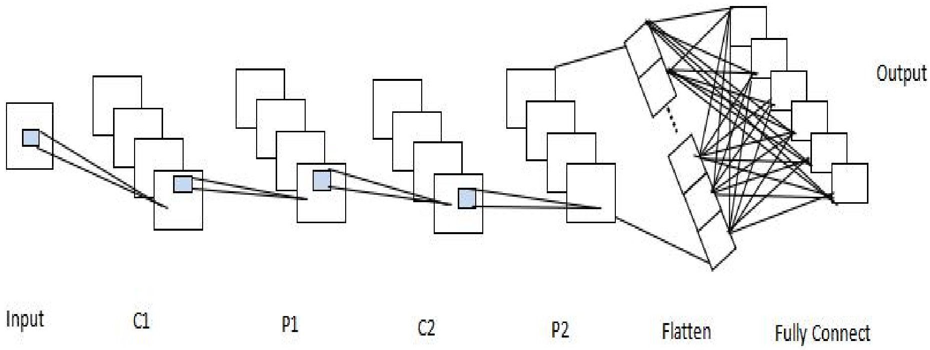


Figure 3.1 Structure of CNN

The convolution layer (C layer): C layer is used to extract the local features of each neuron. Each C layer has multiple feature matrices. Each characteristic matrix can be regarded as a plane. The convolution function of the same plane is the same. It has the characteristics of rotation, displacement invariant and weight sharing, and has strong robustness. Different plane convolution functions by changing the numbers guarantee sufficient feature extraction. The convolution process of the C layer is shown in Fig 3.2, in which the input feature matrix of the previous layer is assumed to be $(N \times N)$, the size of the convolution kernel is $(K \times K)$ for the two-dimensional convolution operation, and the output characteristic matrix $(M \times M)$ is obtained after the activation function, and the relation of the three parameters is $\mathrm{M} = (\mathrm{N} - \mathrm{K})/\mathrm{stride} + 1$. Stride is removed step in each computation step. The calculation formula is shown as formula 3.1.

$$x_j^l = f(\sum_{i \in Mj} x_i^{l-1} \times k_{ij}^l + B_j^l) \tag{3.1}$$

Among them, M_j is the input feature, l is the lth layer in the network. k represents the convolution kernel, it is made of $s \times s$ which is the kernel size. and B is the additive bias of each feature map. f is a nonlinearity active function, typically it is $sigmiod$ function or Rectified Linear Unit (ReLU).

The process as shown in Figure 3.2. Giving an example with the input shape is 4×4 , and using 4 filters to do convolution operations, the stride is 2. It means every step will jump 2 columns or two rows. Will compute the sum of each kernel processed data. And add bias. Through the active function, under the demands of the active function will be active. Getting the four different features map. It is the forward convolution operation process. For the backpropagation process will be introduced the next.

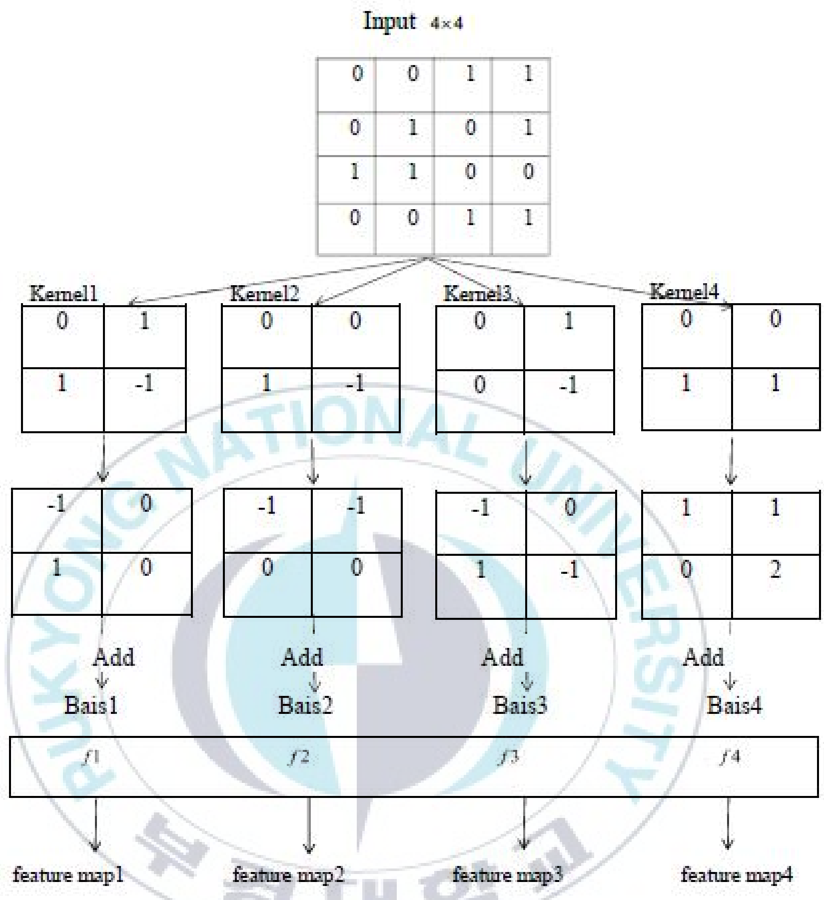


Figure 3.2 Forward convolution operation process with input 4×4

Pooling layer (P layer): the P layer scales the data of the upper layer to reduce its dimension, the feature of the extraction has the scale invariance, the general pooling method has the maximum pooling and minimum pooling, but in order to prevent overfitting, the average pool method was used to adopt[45]. we can consider the pooling layer as the second extraction features. The calculation method of its neuron can be expressed as formula 3.2.

$$x_j^i = f(\beta_j^l down(x_i^{l-1} + b_j^i)) \tag{3.2}$$

$down(.)$ is a downsampling function, and β is a multiplicative bias of the network. Mean pooling process is shown in Figure 3.3. The shape of the feature map is 4×4, after mean polling operation process. The shape of the output feature map is 2×2. Utilizing the polling operation can effectively reduce the dimension of the feature map.

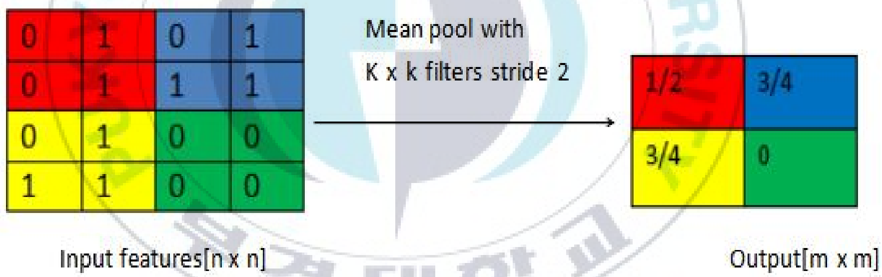


Figure 3.3 Down polling process

Output layer: before output, usually adopting the full connection layer to flatten two-dimension feature map into a one-dimension feature. After flattening, the features will be transformed into a vector. And to discriminate between n classes a full connected layer output layer with n neurons is added. The following formula 3.3 will be used for connecting fully.

$$O = f(b_o + w_o f_v) \tag{3.3}$$

where b_0 is a bias vector and w_0 is a matrix of weight. f_v is the feature vector. For fault detection utilize $sigmoid$ function[46] as the active function was described in formula 3.4, and fault diagnosis utilize $soft\max$ function[47] as the active function was expressed in formula 3.5. In $sigmoid$ function, the output features will be transformed into (0, 1), and in $soft\max$ function, the features vector will be transformed into [0, 1], it is the probabilities of classes. Adopting the maximum probability as a result.

$$S(x) = \frac{1}{1+e^{-x}} \tag{3.4}$$

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{3.5}$$

Parameter training: CNN is a supervised deep learning method. The training of the network is similar to the traditional neural network. k_{ij}^l , B_j^l b_o w_o of the model are learnable parameters. Learning is done with BP algorithm used to propagate the error direction as shown in[48]. Finally, to minimize the loss, the parameters between layers are adjusted by the gradient descent method according to the loss function.

3.3 Proposed Improved Convolutional Neural Network for Fault Diagnosis

CNN was widely applied in many of fields due to its powerful feature extraction ability. Different from the normal data-driven methods used for fault diagnosis. This study proposed a new approach for diagnosis using an image processing method based on CNN. Proposed CNN based method is different from normal CNN based method. The typical CNN contains an input layer, convolution layer, pooling layer, and flatten layer, output layer. In this study, presented an improved CNN based method for fault diagnosis. The structure of processed CNN as shown in Fig4.

From Figure 3.4, the proposed CNN based method adopt input layer, convolution layer, and global average layer instead of the flattening layer and fully connected layer. It can reduce the training time, and simply the network. The key idea of the proposed CNN based method is adopting the Add layer to prevent the hidden key information loss with the depth and complexity improving. Add layer add the previous convolution layer and the next convolution layer as the input of the next convolution layer. In the proposed CNN based method, applying the image processing method process the input original data. And through the feature extraction by the proposed CNN network, get the best performance for fault diagnosis.

Input layer: Named as input, the type of input is Input Layer. In this layer, the original data set will be transformed into a two-dimension image data set. Transformed data as the network input.

Convolution layer: It is named with the prefix con2_, same to aforementioned typical CNN. In this layer, will extract the low level and abstract features hidden in the data set through the convolution operations.

Add layer: The key idea of proposed improved CNN, it is named with prefix add_. Containing two parts, one adds the input layer and the results after the first convolution operation. Another one is to add the previous convolution results and the current convolution results as the input of the next convolution layer. By adding operations, reduce the key feature information loss with the depth of network increasing. Ensure more hidden feature network can get.

Dropout layer: To prevent the overfitting, dropout layer was applied. Failure of some nodes by setting a hyperparameter to overcome the overfitting during the training process.

Global average pooling layer: [49] presented a global average pooling firstly. It reduces the dimension from 3D to 1D. Therefore, Global pooling outputs 1 response for every feature map. This can be the maximum or the average or whatever other pooling operation you use. It is often used at the end of a convolutional neural network to get a shape that works with dense layers. Therefore, no flatten layer must be applied. Different from the traditional CNN based method, in this study, did not adopt the pooling layer and flatten layer. Utilizing the global average polling instead of flattening layer. The difference between flattening and global average pooling layer as shown in Figure 3.5 and Figure 3.6. Normal CNN adopts flatten operations to transform two-dimension features into a one-dimension feature vector and then do fully connected operations. In proposed CNN based fault diagnosis method, adopting global average pooling layer directly get the outputs nodes without flatten and full connection operations. By calculating the average of each feature map as the output nodes, it can reduce the overfitting and save training time effectively, but obtain the high performance.

In this study, the laboratory is based on the TensorFlow Keras API. The experimental platform is Ubuntu 16.04.3 64 bits, memory 23.4 GB, Intel (R) i7-700 CPU, processing speed 3.6 GHz x. And the activation function in convolutional layer adopts ReLu function, and the stride adopts the default value, which is 1, and the kernel size is 3 by 3. All CNN based method adopts one max pooling layer, and one flatten layer, one dropout layer which value was set as 0.5. Loss function is categorical_crossentropy function. The optimizer is an Adaptive Moment Estimation. Different structure of CNN based method only by changing the numbers of convolution layers.

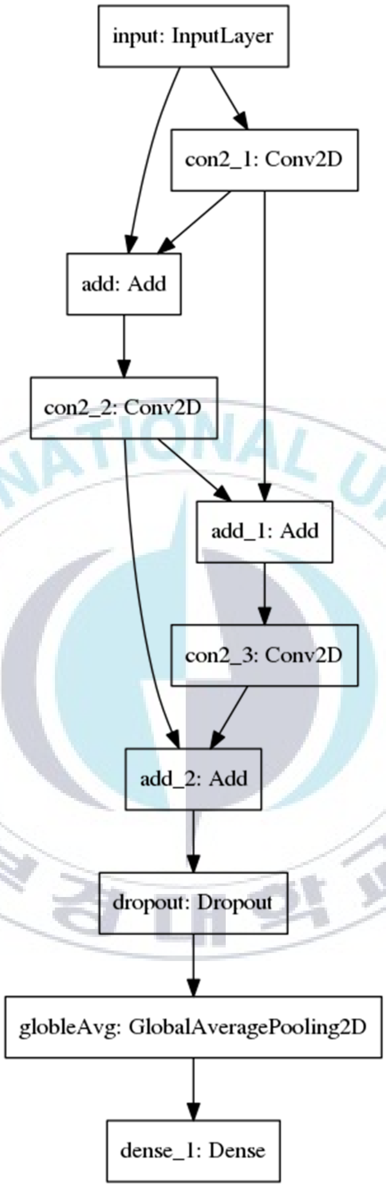


Figure 3.4 The structure of the proposed improved CNN for fault diagnosis

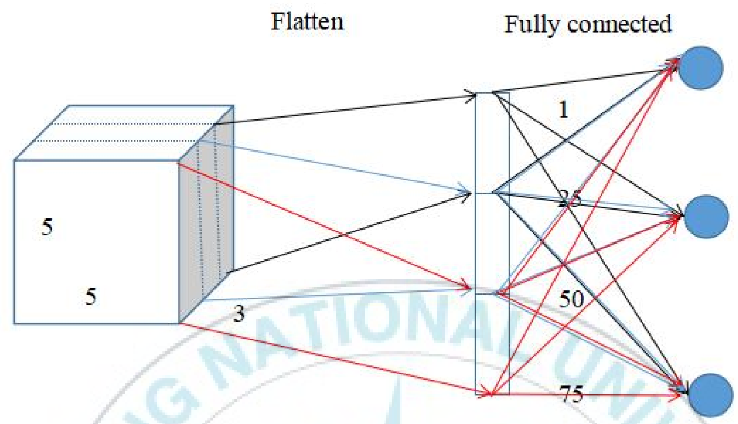


Figure 3.5 Flatten and fully connected process in CNN

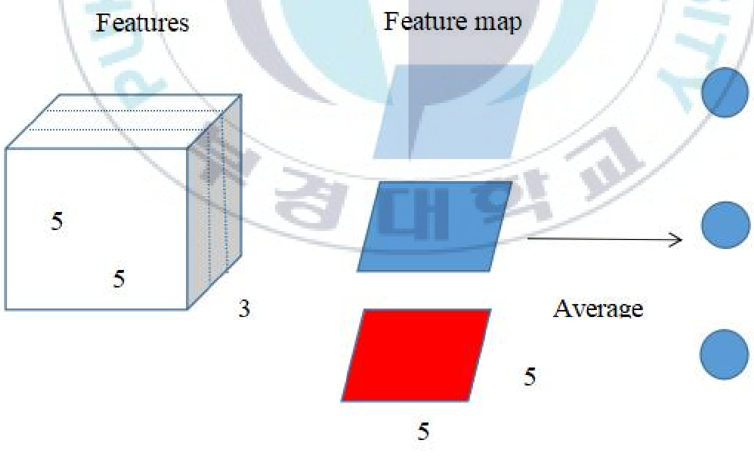


Figure 3.6 Average pooling process in the network in network

Chapter 4.

Case Study

4.1 Case Study 1: Bearing fault diagnosis

4.1.1 Bearing Data Description

The bearing is one of the key components in industrial production processing. The bearing is easy to damage because of overload, wear, corrosion and other reasons. More than 50% of rotating machine faults are related to bearing faults. The early bearing weak is complex and difficult to detect. Therefore, the morning and analysis of bearing fault diagnosis are important. In this study, adopting bearing vibration signals for fault diagnosis. The data set is from bearing fault diagnosis competition[50]. The data set has 792 samples, every sample is vibration signals which was continuous sampling according to the time series. It contains 6000 columns in each sample. And the fault types are ten, from 0 to 9. The special information about the fault type as shown in Table 4.1. There are three kinds of faults according to the diameter of bearing. For diameter1, it has an outer ring fault, inner ring fault, and ball fault, it was labeled as 1, 2, 3. Same to the diameter1, diameter2, diameter3 also has three types of fault. They are labeled as 4, 5, 6, 7, 8. And the label 0 represents there is no fault.

Table 4.1 The fault types of bearing

Type	Outer ring fault	Inner ring fault	Ball fault	Normal
Diameter1	1	2	3	0
Diameter2	4	5	6	
Diameter3	7	8	9	

Giving the different fault types of distribution of this data set. As shown in Figure 4.1. The results show that different bearing fault types have a different rate. This figure shows each fault type is not balanced, specifically, the type of ‘normal’, ‘diameter 3 outer ring fault’, ‘diameter 3 ball fault’ take up a lot. To verify the proposed CNN has the ability to process the imbalanced data set. There, do not need to preprocess.

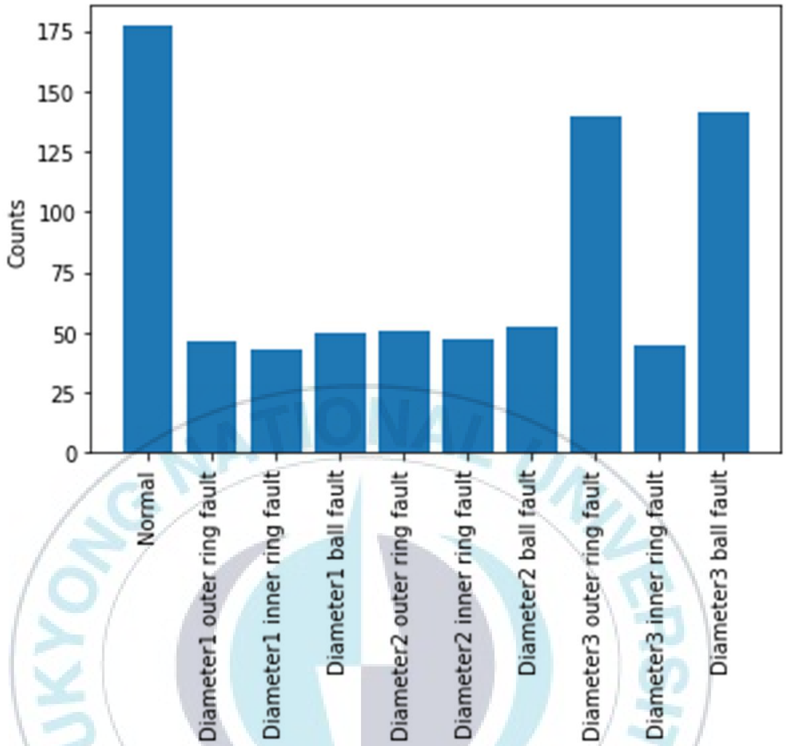


Figure 4.1 The fault types of distribution

Also, the original vibration signals were given in Figure 4.2. From Fig8, know the difference directly among the different fault types. The range of values also can be seen clearly. The change range of diameter3 fault is the largest which is from -5 to 5.

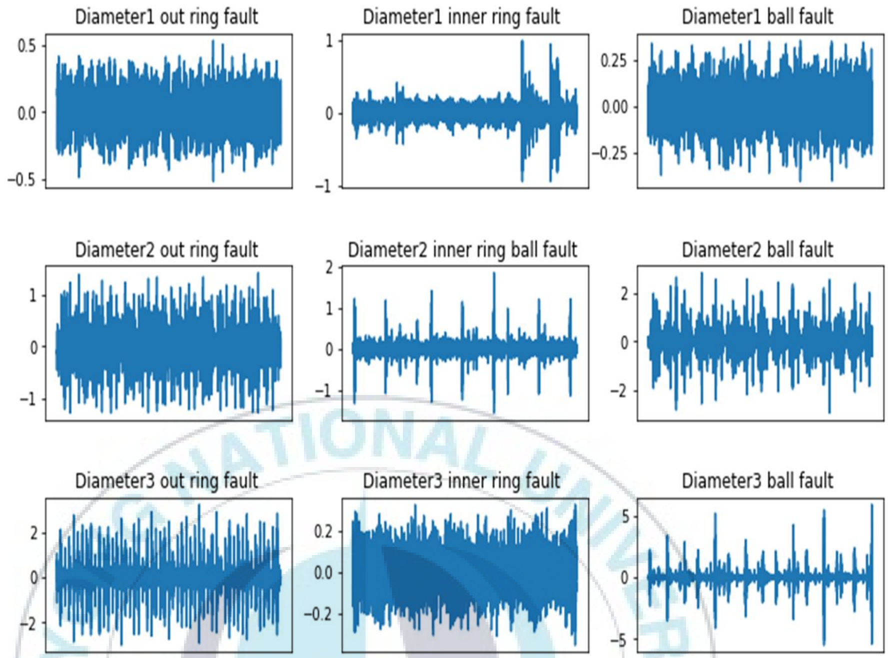


Figure 4.2 Each type of faults examples from original data

4.1.2 Bearing Fault Diagnosis Using Improved CNN

In this section, the whole process of fault diagnosis was drawn. The process of fault diagnosis using CNN has four steps.

First, because CNN requires the data shape is tensor, we need to transform the original data into a tensor. And then to train and verify the model, split the data into training data and testing data. The third is constructing the CNN model. At last, evaluating the trained model. The detailed description as shown in the following:

(1) Transform the data into tensor

CNN is good at processing multidimensional tensor, but original data set is a two-dimensional matrix. First, need to be transformed. In bearing fault diagnosis case, the shape of the original data set is [792,6000], in which every sample will be transformed into a three-dimensional tensor. The transformed shape will show like [792, length, width, 1], the first digit 792 is the number of samples, the relation between length and width is $length \times width = 6000$, which means each sample was treated as one image with 6000 pixels. And the last parameter 1 is the number of channels, in this study, the data has no real meanings of an image, so set the channel is 1. All of the samples will be treated as gray images after transformation. In this study, adopting the different shape of the image to train the model. Here, an example was given about reshaping tensor data with shape [100, 60]. The visualization of transformed data as shown in Figure 4.3.

Figure 4.3 utilizes the same data in Figure 4.2. From Figure 4.3, The fault of diameter1 inner, diameter 2 out ring, diameter3 ball fault are noisy than others. It is difficult to identify which type of fault it is belonging to transformed data visualization directly.

Figure 4.3 Transformed data visualization

The aforementioned transformation is related to the features. For the label, in this study called fault type, also need to reshape to satisfy the requirement of the CNN model. In this study, adopting one-hot encoding to process the label of fault type. In digital circuits and machine learning, [51] explains one-hot is a group of bits among which the legal combinations of values are only those with a single high (1) bit and all the others low (0). One-hot encoding eliminates the impact of different priorities and enables only one state of each step to be activated. In bearing fault diagnosis case. Totally having ten different types of fault, the one-hot encoding results as shown in Table 4.2.

Table 4.2 One-hot encoding for bearing fault diagnosis

label	0	1	2	3	4	5	6	7	8	9
0	1	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0
2	0	0	1	0	0	0	0	0	0	0
3	0	0	0	1	0	0	0	0	0	0
4	0	0	0	0	1	0	0	0	0	0
5	0	0	0	0	0	1	0	0	0	0
6	0	0	0	0	0	0	1	0	0	0
7	0	0	0	0	0	0	0	1	0	0
8	0	0	0	0	0	0	0	0	1	0
9	0	0	0	0	0	0	0	0	0	1

(2) Splitting the tensor into the training set and testing set

After transforming the data, it is time to split data into the training part and testing part. In this study, 80% of each data set was selected for the training set, and 20% of each data set was selected for the testing data set. To ensure the comparability of different models. The splitting data sets are the same for different models using the same data set.

(3) Constructing the model

This study proposed a new method based on CNN for fault diagnosis, as described in section 3.3. According to the description of section 3.3, built the model. In this improved model. Contains one input layer, which shape has been introduced in the previous step (1). It is [100,60,1], three convolution layers, and 3 Add layers, one dropout layer for reducing the overfitting. One global average layer. All of the kernel sizes were set as [3,3], all of the activation functions in convolution layer are Rectified Linear Unit (ReLu), the definition of ReLu as shown in formula 4.1[52]. It means if the feature map elements are greater than 0, will take the element as the output, else will take 0 instead of that element.

$$f(x) = \max(0, x) \tag{4.1}$$

Proposed CNN based method adopts $soft\max$ as the classifier and $\text{categorica l\_crossent ropy}$ a loss function. $soft\max$ already introduced in section 3.1. $\text{categorical\_crossentropy}$ as shown in formula 4.4. In formula 4.2, z is output through the network without activation. Formula 4.3 shows the a is the real output through the network by activation function σ active. y is the real label in the data set.

$$z = \sum w_j x_j + b \tag{4.2}$$

$$a = \sigma(z) \tag{4.3}$$

$$C = -\frac{1}{n}[y \ln a + (1-y)\ln(1-a)] \tag{4.4}$$

In the process of training the neural network, update w b by gradient descent algorithm, calculate the derivative of the loss function to w and b . The process can be expressed as formula 4.5, 4.6.

$$\frac{\partial C}{\partial w} = \frac{1}{n}\sum_x (\sigma(z) - y) \tag{4.5}$$

$$\frac{\partial C}{\partial b} = \frac{1}{n}\sum_x (\sigma(z) - y) \tag{4.6}$$

Utilizing formula 4.5 and 4.6 can update w and b in the next forward processing. Also, the influence and effectiveness of different depth in the proposed framework will be discussed in the next chapter.

(4) Evaluating the model

The case bearing diagnosis problem is naturally a multi-classification problem (in our case, it has 10 different states). Adopt the F1 score[53] for evaluating the trained model.

F1 score can be expressed as shown in formula 4.7. Generally, using the F1 score to evaluate the binary-classification. Precision is the percentage of predicting accurate classifications can be expressed in formula 4.8. The recall is the percentage of related results. The precision and recall can be expressed as shown in formula 4.8 and 4.9. TP is the true positive rate (The prediction is positive and the real label is positive also). FP is the false positive rate (prediction is positive but the real label is negative). TN is the true negative rate (prediction is negative and the real label is negative also). FN is the false negative rate (prediction is negative but the real label is positive). Also, to evaluate the whole multi-classification ability, adopting the accuracy which was expressed as shown in formula 4.10. False positive rate is important for fault diagnosis because of different fault caused by different reasons. So, this study adopted FPR to evaluate the model which was expressed as shown in formula 4.11. All of the aforementioned indexes are for each class. For whole multi-classification effectiveness, compute the meaning of accuracy of each class.

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{4.7}$$

$$precision = \frac{TP}{TP + FP} \tag{4.8}$$

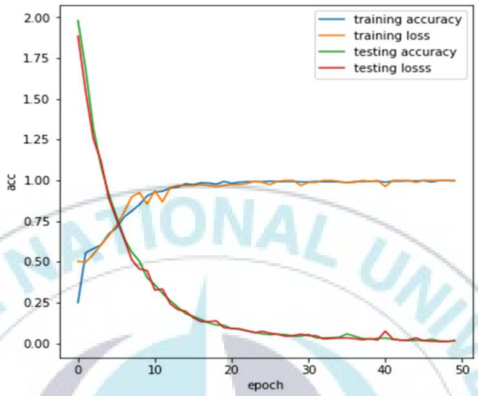
$$recall = \frac{TP}{TP + FN} \tag{4.9}$$

$$Accuracy = (TP + TN)/(TP + TN + FN + FP) \tag{4.10}$$

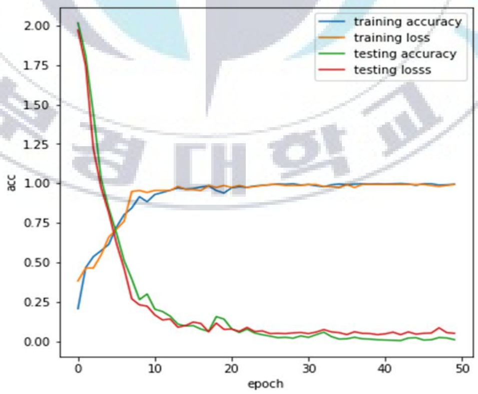
$$FPR = FP/(FP + TN) \tag{4.11}$$

The results using improved CNN in this case study as shown in Figure 4.4 (a). And the results using normal CNN which has 9 layers as shown in Figure 4.4 (b). It can be seen both of normal CNN based method and improved CNN based method testing accuracy became stable after 20 epoch, and there is no overfitting. However, improved CNN method is more stable than normal CNN based method. And the results show that the accuracy of improved CNN is up to 1 after 21 epoch. But for the normal CNN, the accuracy it got is 0.9874 after 50 epoch. And the high accuracy it got is 0.9874 during the whole 50 epoch. Which means proposed CNN based method takes less time can get higher accurate results. The loss of normal CNN based method is 0.0678 after 50 epoch in the testing data set. The proposed improved CNN based method is 0.0151.

The confusion matrix was given in Figure 4.5. (a) is improved CNN based method, (b) is the normal CNN based method. In Figure 4.5, the fault types are represented as OF1, IF1, BF1, OF2, IF2, BF2, OF3, IF3, BF3. OF1 means outer ring fault in diameter1, IF1 is inner ring fault in diameter1, BF1 is ball fault in diameter1. The others are the same. And NO is no fault. Both of normal CNN and improved CNN can identify the types of NO, OF1, BF1, OF2, BF2, OF3, IF3, BF3. But normal CNN cannot identify IF1, IF2 very well. The fault positive rate (FPR) of IF1 is 9.09%, and FPR of IF2 is 14.29%. It means normal CNN based method usually confused about IF1 and IF2. Improved CNN based method performed very well for bearing fault diagnosis.

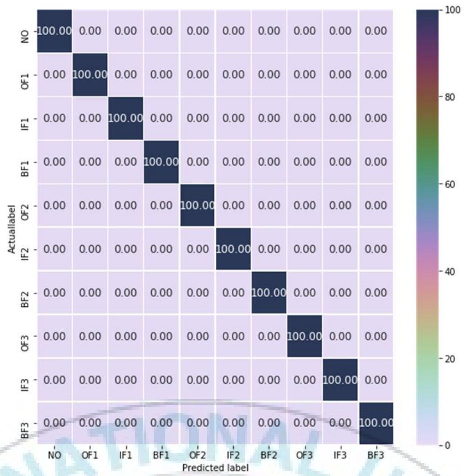


(a)

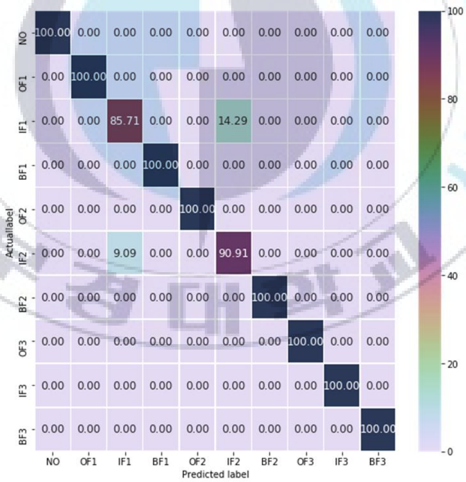


(b)

Figure 4.4 The accuracy of normal CNN and improved CNN based method for bearing fault diagnosis



(a)



(b)

Figure 4.5 The confusion matrix of normal CNN and improved CNN based method

The F1 score was given in Figure 4.6. The green bar is for improved CNN, the orange bar is for CNN. It can be seen easily improved CNN F1_score is 1 for all of the fault types. However, CNN only is up to 0.8517 in IF1, 0.9091 in IF2.

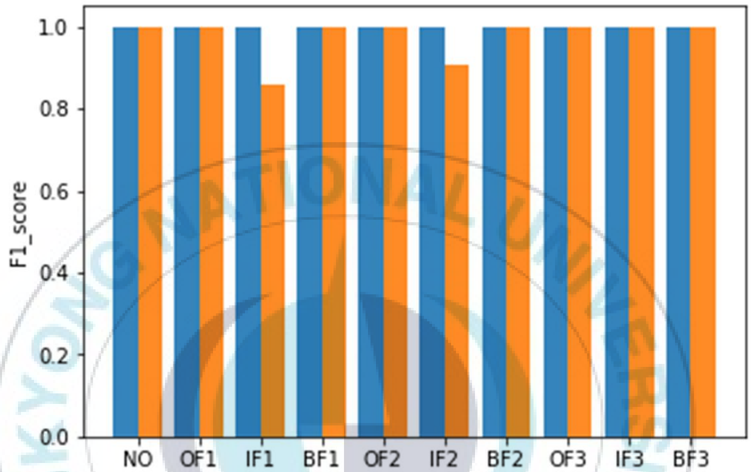


Figure 4.6 The F1 score for each bearing fault types using improved CNN and normal CNN based method

4.1.3 Comparing With Other Methods for Bearing Fault Diagnosis

The proposed improved CNN based fault diagnosis has been discussed in section 4.1.2. It proves that Improved CNN based method is better than the normal CNN based method. In order to prove the overall superiority, there give some compared results with other deep learning methods（DNN）. Also comparing with the traditional machine learning methods such as SVM, Random Forest, and combined PCA-SVM, PCA-Random Forest. The results as shown in Table 4.3. CNN-10 means it has 10 layers in networks, CNN-11 means it has 11 layers in the network; DNN-512-218-64-10 means it has three hidden layers in the network, and the nodes are 512, 218, 64, 10 means the last output (the numbers of classes). DNN-512-128-10 means it has two hidden layers in the network.

From Table 4.3, Traditional machine learning algorithms perform poorly. SVM totally cannot identify OF1, IF2, and cannot perform well for other fault types. After extracting the principal components by PCA (the set parameter is 0.95), combined PCA-SVM improve a little, but it still cannot identify IF2, and the performance is poor. The accuracy of SVM is 0.3459, PCA-SVM is 0.3584 from Table 4.4. Comparing with SVM, Random Forest is better than SVM, the accuracy of Random Forest is 0.3962, PCA-Random Forest is 0.4716. But Random Forest cannot identify IF2, IF3. Combined PCA-Random Forest is still poor for fault diagnosis. Because of the samples is not enough for traditional machine learning methods, they cannot perform well for fault diagnosis.

Table 4.3 The results of bearing fault diagnosis comparing with other methods.

Fault types	NO	OF1	IF1	BF1	OF2
Improved CNN	1.0	1.0	1.0	1.0	1.0
CNN-9	1.0	1.0	0.86	1.0	1.0
CNN-10	1.0	1.0	0.67	0.95	1.0
CNN-11	1.0	0.90	0.85	0.90	1.0
DNN-512-256-128-10	1.0	0.59	0.2	0.56	0.53
DNN-512-128-10	1.0	0.56	0.31	0.30	0.44
SVM	0.52	0.0	0.17	0.11	0.44
Random Forest	0.79	0.13	0.27	0.13	0.11
PCA-SVM	0.70	0.17	0.10	0.15	0.42
PCA-RF	1.0	0.24	0.15	0.22	0.25
Fault types	IF2	BF2	OF3	IF3	BF3
Improved CNN	1.0	1.0	1.0	1.0	1.0
CNN-9	0.91	1.0	1.0	1.0	1.0
CNN-10	0.87	1.0	1.0	1.0	1.0
CNN-11	0.91	1.0	1.0	1.0	1.0
DNN-512-256-128-10	0.31	0.28	0.53	0.84	0.35
DNN-512-128-10	0.17	0.24	0.5	0.7	0.42
SVM	0.0	0.1	0.30	0.15	0.13
Random Forest	0.0	0.14	0.46	0.0	0.15
PCA-SVM	0.0	0.4	0.25	0.23	0.06
PCA-RF	0.0	0.0	0.43	0.70	0.36

For CNN based method, CNN is good at processing the vibration signals, both of accuracy and F1 score is much higher than other methods. However, with the depths increasing, the results do not improve in normal CNN. The accuracy of CNN-9 is 0.9874, CNN-10 is 0.9748, and the CNN-11 is 0.9748. Using the normal CNN based method, how to find the suitable depth of network is difficult. And the improved CNN based method got an outstanding performance for bearing fault diagnosis.

Table 4.4 The accuracy of different methods for bearing fault diagnosis

Methods	Accuracy
Improved CNN	1.0000
CNN-9	0.9874
CNN-10	0.9748
CNN-11	0.9748
DNN-512-256-64-10	0.6038
DNN-512-128-10	0.5660
SVM	0.3460
PCA-SVM	0.3580
Random Forest	0.3960
PCA-Random Forest	0.4716

At last, giving the FPR of bearing fault diagnosis using different methods in Table 4.3. The results show in Figure 4.7. The improved CNN method has the lowest FPR, it is up to 0. It means it has no error for bearing fault diagnosis. And when using normal CNN, CNN-9 has the lower FPR than CNN-10, CNN-11. DNN based method get better performance than the traditional machine learning method. With the depth of CNN layer increasing, may not get the higher performance (the lower FPR) for bearing fault diagnosis.

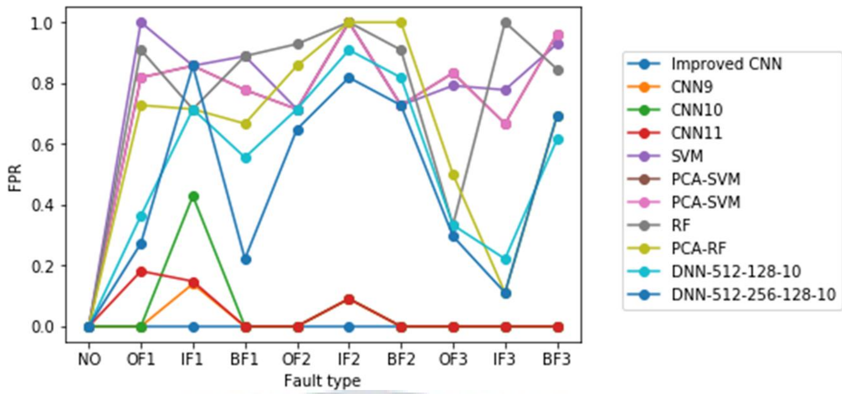


Figure 4.7 The FPR using different methods for bearing fault diagnosis

4.2 Case Study 2: Semiconductor Manufacturing Process Fault Diagnosis

4.2.1 Semiconductor Manufacturing Process Data Description

The data set presented in this case is about the semiconductor manufacturing process signals collected from sensors[54]. Each example represents a single production entity with the associated measured features and the labels represent a simple pass/filed for in housing line testing. The fault status -1 corresponds to represent to a pass and 1 corresponds to a fail. The data consisting of 1567 examples each with 591 features. And each sample has one corresponding label. As with any real-life data situations this data contains missing values varying in intensity depending on the individual's features.

4.2.2 Semiconductor Manufacturing Process Fault Detection Using Improved CNN

In this case, the steps are the same as the aforementioned steps described in 4.1.2. But something is different. Adopting the type of input tensor is [59,10]. And first adopting the mean of each feature to instead of the missing values which existed in each column. We only use one sum layer can get the best results using the proposed CNN based method. The utilized structure as shown in Figure 4.8, adopt one additional layer between the input layer can extract the features totally.

The training and testing results of semiconductor manufacturing process fault detection using improved and normal CNN based method as shown in Figure 4.9. Figure 4.9 (a) is for improved CNN method, (b) is for normal CNN based method. Improved CNN based method get the accuracy is 1.0 from the first epoch in testing data set. But improved CNN forms the second epoch get 1.0 of accuracy. It means improved method has a stronger feature extraction ability than normal CNN based method. (Especially, this normal CNN based CNN method adopts 5 layers, contains two convolution layer, one Max pooling layer, one flatten layer, one flatten layer). Accuracy in training and testing data set is the same, and the loss is the same also using improved CNN based method.

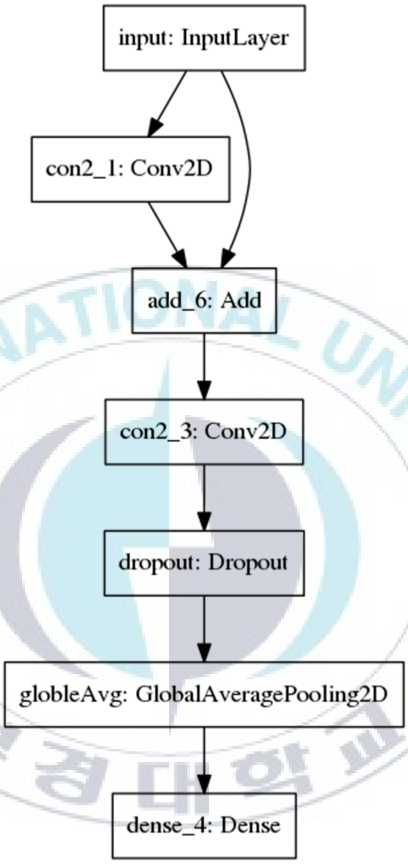
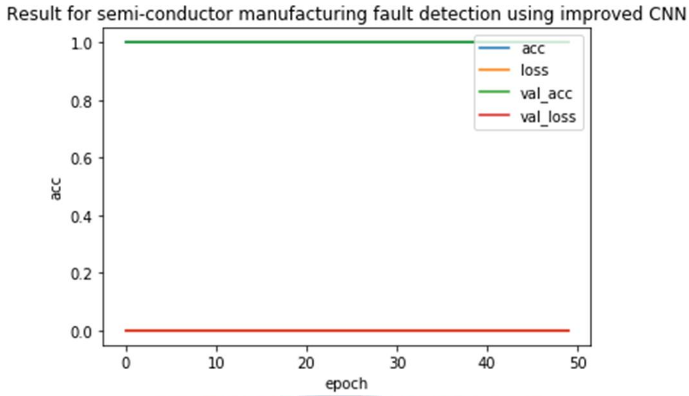
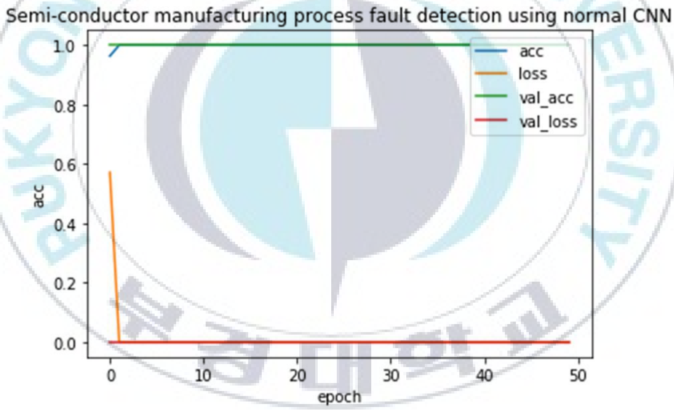


Figure 4.8 The structure of semiconductor manufacturing process fault detection using improved CNN



(a)



(b)

Figure 4. 9 The result of semiconductor manufacturing process detection using improved CNN and normal CNN based method.

4.2.3 Comparing With Other Methods for Semiconductor Manufacturing Process

As the steps in last section 4.1.3, also compared improved CNN based method for semiconductor manufacturing process detection with other methods such as normal CNN, DNN, SVM, PCA-SVM, Random Forest, PCA-Random Forest.

Table 4.5 The accuracy of different methods for semiconductor manufacturing fault detection

Methods	Accuracy	Time per epoch
Improved CNN	1.0000	4s
CNN-5	1.0000	1s
DNN-512-256-64-2	1.0000	9s
DNN-512-128-2	1.0000	5s
SVM	0.9102	-
PCA-SVM	0.9200	-
Random Forest	0.9267	-
PCA-Random Forest	0.9299	-

The deep learning methods perform very well for the semiconductor manufacturing process, Improved CNN, normal CNN, DNN-512-256-64-2, DNN-128-2 got accuracy is 1.0. But different methods consumed different time, the normal CNN used the least time per epoch, but the first epoch is not up to 1.0. Have compared in last section 4.2.2. And DNN based method consumed more time to the trained model.

For traditional machine learning based method, the accuracy is around 0.92, it is lower than deep learning based method. Can conduct that the data set is more complex, the deep learning method is more suitable for fault diagnosis. Especially, proposed improved CNN based fault diagnosis show the outstanding performance among these methods. Previous data set shape is [100, 60], it is more complex than semiconductor manufacturing process data set. The ability of feature extraction obviously can be seen. And the FPR of semiconductor manufacturing fault detection using improved CNN is 0.0, because of it predicted values are totally right.

4.3 Case Study 3: Air Pressure System Failure Detection

4.3.1 Air Pressure System Data Description

This data set obtained from 0 APS failure and Operational Data for Scania Trucks. The data set consists of data collected from heavy Scania trucks in everyday usage. The system in focus is the Air Pressure System (SAS) which generates pressurized air that is utilized in various functions in a truck, such as braking and gear changes0. It has two classes: a negative class and positive class. The positive class consists of the component of failures for a specific component of the APS system. The negative class consists of trucks with failures for components not related to the APS. The data shape has 60000 examples and has 171 attributes. And have many missing values in the data set.

4.3.2 Air Pressure Fault Detection Using improved CNN

Same to the previous two case study, using the improved CNN based method for fault diagnosis. The training and testing accuracy and loss as shown in Figure 4.10. From the figure, know both of training accuracy and testing accuracy are 0.9766, and the loss of training and testing are the same form the iteration 1 to iteration 50. Using improved CNN based method for air pressure system fault detection, only need train one iteration can get the best results. The normal CNN based method get the final accuracy is 0.9501. It is lower than the proposed improved CNN based method.

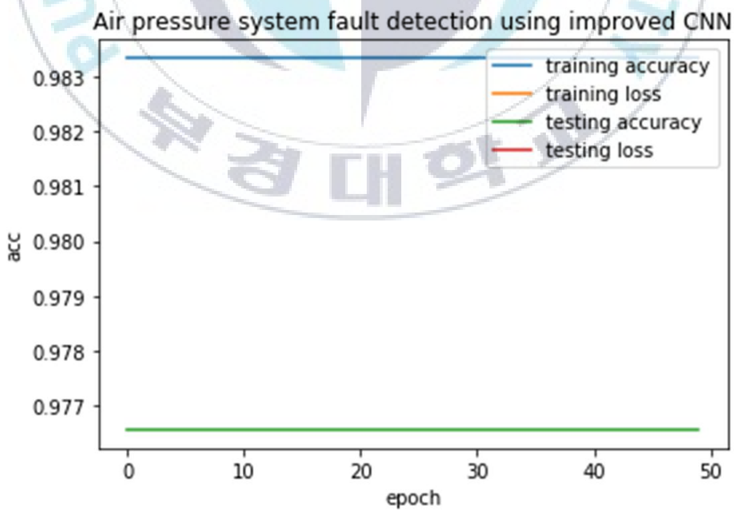


Figure 4.10 The results of air pressure system fault detection using improved CNN

4.3.3 Comparing With Other Methods for Air Pressure Fault Detection

This data set contains too many NA values. So the traditional machine learning method cannot process directly, need to transform NA values into some digits using some special method. But how to transform the data is a big problem, it is needing to think more.

Chapter 5.

Conclusion

Fault diagnosis is more and more important during the industry manufacturing processing. Using less time to do fault diagnosis accurately is the final purpose. However, traditional methods are relying on experts, and experience, massive knowledge. With the rapid development of AI technology, data-driven based method fault diagnosis has been a hot topic. But there is still existing such as much time consumption, the lowest accuracy for fault diagnosis. To overcome these drawbacks, this study proposed a new approach based deep CNN using image processing technology.

Different from the normal CNN based method, the proposed method did not adopt pooling layer to reduce the dimension of feature map after each convolution layer and did not adopt the traditional flatten layer to convert the feature map into a feature vector for fault diagnosis. Instead of adopting the global average pooling layer at the last layer for converting the feature map into vector directly, thus, it reduces the more time to train the model. The key idea of the proposed method is adopting add a layer between the previous convolution layer and the next convolution layer, the output of the added layer will be used as the input of the next layer. The experiments proved that by adding the two layers, can extract the hidden feature fully. Using three different public data set for testing the proposed CNN based method. The results of accuracy are 1.0, 1.0, 0.9766. Comparing with normal CNN based method, it shows a great improvement in the accuracy and time cost. Also comparing with traditional machine learning methods such as SVM, PCA-SVM , Random Forest , PCA-Random Forest, the improved method exhibits a giant preference. Especially improved CNN based method can be used for data which has a great of missing values in the data set which caused by machine or environment. It means improved CNN based method is stable. That is very important for fault diagnosis, because of the complex and changing production environment.

However, the results show the improved CNN based method used more time to train model due to did not adopt the pooling layer to reduce the dimension, in the featured study, will focus on adjusting the structure of CNN based method to overcome spending much more time to train the model. By changing the classifier of the last layer and the definition of the loss function can get the highest accuracy for fault diagnosis early also.

References

[1] Heng S. Industry 4.0: Huge potential for value creation waiting to be tapped. Deutsche Bank Research.

[2] The Industrial Internet Consortium: A Global Nonprofit Partnership of Industry, Government and Academia, 2014, http://www.iiconsortium.org/about-us.html.

[3] Premier of the State of China and K.Q.Li, "Report on the work of the government," in *Proceedings of 3rd Session of the 12th National People's Congress*, March 2015.

[4] Zhang Lijuan, Song Weiwei. How to deploy the intelligent manufacturing industry in Korea. Science and Technology China (4), 2016, 42-45.

[5] Wang, S. et al. Implementing Smart Factory of Industry 4.0: An Outlook, *International Journal of Distributed Sensor Networks*. 2016, doi: 10.1155/2016/3159805.

[6] Li Fu, Yaozu Fan. Research on intelligent track prediction system [J]. *J of Aeronautics*, 2000,(4):299-302.

[7] Han Xiao, Deyun Xiao. Overview of data-driven fault diagnosis methods [J]. *J of control and decision*, 2011, 26(1): 1-9.

[8] Kui Wu, Xianyong Wang, Jie Sun et al. Fault Detection Based on Deep Learning[J]. *J of Computer Measurement and Control*, 2017, 35(3): 481-495.

[9] Yu Peng, Datong Liu. Summary of data-driven fault prediction and health management[J]. *J of mechanical engineering*, 2013,

49(1): 63-72.

[10] Gao H, Liang L, Chen X, et al. Feature extraction and recognition for rolling element bearing fault utilizing short-time Fourier transform and non-negative matrix factorization[J]. *Chinese Journal of Mechanical Engineering* (English Edition), 2015, 28(1):96-105.

[11] G Chen, L Qie, A Zhang, et al. Improved CICA algorithm used for single channel compound fault diagnosis of rolling bearings. *Chinese Journal of Mechanical Engineering*, 2016, 29(1): 204–211.

[12] M Riera-Guasp, J A Antonino-Daviu, G A Capolino. Advances in electrical machine, power electronics, and drive condition monitoring and fault detection: state of the art. *IEEE Transactions on Industrial Electronics*, 2015, 62(3):1746–1759.

[13] M H Drif, A J Cardoso. Stator fault diagnostics in squirrel cage three-phase induction motor drives using the instantaneous active and reactive power signature analyses. *IEEE Transactions on Industrial Informatics*, 2014, 10(2):1348–1360.

[14] Y Wang, F Zhang, T Cui, et al. Fault diagnosis for manifold absolute pressure sensor (MAP) of diesel engine based on Elman neural network observer. *Chinese Journal of Mechanical Engineering*, 2016, 29(2): 386–395.

[15] J Antonino-Daviu, S Aviyente, E G Strangas, et al. Scale invariant feature extraction algorithm for the automatic diagnosis of rotor asymmetries in induction motors. *IEEE*

Transactions on Industrial Informatics, 2013, 9(1): 100–108.

[16] J Faiz, V Ghorbanian, BM Ebrahimi. EMD-based analysis of industrial induction motors with broken rotor bars for identification of operating point at different supply modes. *IEEE Transactions on Industrial Informatics*, 2014, 10(2): 957–966.

[17] Lv, Feiya , et al. "Fault diagnosis based on deep learning." *American Control Conference American Automatic Control Council (AACC)*, 2016.

[18] Shao S Y, Sun W J, Yan R Q, et al. A Deep Learning Approach for Fault Diagnosis of Induction Motors in Manufacturing[J]. *Chinese Journal of Mechanical Engineering*, 2017, 30(6):1347-1356.

[19] GeM, Du R, Zhang G, et al. Fault diagnosis using support vector machine with an application in sheet metal stamping operations[J]. *Mechanical Systems & Signal Processing*, 2004, 18(1):143-159.

[20] Kankar PK, Sharma SC, Harsha S P. Fault diagnosis of ball bearings using machine learning methods[J]. *Expert Systems with Applications*, 2011, 38(3):1876-1886.

[21] Ge M, Du R, Zhang G, et al. Fault diagnosis using support vector machine with an application in sheet metal stamping operations[J]. *Mechanical Systems & Signal Processing*, 2004, 18(1):143-159.

[22] Manjurul Islam MM, Kim J M. Reliable multiple combined

fault diagnosis of bearings using heterogeneous feature models and multiclass support vector Machines[J]. *Reliability Engineering & System Safety*, 2018: S0951832017309730.

[23] Maragoudakis M , Loukis E , Pantelides P P . Gas Turbine Fault Diagnosis using Random Forests[C]// *Conference on Ecai: European Conference on Artificial Intelligence*. IOS Press, 2008.

[24] Shrivastava R , Mahalingam H , Dutta N N . Application and Evaluation of Random Forest Classifier Technique for Fault Detection in Bioreactor Operation[J]. *Chemical Engineering Communications*, 2017, 204(5):591-598.

[25] Shah A M , Bhalja B R . Fault discrimination scheme for power transformer using random forest technique[J]. *IET Generation, Transmission & Distribution,* 2016.

[26] Bengio, Y. Learning Deep Architectures for AI[J]. Foundations and Trends? in Machine Learning, 2009, 2(1):1-127.

[27] Jia Y, Shelhamer E, Donahue J, et al. Caffe: *Convolutional Architecture for Fast Feature Embedding[J]*. 2014.

[28] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. 2015.

[29] Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions[C]// *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2015.

[30] Cai Y, Wang H, Chen X, et al. Vehicle detection based on visual saliency and deep sparse convolution hierarchical

model[J]. *Chinese Journal of Mechanical Engineering*, 2016, 29(4):765-772.

[31] Lecun Y, Bengio Y, Hinton G . *Deep learning[J]. Nature*, 2015, 521(7553):436.

[32] Deng L, Hinton G, Kingsbury B. New types of deep neural network learning for speech recognition and related applications: an overview[C]// *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013.

[33] Zhang Q, Yang L T, Chen Z. Deep Computation Model for Unsupervised Feature Learning on Big Data[J]. *IEEE Transactions on Services Computing*, 2015:1-1.

[34] Wei-Hua LI, Shan WP, Zeng X Q. Bearing fault identification based on deep belief network[J]. *Journal of Vibration Engineering*, 2016.

[35] Zhang Z, Zhao J . A deep belief network based fault diagnosis model for complex chemical processes[J]. *Computers & Chemical Engineering*, 2017: S0098135417301059.

[36] H. Zhao, S. Sun and B. Jin, "Sequential Fault Diagnosis Based on LSTM Neural Network," in *IEEE Access*, vol. 6, pp. 12929-12939, 2018.

[37] Wu H , Zhao J . Deep convolutional neural network model based chemical process fault diagnosis[J]. *Computers & Chemical Engineering*, 2018.

[38] Wei D, Gong Q, Li W, et al. Research on internal and external

fault diagnosis and fault-selection of the transmission line based on the convolutional neural network[J]. *Proceedings of the CSEE*, 2016.

[39] Chen Z Q, Li C, Sanchez R V. Gearbox Fault Identification and Classification with Convolutional Neural Networks[J]. *Shock and Vibration*,2015,(2015-10-13), 2015, 2015(2):1-10.

[40] Wen L, Li X, Gao L, et al. A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method[J]. *IEEE Transactions on Industrial Electronics*, 2018, 65(7):5990-5998.

[41] Guoqian J, Haibo H, Jun Y, et al. Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox[J]. *IEEE Transactions on Industrial Electronics*, 2018:1-1.

[42] Hao Ren, Jianfeng Qu, Yi Chai, et al. Research status and challenges of deep learning in the field of fault diagnosis[J]. *J of Control and Decision*, 2017, 32(8): 1345-1358.

[43] Peng H K, Marculescu R. Multi-scale compositionality: identifying the compositional structures of social dynamics using deep learning[J]. Plos One, 2015, 10(4):e0118309.

[44] Chen Z Q, Li C, Sanchez R V. Gearbox Fault Identification, and Classification with Convolutional Neural Networks[J]. *Shock and Vibration*,2015,(2015-10-13), 2015, 2015(2):1-10.

[45] Chen Xianchang. Research on algorithm and application of deep learning based on the convolutional neural network[D]. Hangzhou：Zhejiang Gongshang

University, 2014(in Chinese)

[46] https://en.wikipedia.org/wiki/Sigmoid_function

[47] https://en.wikipedia.org/wiki/Softmax_function

[48] Jake Bouvrie. Notes on convolutional neural networks. Technical report, 2006.

[49] Lin Min, Chen Qiang, Yan, Shuicheng. Network In Network. 2014, 1–10.

[50] http://madnet.org:8765/explore.html?id=11&t=0.5290157372286013

[51] Harris, David, and Harris, Sarah. Digital design and computer architecture (2nd ed.). San Francisco, Calif.: Morgan Kaufmann. p. 129. ISBN 978-0-12-394424-5.

[52] Wikipedia.org/wiki/Rectifier_(neural_networks)

[53] He H, Garcia E A. Learning from Imbalanced Data[J]. *IEEE Transactions on Knowledge & Data Engineering*, 2009, 21(9):1263-1284.

[54] Madnet.org:8756/explore.html?id=11&t=0.881342595095301

[55] https://archive.ics.uci.edu/ml/machine-learning-databases/00421.

Acknowledgments

The writing of the graduation thesis is nearing completion, and the two-year study abroad life is coming to an end. In the past two years, not only learned a lot of professional knowledge, but also met many talented and interesting people. Looking back, the past is still there. Every minute I experienced in the Pukyong national university happily every day. There is no banquet in the world, I am only grateful at this moment. Thank you for appearing in my life.

First of all, I would like to say thank you to my adviser professor Chang-soo Kim. Being your student is my glory. Can't forget the first time we met. When I first arrived in Korea, everything seemed so strange. It is your smile that dispels my inner tension and uneasiness; Can't forget the time we discussed the paper with you, you require me to present about related work so that I can be more confident to present right now. Can't forget every week's seminar, your inspiration and guidance make my thoughts clear and open; Can't forget to go abroad with the professor to attend the conference. Thank you for taking me to open my eyes. Every time you leave the lab, you will say goodnight to us. I feel how warm it is in Busan.

In the whole process of completing the graduation thesis, I would like to thank Dr. Huang for her inspiration at the group meeting. Thanks to every laboratory partner, you are the source of happiness

for my study, especially Palash. I have known for two years and have an international friendship.

Then, I would like to thank all the instructors, it is my pleasure to get your cultivation in student life. On the road to the academic hall, you are a bright light that illuminates me. Thank you again for your enthusiasm and guidance.

Once again, I want to thank my girlfriend for her support and understanding. I have been in love for five years with you but I studied abroad for two years. Thank you.

Finally, I want to thank my families. Without you, there is no achievement I have achieved today. I am grateful to my parents for their half-hearted efforts. Your deep love for your son is pale in any words in the world. Thanks to my loved ones, you are against me. Care and tolerance made me realize that my blood is thicker than water.

I will send my most sincere wishes to my professors and classmates. Hope all of you happy every day.

Because of my shallow knowledge, there are many shortcomings in my graduation thesis, and I hope you will not hesitate to give me your advice.