

## Article

# An Improved Fault Diagnosis Approach for Pumps Based on Neural Networks with Improved Adaptive Activation Function

Fangfang Zhang <sup>1</sup>, Yebin Li <sup>1</sup>, Dongri Shan <sup>2,3,\*</sup>, Yuanhong Liu <sup>4</sup> and Fengying Ma <sup>1</sup>

<sup>1</sup> School of Information and Automation Engineering, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250300, China; zhff4u@qlu.edu.cn (F.Z.); 10431210506@stu.qlu.edu.cn (Y.L.); mfy@qlu.edu.cn (F.M.)

<sup>2</sup> School of Mechanical Engineering, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250300, China

<sup>3</sup> System Control and Information Processing Lab, Aerospace Information University, Jinan 250200, China

<sup>4</sup> School of Information and Electrical Engineering, Northeast Petroleum University, Daqing 163319, China; liuyuanhong@nepu.edu.cn

\* Correspondence: shandongri@qlu.edu.cn

**Abstract:** Due to the complex underground environment, pumping machines are prone to producing numerous failures. The indicator diagrams of faults are similar to a certain degree, which produces indistinguishable samples. As the samples increase, manual diagnosis becomes difficult, which decreases the accuracy of fault diagnosis. To accurately and quickly judge the fault type, we propose an improved adaptive activation function and apply it to five types of neural networks. The adaptive activation function improves the negative semi-axis slope of the Rectifying linear unit activation function by combining the gated channel conversion unit to improve the performance of the deep learning model. The proposed adaptive activation function is compared to the traditional activation function through the fault diagnosis data set and the public data set. The results show that the activation function has better nonlinearity and can improve the generalization performance of the deep learning model and the accuracy of fault diagnosis. In addition, the proposed adaptive activation function can also be well-embedded into other neural networks.

**Keywords:** deep learning; fault diagnosis; adaptive activation function; pumping unit



**Citation:** Zhang, F.; Li, Y.; Shan, D.; Liu, Y.; Ma, F. An Improved Fault Diagnosis Approach for Pumps Based on Neural Networks with Improved Adaptive Activation Function. *Processes* **2023**, *11*, 2540. <https://doi.org/10.3390/pr11092540>

Academic Editor: Jie Zhang

Received: 1 August 2023

Revised: 21 August 2023

Accepted: 23 August 2023

Published: 24 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The fault diagnosis of the pumping unit in the process of petroleum collection has been a critical research topic. Due to the complex underground environment, during the reciprocating movement of the sucker rod, there are many unknown factors that are prone to result in the failure of the pumping machine and then form a safety hazard. Load (P) and displacement (S) are the parameters generated when the donkey head of the pumping unit moves up and down, and the closed curve formed by them is the indicator diagram. It can reflect the influence of gas, oil, water, sand, wax, and other factors on the pumping unit in real time [1]. If the pump is in the fault state for a long time, the wear of the pump will be aggravated, and the service life of the equipment will be further reduced. Therefore, the fault should be quickly and accurately detected so as to take corresponding fault tolerant control measures to ensure the normal operation of the equipment [2,3].

The traditional method for fault diagnosis of a pumping unit is to measure the load change with displacement at the suspension point, draw the suspension point indicator diagram, and then diagnose the working condition of the pumping unit according to the shape of the indicator diagram. The disadvantages of the traditional method are as follows: first, the fault of the pumping unit is judged mainly by the method of manual identification of indicator diagram, which has a great influence on human factors and low recognition accuracy. Secondly, due to the large number and wide distribution of

pumped wells, manual detection of faults is time-consuming and laborious. However, given the complexity and high cost of pump units, there is less tolerance for performance degradation, productivity decrease, and safety hazards; therefore, it is necessary to detect and identify all potential faults rapidly [4,5]. This means that it is imperative to replace the manual fault diagnosis of the pumping unit with a computer.

With the continuous progress of deep learning technology, using deep learning technology for fault diagnosis has become a new trend. For example, Convolution Neural Network [6], Generation Adversarial Network [7], and Long Short Term Memory [8] et al. show superior performance in fault diagnosis.

At present, the deep learning technology used in fault diagnosis of pumping units is to classify the indicator diagram of different kinds of faults. Automatic features extracted from raw data are an outstanding advantage of deep learning technology and will not depend on the diagnostic knowledge of specialists [9].

In 2018, Y. Duan [10] proposed an improved Alexnet model to realize the automatic recognition of indicator graphs and compared it with the current common neural network model. In 2019, J. Sang [11] proposed a PSO-BP neural network algorithm aiming at the problems of slow convergence and unstable results of the traditional BP neural network algorithm; they designed the adjustment rules of the inertia weight and learning factor of the PSO algorithm and adjusted the weight coefficient of the output layer and the hidden layer of the BP neural network algorithm. In 2020, L. Zhang [12] used Freeman chain code and differential code to extract the characteristics of dynamometer card data of the pumping unit group. Then, a diagnosis model based on the BP neural network was proposed, and the fault type of the pump group could be automatically identified according to the dynamometer card. In 2022, H. Hu [13] proposed a model based on the ResNet-34 residual network to identify the indicator diagrams, which added a residual block structure to the traditional convolution neural network to establish a direct connection between the upper layer input and the lower layer output and achieved the recognition and classification of six power diagrams through parameter adjustment. In the same year, T. Bai [14] proposed a fault diagnosis method based on a time series transformation generative adversarial network (TSC-DCGAN).

Because of the complexity of the pumps' working conditions, there are different shapes of indicator diagrams in different working states. The indicator diagrams of different kinds of faults are similar to a certain degree; thus, indistinguishable samples are produced. This will lead to the poor generalization ability of deep learning models and difficulty in navigating between indistinguishable samples. The function of the activation function is to carry out the nonlinear transformation of data and solve the problem of the insufficient expression and classification ability of the linear model. If the network is all linear transformation, then the multi-layer network can be directly converted into a layer of neural network through matrix transformation. Therefore, the existence of the activation function can make the deep learning model perform better with the increase in the number of layers. Therefore, we will propose a new activation function to improve the generalization performance of the deep learning model so that the faults of the pumping unit can be distinguished in a high dimensional space.

The calculation of the Sigmoid activation function is large, and the transformation of the Sigmoid saturation region is slow. The derivation approaches 0, resulting in the disappearance of the gradient. The output value of the Sigmoid function is always greater than 0, which results in a slow convergence rate of the training model. The Tanh activation function solves the zero-centered output problem, but the gradient disappearance problem and the power operation problem still exist. Rectifying linear unit (ReLU) [15], which has low computational complexity and fast convergence speed, can solve the problems of gradient disappearance and gradient saturation; there is also the phenomenon of Dead ReLU. In recent years, there have been many improved versions of ReLU (rectified linear unit). To solve the Dead ReLU phenomenon, the negative part of ReLU is substituted for a

non-zero slope, and Leaky ReLU (LReLU) [16] is proposed. Hence, LReLU is more inclined to activate in the negative area.

In deep learning, the selection of activation function is generally determined according to the specific situation, and there is no fixed choice. As the adaptive activation function can be automatically adjusted to adapt to the network structure and practical problems, it has been widely developed. Parametric Rectified Linear Unit (PReLU) [17] is also used to solve the Dead ReLU phenomenon. The slope of the negative part can be obtained by learning from the data rather than from defined fixed values. Therefore, PReLU has all the advantages of ReLU in theory and is more flexible than Leaky ReLU. In 2017, the Swish activation function was proposed. It has the characteristics of possessing a lower bound, no upper bound, and being non-monotonic. It is very smooth with its first derivative [18], and its performance is better than ReLU in many aspects. In 2021, H. Hu [19] proposed a new scheme to explore the optimal activation function with greater flexibility and adaptability by adding only a few parameters on the basis of traditional activation functions such as Sigmoid, Tanh, and ReLU. This method avoids local minima by introducing a few parameters into a fixed activation function. In the same year, M. Zhao [20] used the specially designed subnetwork of Resnet-APReLU as an embedded module in order to adaptively generate the multiplicative coefficient in nonlinear transformation.

Based on the above discussion, an adaptive activation function is designed with the gated channel Transmission Unit module (GCT) [21]. Compared with the traditional activation function, the adaptive activation function can effectively avoid gradient disappearance and Dead ReLU problems. Compared with the common adaptive activation function, the designed adaptive activation function combined with the GCT module can obtain global information through less computation, which further improves the performance of the deep learning model. The main contributions are as follows:

1. We propose an improved adaptive activation function. Each layer of deep learning generates different activation functions, improves the generalization performance of the deep learning models, and has strong adaptability to different deep learning models.
2. We apply the proposed activation function to the fault diagnosis of the pumping unit so as to better extract features from the contours of the indicator diagram. The proposed activation function improves the accuracy of fault diagnosis and has a better search ability, which is verified and compared with AlexNet [22], VGG-16 [23], GoogleNet [24], ResNet [25] and DenseNet [26].
3. The proposed activation function is extended to the public datasets CIFAR10, which proves that the proposed activation function is suitable and universal.

The rest of this paper is organized as follows. In Section 2, we introduce the pumping unit data set. In Section 3, we introduce the common adaptive activation function and propose the composition of our adaptive activation function. In Section 4, the experimental analysis and the discussion on the pumping unit failure data set and the public data set are presented. In Section 5, we conclude the paper.

## 2. Experiment Design and Measurement

### 2.1. Introduction to Pumping Unit

At present, about 80% of oil wells in most oil fields in China use rod pumping equipment, and the most widely used is the beam pumping unit [27]. The failure data set of the pumping unit comes from the real data generated by the pumping unit operation in a certain oil field in Northeast China. The pumping unit is a part of a rod pumping unit. Rod pumping equipment is mainly composed of three parts: an oil pumping unit, a well pumping pump, and a sucker rod. The rod pumping equipment is shown in Figure 1.

The pumping unit is driven by a motor, and through the reducer transmission system and the execution system, the rod and the pump plunger are driven to move up and down before, finally, the crude oil is lifted from the well to the surface. The operation of the pumping unit is shown in Figure 2.



**Figure 1.** Pumping machine equipment.



(a) Up stroke

(b) Down stroke

**Figure 2.** The operation of the pumping unit.

## 2.2. Fault Types of Pumping Unit

The fault data set of the pumping unit consists of nine types of indicator diagrams: normal, insufficient fluid supply, contains sand, piston stuck, gas influence, pump up touch, pump down touch, double valve leakage, and pumping rod detachment. The following details will be introduced:

### 1. Normal

The pump work diagram made by normal operation refers to the position shift of the end suspension point relative to the lower dead point as the transverse setting mark, the self-weight force of the rod, and the cumulative load received by the pump plug as the longitudinal setting mark. Drawn in parallel quadrilateral shape.

2. **Insufficient fluid supply**

The shortage of liquid supply is due to the insufficient amount of crude oil in the well, and the plunger pump inhales a large amount of air while drawing crude oil each time. As a result, a large amount of gas in the pump cannot be fully operated.
3. **Contains sand**

Because the well contains sand, the plunger creates an additional resistance in an area during movement. The additional resistance on the up stroke increases the load at the suspension point and on the down stroke at the same position. The increased resistance reduces the load at the suspension point. Because the distribution of sand particles in the pump barrel is not the same, its influence on the load varies greatly in various places, so it will lead to severe fluctuations in the load in a short time.
4. **Piston stuck**

When the pump plunger is stuck near the bottom dead point, the rod is in a stretched state during the up stroke and the down stroke since the whole stroke is actually the process of elastic deformation of the rod; the well work diagram at this time is approximately an oblique line.
5. **Gas interference**

Gas interference is a situation when the gas percent in the oil of the pumping well is high while the crude oil percent is relatively low. This causes the pump barrel to extract most of the gas, resulting in a significant difference between the actual load and the theoretical load.
6. **Pump down touch**

When the anti-impact distance is too large, the piston running up is approaching the upper dead point, and the continuous upward movement of the piston collides with the moving valve, which leads to the sudden loading of the piston and the bunching at the upper dead point.
7. **Pump up touch**

When the anti-impact distance is too small, it is attached to the lower dead point, and the piston moves down and collides with the fixed Val, resulting in sudden unloading of the piston and bunching at the lower dead point.
8. **Double valve leakage**

Double valve leakage refers to the situation where both the moving valve leakage and the fixed valve leakage happen at the same time, and the leakage may be caused by a combination of multiple faults.
9. **Pumping rod detachment**

The pumping unit's power cannot be transmitted to the pump due to the detachment of the sucker rod, resulting in the inability to extract oil.

The failure of the pumping unit will cause great economic losses and security risks. Therefore, rapid and accurate fault diagnosis of the pumping unit is very necessary. The fault diagnosis process in this paper is as follows: Firstly, the displacement and load data of the pumping unit are collected by a wireless dynamometer. Secondly, the indicator diagram of various faults is drawn from the collected data. Finally, the indicator diagram is preprocessed, and then the indicator diagram is input into the deep learning model to output the fault type. The fault diagnosis flow chart of the pumping unit in this paper is shown in Figure 3.

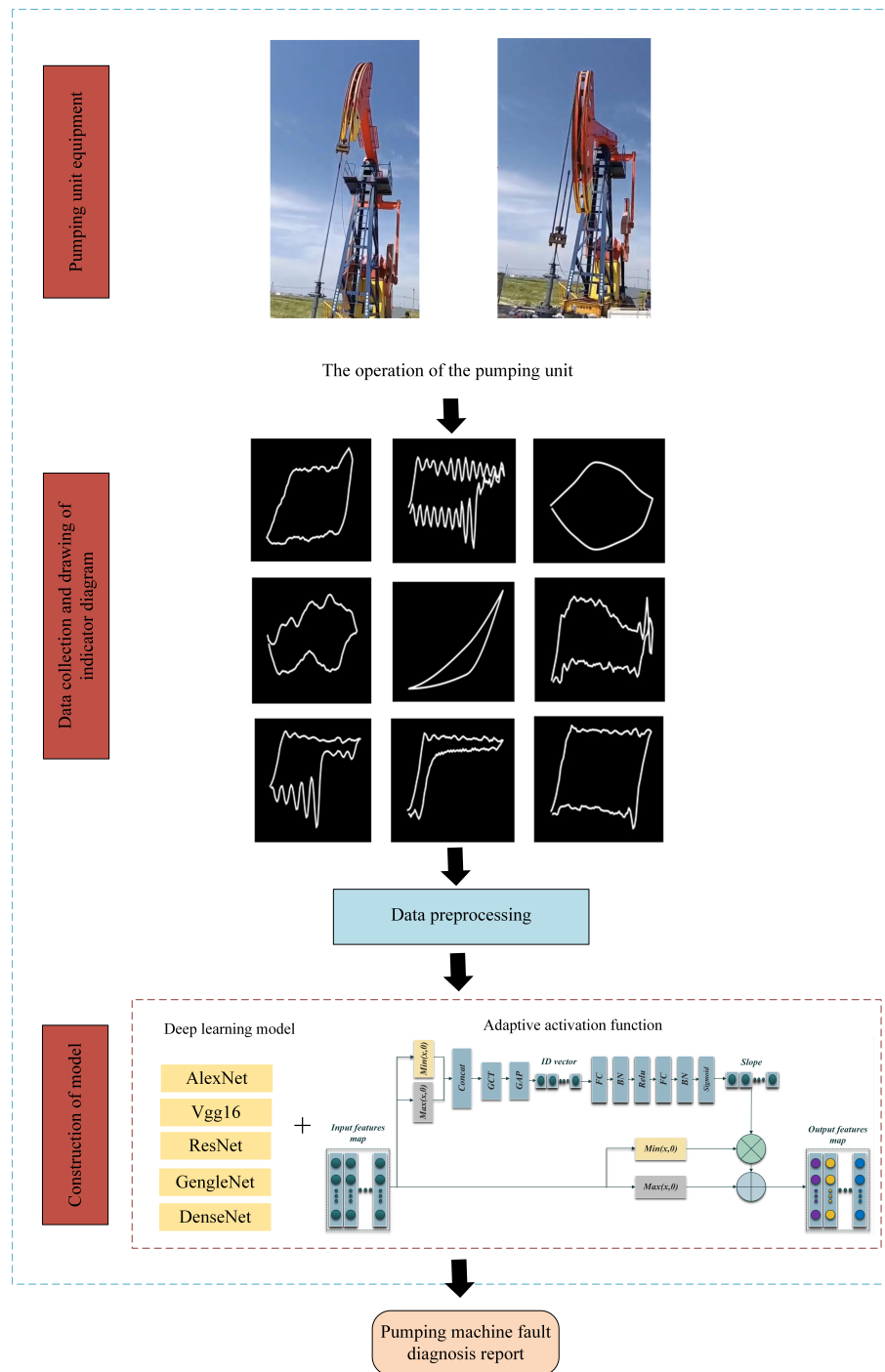


Figure 3. Fault diagnosis flow chart of the pumping unit.

### 3. Theoretical Analysis

#### 3.1. Common Adaptive Activation Functions

The PReLU activation function is a further improvement on the fixed predefined slope of LeakyReLU, which can be changed by backpropagation. It has better adaptability than LReLU. The formula of the activation function is as follows:

$$f(x) = \text{Max}(x, 0) + \delta \text{Min}(x, 0) \tag{1}$$

where  $x$  is the input,  $\delta$  is the trainable multiplicative coefficient (i.e., slope). Each layer has its own  $\delta$ , which improves the nonlinear capability. In PReLU,  $\delta$  in Equation (1) is the

learnable parameter during training, but it is  $\delta$  constant and cannot be adjusted during testing.

The design of the Swish activation function is inspired by the Long Short-term Memory neural network. The Swish activation function can prevent the gradient from gradually approaching zero and leading to saturation during training. It plays an important role in optimization and generalization. The formula of the Swish activation function is as follows:

$$f(x) = x \text{Sigmoid}(\zeta x) \tag{2}$$

where  $\beta$  is the learnable parameter or constant. When  $\zeta = 0$ , the Swish activation function becomes the linear function  $f(x) = x/2$ . When  $\zeta = \infty$ , the Swish activation function becomes 0 or  $x$ , which is equivalent to the ReLU activation function. Therefore, the Swish activation function can be considered as a smooth function between the linear function and the ReLU activation function.

Compared with ReLU, the Mish activation function is smoother at the origin [28]. The formula is as follows:

$$f(x) = x \tanh(\text{In}(1 + e^x)) \tag{3}$$

From Equation (3), the Mish Activation function has no upper limit, but only a lower limit, which can ensure no saturated region; thus, there will be no vanishing gradient during the training. At the same time, it has a faster convergence speed.

### 3.2. The Structure of Adaptive Activation Functions

The structure of adaptive activation functions is shown in Figure 4. The input of the subnetwork is concatenated by the one-dimensional vector obtained from the two inputs. The two inputs are positive features after separation and negative features after separation. The separation of positive and negative features can highlight the key features. The following calculation paths are GCT→GAP→FC→Batch Normalization(BN)→ReLU→FC→BN→Sigmoid→Scales. The function of each layer is described in the following section.

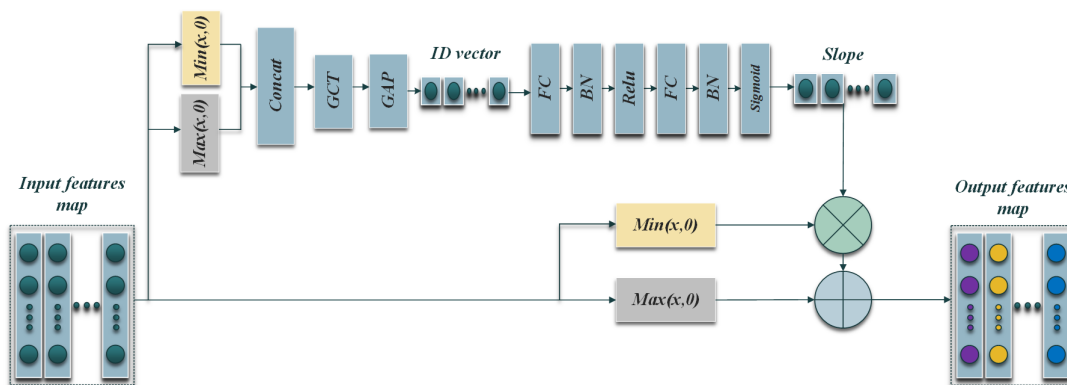


Figure 4. Graph of adaptive activation functions.

GCT combines normalization methods and attention mechanisms, which makes it easy to analyze the relationships (competition or cooperation) between channels. As shown in Figure 5, the GCT module introduces three trainable parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  to evaluate the communication channels. Among them,  $\alpha$  helps embed the output adaptive ability, while  $\beta$  and  $\gamma$  are used to control the activation threshold, which determines the behavior of GCT in each channel.  $h$  and  $w$  are the dimensions of feature vectors,  $c$  is the number of channels, and L2-norm is the normalization of L2.

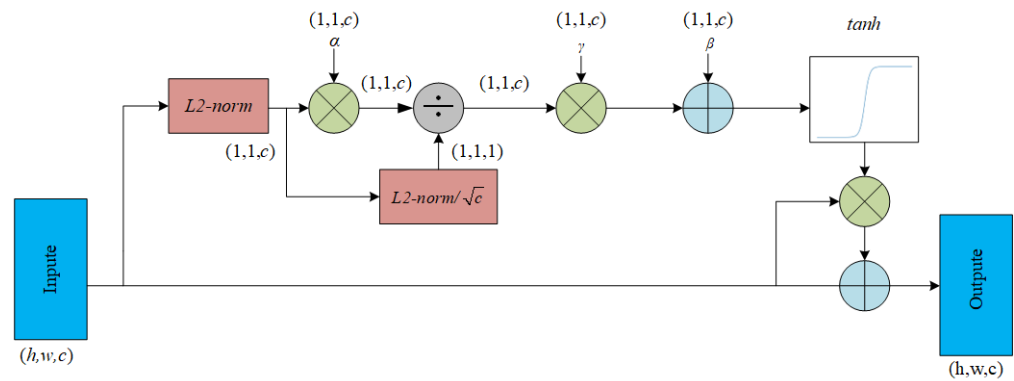


Figure 5. GCT structure drawing.

Global average pooling (GAP) can replace the fully connected (FC) layer to achieve dimensionality reduction. In particular, it retains the spatial information extracted from the previous convolution layers and pooling layers and can also strengthen the relationship between categories and feature maps [29].

ReLU is selected as the activation function of FC in the first layer to reduce the computational complexity and keep the gradient value within a reasonable range for feature extraction. The formula is as follows:

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (4)$$

Then, we add the BN layer, which is a way to unify the scattered data and is similar to normal data standardization. It is also a way to optimize the neural network. The data with unified specifications can make it easier to learn the rules in the data for the deep learning model [30] and can also solve the problem of vanishing gradient. The normalization is described by the following formula:

$$\mu = \frac{1}{N_{batch}} \sum_{i=1}^{N_{batch}} x_i \quad (5)$$

$$\sigma^2 = \frac{1}{N_{batch}} \sum_{i=1}^{N_{batch}} (x_i - \mu)^2 \quad (6)$$

$$\hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \quad (7)$$

$$y_i = \psi \hat{x}_i + \theta \quad (8)$$

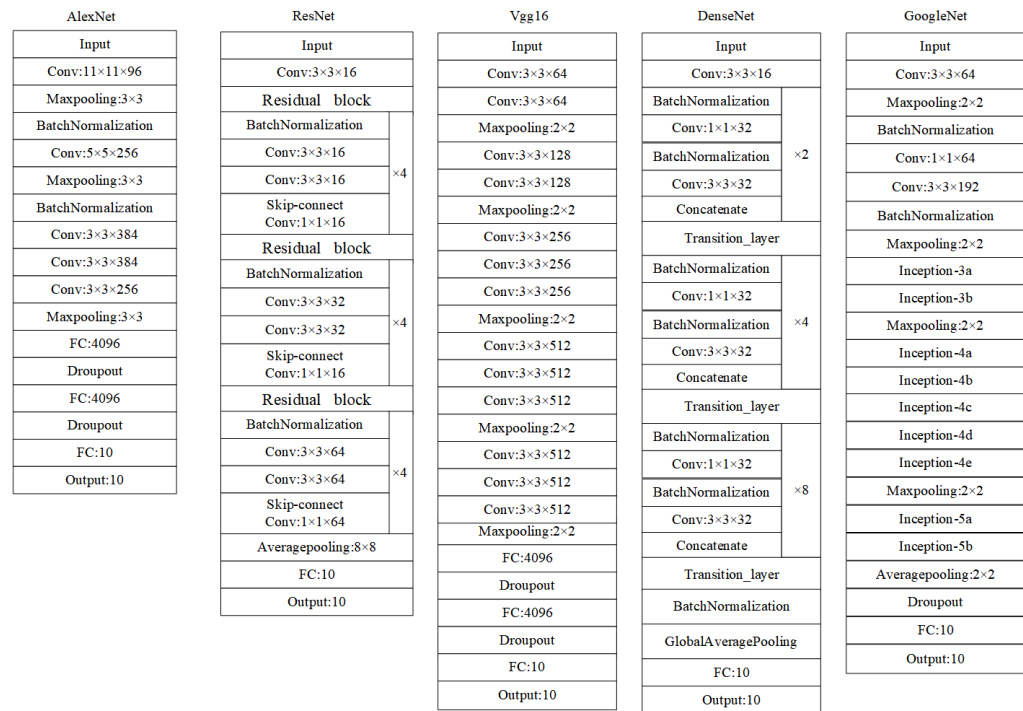
where  $x_i$  and  $y_i$  are the observed input and output of each Nbatch,  $\mu$  represents the mean of the input,  $\sigma^2$  represents the variance of the input,  $\epsilon$  is a constant near zero and  $\theta$ ,  $\psi$  are learnable parameters governing the scaling and shifting distributions. The activation function of the second FC layer is Sigmoid, which can limit the output value during the interval (0, 1) and prevent excessive slope from affecting the performance of the activation function.

To summarize the above contents, the proposed adaptive activation function has the ability to automatically learn complex features. Different nonlinear transformation is applied to different inputs to improve the generalization performance of the deep learning model, which will solve the problems in extracting the feature contour of the indicator diagram and the sparsity of the indicator diagram in pumping unit fault diagnosis. The following experimental simulation will verify the effectiveness of the designed adaptive activation function.



#### 4. Experimental Simulation

This section mainly verified the performance of our activation function, which was tested on AlexNet, VGG-16, GoogleNet, ResNet, and DenseNet. The structure diagrams of these five networks are shown in Figure 6. Moreover, we compared our activation function with the traditional activation functions such as ReLU, Sigmoid, Tanh, LReLU, and PReLU.



**Figure 6.** Network architectures of AlexNet, VGG-16, GoogleNet, ResNet, and DenseNet.

The experiment is mainly divided into two parts. The first part is the simulation of the fault diagnosis data set of the pumping unit. This will prove that the proposed adaptive activation function can extract the features of the indicator diagram and solve the sparsity problem of the indicator diagrams. The improvement in fault diagnosis accuracy indicates that indistinguishable samples are correctly classified. The second part is to verify the superiority of the designed adaptive activation function on the public data set CIFAR10.

##### 4.1. The Data Set of Pumping

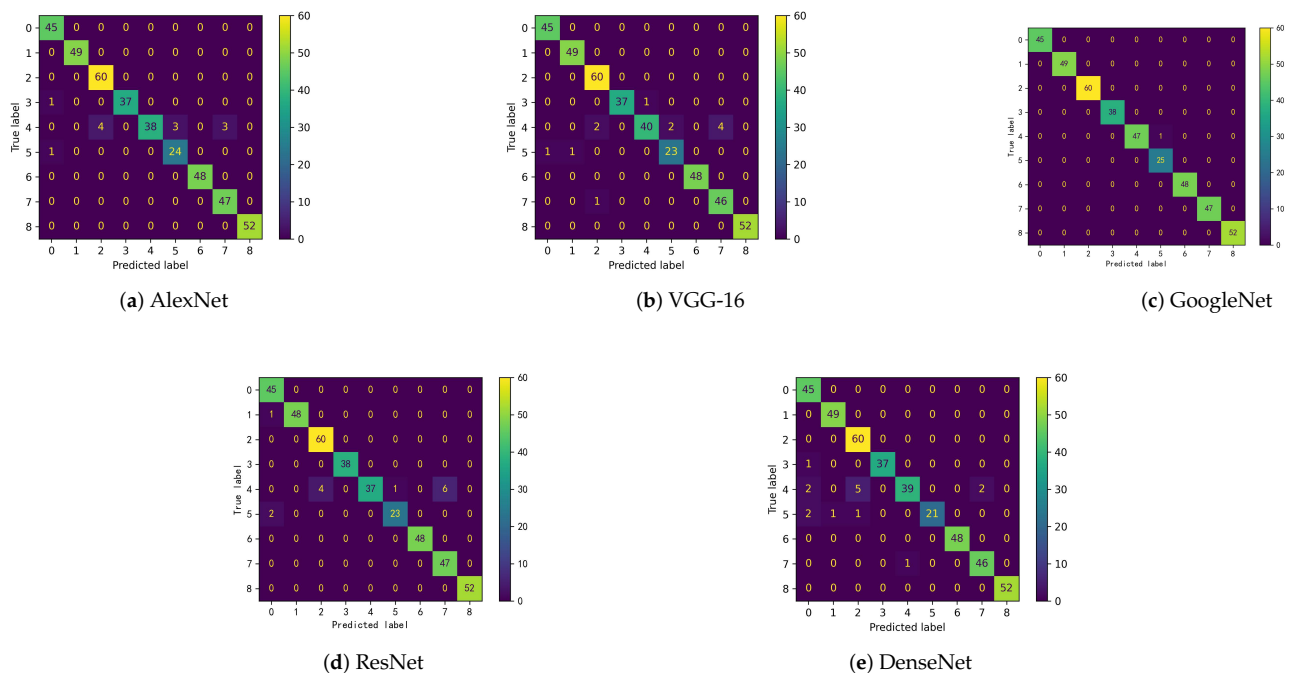
Adaptive Moment Estimation (Adam) was used here, and the initial learning rate was 0.001. The epoch of training was no less than 200. The average accuracy of each model is shown in Table 1. In the fault diagnosis data set of the pumping unit, the adaptive activation function proposed in this paper has the greatest accuracy improvement in the ResNet model. Compared to the traditional activation functions ReLU, Tanh, Sigmoid, and LReLU, the average accuracy of the ResNet model with our activation function, respectively, increased by 1.7%, 5.09%, 5.72%, and 2.54%. Compared to the adaptive activation functions PReLU, Mish, and Swish, the average accuracy of the ResNet model with our activation function, respectively, increased by 1.46%, 1.75%, and 1.56%.

The confusion matrix is a common index and visualization tool to evaluate the results of the classification model, and it can judge the advantages and disadvantages of classifiers. The rows of the matrix represent the true value, and the columns of the matrix represent the predicted value. The confusion matrix can, respectively, count the number of the wrong classification and the right classification and then display the results in a matrix. Figure 7 shows the confusion matrix of the five models for the pumping unit fault data set. It can be seen that the designed adaptive activation function can effectively represent the mapping relationship between the displacement and the load in the indicator diagram and extract

the features of the indicator diagram; thus, those indistinguishable samples are correctly classified. Taking GoogleNet as an example, the accuracy of each type of fault is given in Table 2, proving that the proposed adaptive activation function can improve the accuracy of each type of fault diagnosis.

**Table 1.** Classification precision of various activation functions for different models on pumping machine fault diagnosis.

Methods	AlexNet (%)	VGG-16 (%)	GoogleNet (%)	ResNet (%)	DenseNet (%)
Ours	97.91 ± 0.4997	97.57 ± 0.2658	99.13 ± 0.1941	<b>97.82 ± 0.4061</b>	97.52 ± 0.4706
ReLU	97.33 ± 0.5534	96.94 ± 0.9537	98.56 ± 0.3220	<b>96.12 ± 0.6140</b>	96.89 ± 0.0970
Sigmoid	94.51 ± 0.4930	96.41 ± 0.3567	96.17 ± 0.2831	<b>92.14 ± 0.2145</b>	94.17 ± 0.6584
Tanh	96.41 ± 0.3220	97.04 ± 0.4706	97.91 ± 0.3632	<b>94.66 ± 0.3070</b>	95.05 ± 0.4231
LReLU	97.48 ± 0.8209	97.14 ± 0.5405	98.74 ± 0.2830	<b>95.28 ± 0.9029</b>	96.36 ± 0.5091
PReLU	97.43 ± 0.6254	97.04 ± 0.3220	98.74 ± 0.6584	<b>96.36 ± 0.5091</b>	97.04 ± 0.3883
Mish	97.23 ± 0.6063	96.02 ± 0.6254	98.74 ± 0.2830	<b>96.07 ± 0.3883</b>	97.17 ± 0.1144
Swish	97.72 ± 0.3292	95.15 ± 0.8547	98.74 ± 0.1816	<b>96.26 ± 0.3943</b>	96.75 ± 0.5661



**Figure 7.** The confusion matrix of the five models of the pumping unit fault diagnosis dataset.

**Table 2.** Diagnosis accuracy of various fault types in pumping unit in GoogleNet.

Fault	Ours (%)	ReLU (%)	Sigmoid (%)	Tanh (%)	LReLU (%)	PReLU (%)	Mish (%)	Swish (%)
Pump up touch	100	98	94	94	98	98	100	96
Pumping rod detachment	100	100	100	100	100	100	100	100
Insufficient liquid supply	100	100	98	98	100	100	98	97
Contain sand	100	100	100	100	100	100	100	100
Piston stuck	100	100	100	100	100	100	100	100
Gas influence	96	96	96	92	96	100	96	93
Double valve leakage	100	100	100	100	100	100	100	100
Pump down touch	100	96	96	98	96	96	96	96
Normal	100	100	100	100	100	100	100	100

#### 4.2. CIFAR10

We used the CIFAR10 data set to conduct experiments and analyzed AlexNet, VGG-16, GoogleNet, ResNet, and DenseNet models with our activation function and the traditional

activation functions. We augmented the data to reduce overfitting. The Adam was used with an initial learning rate of 0.001. The epoch of training was no less than 200. The average accuracy of each model is shown in Table 3, where the designed activation function improves the performances of those. Among them, AlexNet, VGG-16, and DenseNet have a good performance. Compared with the traditional activation functions ReLU, Tanh, Sigmoid, LReLU, and adaptive activation functions PReLU, Mish, and Swish, our activation function in the AlexNet model is improved, respectively, by 1.84%, 4.11%, 5.45%, 0.79%, 2.74%, 2.07%, 1.91%; our activation function in VGG-16 model is improved, respectively, by 3.1%, 4.54%, 4.45%, 0.48%, 2.02%, 3.94%, and 4.69%; our activation function in DenseNet model is improved, respectively, by 1.88%, 5.1%, 9.63%, 1.07%, 0.61%, 0.37%, and 0.35%. The above data indicate the superiority of the proposed activation function.

**Table 3.** Classification precision of various activation functions for different models on CIFAR10.

Methods	AlexNet (%)	VGG-16 (%)	GoogleNet (%)	ResNet (%)	DenseNet (%)
Ours	<b>91.10 ± 0.0445</b>	<b>93.86 ± 0.0406</b>	90.35 ± 0.1070	91.73 ± 0.0231	<b>92.30 ± 0.0681</b>
ReLU	<b>89.26 ± 0.0576</b>	<b>90.76 ± 0.0987</b>	89.00 ± 0.0337	90.27 ± 0.0034	<b>90.42 ± 0.0846</b>
Sigmoid	<b>85.65 ± 0.0365</b>	<b>89.32 ± 0.1127</b>	87.23 ± 0.0485	88.06 ± 0.0835	<b>82.67 ± 0.2110</b>
Tanh	<b>86.99 ± 0.2432</b>	<b>89.41 ± 0.0189</b>	83.69 ± 0.0402	88.68 ± 0.0414	<b>87.20 ± 0.0745</b>
LReLU	<b>90.31 ± 0.2147</b>	<b>93.38 ± 0.0414</b>	89.70 ± 0.0527	91.24 ± 0.1059	<b>91.23 ± 0.0684</b>
PReLU	<b>88.36 ± 0.0436</b>	<b>91.84 ± 0.0633</b>	89.31 ± 0.0454	91.08 ± 0.0637	<b>91.69 ± 0.0755</b>
Mish	<b>89.07 ± 0.0847</b>	<b>89.92 ± 0.0577</b>	88.64 ± 0.0729	91.12 ± 0.7960	<b>91.97 ± 0.0758</b>
Swish	<b>89.19 ± 0.0628</b>	<b>89.19 ± 0.0618</b>	88.94 ± 0.1161	90.93 ± 0.0850	<b>91.95 ± 0.0893</b>

## 5. Conclusions

In this paper, a new adaptive activation function is designed and applied to five models of neural networks. Specifically, the adaptive activation function improves the negative semi-axis slope of the ReLU activation function by combining the gated channel conversion unit to enhance the performance of the deep learning model. The activation function in each layer of a neural network is unique; thus, the input signal of each layer has a unique nonlinear transformation.

Therefore, compared with the traditional fixed activation function, our activation function has a better nonlinear transformation ability, and it can be well-embedded into five models. Through the fault diagnosis data set of the pumping unit, it is proven that our activation function can effectively display the mapping relationship between displacement and load in the indicator diagram, thus extracting the features of the indicator diagram and solving the sparsity problem of the indicator diagrams. Indistinguishable samples are correctly classified. Through the CIFAR10 dataset, the superiority and universality of our adaptive activation function are verified.

In short, the proposed adaptive activation function increases the accuracy of fault diagnosis and has a better generalization performance and search ability. Moreover, the proposed adaptive activation functions also can be well-embedded into other models of neural networks.

The pumping unit works in a complex environment in the field, and the data collected should include environmental noise. In the future, we will study how to filter out the noise by extracting the time domain and frequency domain information and then fuse the time domain features and frequency domain features to improve the fault diagnosis accuracy of the pumping unit and anti-noise performance.

**Author Contributions:** Conceptualization, Y.L. (Yebin Li) and F.Z.; Funding acquisition, D.S.; Validation, D.S. and F.M.; Resources, Y.L. (Yuanhong Liu); Data Curation, Y.L. (Yebin Li); Writing—original draft, Y.L. (Yebin Li); Writing—review and editing, F.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the Project of Shandong Provincial Major Scientific and Technological Innovation (Nos. 2019JZZY010444, 2019TSLH0315), the Project of 20 Policies of Facilitate Scientific Research in Jinan Colleges (No. 2019GXRC063) and the industry-university-research collaborative innovation fund project of Qilu University of Technology (Shandong Academy of Sciences) (Nos. 2021CXY-13, 2021CXY-14).

**Institutional Review Board Statement:** Studies not involving humans or animals.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Du, J.; Liu, Z.; Song, K.; Yang, E. Fault diagnosis of pumping machine based on convolutional neural network. *J. Univ. Electron. Sci. Technol. China* **2020**, *49*, 751–757.
2. Zhang, R.; Gao, F. A New Synthetic Minmax Optimization Design of  $H_\infty$  LQ Tracking Control for Industrial Processes under Partial Actuator Failure. *IEEE Trans. Reliab.* **2020**, *69*, 322–333. [[CrossRef](#)]
3. Shi, H.; Li, P.; Su, C.; Wang, Y.; Yu, J.; Cao, J. Robust constrained model predictive fault-tolerant control for industrial processes with partial actuator failures and interval time-varying delays. *J. Process. Control* **2019**, *75*, 187–203. [[CrossRef](#)]
4. Gao, Z.; Cecati, C.; Ding, S.X. A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3757–3767.
5. Gao, Z.; Cecati, C.; Ding, S.X. A survey of fault diagnosis and fault-tolerant techniques—Part II: Fault diagnosis with knowledge-based and hybrid/active approaches. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3768–3774.
6. Tang, A.; Zhao, W. A Fault Diagnosis Method for Drilling Pump Fluid Ends Based on Time–Frequency Transforms. *Processes* **2023**, *11*, 1996.
7. Fu, Z.; Zhou, Z.; Yuan, Y. Fault Diagnosis of Wind Turbine Main Bearing in the Condition of Noise Based on Generative Adversarial Network. *Processes* **2022**, *10*, 2006. [[CrossRef](#)]
8. Agarwal, P.; Gonzalez, J.I.M.; Elkamel, A.; Budman, H. Hierarchical Deep LSTM for Fault Detection and Diagnosis for a Chemical Process. *Processes* **2022**, *10*, 2557. [[CrossRef](#)]
9. Xu, G.; Liu, M.; Jiang, Z.; Shen, W.; Huang, C. Online fault diagnosis method based on transfer convolutional neural networks. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 509–520. [[CrossRef](#)]
10. Duan, Y.; Li, Y.; Sun, Q.; Xu, D. Improved alexnet model and its application in well dynamogram classification. *Comput. Appl. Softw.* **2018**, *35*, 6.
11. Sang, J. Research on pump fault diagnosis based on pso-bp neural network algorithm. In Proceedings of the 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 24–26 May 2019; pp. 1748–1752.
12. Zhang, L.; Du, Q.; Liu, T.; Li, J. A fault diagnosis model of pumping unit based on bp neural network. In Proceedings of the 2020 International Conference on Networking and Network Applications (NaNA), Haikou, China, 10–13 December 2020; pp. 454–458.
13. Hu, H.; Li, M.; Dang, C. Research on the fault identification method of oil pumping unit based on residual network. In Proceedings of the 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), Virtual Conference, 15–17 April 2022; pp. 940–943.
14. Bai, T.; Li, X.; Ding, S. Research on electrical parameter fault diagnosis method of oil well based on tsc-dcgan deep learning. In Proceedings of the 2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Xian, China, 15–17 July 2022; pp. 753–761.
15. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML), Haifa, Israel, 21–24 June 2010.
16. Maas, A.L. Rectifier Nonlinearities Improve Neural Network Acoustic Models. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 17–19 June 2013.
17. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1026–1034.
18. Ramachandran, P.; Zoph, B.; Le, Q.V. Searching for Activation Functions. *arXiv* **2017**, arXiv:1710.05941.
19. Hu, H.; Liu, A.; Guan, Q.; Qian, H.; Li, X.; Chen, S.; Zhou, Q. Adaptively customizing activation functions for various layers. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, 1–12. [[CrossRef](#)] [[PubMed](#)]
20. Zhao, M.; Zhong, S.; Fu, X.; Tang, B.; Dong, S.; Pecht, M. Deep residual networks with adaptively parametric rectifier linear units for fault diagnosis. *IEEE Trans. Ind. Electron.* **2021**, *68*, 2587–2597. [[CrossRef](#)]
21. Yang, Z.; Zhu, L.; Wu, Y.; Yang, Y. Gated channel transformation for visual recognition. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 11791–11800.
22. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2012**, *60*, 84–90. [[CrossRef](#)]

23. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
24. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 70–778.
26. Huang, G.; Liu, Z.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
27. Cao, L.; Zhao, T. Pumping unit design and control research. In Proceedings of the 2019 IEEE International Conference on Mechatronics and Automation (ICMA), Tianjin, China, 4–7 August 2019; pp. 1738–1743.
28. Misra, D. Mish: A self regularized non-monotonic neural activation function. *arXiv* **2019**, arXiv:1908.08681.
29. Lin, M.; Chen, Q.; Yan, S. Network in network. *arXiv* **2013**, arXiv:1312.4400.
30. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv* **2015**, arxiv:1502.03167.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.