Fault Diagnosis of Vertical Pumping Unit Based on Characteristic Recalibration Residual Convolutional Neural Network

Youxiang Duan[1], Cheng Chang[1](✉), Qifeng Sun[1], Chengze Du[1], and Zhengyang Li[2]

[1] School of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266580, China
[2] School of Precision Instruments and Optoelectronics Engineering, Tianjin University, Tianjin 300072, China

Abstract. Rod pumping unit is the main equipment of oil exploitation. The automatic and intelligent of the management, control, running of pumping units are important goals for the construction of the smart oil field. The continuous development of network and deep learning technology provides strong support for the realization of intelligent pumping unit systems. The use of real-time collected pumping unit operating data for working condition supervision and intelligent analysis and decision-making has become an important part of the new generation of vertical pumping unit systems. This article is based on the collected operating data such as the dynamometer diagram of the pumping unit, using the deep learning technology, the intelligent working condition analysis, and the fault diagnosis model of the new generation vertical pumping unit are established. The training time of the model is short while the accuracy of identification and classification is high, which meets the practical application requirements well.

Keywords: Vertical pumping unit · Deep learning · Dynamometer card · Fault analysis

1 Introduction

With the rapid development and widespread application of technologies such as network information and artificial intelligence [1], intelligent oilfields are replacing conventional oilfields as the main target for the construction of next-generation oilfields [2]. Intelligent supervision of oil production is a top priority for future oilfield construction [3].

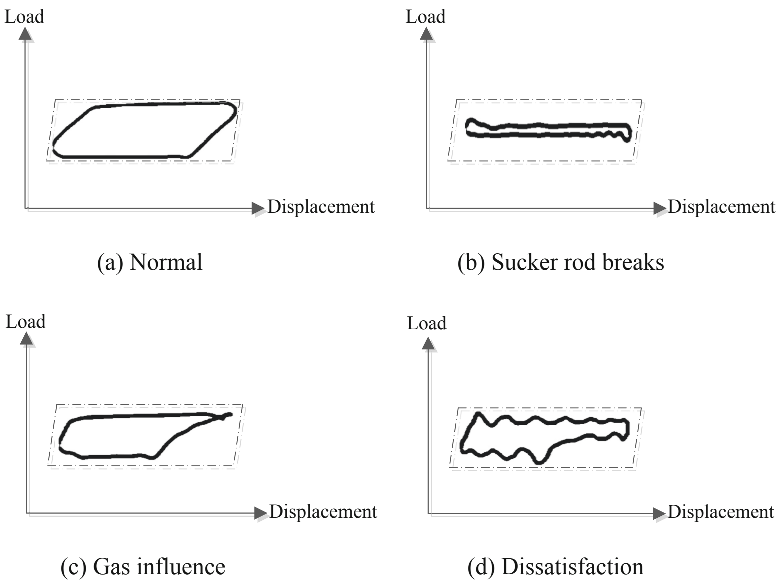
Oil is mainly lifted from the underground to the ground by a rod pumping unit, which is divided into a beam type and a beamless type. Due to the shortcomings of beam pumping units such as low efficiency and high consumption, and difficult to balance during long strokes [4]. The new generation of none-beam pumping units is gradually replacing beam pumping units. The vertical pumping unit is a tower pumping unit, with high efficiency and low consumption, safe and stable [5], and has been popularized and applied in domestic and foreign oil fields.

H. Song and D. Jiang (Eds.): SIMUtools 2020, LNICST 370, pp. 545–556, 2021.
https://doi.org/10.1007/978-3-030-72795-6_44

The pumping unit usually works in harsh and remote areas, and its operation is affected by many factors such as wind, snow, rain, and other severe weather, gearbox gear wear, and motor over-time operation. If failures are not found and diagnosed in time, it will seriously affect oil production and even cause production accidents.

Therefore, it is of great significance to intelligently analyze and diagnose the operating conditions of pumping units [6]. For many years, the intelligent analysis and diagnosis of pumping unit conditions using new computer technology have been an important part of oilfield application research. Artificial intelligence technologies such as neural networks and support vector machines have been used in the intelligent diagnosis of pumping unit failures [7, 8]. But the application effect did not meet people's expectations. The development of artificial intelligence technology, especially the emergence of deep learning technology, has provided new solutions for intelligent analysis and diagnosis of pumping unit failures [9].

The dynamometer method is the most effective fault analysis and diagnosis method for pumping units today [10]. The dynamometer diagram is a closed curve of the change of the suspension point load with the displacement of the suspension point in a pumping cycle. The shape characteristics of the dynamometer diagram can reflect the operating conditions of the pumping unit [11]. Although the vertical pumping unit and the beam pumping unit have different structures, the principle of oil extraction is the same. Therefore, the dynamometer method is used to analyze the operating conditions of the vertical pumping unit, and the fault diagnosis of the vertical pumping unit is also effective. However, different mechanical structures of pumping units have different shape characteristics of the dynamometer diagram.



(a) Normal (b) Sucker rod breaks

(c) Gas influence (d) Dissatisfaction

Fig. 1. Vertical pumping unit power diagram

Figure 1 is a typical dynamometer diagram of a vertical pumping unit collected and summarized. Figure 1 (a) is a dynamometer diagram of normal operation, and Fig. 1 (b) is a dynamometer diagram of a broken sucker rod. Figure 1 (c) is the dynamometer diagram of gas influence, and Fig. 1 (d) is the dynamometer diagram of dissatisfaction.

The experimental method of pumping unit fault diagnosis is as follows: First, collect the pumping unit dynamometer diagram data and use the data to draw the dynamometer diagram. Then, mark the dynamometer diagram according to the working condition of the pumping unit represented by the dynamometer diagram. Finally, perform a classification and recognition experiment on the dynamometer diagram, divide the dynamometer diagram data into a training set, a validation set, and a test set, use the training set to train the model, use the validation set to determine the model hyperparameters, select the optimal model, and use the test set perform performance evaluation on the trained model.

The significance of applying deep learning model to the analysis of pumping unit operating conditions is to use the deep learning model to analyze the dynamometer diagram instead of manual analysis by a large number of experienced professional staff. The analysis results are no longer affected by the staff's experience and knowledge, and the personal workload is greatly reduced Small, able to meet the needs of intelligent and automated oilfield construction.

The deep learning model can not only be applied to the fault diagnosis of pumping units but also be applied to the data processing of pumping units, for example, the power diagram converse to the dynamometer diagram. It has useful and broad application prospects.

2 Fault Diagnosis Model of Vertical Pumping Unit Based on Feature Recalibration Residual Convolutional Neural Network

2.1 Model Structure

Convolutional neural network (CNN) is a kind of feedforward neural network with convolution operation and deep structure. It consists of a convolutional layer, pooling layer, fully connected layer, etc. [12]. Different from other automatic image recognition and classification methods, CNN can automatically extract and filter features, without the need to manually extract features. Through the use of convolutional continuous extraction, more abstract features can be obtained. Using the obtained abstract features can complete a variety of tasks and perform well in the field of image recognition and classification [13]. The dynamometer diagram can be regarded as a special image, so it is feasible to analyze the dynamometer diagram with a CNN [14].

According to fault diagnosis needs, draw on the deep learning model with good image recognition and classification effect, such as, draw on the residual ideas and modularization in ResNet and DenseNet, draw on the SE substructure in SENet, etc. [15–17], and improve based on these, this paper proposes a feature recalibration residual CNN model based on dynamometer diagram data, which aims to enhance model performance and shorten training time.

Feature recalibration residual CNN model has 14 layers, and its module structure is shown in Fig. 2, where the convolutional layer contains 16 3 × 3 convolution kernels. 5 SE-residual modules (The residual module is embedded in the Squeeze-and-Excitation substructure) contains 2 convolutional layers, the convolution kernel sizes are 1 × 1 and 3 × 3, and the number of convolution kernels is 32, 64, 64, 128, 128. The convolution layer sets the convolution step size to 1, adding L2 regularization, and LeakyReLu as the activation function, the pooling filter size is 2 × 2, and the pooling step size is set to 2, the neurons of the fully connected layer The numbers are 1024, 512, and 4, and the activation functions are LeakyReLu, LeakyReLu, and softmax. The input of the model is the dynamometer diagram, and the output is the type of the dynamometer diagram (fault type).

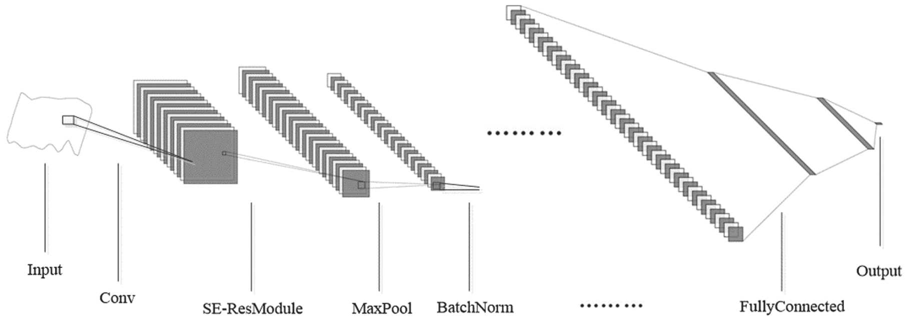


Fig. 2. Feature recalibration residual CNN model

The feature recalibration residual CNN model strengthens the effect of effective feature maps by embedding substructures, and the performance is significantly improved. Compared with the classical deep learning model to improve the performance by increasing the depth, the feature recalibration residual CNN model embeds the substructure adds little calculation and the modular structure is easy to modify to meet the application requirements.

2.2 Residual Module

The residual module designed in this paper consists of 1 × 1 convolutional layer, 3 × 3 convolutional layer, intermediate operations (BN is batch normalization, LeakyReLu is activation operation, Ave_pool is average pooling), and 1 The identity mapping is composed as shown in Fig. 3. Among them.

The module input is x. The module output is H (x) = F (x) + x. The formula changes to F (x) = H (x) − x.

F (x) represents the residual, the module fits between the input and output is the residual F (x).

Unlike most residual networks, this structure consists of 1 × 1 convolutional layer, 3 × 3 convolutional layer, and intermediate operations. It not only reduces the dimensions of extracted features, reduces the number of parameters, but also increases non-linear factors, and accelerates the model convergence. To play an important role in simplifying the model and enhancing the ability to express the model.

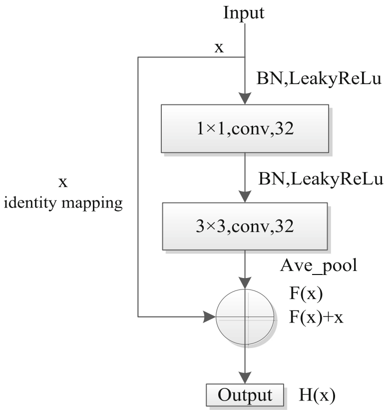


Fig. 3. Residual module

2.3 SE-Residual Module

After the residual module is built, the SE substructure is embedded in it, forming the main part of the feature recalibration residual convolution neural network model, namely the SE-residual module, as shown in Fig. 4 (in the Fig, c is the number of feature channels, H is the feature map height and w is the feature map width).

In the SE substructure, the first step is Squeeze operation, which uses average pooling to compress the features. Then, the fully connected layer is used to reduce the feature's dimension to 1/16 (16 is the setting parameter value, which can be modified). After the second fully connected layer, the feature dimensions are restored. Compared with only one fully connected layer, it can add more non-linearity, enhance the relationship between the channels, and reduce the number of parameters. The second step is Excitation operation, which uses the Sigmod function to calculate the weight for each feature channel, and the weight represents the importance of each feature channel. The third step is Scale operation, which uses multiplication to weigh each channel feature to complete the re-calibration of the original feature.

Squeeze operation expression (1).

$$Z_c = F_{sq}(U_c) = \frac{1}{H \times W} \sum\nolimits_{i=1}^{H} \sum\nolimits_{j=1}^{W} U_c(i,j) \tag{1}$$

F_{sq} is the Squeeze function, that is, average pooling and U_c is a feature map with height and width H × W. The input of H × W × C is converted into the output of 1 × 1 × C by the Squeeze function.

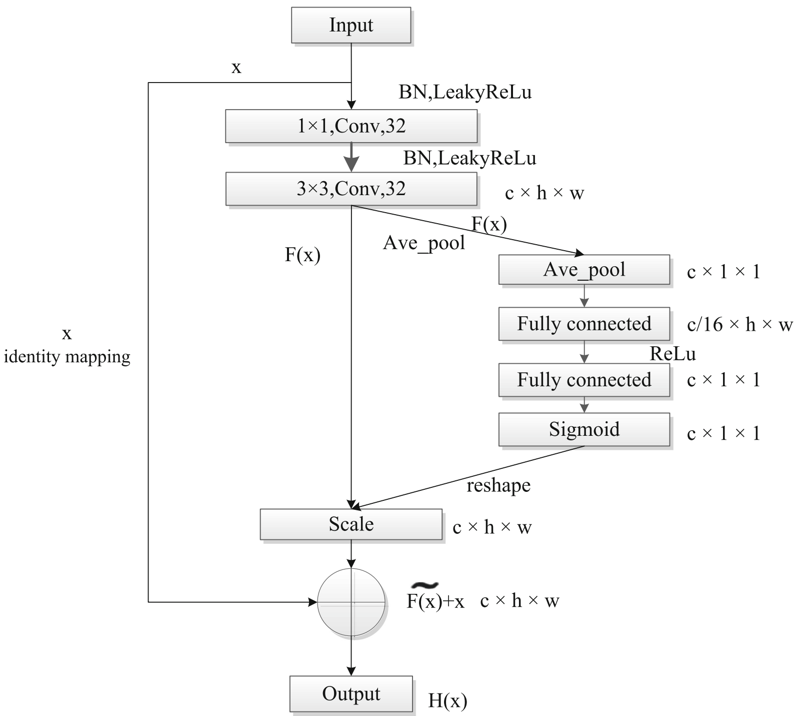


Fig. 4. SE-residual module

Excitation operation expression (2).

$$s = F_{ex}(z, W) = \sigma(W_2\delta(W_1z)) \tag{2}$$

Fex is the Excitation function, z in formula (2) is Z_c in formula (1), and W_{1z} is the first fully connected calculation operation. The dimension of W_1 is C/16 × C. Taking 16 in this paper is to reduce the number of channels to the original 1/16. After the ReLu activation operation and the second fully connected layer are multiplied with W_2, the dimension of W_2 is C × C/16, and the output dimension is C × 1 × 1. Finally, s is obtained after passing the sigmoid function.

Scale operation expression (3).

$$F_{scale}(U_c, S_c) = S_c \cdot U_c \tag{3}$$

F_{scale} is the Scale function representing U_c multiplied by S_c, U_c is a two-dimensional matrix, and Sc is the weight.

Compared with the residual module, the performance of the SE residual module is improved. The added parameters only exist in the 2 fully connected layers, and the increased amount of computers is almost negligible. At the same time, the structure of the SE residual module is very simple and easy to implement, without introducing new functions or layers.

2.4 Model Features

Using the idea of residual to build the model changes the way of forward and backward information transmission. During model training, if the feature represented by x is already very mature, that is, an increase or decrease in x will increase the loss of the model. At this time, F(x) will tend to 0, x will continue to transmit information from the identity mapping path. This is conducive to the training of deep networks and solves the gradient descent problem to a certain extent. As the depth of the model deepens, its expression ability becomes stronger and the classification accuracy of the test set is higher.

The feature recalibration through the SE substructure is mainly based on the model's continuous learning process through the loss value. The feature weight calibration increases the weight of valid features, reduces the weight of invalid features, and trains a model with stronger performance. The SE substructure is simple to implement and suitable for embedding various deep learning models.

Adding a batch normalization operation between the convolutional layer can reduce the number of model parameters, improve the accuracy of model training, and reduce the problem of gradient descent.

The main steps of batch normalization.

1. Input.

$$x : \beta = \{x_1, \ldots, x_m\} \tag{4}$$

2. Output.

$$\{y_i = BN_{\gamma,\beta}(x_i)\} \tag{5}$$

3. Calculating the mean of batch data.

$$\mu_\beta = \frac{1}{m}\sum\nolimits_{i=1}^{m} x_i \tag{6}$$

4. Calculate the batch data variance.

$$\sigma_\beta^2 = \frac{1}{m}\sum\nolimits_{i=1}^{m} (x_i - \mu_\beta)^2 \tag{7}$$

5. Normalize.

$$\hat{x}_i = \frac{(x_i - \mu_\beta)}{\sqrt{\sigma_\beta^2 + \varepsilon}} \tag{8}$$

6. Scale change and migration processing.

$$y_i = y\hat{x}_i + \beta = BN_{\gamma,\beta}(x_i) \tag{9}$$

7. Return learning parameters γ and β.

The model uses the LeakyReLu function as the activation function. Compared with the ReLu function, the LeakyReLu function can modify the data distribution and update the network parameters when the input is negative [18]. LeakyReLu function formula such as (10).

$$y_i = \begin{cases} x_i & if\ x_i \geq 0 \\ \frac{x_i}{a_i} & if\ x_i < 0 \end{cases} \tag{10}$$

a_i is a fixed parameter in the interval (1,+∞).

The model uses Adam as the optimization function. Compared with SGD, Momentum, AdaGrad, and other optimization functions, the Adam optimization function is easy to implement, the range of hyperparameter adjustment is small, the update step can be limited to a certain range, and the learning rate does not need to be manually adjusted. Learning model [19], parameter description of Adam's optimization function.

α is the Step size.
β_1 is the first-order moment attenuation coefficient, initialized to 0.9.
β_2 is the second-order moment attenuation coefficient, initialized to 0.999.
f(θ) is the objective function. θ is the parameter to be optimized.
t is the number of update steps, initialized to 0.
g_t is Gradient for f (θ) derived from θ.
m_t is the first moment of g_t, which is the expectation of g_t.
v_t is the second moment of g_t, which is the expectation of g_t^2.
m_t is the offset correction of m_t, because m_t is initialized to 0, which will cause b to be offset to 0.

Similarly, $\widehat{v_t}$ is the offset correction for v_t.

The main update steps of the Adam optimization function.

Update steps.

$$\mathrm{t} = \mathrm{t} + 1 \tag{11}$$

Calculate the gradient of f (θ) to the parameter θ.

$$g_t = \nabla_\theta f_t(\theta_{t-1}) \tag{12}$$

Calculate the first moment of the gradient.

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \tag{13}$$

Calculate the second moment of the gradient.

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \tag{14}$$

Correct the first moment.

$$m_t = m_t / (1 - B_1^t) \tag{15}$$

Correct the second-order moment.

$$v_t = v_t/(1 - B_2^t) \tag{16}$$

Update parameters.

$$\theta_t = \theta_{t-1} - \alpha \cdot m_t/(\sqrt{v_t} + \varepsilon) \tag{17}$$

The model uses the variance scaling metshod to initialize the weights, which has a stronger generalization ability than the conventional initialization method. The model is finally 3 fully connected layers, which not only plays a role in classification but also reduces the impact of feature distribution on classification [20].

3 Experiment and Analysis

3.1 Data Source

The dynamometer data set used in this article is from a domestic oil field using a new vertical intelligent pumping unit. The data collection period is from 12:00 on January 15, 2019, to 21:00 on August 20, 2019. After the manual screening, The classification and data enhancement totaled 19,000 dynamometer maps, of which 4,084 were “working normally”, 2,697 were “filled with dissatisfaction”, 8083 were “gas impact”, and 4,136 were “pump rod disconnection”.

The data set is divided using a random extraction method. The training set accounts for 90% and the test set accounts for 10% for the model.

3.2 Model Construction

This experiment uses NVIDIA Tesla P100 GPU graphics card and uses tflearn and sklearn as platforms to develop, and build logistic regression model, random forest model, XGBoost model, logistic regression, random forest, XGBoost integrated model, 10-layer ResNet model, 53-layer The DenseNet model and the feature recalibration residual CNN model proposed in this paper.

The last fully connected layer is classified using the Softmax function, the loss function is cross-entropy, and the optimization function is Adam.

3.3 Experiment and Analysis

Use the prepared data set to train the logistic regression model, random forest model, XGBoost model, integrated model, ResNet model, DenseNet model, and feature recalibration residual CNN model. Use the trained model to show power on the test set Graph classification experiment. To reduce the error of the experiment, the average value obtained from multiple experiments is shown in Fig. 5, Fig. 6, Table 1, and Table 2.

From Fig. 5, Fig. 6, Table 1, and Table 2, the average accuracy, average accuracy, average recall, and average f1 scores of the feature recalibration residual CNN model on the test set compared with the rest of the models are the highest. The loss rate is the lowest, the average training time is very short, and the model obtained has a strong generalization ability, which can be applied to the application requirements of vertical pumping unit fault monitoring and analysis.

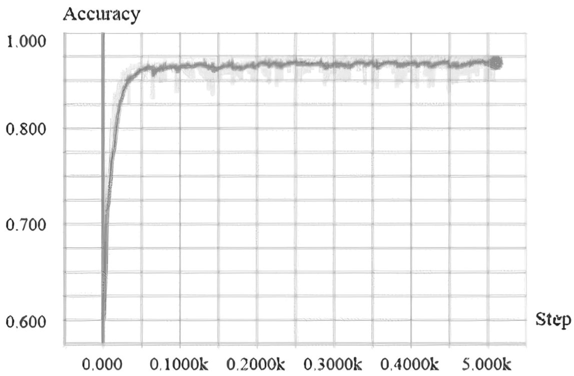


Fig. 5. Accuracy of the model test set in this paper

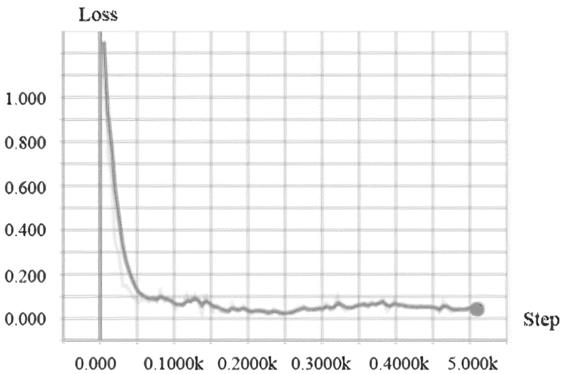


Fig. 6. Loss rate of the model test set in this paper

Table 1. Model accuracy, recall rate, f1 score of this paper

Class	Precision	Recall	F1-score
1	0.98	0.97	0.98
2	0.98	0.98	0.98
3	0.99	0.98	0.98
4	0.99	0.99	0.99
Total	0.98.5	0.98	0.98

Table 2. Comparison of experimental results. The first five columns of data in the table are the average of the model test set and the last column is the average training time of the model.

Model	Accuracy	Loss rate	Precision	Recall	f1 score	Time
Logistic regression	80%	10.9%	80%	80%	80%	27 min
Random forest	92%	10%	93%	93%	93%	23 min
XGBoost	70%	15%	73%	69%	66%	45 min
Integration	87%	7%	88%	87%	88%	1 h 35 min
ResNet	96%	13%	96%	96%	96%	2 h
DenseNet	97.5%	6.8%	97.5%	97.5%	98%	3 h
this article	98.4%	5.5%	98.5%	98%	98%	30 min

4 Total Knots

In oilfields at home and abroad, especially some tight and low-permeability special oil and gas reservoirs, vertical intelligent pumping units are increasingly used. It is of great significance to optimize the production decision by monitoring the oil and gas production process by analyzing the vertical pumping unit dynamometer diagram. In this paper, the new deep learning technology is used to analyze the working conditions of vertical pumping units, and a feature recalibrated residual convolution neural network model is designed to realize the automatic recognition and classification of the dynamometer diagram of vertical intelligent pumping units. In this paper, the residual idea is added to the model to effectively avoid vanishing gradient problem. The batch normalization layer, activation layer, and average pooling layer in the residual module can not only improve the training speed but also enhance the nonlinear ability of the model. The SE substructure is embedded in the residual module, which makes the weight of the effective feature graph increase and the weight of the invalid feature graph decreases continuously, which enhances the learning ability of the model. Experimental results show that compared with other models, The average classification accuracy, average accuracy, average recall, and average f1 score of the model on the test set are the highest, the average loss rate is the lowest, and the average training time is short, which meets the needs of applications such as accuracy and real-time performance. It has made useful explorations for the application of deep learning new technologies and the construction of smart oilfields.

References

1. Jufu, F., Jian, Y.: Foreword of the frontier progress of artificial intelligence. Comput. Res. Dev. **56**(08), 1604 (2019)
2. Yaowei, Z., Li, Z.: Application and development of intelligent oil field and intelligent drilling and production technology. Low Carbon World (02), 71–72 (2019)
3. Changshan, L.: Intelligent management mode in petrochemical production process. Zhonghua Constr. **35**(08), 90–91 (2019)

4. Liuzhu, Z.: Analysis of energy-saving and consumption-reducing measures at the oil production site. Cleaning World **35**(05), 31–32 (2019)
5. Chunming, X.: Intelligent control system of vertical energy-saving pumping unit. Electromech. Eng. **31**(04), 486–489 (2014)
6. Zhigang, T.: Research on fault dynamometer diagnosis method based on convolutional neural network. Xi'an University of Science and Technology, Xi'an (2019)
7. Shuguang, L., Chenhui, J., Yamin, T., et al.: An ART2-based pumping unit fault diagnosis. In: Chinese Control and Decision Conference 2015, Qingdao, pp. 2955–2958 (2015)
8. Zhidan, Z., Haojie, F., Penghui, L.: Application of CNN-SVM model in fault diagnosis of pumping unit wells. J. Henan Polytech. Univ. (04), 112–117 (2018)
9. Zhou, X., Zhao, C., Liu, X.: Application of cnn deep learning to well pump troubleshooting via power cards. In: Abu Dhabi International Petroleum Exhibition & Conference 2019, Abu Dhabi (2019)
10. Zhidan, Z., Penghui, L., Miaomiao, G.: Research on fault diagnosis of rod pumping unit in petroleum production. Comput. Simul. (02), 443–447 (2016)
11. Zhou, W., Li, X., Yi, J., et al.: A novel UKF-RBF method based on adaptive noise factor for fault diagnosis in pumping unit. IEEE Trans. Ind. Inform. **15**(03), 1415–1424 (2018)
12. Szegedy, C., Liu, W., Jia, Y., et al.: Going deeper with convolutions. In: Computer Vision and Pattern Recognition 2015, Boston, Massachusetts, pp. 1–9 (2015)
13. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Computer Vision and Pattern Recognition 2014, Columbus, Ohio (2014)
14. Tong, X.: Research on oilfield intelligent fault diagnosis system based on optimized RBF network. Xi'an Shiyou University, Xi'an (2019)
15. He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition. In: Computer Vision and Pattern Recognition 2016,Las Vegas, Nevada, pp. 770–778 (2016)
16. Huang, G., Liu, Z., Der Maaten, L.V., et al.: Densely connected convolutional networks. In: Computer Vision and Pattern Recognition 2017, Honolulu, Hawaii, pp. 2261–2269 (2017)
17. Hu J, Shen L, Albanie S, et al.: Squeeze-and-excitation networks. IEEE Trans. Pattern Anal. Mach. Intell. 7132–7141 (2018)
18. Gongpeng, W., Meng, D., Changyong, N.: Stochastic gradient descent algorithm based on convolutional neural network. Comput. Eng. Des. **39**(02), 441–445 (2018)
19. Hui, Z.: Research and improvement of optimization algorithms in deep learning. Beijing University of Posts and Telecommunications, Beijing (2018)
20. Fuyu, H., Jingzhong, W., Linhao, Z.: Public opinion classification method of heterogeneous data based on CNN and LSTM. J. Comput. Syst. **28**(06), 141–147 (2019)