Multi-Source Consistency Deep Learning for Semi-Supervised Operating Condition Recognition in Sucker-Rod Pumping Wells

Jianguo Yang[1], Bin Zhou[2]*, Muhammad Tahir[3]*, Min Zhang[4], Xiao Zheng[5], Xinqian Liu[6]
School of Mechanical Engineering, Shandong University of Technology, Zibo, China[1]
School of Computer Science and Technology, Shandong University of Technology, Zibo, China[2, 4, 5, 6]
Department of Computer Science-Faculty of Engineering Science and Technology (FEST), Iqra University, Main Campus, Defense View, Karachi 75500, Sindh, Pakistan[3]

Abstract*—How making full use of the multiple measured information sources obtained from the sucker-rod pumping wells based on deep learning is crucial for precisely recognizing the operating conditions. However, the existing deep learning-based operating condition recognition technology has the disadvantages of low accuracy and weak practicality owing to the limitations of methods for handling single-source or multi-source data, high demand for sufficient labeled data, and inability to make use of massive unknown operating condition data resources. To solve these problems, here we design a semi-supervised operating condition recognition method based on multi-source consistency deep learning. Specifically, on the basis of the framework of WideResNet28-2 convolutional neural network (CNN), the multi-head self-attention mechanism and feedforward neural network are first used to extract the deeper features of the measured dynamometer cards and the measured electrical power cards, respectively. Then, the consistency constraint loss based on cosine similarity measurement is introduced to ensure the maximum similarity of the final features expressed by different information sources. Next, the optimal global feature representation of multi-source fusion is obtained by learning the weights of the feature representations of different information sources through the adaptive attention mechanism. Finally, the fused multi-source feature combined with the multi-source semi-supervised class-aware contrastive learning is exploited to yield the operating condition recognition model. We test the proposed model with a dataset produced from an oilfield in China with a high-pressure and low permeability thin oil reservoir block. Experiments show that the method proposed can better learn the critical features of multiple measured information sources of oil wells, and further improve the operating condition identification performance by making full use of unknown operating condition data with a small amount of labeled data.

Keywords*—Operating condition recognition of sucker-rod pumping wells; multi-source consistency learning; semi-supervised learning; CNN; attention mechanism

I. INTRODUCTION

In the process of oil production, the sucker-rod pumping well production system can obtain a large amount of operating condition data from multiple information sources, such as dynamometer cards, electrical parameters, etc. Meanwhile, massive unknown operating condition data can also be obtained. It is crucial to timely and accurately recognize the operating conditions by effectively using and processing these different high-dimensional heterogeneous information sources. Recently, deep learning has become a research hotspot in oil production engineering and information technology due to its powerful ability to handle complex data structures and extract deep hidden features, thus improving the accuracy of model classification [1-2]. However, deep learning usually relies mainly on a great amount of labeled data for model training in order to maintain high accuracy. But the acquisition of such a large number of labeled data is impractical, and this will also cause a serious waste of massive unlabeled data resources. Therefore, exploring the integration of multiple information sources, massive unlabeled operating condition data, and deep learning to optimize the operating condition recognition with few labeled data is much of important for further improving the accuracy and extending the application of operating condition identification.

Existing deep learning-based sucker-rod pumping well operating condition recognition technology is mainly achieved by dynamometer cards or electrical parameters alone [3-4], or the effective combination of the two kinds of information as well as the oil well production information with artificial intelligence methods [5-6]. These studies have achieved good results, but also hold some limitations: (1) Relying on a single data source, which may easily cause false positives in electromechanical hydraulic coupling complex systems. (2) Poor fusion strategy and low robustness when integrating multiple information sources, which can degrade the performance and lower the robustness due to the uncertainty of production statistical data [7]. (3) Massive labeled samples requirement [8-9] and few unlabeled samples considered [10-11] during model training, which can greatly weaken the practical value of operating condition recognition technologies. (4) The feature extraction strategy is mainly realized by combining mechanism analysis [12-14], which will cause precision errors in the calculation of model feature parameter values and affect the recognition effect. Besides, the problems of "zero division" (appears during the electrical parameter timing signal recognition) and damping coefficient (appears during the pump dynamometer card recognition) will also increase the accuracy error.

Multi-source information fusion [15] is an information integration technology that can achieve more comprehensive and accurate feature representation by integrating features from

different information sources. By combining the unique advantages of each data source, it overcomes the limitation of a single data source, reduces the prediction error, and therefore significantly improves the robustness. Compared with traditional multi-feature connection strategy [16], multi-source information fusion can smartly integrate multi-source features and effectively address the challenge caused by simple connection fusion, such as higher feature dimension, information redundancy and inability to capture complex relationships among features. In the fusion process, following the consistency principle [17], multi-source information fusion can keep the logical consistency of the fused features and reduce information redundancy and conflict. Further, taking the cosine similarity measure [18] as a consistency measure, a more accurate measurement instead of the traditional Euclidean distance, can further improve the reliability and effectiveness of fusion results. In order to realize efficient multi-source information fusion, a powerful feature extractor is also needed to capture the key features of each information source. With wide network structure, WideResNet28-2 (WRN28-2) [19] can capture more abundant and diversified feature information, thus providing basic support for subsequent fusion. Referring to the encoder design in Transformer architecture [20], the introduction of multi-head self-attention mechanism and feedforward neural network can facilitate the deep interaction between multi-source features, significantly enhance the representation capability of multi-source features, and provide more abundant feature information for the fusion process. Studies have shown [21] that the introduction of an attention mechanism can further improve the effect of feature fusion. The attention mechanism can automatically learn and dynamically adjust the weights of features of various data sources and focus its attention on more critical information in the classification task, thus effectively improving the classification performance and overall effect.

Semi-supervised deep learning [22] trains the model by using a few labeled data and lots of unlabeled data, with the aim of enhancing the generalization ability of the model. It has four main learning strategies: pseudo-label learning [23], consistency regularization [24], deep learning-based generation [25], and graph-based strategy [26]. In pseudo-label learning, the model uses its own predictions of unlabeled samples to expand the training data. Compared with other methods, it has high flexibility and strong generalization, making it applicable for a variety of practical application scenarios, therefore is widely used in semi-supervised deep learning. Contrastive learning [27] is an effective unsupervised learning method. It learns distinguishing feature representations by comparing sample pairs, therefore the learned features achieve improved generalization on unlabeled data. Studies have shown that the combination of class-aware contrastive learning, pseudo-label learning and consistency regularization learning [28] can effectively solve the common confirmation bias problem in semi-supervised deep learning and improve accuracy. By minimizing the distance of samples of the same category in feature space and maximizing the distance between samples of different categories, this combination strategy can not only enhance differentiation between categories, but also significantly improve the accuracy of semi-supervised learning for tasks such as image recognition.

As described above, to solve the limitations of existing deep learning-based operating condition recognition research, this paper proposes a semi-supervised sucker-rod pumping well operating condition recognition method based on multi-source consistency deep learning. Specifically, in order to avoid damping coefficient, "division by zero" and statistical data uncertainty, we select the measured dynamometer cards and the measured electrical power cards (both of which are high-dimensional binary images) as multi-source data. We adopt the framework of WRN28-2 convolutional neural network, and achieve the maximum consistency interactive learning based on features complements of different information sources through the combination of multi-head self-attention mechanism, feedforward neural network and cosine similarity measurement. Then, we apply the adaptive attention mechanism to achieve the optimal global feature representation of multi-source fusion by learning the weights of feature representations of different information sources, thus improving the existing operating condition feature extraction technology by capitalizing on all the multi-source information. The combination of multi-source fusion features and semi-supervised class-aware contrastive learning can further utilize the massive unlabeled data and a few multi-source labeled data to improve the operating condition recognition accuracy and extend its applications.

The contributions of this paper are as follows:

1) An efficient multi-source feature fusion deep learning algorithm for the operating condition is proposed. Based on the framework of the WRN28-2 convolutional neural network, the algorithm uses a multi-head self-attention mechanism and feedforward neural network to extract the deep image features of each measured information source respectively, and introduces consistency constraint loss based on cosine similarity measurement to ensure the maximum similarity of the final feature representations from different information sources. Then, the optimal global feature representation of multi-source fusion is obtained by using an adaptive attention mechanism to learn the weights of feature representations of different information sources.

2) A multi-source semi-supervised class-aware contrastive learning operating condition recognition method is proposed, which effectively integrates the multi-source feature fusion deep learning algorithm with the semi-supervised class-aware contrastive algorithm. Its objective function consists of four parts: supervised loss function, multi-source information consistency constraint loss function, unsupervised loss function and class-aware contrastive loss function. The recognition bias of the same category in multi-source feature space is reduced by class-aware contrastive learning, and the discriminant ability of the semi-supervised deep learning model is highly enhanced.

3) The effectiveness and practicability of the proposed method is verified experimentally. Compared with state-of-the-art multi-source semi-supervised learning recognition methods, traditional single-source recognition methods or multi-feature connection recognition methods, different attention mechanism learning recognition methods, and different contrastive learning

recognition methods, the proposed method exhibits improved recognition accuracy in different proportions of labeled training data, especially with 10\% increase. In addition, key hyperparameter analysis and ablation experiments further verify the effectiveness and practicability of the proposed method.

II. Related Works

In this part, we first give a short introduction to the information sources of the sucker-rod pumping well production system in terms of the operating principle of the system, and then we present the related works on operating condition recognition research respectively, using dynamometer cards, electrical parameters and multiple information sources.

A. Information Source Analysis of the Sucker-Rod Pumping Well Production System

The architecture diagram of the sucker-rod pumping well production system [29] is shown in Fig. 1.

It can be observed that the system is composed of 15 parts: (1) Standing valve; (2) Pump barrel; (3) Traveling valve; (4) Plunger; (5) Dynamic fluid level; (6) Sucker rod; (7) Casing; (8) Tubing; (9) Beam hanger; (10) Horse head; (11) Beam; (12) Link rod; (13) Crank; (14) Induction motor; (15) Electronic control cabinet.

The operating principle of the system is as follows: First, the high-speed rotating motion of the induction motor will be converted into the mechanical up-and-down swing of the beam by the crank connecting to the link rod. Then, the sucker rod connected to the beam hanger suspended from the horse head drives the pump plunger in the underground wellbore to move up and down through the mechanical swing of the beam. Finally, the mixed fluid in the wellbore is pumped to the surface through the tubing.

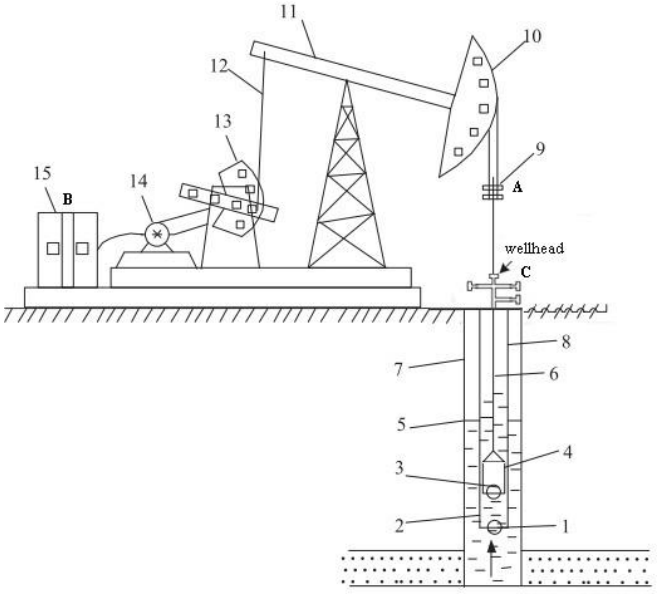


Fig. 1. The architecture diagram of the sucker-rod pumping well production system.

As the key downhole production equipment, the oil well pump mainly consists of three parts: pump barrel, plunger, and valves (see (1) and (3) in Fig. 1). One operation of the oil well pump includes valves opening and closing during one stroke.

In the sucker-rod pumping well production system, the measured information mainly comes from three sources: dynamometer cards (see A in Fig. 1), electronic parameters (see B in Fig. 1), wellhead data (wellhead temperature, wellhead pressure, etc., see C in Fig. 1). The measured dynamometer cards are binary images composed of the displacement and load of the polish rod (see (9) in Fig. 1), which are usually with high dimension and strong noise, and can reflect the situation of the wellbore and the stratum in real-time. The existing operating condition recognition technology using dynamometer cards mainly uses the pump dynamometer card data (transformed from the measured dynamometer card data after denoising, but may occur the damping coefficient problem). The measured electrical parameters are time-series signals, which can reflect the situation of the ground and the stratum in real-time. It is worth mentioning that the "division by zero" problem may occur when using the electronic parameters for operating condition recognition.

In the production process, there are also some non-measured information sources, such as dynamic fluid level (see (5) in Fig. 1), liquid-producing capacity, working time, etc. These statistical production data can also be regarded as information sources. However, the present research on operating condition recognition usually focuses on effectively utilizing the aforementioned two types of information sources.

B. The Operating Condition Recognition Technology Review

In general, the research on the operating condition recognition technology of the sucker-rod pumping well production system is mainly focused on three aspects: dynamometer card-based, electronic parameter-based and multi-source-based.

1) The dynamometer card-based recognition technology: The dynamometer card-based recognition method is the most widely used one, which is mainly implemented by ground or pump dynamometer cards combined with some artificial intelligence technology. For example, Li et al. [30] employed the moment feature and the “four-point method” to extract pump dynamometer card features and used SVM combined with particle swarm optimization to recognize working conditions. Zheng et al. [31] used the moment feature and the astr polygon decomposition centroid localization algorithm to extract ground dynamometer card features and used HMM combined with clonal selection optimization to recognize working conditions. Zhang et al. [32] extracted the feature of dynamometer cards by the fast discrete curvelet transform and utilized a sparse multi-graph regularized extreme-learning machine to recognize working conditions. The above working condition recognition methods involve a large number of complex feature parameter calculation and need a large amount of labeled data for training, thus greatly affecting the working

condition recognition performance. To this end, recently, deep learning is applied to extract the features in order to improve the recognition performance. Ye et al. [3] proposed an improved CNN to automatically obtain the deep feature expression of the dynamometer card data, overcoming the limitations of traditional feature extraction methods and achieving better recognition accuracy. But this method requires lots of labeled working condition data for model training. In the case of few-shot learning, He et al. [10] proposed to compress the raw working condition data through a 4-dimensional time-frequency feature extraction method, and then optimize parameters of the convolutional shrinkage neural network through meta-learning. This method overcomes the limitation of a large number of labeled data requirements for training, but it reveals the neglect of the massive unknown working condition data. In addition, there are some precision errors in the calculation of the characteristic data value based on the mechanism analysis.

2) The electrical parameter-based recognition technology: The electronic parameter-based operating condition recognition technology works with lower cost and higher reliability. Zheng et al. [33] proposed to extract feature data of the measured electric power signal based on mechanism analysis and accomplish the operating condition recognition by HMM. In addition, Chen et al. [34] deduced the electric power card model by considering the angular velocity of the crank and the friction and inertia of the 4-bar linkage and implemented the operating condition recognition by establishing a feature atlas of electric power cards. In order to further improve the working condition recognition effect based on electrical parameter sequence signals, researchers also introduced deep learning technology to extract deeper features of electrical parameter data. Wei et al. [4] realized a deep and broad learning system by motor power data, utilizing CNN to extract features of the electric power signal and employing the broad learning system to recognize the operating condition. However, because of " division by zero", the recognition accuracy based on electronic parameters is far from satisfactory. In addition, the current deep learning-based working condition recognition technology using electrical parameter data also needs a lot of labeled working condition training data.

3) The multiple information source-based recognition technology: Although it is easy to classify operating conditions using single-source data, it is also prone to error. Therefore, researchers started to develop operating condition recognition models based on multiple information sources. Zheng et al. [35] applied seven feature data such as sucker-rod specifications, pump diameter, stroke, speed, moisture content, gas-liquid ratio and working condition to a HMM model. Zhang et al. [36] adopted feature data such as dynamometer cards, power, dynamic fluid level, and liquid-producing capacity. Liu et al. [37] selected feature data such as dynamometer cards, work day time, and liquid-producing capacity. However, all the above operating condition recognition methods use non-measured production statistics data, which will affect the robustness of the recognition model. Zhou et al. [38] put forward a semi-supervised operating condition recognition technology integrating four measured data, namely, ground dynamometer cards, electric power signal, wellhead temperature, and wellhead pressure. The technical model has good performance and wide application, but the recognition accuracy of multi-classification conditions needs to be improved. Deep learning technology can improve model performance by deeply mining multi-source fusion feature expressions. Li et al. [5] developed an improved deep learning-based operating condition recognition method by the fusion of Fourier descriptor features and graphic features based on dynamometer cards. Abdurakipov et al. [7] used pump dynamometer cards, pump inlet pressure and temperature production data to establish an operating condition identification model with transformer. In summary, at present, the multi-source operating condition recognition technology based on deep learning mostly relies on lots of labeled data, and the multi-source fusion technology is mostly traditional. Recently, in the multi-source semi-supervised deep learning classification recognition research, Wu et al. [39] proposed a multi-view semi-supervised learning method based on graph convolutional network to improve model performance under label-scarce situations. This method maps data from multiple views into a low-dimensional space and uses Laplacian embedding technology to preserve the structural information between data, thereby achieving the fusion and representation learning of multi-view data. In addition, this method implies that we can try to explore the best consistent connection relationship between multiple views to obtain better recognition effects.

To sum up, the research on the operating condition recognition technology has achieved remarkable results, especially with the introduction of deep learning technology, which overcomes the limitations of traditional manual design features. However, in order to establish an efficient and practical operating condition recognition model based on deep learning, limitations such as the requirement of lots of labeled data, the neglect of massive unknown data, and the improvement of the multi-source fusion strategy remain for further handling.

III. Multi-Source Consistency for Semi-Supervised Deep Learning

This section introduces our proposed multi-source consistency deep learning for semi-supervised operating condition recognition method from sucker-rod pumping wells, emphasizing the integration of a multi-source consistency strategy, which includes WRN28-2 convolutional neural network backbone, multi-head self-attention mechanism, feedforward neural network, and an adaptive attention-based feature fusion strategy. Furthermore, this section describes how our model employs the semi-supervised learning technique to make efficient use of limited labeled data alongside a large number of unlabeled data, thereby enhancing recognition performance.

A. Problem Definition

Before introducing our method, the key symbols and variables used in this paper are clearly described in Table I.

TABLE I. NOTATION DESCRIPTION

Notation	Description
$\mathcal{V}$	the number of information sources
D	multi-source data set
L^v	labeled data set of the v^{th} data source
U^v	unlabeled data set of the v^{th} data source
l	the number of labeled data
m	the number of unlabeled data
A_w	weak data enhancement method
A_s	strong data enhancement method
F	backbone network model
G	global feature after multi-source feature fusion
H	multi-source data feature representation after multi-source information consistency learning
ω^v	the attention weight of the data feature of the v^{th} data source
S^v	cosine similarity of data from the v^{th} data source
$\|\cdot\|_2$	the l_2 norm

Here, we study a semi-supervised operating condition recognition task based on multi-source data. Suppose there are $\mathcal{V}$ different data sources, and each data source contains the C class. Multi-source data set D consists of labeled data set L and unlabeled data set U, i.e. $D = \{L, U\}$. For labeled data set L, it is defined as $L = \{L^v | v = 1,2,\dots,\mathcal{V}\} = \{(x_i^v, y_i) | i = 1,2,\dots,l\}$, where L^v is the labeled data set of the v^{th} data source, x_i^v is the i^{th} labeled image of the v^{th} data source, $y_i \in \{1,2,\dots,C\}$ is the label corresponding to the i^{th} image, and l is the number of labeled data. Unlabeled data set U is defined as $U = \{U^v | v = 1,2,\dots,\mathcal{V}\} = \{u_i^v | i = l+1, l+2,\dots,l+m\}$, where U^v is the unlabeled data set of the v^{th} data source, u_i^v is the i^{th} unlabeled image of the v^{th} data source, m is the number of unlabeled data. In order to increase the data diversity and improve the accuracy of semi-supervised learning, this paper adopts two data enhancement methods: weak enhancement $A_w(\cdot)$ and strong enhancement $A_s(\cdot)$. Weak enhancement usually includes small-angle rotation, translation, scaling, etc. Strong enhancement includes cropping, color transformation, etc. For labeled data set L and unlabeled data set U, the weak enhanced data are represented as $A_w(L)$ and $A_w(U)$, and the strong enhanced data are represented as $A_s(L)$ and $A_s(U)$, respectively. In this paper, WRN28-2 model is used as backbone network $F(\cdot)$ to learn feature representations of multi-source data, which are used for subsequent consistency learning and semi-supervised learning tasks.

B. Model Overview

This paper proposes a semi-supervised sucker-rod pumping well operating condition classification method based on multi-source information consistency deep learning. The model structure is shown in Fig. 2. The algorithm consists of the following two key steps: multi-source information consistency learning and multi-source semi-supervised learning. In multi-source information consistency learning, firstly, the WRN28-2 convolutional neural network is used to extract the feature representation of multi-source data. For the multi-source feature representation of the same sample, the dependence relationship in the feature representation is captured by multi-head self-attention mechanism, and the deeper feature is extracted by non-linear transformation with feedforward neural network. In order to ensure the consistency of multi-source information, this paper introduces a multi-source consistency constraint loss function based on the cosine similarity to strengthen the correlation among different information sources. In addition, the attention mechanism is used to learn the weights of feature representations from different information sources to achieve an effective fusion of multi-source feature representations, so as to obtain the optimal global feature representation of each sample. In multi-source semi-supervised learning, two enhancement strategies, weak enhancement and strong enhancement, are applied to the unlabeled data respectively, and the corresponding predicted values are generated, and the unlabeled data with high confidence are selected to generate pseudo labels. To improve the prediction accuracy, consistency regularization is used to calculate the cross-entropy loss of the predicted value corresponding to the strong enhancement and the pseudo-label corresponding to the weak enhancement. In view of the possibility of pseudo label method being affected by confirmation bias in the training process, this paper introduces the class-aware contrastive learning method [28], aggregates samples of the same category in feature space, and disperses samples of different categories to reduce confirmation bias and improve the discriminant ability.

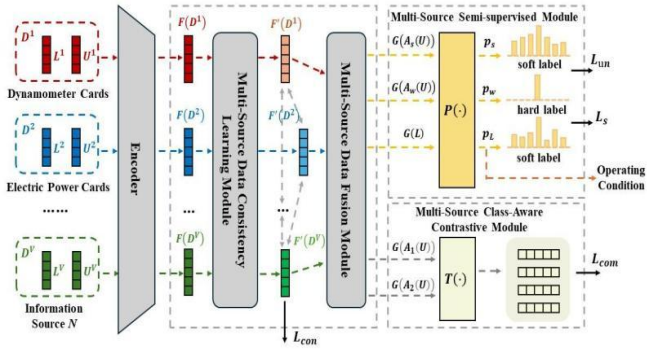


Fig. 2. Model structure diagram.

C. Consistency Learning of Multi-source Information

In this paper, the WRN28-2 backbone network is used to learn the feature representation $F(D)$ of multi-source data set D, where $F(D) = \{F(L), F(U)\}$, and the expressive power of the feature representation is enhanced by multi-head self-attention mechanism. Then, the multi-source consistency constraint loss function based on cosine similarity is used to ensure that the feature representation of the multi-source data is consistent in the feature space, thereby improving the generalization ability. The multi-source information consistency learning model structure is shown in Fig. 3.

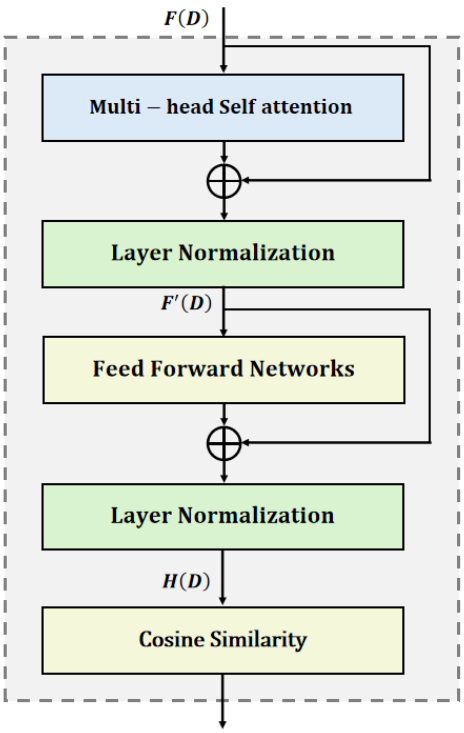


Fig. 3. Structure diagram of multi-source information consistency learning model.

Firstly, the feature representation of the multi-source data is sent into the multi-head self-attention module (MHSAM), and the feature representation ability of the multi-source data is enhanced by capturing the information of the input feature representation in their respective subspaces. Then, we implement residual connections between the input feature $F(D)$ and output feature $F'(D)$ of the MHSAM and carry out Layer Normalization (LN), as shown in Eq. (1):

$$F'(D) = LN\left(F(D) + MHSAM\left(F(D)\right)\right) \quad (1)$$

Then, the deeper features are extracted by nonlinear transformation through the feedforward network module (FFNM). Similarly, the residual connections and layer normalization are used to stabilize the training process, as shown in Eq. (2):

$$H(D) = LN\left(F'(D) + FFNM\left(F'(D)\right)\right) \quad (2)$$

where the FFNM consists of two linear mapping layers and Relu activation function, the former of which is $FFNM(x) = Relu(W_1 * W_2 * x)$, W_1 and W_2 are the weights of the linear mapping layer. $H(D)$ is the feature representation of multi-source data D after consistency learning of multi-source information. The above multi-source information consistency learning method is inspired by the Transformer network structure, and the Transformer coding layer is used to realize the relevant calculation.

There is an intrinsic consistency between different feature representations of multi-source data. To make full use of the consistency and improve the generalization of the model, we design a multi-source consistency constraint loss function based on cosine similarity to maximize the cosine similarity between feature representations of different sources, so as to ensure their consistency in feature space.

Specifically, for the multi-source feature representation $H(D)$ of the same sample, the l_2 norm ($\|\cdot\|_2$) is first converted into the standard feature representation $\|H(D)\|_2$ to eliminate the influence caused by the magnitude difference between the multi-source feature representation vectors. Then, the cosine similarity between the multi-source feature representation vectors is calculated to obtain the degree of consistency between the multi-source feature representations, which is calculated as in Eq. (3):

$$S^v = \|H(D^v)\|_2 \cdot \|H(D^v)\|_2^T \quad (3)$$

where S^v represents the cosine similarity of the v^{th} data source, and $(\cdot)^T$ represents the transpose of the vector.

Then, to ensure the consistency of feature representations from different data sources in feature space and enhance the generalization ability of the model, the difference matrix between any two data source similarity matrices is calculated, and the square of the norm of the difference matrix is calculated as the constraint loss function of multi-source information consistency learning, as shown in Eq. (4):

$$L_{con} = \sum_{i,j \in v} \left\| S^i - S^j \right\|_2^2 \quad (4)$$

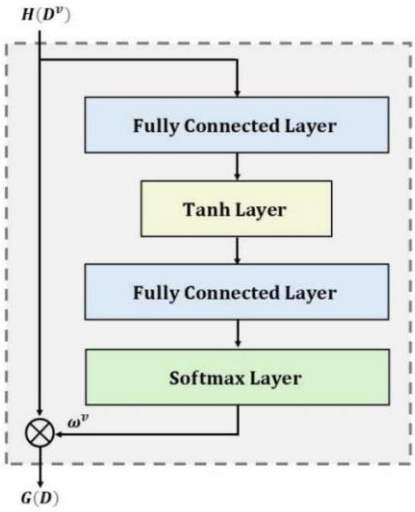


Fig. 4. Structure diagram of multi-source feature fusion model.

To further improve the performance, we further apply a multi-source feature fusion method based on adaptive attention mechanism, the model structure diagram is shown in Fig. 4. The adaptive attention mechanism is used to learn the weights of different data source feature representations, and the multi-source feature representations are then fused into a unified global feature representation to obtain the optimal global characteristics of each sample. Specifically, for the v^{th} data source, the data feature is represented as $H(D^v)$. In this paper, a network structure consisting of fully connected layer, followed by $tanh$ activation function and then by fully connected layer is adopted. The nonlinear combination of $H(D^v)$ is obtained by the network structure learning and the

corresponding attention score W_{out}^{v} is generated, as shown in Eq. (5):

$$W_{out}^{v} = W_2 \cdot [tanh(W_1 \cdot H(D^v) + b_1)] + b_2 \quad (5)$$

where W_1 and W_2 represent the weight of the two fully connected layers respectively, b_1 and b_2 are the corresponding bias vectors.

Then, the SoftMax function is used to normalize the attention score W_{out}^{v} , and the attention weight ω^v of each data source feature representation is obtained. The calculation is shown in Eq. (6):

$$\omega^v = softmax(W_{out}^{v}) = \frac{exp(W_{out}^{v})}{\sum_{n=1}^{N} exp(W_{out}^{v})} \quad (6)$$

where exp is the natural exponential function, and $\sum_{v=1}^{\mathcal{V}} \omega^v = 1$.

Finally, according to the normalized attention weight ω^v, the feature representation of each data source is fused into a unified global feature representation, and the fusion process is achieved by weighted summation, as shown in Eq. (7):

$$G(D) = \sum_{v=1}^{\mathcal{V}} \omega^v \cdot H(D^v) \quad (7)$$

where $G(D)$ is the global feature representation after fusion, and $G(D) = \{G(L), G(U)\}$.

D. Multi-Source Data Semi-supervise Module

After completing the multi-source information consistency learning and the multi-source feature fusion, this paper implements pseudo-label-based semi-supervised learning of multi-source data by referring to Fix-Match [40]. Two enhancement strategies, weak enhancement $G\big(A_w(U)\big)$ and strong enhancement $G\big(A_s(U)\big)$, are applied to unlabeled data, respectively. Then, corresponding predicted values p_w and p_s are generated using a fully connected layer $P(\cdot)$, which are calculated as shown in Eq. (8):

$$p_w = P\left(G\big(A_w(U)\big)\right)$$
$$p_s = P\left(G\big(A_s(U)\big)\right) \quad (8)$$

Then, weakly enhanced unlabeled data with high confidence predicted value are screened out, and the corresponding pseudo-label q_w are obtained using $argmax$ function, as shown in Eq. (9) :

$$q_w = argmax(\hat{p}_w) \quad (9)$$

where $\hat{p}_w$ represents the part of p_w, $\max(\hat{p}_w) > t$, $\max(\cdot)$ is the maximum function and t is the confidence threshold.

Finally, to increase the prediction accuracy, the consistency regularization is adopted to compute the cross-entropy loss between the predicted value of the strong enhancement and the pseudo-label of the weak enhancement. The detailed calculation of the semi-supervised loss L_{un} is shown in Eq. (10):

$$L_{un} = -\sum_{c=1}^{C} q_w^c \log(p_s^c) \quad (10)$$

where c represents the c^{th} class, and C is the total number of sample classes.

The supervised loss function L_s uses the cross-entropy loss of multi-source labeled data L, which is calculated as shown in Eq. (11):

$$L_s = -\sum_{c=1}^{C} y^c \log(p_L) \quad (11)$$

where p_L is the predicted value of labeled samples generated by $P(\cdot)$, $p_L = P\big(G(L)\big)$.

Pseudo-label based semi-supervised learning methods need to generate pseudo-labels independently in the process of semi-supervised learning, and are susceptible to confirmation bias in the training process. To solve this, this paper introduces a multi-source data class-aware contrastive learning method, aggregates samples of the same category in feature space, and drives samples of different categories away from each other. By the contrastive learning, the robustness of the model is increased and the impact of confirmation bias is reduced.

First, the unlabeled data are enhanced by two different data enhancement methods $A_1(\cdot)$ and $A_2(\cdot)$ respectively, obtaining $A(U) = [A_1(U), A_2(U)]$, and then the high-dimensional feature representation is mapped to a low-dimensional embedding vector through the projection layer $T[\cdot]$ composed of two fully connected layers. The specific calculation is shown in Eq. (12):

$$T[G(A(U))] = \{T[G(A_1(U))], T[G(A_2(U))]\} \quad (12)$$

The class-aware contrastive learning loss consists of two parts: narrowing the aggregation distance of enhanced data from the same category samples and widening the aggregation distance of enhanced data from the different category samples. The specific calculation is shown in Eq. (13):

$$L_{com} = \sum_{i \in I} \left[log \sum_{a \in O(i)} \left(exp\left(\frac{T\left[G\left(A^i(U)\right)\right] \cdot T\left[G\left(A^a(U)\right)\right]}{\tau} \right) - exp\left(\frac{T\left[G\left(A_1^i(U)\right)\right] \cdot T\left[G\left(A_2^i(U)\right)\right]}{\tau} \right) \right) \right] \quad (13)$$

where $I = \{1,2,\dots,M\}$, M is the number of unlabeled samples. $O(i)$ is the index of $M-1$ unlabeled samples except the index i. $T\left[G(A_1^i(U))\right]$ and $T\left[G(A_2^i(U))\right]$ are a set of positive sample pairs that are output from the i^{th} unlabeled sample after two data enhancement transformations and then through the projection layer. $T[G(A^i(U))]$ and $T[G(A^a(U))]$ are the corresponding negative sample pairs. τ is the temperature coefficient.

E. Objective Function

The loss function set includes four parts: supervised loss function L_s of labeled data, multi-source information consistency constraint loss function L_{con}, unsupervised loss function L_{un} and class-aware contrastive loss function L_{com}. The total loss function L is calculated using Eq. (14):

$$L = L_s + \lambda_{con} L_{con} + \lambda_{un} L_{un} + \lambda_{com} L_{com} \quad (14)$$

where λ_{con}、λ_{un} and λ_{com} are the weights of multi-source information consistency constraint loss, unsupervised loss and class-aware contrastive loss, respectively.

IV. Experimental Results

The experimental dataset is obtained from an oilfield in China with a high-pressure and low-permeability thin oil reservoir block. The operating condition samples used in the experiment are selected strictly according to the operation records of the oil wells. The measured dynamometer cards (binary images composed of polish rod displacement and load) and measured electric power cards (binary images composed of polish rod displacement and motor power) are included in each operating condition sample. The measured dynamometer card data and electric power card data are composed of the data points collected at the actual corresponding acquisition time on the production site. The sample set is built by the accumulation of operating condition samples of 60 sucker-rod pumping wells for 3 years and includes 11 typical operating conditions, each operating condition has 150 samples, for a total of 1650 samples. The 11 typical operating conditions include normal, assist-blowing, lack of supply liquid, rod cutting, stuck pump, traveling valve failing, wax precipitation, pump leakage, tubing leakage, standing valve leakage, traveling valve leakage.

The experiment is run on Python 3.8.10/pytorch 1.10.0 and Linux 5.4.0/24 GB NVIDIA RTX3090 GPU. The training set and the test set each have 825 samples, of which each operating condition contains 75 samples. The experimental results are averaged 10 times.

To verify the effectiveness and practicability, experiments are carried out from five aspects: (1) The verification of the superiority of the basic model. The learning models based on the classical residual network are compared in the case of full labeled training samples. (2) The verification of the effectiveness of the proposed method. On the basis of the first part of the experiment, different operating condition recognition methods are compared in terms of the operating condition recognition effect with full labeled training samples. (3) The verification of the practicability. On the basis of the second part of the experiment, different operating condition recognition methods are compared in terms of the operating condition recognition effect with different proportions labeled training samples. (4) Hyperparameter analysis. The impact of the four main hyperparameters of the objective function λ_{con}、λ_{un}、λ_{com} and the number of heads of the multi-head attention mechanism are analyzed, respectively. (5) Ablation test.

A. The Basic Model Analysis

To verify the superiority of the designed infrastructure, in the case of full labeled training samples, based on the two main learning models of the depth and width of the classical residual network combined with class-aware contrastive learning (CCL), we compare the results of various operating condition recognition methods based on traditional single-source operating condition recognition methods (WRN28-2/10_CCL, Resnet18/34/50/101/152_CCL,), traditional multi-source connection operating condition recognition methods (MCWRN28-2/1_CCL, MCResnet18/34/50/101/152_CCL), multi-source interactive consistent learning operating condition recognition methods (MIWRN28-2/10_CCL), and multi-source data online augmented interactive consistent learning operating condition recognition methods (MAIWRN28-2_CCL (the proposed method in this paper), MAIWRN28-10_CCL). In the proposed method, the pseudo-label threshold τ is 0.95; the learning rate is 0.02; the number of heads of the multiple self-attention mechanisms in the transformer encoder layer is 2; the middle layer dimension of FFNM is set to 512; λ_{un} and λ_{com} are set to 1, and λ_{con} is set to 10. The comparison results of various operating condition recognition methods based on the classical residual network combined with CCL are shown in Table II.

TABLE II. The Comparison Results of Various Operating Condition Recognition Methods based on the Classical Residual Network Combined with CCL (Average Accuracy, %)

Operating condition recognition methods	Measured dynamometer cards	Measured electrical power cards	Measured dynamometer cards and electrical power cards
Resnet18_CCL	93.67	95.83	--
Resnet34_CCL	92.87	95.25	--
Resnet50_CCL	88.55	94.71	--
Resnet101_CCL	84.09	94.27	--
Resnet152_CCL	89.79	95.27	--
WRN28-2_CCL	93.72	**95.93**	--
WRN28-10_CCL	**94.34**	95.67	--
MCResnet18_CCL	--	--	96.55
MCResnet34_CCL	--	--	96.46
MCResnet50_CCL	--	--	95.39
MCResnet101_CCL	--	--	95.03
MCResnet152_CCL	--	--	94.18
MCWRN28-2_CCL	--	--	97.55
MCWRN28-10_CCL	--	--	97.23
MIWRN28-2_CCL	--	--	97.96
MIWRN28-10_CCL	--	--	97.41
MAIWRN28-2_CCL(Ours)	--	--	**98.79**
MAIWRN28-10_CCL	--	--	98.34

From Table II:

1) Among the operating condition recognition methods based on the classical residual network combined with CCL, recognition methods based on WideResNet have significantly better effects than those based on Resnet, which reflects that WideResNet can obtain better performance by increasing network width and reducing network depth.

2) Among the traditional single-source operating condition recognition methods, in terms of recognition effects, recognition methods based on the measured electrical power cards are obviously superior to those based on the measured dynamometer cards, which shows that the measured electrical power cards adopted in this paper can better overcome the influence of "zero division" on the operating condition recognition accuracy. In addition, the traditional single-source

operating condition recognition methods based on WideResNet are more accurate than those based on ResNet. Meanwhile for the measured dynamometer cards, WRN28-10_CCL has the best recognition effect, and for the measured electrical power cards, WRN28-2_CCL has the best recognition accuracy.

3) Among the traditional multi-source connection operating condition recognition methods, MCWRN28-2/10_CCL, MCResnet18/34_CCL have better recognition accuracy than those based on traditional single-source methods, but MCResnet50/101/152_CCL are the opposite, which indicates the technology limitations of the traditional multi-source connection operating condition recognition methods. In addition, MCWRN28-2/10_CCL have better recognition effects than MCResnet18/34_CCL.

4) Among the operating condition recognition methods based on multi-source consistency learning, MIWRN28-2/10_CCL are more accurate than those recognition methods based on traditional single source and traditional multi-source connection, which reflects the advantage of the multi-source consistency learning method proposed in this paper. Further, this paper carries out the online augmented processing on the multi-source consistency learning to further improve the recognition accuracy. MAIWRN28-2/10_CCL well reflect this strategy for the operating condition recognition, of which MAIWRN28-2_CCL performs best and is selected and used in subsequent experiments in this paper.

B. Multi-Source Learning Operating Condition Recognition with Full Labeled Training Samples

To verify the effectiveness, on the basis of the first part of the experiment, the method proposed in this paper is compared with the interpretable multi-view graph convolutional network recognition method (IMvGCN) [39], and the multi-source consistency learning operating condition recognition methods based on different contrastive learning (MAIWRN28-2_InfoCL), based on different attention mechanism fusion (MAIWRN28-2_SELF_CCL, MAIWRN28-2_SE_CCL). The comparison results of different multi-source learning operating condition recognition methods with full labeled training samples are shown in Table III.

From Table III:

1) Compared with the multi-source learning operating condition recognition method based on InfoNCE contrastive learning, our method improves the recognition accuracy by about 1.2%, showing the superiority of CCL adopted in this paper.

2) Compared with the multi-source learning operating condition recognition methods based on SE attention mechanism and self-attention mechanism fusion, the proposed method improves the recognition accuracy by about 0.5% and 0.6%, respectively, showing the advantage of fusion learning of the attention mechanism adopted in this paper.

3) Compared with IMvGCN, our method improves the recognition accuracy by about 2.2%, showing the superiority of the multi-source consistency learning adopted in this paper.

4) Compared with the multi-source learning operating condition recognition methods based on different multi-source learning algorithms, different contrastive learning algorithms and different attention mechanism fusion algorithms, our method has higher recognition accuracy, thus verifying the effectiveness of the proposed method.

TABLE III. THE COMPARISON RESULTS OF VARIOUS OPERATING CONDITION RECOGNITION METHODS BASED ON THE CLASSICAL RESIDUAL NETWORK COMBINED WITH CCL (AVERAGE ACCURACY, %)

Methods	Measured dynamometer cards and measured electrical power cards
MAIWRN28-2_InfoCL	97.63
MAIWRN28-2_SELF_CCL	98.17
MAIWRN28-2_SE_CCL	98.32
IMvGCN	96.57
MAIWRN28-2_CCL(Ours)	**98.79**

C. Multi-Source Learning Operating Condition Recognition with Different Proportions of Labeled Training Samples

To verify the practicability, on the basis of the two parts of the experiment, we add the multi-source consistency learning semi-supervised CCL recognition method based on CoMatch [41] (MAIWRN28-2_CoMatch_CCL), and compare the recognition results in 5 groups of labeled training samples with different proportions (1%, 10%, 30%, 50%, 70%). The comparison results of different multi-source learning operating condition recognition methods with different proportions of labeled training samples are shown in Table IV.

TABLE IV. COMPARISON RESULTS OF DIFFERENT MULTI-SOURCE LEARNING OPERATING CONDITION RECOGNITION METHODS WITH DIFFERENT PROPORTIONS OF LABELED TRAINING SAMPLES (AVERAGE ACCURACY, %)

Methods	1% (n=11)	10% (n=88)	30% (n=25)	50% (n=41)	70% (n=58)
MAIWRN28-2_InfoCL	80.64	91.03	93.64	95.45	96.86
MAIWRN28-2_SELF_CCL	81.17	92.45	95.55	96.08	97.18
MAIWRN28-2_SE_CCL	81.45	92.85	94.81	96.53	97.52
MAIWRN28-2_CoMatch_CCL	79.88	91.12	94.34	95.84	97.56
IMvGCN	**83.79**	92.68	93.65	95.81	96.35
MAIWRN28-2_CCL(Ours)	82.98	**94.13**	**96.18**	**97.20**	**98.25**

From Table IV:

1) In five groups of different proportions labeled training samples, compared with various multi-source semi-supervised learning recognition methods (MAIWRN28-2_SE_CCL, MAIWRN28-2_CoMatch_CCL, MAIWRN28-2_SELF_CCL, MAIWRN28-2_InfoCL, IMvGCN), our method has obtained higher accuracy (except 1%), reflecting the advantage of the semi-supervised learning algorithm adopted.

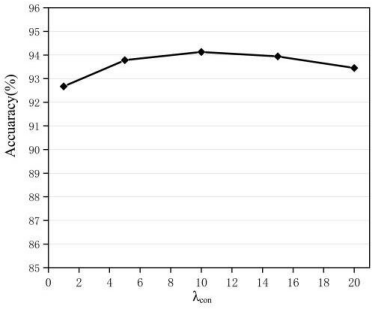
2) Among the recognition results of five groups of labeled training samples with different proportions, 10% (8 labeled training samples for each class) has a particularly significant recognition effect, which shows that our method can effectively

use a large number of unlabeled training samples to further improve the recognition accuracy, thus verifying the practicability of our method.

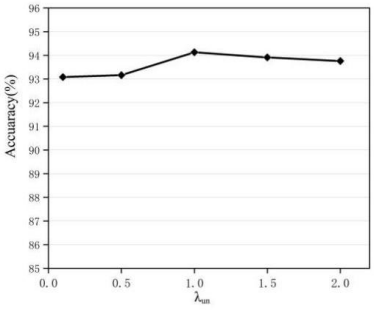
D. Key Hyperparameter Analysis

To explore the influence of hyperparameters on the performance of the operating condition recognition model, an extensive hyperparameter analysis of the proposed method is carried out with 10% proportion labeled training samples. We focus on four major hyperparameters: multi-source information consistency constraint loss weight λ_{con}, unsupervised loss weight λ_{un}, class-aware contrastive loss weight λ_{com}, and the number of heads of multi-head attention in Transformer coding layer. The value range of λ_{con} is [1,5,10,15,20], the value range of λ_{un} is [0.1, 0.5, 1,1.5, 2], the value range of λ_{com} is [0.1, 0.5, 1,1.3, 1.5], the value range of the head number of multi-head attention is [1,2,4,8,16].

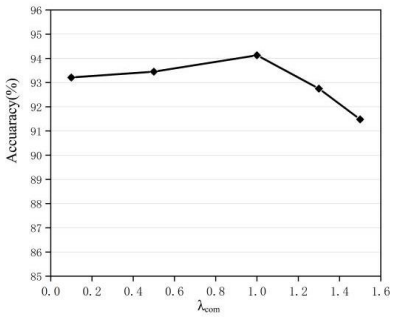
Fig. 5 shows the classification accuracy on the selected data set in different hyperparameter settings. It can be seen from Fig. 5 (a) and (b) that our method shows high stability within certain parameter ranges of λ_{con} and λ_{un}, that is, when λ_{con} , λ_{un} change within a certain ranges, the model accuracy fluctuates only slightly. It can be seen from Fig. 5(c), with the increase of λ_{com}, the impact on the model gradually increases. It can be seen from Fig. 5(d), when the number of heads of attention varies within the value range, the accuracy shows a trend of first increasing and then decreasing, and when the head=2, the accuracy is optimal. By selecting the optimal hyperparameter configuration, the overall performance of our proposed method can be effectively improved.



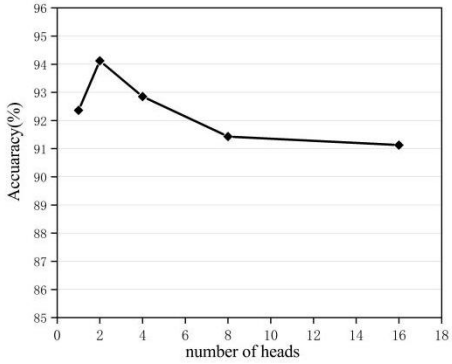
(a) Accuracy with different λ_{con}, λ_{un}=1.0, λ_{com}=1.0, multi-head attention with 2 heads.



(b) Accuracy with different λ_{un} , λ_{con}=10, λ_{com}=1.0, multi-head attention with 2 heads.



(c) Accuracy with different λ_{com} , λ_{con}=10, λ_{un}=1.0, multi-head attention with 2 heads.



(d) Accuracy with different heads of multi-head attention, λ_{com}=1.0, λ_{con}=10, λ_{un}=1.0.

Fig. 5. Key hyperparameter analysis of the proposed method in 10% proportion labeled training samples.

E. Ablation Experiments

Ablation experiments are conducted from three aspects with 10% proportion of labeled training samples: (1) the verification of the effectiveness of the multi-feature fusion algorithm based on attention mechanism. We compare the operating condition recognition results with the pooling learning multi-feature fusion method (MAIWRN28-2_Pooling_CCL) and the fixed-weight multi-feature fusion method (MAIWRN28-2_Fixed-weight_CCL). (2) the verification of the effectiveness of the contrastive learning algorithm. We compare with the multi-feature fusion method based on non-contrastive learning (MAIWRN28-2). (3) the verification of the effectiveness of online augmented data processing. We compare with the multi-feature fusion method without online augmented data processing (MIWRN28-2_CCL). The comparison results of ablation experiments are shown in Table V.

TABLE V. THE COMPARISON RESULTS OF VARIOUS OPERATING CONDITION RECOGNITION METHODS BASED ON THE CLASSICAL RESIDUAL NETWORK COMBINED WITH CCL (AVERAGE ACCURACY, %)

Methods	10% labeled (n=88)
MIWRN28-2_CCL	91.27
MAIWRN28-2	90.18
MAIWRN28-2_Fixed-weight_CCL	90.46
MAIWRN28-2_Pooling_CCL	92.18
MAIWRN28-2_CCL(Ours)	**94.13**

From Table V:

1) Compared with the multi-feature fusion method based on pooling learning and fixed weights, the multi-feature fusion method proposed in this paper can obtain better global feature representation by adaptive adjustment of the feature weights of different information sources through the attention mechanism, thus improving the condition recognition effect.

2) Compared with the multi-feature fusion method based on non-contrastive learning, the proposed method can effectively solve the common confirmation bias problem in semi-supervised deep learning through class-aware contrastive learning, thus further improving the model performance.

3) By using the augmented technique based on geometry operations such as rotation, scaling and expansion, our method can significantly expand the size and diversity of the training set without the need for collecting additional data, thereby improving the model's performance on unknown data.

V. Conclusion and Future Work

In order to solve the limitations of operating condition recognition in the context of big data oil production, such as the bottleneck of traditional single or multi-source operating condition recognition technology, high demand for lots of labeled operating condition data, and inability to make use of massive unknown operating condition data resources, we propose a semi-supervised sucker-rod pumping well operating condition recognition method based on multi-source consistency deep learning. The proposed approach first draws on the design idea of the Transformer coding layer and extracts the deep features of the measured dynamometer cards and the measured electrical power cards, respectively, through a multi-head self-attention mechanism and feedforward neural network. Then, combined with cosine similarity measurement, it realizes the maximum consistency of interactive learning on the complementary feature information of different information sources. Further, the optimal global feature representation of multi-source fusion is obtained by self-learning the weights of the feature representations of different information sources by the attention mechanism. Finally, the multi-source fusion feature is combined with the multi-source semi-supervised class-aware contrastive learning to build the operating condition recognition model and carry out the operating condition recognition. A large number of operating condition recognition comparison experiments show our method can not only mine the deep characteristics of multiple measured information source data but also effectively utilize massive unknown operating condition data in a small amount of labeled operating condition data to further improve the recognition effect and engineering practicability. Key hyperparameter analysis and ablation experiments further verify the effectiveness of the proposed method.

Future Work

The future directions of this paper worth further exploration are: (1) The study of the influence of heterogeneous data from multiple sources on the universality and generalization of the operating condition recognition model. In this paper, two kinds of image information sources, dynamometer cards and electrical power cards, are used to carry out research by combining the adaptive interactive learning fusion method based on the attention mechanism. The sucker-rod pumping well production system is a complex nonlinear system with mechanical-electrical-hydraulic coupling, and operating conditions are complex and changeable. Therefore, it is worth studying more comprehensive and appropriate information sources and more effective fusion learning methods. (2) The exploration of more efficient semi-supervised deep learning methods to further optimize the operating condition recognition model performance in insufficient labeled samples.

Acknowledgment

This work was supported by the Natural Science Foundation of Shandong Province under Grant ZR2021MF031.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] Z. Huang, K. Li, C. Ke, H. Duan, M. Wang, S. Bing, "An intelligent diagnosis method for oil-well pump leakage fault in oilfield production Internet of Things system based on convolutional attention residual learning," Engineering Applications of Artificial Intelligence, vol. 126, p., 106829, 2023.

[2] W. Wu, X. Xing, H. Wei, B. Li, X. Wang, "Fault diagnosis of pumping system based on multimodal attention learning (CBMA Learning)," Journal of Process Control, vol. 128, p. 103006, 2023.

[3] Z. W. Ye, Q. J. Yi, "Working-condition diagnosis of a beam pumping unit based on a deep-learning convolutional neural network," Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, vol. 236, no. 5, pp, 2559-2573, 2022.

[4] J. L. Wei, X. W. Gao, "Fault diagnosis of sucker rod pump based on deep-broad learning using motor data," IEEE Access, vol. 8, pp. 222562-222571, 2020.

[5] J. N. Li, J. Shao, W. Wang, W. H. Xie, "An evolutional deep learning method based on multi-feature fusion for fault diagnosis in sucker rod pumping system," Alexandria Engineering Journal, vol. 66, pp. 343-355, 2023.

[6] Y. He, Z. Guo, X. Wang, W. Abdul, "A Hybrid Approach of the Deep Learning Method and Rule-Based Method for Fault Diagnosis of Sucker Rod Pumping Wells," Energies, vol. 16, no. 7, p. 3170, 2023.

[7] S. S. Abdurakipov, M. Dushkin, D. Del'tsov, E. B. Butakov, "Diagnostics of Oil Well Pumping Equipment by Using Machine Learning," Journal of Engineering Thermophysics, vol. 33, no. 1, pp. 39-54, 2024.

[8] H. Li, H. Niu, Y. Zhang, Z. Yu, "Research on indirect measuring method of dynamometer diagram of sucker rod pumping system based on long-short term memory neural network," Journal of Intelligent & Fuzzy Systems, vol. 45, no. 3, pp. 4301-4313, 2023.

[9] X. Wang, Y. He, F. Li, Z. Wang, X. Dou, H. Xu, et al., "A working condition diagnosis model of sucker rod pumping wells based on big data deep learning," International petroleum technology conference (IPTC), Beijing, China, 2019, pp. 317–326.

[10] Y. P. He, C. Z. Zang, P. Zeng, M. X. Wang, Q. W. Dong, G. X. Wan, et al., "Few-shot working condition recognition of a sucker-rod pumping system based on a 4-dimensional time-frequency signature and meta-learning convolutional shrinkage neural network," Petroleum Science, vol. 20, no. 2, pp. 1142-1154, 2023.

[11] Z. Ma, Y. Chen, Y. Fan, X. He, W. Luo, J. Shu, "An improved AoT-DCGAN and T-CNN hybrid deep learning model for intelligent diagnosis of PTCs quality under small sample space," Applied Sciences, vol. 13, no. 15, p. 8699, 2023.

[12] D. Z. Hao and X. W. Gao, "Unsupervised Fault Diagnosis of Sucker Rod Pump Using Domain Adaptation with Generated Motor Power Curves," Mathematics, vol. 10, no. 8, p. 1224 ,2022.

[13] Y. P. He, H. B. Cheng, P. Zeng, C. Z. Zang, Q. W. Dong, G. X. Wan, et al., "Working condition recognition of sucker rod pumping system based on 4-segment time-frequency signature matrix and deep learning," Petroleum Science, vol. 21, no.1, pp. 641-653, 2024.

[14] R. Zhao, C. Wang, H. Zhao, C. Xiong, J. Shi, X. Zhang, et al., "Research and Application of Rod Pump Working Condition Diagnosis and Virtual Production Metering Based on Electric Parameters," In SPE Middle East Oil and Gas Show and Conference , SPE, 2021.

[15] P. Zhang, T. Li, G. Wang, C. Luo, H. Chen, J. Zhang et al., "Multi-source information fusion based on rough set theory: A review," Information Fusion, vol. 68, pp. 85-117, 2021.

[16] J. Huang, Z. Chen, Q. J. Wu, C. Liu, H. Yuan, W. He, "CATFPN: Adaptive feature pyramid with scale-wise concatenation and self-attention," IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 12, pp. 8142-8152, 2021.

[17] A. Kumar, P. Rai, H. Daume, "Co-regularized multi-view spectral clustering," Advances in neural information processing systems, vol. 24, 2011.

[18] G. Qian, S. Sural, Y. Gu, S. Pramanik, "Pramanik. Similarity between Euclidean and cosine angle distance for nearest neighbor queries," In Proceedings of the 2004 ACM symposium on Applied computing, 2004, pp. 1232-1237.

[19] S. Zagoruyko, N. Komodakis, "Wide Residual Networks," In British Machine Vision Conference 2016. British Machine Vision Association, 2016, arxiv preprint arxiv:1605.07146.

[20] Y. Liu, Y. Zhang, Y. Wang, F. Hou, J. Yuan, J. Tian, et al. "A survey of visual transformers," IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 6, pp. 7478 - 7498, 2023.

[21] Z. Wang, J. Xuan, T. Shi, "Multi-source information fusion deep self-attention reinforcement learning framework for multi-label compound fault recognition," Mechanism and Machine Theory, vol. 179, p. 105090, 2023.

[22] X. Yang, Z. Song, I. King, Z. Xu, "A survey on deep semi-supervised learning," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 9, pp. 8934-8954, 2022.

[23] B. Zhang, Y. Wang, W. Hou, H. Wu, J. Wang, M. Okumura, et al., "Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling," Advances in Neural Information Processing Systems, vol. 34, pp. 18408-18419, 2021.

[24] Y. Fan, A. Kukleva, D. Dai, B. Schiele, "Revisiting consistency regularization for semi-supervised learning," International Journal of Computer Vision, vol. 131, no. 3, pp. 626-643, 2023.

[25] W. Ma, F. Cheng, Y. Xu, Q. Wen, Y. Liu, "Probabilistic representation and inverse design of metamaterials based on a deep generative model with semi-supervised learning strategy," Advanced Materials, vol. 31, no. 35, p. 1901111, 2019.

[26] Z. Song, X. Yang, Z. Xu, I. King, "Graph-based semi-supervised learning: A comprehensive review," IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 11, pp. 8174-8194, 2022.

[27] Y. Gan, H. Zhu, W. Guo, G. Xu, G. Zou, "Deep semi-supervised learning with contrastive learning and partial label propagation for image data," Knowledge-Based Systems, vol. 245, p. 108602, 2022.

[28] F. Yang, K. Wu, S. Zhang, G. Jiang, Y. Liu, F. Zheng, et al., "Class-aware contrastive semi-supervised learning," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 14421-14430.

[29] A. Zhang, X. W. Gao, "Supervised dictionary-based transfer subspace learning and applications for fault diagnosis of sucker rod pumping systems," Neurocomputing, vol. 338, pp. 293-306, 2019.

[30] K. Li, X. W. Gao, Z. Tian, "Using the curve moment and the PSO-SVM method to diagnose downhole conditions of a sucker rod pumping unit," Petroleum Science, vol. 10, pp. 73–80, 2013.

[31] B. Y. Zheng, X. W. Gao, "Sucker rod pumping diagnosis using valve working position and parameter optimal continuous hidden Markov model," Journal of Process Control, vol. 59, pp. 1–12, 2017.

[32] A. Zhang, X. W. Gao, "Fault diagnosis of sucker rod pumping systems based on Curvelet Transform and sparse multi-graph regularized extreme learning machine," International journal of computational intelligence systems, vol. 11, pp. 428–437, 2018.

[33] B. Y. Zheng, X. W. Gao, X. Y. Li, "Fault detection for sucker rod pump based on motor power," Control Engineering Practice, vol. 86, pp. 37–47, 2019.

[34] D. C. Chen, R. Q. Zhou, H. X. Meng, Y. Peng, F. Chang, D. Jiang et al., "Fault diagnosis model of the variable torque pumping unit well based on the power-displacement diagram," IOP Conference Series: Earth and Environmental Science, 2019.

[35] B. Y. Zheng, X. W. Gao, X. Y. Li, "Diagnosis of Sucker Rod Pump based on generating dynamometer cards," Journal of Process Control, vol. 77, pp. 76–88, 2019.

[36] R. Zhang, Y. Yin, L. Xiao, "A real-time diagnosis method of reservoir-wellbore-surface conditions in sucker-rod pump wells based on multidata combination analysis," Journal of Petroleum Science and Engineering, vol. 198, p. 108254, 2021.

[37] S. Liu, C. S. Raghavendra, Y. Liu, K. Yao, O. Balogun, L. Olabinjo, et al., "Automatic early fault detection for rod pump systems," SPE Annual Technical Conference and Exhibition: OnePetro, Denver, Colorado, USA, 2011.

[38] B. Zhou, R. Niu, S. Yang, J. G. Yang, W. W. Zhao, "Multisource working condition recognition via nonlinear kernel learning and p-Laplacian manifold learning," Heliyon, vol. 10, no. 5, p. E26436, 2024.

[39] Z. Wu, X. Lin, Z. Lin, Z. Chen, Y. Bai, S. Wang, "Interpretable graph convolutional network for multi-view semi-supervised learning," IEEE Transactions on Multimedia, vol. 25, pp. 8593-8606, 2023.

[40] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, et al., "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," Advances in neural information processing systems, vol. 33, pp. 596-608, 2020.

[41] I. Hernandez-Sequeira, R. Fernandez-Beltran, Y. Xu, P. Ghamisi, F. Pla, "Semi-supervised classification for remote sensing datasets," International Conference on Image Analysis and Processing, Cham: Springer Nature Switzerland, 2023, pp. 463-474.