Entrepreneurship and Innovation
2025, VOL. 9, NO. 1, 74-87
DOI: 10.47297/wspeiWSP2516-253515.20250901

The spatiotemporal attention logging parameter prediction model based on historical feature separation of prior knowledge

Chenguang Zhu

College of Geophysics and Petroleum Resources Yangtze University, Wuhan, Hubei, 43029, China

ABSTRACT
In oil and gas exploration and development, logging is fundamental for reservoir identification and decision-making, relying on data about the physical properties of geological formations. Accurate analysis of logging reservoir parameters is critical, but existing data-driven prediction models often lack the capability to effectively extract reservoir parameter features, leading to weak generalization in practical scenarios. This paper addresses these limitations by proposing a novel model, SMHDS, specifically designed to predict high-information-density data, such as logging parameters. The model partitions the data into historical and current data, using the latter as prior knowledge and focusing on input parameters most relevant to the prediction target. It improves upon traditional attention mechanism by embedding spatial and temporal information into the attention mechanism, reallocating parameter weights for enhanced prediction accuracy.For the above method, this paper conducts three sets of experiments on actual data in an exploration area, with porosity as the target parameter. The performance of the model is tested from three perspectives: the rationality of the model structure, the accuracy of model prediction, and the generalization ability of the model. It is compared with three similar algorithms (CNN-LSTM, LSTM, MLP). The results indicate that the spatiotemporal model based on historical feature separation of prior knowledge, exhibits rational weight allocation for regression problems with logging parameters as the target. It effectively utilizes the relevant features of input parameters, resulting in smaller prediction errors compared to actual results, and demonstrates good data generalization capability.

KEYWORDS
Porosity; Well Logging; Feature Fusion; Deep Learning; Attention Mechanism.

In oil and gas exploration and development, logging holds significant research value by providing key data on the physical properties of formations surrounding the wellbore. Logging is essential for reservoir identification, evaluating reservoir quality, determining hydrocarbon-bearing zones, and designing effective oil and gas development plans, serving as the foundation for accurate reservoir characterization and optimized development [1-4]. Techniques for predicting logging reservoir parameters further enhance the application value of logging. By utilizing existing logging data to predict difficult-to-obtain or missing parameters, these techniques complement incomplete datasets, improving the accuracy and reliability of logging data while reducing costs and risks associated with logging. Furthermore, reservoir parameter prediction through logging provides additional support for data-driven reservoir modeling and analysis [5], making reservoir characterization and production optimization more precise and effective. This technology is of significant importance for enhancing the benefits of oil and gas field development.

Geological researchers typically analyze geological conditions by establishing empirical formulas or simplified stratigraphic models [6-8]. While these methods perform well in formations with good homogeneity, the strong anisotropy of formations [9] results in highly nonlinear relationships among logging parameters and with analysis targets [10], leading to significant limitations of these methods when facing relatively complex geological environments.

With the continuous development of logging technology, large and complex logging data require more advanced analysis and prediction methods. In recent years, deep learning, as a cutting-edge artificial intelligence technology, has been applied and researched in various fields [11-13]. The introduction of deep learning algorithms into the field of geophysics provides a new approach for studying logging parameters with high dimensionality and complex nonlinear relationships, offering significant benefits to geoscientists by providing large-scale, high-quality data, enabling them to more comprehensively and accurately understand the properties of underground rocks [14].

This paper focuses on the prediction of logging parameters from the perspective of deep learning. In this field, there have been studies where researchers used a composite network consisting of multiple BP (Backpropagation Neural Network) subnetworks for lithology classification [15]. Additionally, there are researchers who have established nonlinear relationships between spontaneous potential (SP), resistivity (RT), natural gamma (GR), and acoustic (AC) using the BP algorithm to reconstruct the acoustic curve [16]. Other scholars have also made improvements to the BP algorithm by increasing the parameters of hidden layers, optimizing the backpropagation process, and reconstructing logging parameters as well as calculating porosity in sandstone reservoirs [17-18]. Further, there are those who have added fuzzy logic and ensemble algorithm mechanisms to the MLP (Multilayer Perceptron) algorithm to predict shear wave velocity and PHIT with the aim of improving the generalization capability of the algorithm [19-20]. The algorithms employed by the aforementioned scholars have enhanced the nonlinear mapping ability between logging parameters compared to conventional empirical formulas [21-23], but such methods are limited to exploring the relationship between logging

parameters at the same depth, with less ability to investigate the nonlinear relationship between logging parameters at neighboring depths.

To extract spatial features of logging parameters in the neighborhood, some scholars and have approached this from the perspective of computer vision, treating logging parameter sequences as a special type of image data [24-25]. They have used the local receptive fields of convolutional neural networks (CNN) to extract longitudinal and transverse features of logging parameters through convolutional kernels. This idea is innovative and provides a new perspective for exploring the features of logging parameters. However, there are also shortcomings. The unit value of logging data carries a higher information density compared to conventional image pixels. Filtering logging parameters with convolutional kernels may lead to a loss of potential information entropy, and the convolution process, limited by kernel size, might not effectively preserve the context of logging parameters, resulting in a loss of relationships.

From the perspective of establishing the relationship between logging parameters in depth, researchers treat the depth of logging parameters as time and attempt algorithms based on the recurrent neural network (RNN) structure. Such algorithms typically allow features at the current time step to be influenced by the previous time step. Dongxiao Z et al. (2018) reconstructed logging parameters using cascaded Long Short-Term Memory (LSTM) [26]. Chen W et al. (2020) added network layers based on a single-layer LSTM and used multi-layer LSTM to predict porosity [27]. Some researchers have also improved the architecture of Bi-LSTM (Bidirectional Long Short-Term Memory) by incorporating convolution operations into each LSTM cell unit, enhancing the model's ability to extract spatial features [28]. However, these algorithms still have limitations. Due to the recursive nature of these networks, the weights from neighboring depths that can influence the current time step tend to weaken as the distance increases.

Considering the continuity and cyclicity of formations, which leads to a strong correlation between early and late data in logging sequences [29], some researchers have utilized global attention mechanisms to extract spatial features of logging parameters, enabling the generation of multiple logging parameters [30-31]. However, there are still areas for improvement in these methods. Although the attention mechanism, compared to recursive structures, benefits from parallel computing, it introduces more parameters due to its global correlation characteristics, resulting in increased computational demands. Additionally, the more complex feature interaction processes in these models may make it difficult to focus on key information while introducing more irrelevant features. Thus, balancing the highlighting of important information and suppressing redundant information has become a new research focus. This paper conducts detailed research on the aforementioned issues and conducts extensive experimental studies on (1) the selection of logging parameter attention interaction and (2) the transmission of logging parameter feature information in the model. Based on the analysis of the experimental results, the following contributions are made:

(1)Previous scholars treated logging parameters as "word vectors," while this paper, oriented toward the relationships between logging parameters, improves the structure of the attention module, making it more suitable for sequence prediction problems of logging parameters. (2)By introducing the prior knowledge of separating historical features from current features, a sequence attention model tailored for high-density feature extraction is proposed.

In summary, the current research hotspot in logging parameter prediction is how to highlight key information and suppress redundant information while extracting more logging parameter features. Faced with these challenges, this paper addresses the problem from two perspectives: 1) the selection of logging parameter attention interactions; 2) the transmission of logging parameter feature information in the model. The model designed in this paper (SMHDS) is a sequence attention model tailored for high-density feature extraction. By improving the calculation dimension of the attention module, the model focuses more on the spatial features between logging parameters. Through two model branches, the interference of redundant information is reduced, enabling the learning of sequential features of historical logging parameters and spatial features corresponding to target depths. And attention interaction is conducted at the end of the model to concentrate the weights of logging parameter features on key parts in both sequence and space. In the experimental section, this paper tested the rationality of the model modules, feature extraction capability, and generalization ability of the model on four datasets, demonstrating superior performance compared to existing methods.

The main contributions of this study are as follows: 1) Previous scholars treated logging parameters as "word vectors," while this paper, oriented toward the relationships between logging parameters, improves the structure of the attention module, making it more suitable for sequence prediction problems of logging parameters; 2) Introducing prior knowledge of separating historical features from current features reduces the loss of current depth features during the information transmission process in the model.

1 Methods and Principles

The construction of the model in this paper underwent extensive experimentation and computational screening, including but not limited to parameter adjustments for each module and comparisons with similar algorithms for each module's functionality. Ultimately, based on the main functions of the modules, it was divided into four key components.

Each component has specific functions and purposes to ensure the effectiveness and accuracy of the entire model in the prediction task. The structural diagram is shown in Figure 1:

1) Data Encoding Layer: This component receives the cleaned raw data and effectively encodes it. At this stage, this component first maps the raw data into feature vectors to ensure consistency among different parameters, facilitating the model to process the data. Then, the feature vectors are segmented, extracting the features of the current depth as prior knowledge for information focusing in the historical constraint layer, while the remaining historical features are used for feature attention interaction in the subsequent components.

2) Historical Feature Attention Interaction Layer: This component receives historical features from the data encoding layer and interacts through self-attention mechanism. Through this component, the model can automatically learn and adjust the weighted relationships between different parameters at different historical depths. The feature weight values of different parameters in the interacted historical feature matrix are adjusted based on their correlation with the predicted porosity. The higher the correlation, the higher the weight; conversely, the lower the weight. Through this mechanism, the model can more effectively utilize historical information and integrate it into the prediction task.

3) Temporal Embedding Layer: The goal of this component is to embed the positional relationships between different depth features in the historical feature matrix. The historical feature matrix transmitted by the historical feature attention interaction layer only embeds the weight relationships between logging parameters, without considering the positional feature relationships between depths. Through this component, the model can capture the temporal information between depth features and further extract the trend characteristics of logging parameters in the depth direction. This temporal embedding makes the model more robust and better able to extract the changing trends of logging parameters.

4) Historical Constraint Layer: This component embeds prior knowledge into the weighted information of historical features through an attention module. By preserving the importance of prior knowledge while adding constraints to the weights of historical features, the robustness of the model is further improved.

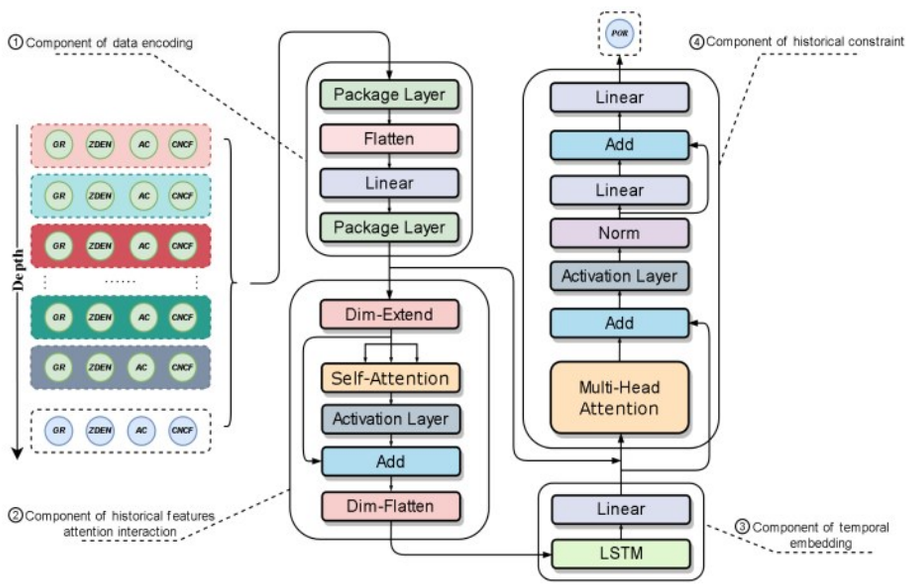

Figure 1 Architecture of our SMHDS Network: The input to the network is 2-D logging reservoir parameters, and the final layer of the network outputs the target parameter value (here predicting porosity). Component one consists of a linear layer with dimensions (16, 32, 16). Components two and four use Attention structures with input and output dimensions of 16. Component three features an LSTM with two layers, each having 128 memory units, with the activation function being the Sigmoid function.

1.1 Data Encoding Layer

The data encoding layer consists of two modules based on functionality. The first module, Flatten, is implemented by neural network linear layers. Its purpose is to perform linear transformation on the logging curve parameters, converting the single-valued logging curve parameter into a dense floating-point vector. This facilitates better exploration of logging curve parameter information during the attention interaction stage. The linear transformation formula is as follows:

$$Y_{n\times o} = X_{n\times 1} W_{1\times o\_dim} + b \#(1)$$

Where n is the number of input samples, 1 is the default number of features for logging curve parameters, o_dim is the length of the vector after linear transformation, and the vector b is a bias vector of length o_dim.

The second module aims to repackage the structure of the logging parameter dataset. After data cleaning, the dataset is reshaped into a three-dimensional vector, with the second and third dimensions storing the longitudinal and lateral

information of the logging curve parameters, respectively. Based on practical experience, it is known that logging parameters corresponding to the target parameter depth have the greatest influence weight. Therefore, this paper aims to maximize the retention of information at the depth where the target parameter is located. In the process of data transmission, the model treats the information at the depth of the target parameter as prior knowledge. By reshaping the data structure, the original dataset is separated into two sets: target depth features and historical depth features. These two sets are input as two branches into the model respectively, to avoid contaminating the weight of logging parameters corresponding to the depth of the target parameter with redundant information. Finally, they interact in component four to ultimately predict the target parameter.

1.2 Historical Feature Attention Interaction Layer

The primary role of the Historical Feature Attention Interaction Layer is to embed spatial domain information between different parameters at the same depth of historical logging parameters. Experimental research has revealed that, in addressing the logging parameter prediction problem, the feature extraction part of the model's shallow layers should reduce the complexity of feature information. Component two is centered around this conclusion and conducts relevant work.

Previous scholars constructed attention structures with the depth of logging parameters as "sentences" and logging parameters at the same depth as "word vectors." This approach may be influenced by the similarity between the natural structure of collected data and the corpus structure in natural language processing. However, experimental results indicate that, for logging parameter problems, treating logging parameters at the same depth as "sentences" and the logging parameters themselves as "word vectors" leads to better predictive performance for the model.

Component two is an improved self-attention structure. This structure transforms the three-dimensional feature vectors passed from component one into four-dimensional vectors through a dimension transformation module and linear transformation. Its purpose is to focus the self-attention module on logging parameters at the same depth. Although this process can also be achieved by transposing three-dimensional vectors, increasing the dimensionality and then linearly transforming them can enhance the information content of the logging parameters themselves. Through experimental analysis, the self-attention structure of component two eliminates positional encoding, enabling the model to achieve better performance.

In essence, the process of attention interaction is based on content addressing, i.e., finding the state that best matches the given state from certain state sets in the model transmission. This process can be expressed by the following formula:

$$Attention(Q,K,V)=\text{softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V\#(2)$$

The core idea of the attention mechanism is illustrated in Figure 2.

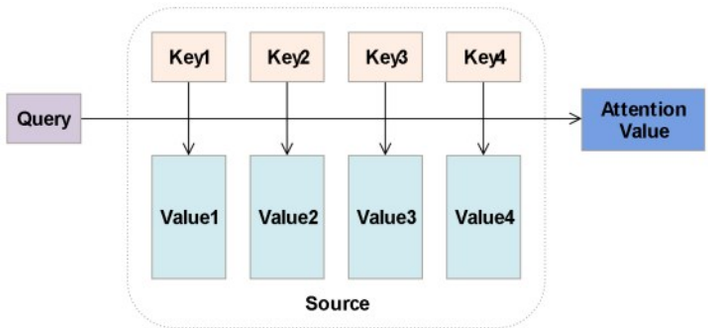

Figure 2 Structure of Attention Mechanism

The main component of the Historical Feature Attention Interaction Layer is a self-attention interaction module, with the relevant parameters represented as follows:

$$Q^{*}=\{Q_1,Q_2,\ Q_3,\cdots,\ Q_d\}\#(3)$$

$$Q_i=\{a_1,a_2,\ a_3,a_4,\cdots,\ a_{o\_dim}\}\#(4)$$

$$K^{*}=\{K_1,K_2,\ K_3,\cdots,\ K_d\}=V^{*}\#(5)$$

$$K_i=\{\theta_1,\theta_2,\ \theta_3,\cdots,\ \theta_{o\_dim}\}\#(6)$$

$$V_i=\{\beta_1,\beta_2,\ \beta_3,\cdots,\ \beta_{o\_dim}\}\#(7)$$

Formula (3) denotes Q^{*} as the query vector (Query), while K^{*} (Formula 5) represents the key vector (Key), and V^{*}

(Formula 5) represents the value vector (Value). d signifies the number of logging parameters at the same depth. This is different from how earlier scholars treated d as the length of the depth sequence of imported data. α,θ ,and β represent the feature components corresponding to each element of the Q^* , K^*,and V^* vectors, respectively. This module takes inputs of logging parameters corresponding to historical depths, passes through three different weight matrices, and obtains the final Q^*, K^*, and V^* vector matrices. However, to reduce the computational parameter count, the K^* and V^* matrices utilize the same vector matrix. The interaction process is depicted in Figure 3, where the left side represents the feature interaction method of earlier scholars, and the right side illustrates the improved approach in this paper:

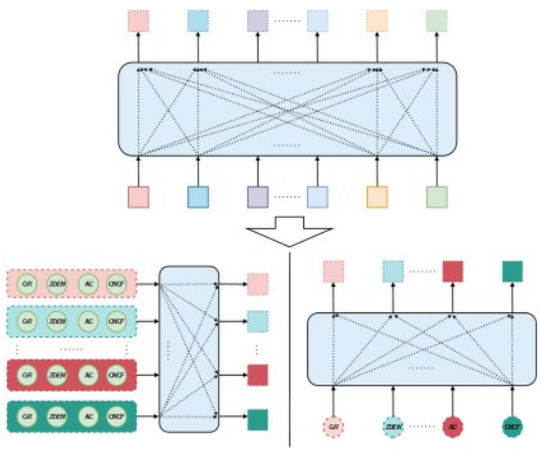

Figure 3

Figure 3 Self-Attention Interaction: On the left is the conventional self-attention structure, which takes multiple depth layers of reservoir parameters as input. On the right is the modified structure proposed in this paper, designed for a single depth layer of reservoir parameters, used in Component Two.

d_k refers to the length of each logging parameter feature after feature encoding, i. e., o_dim in Formula (1). QK^T represents the relationship matrix between logging parameters at each depth other than the target depth. Since the magnitudes of relationship weights differ, directly averaging them would lead to imbalanced magnitudes. Hence, the Softmax function is used to normalize these relationship weights. From the core idea perspective, the above computation can be viewed as a weighted averaging process.

In this and subsequent layers, for the purpose of accelerating model convergence and minimize information loss during transmission, and to focus more on the differences before and after the execution of network layers. Based on experimental results analysis, this paper introduces multiple skip connection modules into the model, as follows:

$$H(x) = x + F(x) \#(8)$$

1.3 Temporal Embedding Layer

The Temporal Embedding Layer consists of an LSTM module and a linear layer. LSTM, as shown in Figure 4, is a variant of the RNN network. LSTM has a chain-like structure similar to RNN, which can extract contextual information from the data and its own correlation, suitable for processing sequential data. However, compared to the RNN structure, LSTM addresses the issues of gradient explosion and long-term dependency by introducing gate structures in memory units.

The Historical Feature Attention Interaction Layer embeds spatial domain information for the data along the horizontal axis, but for the model, positional information in the depth direction is also an important research consideration.

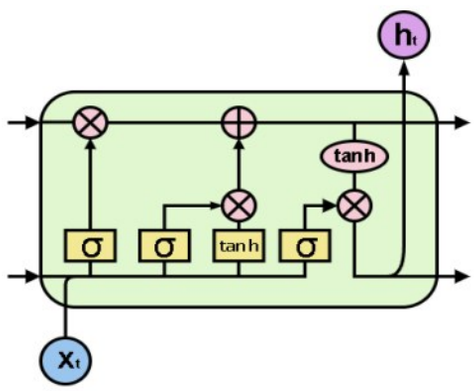

Figure 4 LSTM Memory Cell

For the logging parameters studied in this paper, although there is no temporal sequence in the physical sense for logging parameters at different depths, logging parameters between adjacent depths exhibit similar characteristics. Therefore, the data passed by Component 2 is regarded as a special temporal data sequence, and LSTM is utilized to embed depth-wise sequential information for logging parameters.

LSTM consists of three gate structures: the forget gate, the input gate, and the output gate. The forget gate controls the information to be discarded from the previous memory cell, the input gate controls the information to be stored in the current memory cell, and the output gate controls the information to be outputted from the current memory cell. Let the output result of the Historical Feature Attention Interaction Layer in Component 2 be x_i, and its set be X, where d represents the sequence length, i.e.,:

$$X=\{x_1, x_2, x_3, \cdots, x_{d-1}\}\#(9)$$

The calculation of the forget gate is represented as:

$$f_t=\sigma\left(W_f*\left[h_{t-1}, x_t\right]+b_f\right), x_t\in X\#(10)$$

The calculation of the input gate is represented as:

$$i_t=\sigma\left(W_i*\left[h_{t-1}, x_t\right]+b_i\right), x_t\in X\#(11)$$

$$\tilde{C}_t=\tanh\left(W_c*\left[h_{t-1}, x_t\right]+b_c\right), x_t\in X\#(12)$$

The calculation of the output gate is represented as:

$$o_t=\sigma\left(W_o*\left[h_{t-1}, x_t\right]+b_o\right), x_t\in X\#(13)$$

Where σ is the activation function, W and b are the weight and bias matrices, h_{t-1} is the previous time step's hidden state, x_t is the input value at the current time step, f_t is the output of the forget gate, i_t is the output of the input gate, o_t is the output of the output gate, and $\tilde{C}_t$ is the temporary memory state, i.e., the new information controlled by the input gate.

The output of each memory cell is h_t and C_t,and h_t is determined by the current memory state C_t. This relationship is represented by the formula:

$$C_t=f_t*C_{t-1}+i_t*\tilde{C}_t\#(14)$$

$$h_t=o_t*\tanh\left(C_t\right), t\in\{1, 2, 3, \cdots, d-1\}\#(15)$$

1.4 The Historical Constraint Layer

The Historical Constraint Layer consists of an Inter-Attention module and a feedforward network module. Through the computations of the preceding three components, the model obtains a feature vector containing historical trend information of the target parameter. This vector embeds the weighted relationships between historical well logging parameters at the same depth and the positional relationships of well logging parameters at different depths. In this component, this vector is used as the input for Q^* , while the target depth feature vector, transmitted solely from Component One, serves as the input for K^* and V^* . Through the Inter-Attention module, the relationship between this vector and the target depth feature vector is computed, and the weights of the features in this vector are reassigned to strengthen the feature components highly relevant to the target parameter.

The interaction mode of the Inter-Attention module is illustrated in Figure 5:

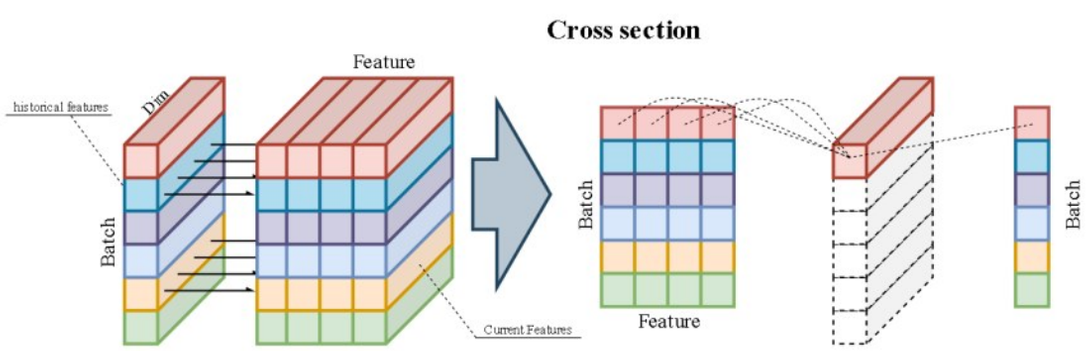

Figure 5

Figure 5 Inter-Attention Interaction: This structure takes both historical features and current features as input. The historical features, which are used as Q, come from the outputs of the first three components. The current features, used as K and V, are derived from the transmission of the first component.

The feedforward network module connected with the Inter-Attention module primarily serves to project the output matrix of the Inter-Attention module through a linear transformation into a higher-dimensional space, enhancing the

model's non-linear expression capability and convergence speed, the linear layer in 1.3 is the same as here.

2 Experiment and analysis

This article conducted four sets of experiments to evaluate the performance of the spatiotemporal model based on historical feature separation of prior knowledge (SMHDS). The first set of experiments is ablation experiments, aimed at analyzing the contribution of each module of the model to the final performance. The second set of experiments is comparative experiments, testing the effect of the improved attention module proposed in this article compared to the conventional attention module on the well log parameter prediction problem. The third set of experiments is also a comparative experiment, explored the model's capability in feature extraction and generalization for well logging parameters by considering the domain of both training and prediction data as variables. Additionally, three control groups were set up based on the model structures proposed by previous researchers, namely MLP, LSTM, and CNN-LSTM. Parameters of these three models were extensively experimented on real data to ensure their relative optimality for such prediction problems. Among them, it was found through experiments that using one-dimensional convolution to extract features in CNN-LSTM effectively prevents information loss between parameters at the same depth.

2.1 Experimental Preparation

The data for this study was collected from multiple wells in a specific exploration area and divided into four subsets labeled A, B, C, and D.

Dataset A comprises 644 discrete core data points from the exploration area, which were concatenated with corresponding well log interpretation data. This dataset was split into training and testing sets in an 8:2 ratio. Dataset B consists of data from a specific layer of well B, covering the depth range from 3342 m to 3494 m, with the interval 3395.2 m to 3445.8 m reserved for testing and the rest for training. Similarly, datasets C and D correspond to adjacent wells C and D, with layer depths ranging from 2797.9 m to 2882.9 m and 2793 m to 2842 m, respectively. Dataset C was allocated for training, while dataset D was designated for testing.

Prior to model training, datasets B, C, and D underwent stratified processing based on the trend of the Gamma Ray (GR) values to identify sandstone reservoir sections according to the following rules:

(1) Outliers were removed from the datasets.

(2) The mud content was calculated using the Gamma Ray (GR).

(3) Reservoir sections were defined based on a mud content of less than 40%, and these reservoir interval data were selected as the training data for the model.

After processing, Dataset B contains 1281 samples, Dataset C contains 951 samples, and Dataset D contains 379 samples.

In this study, the model takes GR (Gamma Ray, api), ZDEN (Bulk Density, g/cm3), AC (Acoustic Slowness, μs/ft), and CNCF (Compensated Neutron, v/v) as input parameters, with POR (Porosity, %) as the target for prediction. The correlation between these four curves and porosity (POR) is illustrated in Fig 6. It can be observed that AC and CNCF exhibit a significant positive correlation trend with porosity, while ZDEN shows a notable negative correlation trend. GR is meaningful for lithology classification of sandstone and mudstone, hence it is also used as input to provide constraints on data boundaries, reduce interference from dirty data that hasn't been filtered out.

Figure 6 Heatmap of Porosity Correlation with Input Parameters

$$r_{xy}=\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}\#(16)$$

Formula 16 represents the calculation formula for the heatmap coefficient, where r_{xy} denotes the Pearson correlation coefficient between variables x and y. x_i and y_i respectively represent the ith observation of variables x and y, while $\bar{x}$ and $\bar{y}$ denote the mean values of variables x and y respectively. n represents the sample size.

Due to the diverse sources of the collected raw data from various measurement tools, there existed differences in the dimensional scales of the measured physical quantities. Consequently, direct comparisons among different well logging parameters lacked physical significance. Therefore, the data were subjected to min-max normalization in this study. This normalization process, achieved through linear transformation of the original data, mapped the data values to the interval [0,1], expressed as follows:

$$x'=\frac{x-\min(x)}{\max(x)-\min(x)}\#(17)$$

For the results of the three sets of experiments, three evaluation metrics were employed in this study: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R-Square).

MAE calculates the average absolute error between the model's predicted values and the actual values, used to assess the proximity of the model's results to the true values. A smaller MAE indicates a more accurate predictive performance of the model.

RMSE, relative to MAE, is more sensitive to outliers in the data. By incorporating multiplication, it amplifies the deviation of outliers, serving to measure the volatility of the data. A smaller RMSE indicates better stability of the model. R-Square reflects the correlation between the model and the target. The closer the R-Square value is to 1, the better the model's ability to extract features. The calculation formulas are as follows:

$$MAE\left(y,\tilde{y}\right)=\frac{1}{n_{samples}}\sum_{i=1}^{n_{samples}}\left|y_i-\tilde{y}_i\right|\#(18)$$

$$RMSE\left(y,\tilde{y}\right)=\sqrt{\frac{1}{n_{samples}}\sum_{i=1}^{n_{samples}}(y_i-\tilde{y}_i)^2}\#(19)$$

$$R^2\left(y,\tilde{y}\right)=1-\frac{\sum_{i=1}^{n_{samples}}(y_i-\tilde{y}_i)^2}{\sum_{i=1}^{n_{samples}}(y_i-\bar{y})^2},\ \bar{y}=\frac{\sum_{i=1}^{n_{samples}}y_i}{n_{samples}}\#(20)$$

2.2 Experiment 1

Experiment 1 utilizes Dataset A as the sample set to evaluate the contribution of each module of the model to the overall performance.

Table 1 Ablation Experiment

Component	Choice						
Historical inputs	√		√	√	√	√	√
Self- Attention	√	√		√	√	√	√
Timing module	√	√	√		√	√	√
Inter Attention	√	√	√	√		√	√
Skip-connection	√	√	√	√	√		√
LayerNorm	√	√	√	√	√	√	
MAE/%	0.483	0.869	0.884	0.765	0.745	1.048	0.728
RMSE/%	0.662	1.382	1.237	1.094	1.093	1.409	1.148
R-Square/%	0.706	0.075	-2.117	-0.314	0.137	-4.035	-0.098

To ensure the trimmed model remains executable, the pruned parts are replaced by linear layers and undergo parameter fine-tuning.

Table 2 presents the results of the ablation experiment. From the table, it can be observed that the model's prediction of the target parameter POR on dataset A is not accurate enough, with many samples underfitting, as depicted in Figure 7. The emergence of the above phenomenon is mainly attributed to the limited and discrete samples in dataset A, which span multiple layers and exhibit poor correlation between samples, resulting in insufficient accuracy in learning sample features by the model.

After removing the historical feature separation structure, the model's MAE increases from 0.483 to 0.869, RMSE fluctuates from 0.662 to 1.382, and R-Square decreases from 0.706 to 0.075. This indicates that separating the current and historical features helps the model focus on key information. Furthermore, after pruning the skip connection module, the model's R-Square interpretability decreases to -4.035, further demonstrating the high information density of well logging parameters mentioned earlier.

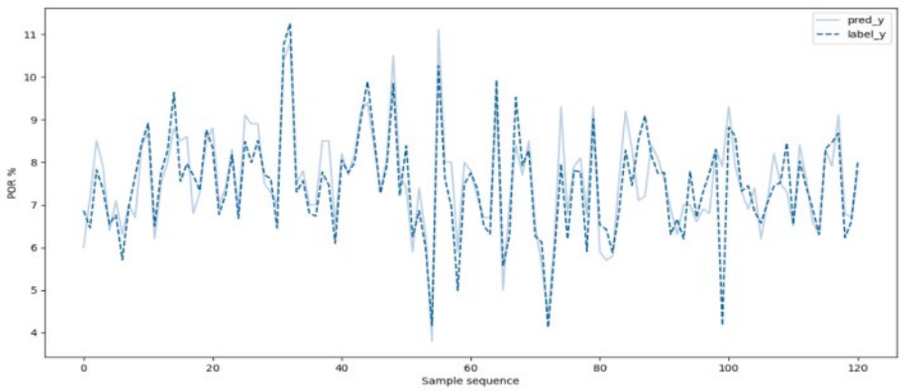

Figure 7 Prediction Results of Discrete Core Data

Figure 8 illustrates the histogram depicting the redistribution of weights in the historical constraint layer. The correlation values between well logging parameters and porosity (as shown in Fig 6) are incorporated for reference. It can be observed that the historical features exhibit the highest correlation with ZDEN and the lowest correlation with GR. This outcome aligns closely with the actual correlation analysis between porosity and the four input parameters presented in Fig 6. It indicates that the model built in this study has acquired a certain ability to focus on information through the learning process of the first three components.

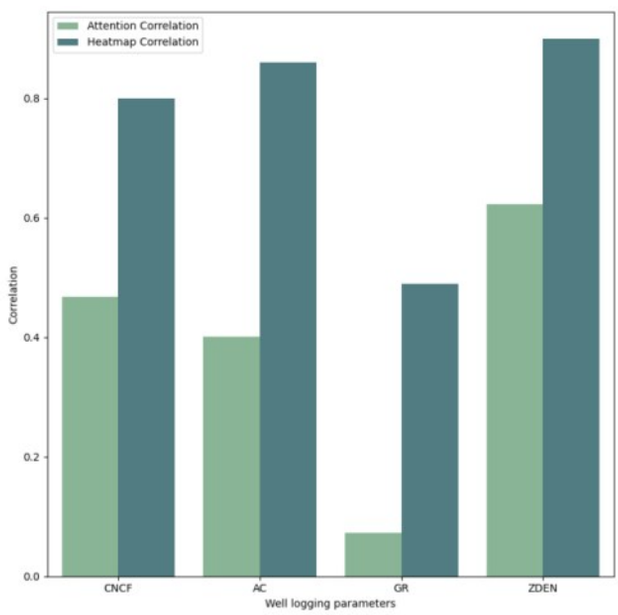

Figure 8 Relationship of Parameter Weights in the Historical Constraint Layer

2.3 Experiment 2

The paper constructs a lightweight neural network backbone consisting of three layers of MLP. The backbone accepts inputs of length Seqlen * o_dim and maps the weighted average features from the attention module to the output through three layers with 64, 32, and 1 neuron respectively. Seqlen represents the length of the depth sequence of input logging parameters, while o_dim represents the length of the feature vector of logging parameters in the attention module. Experiment 2 involves concatenating the backbone with the improved attention module proposed in this paper and the conventional attention module separately, followed by conducting comparative experiments on dataset B.

Referring to Table 2, the average absolute error of the improved attention module in this paper decreased by 0.4138

Table 2 Evaluation of Attention Modules

Method	Backbone	MAE/%	RMSE/%	R-Square/%
Proposed Attention	Dense (***Seqlen*o_dim***,64,32,1)	0.1354	0.3161	0.9962
Attention		0.5492	0.7688	0.9758

compared to the conventional model; the stability of the improved attention module increased by 59% compared to the conventional attention module, and the feature extraction capability of the improved attention module increased by 2% based on the conventional attention module. From Figure 9, it can be clearly seen that under other parameters consistent conditions, the conventional attention module exhibits more underfitting peaks, which demonstrates that for well logging parameters, when the attention module extracts shallow features, introducing depth-oriented sequential information will increase the complexity of the model in mining key features of well logging parameters. This also indirectly suggests that researchers focusing on using well logging parameter sequences as the target of attention module attention should build models with more complex network parameters and layers to separate the impact of sequence information on well logging parameters.

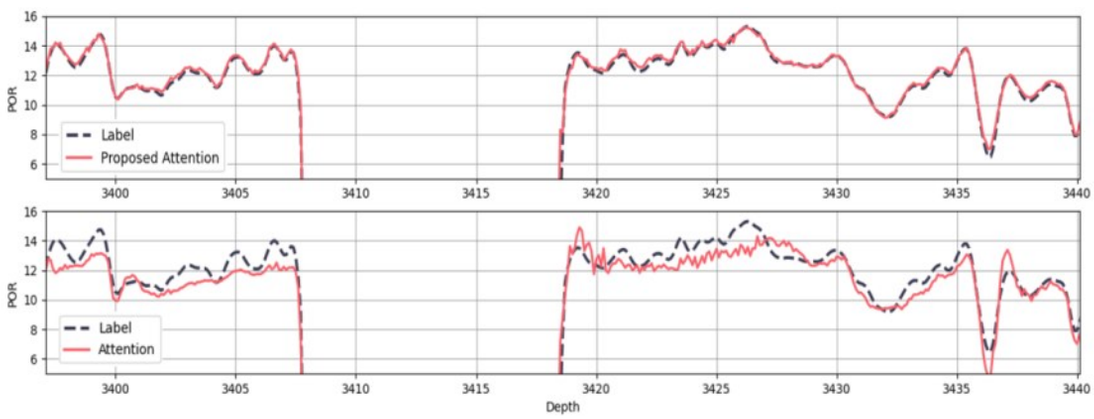

Figure 9

Figure 9 Attention Module Effect Comparison (Well B): The upper section shows the Proposed Attention structure, which is the modified structure presented in this paper. The lower section displays the Conventional Attention structure.

2.4 Experiment 3

Experiment 3 first evaluated the model's feature extraction capability by using the same-well data from dataset B to mitigate the impact of data distribution on feature extraction.The comparison between the predicted results of the model proposed in this paper and those of the control group methods with the actual values is shown in Figure 10. The evaluation comparison is presented in Table 3.

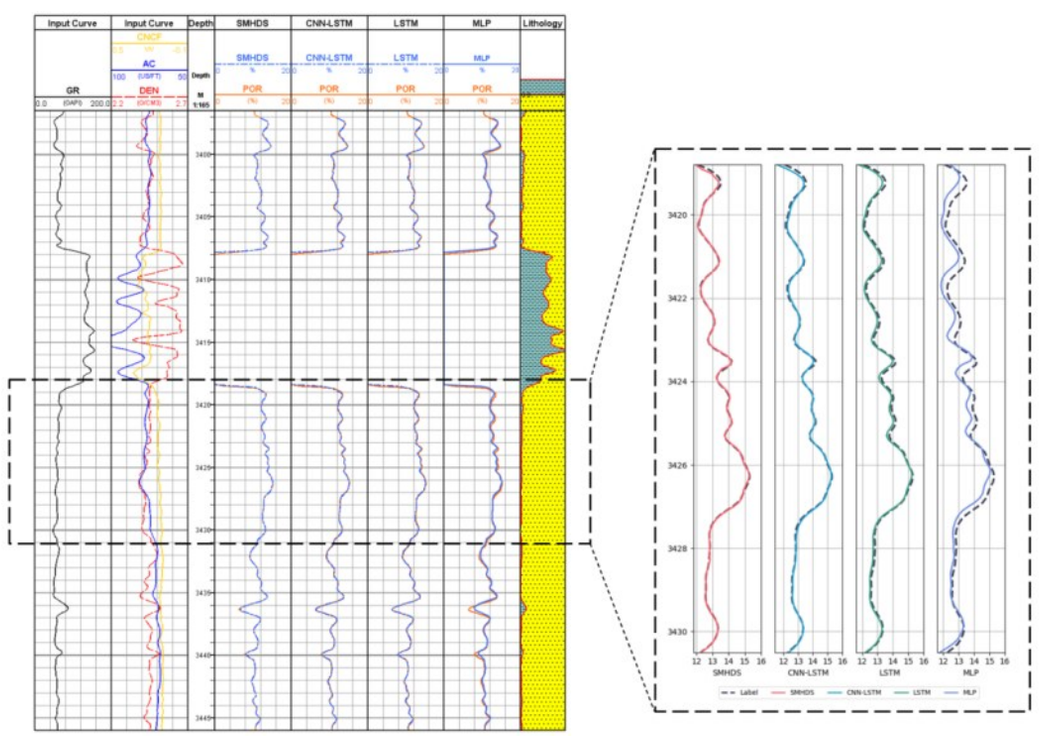

Figure 10 B Well Porosity Prediction Results Compared with Optimization Results (Same well comparison group)

Table 3 Model Performance Evaluation Table (Same well comparison group)

Method	MAE/%	RMSE/%	R-Square/%
MLP	0.3032	0.4171	0.9190
LSTM	0.0856	0.1002	0.9964
CNN-LSTM	0.0554	0.0835	0.9975
SMHDS	0.0404	0.0697	0.9982

According to the experimental results, it is observed that both the proposed model and the three control models exhibit good predictive capabilities in the same well prediction experiment. Taking the depth range of 3418-3431 m in dataset B as an example (see the magnified area in Fig 10), it can be observed that MLP has the lowest fitting accuracy. This is because MLP only extracts features for a single depth and does not consider the influence of depth direction information. In contrast, the improved CNN-LSTM introduces more spatial information, resulting in slightly higher prediction accuracy on the same well test set compared to the LSTM structure alone. The model proposed in this paper introduces temporal and spatial information through the attention module and temporal embedding module, achieving the highest fitting accuracy for the porosity in the test set, demonstrating excellent feature extraction capabilities in the study of well logging parameters.

Based on these findings, the paper further investigation was conducted into the model's generalization ability. By selecting datasets C and D from two different wells, which introduced variations in data distribution, the model's generalization ability across wells was evaluated. The results are illustrated in Figure 11, with the evaluation comparison presented in Table 4.

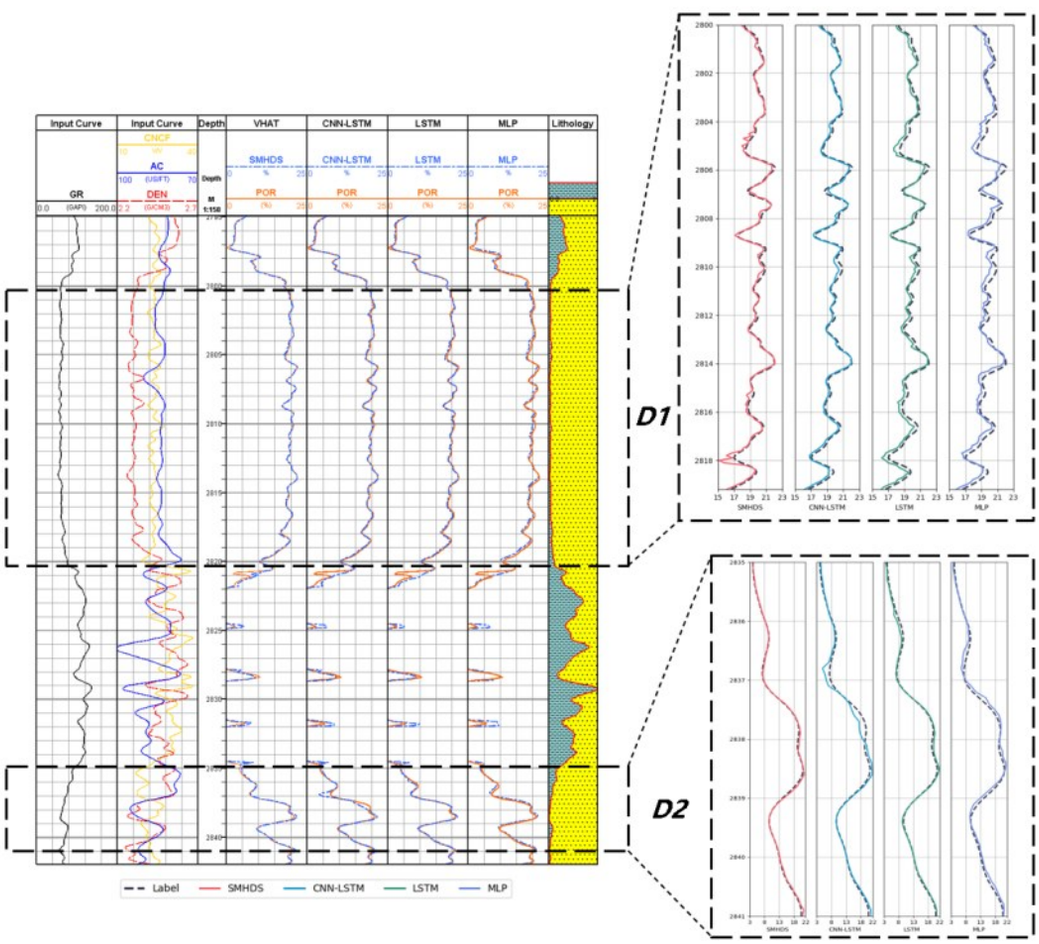

Figure 11 D Well Porosity Prediction Results Compared with Optimization Results (Different well comparison group)

According to Figure 11, it can be observed that in the cross-well prediction test, the four models exhibit similar fitting trends. The underfitting phenomenon of MLP is more pronounced, as exemplified in the enlarged area of Fig 11, where it can be seen that the predicted results for the 2800-2820 m interval of dataset D are noticeably lower than the actual values for all four methods. Through the analysis of Fig 12, it is found that the porosity values in section D1 are more concentrated, with the lower quartile being 19.05, higher than the upper quartile of section C, indicating that less than 25% of the samples in section C have porosity values similar to 75% of the samples in section D1. Moreover, from the overall situation, the porosity values in section D1 are generally higher than those in section C, leading to the models'

predicted results being underestimated, consistent with the actual prediction errors. In the 2835-2841m interval of dataset D, there is more overlap between the porosity values and those of section C, but according to the median of section D2, it can be inferred that there is a wide distribution of samples in section D2, causing 25% of the samples in section C to correspond to 50% of the samples in section D2, indicating the uneven distribution issue in the correspondence of samples between segments C and D.

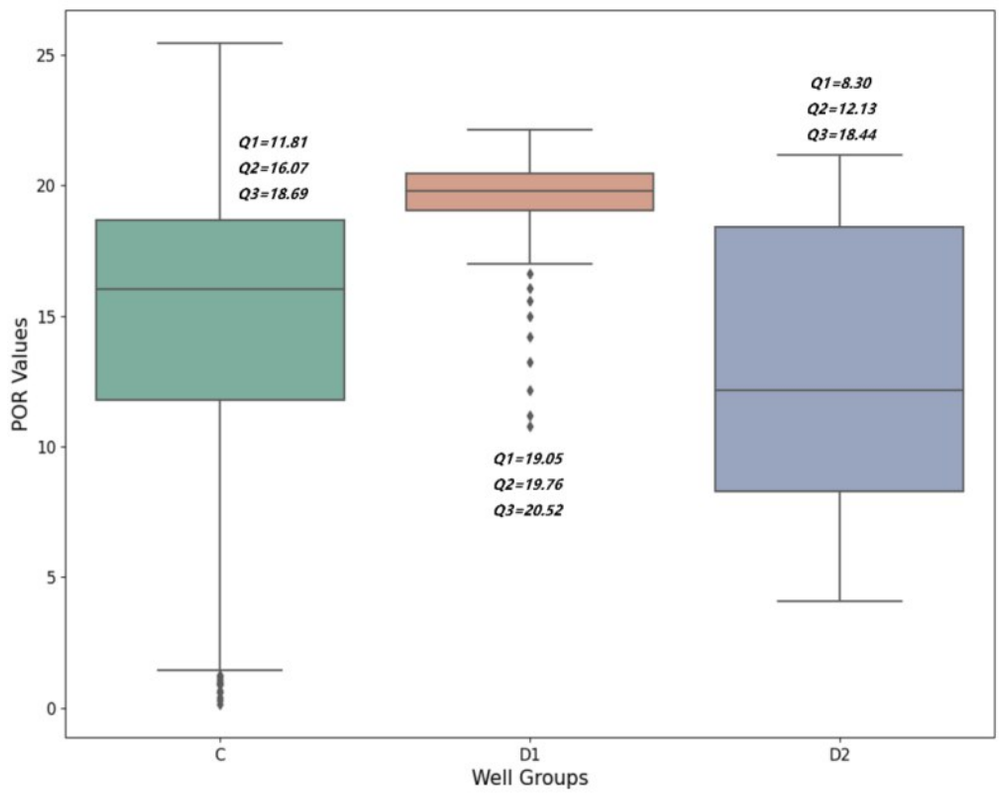

Figure 12 The box plot of porosity in the D1 and D2 segments of well C and D

From Figure 11, it can be observed that CNN-LSTM performs better than LSTM in predicting the D1 segment, but weaker in the D2 segment. This phenomenon contradicts the evaluation results of the feature extraction capability in the same well group. Based on the analysis of the sand-mud ratio in the D2 segment, it is inferred that the increase in mud content at partial depths may cause significant differences in the patterns of well logging parameters extracted by the CNN feature extraction layer at different time steps, thereby reducing the model's resistance to data distribution fluctuations. Comparing the R-Square values of MLP in the same well and different well groups, it is found that MLP is less affected by dimensional fluctuations. This result may be attributed to the fact that MLP only accepts data from a single depth, without considering changes in the overall trend, thus reducing the overall impact of dimensional fluctuations. However, The model proposed in this paper still demonstrates the best predictive performance in cross-domain testing, indicating that it possesses strong feature extraction capability and good generalization ability.

Table 4 Model Performance Evaluation Table (Different well comparison group)

Method	MAE/%	RMSE/%	R-Square/%
MLP	0.5853	0.8824	0.9500
LSTM	0.4036	0.6239	0.9903
CNN-LSTM	0.4647	0.7563	0.9858
SMHDS	0.2729	0.4680	0.9945

3 Conclusion

The paper proposes a spatiotemporal model based on historical feature separation of prior knowledge (SMHDS) for focusing on key information in well logging parameter prediction models. Through three sets of experiments on four sets of actual data from a certain area, the following conclusions are drawn:

Separating historical features from current features is more conducive to focusing on key information in well logging parameter prediction problems.

Well logging parameters at the same depth as the focus of attention provide better assistance to the model's feature

extraction capability than well logging curves with depth as the axis.

Compared with similar models, the information transmission process in the proposed model structure in this paper is more reasonable for well logging parameter prediction problems, thus exhibiting better feature extraction capability and generalization ability.

The core issue in well logging parameter prediction is how to preserve the original information of well logging parameters while identifying their correlation with target parameters. Therefore, further research in this area should focus on retaining the original information of well logging parameters as much as possible during the model's information transmission process, while also exploring the relationships between well logging parameters to the greatest extent. Additionally, since models for such problems are generally supervised and heavily rely on the quality and rationality of input data, new solutions can also be explored from the perspectives of optimizing input data and establishing connections between source domain data and target domain data to provide new insights for addressing the aforementioned issues.

References

[1] Avseth P , Skjei N , Skålnes Å . Rock physics modelling of 4D time - shifts and time - shift derivatives using well log data – a North Sea demonstration[J].Geophysical Prospecting,2013,61(2):380-390.

[2] Saleh S ,Jahr T ,Jentzsch G , et al.Crustal evaluation of the northern Red Sea rift and Gulf of Suez, Egypt from geophysical data: 3-dimensional modeling[J].Journal of African Earth Sciences,2006,45(3):257-278.

[3] Louis G. Chombart; Well logs in carbonate reservoirs. Geophysics 1960,25 (4): 779–853.

[4] Aslan Gassiyev, Feifei Huang, Evgeni M. Chesnokov; Using the pair-correlation function as a tool to identify the location for shale gas/oil reservoir based on well-log data. Geophysics 2016,81 (2): D91–D109.

[5] Lei Lin, Hong Huang, Pengyun Zhang, Weichao Yan, Hao Wei, Hang Liu, Zhi Zhong; A deep-learning framework for borehole formation properties prediction using heterogeneous well-logging data: A case study of a carbonate reservoir in the Gaoshiti-Moxi area, Sichuan Basin, China. Geophysics 2023,89 (1): WA295–WA308.

[6] ZHANG Zhaohui,GAO Chuqiao,LIU Juanjuan. Calculation method of porosity based on formation component analysis[J].Lithologic Reservoirs, 2012, 24(1): 97-99.

[7] JIN Yunzhi, GAO Chuqiao, GAO Yongde, et al. Quantitative Calculation of Petrophysical Components of Igneous Rocks Using Formation Elemental Logging Data[J]. Well Logging Technology, 2018, 42(5): 521-524.

[8] ZHENG Qibin, GAO Chuqiao, ZHAO Bin. Overview of Digital Core Modeling Methods and Applications[J]. Energy and Environmental Protection, 2017, 39(12): 145-148.

[9] Mohaghegh S,Arefi R,Ameri S,et al. 1996. Petroleum reservoir characterization with the aid of artificial neural networks. Journal of Petroleum Science and Engineering,16(4):263-274.

[10] Jian H ,Chenghui L ,Zhimin C , et al.Integration of deep neural networks and ensemble learning machines for missing well logs estimation[J]. Flow Measurement and Instrumentation,2020,73(prepublish).

[11] Kpüklü, Okan, Wei X , Rigoll G .You Only Watch Once: A Unified CNN Architecture for Real-Time Spatiotemporal Action Localization[J]. 2019.

[12] Li K , Zhiwei L , Lingyu M , et al. YOLO-FA: Type-1 fuzzy attention based YOLO detector for vehicle detection[J]. Expert Systems With Applications,2024,237(PB).

[13] Liu P ,Quan F ,Gao Y , et al.Green energy forecasting using multiheaded convolutional LSTM model for sustainable life[J].Sustainable Energy Technologies and Assessments,2024,63103609.

[14] Li Ning, Xu Binsen, Wu Hongliang, Feng Zhou, Li Yusheng, Wang Kewen, Liu Peng. Application status and prospects of artificial intelligence in well logging and formation evaluation[J]. Acta Petrolei Sinica, 2021, 42(4): 508-522.

[15] Fung C C , Wong K W , Eren H ,et al.Modular artificial neural network for prediction of petrophysical properties from well log data[J].IEEE, 1996.

[16] YANG Zhili, ZHOU Lu, PENG Wenli, et al. Application of BP Neural Network Technology in Reconstruction of Sonic Logging Curves[J]. Journal of Southwest Petroleum University (Natural Science Edition), 2008(01): 63-66,13.

[17] Mo X , Zhang Q , Li X .Well logging curve reconstruction based on Genetic Neural Networks[C]//International Conference on Fuzzy Systems & Knowledge Discovery.IEEE, 2016.

[18] YANG Liuqing, ZHA Bei, CHEN Wei. Prediction Method of Porosity in Sandstone Reservoirs Based on Deep Neural Network[J]. China Science and Technology Paper, 2020, 15(1): 73-80.

[19] Asoodeh M ,Bagheripour P .Prediction of Compressional, Shear, and Stoneley Wave Velocities from Conventional Well Log Data Using a Committee Machine with Intelligent Systems[J].Rock Mechanics and Rock Engineering,2012,45(1):45-63.

[20] Wakeel H ,Miao L ,Muhammad A , et al.Machine learning - a novel approach to predict the porosity curve using geophysical logs data: An example from the Lower Goru sand reservoir in the Southern Indus Basin, Pakistan[J].Journal of Applied Geophysics,2023,214.

[21] Faust L. Y. (1953). A VELOCITY FUNCTION INCLUDING LITHOLOGIC VARIATION. GEOPHYSICS, 18(2), 271–288.

[22] Eskandari H , Rezaee M R , Mohammadnia M .Application of Multiple Regression and Artificial Neural Network Techniques to Predict Shear Wave Velocity from Wireline Log Data for a Carbonate Reservoir, Southest Iran[J]. 2004.

[23] Castagna J P .Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks[J].Geophysics, 1985, 50.

[24] DUAN Youxiang, LI Gentian, SUN Qifeng. Application Research of Convolutional Neural Network in Reservoir Prediction[J]. Journal on Communications, 2016, 37(S1): 1-9.

[25] Zhong Z , Carr T R , Wu X ,et al. Application of a convolutional neural network in permeability prediction: A case study in the Jacksonburg-Stringtown oil field, West Virginia, USA[J].GeoScienceWorld, 2019(6).

[26] Dongxiao Z ,Yuntian C ,Jin M .Synthetic well logs generation via Recurrent Neural Networks[J].Petroleum Exploration and Development,2018, 45(4):629-639.

[27] ChenW ,YangL ,ZhaB , et al.Deep learning reservoir porosity prediction based on multilayer long short-term memory network[J].Geophysics, 2020,85(4).

[28] Pham N , Wu X , Naeini E Z .Missing well log prediction using convolutional long short-term memory network[J].Geophysics, 2020, 85(4):1-55.

[29] Lei Lin, Hao Wei, Tiantian Wu, Pengyun Zhang, Zhi Zhong, Chenglong Li; Missing well-log reconstruction using a sequence self-attention deep-learning framework. Geophysics 2023,88 (6): D391–D410.

[30] Liqun S , Yanchang L , Min T , et al. CNN-BiLSTM hybrid neural networks with attention mechanism for well log prediction[J]. Journal of Petroleum Science and Engineering,2021,205.

[31] Liuqing Yang, Sergey Fomel, Shoudong Wang, Xiaohong Chen, Wei Chen, Omar M. Saad, Yangkang Chen; Porosity and permeability prediction using a transformer and periodic long short-term network. Geophysics 2023,88 (1).