# Lithosphere

*Research Article*

# Method for Predicting Transverse Wave Velocity Using a Gated Recurrent Unit Based on Spatiotemporal Attention Mechanism

**Jiaxin Huang**[1] **, Gang Gao**[1] **, Xiaoming Li,**[2] **Yonggen Li**[2] **and Zhixian Gui**[1]

[1]*Key Laboratory of Exploration Technologies for Oil and Gas Resources (Yangtze University), Ministry of Education, Wuhan 430100, China*

[2]*Research Institute of Petroleum Exploration and Development, CNPC, Beijing 100083, China*

Correspondence should be addressed to Gang Gao; 500871@yangtzeu.edu.cn

Transverse wave velocity plays an important role in seismic exploration and reservoir assessment in the oil and gas industry. Due to the lack of transverse wave velocity data from actual production activities, it is necessary to predict transverse wave velocity based on longitudinal wave velocity and other reservoir parameters. This paper proposes a fusion network based on spatiotemporal attention mechanism and gated recurrent unit (STAGRU) due to the significant correlation between the transverse wave velocity and reservoir parameters in the spatiotemporal domain. In the case of tight sandstone reservoirs in the Junggar Basin, the intersection plot technique is used to select four well logging parameters that are sensitive to transverse wave velocity: longitudinal wave velocity, density, natural gamma, and neutron porosity. The autocorrelation technique is employed to analyze the depth-related correlation of well logging curves. The relationship between the spatiotemporal characteristics of these well logging data and the network attention weights is also examined to validate the rationale behind incorporating the spatiotemporal attention mechanism. Finally, the actual measurement data from multiple wells are utilized to analyze the performance of the training set and test set separately. The results indicate that the predictive accuracy and generalization ability of the proposed STAGRU method are superior to the single-parameter fitting method, multiparameter fitting method, Xu-White model method, GRU network, and 2DCNN-GRU hybrid network. This demonstrates the feasibility of the transverse wave velocity prediction method based on the spatiotemporal attention mechanism in the study of rock physics modeling for tight sandstone reservoirs.

## 1. Introduction

Transverse wave velocity is an important parameter for evaluating the physical properties and structures of underground media in seismic exploration. It plays an indispensable role as fundamental information in prestack seismic inversion, fluid identification, and AVO analysis [1–4]. However, due to the high cost of exploration or limited acquisition techniques, actual seismic data often lack transverse wave velocity information, especially in many areas and older wells [5]. Therefore, it is extremely important to achieve high-precision and low-cost prediction of transverse wave velocity using other well logging data.

Researchers both domestically and internationally have conducted in-depth discussions on transverse wave velocity prediction, mainly using two methods: the empirical formula method [6–8] and the rock physics model method [9–13]. However, empirical formulas vary depending on the region and lithology, resulting in limited accuracy and insufficient generalization. The rock physics model method is complex, involves multiple parameters, and has low computational efficiency, and in complex reservoirs, some parameters are difficult to obtain accurately, which limits the application of rock physics models [14].

Deep learning has had a significant impact in various fields such as speech recognition, natural language processing, and facial recognition, by establishing complex

nonlinear relationships between input and output data [15–17]. With the rapid development of deep learning in recent years, more and more experts have applied convolutional neural networks (CNNs) and gated recurrent unit (GRU) to the field of geophysics, achieving promising research results in fault recognition, lithology classification, reservoir parameter inversion, and other geological problems [18–23]. In the modeling of transverse wave velocity, researchers mainly use recurrent neural networks (RNN) and CNN to establish the nonlinear mapping between input and output. Considering the powerful spatial feature extraction capabilities of CNN, using CNN to predict transverse wave velocity [24] improves the prediction accuracy. Well logging data exhibit regularity in the depth direction, and compared with CNN, RNN is more suitable for handling conventional well logging data. Some researchers [25, 26] have proposed methods using Long Short-Term Memory (LSTM) networks to address the lack of transverse wave velocity data. This method fully considers the temporal characteristics of conventional well logging data and has achieved good prediction results in carbonate and sandstone reservoirs. Compared with LSTM networks, GRU can reduce the number of network parameters and has been widely used for predicting transverse wave velocity and porosity [20, 27, 28]. However, the aforementioned methods only focus on the spatial or temporal characteristics of conventional well logging data, neglecting the impact of spatiotemporal features on transverse wave velocity. To comprehensively consider the influence of spatiotemporal features of conventional well logging data on transverse wave velocity, some researchers have proposed fusion networks composed of CNN and LSTM or GRU [29–31]. Although these methods improve the performance of the network, they do not highlight the importance of spatiotemporal features on transverse wave velocity. Therefore, there is a high requirement for the weight distribution of the extracted spatiotemporal features from the network.

In recent years, attention-based neural networks have been applied in various fields such as machine translation [32], text classification [33], power load forecasting [34], and Earth sciences. These researchers believe that attention mechanisms can improve the network's sensitivity to important features. Kavianpour et al. [35] developed an attention-based CNN-BILSTM fusion network for earthquake prediction and achieved good prediction results. Mousavi [36] (2020) developed an attention-based deep learning model for simultaneous earthquake detection and phase picking, which can detect multiple earthquakes and accurately pick phases like human analysts. To improve the accuracy of earthquake prediction, Banna et al. [37] (2021) incorporated attention mechanisms into a bidirectional LSTM structure. Bai [38] proposed an attention-based LSTM-FCN network model that improves the accuracy of earthquake event detection and localization. Shan et al. [39] provided a fusion network based on CNN-BILSTM for predicting well logging data to reduce drilling costs.

The above literature indicates that deep learning networks have been widely applied in earthquake phase recognition, lithology identification, and reservoir parameter inversion. However, attention-based neural networks are rarely used for transverse wave velocity prediction. Additionally, due to the correlation between transverse wave velocity and the spatiotemporal characteristics of conventional well logging data, a spatiotemporal attention-based GRU fusion network (STAGRU) is proposed. This network mainly includes GRU layers, spatial attention layer, temporal attention layer, and fully connected layers. In this study, the tight sandstone reservoir in the Junggar Basin is taken as the research object. Based on the STAGRU fusion network, the training and prediction process of transverse wave velocity is established, and the weight distribution of the attention layers is analyzed. Finally, computations were conducted for single-parameter fitting, multiparameter fitting, and the Xu-White model method, followed by training for the STAGRU, GRU network, and 2DCNN-GRU hybrid network models. Subsequently, optimization was performed on the parameters of the STAGRU network to obtain predictive results. The effectiveness of using the STAGRU network for transverse wave velocity prediction is verified.

## 2. Methodology and Techniques

*2.1. Gated Recurrent Unit.* The GRU is a variant of the RNN similar to the LSTM, proposed by Kyunghyun Cho (2014) [40]. On the one hand, it overcomes the issues of vanishing or exploding gradients in traditional RNNs and resolves the problem of long computation time in LSTMs [41]. On the other hand, GRU has fewer training parameters, faster convergence speed, and the ability to handle nonlinear and time series problems [42]. The GRU hidden layer has two gates: the reset gate ($r_t$) and the update gate ($z_t$; Figure 1), making the model simple and less prone to overfitting. The reset gate and update gate perform retention and forgetting functions based on the current input at time step $t$. When the reset gate ($r_t$) is close to 1, it means that more information is retained. When the update gate ($z_t$) is close to 1, it means that more information is forgotten. Given the input data $x_t$ at time step $t$, the reset gate ($r_t$) and update gate ($z_t$) can be represented as:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \tag{1}$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \tag{2}$$

In the equations, $W_r$ and $W_z$ represent the weight matrices for the reset gate and update gate, respectively. $b_r$ and $b_z$ are the biases. $h_{t-1}$ represents the output of the hidden state at time step $t-1$. The "$\sigma$" symbol denotes the logistic sigmoid function, which maps the output to the range of [0, 1]. The "[]" indicates the concatenation of two matrices.

The new state includes the information controlled by the reset gate and is combined with the update gate to obtain the final output $h_t$:
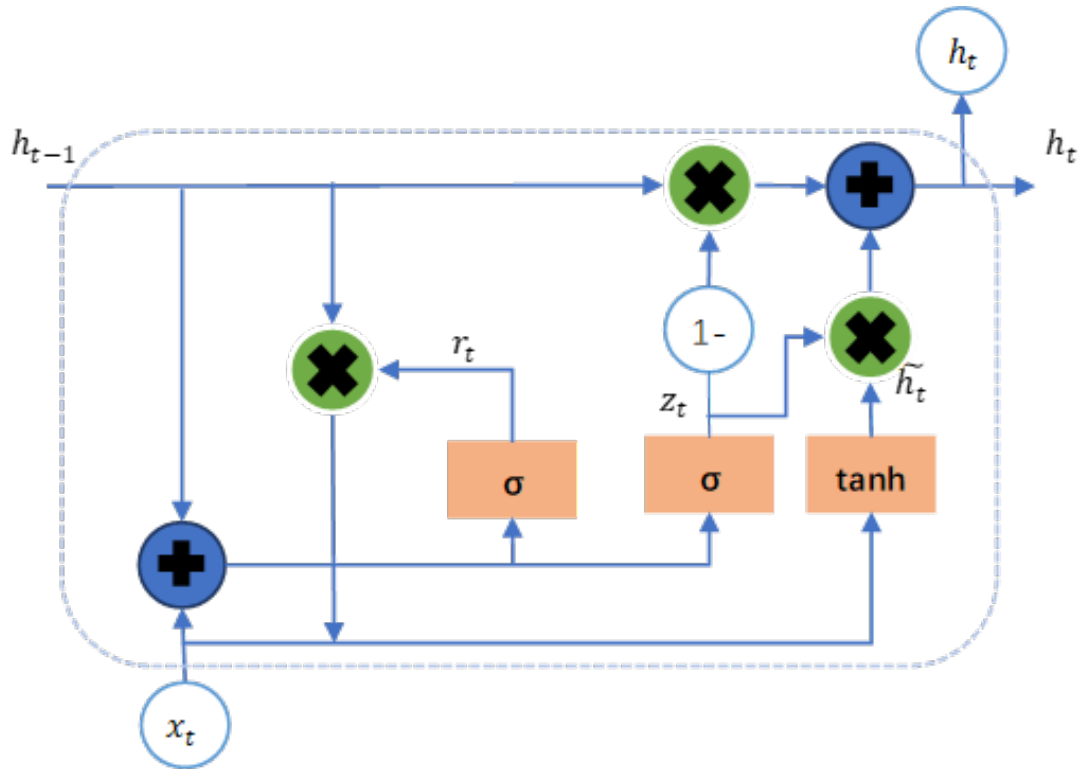
FIGURE 1: Structure of the GRU network.

$$\widetilde{h_t} = \tanh\left(W_h[r_t * h_{t-1}, x_t] + b_h\right) \tag{3}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \bar{h}_t \tag{4}$$

In the equation, $W_h$ and $b_h$ represent the weight matrix and bias of the new state "$\widetilde{h_t}$". "tanh" denotes the hyperbolic tangent activation function. "$*$" represents matrix multiplication. $h_t$ represents the output of the current hidden state.

*2.2. Temporal and Spatial Attention Mechanism.* The well logging data exhibits certain regularities in sedimentary formations. By incorporating the temporal attention mechanism into the GRU network, the sensitivity of important temporal features to the shear wave velocity is enhanced. The time features are inputted into the temporal attention layer, and different weights are assigned to these features to obtain the output of the temporal attention layer. The hidden state $H_n = [H_{1,n}, H_{2,n} \cdots H_{t,n}]$ represents the $t$-dimensional vector of the $n$th spatial feature. The temporal attention weights can be represented as follows:

$$\beta_n^` = softmax\left(W_\beta H_n + b_\beta\right) \tag{5}$$

$$X_n^` = \beta_n^` \odot H_n = [\beta_{1,n} H_{1,n} \beta_{2,n} H_{2,n} \cdots \beta_{t,n} H_{t,n}] \tag{6}$$

where $\beta_n^` = [\beta_{1,n}, \beta_{2,n} \cdots \beta_{t,n}]$ represents the weights of the temporal attention layer. $W_\beta$ and $b_\beta$ denote the weight matrix and bias, respectively. Softmax is the normalization

function, and "$\odot$" denotes the hadamard product. $X_n^`$ represents the weighted result.

Due to the correlation between the shear wave velocity and the spatial features of conventional well logs, the spatial attention mechanism is incorporated into the GRU network to enhance the sensitivity of important spatial features to the shear wave velocity. The spatial features of the conventional well logs are fed into the spatial attention layer, and different weights are assigned to these features to obtain the output of the spatial attention layer. The hidden state $h_t = [h_{t,1}, h_{t,2} \cdots h_{t,m}]$ represents the $m$-dimensional feature vector at time step $t$. The spatial attention weights can be represented as follows:

$$\alpha_t^` = Softmax(W_\alpha h_t + b_\alpha) \tag{7}$$

$$X_t^` = \alpha_t^` \odot h_t = [\alpha_{t,1} h_{t,1}, \alpha_{t,2} h_{t,2} \cdots \alpha_{t,m} h_{t,m}] \tag{8}$$

where $\alpha_t^` = [\alpha_{t,1}, \alpha_{t,2} \cdots \alpha_{t,m}]$ represents the weights of the spatial attention layer, $W_\alpha$ and $b_\alpha$ are the weight matrix and bias, respectively. $X_t^`$ represents the weighted result.

*2.3. The Structure of the STAGRU Fusion Network.* A STAGRU is proposed to address the correlation between shear wave velocity and the spatiotemporal features of conventional well logging data (as shown in Figure 2). The STAGRU fusion network consists of an input layer, two GRU layers, a temporal attention layer, a spatial attention layer, and a fully connected layer. The GRU layers are responsible for extracting the spatiotemporal features from

the conventional well logging data, while the temporal and spatial attention layers enhance the network's sensitivity to important spatiotemporal features. The fully connected layer improves the nonlinearity of the proposed network.

*2.4. Training and Prediction Process of the STAGRU Fusion Network.* The training and prediction process of the proposed STAGRU fusion network can be divided into the following steps:

(1) Data preprocessing: Due to the significant differences among well logging data, the StandardScaler function is applied to standardize the data using the standard deviation. The entire dataset is scaled to have zero mean and unit variance, ensuring that the processed dataset follows a normal distribution, as shown in Equation (9).

$$X_i^` = \frac{X_i - X_m}{X_\sigma} \tag{9}$$

where $X_i$ represents the well logging data, $X_m$, $X_\sigma$ denote the mean and standard deviation of the well logging data, and $X_i^`$ represents the standardized value.

(2) Building the STAGRU Fusion Network: The training set of well logging data is used as the input to the STAGRU Fusion Network, and the target output is the predicted shear wave velocity. After setting the hyperparameters of the network, the network undergoes optimization through iterative training.

(3) Training of the STAGRU Fusion Network: The mean squared error (MSE) is used as the loss function for the network. The STAGRU network undergoes multiple iterations of training to find the optimal parameters, and the network with the lowest loss error is selected.

(4) Prediction of Shear Wave Velocity: The test set of well logging data is inputted into the trained STAGRU Fusion Network to predict the shear wave velocity.

(5) Network Evaluation: The mean absolute error (MAE) and coefficient of determination ($R^2$) are used as evaluation metrics to assess the prediction performance of the network. The calculation methods are as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |(\widetilde{y_i} - y_i)| \tag{10}$$

$$R^2 = \frac{\sum_{i=1}^{n} (\widetilde{y_i} - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{11}$$

where $y_i$ represents the actual value, $\bar{y}$ represents the average of the actual values, $\widetilde{y_i}$ represents the predicted value, and $n$ is the number of samples.

# 3. Experiment and Analysis

*3.1. Data Preparation.* The experiments for this study were conducted on an Intel(R) Core(TM) i5-8250U CPU @ 1.60 GHz 1.80 GHz, with an NVIDIA GeForce 940 MX environment. Python 3.9 was used as the compilation environment, and TensorFlow, version 2.11.0, served as the deep learning platform. The data in this paper were sourced from the Qiugu Formation reservoir in a specific area, specifically selecting five wells. The dataset includes six well logging parameters: longitudinal wave velocity (VP), density (DEN), natural gamma (GR), neutron porosity (DEN), resistivity (RT), and spontaneous potential (SP). The reservoir is primarily composed of sandstone and shale. It is characterized by deep burial, low porosity, low permeability, and a complex pore structure, making it a typical unconventional tight oil and gas reservoir. This paper selected well log parameters measured at a depth of 5420–5480 m from four wells as the training dataset to train the model. Well log parameters measured at a depth of 5520–5590 m from well Y, which was not included in the training, were chosen as the test dataset to evaluate the model's performance. In order to improve the accuracy of the rock physics model and prestack seismic inversion of well and seismic data sets, a GRU fusion network based on the spatiotemporal attention mechanism was used for predicting shear wave velocity.

*3.2. Feature Selection for Data.* Choosing appropriate well logging parameters can improve the predictive performance of the model. Well logging parameters can reflect the reservoir's storage capacity, lithology, permeability, and other characteristics. There exists a certain correlation between different well logging parameters detected in the same formation. Each well logging parameter responds to the geological features from different perspectives and mechanisms. Well logging parameters are measurements of different physical properties of the same rock. Different physical properties can reflect the same petrophysical parameter of the rock (such as porosity, which can be interpreted simultaneously using acoustic, density, and neutron measurements). Therefore, there is a certain correlation between shear wave velocity and other well logging parameters. This is the spatial characteristic among well logging curves. In theory, the predictive accuracy of using deep learning to solve regression problems depends on the correlation between the input and output. Studies have shown that there are nonlinear features between logging parameters and reservoir petrophysical parameters. Figure 3 shows the scatter plot of shear wave velocity (VS) with conventional logging data. The correlations from high to low are as follows: compressional wave velocity (VP), neutron porosity (CNL), gamma ray (GR), density (DEN), resistivity logarithm (RT), and spontaneous potential (SP). Their respective coefficient of determination ($R^2$) values are 0.930, 0.851, 0.647, 0.337, 0.289, and 0.024. Among these well logging parameters, the logarithm of resistivity and spontaneous potential has a relatively low correlation with transverse wave velocity. On the other hand, the remaining well logging parameters, namely, longitudinal wave velocity (VP), neutron porosity (CNL), natural gamma (GR), and density (DEN), exhibit correlations greater than 0.3 with transverse wave velocity. Therefore, in this study, VP, CNL, GR, and DEN are selected as the well logging parameters to predict transverse wave velocity. The correlation analysis (Figure 4) of the selected well logging parameters
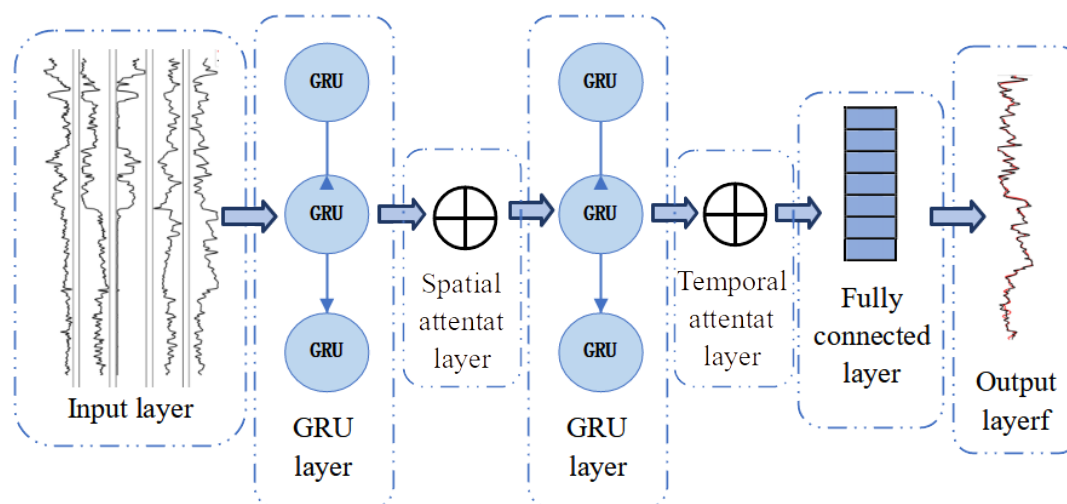
FIGURE 2: Architecture of the STAGRU Fusion Network.

indicates that there are both certain connections and significant differences among them. This suggests that well logging parameters contain diverse and rich information, which serves as the basis for predicting shear waves using well logging parameters. As the formation deposition is gradual, the adjacent data points in well logging curves exhibit correlation, indicating temporal characteristics. The autocorrelation function is used to analyze the autocorrelation of conventional well log curves. Figure 5 represents the autocorrelation coefficients of well log curves, where the x-axis represents the lag, which represents the displacement of the well log curve, and the autocorrelation decreases with increasing lag. The y-axis represents a series of correlation coefficients corresponding to different lag. From the figure, it can be seen that when the lag is 20, the autocorrelation coefficients in descending order are transverse wave velocity (VS), longitudinal wave velocity (VP), neutron porosity (CNL), natural gamma (GR), and density (DEN). Among them, VS, VP, and CNL exhibit autocorrelation coefficients greater than 0.6. To better illustrate the autocorrelation characteristics of well logging curves, on the basis of Figure 5, the relationship between autocorrelation and lag distance of conventional well logging parameters is plotted, with a maximum lag distance of 40 (Figure 6). From this figure, it can be observed that the autocorrelation coefficient reaches above 0.2 at a lag distance of 15. The above analysis indicates that the temporal and spatial features of the conventional logging data have a certain correlation with shear wave velocity.

*3.3. The Interpretability of Attention Weights.* To validate the effectiveness of the attention mechanism in enhancing the network's sensitivity to important spatiotemporal features, two networks were constructed: one with the inclusion of the spatiotemporal attention mechanism and one without. This allowed for an analysis of the weights in the spatiotemporal attention layer.

Figure 7 displays the spatial feature weight distribution with and without the spatial attention layer. It can be observed that when the spatial attention layer is added,

the weight distribution from high to low is VP, CNL, GR, and DEN, which is consistent with the distribution of the correlation coefficients between shear wave velocity and other logging data in Figure 3. The emergence of this consistency can be explained as follows: the spatial attention layer learns the relationships between logging data, especially the correlation with shear wave velocity, to determine the importance of different features in predicting shear wave velocity. If a certain feature is highly correlated with shear wave velocity, the spatial attention layer assigns a higher weight to that feature to ensure that the model captures this information effectively. Conversely, without the inclusion of the spatial attention layer, the model cannot effectively extract the spatial features among different logging data. In the shear wave velocity prediction of the STAGRU fusion network, the spatial attention layer assigns the highest weight to VP, reaching 0.45. This implies that VP has the most significant influence on shear wave velocity. The reason for this phenomenon is that longitudinal wave velocity and shear wave velocity reflect different elastic information in rocks, and they are positively correlated. Particularly in sedimentary formations, the correlation coefficient between the two can exceed 0.9 due to the complexity of the strata and variations in strata properties. Therefore, the introduction of the spatial attention mechanism helps capture this physical correlation, thereby enhancing the prediction performance of shear wave velocity. In summary, the above analysis further validates the rationale for adding the spatial attention mechanism in this study because it enables more accurate capture of the physical relationships between shear wave velocity and other logging data. This contributes to improving the model's performance and reliability.

Figure 8 shows the weight distribution of temporal features with and without the temporal attention layer. Fifteen sampling points were selected from the conventional well logging data as samples, and the transverse wave velocity in the middle of the sample length was chosen as the label. It can be observed that the temporal attention layer assigns different weights to the temporal features
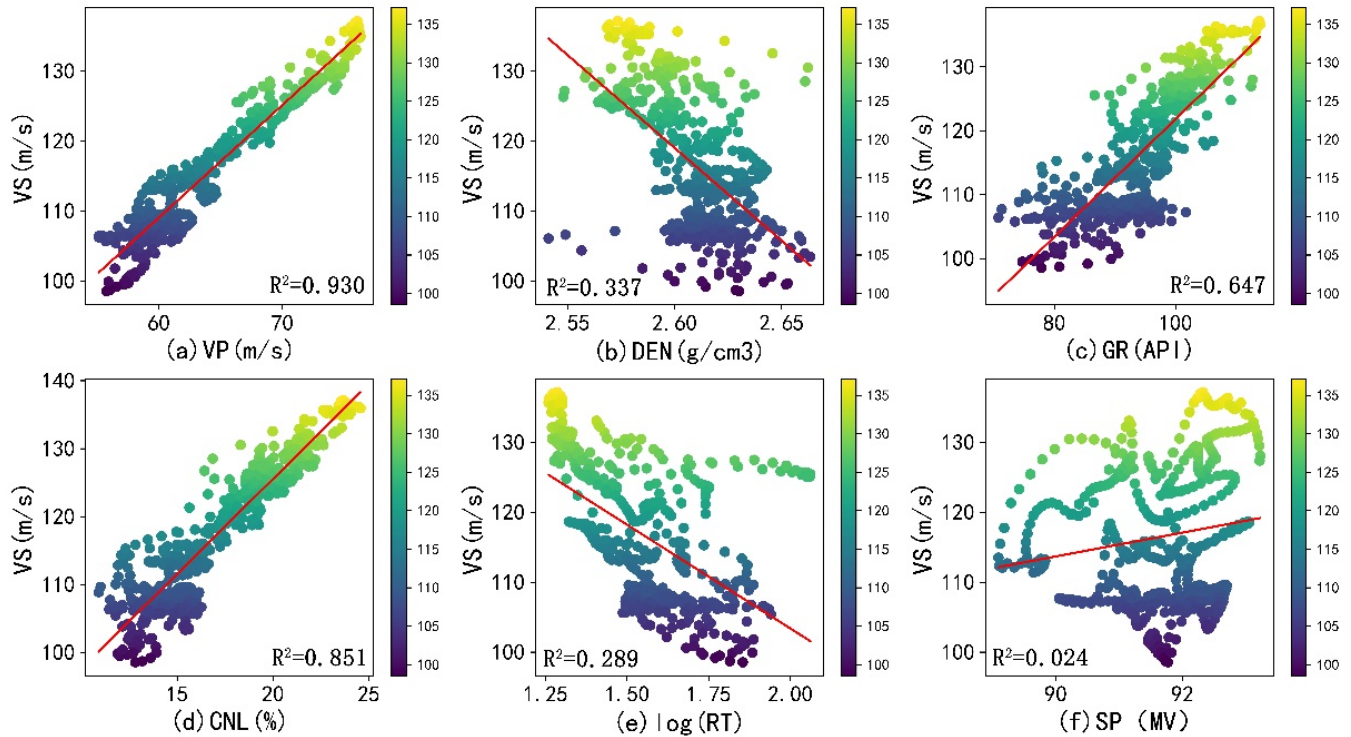
FIGURE 3: Scatter plot of conventional well data and shear wave velocity. (a) VP–VS; (b) DEN–VS; (c) GR–VS; (d) CNL–VS; (e) log (RT)–VS; (f) SP–VS; $R^2$ is the correlation coefficient between the two variables.
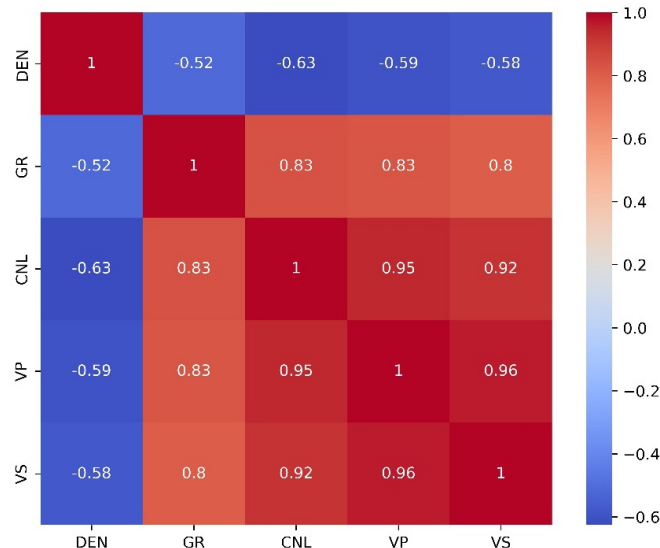


FIGURE 4: Cross-correlation analysis of conventional well data.

of the well logging data. In the transverse wave velocity prediction of the STAGRU fusion network, the sample data at the label position are assigned the highest weight by the temporal attention layer, reaching 0.204. This indicates that it has the greatest impact on the transverse wave velocity. As the distance increases from the label position, the overall trend of the attention weights is decreasing, which is consistent with the autocorrelation distribution of the conventional well logging data shown in Figure 5.

The reason for this phenomenon is the gradual variation in mineral composition in the sedimentary formation, which results in a certain level of autocorrelation in the well logging data along the depth direction. This validates the rationale for incorporating the time attention mechanism in this study.

*3.4. Network Comparison Analysis.* To validate the performance of the proposed STAGRU fusion network in
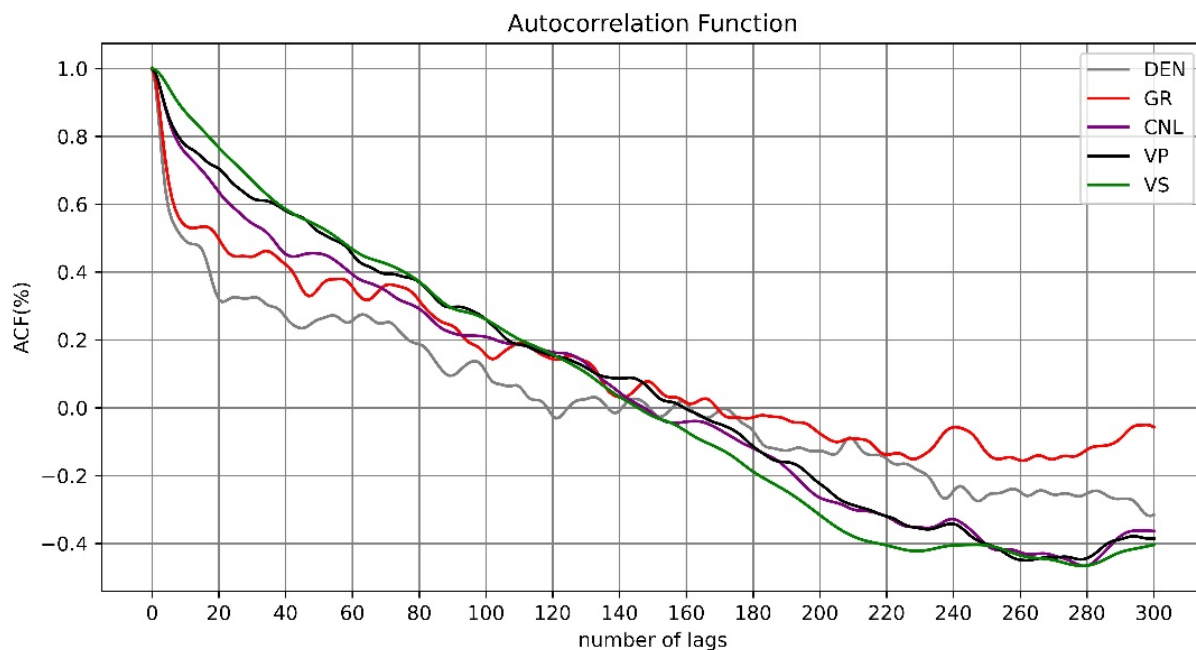
FIGURE 5: Autocorrelation coefficients of conventional well data.
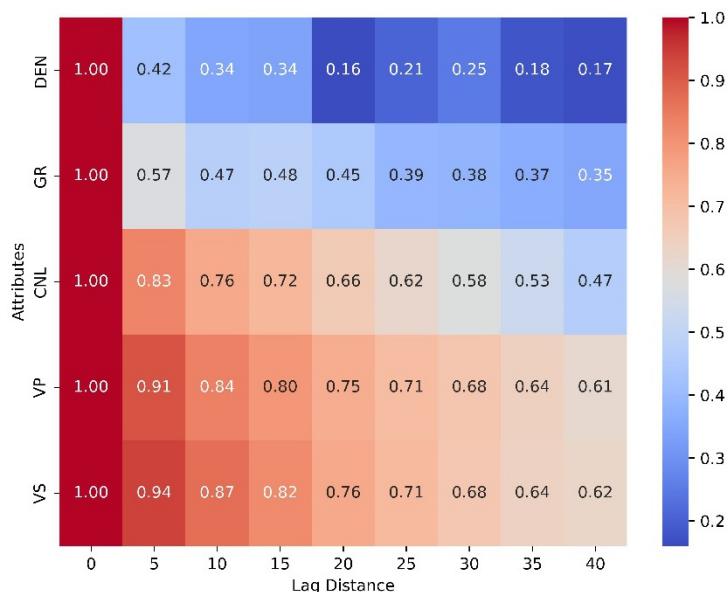


FIGURE 6: Relationship between autocorrelation and lag distance of conventional well data.

this study, network structures for STAGRU, 2DCNN-GRU, and GRU are constructed as shown in Table 1. All networks utilize the Adaptive Moment Estimation (Adam) as the optimization algorithm, which combines the advantages of AdaGrad and RMSProp algorithms. Adam adapts different learning rates for different parameters automatically and can converge well even in unstable objective functions. It can address the issue of rapidly decreasing gradients and demonstrates strong advantages in handling large-scale data and parameter optimization. In addition, Dropout layers are incorporated to randomly drop neurons, reducing overfitting and enhancing the generalization capability of the network. Furthermore, Early Stopping is employed during the training process to prevent overfitting and improve the model's generalization ability.

To further evaluate the accuracy of the STAGRU fusion network in predicting shear wave velocity, single-parameter fitting, multiparameter fitting, and rock physics methods are employed to calculate the shear wave velocity. The results are then compared and analyzed against the predictions
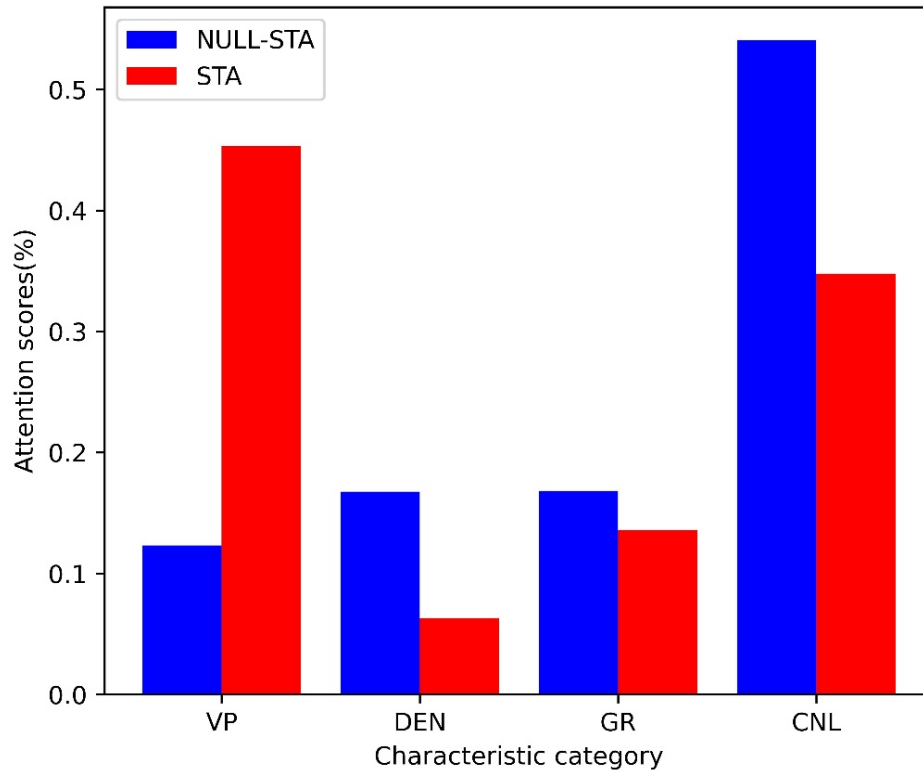
FIGURE 7: The distribution of spatial feature weights with and without spatial attention layer.
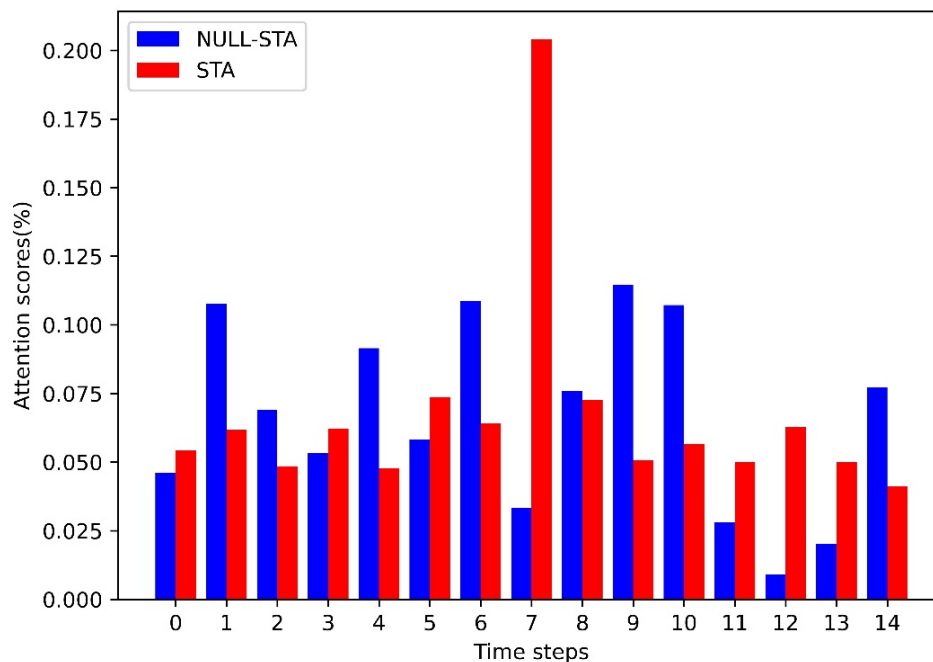


FIGURE 8: The weight distribution of temporal features with and without the time attention layer.

of the STAGRU, 2DCNN-GRU, and GRU methods. Based on the correlation analysis in section 3.2 and following the principles of linear regression, a single-parameter fitting is performed between shear wave velocity and compressional wave velocity, resulting in Equation 12. Subsequently, a multiparameter fitting is conducted between shear wave velocity and compressional wave velocity, density, natural gamma, and neutron porosity, resulting in Equation 13. For the rock physics method, the improved Xu-White model [12] is used. This method utilizes a particle swarm

TABLE 1: STAGRU, 2DCNN-GRU, and GRU network architectures.

| Network name | Network structure | Kernel size of 2DConv layers | Parameters of the gated recurrent unit | Parameters of spatial and temporal attention layers | Parameters of the dense layer | Trainable parameters | Other parameters |
|---|---|---|---|---|---|---|---|
| STAGRU | GRU + Dense + SATT + GRU + Dense + TATT + Dropout + Dense | 0 | 18 + 32 | 4 + 15 | 18 + 32 +1 | 8167 | Return_ sequences=true |
| 2DCNN-GRU | Conv + GRU + GRU + Dense | (15 × 3) | 16 + 32 | 0 | 1 | 25281 | Padding='same' Return_ sequences=true |
| GRU | GRU + GRU + Dense | 0 | 4 + 16 | 0 | 1 | 6769 | Return_ sequences=true |

SATT, spatial attention layer. TATT, temporal attention layer. Dropout, 0.25.

algorithm to estimate the porosity ratio based on the measured compressional wave velocity. From the porosity ratio, the shear wave velocity can be estimated.

$$y_{VS} = 1.609 x_{VP} + 12.498 \qquad (12)$$

$$y_{VS} = 1.461 x_{VP} - 0.589 x_{DEN} + 0.005 x_{GR} \\ + 0.271 x_{CNL} + 18.632 \qquad (13)$$

The well logging data from four wells in a certain area were used to train the STAGRU, 2DCNN-GRU, and GRU networks. Figure 9 displays the loss curves for the three network types on the validation and training sets. It can be observed that as the number of training iterations increases, the loss errors continuously decrease and eventually reach a stable constant value. This indicates that the networks have reached their optimal states. However, the loss error of the STAGRU fusion network is lower than that of the GRU network and 2DCNN-GRU hybrid network, indicating that the STAGRU fusion network is better at capturing the spatiotemporal features and the correlation between conventional well logging data and transverse wave velocity. This suggests that the STAGRU fusion network has an improved sensitivity to important spatiotemporal features.

The red curves in Figure 10, from left to right, represent the prediction results of the single-parameter fitting method, multiparameter fitting method, rock physics modeling method, GRU network, 2DCNN-GRU hybrid network, and STAGRU fusion network on the training set. The MSE, MAE, and correlation coefficient $R^2$ for these five to six methods are shown in Table 2. It can be observed that the predicted transverse wave velocities using deep learning methods closely match the true values, with smaller MSE and MAE compared with the fitting method and rock physics modeling method. This indicates that deep learning has certain advantages in transverse wave velocity prediction, especially in the depth range of 5425–5435 m. Furthermore, the prediction performance of the STAGRU fusion network is better than that of the GRU network and 2DCNN-GRU hybrid network, particularly evident in the

depth range of 5455–5475 m, where STAGRU can handle abrupt changes more effectively. This demonstrates that the STAGRU fusion network exhibits slightly higher predictive performance than the GRU network.

To further validate the predictive accuracy and generalization of the proposed network, the logging data of well Y, which was not involved in the training process, were inputted into the trained model for testing. The prediction results of the single-parameter fitting method, multiparameter fitting method, Xu-White model, GRU, 2DCNN-GRU, and STAGRU were analyzed, as shown in Figure 11. In mudstone sandstone formations, at depths of 5535–5555 and 5520–5545 m in well Y, the predicted values of the single-parameter fitting method, multiparameter fitting method, Xu-White model, GRU network, and 2DCNN-GRU hybrid network show significant discrepancies compared with the measured values. However, the predicted values of the STAGRU network exhibit smaller differences from the measured values. This observation indicates that the proposed network has better predictive performance compared with the other five methods. Table 3 presents the comparative results of the six methods using MSE, MAE, and coefficient of determination ($R^2$) as quantitative evaluation metrics. It can be observed that in Y well, the STAGRU fusion network has the lowest MAE and the highest $R^2$. The evaluation results indicate that the STAGRU fusion network exhibits higher prediction accuracy and generalization capability.

*3.5. Optimization of Network Parameters.* The key to improving the accuracy of the model estimation lies in adjusting the hyperparameters. In order to obtain the optimal hyperparameters, parameters are systematically changed to test and evaluate the model's performance. As the underground formations exhibit certain sedimentation patterns in the vertical depth direction, and there is a certain correlation between sequential sampling points, it indicates that the sample length of the input data to the network affects the prediction of shear wave velocity in deep learning. To select the optimal sample length and achieve higher accuracy for the STAGRU fusion network, while
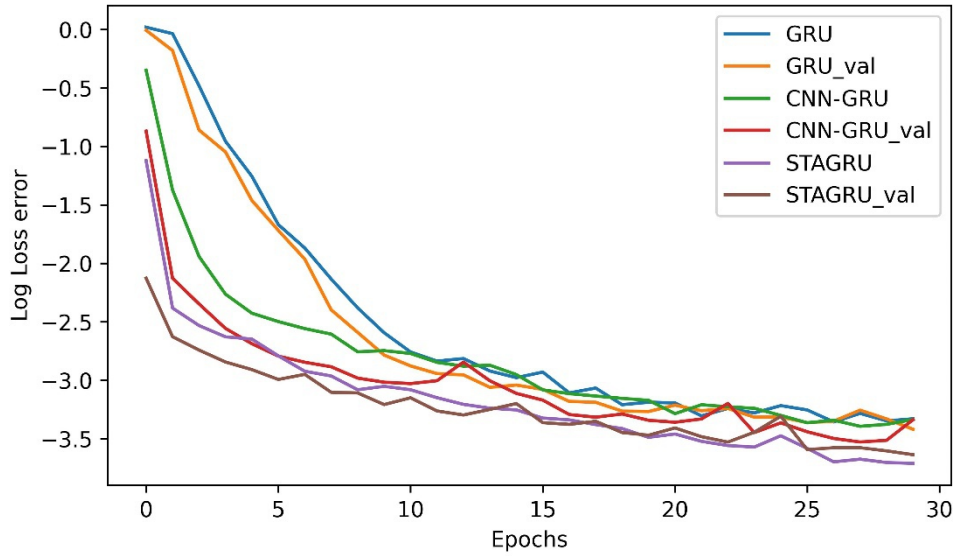
FIGURE 9: The logarithmic loss error curves for the STAGRU network compared with the 2DCNN-GRU and GRU networks (the term "val" represents the validation set).
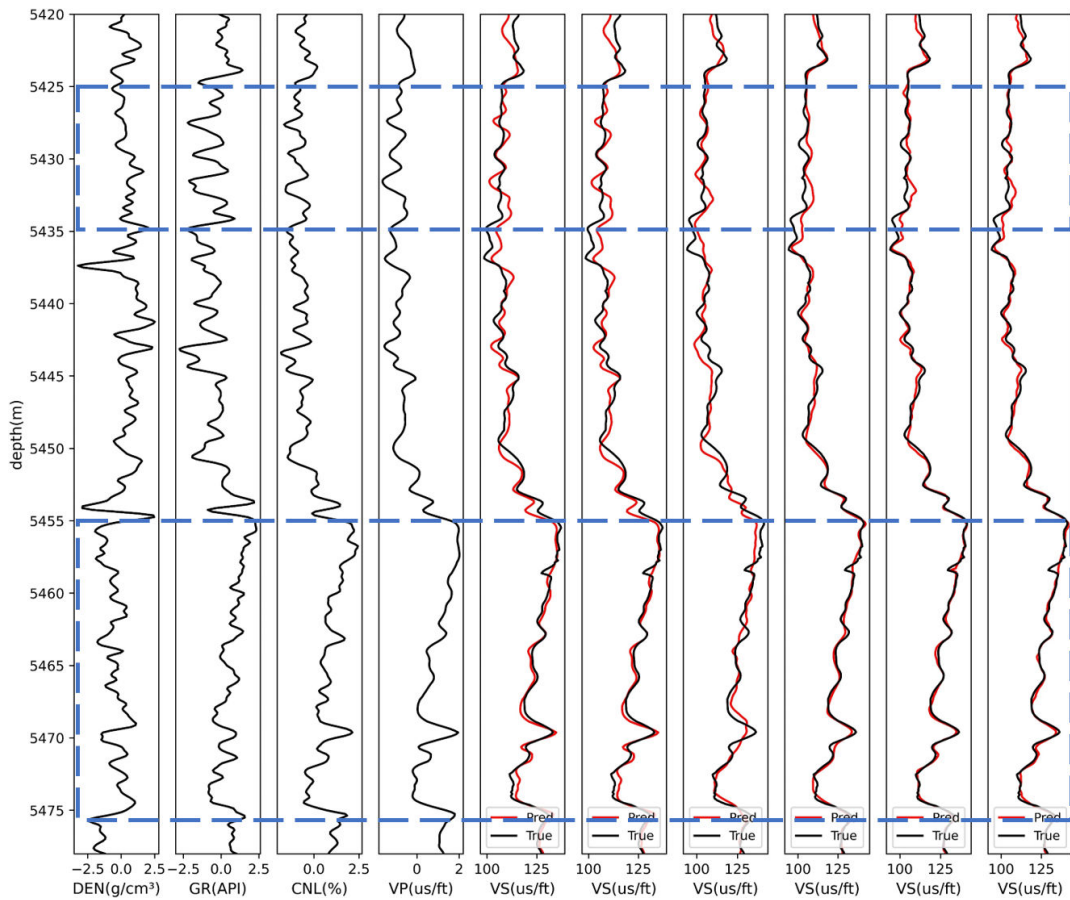


FIGURE 10: Prediction results of single-parameter fitting method, multiparameter fitting method, Xu-White model method, GRU, 2DCNN-GRU, and STAGRU networks on the training set.

keeping the GRU parameters at 18 and 32, and the input for the spatial attention layer at 4, experiments were conducted with sample lengths set to 5, 20, and 40. The structure of the STAGRU fusion network obtained in this experiment
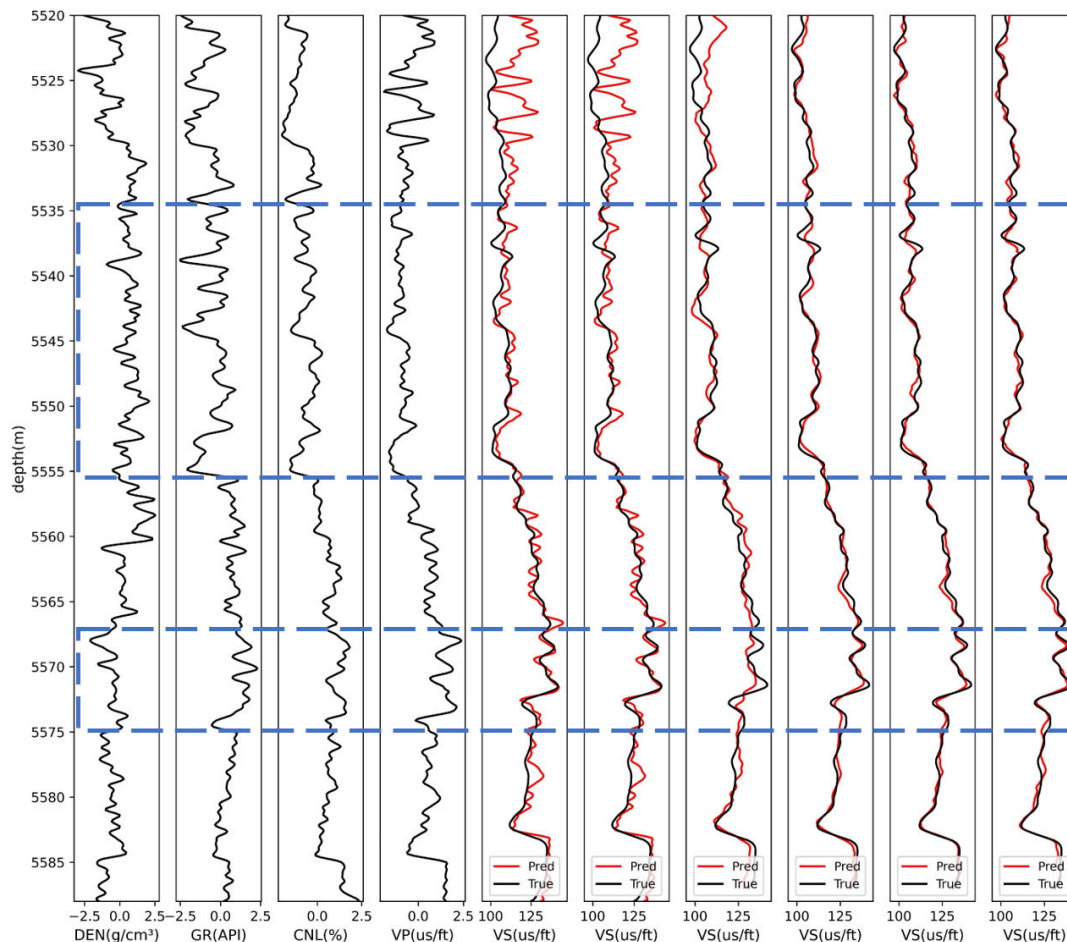
FIGURE 11: Prediction results of single-parameter fitting method, multiparameter fitting method, Xu-White model method, GRU, 2DCNN-GRU, and STAGRU networks on the test set.

TABLE 2: Comparison of prediction results of single-parameter fitting method, multiparameter fitting method, Xu-White model method, GRU, 2DCNN-GRU, and STAGRU networks on the training set.

| Method | MSE | MAE | $R^2$ | Training time (s) |
|---|---|---|---|---|
| Single-parameter fitting | 4.811 | 2.078 | 0.930 | |
| Multiparameter fitting | 4.454 | 2.003 | 0.935 | |
| Xu - White model | 3.412 | 1.603 | 0.949 | |
| GRU | 2.369 | 1.240 | 0.974 | 12.03 |
| 2DCNN-GRU | 2.297 | 1.207 | 0.976 | 9.81 |
| STAGRU | 1.546 | 1.021 | 0.981 | 10.55 |

TABLE 3: Comparison of prediction results of single-parameter fitting method, multiparameter fitting method, Xu-White model method, GRU, 2DCNN-GRU, and STAGRU networks on the test set.

| Method | MSE | MAE | $R^2$ | Testing time (s) |
|---|---|---|---|---|
| Single-parameter fitting | 4.965 | 4.675 | 0.618 | |
| Multiparameter fitting | 4.548 | 4.244 | 0.691 | |
| Xu - White model | 3.612 | 3.099 | 0.812 | |
| GRU | 2.943 | 1.376 | 0.853 | 0.37 |
| 2DCNN-GRU | 2.248 | 1.199 | 0.870 | 0.24 |
| STAGRU | 1.874 | 1.172 | 0.879 | 0.39 |

is shown in Table 4. The prediction and evaluation results of the STAGRU fusion network in this experiment are presented in Figure 12 and Table 5, respectively. From the figure, it can be observed that the prediction performance of STAGRU varies with different sample lengths. When the sample length is set to 20, the network achieves the best prediction performance, with a corresponding $R^2$ coefficient

of 0.884. This further confirms the excellent performance of the STAGRU fusion network.

## 4. Conclusion

Due to the complex pore structure of the mudstone reservoir in the Junggar Basin, conventional networks have

TABLE 4: Structure of the STAGRU fusion network in the experiments.

| Network name | Network structure | Sample length | Number of GRU units | Parameters of the dense layer | Parameters of spatial and temporal attention layers | Trainable parameters | Other parameters |
|---|---|---|---|---|---|---|---|
| STAGRU | GRU + Dense + SATT + GRU + Dense + TATT + Dropout + Dense | 5 | 18 + 32 | 18 + 32 + 1 | 4 + 5 | 7847 | Return_sequences=true |
| | | 20 | | | 4 + 20 | 8327 | |
| | | 40 | | | 4 + 40 | 8967 | |

SATT, spatial attention layer. TATT, temporal attention layer. Dropout, 0.25.
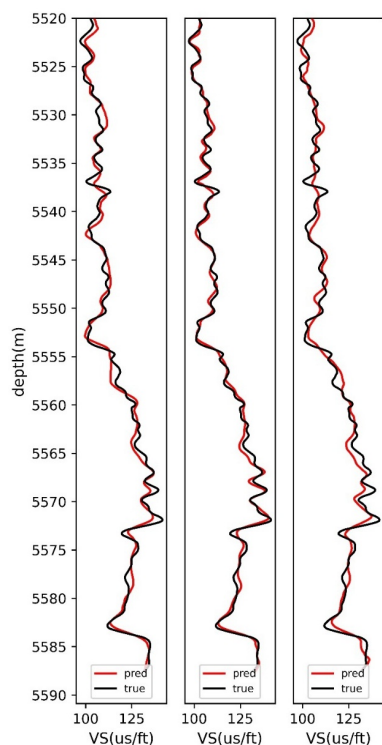


FIGURE 12: Experiments based on the STAGRU fusion network (sequence lengths from left to right: 5, 20, and 40).

TABLE 5: $R^2$ based on STAGRU fusion network experiment.

| Sample length | $R^2$ |
|---|---|
| 5 | 0.875 |
| 20 | 0.884 |
| 40 | 0.869 |

limited sensitivity to important spatiotemporal features. This study proposes a STAGRU. The results demonstrate that the weight distribution of the spatiotemporal attention layer is consistent with the autocorrelation of conventional well logging data and the correlation between well logging data and shear wave velocity. This verifies that the proposed network can improve the sensitivity of the network to important spatiotemporal features and validates the rationality of adding the spatiotemporal attention mechanism. Furthermore, the test results indicate that in the mudstone reservoir, the STAGRU fusion network achieves an $R^2$ value that is 3.6% higher than that of the GRU network and 1.6% higher than that of the 2DCNN-GRU hybrid network. This suggests that the proposed network exhibits superior predictive accuracy compared with the single-parameter fitting method, multiparameter fitting method, Xu-White model, GRU neural network, and CNN-GRU hybrid networks.

It should be noted that while the method proposed in this paper can accurately predict shear wave velocity for the Qigu Formation reservoir in this specific area, further research is needed if this method is to be applied to well log data from reservoirs outside of the Qigu Formation.

## Data Availability

Data used to support the results of this study can be found in this manuscript text.

## Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

## References

[1] H. Wang, J. Ma, L. Li, et al, "Research on prediction method of shear wave velocity in tight reservoirs: A case study of the Sulige gas field," *Progress in Geophysics*, vol. 34, no. 4, pp. 1521–1529, 2019.

[2] Y. Chen, Z. Sun, and K. Xu, "Research on pre-stack multi-parameter seismic inversion method for shale gas reservoirs," *Petroleum Geophysics*, vol. 61, no. 6, pp. 1016–1027, 2022.

[3] G. Gao, Z. He, J. Cao, et al, "Application of two-term elastic impedance inversion method in deep carbonate reservoir prediction," *Petroleum Geophysical Exploration*, vols. 506, 330–331, no. 3, pp. 450–457, 2013.

[4] Q. Liu, W. Hu, and Q. Li, "Application of pre-stack AVO Optimized inversion based on Xu-white model in deep-sea Turbidite reservoirs," *Mineral Resources Exploration*, vol. 14, no. 2, pp. 214–222, 2023.

[5] W. Li, H. Wang, Z. Yao, et al, "Shear wave velocity inversion and fluid substitution based on constrained conditions," *Chinese Journal of Geophysics*, vol. 52, no. 3, pp. 785–791, 2009.

[6] J. P. Castagna, M. L. Batzle, and R. L. Eastwood, "Relationships between Compressional-wave and shear-wave velocities in Clastic silicate Rocks[Z]," *GEOPHYSICS*, vol. 50, no. 4, pp. 571–581, 1985.

[7] . De-hua HanA. Effects of porosity and clay content on wave velocities in sandstones[Z]: 1-2.

[8] M. L. Greenberg and J. P. Castagna, "Shear-wave velocity estimation in porous rocks: theoretical formulation, preliminary verification and Applications1[J]," *Geophysical Prospecting*, vol. 40, no. 2, pp. 195–209, 1992. https://onlinelibrary.wiley.com/toc/13652478/40/2.

[9] S. Xu and R. E. White, "A new velocity model for Clay-Sand Mixtures[J]," *Geophysical Prospecting*, vol. 43, pp. 91–118, 1995.

[10] M. W. Lee, "A simple method of predicting S-wave Velocity[J]," *GEOPHYSICS*, vol. 71, no. 6, pp. F161–F164, 2006.

[11] G. Zhang, C. Li, X. Yin, et al, "Estimation method of carbonate rock shear wave velocity based on modified Xu-white model," *Petroleum Geophysical Exploration*, vol. 47, no. 5, pp. 717–722, 2012.

[12] Y. Dou, G. Gao, L. Liang, et al, "Prediction of shear wave velocity based on Xu-white model," *Xinjiang Petroleum Geology*, vol. 37, no. 1, pp. 83–87, 2016.

[13] K. L. Zhang, "Estimation method of shale gas reservoir shear wave velocity based on rock physics model, " *Science Technology and Engineering*, vol. 19, no. 11, pp. 10–15, 2019.

[14] S. Wang, G. Yang, and X. Mu, "Prediction of shear wave velocity based on deep feedforward neural network method," *Petroleum Geology and Recovery Efficiency*, vol. 29, no. 1, pp. 80–89, 2022.

[15] K. Yu, L. Jia, Y. Chen, et al, "Deep learning: yesterday, today, and tomorrow," *Journal of Computer Research and Development*, vol. 50, no. 9, pp. 1799–1804, 2013.

[16] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *IEEE Computational Intelligence Magazine*, vol. 13, no. 3, pp. 55–75, 2018.

[17] "Shenghua GAO, Yuting Zhang, Kui Jia, Jiwen LU, member, IEEE, and Yingying Zhang [J]," , 1–509, n.d.

[18] X. J. Xiong, Y. Ma, et al, "Seismic fault detection with Convolutional neural Network[J], " *Geophysics*, vol. 83, pp. 97–O103, 2018.

[19] Y. Imamverdiyev and L. Sukhostat, "Lithological Facies classification using deep Convolutional neural Network[J]," *Journal of Petroleum Science and Engineering*, vol. 174, pp. 216–228, 2019.

[20] J. Wang, J. Cao, J. You, et al, "Prediction of reservoir porosity-permeability parameters based on Gated recurrent unit neural network," *Petroleum Geophysical Exploration*, vol. 59, no. 4, pp. 616–627, 2020.

[21] Z. Dong, L. Wang, Z. Du, et al, "Seismic impedance inversion based on closed-loop CNN-GRU and active learning," in *Proceedings of the 2021 Symposium on Exploration Techniques of China Petroleum Society*, pp. 1062–1065, China Academic Journal Electronic Publishing House, 2021.

[22] H. Jo, Y. Cho, M. Pyrcz, H. Tang, and P. Fu, "Machine-learning-based porosity estimation from Multifrequency Poststack seismic data," *GEOPHYSICS*, vol. 87, no. 5, pp. M217–M233, 2022.

[23] H. Tang, P. Fu, H. Jo, et al., "Deep learning-accelerated 3d carbon storage reservoir pressure forecasting based on data assimilation using surface displacement from Insar," *International Journal of Greenhouse Gas Control*, vol. 120, p. 103765, 2022.

[24] Q. Ma, X. Zhang, C. Zhang, et al, "Prediction of shear wave velocity based on one-dimensional Convolutional neural network," *Lithologic Reservoirs*, vol. 33, no. 4, pp. 111–120, 2021.

[25] J. You, J. Cao, X. Wang, and W. Liu, "Shear wave velocity prediction based on LSTM and its application for morphology identification and saturation inversion of gas Hydrate[J], " *Journal of Petroleum Science and Engineering*, vol. 205, p. 109027, 2021.

[26] H. Zhou, Z. Wu, X. Zhang, et al, "Prediction of shear wave based on LSTM recurrent neural network," *Fault-Block Oil & Gas Field*, vol. 28, no. 6, pp. 829–834, 2021.

[27] Y. Sun and Y. Liu, "Prediction of shear wave velocity using GRU neural network," *Petroleum Geophysical Exploration*, vol. 55, no. 3, pp. 484–492, 2020.

[28] X. Zhou, J. Cao, X. Wang, et al, "Acoustic logging curve reconstruction technique based on Bidirectional Gated recurrent unit neural network," *Progress in Geophysics*, vol. 37, no. 1, pp. 357–366, 2022.

[29] J. Wang and J. Cao, "Data-driven S-wave velocity prediction method via a deep-learning-based deep Convolutional Gated recurrent unit fusion Network[J]," *GEOPHYSICS*, vol. 86, no. 6, pp. M185–M196, 2021.

[30] J. Wang, J. Cao, S. Zhao, and Q. Qi, "S-wave velocity inversion and prediction using a deep hybrid neural Network[J]," *Science China Earth Sciences*, vol. 65, no. 4, pp. 724–741, 2022.

[31] T. Chen, G. Gao, P. Wang, et al., "Prediction of shear wave velocity based on a hybrid network of two-dimensional Convolutional neural network and Gated recurrent Unit[Z]," *Geofluids*, vol. 2022, pp. 1–14, 2022.

[32] H. Pham, C. D. Manning, and M. T. Luong, "Effective approaches to attention-based neural machine Translation[J]," *Computer Ence*, vol. 2015, pp. 1–2, n.d.

[33] J. Deng, L. Cheng, and Z. Wang, "Attention-based Bilstm fused CNN with gating mechanism model for Chinese long text Classification[J]," *Computer Speech & Language*, vol. 68, p. 101182, 2021.

[34] J. Du, Y. Cheng, Q. Zhou, J. Zhang, X. Zhang, and G. Li, "Power load forecasting using Bilstm-Attention[J]," *IOP Conference Series*, vol. 440, no. 3, p. 032115, 2020.

[35] P. Kavianpour, E. Jahani, and M. Kavianpour, "A CNN-Bilstm model with attention mechanism for earthquake Prediction[J]," 1–2, 2021.

[36] S. M. Mousavi, W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C. Beroza, "Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase Picking[J]," *Nature Communications*, vol. 11, no. 1, 2020.

[37] Md. H. A. Banna, T. Ghosh, Md. J. A. Nahian, et al., "Attention-based bi-directional long-short term memory network for earthquake Prediction[J]," *IEEE Access*, vol. 9, pp. 56589–56603, 2021.

[38] T. Bai and P. Tahmasebi, "Attention-based LSTM-FCN for earthquake detection and Location[J]," *Geophysical Journal International*, vol. 228, no. 3, pp. 1568–1576, 2021.

[39] L. Shan, Y. Liu, M. Tang, M. Yang, and X. Bai, "CNN-Bilstm hybrid neural networks with attention mechanism for well log Prediction[J]," *Journal of Petroleum Science and Engineering*, vol. 205, no. 2, p. 108838, 2021.

[40] B. V. K. Merriënboer, D. Bahdanau, and Y. H. Bengio, "On the properties of neural machine translation: Encoder-Decoder Approaches[J]," 2014. https://doi.org/10.48550/arXiv.1409.1259.

[41] IEEE, and P. Frasconi, P. Simard, et al, "Learning long-term dependencies with gradient descent is Difficult[J]," 1–2, n.d.

[42] Y. Liu, "Research on time series prediction based on GRU neural network," Master's thesis, Chengdu University of Technology, 2017.