GeoScienceWorld
Lithosphere
Volume 2022, Article ID 4701851, 12 pages
https://doi.org/10.2113/2022/4701851

Lithosphere

Research Article

Shear-Wave Velocity Prediction Method via a Gate Recurrent Unit Fusion Network Based on the Spatiotemporal Attention Mechanism

Tengfei Chen,[1] **Gang Gao**,[1] **Yonggen Li**,[2] **Peng Wang**,[1] **Bin Zhao**,[1] **Zhixian Gui**,[1] **and Xiaoyan Zhai**[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; 517906306@qq.com

Received 17 June 2022; Revised 18 November 2022; Accepted 25 November 2022; Published 31 December 2022

Academic Editor: Shuangpo Ren

Compression-wave velocity and shear-wave velocity are important elastic parameters describing deeply tight sandstone. Limited by cost and technical reasons, the conventional logging data generally lack shear-wave velocity. In addition, the existing rock physics theory is difficult to accurately establish the rock physics models due to the complex pore structure of tight sandstone reservoir. With the rapid development of the artificial intelligence, the attention mechanism that can increase the sensitivity of the network to important characteristics has been widely used in machine translation, image processing, and other fields, but it is rarely used to predict shear-wave velocity. Based on the correlation between the shear-wave velocity and the conventional logging data in the spatiotemporal direction, a gate recurrent unit (GRU) fusion network based on the spatiotemporal attention mechanism (STAGRU) is proposed. Compared with the convolutional neural network (CNN) and gate recurrent unit (GRU), the network proposed can improve the sensitivity of the network to important spatiotemporal characteristics using the spatiotemporal attention mechanism. It is analyzed that the relationship between the spatiotemporal characteristics of the conventional logging data and the attention weights of the network proposed to verify the rationality of adding the spatiotemporal attention mechanism. Finally, the training and testing results of the STAGRU, CNN, and GRU networks show that the prediction accuracy and generalization of the network proposed are better than those of the other two networks.

1. Introduction

Shear-wave velocity plays a crucial role in prestack elastic inversion, reservoir sensitive parameter analysis, and in situ stress analysis [1–5]. Limited by cost and current acquisition technology, shear-wave velocity is generally lacking in conventional logging data in most areas, especially in old wells. Therefore, developing a reliable and low-cost prediction method of shear-wave velocity is extremely important.

Previously, the empirical formula [6, 7] and the petrophysical model [8, 9] were used to predict shear-wave velocity. However, the empirical formula varies with the region and lithology, which leads to the problem of insufficient generalization. The prediction accuracy of petrophysical model depends on the calculation accuracy of each parameter, but some parameters in complex reservoirs are difficult to obtain accurately [10], which limits the application of petrophysical model.

Deep learning can establish the nonlinear-mapping relationship between input and output, so it has been widely used in face recognition, machine translation, image classification, and other fields [11–13]. In recent years, some geophysicists have introduced deep learning into the field of geology and achieved some research results in the fields of petrophysics, lithology identification, and reservoir parameter inversion [14–16]. Among them, in the aspect of shear-wave velocity modeling, scholars mainly use recurrent neural network (RNN) [17] and convolutional neural network (CNN) [18]

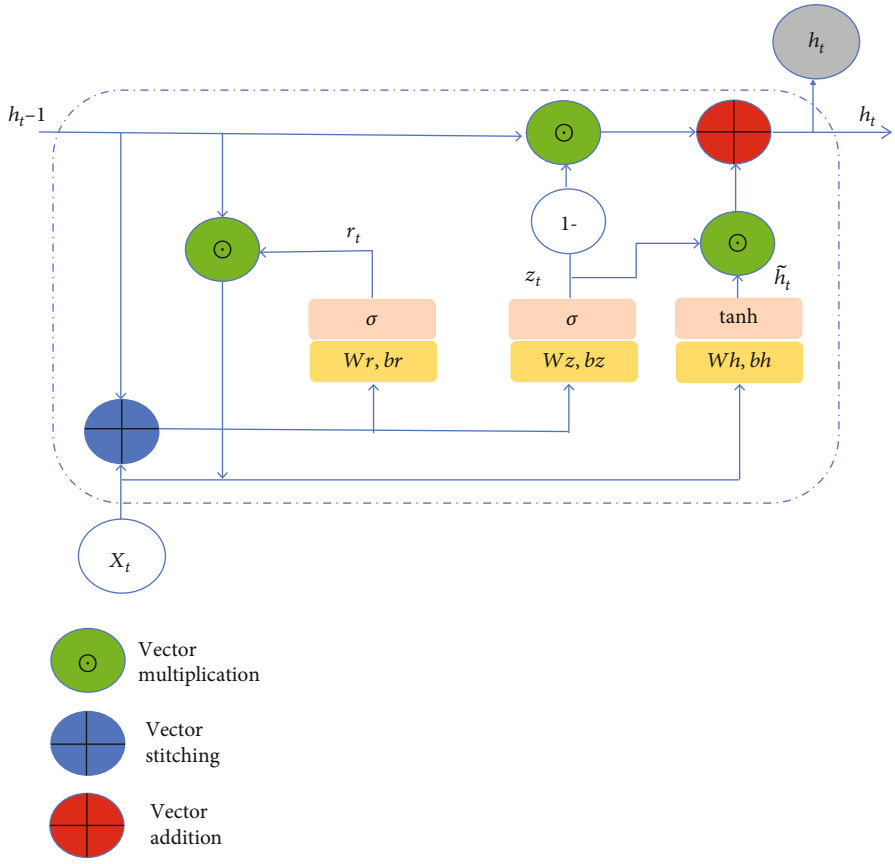

Figure 1: Structure of GRU network.

to establish the nonlinear mapping between input and output [19–21]. Considering the powerful function of CNN extracting spatial characteristics, Ma et al. [22] used CNN to predict the shear-wave velocity, and the prediction accuracy of the shear-wave velocity has been improved. The logging data has regularity in the depth direction, and compared with the CNN, the RNN is more suitable for dealing with conventional logging data. Some scholars [23, 24] proposed a method to deal with the lack of shear-wave velocity using a long short-term memory (LSTM) network [25], which fully considered the temporal characteristics of the conventional logging data and achieved good prediction results in carbonate and sandstone reservoirs. Compared with the LSTM network, GRU [26] can reduce the network training parameters and has been widely used to predict shear-wave velocity and porosity [21, 27]. However, the above methods only focus on the spatial characteristics or temporal characteristics of the conventional logging data and ignore the influence of the spatiotemporal characteristics of the conventional logging data on shear-wave velocity. To comprehensively consider the impact of the spatiotemporal characteristics of the conventional logging data on shear-wave velocity, the fusion network composed of CNN and LSTM or GRU is suggested by some scholars [28–30]. Although the above method improves the performance of the network, it does not highlight the impact of important spatiotemporal characteristics on shear-wave velocity, so there is a higher requirement for the weight distribution of spatiotemporal characteristics extracted from the network.

In recent years, the neural networks based on the attention mechanism have been used in machine translation, power load, and geoscience [31–36], and these scholars claim that the attention mechanism can improve the sensitivity of the network to important characteristics. Kavianpour et al. [37] developed an attention-based CNN-BILSTM fusion network to predict earthquakes and achieved good prediction results. To improve the accuracy of the seismic event detection and localization, an attention mechanism-based LSTM-FCN fusion network is proposed by Bai and Pejman [38]. Compared with the ConvNetQuake model [39], the prediction accuracy and classification performance of the network are effectively improved. Shan et al. [40] provided a fusion network based on the CNN-BILSTM, which is used to predict the logging data to reduce drilling costs.

The above literatures show that deep learning network has been widely used in seismic facies, lithology identification, reservoir parameter inversion, and other fields, but few neural networks based on the attention mechanism are used to predict shear-wave velocity. In addition, shear-wave velocity has a certain correlation with spatiotemporal characteristics of the conventional logging data; a GRU fusion network based on the spatiotemporal attention mechanism (STAGRU) is proposed, which mainly includes the GRU layer, the spatial attention layer, the temporal attention

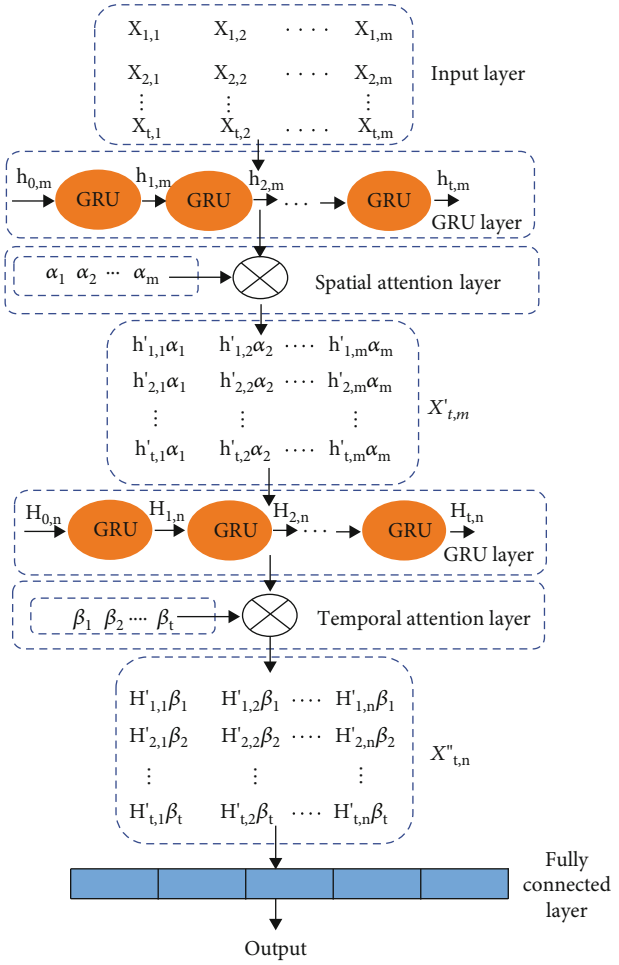

Figure 2: Structure of STAGRU fusion network.

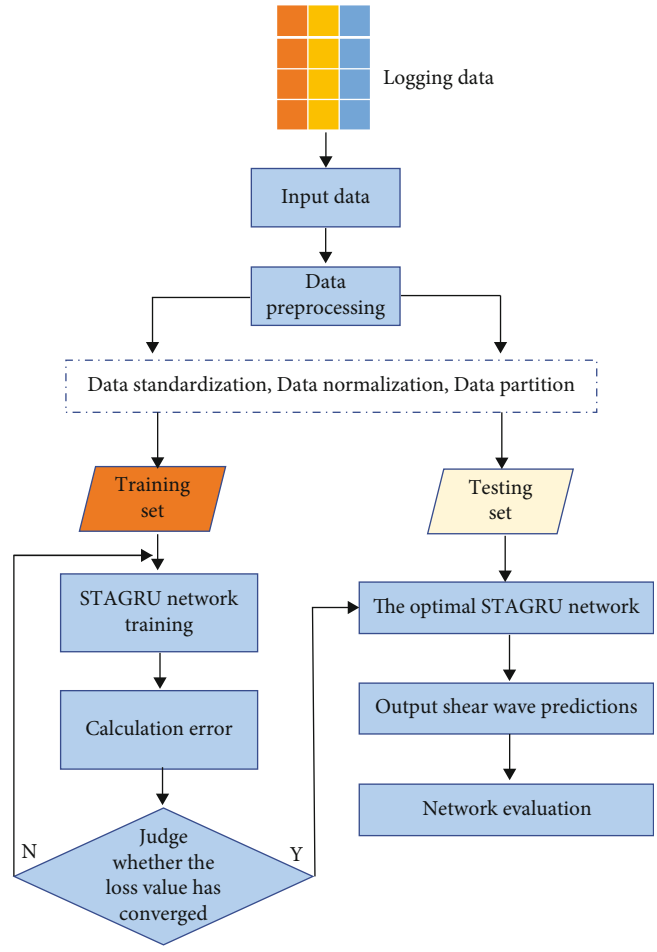

Figure 3: The shear-wave velocity training and prediction flow chart of STAGRU fusion network.

layer, and the fully connected layer. In this paper, the tight sandstone reservoir in the Junggar Basin is used as the research object. Based on the STAGRU fusion network, a training and prediction flow about the shear-wave velocity is established, and the weight distribution of the attention layer is analyzed. Finally, the training and prediction results of the STAGRU, CNN, and GRU networks show the network proposed has higher prediction accuracy and generalization.

2. Methods

2.1. Gate Recurrent Unit (GRU). The gate recurrent unit (GRU) is a variant of the recurrent neural network (RNN), which not only solves the problem of vanishing or exploding gradients in the traditional RNN [41] but also solves the problem of long calculation time of the LSTM. On the other hand, it sets up two gates in the hidden layer—reset gate (r_t) and update gate (z_t) (Figure 1). The reset gate and update gate perform retention and forgetting functions, respectively, according to the input of the current moment. When the reset gate (r_t) is closer to 1, it means that more information is preserved. When the update gate (z_t) is close to 1, it means that more information is forgotten. When the logging data x_t is input at time t, the reset gate (r_t) and update gate (z_t) can be expressed as

$$\begin{aligned} r_t &= \sigma(W_r[h_{t-1}, x_t] + b_r), \\ z_t &= \sigma(W_z[h_{t-1}, x_t] + b_z), \end{aligned} \tag{1}$$

where W_r and W_z are weight matrices of the reset gate and the update gate, respectively; b_r and b_z are the biases; h_{t-1} is the output of the hidden state at time $t-1$; "σ" is the logistic sigmoid function, it can map the output to the range of [0,1]; and "[]" denotes two matrices are concatenated.

The new state $\tilde{h}_t$ contains the information controlled by the reset gate and is combined with the update gate to get the final output h_t:

$$\begin{aligned} \tilde{h}_t &= \tanh\left(W_h[r_t * h_{t-1}, x_t] + b_h\right), \\ h_t &= (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t, \end{aligned} \tag{2}$$

where W_h and b_h are weight matrix and bias of the new state $\tilde{h}_t$, respectively; "tanh" is the activation function; "*" means matrix multiplication; and h_t is the output of the current hidden state.

2.2. Spatial Attention Mechanism. Shear-wave velocity has a certain correlation with spatial characteristics of the

(a) (b) (c)

Figure 4: The cores of commercial oil wells. (a) Horizontal-bedding fracture. (b) Cross-bedding fracture. (c) Deformed-bedding fracture.

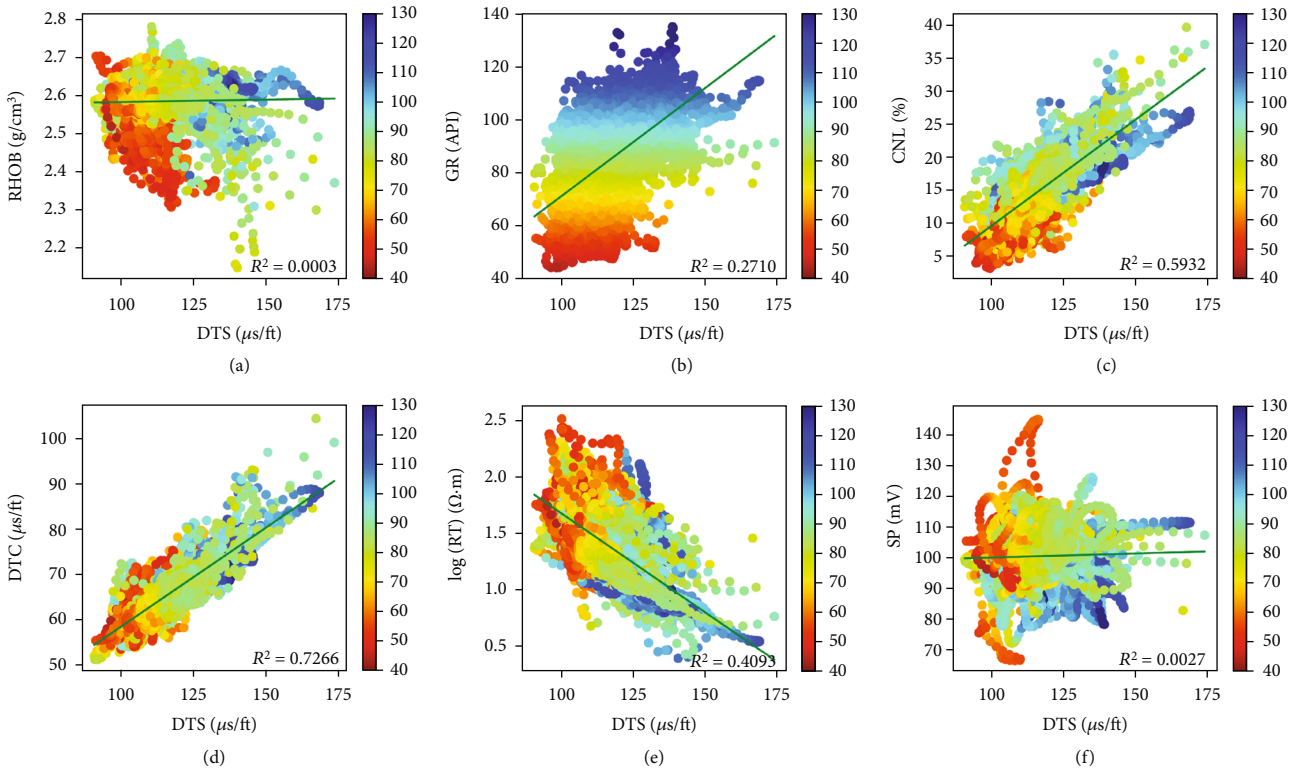

Figure 5: The cross plot of conventional logging data against shear-wave velocity (a) RHOB–DTS. (b) GR–DTS. (c) CNL–DTS. (d) DTC–DTS. (e) log (RT)–DTS. (f) SP–DTS. (R is the correlation coefficient).

conventional logging data. The spatial attention mechanism is added on the basis of the GRU network to improve the sensitivity of important spatial characteristics on the shear-wave velocity. The spatial characteristics of the conventional logging data are input to the spatial attention layer and the different weights of this layer are assigned to the spatial characteristics to obtain the output of the spatial attention layer. The hidden state $h_t = [h_{t,1}, h_{t,2}, h_{t,3} \cdots, h_{t,m}]$ is the m -dimensional characteristic vector at the t-th time step. The spatial attention weights can be expressed as

$$\begin{aligned} \alpha_t' &= \text{Soft max}\,(W_\alpha h_t + b_\alpha), \\ X_t' &= \alpha_t' \odot h_t = [\alpha_{t,1} h_{t,1}, \alpha_{t,2} h_{t,2} \cdots \alpha_{t,m} h_{t,m}], \end{aligned} \tag{3}$$

where $\alpha_t' = [\alpha_{t,1}, \alpha_{t,2} \cdots \alpha_{t,m}]$ is the weights of the spatial attention layer. W_α and b_α are the weight matrix and bias, respectively. The softmax is the normalization function. "$\odot$" denotes that the dot product. X_t' represents the weighted result.

2.3. Temporal Attention Mechanism. Since the logging data has regularity in the sedimentary formation, the temporal attention mechanism is added on the basis of the GRU network to improve the sensitivity of important temporal characteristics on the shear-wave velocity. The temporal characteristics are input to the temporal attention layer and the different weights of this layer are assigned to the temporal characteristics to obtain the output of the temporal attention layer. The hidden state $H_n = [H_{1,n}, H_{2,n}, H_{3,n} \cdots, H_{t,n}]$ is the t-dimensional vector of the n-th spatial characteristic. The temporal attention weights can be expressed as

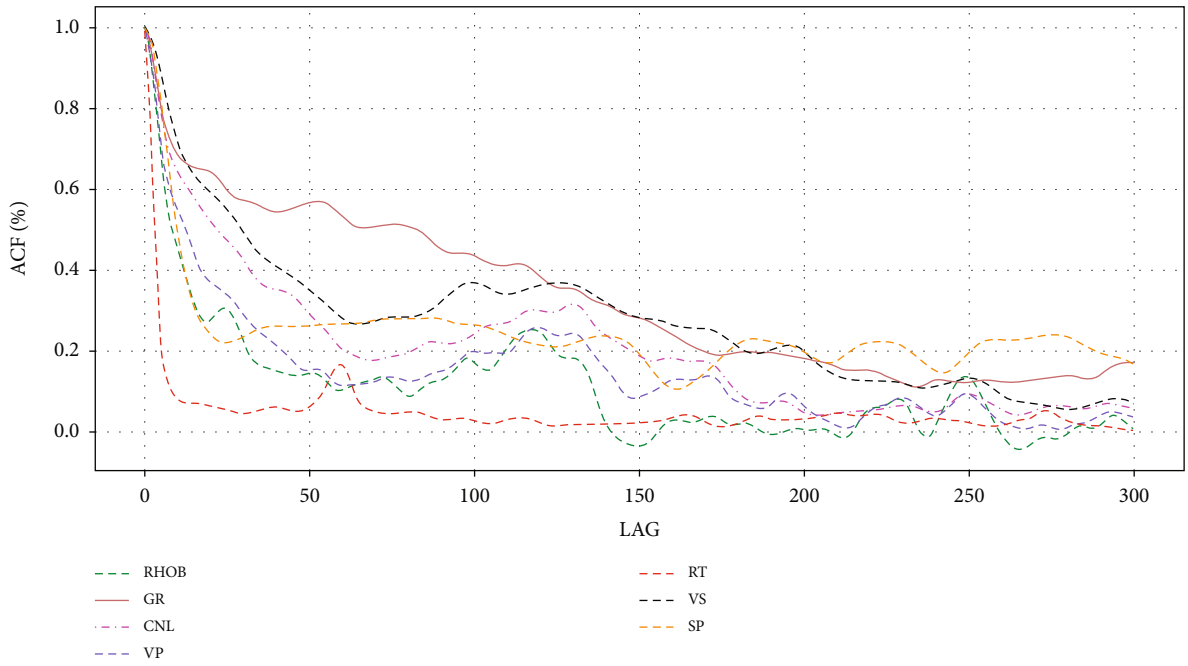

Figure 6: Autocorrelation coefficients of the conventional logging data.

$$\beta_n' = \mathrm{softmax}\left(W_\beta H_n + b_\beta\right),$$
$$X_n'' = \beta_n' \odot H_n = \left[\beta_{1,n} H_{1,n} \beta_{2,n} H_{2,n} \cdots \beta_{t,n} H_{t,n}\right], \tag{4}$$

where $\beta_n' = [\beta_{1,n}, \beta_{2,n} \cdots \beta_{t,n}]$ is the weights of the temporal attention layer. W_β and b_β are the weight matrix and bias, respectively. X_n'' represents the weighted result.

2.4. The Structure of the STAGRU Fusion Network. Since the shear-wave velocity has a certain correlation with spatiotemporal characteristics of the conventional logging data, a GRU fusion network based on the spatiotemporal attention mechanism is suggested (Figure 2). The STAGRU fusion network consists of an input layer, two GRU layers, a spatiotemporal attention layer, and a fully connected layer. The GRU layer is used to extract the spatiotemporal characteristics of the conventional logging data. The spatiotemporal attention layer is used to improve the sensitivity of the network to important spatiotemporal characteristics. The fully connected layer is used to increase the nonlinearity of the network proposed.

2.5. Training and Prediction Process of the STAGRU Fusion Network. The detailed process of training and prediction of the STAGRU fusion network in this study can be divided into the following steps (Figure 3):

(1) *Input Training Data*. The training set of the logging data is input to the STAGRU fusion network.

(2) *Data Preprocessing*. Due to the large gap between logging data, the StandardScaler and MinMaxScaler functions are used to preprocess the logging data and map the logging data to the range of [0,1], as shown in

Table 1: Parameters' setting of the STAGRU fusion network.

Parameters	Values
The number of hidden units in the first layer of GRU	25
Spatial attention weights	6
The number of hidden units in the second layer of GRU	25
Temporal attention weights	25

$$Y_i = \frac{X_i - X_m}{X_\sigma},$$
$$Y_i' = \frac{Y_i - Y_{\min}}{Y_{\max} - Y_{\min}}, \tag{5}$$

where X_i is the logging data, X_m and X_σ are the mean and variance of the logging data, respectively. Y_i is the standardized value; $Y_{\min}$ and $Y_{\max}$ are the minimum and maximum values of the standardized data, respectively. Y_i' is the normalized value.

(1) *STAGRU Fusion Network Training*. The mean square error (MSE) is used as the loss function of the network and the network proposed with the lowest loss error is selected.

(2) *Input Testing Data*. The testing set of the logging data is input to the STAGRU fusion network.

(3) *Network Evaluation*. The mean absolute error (MAE) and coefficient of determination (R^2) are used as the evaluation metrics of the network. The operation is calculated as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|(\tilde{y}_{ii} - y_i)|,$$
$$R^2 = \frac{\sum_{i=1}^{n}(\tilde{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{6}$$

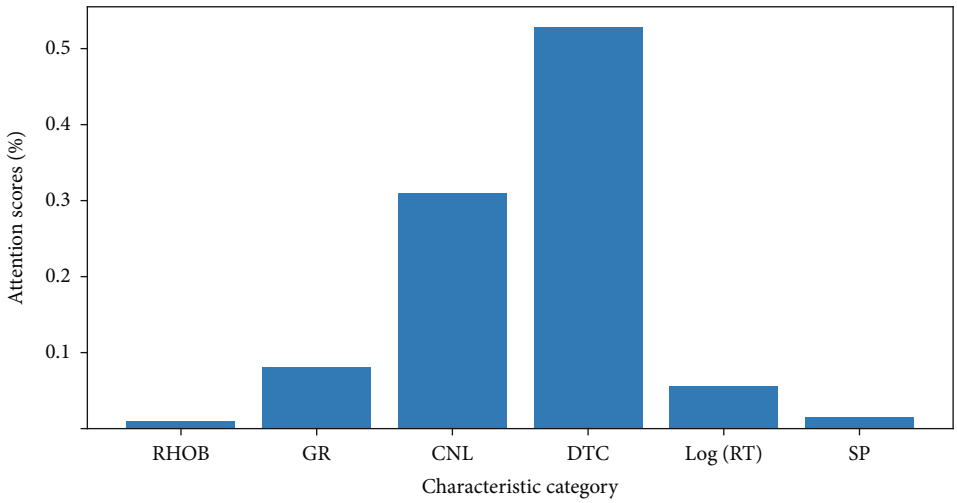

Figure 7: The weights of spatial characteristics in the spatial attention layer.

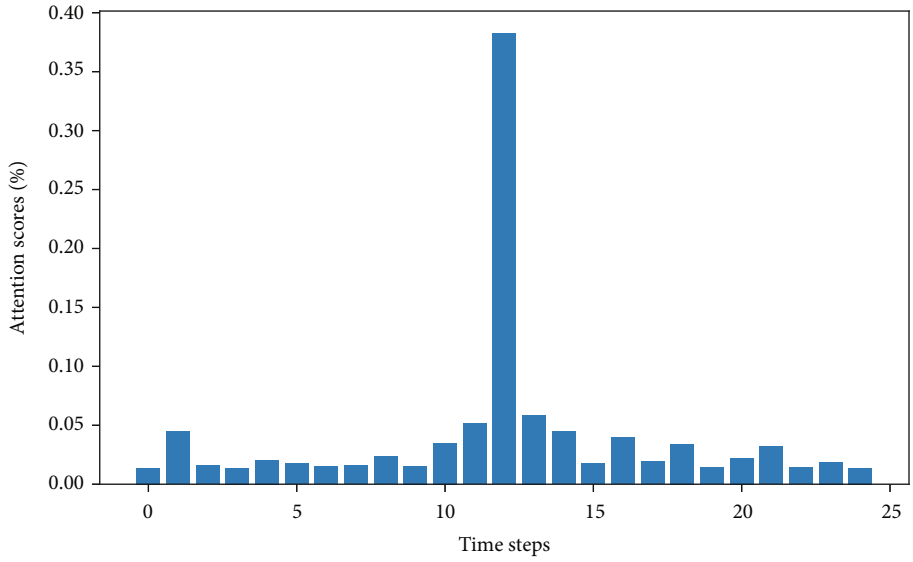

Figure 8: The weights of temporal characteristics in the temporal attention layer.

where y_i represents the real value, $\bar{y}$represents the mean of real value, $\tilde{y}_i$ represents the predicted value, and n is the number of samples.

3. Example Analysis

3.1. Dataset Introduction. The dataset used in this paper comes from the Jurassic Badaowan Formation in Junggar Basin, mainly composed of sandstone and mudstone. The reservoir in the area of deep burial, low porosity, low permeability, and complex pore structure is a typical unconventional tight hydrocarbon reservoir. The pore structure of commercial oil wells generally includes the fractures of horizontal, cross, and deformed bedding (Figure 4), which is the main cause of rock anisotropy. Therefore, the existing rock physics theory is difficult to accurately establish the rock physics models of tight sandstone reservoir. In order to improve the accuracy of the rock physics models and the prestack seismic inversion integrating wells with seismic data, a GRU fusion network based on the spatiotemporal attention mechanism is used to predict shear-wave velocity. There are 15 wells in the area, of which 5 wells contain the shear-wave velocity, which are marked as NY-1, NY-5, NY-9, NY-11, and NY-15, respectively. The logging data in NY-1, NY-5, and NY-9 wells is spliced to train the network, and the logging data in NY-11 and NY-15 wells is used to verify the generalization of the network.

3.2. Data Feature Selection. It is a typical regression problem that uses deep learning to solve the lack of shear-wave velocity. In theory, the prediction accuracy of using deep learning for regression problem is related to the correlation between input and output. Figure 5 shows the cross plot of shear-wave velocity and conventional logging data that the correlations from high to low are as follows: compression-wave velocity (VP), neutron porosity (CNL), resistivity (RT),

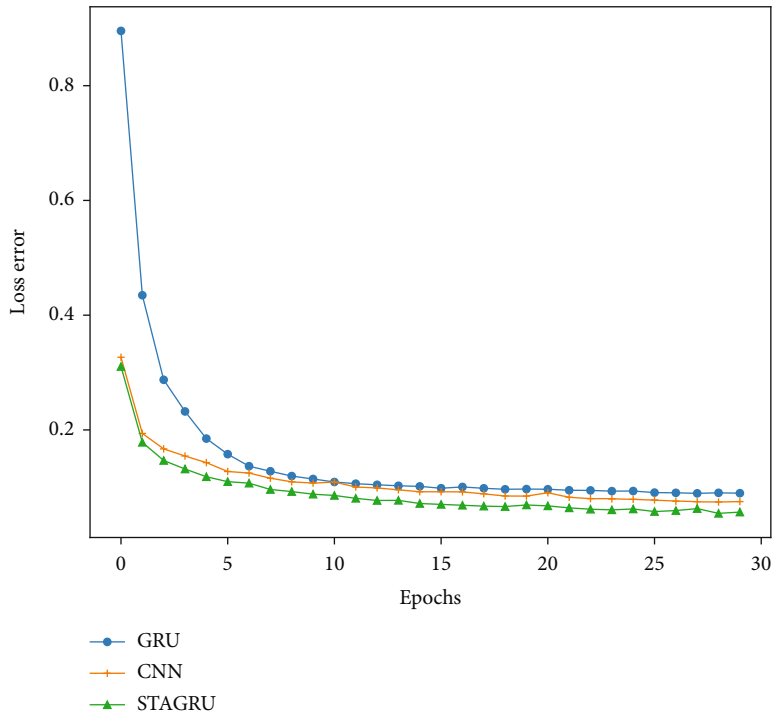

Figure 9: Loss error curves of the STAGRU, GRU, and CNN networks.

gamma (GR), natural potential (SP), and density (RHOB) with coefficients of 0.7266, 0.5932, 0.4093, 0.2710, 0.0027, and 0.0003, respectively. On the other hand, since the logging data has regularity in the sedimentary formation, the autocorrelation function (ACF) is used to analyze the autocorrelation of the conventional logging data. Figure 6 shows the logging data lags 300 that the autocorrelation of the logging data decreases continuously with the increase of lag. When the lag is 50, the autocorrelation from high to low is GR, VS, CNL, SP, VP, RHOB, and RT, respectively. The above analyses show that the spatiotemporal characteristics of the conventional logging data have a certain correlation with shear-wave velocity.

3.3. Interpretability of Attention Weights. In order to improve the sensitivity of the network to important spatiotemporal characteristics, a GRU fusion network based on the spatiotemporal attention mechanism is suggested and the parameters are shown in Table 1. Meanwhile, the weights of the spatial and temporal attention layer are analyzed in Figures 7 and 8, respectively.

Figure 7 shows the weight distribution of the spatial attention layer of the STAGRU fusion network. It can be seen that the weight distribution of the spatial attention layer from high to low is VP, CNL, RT, GR, SP, and RHOB, which is consistent with the distribution of the correlation coefficient in Figure 5. In the prediction of the shear-wave velocity of the STAGRU fusion network, the VP is assigned the largest weight ratio by the spatial attention layer, reaching 0.52, which means that the VP has the greatest influence on the shear-wave velocity. The reason for the above phenomenon is that compression-wave velocity and shear-wave velocity reflect the elastic information of the rock from different aspects, and the two are positively correlated, especially in the sedimentary formation; the correlation coefficient of them can reach more than 0.7, which verifies the rationality of adding the spatial attention mechanism in this study. This is the main reason that the empirical formula can better fit the shear-wave velocity according to the compression-wave velocity.

Figure 8 shows the weight distribution of the temporal attention layer of the STAGRU fusion network. The 25 sampling points in the conventional logging data are selected as a sample and the shear-wave velocity in the middle of the sample length is selected as the label. It can be seen that the different weights are assigned to the temporal characteristics of the logging data by the temporal attention layer. In the prediction of the shear-wave velocity of the STAGRU fusion network, the sample data at the label position is assigned the largest weight ratio by the temporal attention layer, reaching 0.3723, which means that it has the greatest impact on the shear-wave velocity. With the increasing distance on both sides of the label, the general trend of the attention weights is decreasing, which is consistent with the distribution of the autocorrelation of the conventional logging data in Figure 6. The reason for this phenomenon is that the mineral composition of the sedimentary formation is gradually changing, so the logging data has a certain autocorrelation in the depth direction, which verifies the rationality of adding the temporal attention mechanism in this study.

4. Network Comparative Analysis

To verify the performance of the STAGRU fusion network, the prediction results of the STAGRU, CNN, and GRU

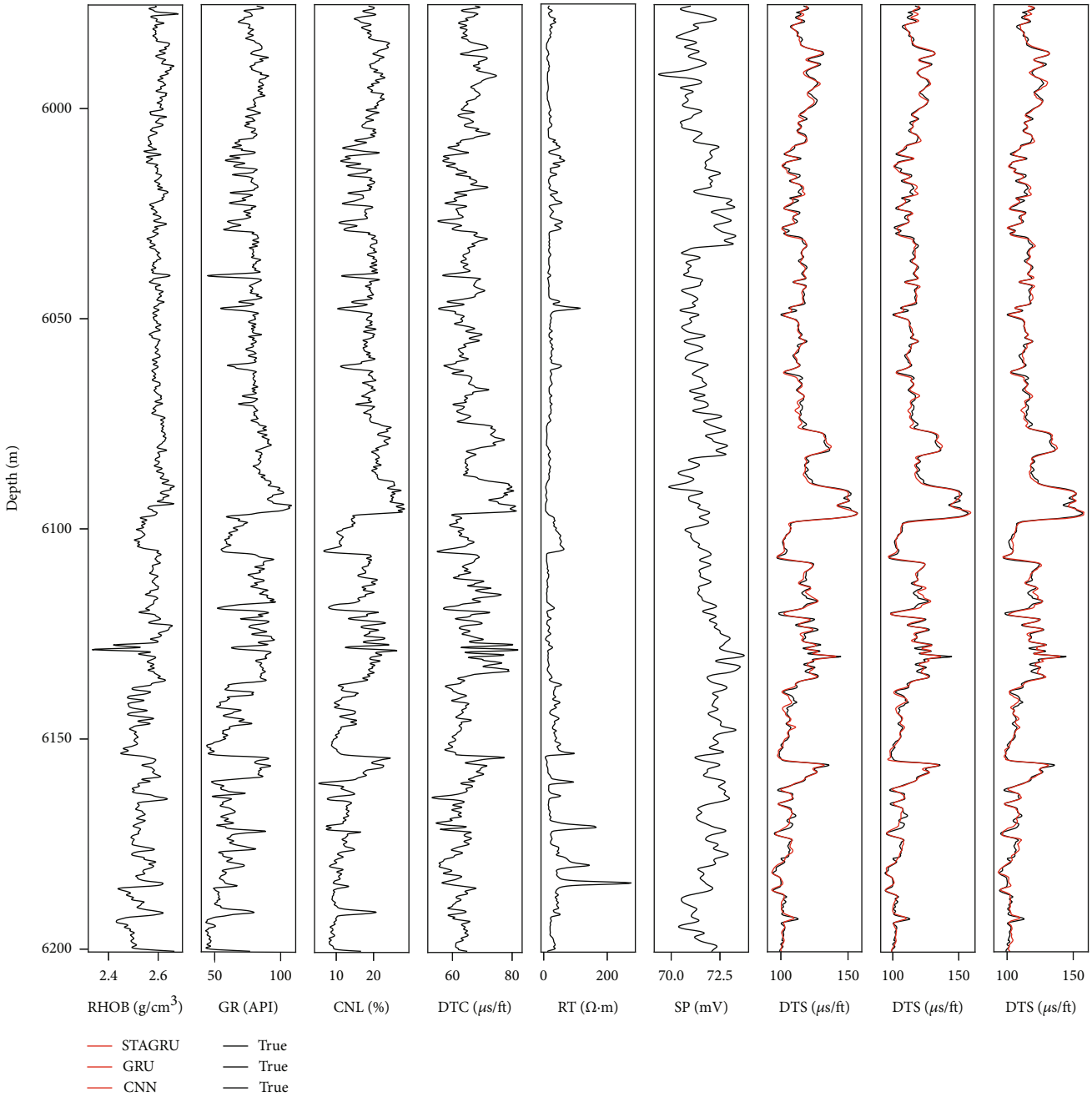

Figure 10: Prediction results of the STAGRU, GRU, and CNN networks in the training set.

networks are analyzed, respectively. All networks use adaptive moment estimation (Adam) as the optimization algorithm, which combines the advantages of the AdaGrad and RMSProp algorithms [42, 43], and has the strong advantages in handling large-scale data, parameter optimization, and nonstationary objectives. At the same time, the dropout layer is added to randomly discard neurons to increase the generalization of the network.

4.1. Training Set Analysis. The logging data in NY-1, NY-5, and NY-9 wells is used to train the CNN, GRU, and STAGRU networks. Figure 9 shows the loss errors of the three networks; it can be seen that the loss error decreases continuously with the increase of training times, and finally reaches a stable and constant value, which means that the network has reached the optimal state. However, the loss

Table 2: Comparison of the prediction results of the GRU, CNN, and STAGRU in training set.

Networks	MAE	R^2
GRU	0.154	0.970
CNN	0.149	0.972
STAGRU	0.135	0.978

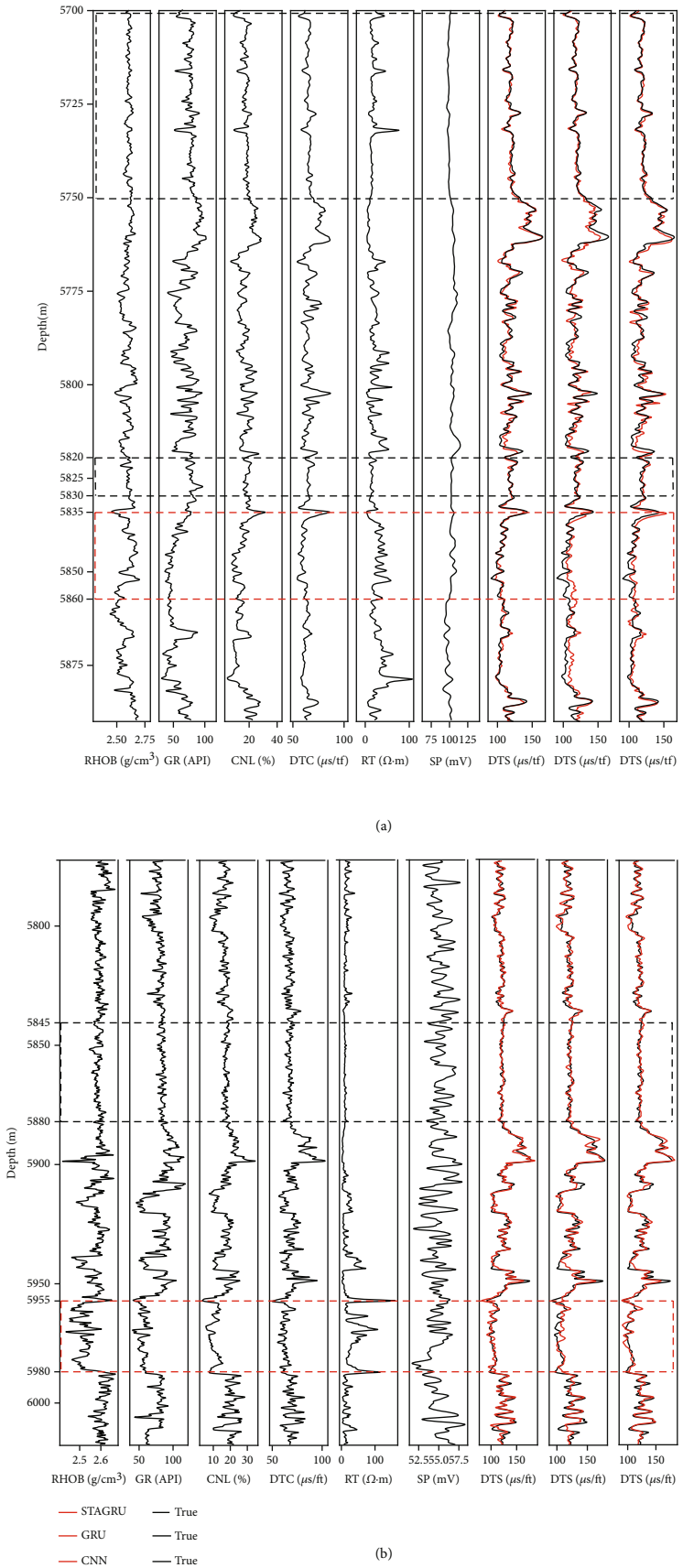

Figure 11: Prediction results of the STAGRU, GRU, and CNN networks in testing set. (a) NY-11 well; (b) NY-15 well.

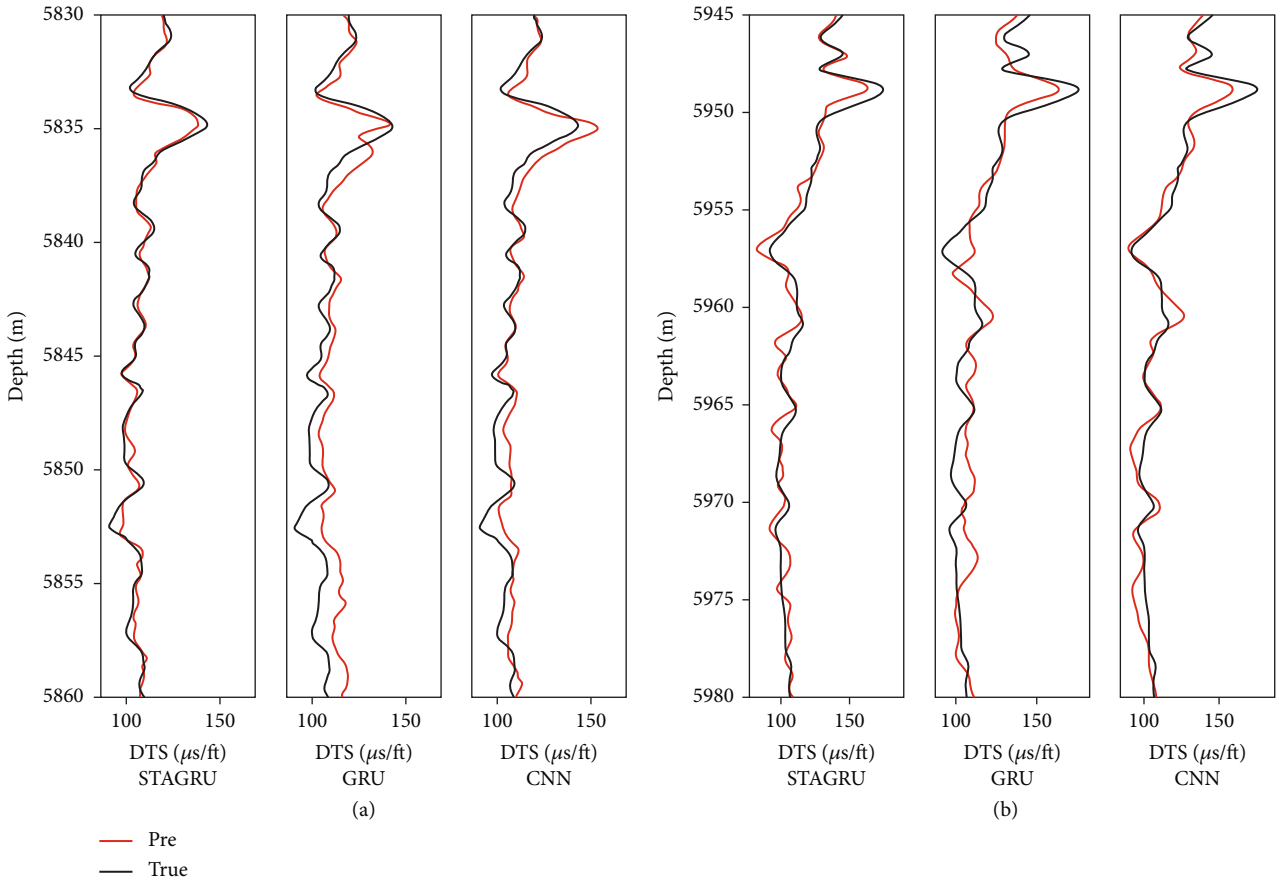

Figure 12: The local amplification of shear-wave velocity predictions of the STAGRU, GRU, and CNN networks in the sandstone formation. (a) NY-11 well; (b) NY-15 well.

error of the STAGRU fusion network is lower than that of the other two networks, indicating that the STAGRU fusion network can better mine the correlation between the spatiotemporal characteristics of the conventional logging data and the shear-wave velocity and improve the sensitivity of the network to important spatiotemporal characteristics. The predictions and evaluation results of the three networks in training set are shown in Figure 10 and Table 2, respectively. It can be seen that the prediction performance of the STAGRU fusion network is slightly higher than that of the GRU and CNN networks.

4.2. Testing Set Analysis. In order to better verify the prediction accuracy and generalization of the network proposed, the prediction results of the STAGRU, GRU, and CNN networks are analyzed in the NY-11 and NY-15 wells as shown in Figure 11. In the mudstone formation, the predicted values of the three networks are almost consistent with the real values, such as the logging data at depths from 5700 to 5750 m in the NY-11 well and from 5845 to 5880 m in the NY-15 well. However, in the sandstone formation, the prediction performance of the network proposed is higher than the other two networks, such as the logging data at depths from 5835 to 5860 m in the NY-11 well and from 5955 to 5980 m in the NY-15 well. It is concluded that the network proposed has a higher prediction accuracy from the local amplification of the STAGRU, CNN, and GRU network predictions in the formation of sandstone in the NY-11 and NY-15 wells (Figure 12). The prediction performance of the CNN is better than that of the GRU, but the predicted values of the STAGRU fusion network are almost consistent with the real values at depths from 5945 to 5950 m in the NY-15 well, which shows the network proposed can make up for the problems of information loss of the GRU and the poor global perception of the CNN. Table 3 shows the comparison results of the three networks that are quantitatively evaluated using the MAE and R^2. It can be seen that the MAE of the STAGRU fusion network is

Table 3: Comparison of the prediction results of the GRU, CNN, and STAGRU in testing set.

Well names	Networks	MAE	R^2
NY-11	GRU	0.225	0.824
	CNN	0.207	0.831
	STAGRU	0.180	0.917
NY-15	GRU	0.198	0.841
	CNN	0.188	0.860
	STAGRU	0.184	0.897

the lowest and the R^2 is the highest in the NY-11 and NY-15 wells. It can be seen from the evaluation results that the STAGRU fusion network has higher prediction accuracy and generalization.

5. Conclusion

Due to the complex pore structure of the tight sandstone reservoir in the Junggar Basin and the conventional network is insufficient sensitivity to important spatiotemporal characteristics, a GRU fusion network based on the spatiotemporal attention mechanism is developed in this paper. The result shows the weight distribution of the spatiotemporal attention layer is consistent with the autocorrelation of the conventional logging data, and the correlation between the conventional logging data and the shear-wave velocity, which verifies the network proposed, can improve the sensitivity of the network to important spatiotemporal characteristics and the rationality of adding the spatiotemporal attention mechanism. In addition, the test cases show that in the formation of sandstone in NY-11 and NY-15 wells, the R^2 of the STAGRU fusion network is 9.7% and 8.8% higher than that of the GRU and is 7.8% and 6.3% higher than that of the CNN, indicating that the network proposed has better prediction accuracy and generalization.

Data Availability

All data had been shown in the article.

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.

Authors' Contributions

Each author has contributed to the present paper. Tengfei Chen was responsible in drafting the article, programming, and in performing the experiments; Gang Gao was responsible in conceiving the method, directing the experiments, and revising the article; Yonggen Li was responsible for the interpretation of the data; Peng Wang was responsible for the analysis of the data; Bin Zhao supervised the experiments; Zhixian Gui was responsible in revising the article; and Xiaoyan Zhai was responsible for checking the figures.

Acknowledgments

This work is jointly supported by the State Key Program of National Natural Science Foundation of China (Grant No. 42030805) and the Scientific Research and Technology Development Project of China National Petroleum Corporation (Grant No. 2021DJ3704).

References

[1] X. Y. Yin, G. S. Cheng, and Z. Y. Zong, "Non-linear AVO inversion based on a novel exact PP reflection coefficient," *Journal of Applied Geophysics*, vol. 159, pp. 408–417, 2018.

[2] G. Y. Zhang, Z. Z. Wang, and C. Y. Lin, "Seismic reservoir prediction method based on wavelet transform and convolutional neural network and its application," *Journal of China University of Petroleum (Edition of Natural Science)*, vol. 44, no. 4, pp. 83–93, 2020.

[3] O. Jørgensen and D. Burns, "Novel finite-element approach applied to borehole quadrupole dispersion analysis in stress-sensitive formations," *Geophysics*, vol. 78, no. 6, pp. D499–D509, 2013.

[4] G. Gao, Z. H. He, and J. X. Cao, "The new two-term elastic impedance inversion and its application to predict deep gas-bearing carbonate reservoirs," *Oil Geophysical Prospecting*, vol. 48, no. 3, pp. 450–457, 2013.

[5] G. Gao, Y. H. Yang, B. Zhao, H. L. Duan, X. Y. Wang, and Y. Z. Wei, "A method for establishing and directly extracting sensitive identification factors of unconsolidated sandstones," *Oil Geophysical Prospecting*, vol. 54, no. 6, pp. 1329–1347, 2019.

[6] J. P. Castagna, M. L. Batzle, and R. L. Eastwood, "Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks," *Geophysics*, vol. 50, no. 4, pp. 571–581, 1985.

[7] M. L. Greenberg and J. P. Castagna, "Shear-wave velocity estimation in porous rocks: theoretical formulation, preliminary verification and applications1," *Geophysical Prospecting*, vol. 40, no. 2, pp. 195–209, 1992.

[8] G. T. Kuster and M. N. Toksöz, "Velocity and attenuation of seismic waves in two-phase media: part I. theoretical formulations," *Geophysics*, vol. 39, no. 5, pp. 587–606, 1974.

[9] S. Xu and R. E. White, "A new velocity model for clay-sand mixtures1," *Geophysical Prospecting.*, vol. 43, no. 1, pp. 91–118, 1995.

[10] Y. H. Yang, X. Yin, G. Gao, Z. X. Gui, and B. Zhao, "Shear-wave velocity estimation for calciferous sandy shale formation," *Journal of Geophysics and Engineering*, vol. 16, no. 1, pp. 105–115, 2019.

[11] D. S. Trigueros, L. Meng, and M. Hartnett, "Face recognition: from traditional to deep learning methods," 2018, https://arxiv.org/abs/1811.00116.

[12] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing [Review Article]," *ieee Computational intelligen Ce magazine*, vol. 13, no. 3, pp. 55–75, 2018.

[13] J. Hu, L. Shen, G. Sun, and S. Albanie, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, Salt Lake City, 2018.

[14] Z. Y. Wu, X. Zhang, C. L. Zhang, and H. Y. Wang, "Lithology identification based on L STM recurrent neural network," *Lithologic reservoirs*, vol. 33, no. 3, pp. 120–128, 2021.

[15] H. Song, W. Chen, M. J. Li, and H. Y. Wang, "A method to predict reservoir parameters based on convolutional neural network-gated recurrent unit (CNN-GRU)," *Petroleum Geology and Recovery Efficiency*, vol. 26, no. 5, pp. 73–78, 2019.

[16] D. Cao, Y. Su, and R. Cui, "Multi-parameter pre-stack seismic inversion based on deep learning with sparse reflection coefficient constraints," *Journal of Petroleum Science and Engineering*, vol. 209, article 109836, 2022.

[17] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[18] Y. LeCun, B. Boser, J. S. Denker et al., "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[19] Y. Zhang, C. L. Zhang, Q. Y. Ma, X. Zhang, and H. Zhou, "Automatic prediction of shear wave velocity using convolutional neural networks for different reservoirs in Ordos Basin," *Journal of Petroleum Science and Engineering*, vol. 208, article 109252, 2022.

[20] Z. Zhang, S. Hu, K. Zhang, Q. Y. Ma, and Y. N. Jiang, "Shear wave velocity prediction of complex volcanic reservoirs based on deep learning," in *International Conference on Computational Modeling, Simulation, and Data Analysis (CMSDA 2021)*, vol. 12160, pp. 553–558, Sanya, China, 2022.

[21] Y. H. Sun, "Prediction of S-wave velocity based on GRU neural network," *Oil Geophysical Prospecting*, vol. 55, no. 3, 2020.

[22] Q. Y. Ma, X. Zhang, C. L. Zhang, H. Zhou, and Z. Y. Wu, "Shear-wave velocity prediction based on one-dimensional convolutional neural net-work," *Lithologic Reservoirs*, vol. 33, no. 4, pp. 111–120, 2021.

[23] H. Zhou, Z. Y. Wu, X. Zhang, C. L. Zhang, and Q. Y. Ma, "Shear wave prediction method based on LSTM recurrent neural network," *Fault-Block Oil& Gas Field*, vol. 28, no. 6, pp. 829–834, 2021.

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

[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[26] K. Cho, B. V. Merriënboer, D. Bahdanau, and Y. H. Bengio, "On the properties of neural machine translation: encoder-decoder approaches," 2014, https://arxiv.org/abs/1409.1259.

[27] J. Wang, J. Cao, J. You, J. Liu, and X. Zhou, "Prediction of reservoir porosity, permeability, and saturation based on a gated recurrent unit neural network," *Geophysical Prospecting for Petroleum*, vol. 59, no. 4, pp. 616–627, 2020.

[28] J. Wang and J. X. Cao, "Data-driven S-wave velocity prediction method via a deep-learning-based deep convolutional gated recurrent unit fusion network," *Geophysics*, vol. 86, no. 6, pp. M185–M196, 2021.

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

[30] T. F. Chen, G. Gao, P. Wang, B. Zhao, Y. G. Li, and Z. X. Gui, "Prediction of shear wave velocity based on a hybrid network of two-dimensional convolutional neural network and gated recurrent unit," *Geofluids*, vol. 2022, Article ID 9974157, 14 pages, 2022.

[31] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, p. 30, 2017.

[32] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, https://arxiv.org/abs/1409.0473.

[33] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2018, https://arxiv.org/abs/1810.04805.

[34] H. Choi, K. Cho, and Y. Bengio, "Fine-grained attention mechanism for neural machine translation," *Neurocomputing*, vol. 284, pp. 171–176, 2018.

[35] J. Du, Y. Cheng, Q. Zhou, J. Zhang, X. Zhang, and G. Li, "Power load forecasting using BiLSTM-attention," *IOP Conference Series: Earth and Environmental Science*, vol. 440, no. 3, article 032115, 2020.

[36] L. Yang, S. Wang, X. Chen et al., "High-fidelity permeability and porosity prediction using deep learning with the self-attention mechanism," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, 2022.

[37] P. Kavianpour, M. Kavianpour, E. Jahani, and A. Ramezani, "A CNN-BiLSTM model with attention mechanism for earthquake prediction," 2021, https://arxiv.org/abs/2112.13444.

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

[39] T. Perol, M. Gharbi, and M. Denolle, "Convolutional neural net-work for earthquake detection and location," *Science Advances*, vol. 4, no. 2, article e1700578, 2018.

[40] L. Q. Shan, Y. C. Liu, M. Tang, M. Yang, and X. Y. Bai, "CNN-BiLSTM hybrid neural networks with attention mechanism for well log prediction," *Journal of Petroleum Science and Engineering*, vol. 205, article 108838, 2021.

[41] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.

[42] H. Saito and M. Kato, "Machine learning technique to find quantum many-body ground states of bosons on a lattice," *Journal of the Physical Society of Japan*, vol. 87, no. 1, article 014001, 2018.

[43] L. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Transactions on Medical Imaging*, vol. 36, no. 4, pp. 994–1004, 2017.