ACCSCIENCE PUBLISHING

Journal of Seismic Exploration

ARTICLE

A precise picking method for seismic first arrivals based on the residual long short-term memory network driven by time-frequency dual domain data

Ziyu Qin[1], Xianju Zheng[1]*, and Wenhua Wang[2]

[1]Department of Software Engineering, School of Computer Engineering, Chengdu Technological University, Chengdu, Sichuan, China

[2]Department of Intelligence Science and Technology, School of Computer Science, Chengdu Normal University, Chengdu, Sichuan, China

Abstract

First-arrival picking of seismic data is one of the key steps in seismic data processing. When seismic data have low signal-to-noise ratio (SNR) and weak first-arrival energy, accurately and efficiently picking first arrivals remain a critical challenge for most automatic picking methods. To address this issue, this paper proposes a Multi-perspective Residual Long Short-Term Memory (M-Res-LSTM) network. This network integrates the spatial feature extraction advantage of Residual Networks and the temporal dynamic modeling capability of LSTM networks, while introducing a coordinate attention mechanism. Through multi-perspective learning in both time and frequency domains, it effectively improves the reliability of automatic first-arrival picking. First, this paper elaborates on the core principle of the M-Res-LSTM network for automatic first-arrival picking: the amplitude, frequency, and phase features of seismic data are used as network inputs, and the accurately picked first arrivals manually serve as network outputs. After training the network using a supervised learning approach, the well-trained model is applied to perform automatic first-arrival picking. Second, an analysis of the network's hyperparameters is conducted to determine the optimal parameter configuration. Finally, automatic first-arrival picking tests are carried out on seismic datasets with different characteristics, and the picking results are compared with those obtained by the energy ratio method, single-input Res-LSTM, and Swin-Transformer. The results demonstrate that the proposed M-Res-LSTM method maintains good stability and accuracy even in complex scenarios with low first-arrival energy and poor SNR.

Keywords: Automatic first-arrival picking; Time-frequency dual domain; Multi-perspective learning; Res-LSTM; Attention mechanism

***Corresponding author:**
Xianju Zheng
(zhengxju@cdtu.edu.cn)

Citation: Qin Z, Zheng X, Wang W. A precise picking method for seismic first arrivals based on the residual long short-term memory network driven by time-frequency dual domain data. *J Seismic Explor*. doi: 10.36922/JSE025330058

Received: August 16, 2025

1st revised: September 1, 2025

2nd revised: September 8, 2025

Accepted: October 29, 2025

Published online: November 18, 2025

1. Introduction

First-arrival waves refer to the seismic waves that propagate through subsurface media and reach geophones first, typically existing in the form of direct waves or refracted waves. In the seismic data processing workflow, the travel time of first-arrival waves is

of crucial significance, as it can provide core foundational data for near-surface inversion work. Although the method of manual picking of the first arrival has high accuracy, it is time-consuming and labor-intensive. Moreover, the accuracy of the picking is also affected by the experience of the interpreters.[1] To pick the first arrivals efficiently and accurately and reduce the workload of data processing personnel, scholars have proposed different semi-automatic or automatic first-arrival picking methods.

Given the similar characteristics of adjacent traces, existing studies have proposed a method to determine the first-arrival time through cross-correlation operations between adjacent traces.[2-4] The selection of the standard trace has a great influence on the first-arrival picking results of this type of method, and its effect needs to be further improved when the signal-to-noise ratio (SNR) of the data is low. The algorithm based on energy characteristics possesses robust anti-noise performance and achieves favorable automatic picking results, and has also achieved good results in the processing of actual data.[5-8] Since this method is greatly affected by the window length, many picking methods based on multiple time windows have also been developed.[9,10] Another common method is the Akaike Information Criterion (AIC), which discriminates the first arrivals using the difference in AIC values between the seismic signal and the noise.[11-13] However, the picking results are not satisfactory under a low SNR. The fractal dimension algorithm[14-17] has relatively good anti-noise ability, but it does not take into account the similarity between seismic traces. Clustering-based methods identify first arrivals in an unsupervised manner based on the characteristic differences between first arrivals and noise. However, the inherent temporal connections between different subsequences are not considered, making it difficult to distinguish low-amplitude signals from noise under low SNR conditions.[18,19]

Transforming seismic signals into other domains or spaces can further highlight the differences between first-arrival waves and noises. Performing τ-p transformation,[20] wavelet transformation,[21] shearlet transformation,[22-24] etc., are also commonly used means to enhance the precision of first-arrival picking; Beyond the common shot gather, picking first arrivals on common offset gathers and common receiver gathers also offers distinct advantages.[5,25] In addition to time-space domain signals, neural networks incorporate attributes such as amplitude, frequency, time-frequency characteristics, short-term average/long-term average (STA/LTA) ratios, and data distribution skewness as inputs,[26-31] thereby facilitating more effective capture of relevant features by the networks.

Deep learning can efficiently extract the internal laws of data, construct multi-level data representations, and is more friendly to massive data. Leveraging the capabilities of generative adversarial networks,[32] convolutional neural networks,[1,33] UNet,[34-37] recurrent neural networks, meta-learning,[38,39] transformers,[40-42] transfer learning,[43] and various hybrid networks[38,44] in extracting complex features from seismic data, deep learning algorithms have emerged as a critical force in the picking of seismic first arrivals. Similar to conventional methods, to further improve the picking accuracy, scholars have attempted to input data with different attributes into the network, such as time-frequency domain data and STA/LTA feature maps.[44,45] The rich data features provide more information references for the model.

Studies indicate that current first-arrival picking operations are typically performed exclusively within either the time-space domain or a single transformed domain. The energy features and time-frequency features of seismic data serve as valid criteria for first-arrival identification, and the integration of multiple sets of feature data can, further, enhance the precision of first-arrival extraction. Building on this insight, this study comprehensively leverages the data features from both the time domain and frequency domain, proposes a Multi-perspective Residual Long Short-Term Memory (M-Res-LSTM) network tailored for first-arrival picking, and elaborates in detail on the complete workflow of automatic first-arrival picking. Finally, experiments on automatic first-arrival picking were conducted on real seismic datasets with distinct characteristics using this network, thereby verifying the effectiveness of the proposed method.

2. Methods

2.1. Network architecture

To fully leverage the spatiotemporal and time-frequency characteristics of seismic signals, this study proposes the M-Res-LSTM network for automatic first-arrival picking. Figure 1 shows its structural schematic with key features.

2.1.1. Branch structure

The network comprises four parallel branch modules, each consisting of m residual modules and n residual LSTM modules. These branches process four types of two-dimensional (2D) input data, namely, time-space domain seismic signals, instantaneous amplitude, frequency, and phase features, thereby enabling multi-perspective extraction of seismic information. A parameter-sharing mechanism between LSTM branches enhances training efficiency and generalization.

Residual modules retain spatial information through shortcut paths, mitigating degradation in deep networks through direct gradient flow.[46] For seismic data, this

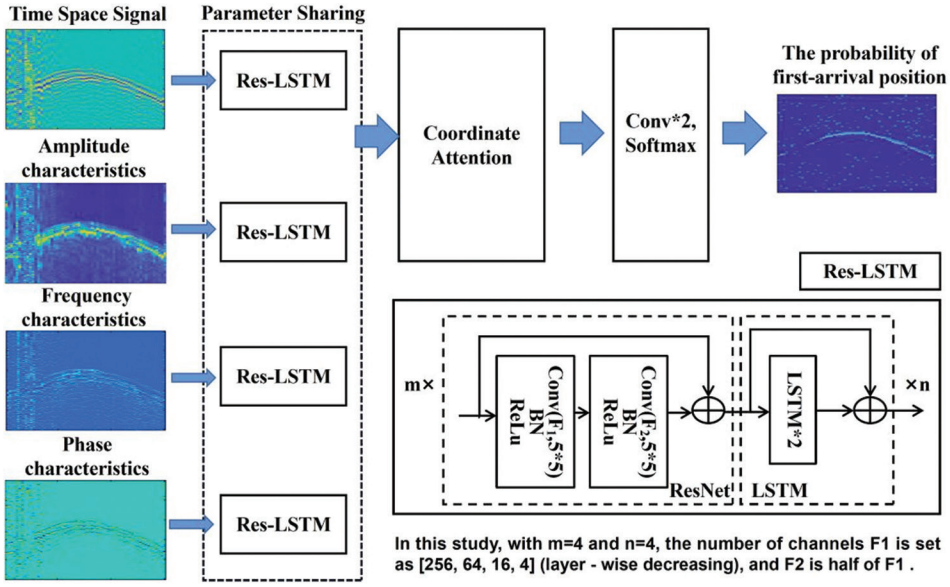

Figure 1. Schematic diagram of the M-Res-LSTM network model structure. "m" and "n" represent the numbers of Residual Networks and LSTM modules, respectively.
Abbreviation: LSTM: Long Short-Term Memory; M-Res-LSTM: Multi-perspective Residual Long Short-Term Memory; Res-LSTM: Residual Long Short-Term Memory.

preserves inter-trace correlations and nonlinear spatial features. For input data $X_k \in R_{H\times W}, k=1,2,3,4$, the output of the *i*-th residual layer is:

$$Z_i^{res} = X_k + F_i\left(X_k, \theta_{res}^i\right), i=1,2\cdots m \quad \text{(I)}$$

Where F_i denotes nonlinear transformations (convolution, batch normalization, and activation) with parameters θ_{res}^i .

Following residual processing, outputs are converted to sequences for residual LSTM modules. LSTM's gating mechanisms (input, forget, and output) capture temporal dynamics. For *n* residual LSTM layers, the recurrence relation is:

$$h_j^t = x_{j-1}^t + LSTM\left(x_t^j, x_j^t, \phi_j\right), j=1,2\cdots n \quad \text{(II)}$$

where h_j^t is the hidden state of the *j*-th layer at time *t*, ϕ_j are layer parameters, and the *n*-th layer output h_n^t represents temporal features, which are reshaped to 2D for fusion.

2.1.2. Coordinate attention module

To adapt to the requirements of the first-arrival picking task, multi-domain features extracted from the four branches are fused through the coordinate attention mechanism.[47] Unlike the standard coordinate attention, the proposed method first extracts multi-dimensional visual features tailored to the task characteristics, then assigns adaptive weights to enhance critical information. Meanwhile, it not only additionally designs a convolutional gating structure specifically for suppressing high-amplitude noise in seismic data but also further integrates a residual connection to prevent first-arrival signals from being excessively suppressed. The detailed process is as follows:

First, global pooling is performed on the concatenated feature $F \in R_{H\times W\times C}$ along the width (W) and height (H) directions:

$$z_h(k) = \frac{1}{W}\sum_{0\le j<W} F(h,i,k), z_w(\omega) = \frac{1}{H}\sum_{0\le j<H} F(j,\omega,k) \quad \text{(III)}$$

After aligning the dimensions of x_w with x_h through a transposition operation, the two are concatenated along the height direction. The concatenated result is processed by a custom activation function and 1×1 convolution to obtain the feature *m*. The processed feature *m* is split into height-related and width-related components. Meanwhile, a convolutional gating structure is designed based on the original branch features to generate a screening mask g_i. Finally, branch attention weights are generated and fused:

$$branch\_att = \frac{a_h + a_w}{2} \odot g_i \quad \text{(IV)}$$

Here, a_h and a_w are the branch attention weights, respectively; and e denotes element-wise multiplication.

Each branch feature is multiplied by its corresponding attention weight, and the weighted branch features are summed to obtain a fused representation:

$$F_{fused} = \sum_{i=1}^{4} \left(b_i \bullet branch\_att_i \right) \quad \text{(V)}$$

Where b_i is the feature of the i-th branch, and $branch\_att_i$ is its corresponding attention weight.

This process not only retains the feature advantages of each branch in specific domains but also mitigates the interference of redundant information and noise through weight modulation.

To further focus on regions with concentrated first-arrival wave energy, coordinate attention computation is re-applied to F_{fused}: first, global average pooling is performed along the width and height directions, with the pooling formulas as follows:

$$z'_h(k) = \frac{1}{W} \sum_{0 \le j < W} F_{fused}(h,i,k), z'_w(\omega) = \frac{1}{H} \sum_{0 \le j < H} F_{fused}(j,\omega,k) \quad \text{(VI)}$$

After dimension alignment, feature concatenation, channel compression, and component splitting, the base spatial attention weights c_h and c_w are obtained. Concurrently, a convolutional gating structure is designed based on F_{fused} to generate the spatial screening mask g_s. The spatially attentive weights with enhanced noise robustness are derived through the following formulas:

$$\alpha = c_h \odot g_s, \beta = c_w \odot g_s \quad \text{(VII)}$$

Finally, the spatial attention weights are combined with F_{fused} through residual connection to enhance the signals in key regions. The calculation formula for the final output feature $\hat{F}$ is as follows:

$$\check{F}(h,\omega,k) = F_{fused}(h,i,k) + F_{fused}(h,i,k) \cdot \alpha(h,k) \cdot \beta(\omega,k) \quad \text{(VIII)}$$

By generating screening masks with the same dimension as the attention weights through convolutional gating, this method can effectively suppress high-amplitude noise in seismic data, accurately focus on regions with significant first-arrival wave energy variations, and remarkably improve the processing performance of seismic data with low SNR.

2.1.3. Output layer

Fused features $\hat{F}$ are passed through two convolution layers and a Softmax activation to predict first-arrival positions:

$$P = Soft\max\left(Conv_2\left(Conv_1\left(\hat{F}, \omega_1 \right), \omega_2 \right) \right) \quad \text{(IX)}$$

Where ω_1, ω_2 are convolution parameters.

The network takes the amplitude, frequency, and phase characteristics of seismic data as inputs and incorporates a coordinate attention mechanism to achieve feature fusion across different branches. By assigning distinct weights to multiple features, it allocates varying levels of attention to them, thereby enhancing task-critical features while suppressing those irrelevant to the current task. This mechanism effectively mitigates the mispicking of first arrivals.

In recent studies, a multitude of innovative networks have been proposed for seismic first-arrival picking, including those based on Transformer,[40] meta-learning,[38] and multi-stage network[45] architectures. Table 1 presents a comparison between the method proposed in this study and the aforementioned methods, focusing on their characteristics including network architecture, input, advantages, and dependency conditions.

Compared with other networks, M-Res-LSTM still possesses unique characteristics and advantages:

(i) In terms of feature input types, M-Res-LSTM innovatively incorporates amplitude, frequency, and phase information, providing more comprehensive feature support for first-arrival picking

(ii) M-Res-LSTM introduces the coordinate attention mechanism, which includes a time-frequency domain weight allocation layer. By calculating the weights of time-domain and frequency-domain features, it can more accurately capture key information in the time-frequency domain and improve the ability to identify weak first-arrival signals and the like

(iii) In terms of feature fusion methods, M-Res-LSTM employs parallel branches and adaptive weight fusion, enabling collaborative optimization of multi-domain features such as time and frequency, thus enhancing the effectiveness and flexibility of feature fusion.

2.2. Dataset and training

To apply the M-Res-LSTM network for automatic first-arrival picking, seismic data are first transformed to generate profiles containing amplitude, phase, and frequency features. Subsequently, these feature profiles are partitioned into three subsets: the training set, validation set, and test set, with the respective proportions accounting for 80%, 10%, and 10% of the total dataset, respectively. Finally, the network is trained using the training set, the optimal generalization of the model is achieved on the

Table 1. Comparison of different network features

Comparison	M-Res-LSTM	Res-LSTM	Swin-Transformer[40]	Meta-learning[38]	MSSPN[45]
Architecture	Multi-branch parallel architecture and realizing dynamic fusion of time-frequency dual-domain features via the coordinate attention mechanism	Single-branch cascaded architecture	U-shaped hierarchical self-attention architecture	Dual-loop meta-training framework	Four-stage cascaded segmentation architecture
Input	Spatiotemporal signals, amplitude, frequency, phase	Spatiotemporal domain signals	Spatiotemporal domain signals	Spatiotemporal domain signals	Spatiotemporal domain signals and STA/LTA feature maps
Advantages	Multi-input supports the attention mechanism for accurate noise suppression; residual-temporal modeling adapts to seismic wave propagation	Concise architecture, low computational overhead, and easy reproduction	SW-MSA adapts to drastic changes in local first arrivals; Dilated convolution expands the receptive field	Only 5–20% of manual labels are required, reducing costs; weight matrix filters label noise	VCTE effectively narrows the first-arrival range; mixed loss enhances the continuity of first arrivals
Dependency Conditions	High-quality labels	High-quality labels	High-quality labels	Allowing partial low-quality labels	Requiring prior information and high-quality labels

Abbreviations: LTA: Long-term average; LSTM: Long-Short-Term Memory; M-Res-LSTM: Multi-perspective Residual Long-Short-Term Memory; MSSPN: Multistage segmentation picking network; Res-LSTM: Residual Long-Short Term Memory; SW-MSA: Shifted window-multihead self-attention; STA: Short-term average; VCTE: Velocity-constrained trend estimation.

validation set, and the first-arrival picking test is carried out on the test set. In this study, common shot gather records are selected as the original seismic data. Theoretically, common receiver gather records, common offset gather records, and common midpoint gather records are all optional. The main steps of the model training process are described as follows.

2.2.1. Generation of feature profiles

The three instantaneous profiles calculated using the Hilbert transform can reflect various characteristics of seismic data. The analytic signal of a seismic signal can be expressed as:[48]

$$\tilde{x}(t) = x(t) + ix^{*}(t) \tag{X}$$

Where $x(t)$ denotes the real component of the complex trace, corresponding to the actual seismic record trace; $x^{*}(t)$ represents the imaginary component of the complex trace, which is orthogonal to the real component.

Instantaneous amplitude reflects the energy intensity of a seismic signal at a certain moment, which is usually related to the reflection coefficient of the stratum. A high instantaneous amplitude may indicate lithologic abrupt changes or the presence of fluids.[49] Instantaneous amplitude can be expressed as:

$$A(t) = \sqrt{x^{2}(t) + x^{*2}(t)} \tag{XI}$$

Instantaneous phase describes the phase state of a signal, which can identify stratal continuity and structural features. Abrupt changes may indicate faults or unconformities:

$$\theta(t) = tg^{-1}\left[x^{*}(t) / x(t)\right] \tag{XII}$$

Instantaneous frequency is the time rate of change of the phase and can assist in identifying lithology and predicting reservoirs:

$$\omega(t) = d\theta(t) / dt \tag{XIII}$$

In the first-arrival picking process, the energy mutation points of instantaneous amplitude facilitate the identification of the first-arrival wavefront; the jump characteristics of instantaneous phase can enhance the identification of interfaces at the first-arrival time; and the high-frequency concentration characteristics of instantaneous frequency help distinguish signal differences between first-arrival waves and subsequent reflected waves, thereby providing abundant information for improving the accuracy of first-arrival picking.

Since first arrivals correspond to the high-amplitude signals first received by geophones in seismic records, amplitude serves as their primary characteristic. However, seismic data may contain noise with frequency and phase similar to those of effective signals. To suppress such noise, it is necessary to constrain the instantaneous frequency and instantaneous phase through amplitude (*i.e.*, seismic data) to obtain frequency and phase characteristic data.

2.2.2. Data normalization and cropping

Seismic data and feature data exhibit differences in dimensions and numerical magnitudes, so it is necessary to normalize all types of data to enable the network to better capture the first-arrival features. In this study, the maximum absolute value normalization method is adopted, defined as follows:

$$x = x / \max(abs(x)) \quad \text{(XIV)}$$

Before the network training, the seismic data are segmented into a size of 256 × 256, so as to speed up the training process and eliminate unnecessary information.

2.2.3. Network training

Typically, identifying the actual location of the first arrivals (i.e., the onset) poses a challenge. Therefore, in practical processing, a fixed phase (such as a wave crest or a wave trough) is selected as the first-arrival position.[5] In this study, the network treats first-arrival picking as a binary segmentation task. During sample preparation, the first peak value of the first-arrival wave is taken as the first-arrival position and labeled as 1 (the first category), while positions corresponding to all other time points are labeled as 0 (the second category). Accordingly, the network employs a binary cross-entropy loss function to quantify the discrepancy between the network output and the labels for classification purposes. The loss is defined as the sum of the losses of all pixels in the training samples within a mini-batch, expressed as follows:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} y_i \log\left(p\left(y_i\right)\right) + \left(1-y\right)\log\left(1-p\left(y_i\right)\right) \quad \text{(XV)}$$

Where y stands for the binary label (either 0 or 1), while $p(y)$ denotes the probability that the output corresponds to label y.

The network undergoes training through the Adaptive Moment Estimation (Adam) algorithm,[50] combined with the back-propagation approach. When the validation set loss decreases and stays stable over a certain number of iteration cycles, training halts, and the weights are saved as training results. The saved optimal weights are then applied to predict the first arrivals in the test set according to a specific formula:

$$y_{test} = \Phi\left(\theta_{opt}, x_{test}\right) \quad \text{(XVI)}$$

With Φ representing the function of the network.

2.2.4. Performance evaluation

Pixel accuracy serves as a widely used indicator in semantic segmentation,[51] defined as the proportion of correctly labeled pixels relative to the total number of pixels:

$$PA = \sum_{i=0}^{k} p_{ii} / \sum_{i=0}^{k}\sum_{j=0}^{k} p_{ij} \quad \text{(XVII)}$$

Where p_{ii} represents the quantity of pixels for which class i is inferred as class i, and $\sum_{i=0}^{k}\sum_{j=0}^{k} p_{ij}$ represents all the pixel points.

After processing by the deep learning network, the probability of a point being a first arrival ranges from 0 to 1 (with a maximum of 1 and a minimum of 0). To determine the first-arrival position, thresholding is first applied to all seismic traces: if a trace contains no points with probability exceeding the threshold, the entire trace is discarded. For traces containing points with probability exceeding the threshold, the position corresponding to the original maximum probability is designated as the first-arrival position. Verified through tests on multiple datasets with distinct characteristics, the model achieves optimal overall performance when the threshold is set to 0.4.

2.2.5. Hyperparameter analysis

The first-arrival picking results of the M-Res-LSTM network are affected by hyperparameters. Therefore, during the process of training the model, we conducted experiments on some parameters, including the learning rate, the size of the kernel matrix, batch size, and the network depth. During the experiments, only the parameter being tested was changed while other parameters remained the same, and the optimal parameters were determined according to the pixel accuracy value of the validation set. The test results are shown in Figure 2.

Through the experiments, it can be seen that an overly large learning rate will make it difficult for the network to converge, and the phenomenon of back-and-forth oscillation will occur. For this data, a learning rate of 0.005 has the best effect, as shown in Figure 2A; as illustrated in Figure 2B, when the kernel matrix size is 5 × 5 of Residual Networks (ResNet), the model achieves the highest accuracy along with a fast convergence rate; Figure 2C shows that a larger batch size results in better generalization performance, though it accordingly demands more computation time and memory capacity. Using the network structures of 4×ResNet + 4×lstm and 5×ResNet + 5×lstm can both achieve relatively good accuracy, but a deeper network means that more memory will be occupied (Figure 2D). Table 2 shows the

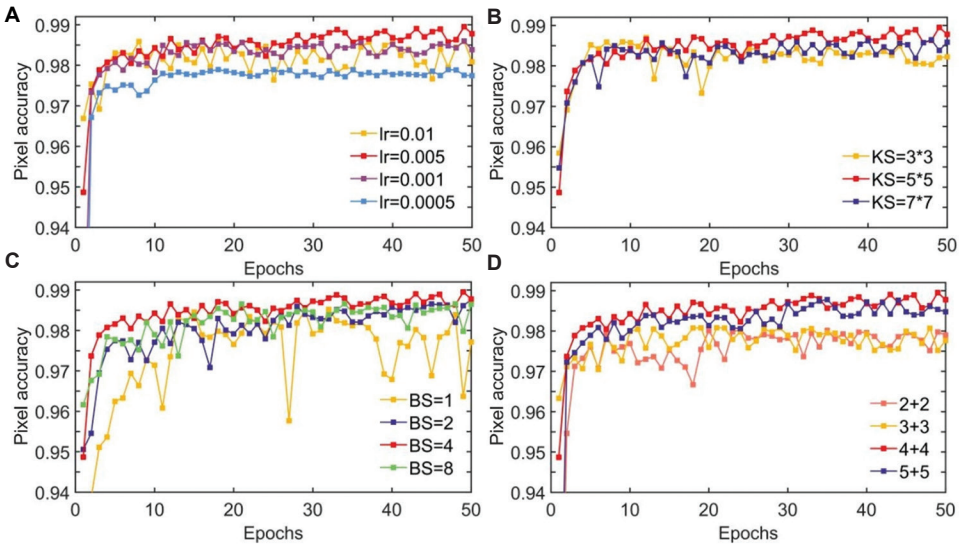

Figure 2. Pixel accuracy curves corresponding to various hyperparameters. (A) Curves of accuracy under different learning rates. (B) Accuracy curves with varying kernel matrix sizes. (C) Accuracy curves for different batch sizes. (D) Accuracy curves across different model depths.

Table 2. Experimental results of hyperparameters at the 50th epoch

Hyperparameters	Variants	PA
Learning rate	0.01	0.981
	0.005	**0.987**
	0.001	0.984
	0.0005	0.977
Batch size	1	0.976
	2	0.987
	4	**0.987**
	8	0.986
Kernel size	(3,3)	0.982
	(5,5)	**0.987**
	(7,7)	0.985
Network depth	2+2	0.979
	3+3	0.976
	4+4	**0.987**
	5+5	0.984

Notes: Values in boldface represent the values corresponding to the optimal performance for each hyperparameter.
Abbreviation: PA: Pixel accuracy.

pixel accuracy values obtained with different parameters when 50 epochs of iteration are carried out. Considering comprehensively the changing trend of the pixel accuracy value with the epoch (Figure 2) and the final accuracy (Table 2), the network will be trained with the parameters of kernel size = 5×5, learning rate = 0.005, batch size = 4, and 4×ResNet + 4×lstm. This parameter combination is expected to obtain the optimal convergence effect.

Table 3 elaborates on the parameter configurations of each module within the M-Res-LSTM network, including the input/output dimensions, channel numbers, and kernel sizes for submodules such as ResNet, LSTM, and coordinate attention. It provides a detailed technical blueprint for the network's architecture, with a total parameter quantity of 5.73 million, thus enabling the network to efficiently tackle seismic data processing tasks.

3. Results and discussion

We utilized the M-Res-LSTM network to perform first-arrival picking on three sets of real seismic data with distinct characteristics, and compared its prediction results with those from manual picking and the traditional energy ratio method. All the training was conducted on a single NVIDIA GeForce GTX 1080 Ti GPU, using the TensorFlow framework.

3.1. Data 1

Data 1 consists of small-scale 3D seismic data acquired in a plain area using dynamite sources. For each shot, 10 receiver arrays were designed, with 60 geophones deployed in each array, and the maximum offset is 1200 m. Each trace of the acquired data contains 501 sampling points, with a sampling interval set to 4 ms, and the effective recording duration of each trace is 2 s. The work area features hilly terrain, with surface elevation varying in the range of 92–160 m. Due to the limited coverage range of a single shot, the impact of topographic relief is relatively minor, and the first arrivals of seismic waves exhibit an overall smooth characteristic, providing a favorable foundation for first-arrival picking. It should be specifically noted

that the acquired seismic data contain strong industrial electrical interference and mechanical interference, and such interference signals have exerted a significant impact on first-arrival picking for some seismic traces.

The results of the energy ratio method (blue circles), manual picking (green triangles), and M-Res-LSTM (red triangles) are displayed on a representative single shot record, as shown in Figure 3. When the seismic trace contains clear first arrivals that can be identified manually, the results of M-Res-LSTM are consistent with those of manual picking. When the first arrivals are indistinguishable even to human interpreters (often due to strong abnormal noise), the model fails to pick them. This is because training labels cannot provide corresponding first-arrival positions for such unidentifiable traces. The energy ratio method attempts to pick every seismic trace, resulting in messy outputs for traces with unrecognizable first arrivals. It is reasonable to abandon picking for traces where first arrivals are unidentifiable (even manually) than to generate incorrect picks, as erroneous first arrivals significantly impact velocity modeling, while the absence of a small number of picks has minimal effect on subsequent processing.

Figure 4 shows the projections of the first-arrival time on the seismic data, instantaneous amplitude, frequency characteristic profile, and phase characteristic profile. After zooming in on the data in the red box, it can be seen that the first arrivals are located at the position of the first continuous strong amplitude and has similar phase and frequency characteristics. The method proposed in this paper can accurately pick the first arrivals through these characteristics. Figure 5 shows the absolute error of each seismic trace relative to the result of manual picking (only comparing the picked seismic traces). It can be seen from the absolute error that the picking effect of the energy ratio method is not as good as that of M-Res-LSTM.

3.2. Data 2

Data 2 used in this study is 2D seismic data acquired in a loess tableland area with dynamite sources. Each shot has 800 receiver channels, and the maximum offset is 8000 m. Each trace contains 751 sampling points with a sampling interval of 2 ms. From the perspective of the work area's geological conditions and data characteristics, the thickness of the loess layer in the work area varies significantly, with surface elevation ranging from 1200 to 1800 m and a maximum elevation difference of 600 m in the region. The severe topographic relief exerts a significant impact on the propagation path of seismic waves—not only causing the first-arrival phase within a single shot to be significantly disturbed by terrain but also leading to a large first-arrival time difference between adjacent receiver channels, which increases the basic difficulty of first-arrival picking. More critically, affected by the strong scattering of the loess layer itself and the energy attenuation of seismic waves, the first-arrival energy of the acquired data is generally weak, with unobvious onset characteristics. This has become the

Table 3. Detailed parameter table of each module in the M-Res-LSTM network

Network module	Submodule	Input dimension	Output dimension	Channel	Kernel size
Branch 1/2/3/4	ResNet	256×256×1	256×256×2	[256,128,64,32,16,8,4,2]	5×5
	LSTM	256×512	256×256×2	-	-
Feature fusion	Coordinate attention	4×[256×256×2]	256×256×4	4	1×1
Output layer	Convolutional Layer1	256×256×4	256×256×2	2	3×3
	Convolutional Layer2	256×256×2	256×256×1	1	1×1

Note: Parameter quantity=5.73 million.
Abbreviation: LSTM: Long Short-Term Memory.

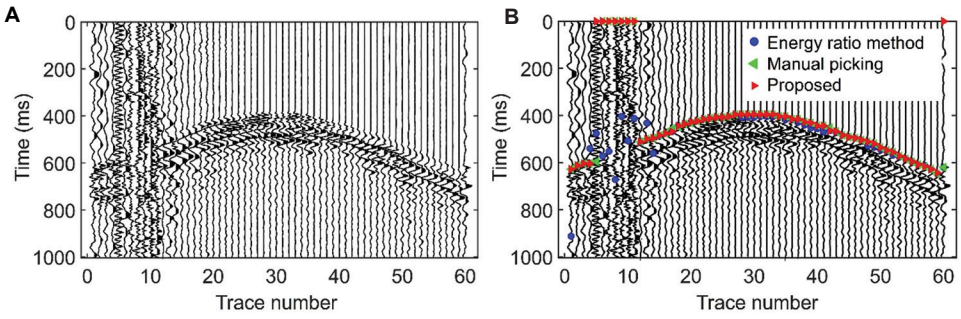

Figure 3. Data 1. (A) Original seismic record. (B) Picking results.

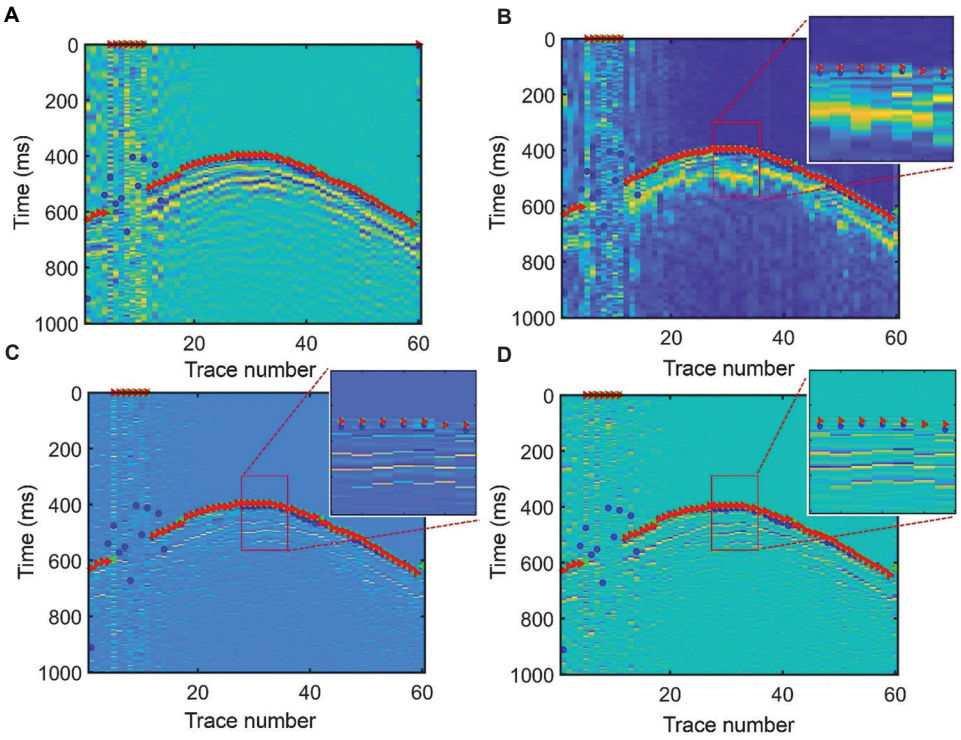

Figure 4. Projections of the seismic first arrivals on Data 1. (A) Original seismic record. (B) Instantaneous amplitude. (C) Frequency characteristic data. (D) Phase characteristic data.

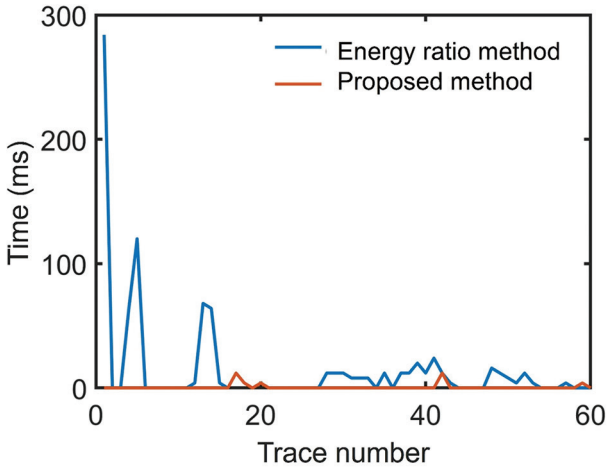

Figure 5. Absolute errors of the picking results between the energy ratio method and the method proposed in this paper (for Data 1)

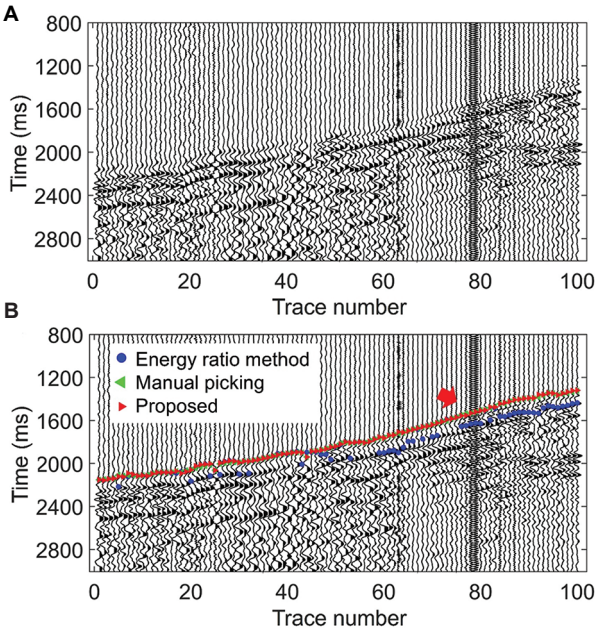

Figure 6. Data 2. (A) Original seismic record. (B) Picking results.

core difficulty in first-arrival picking for this dataset: most conventional picking methods tend to mistakenly identify subsequent phases after the first arrival as the first arrival, resulting in deviations in picking results.

Figure 6A presents a typical original profile, while Figure 6B displays the picking results obtained through manual picking (green triangles), the energy ratio method (blue circles), and the method proposed in this paper (red triangles). As observed from the picking results, the proposed method exhibits high consistency with manual

picking, whereas the energy ratio method erroneously identifies the position of the second peak (with stronger energy) as the first arrival. In addition, the proposed method successfully picks a small number of noisy traces in the seismic profile (at the position of the red arrow). This is attributed to the adoption of multi-trace input for training, enabling the network to infer first-arrival positions based on the characteristics of adjacent traces—analogous to the logic of manual picking. Figure 7 illustrates the absolute errors of the picking results of the two methods. Since the energy ratio method regards the second peak as the first arrivals, it has a relatively large absolute error. Similarly, by projecting the first-arrival time onto the seismic data, instantaneous amplitude, frequency characteristic profile, and phase characteristic profile (Figure 8), it can be seen that the first arrivals exhibit good consistency with these profiles. On magnification (within the white square), the proposed network is shown to accurately capture the amplitude, phase, and frequency characteristics of the first arrivals. Under the joint constraints of these three aspects, the accuracy of the picking is ensured.

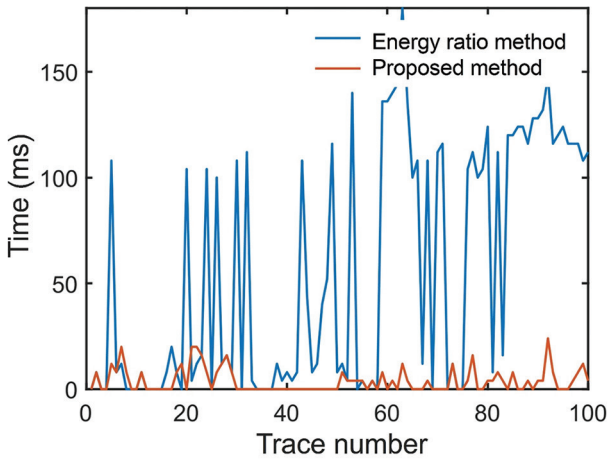

Figure 7. Absolute errors of the picking results between the energy ratio method and the method proposed in this paper (for Data 2)

3.3. Data 3

Data 3 is 3D seismic data acquired in the marginal area of a basin using a vibroseis source. For each shot, 30 receiver arrays were designed, with each array containing 170 receiver channels; the maximum offset exceeds 5000 m, enabling wide-range coverage of deep geological structures. However, due to the inherent limitation of the vibroseis source, namely, its limited excitation energy, the effective seismic wave energy received by geophones far from the source is significantly weakened, resulting in a relatively low overall SNR of the data. From the perspective of the work area's geological and topographic conditions, this region features a typical piedmont zone landscape, with extremely severe surface elevation relief: the elevation

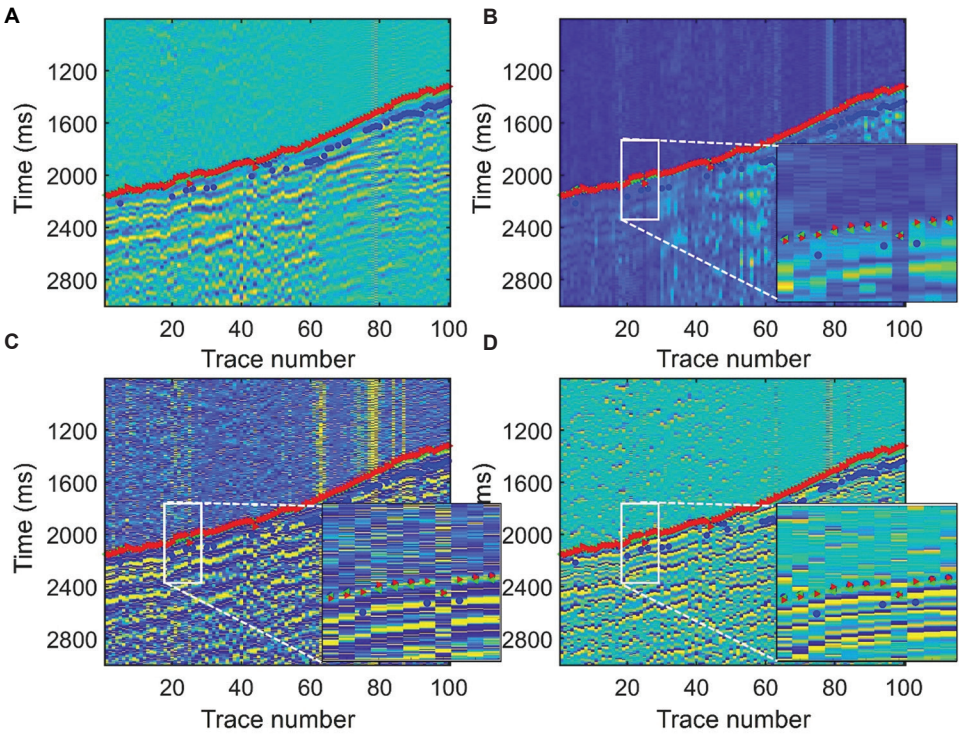

Figure 8. Projections of the seismic first arrivals on Data 2. (A) Original seismic record. (B) Instantaneous amplitude. (C) Frequency characteristic data. (D) Phase characteristic data.

ranges from 300 to 1500 m, and the maximum elevation difference in the area reaches 1200 m. Such severe topographic relief leads to a significant increase in the first-arrival time difference between adjacent receiver channels. Combined with the data's inherent issues—weak first-arrival energy and strong noise interference—this further complicates first-arrival picking.

To verify the cross-work area generalization ability of M-Res-LSTM, the model trained on Data 1 and Data 2 was directly transferred to Data 3 without any fine-tuning. Figure 9A shows a typical shot gather of Data 3, from which it can be seen that the first-arrival signals of some receiver channels are completely submerged in noise, and the energy difference between the first arrivals and background noise is small. Figure 9B compares the picking results of manual picking (yellow triangles), the energy ratio method (green triangles), and the proposed M-Res-LSTM in this study (red triangles). In Data 3, the energy ratio method is significantly affected by noise; in contrast, relying on the time-frequency dual-domain multi-feature constraint and coordinate attention mechanism, M-Res-LSTM still achieves favorable picking performance.

Figure 10 presents the absolute errors of the proposed method relative to manual picking (only valid picked channels are counted). The average absolute error of the energy ratio method reaches 5.9 ms, with the maximum error exceeding 150 ms, which is far beyond the acceptable range for seismic processing. In contrast, the average absolute error of M-Res-LSTM is only 1.34 ms, and the error of more than 87% of the gathers is controlled within 5 ms, which meets the accuracy requirements for near-surface inversion. These results indicate that by virtue of multi-domain feature learning and the attention mechanism, M-Res-LSTM effectively avoids overfitting to the features of the training work areas and can adapt to new work areas with significantly different geological conditions and noise levels.

3.4. Comparison against deep learning-driven first-arrival picking methods

To further verify the effectiveness of the proposed method, this study conducted deep learning-based tests on 1,000,000 seismic traces collected from multiple distinct work areas, in addition to testing traditional methods. The evaluation metrics selected include pixel accuracy, F1-score, first-arrival time deviation (characterized by mean absolute error, MAE), picking rate, and single-trace picking time (unit: milliseconds per trace),[40,45] which are used to comprehensively assess the performance of different networks in the seismic first-arrival picking task. Table 4 presents the performance differences between

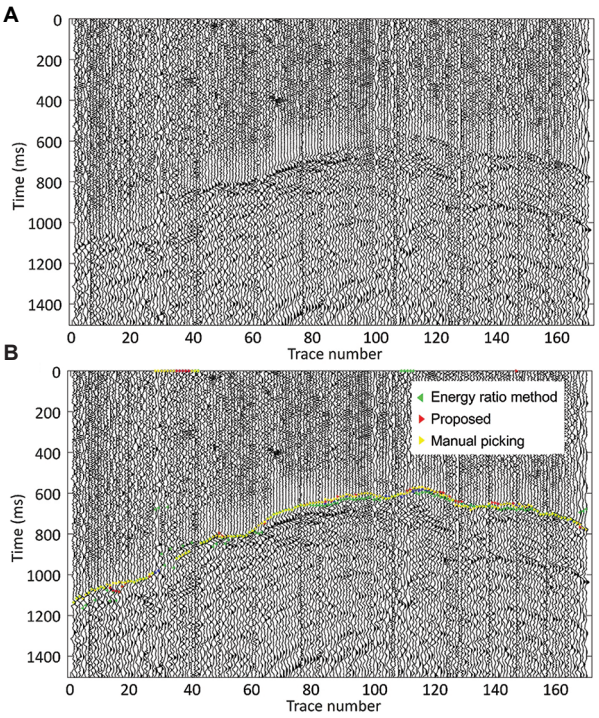

Figure 9. Data 3. (A) Original seismic record. (B) Picking results.

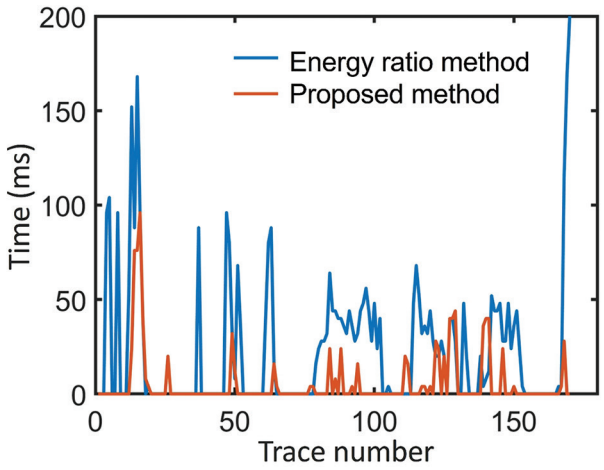

Figure 10. Absolute errors of the picking results between the energy ratio method and the method proposed in this paper (for Data 3)

Table 4. Comparison of picking results of different networks

Network	PA	F1-score	MAE (ms)	Picking rate (%)	Time (ms/trace)
Res-LSTM	0.975	0.942	7.8	92.1	0.65
Swin-Transformer	0.983	0.965	6.5	92.7	0.77
M-Res-LSTM	0.985	0.964	5.9	93.4	0.73

Abbreviation: M-Res-LSTM: Multi-perspective Residual Long Short-Term Memory; MAE: Mean absolute error; PA: Pixel accuracy; Res-LSTM: Residual Long-Short Term Memory.

the proposed M-Res-LSTM network, Res-LSTM-based networks, and Swin-Transformer-based networks[40] in seismic first-arrival picking.

Although Res-LSTM integrates the advantages of ResNet and LSTM networks, its simple cascaded structure prevents it from fully exploiting multi-dimensional information, resulting in limited overall performance and the lowest values across all metrics. Swin-Transformer achieves the highest F1-score by virtue of its self-attention mechanism; however, its single-input design restricts multi-domain feature fusion capability, and the high computational complexity of the self-attention mechanism causes it to underperform M-Res-LSTM in both first-arrival time accuracy and picking comprehensiveness.

The proposed M-Res-LSTM network in this study processes spatiotemporal signals, amplitude, frequency, and phase features in parallel through its multi-branch structure and coordinate attention mechanism, enabling comprehensive capture of spatiotemporal features. As shown in the table data, except for a slightly lower F1-score than Swin-Transformer, M-Res-LSTM outperforms Swin-Transformer in PA, picking rate, and single-trace picking time, with the first-arrival time deviation as low as 5.9 ms. This indicates that while ensuring classification accuracy comparable to Swin-Transformer, M-Res-LSTM achieves better performance in picking efficiency, picking comprehensiveness, and time accuracy through multi-domain feature parallel processing and efficient structural design. It fully verifies the effectiveness of the multi-branch structure and coordinates attention mechanism in the seismic first-arrival picking task, and can better balance accuracy, efficiency, and robustness.

4. Conclusion

The M-Res-LSTM network enables high-precision automatic picking of seismic first arrivals using time-frequency dual-domain features and an attention mechanism. Its multi-branch architecture supports parallel processing of amplitude, frequency, and phase features, thereby fully exploiting the multi-dimensional discriminative information inherent in seismic signals. The combination of residual modules and LSTM not only solves the degradation problem of deep networks but also strengthens the joint capture of spatiotemporal features. The coordinate attention mechanism effectively suppresses noise interference by dynamically adjusting feature weights, reducing the impact of incorrect first arrivals on subsequent velocity modeling.

Compared with traditional methods, manual picking achieves relatively high accuracy but suffers from the drawbacks of being time-consuming and labor-intensive. Furthermore, its results are significantly influenced by the experience of interpreters, making it difficult to meet the requirements of large-scale data processing. In contrast, the traditional energy ratio method enables automated processing yet is highly sensitive to the SNR of data, and tends to produce disorganized picking results or misjudgments in low-SNR scenarios. When compared with existing deep learning methods, the M-Res-LSTM network, leveraging a multi-feature parallel processing mechanism, exhibits superior robustness in practical applications compared to the single-branch Res-LSTM network. Meanwhile, in comparison with the Swin-Transformer network, although the M-Res-LSTM is slightly inferior in terms of pixel accuracy, it demonstrates distinct advantages in the average deviation of first-arrival time, picking rate, and single-trace computation time, thereby effectively balancing the accuracy and efficiency of first-arrival picking. Verified through data processing across different work areas, the proposed method can still obtain relatively ideal picking results even in complex scenarios with low SNR and weak first-arrival energy.

M-Res-LSTM adopts an end-to-end training mode, requiring no manual intervention. Moreover, as the amount of training data increases, its transfer ability to data from new work areas is expected to further improve. However, the complexity of the network structure makes its computation time slightly longer than that of simple models. In the future, efficiency can be optimized through model lightweighting. In addition, this paper verifies the effectiveness of time-frequency dual-domain features. Future research can explore fusion methods of more features or combine transfer learning to solve the training problems in small-sample work areas, promoting the large-scale application of this method under complex surface conditions.

Acknowledgments

None.

Funding

This work was supported in part by the Natural Science Foundation of Sichuan Province (2024NSFSC0808) and in part by the Talent Project of Chengdu Technological University (2024RC028).

Conflict of interest

The authors declare that there are no (potential) conflicts of interest with any institutes, organizations, or agencies that might influence the integrity of the results or the objective interpretation of this submitted work. All research activities and manuscript preparation have been

conducted in an objective and impartial manner, free from any factors that could compromise the validity and fairness of the study.

Author contributions

Conceptualization: Ziyu Qin, Xianju Zheng
Formal analysis: Ziyu Qin
Investigation: Wenhua Wang
Methodology: Ziyu Qin
Supervision: Xianju Zheng
Visualization: Ziyu Qin, Wenhua Wang
Writing–original draft: Ziyu Qin
Writing–review & editing: Xianju Zheng, Wenhua Wang

Availability of data

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Yuan S, Liu J, Wang S, Wang T, Shi P. Seismic waveform classification and first-break picking using convolution neural networks. *IEEE Geosci Remote Sens Lett*. 2018;15(2):272-276.

   doi: 10.1109/LGRS.2017.2785834

2. Peraldi R, Clement A. Digital processing of refraction data study of first arrivals. *Geophys Prospect*. 1972;20(3):529-548.

   doi: 10.1111/j.1365-2478.1972.tb00653.x

3. Molyneux JB, Schmitt DR. First-break timing: Arrival onset times by direct correlation. *Geophysics*. 1999;64(5): 1492-1501.

   doi: 10.1190/1.1444653

4. Gelchinsky B, Shtivelman V. Automatic picking of first arrivals and parameterization of travel time curves. *Geophys Prospect*. 1983;31:915-928.

   doi: 10.1111/j.1365-2478.1983.tb01097.x

5. Coppens F. First arrival picking on common-offset trace collections for automatic estimation of static corrections. *Geophys Prospect*. 1985;33(8):1212-1231.

   doi: 10.1111/j.1365-2478.1985.tb01360.x

6. Qin Z, Pan S, Hu L, Cui Q, Gou Q. First-arrival automatic picking based on improved energy ratio method and outlier detection theory. *Acta Geophysica*. 2021;69(5):1667-1677.

   doi: 10.1007/s11600-021-00639-w

7. Allen RV. Automatic earthquake recognition and timing from single traces. *Bull Seismol Soc Am*. 1978;68(5): 1521-1532.

   doi: 10.1785/bssa0680051521

8. Wong J, Han L, Bancroft J, Stewart RR. *Automatic Time-Picking of First-Arrivals on Noisy Microseismic Data*. Alberta: CSEG Microseismic Workshop; 2009. p. 1-6.

9. Akram J, Eaton DW. A review and appraisal of arrival-time picking methods for downhole microseismic data. *Geophysics*. 2016;81(2):KS71-KS91.

   doi: 10.1190/geo2014-0500.1

10. Pan S, Qin Z, Lan H, Badal J. Automatic first-arrival picking method based on an image connectivity algorithm and multiple time windows. *Comput Geosci*. 2019;123:95-102.

    doi: 10.1016/j.cageo.2018.12.001

11. Leonard M. Comparison of manual and automatic onset time picking. *Bull Seismol Soc Am*. 2000;90(6):1384-1390.

    doi: 10.1785/0120000026

12. Zhang H, Thurber C, Rowe C. Automatic p-wave arrival detection and picking with multiscale wavelet analysis for single component recordings. *Bull Seismol Soc Am*. 2003;93(5):1904-1912.

    doi: 10.1785/0120020241

13. Jin H, Li L, Cheng S, Li X. Time-arrival pickup method of tunnel water inrush microseismic signals based on kurtosis value and AIC method. *Tunnell Undergr Space Technol*. 2024;154:106135.

    doi: 10.1016/j.tust.2024.106135

14. Boschetti F, Dentith MD, List RD. A fractal-based algorithm for detecting first arrivals on seismic traces. *Geophysics*. 1996;61(4):1095-1102.

    doi: 10.1190/1.1444030

15. Chi-Duráur-Dur/1.14440Dur- M, Silva J. Automatic detection of p- and s-wave arrival times: New strategies based on the modified fractal method and basic matching pursuit. *J Seismol*. 2017;21(5):1171-1184.

    doi: 10.1007/s10950-017-9658-0

16. Jiao LX, Moon WM. Detection of seismic refraction signals using a variance fractal dimension technique. *Geophysics*. 2000;65(1):286-292.

    doi: 10.1190/1.1444719

17. Esmaeili S. A new automatic first break picking method based on the STA/LTA fractal dimension algorithm. *J Seismic Explor*. 2022;31(3):253-265.

18. Lan Z, Gao P, Wang P, Wang Y, Liang J, Hu G. Automatic first arrival time identification using fuzzy C-Means and AIC. *IEEE Trans Geosci Remote Sens*. 2022;60:5907613.

    doi: 10.1109/TGRS.2021.3121032

19. Zhang Y, Lan Z, Xue Y, Wang J, Zhu K, He J. Automatic first arrival picking based on self-similarity and multicenter fuzzy clustering. *IEEE Trans Geosci Remote Sens*. 2025;63:5906013.

    doi: 10.1109/TGRS.2025.3538757

20. Mousa WA, Al-Shuhail AA. Enhancement of first arrivals using the ivaltransform on energy-ratio seismic shot records. *Geophysics*. 2012;77(3):V101-V111.

    doi: 10.1190/geo2010-0331.1

21. Gaci S. The use of wavelet-based denoising techniques to enhance the first-arrival picking on seismic traces. *IEEE Trans Geosci Remote Sens*. 2014;52(8):4558-4563.

    doi: 10.1109/tgrs.2013.2282422

22. Cheng Y, Li Y, Zhang C. First arrival time picking for microseismic data based on shearlet transform. *J Geophys Eng*. 2017;14(2):262-271.

    doi: 10.1088/1742-2140/aa5777

23. Sheng GQ, Tang XG, Xie K, Xiong J. Hydraulic fracturing microseismic first arrival picking method based on non-subsampled shearlet transform and higher-order-statistics. *J Seismic Explor*. 2019;28(6):593-618.

24. Dong XT, Jiang H, Li Y, Yang B. Arrival time picking of micro-seismic data by using SPE algorithm. *J Seismic Explor*. 2019;28(5):475-494.

25. Qin Z, Pan S, Chen J, Cui Q, He J. Method of automatically detecting the abnormal first arrivals using delay time. *IEEE Trans Geosci Remote Sens*. 2022;60:4504908.

    doi: 10.1109/tgrs.2021.3118921

26. Murat ME, Rudman AJ. Automated first arrival picking: A neural network approach. *Geophys Prospect*. 1992;40: 587-604.

    doi: 10.1111/j.1365-2478.1992.tb00543.x

27. McCormack MD, Zaucha DE, Dushek DW. First-break refraction event picking and seismic data trace editing using neural networks. *Geophysics*. 1993;58(1):67-78.

    doi: 10.1190/1.1443352

28. Maity D, Aminzadeh F, Karrenbach M. Novel hybrid artificial neural network based autopicking workflow for passive seismic data. *Geophys Prospect*. 2014;62:834-847.

    doi: 10.1111/1365-2478.12125.

29. Maity D, Salehi I. Neuro-evolutionary event detection technique for downhole microseismic surveys. *Comput Geosci*. 2016;86:23-33.

    doi: 10.1016/j.cageo.2015.09.024

30. Mousavi SM, Horton SP, Langston CA, Samei B. Seismic features and automatic discrimination of deep and shallow induced-microearthquakes using neural network and logistic regression. *Geophys J Int*. 2016;207:29-46.

    doi: 10.1093/gji/ggw258

31. Wu Y, Pan S, Lan H, Badal J, Wei Z, Chen Y. Automatic seismic first-break picking based on multi-view feature fusion network. *Geophys Prospect*. 2024;72(9):3547-3559.

    doi: 10.1111/1365-2478.13592

32. Tsai KC, Hu W, Wu X, Chen J, Han Z. Automatic first arrival picking via deep learning with human interactive learning. *IEEE Trans Geosci Remote Sens*. 2020;58(2):1380-1391.

    doi: 10.1109/tgrs.2019.2946118

33. Wang J, Xiao Z, Liu C, Zhao D, Yao Z. Deep learning for picking seismic arrival times. J *Geophys Res Solid Earth*. 2019;124(7):6612-6624.

    doi: 10.1029/2019jb017536

34. Zhu W, Beroza GC. Phasenet: A deep-neural-network-based seismic arrival time picking method. *Geophys J Int*. 2019;216(3):1831-1841.

    doi: 10.1093/gji/ggy423

35. Wu H, Zhang B, Li F, Liu N. Semiautomatic first-arrival picking of microseismic events by using the pixel-wise convolutional image segmentation method. *Geophysics*. 2019;84(3):V143-V155.

    doi: 10.1190/geo2018-0389.1

36. Mardan A, Blouin M, Fabien-Ouellet G, Giroux B, Vergniault C, Gendreau J. A fine-tuning workflow for automatic first-break picking with deep learning. *Near Surface Geophys*. 2024;22:539-552.

    doi: 10.1002/nsg.12316

37. Zhang Z, Yang J. Seismic first break picking based on multi-task learning. *Front Earth Sci*. 2025;13:1601134.

    doi: 10.3389/feart.2025.160113433

38. Li H, Sun Y, Li J, Li H, Dong H. A meta-learning based approach for automatic first-arrival picking. *IEEE Trans Geosci Remote Sens*. 2024;62:5922615.

    doi: 10.1109/tgrs.2024.3436817

39. Yuan P, Hu W, Wu X, Nguyen H. Adaptive first arrival picking model with meta-learning. In: *SEG International Exposition and 90th Annual Meeting*. United States: IEEE; 2020. p. 1486-1490.

    doi: 10.1190/segam2020-3420195.1

40. Jiang P, Deng F, Wang X, Shuai P, Luo W, Tang Y. Seismic first break picking through swin transformer feature extraction. *IEEE Geosci Remote Sens Lett*. 2023;20:7501505.

    doi: 10.1109/LGRS.2023.3248233

41. Chen K, Li M, Li X, *et al*. Enhancing microseismic event detection with TransUNet: A deep learning approach for simultaneous pickings of P-wave and S-wave first arrivals. *Artif Intell Geosci*. 2025;6:100129.

    doi: 10.1016/j.aiig.2025.100129

42. Jiang P, Deng F, Wang X, Luo W, Ye C. 3-D seismic first break picking based on two-channel mask strategy. *IEEE Trans Geosci Remote Sens*. 2024;62:5918115.

    doi: 10.1109/TGRS.2024.3412673

43. Liao X, Cao J, Hu J, You J, Jiang X, Liu Z. First arrival time identification using transfer learning with continuous wavelet transform feature images. *IEEE Geosci Remote Sens Lett*. 2020;17(11):2002-2006.

doi: 10.1109/lgrs.2019.2955950

44. Zhou Y, Yue H, Kong Q, Zhou S. Hybrid event detection and phase-picking algorithm using convolutional and recurrent neural networks. *Seismol Res Lett*. 2019;90(3):1079-1087.

doi: 10.1785/0220180319

45. Wang H, Zhang J, Wei X, Zhang C, Long L, Guo Z. MSSPN: Automatic first-arrival picking using a multistage segmentation picking network. *Geophysics*. 2024;89(3): U53-U70.

doi: 10.1190/geo2023-0110.1

46. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016. p. 770-778.

doi: 10.48550/ARXIV.1512.03385

47. Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. United States: IEEE; 2021.

doi: 10.48550/arXiv.2103.02907

48. Taner MT, Koehler F, Sheriff RE. Complex seismic trace analysis. *Geophysics*. 1979;44(6):1041-1063.

doi: 10.1190/1.1440994

49. Robertson JD, Nogami HH. Complex seismic trace analysis of thin beds. *Geophysics*. 1984;49(4):344-352.

doi: 10.1190/1.1441670

50. Kingma D, Ba J. Adam: A Method for Stochastic Optimization. *In: Conference Paper at the 3rd International Conference for Learning Representations*, San Diego, CA, USA; 2015.

doi: 10.48550/ARXIV.1412.6980

51. Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017;39(4):640-651.

doi: 10.1109/tpami.2016.2572683