GEOPHYSICS, VOL. 88, NO. 4 (JULY-AUGUST 2023); P. E107–E122, 18 FIGS., 3 TABLES.
10.1190/GEO2022-0317.1

DOI:10.1190/geo2022-0317.1

IncepTCN: A new deep temporal convolutional network combined with dictionary learning for strong cultural noise elimination of controlled-source electromagnetic data

Guang Li[1], Shouli Wu[2], Hongzhu Cai[3], Zhushi He[2], Xiaoqiong Liu[2], Cong Zhou[2], and Jingtian Tang[4]

ABSTRACT

When the controlled-source electromagnetic (CSEM) data are contaminated by intense cultural noise and the signal-to-noise ratio (S/N) is lower than 0 dB, the existing denoising methods can hardly achieve good results. To overcome the problem, a new strong-noise elimination method called inception-temporal convolutional network-shift-invariant sparse coding (IncepTCN-SISC) is developed based on deep learning and dictionary learning. First, a novel deep neural network model called IncepTCN is created based on the inception block and temporal convolutional network (TCN). Then, IncepTCN is used to recognize strong-noise segments in the observed signal, which are then discarded. Finally, a dictionary-learning method based on shift-invariant convolutional coding is used to denoise the remaining weak-noise segments. A series of simulated and field data experiments indicate that the new proposed IncepTCN network has obvious advantages in accuracy and efficiency compared with alternative methods. The average recognition accuracy of IncepTCN is 96.5%, which is 25.5%, 3.2%, 1.1%, and 2.0% higher than that of the fuzzy C-means clustering, convolutional neural network (CNN), residual network (ResNet), and the nonimproved TCN, respectively. In addition, the test results of unfamiliar data indicate that the generalization ability of IncepTCN is significantly better than the CNN, ResNet, and nonimproved TCN. This IncepTCN-SISC method can improve the S/N of CSEM data from −5.0 dB to 3.1 dB or from 5.0 dB to 31.9 dB and solve the denoising problem of noisy data below 0 dB to a certain extent. After IncepTCN-SISC processing, the initially distorted apparent resistivity curves become smooth, and the result is better than dictionary learning. This method is intelligent without any manual intervention and is suitable for batch processing of CSEM data.

INTRODUCTION

The controlled-source electromagnetic (CSEM) method includes the controlled-source audio magnetotelluric (CSAMT) method, wide-field electromagnetic (WFEM) method, induced polarization (IP) method, and transient electromagnetic (TEM) method. It is widely used in resource exploration, engineering geologic exploration, geologic disaster detection, and many other fields (Myer et al., 2011; Maclennan and Li, 2013; He, 2018; Hu et al., 2022). Due to controlled sources, CSEM has a better anti-interference ability than the natural source-based MT sounding method. Nevertheless, the results of CSEM exploration are still inevitably affected by cultural noises. With the continuous increase in exploration depth and accuracy requirements, it becomes more important to eliminate cultural noise in CSEM data (Yang et al., 2018; Adrian et al., 2021).

To improve the quality of CSEM data, scholars have proposed many effective cultural noise suppression methods. Among them,

Manuscript received by the Editor 18 May 2022; revised manuscript received 18 December 2022; published ahead of production 13 March 2023; published online 7 June 2023.
[1]China University of Geosciences, Badong National Observation and Research Station of Geohazards, School of Geophysics and Geomatics, Wuhan, China and East China University of Technology, Jiangxi Engineering Technology Research Center of Nuclear Geoscience Data Science and System, Nanchang, China. E-mail: li_guangg@163.com.
[2]East China University of Technology, School of Geophysics and Measurement-Control Technology, Nanchang, China. E-mail: wsl13147161170@163.com; he2648249123@163.com; xiaoqiongliu5288@163.com; zhoucong_522@163.com.
[3]China University of Geosciences, School of Geophysics and Geomatics, Wuhan, China. E-mail: caihongzhu@hotmail.com (corresponding author).
[4]Central South University, Key Laboratory of Metallogenic Prediction of Non-Ferrous Metals and Geological Environment Monitor, Ministry of Education, Changsha, China. E-mail: jttang@mail.csu.edu.cn.

two commonly used schemes are noise separation and data selection. Noise separation includes singular-value decomposition (SVD) (Reninger et al., 2011), wavelet transform (Deo and Cull, 2016), empirical-mode decomposition (EMD) (Liu et al., 2019), wavelet neural networks (Wu et al., 2019), and other methods. In recent years, compressed sensing has successfully implemented seismic signal and ground-penetrating radar (GPR) signal processing (Candès and Wakin, 2008). Sparse representation (Jafari and Plumbley, 2011), which is the critical technology of compressed sensing, has attracted much attention in the electromagnetic community. It is first used to remove periodic or regular components in MT signals (Li et al., 2017) and MT inversion (Nittinger and Becken, 2018).

The CSEM data with periodic characteristics can easily be sparsely represented (most of the amplitudes are very small or equal to zero in a particular transform domain), which is especially suitable for sparse representation denoising. Zhang et al. (2020) denoise the marine CSEM (MCSEM) data via a predefined dictionary. Even so, the atoms in the predefined dictionary are not flexible enough for different kinds of noise. Xue et al. (2020) propose an airborne TEM (ATEM) data denoising method using K-SVD dictionary learning, and its performance is significantly better than the predefined dictionary. The K-SVD algorithm also has achieved good results in the denoising of MCSEM data (Zhang et al., 2022a). Taking advantage of the sparsity of periodic signals, Li et al. (2021a) propose a denoising method of CSEM data based on shift-invariant sparse coding (SISC) dictionary learning (Blumensath and Davies, 2006). The method automatically obtains sparsity and other parameters, and it can adaptively and accurately remove various types of noise, such as power-frequency interference, baseline-drift noise, random noise, and impulsive noise; its performance is better than soft-threshold wavelet denoising (wavelet), mathematical morphological filtering (MMF), and K-SVD dictionary learning. However, the sparse representation denoising method also has its limitation. When there is abnormally large amplitude noise in the time series, its denoising accuracy decreases sharply.

Theoretically, we only need one cycle of high-quality data to obtain all the required information because of the periodic controlled-source signal. Therefore, the data selection method also is suitable for CSEM data processing. The data selection method can be implemented in the frequency, time, or other domains (Rita et al., 2013). In most cases, the selection threshold needs to be set manually, such as the WFEM spectra selection method based on the gray judgment criterion (Mo et al., 2017) and the IP time-series selection method based on the correlation coefficient between the transmitted signal and the received signal (Liu et al., 2017). However, setting the threshold manually can easily cause subjective deviation and is not suitable for batch processing. Rather than using the single correlation coefficient, Li et al. (2021b) use the correlation coefficient, maximum value of amplitude, and sample entropy as the evaluation criteria and implement unsupervised machine learning to obtain the threshold automatically. The method eliminates the subjective deviation and thus improves the reliability. However, unsupervised machine learning can only divide the data into two parts and cannot determine the high-quality one.

Deep learning is a new research direction in the field of machine learning. It shows outstanding performance in classification and prediction. Its performance in many areas far exceeds that of previous technology (Lecun et al., 2015; He et al., 2021; Kaur et al., 2021; Yu and Ma, 2021), and it increasingly has attracted attention in the electromagnetic community in recent years. Li et al. (2020) present an ATEM imaging method based on the long short-term memory (LSTM) technique and error back-propagation scheme. Moghadas (2020) realizes 1D electromagnetic induction data inversion based on a convolutional neural network (CNN). CNN also is used for MT data classification (Li et al., 2022b, 2023). Bang et al. (2021) propose an ATEM imaging method based on a recurrent neural network (RNN). Deep learning also has appeared in the field of CSEM data denoising. Wu et al. (2020) propose a multisource noise removal method for ATEM data via a deep denoising autoencoder (DAE) neural network. Wu et al. (2021) propose a new TEM data denoising method based on LSTM and DAE. Zhang et al. (2022b) classify and denoise MT data based on the residual network (ResNet). Although deep learning in electromagnetic exploration is still in its infancy, its advantages have been demonstrated in recent publications.

The existing deep-learning identification methods of electromagnetic time series mainly include CNN and ResNet, but research shows that CNN has some shortcomings, such as gradient explosion or disappearance (Yang and Li, 2020; Ma et al., 2021; Li et al., 2022a). In recent years, Lea et al. (2017) propose a new deep neural network (DNN) model for time series processing, namely temporal convolutional network (TCN). Similar to LSTM, TCN also is invented for sequence modeling, including time series classification and prediction, but its application in geophysics is rarely reported. This paper constructs a new deep-learning network model called IncepTCN based on TCN and inception block (a multiscale feature extraction network) (Szegedy et al., 2015) and combines it with dictionary learning to solve the problems in dictionary learning and unsupervised machine learning mentioned previously. The main idea is as follows. First, the strong-noise segments are identified by the proposed IncepTCN network and discarded to avoid their contribution to the decline in dictionary-learning accuracy. Then, a dictionary-learning method based on SISC is used to denoise the remaining weak-noise segments. The supervised deep-learning algorithm learns the rules from the manually labeled samples and can accurately distinguish the high-quality and strong-noise segments. Therefore, it also solves the problem that the unsupervised machine-learning method can only classify but cannot judge the categories.

The rest of the paper is organized as follows. First, it introduces the origin, advantages, and basic network structures of CNN, ResNet, TCN and inception, and the construction of the optimized network IncepTCN. Then, through the analysis of synthetic data, the superiority and breakdown point of the proposed method are tested. Subsequently, the performances of five algorithms, i.e., fuzzy C-means (FCM) clustering, CNN, ResNet, TCN, and IncepTCN, are compared in terms of classification accuracy (A), precision (P), recall rate (R), $F_{1\text{-score}}$, and confusion matrix. Subsequently, the proposed method is applied to the measured CSEM data to verify the effectiveness further. Finally, some conclusions and suggestions are presented.

METHOD AND THEORY

Data processing flow

The data processing workflow is shown in Figure 1. First, the observed data are preprocessed by conventional means. Then, IncepTCN is used to divide the preprocessed data into strong- and weak-noise segments. The strong-noise segments are directly

discarded, and the weak-noise segments are retained. Finally, SISC removes the residual noise in the weak-noise segments. The preprocessing mentioned previously includes removing power-frequency interference based on fast Fourier transform (FFT) and removing baseline drift via complementary ensemble EMD (CEEMD). Compared to the FFT-CEEMD-SISC method proposed by Li et al. (2021a), the main difference in our scheme is the addition of the IncepTCN identification. Therefore, the focus of this paper is IncepTCN. For other relevant algorithms, please refer to the relevant literature (Li et al., 2021a).

Temporal convolution networks

The temporal convolution network (Lea et al., 2017) is proposed to improve the performance of sequence modeling. It solves the problems of the varying length of input and output sequences and future information leakage in CNNs. It realizes the function similar to RNNs, which can obtain sequences of arbitrary length and map them to output sequences of the same length. Meanwhile, it avoids the problems of gradient explosion or disappearance and excessive memory occupation of RNNs. Its prominent feature is the use of dilated causal convolution (DCC). Causal convolution can solve the problems of different input and output sequence lengths and future information leakage in the CNN model. Dilated convolution can broaden the receptive field of the convolution kernel and reduce the number of network layers. TCN has simple architecture and no hop connection between layers, allowing very deep networks and a long-term memory. It can predict the time series more accurately combined with historical information (Yang and Li, 2020; Ma et al., 2021; Li et al., 2022a). The preliminary evaluation of TCN shows that its performance on multiple tasks and data sets is better than RNN networks, such as LSTM, and shows longer-term memory (Bai et al., 2018; Hewage et al., 2020).

As shown in Figure 2, the basic TCN is composed of three parts: the input layer, hidden layer, and output layer. The dimension of the input layer is the CSEM sample length multiplied by the number of channels. The hidden layer consists of five residual blocks (Figure 3), containing two 1D DCCs and a 1D convolution. The output of the residual block is aggregated by the results of two 1D DCCs and a 1D convolution. The connection between input and output can avoid the network degradation caused by the increase of layers. The output layer consists of a flattened layer and a fully connected dense layer. The flatten layer flattens the result from the residual block 5 into a 1D sequence; the dense layer contains two neurons and is activated using the softmax function. Softmax function often is used in classification tasks, which maps the output of the network into the interval [0, 1] and represents the normalized category probability. Let the training set has N samples, then the softmax function is defined as follows (Cardarilli et al., 2021):

$$y_{ic} = \frac{e^{y_i}}{\sum_{c=1}^{C} e^{y_c}}, \tag{1}$$

where y_i is the output of the network, $i = 1, \ldots, N, c \in [1, C]$, and C is the number of categories. Then, y_{ic} represents the probability that y_i belongs to the cth category.

DCC is a combination of causal convolution and dilated convolution. Causal convolution obtains historical information, and dilated convolution expands the receptive field of the convolution kernel. The architecture of DCC is shown in Figure 4. Causal convolution means that the mth element of the output sequence only depends on the mth and previous elements in the input sequence, which effectively avoids the influence of future information on the existing state. The dilated convolution enables the network to obtain more extended historical information based on the same network layer. To ensure the consistent length of input and output sequences, the TCN adopts the mechanism of zero filling, and the number of filling is

$$z = (k-1) \times d, \tag{2}$$

where k is the size of the convolution kernel, $d = 2^L$ is the dilated coefficient, and L is the number of network layers before the dilated convolution. Meanwhile, to speed up obtaining historical information, the dilated convolution is used to control the growth rate with the dilated coefficient. Taking time series $X = \{x_1, x_2, \ldots, x_M\}$ as an example, using convolution kernel f with size k, the DCC of the input x_m in the layer L is

$$F_d^L(m) = \sum_{i=0}^{k-1} f_i \cdot x_{m-i\cdot d}^{L-1}, \tag{3}$$

where the superscript indicates the network layer of the current element. The activation function $H_d^L(m)$ after convolution is

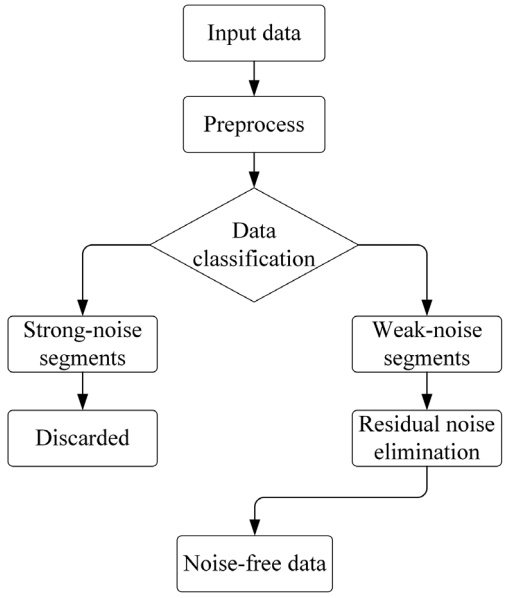

Figure 1. Workflow of the proposed IncepTCN-SISC method.

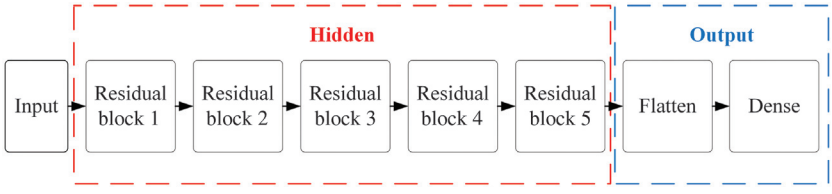

Figure 2. The architecture of basic TCNs.

DOI:10.1190/geo2022-0317.1

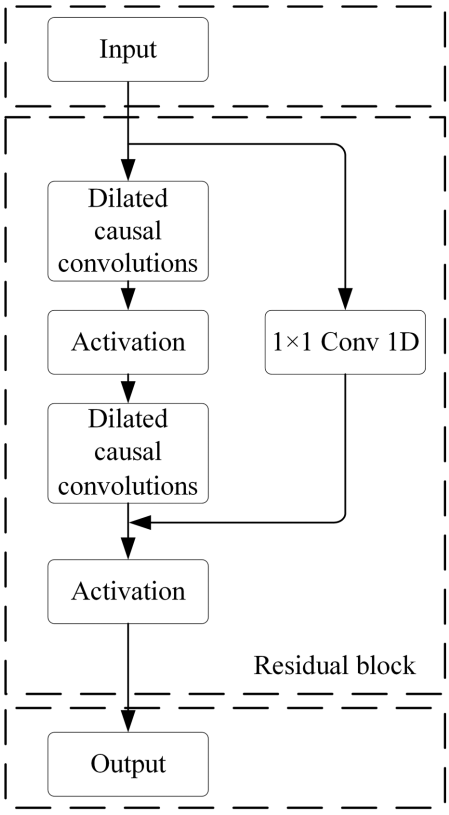

Figure 3. The structure of residual block.

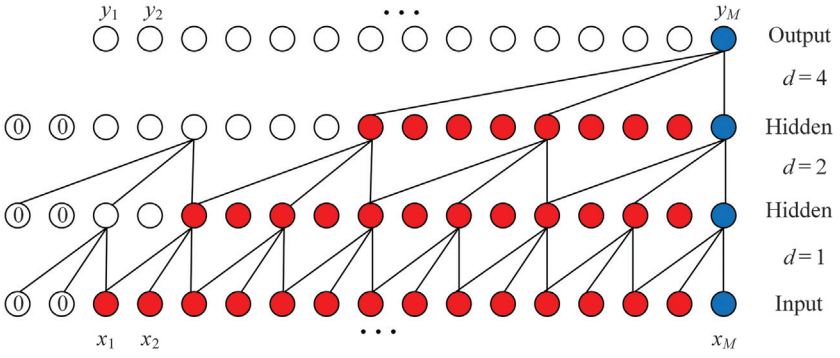

Figure 4. Schematic diagram of DCC.

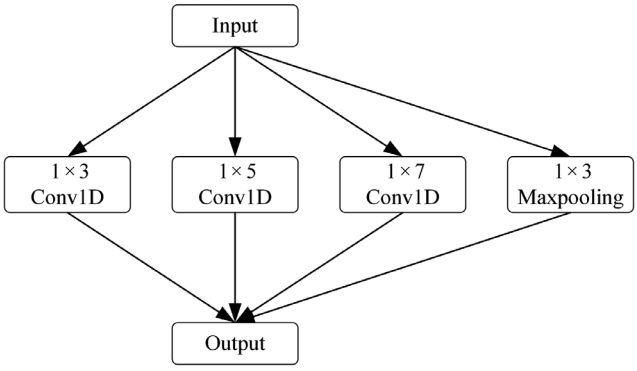

Figure 5. Sketch map of inception block.

$$H_d^L(m) = \mathrm{ReLU}(F_d^L(m) + b), \tag{4}$$

where ReLU is the rectified linear unit, which is one of the commonly used activation functions in the convolution network, and b is the offset term. The expression of ReLU is as follows:

$$\mathrm{ReLU}(t) = \max[0, t] = \begin{cases} 0 & t \le 0 \\ t & t > 0 \end{cases}. \tag{5}$$

TCN uses residual blocks to enhance features further. The activation function $\tilde{H}^L(m)$ after adding residual connection is

$$\tilde{H}_d^L(m) = \mathrm{ReLU}(F_d^L(m) + b + H_d^{L-1}(m)). \tag{6}$$

According to equation 6, the features in layer L can be further strengthened based on layer L-2 when the network learns effective data features in layers L and L-1, which improves the generalization ability of the model to a certain extent and prevents the degradation of the network. Network degradation means that the accuracy of the model reaches saturation as the number of network layers increases. If the number of network layers continues to increase, the accuracy will decline rapidly (He et al., 2016). Even if the network does not learn compelling data features in layers L and L-1, the network can still retain the features and gradients learned in layer L-2.

Inception-temporal convolutional networks

The inception module (Szegedy et al., 2015) is a multiscale feature extraction structure (Figure 5). It contains multiple convolution cores of different scales. The output is the superposition of multiple convolution layers in the depth direction, which is helpful for extracting the features of different scales. The existing literature shows that combining the inception module and dilated convolution can improve the model's accuracy (Fawaz et al., 2020; Kang et al., 2020). Inspired by this, we use the inception module to improve the performance of TCN. In deep learning, multiple modules often are superimposed repeatedly to improve network performance. To obtain the best model, we conduct a series of tests on different combinations of modules.

First, we keep the number of inception modules at one and gradually increase the number of TCN modules to obtain a series of different networks. The networks are trained by the same data and then used to classify 32 groups of test data. There are more than 74,000 samples in the training set, and the test data are from Huidong County, Sichuan Province and Qiaojia County, Yunnan Province. Details will be introduced subsequently. The experimental results are plotted as two curves, as shown in Figure 6. Obviously, with the increase of the number of TCN modules, the training time is significantly increased, but the average accuracy is not significantly improved. The highest average accuracy is 96.2%.

Second, we keep the number of TCN modules unchanged and then gradually increase the number of inception modules. As shown in Figure 7, with the increase of the number of inception modules, the average accuracy gradually increases, but the time consumption also increases significantly. The highest average accuracy reaches 96.7%. In general, more network layers (modules) mean better

nonlinear expression ability, so that more complex features can be learned. However, with the increase of network layers, the number of parameters also will increase dramatically, resulting in a significant increase in time consumption.

Finally, according to the ratio of inception to TCN of 1:2, we form a new module inception-TCN-TCN and gradually increase its number. The experimental results are shown in Figure 8. When the number of inception-TCN-TCN modules is five, the average accuracy reaches 96.5%, and the training takes only 1666.9 s. Although its average accuracy is slightly lower than the highest accuracy of 96.7% shown in Figure 7, its training time is significantly less than that of the network corresponding to the same level of accuracy shown in Figure 7. Considering the accuracy and efficiency, we finally choose the network consisting of five inception modules and 10 TCN modules.

Next, we will introduce the specific structure of the network. Meanwhile, we show the network structures of CNN and ResNet, which are currently the most commonly used and advanced deep-learning classification methods. As shown in Figure 9a, the commonly used CNN is composed of nine layers of five types, including

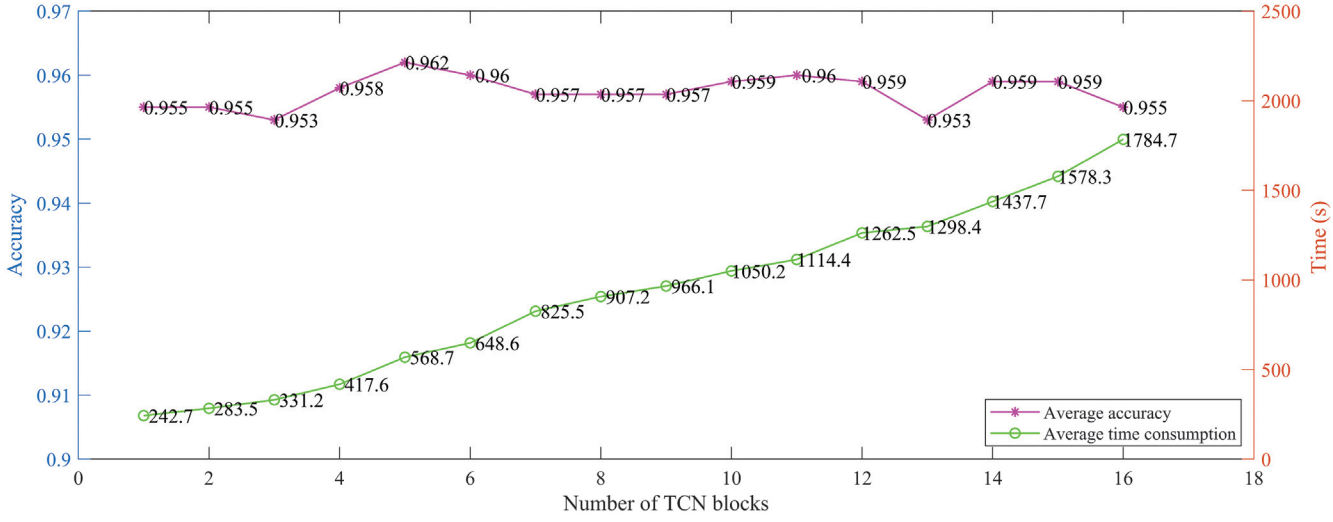

Figure 6. Curves of average accuracy (purple line, left scale) and time consumption (green line, right scale) when the number of TCN modules increases.

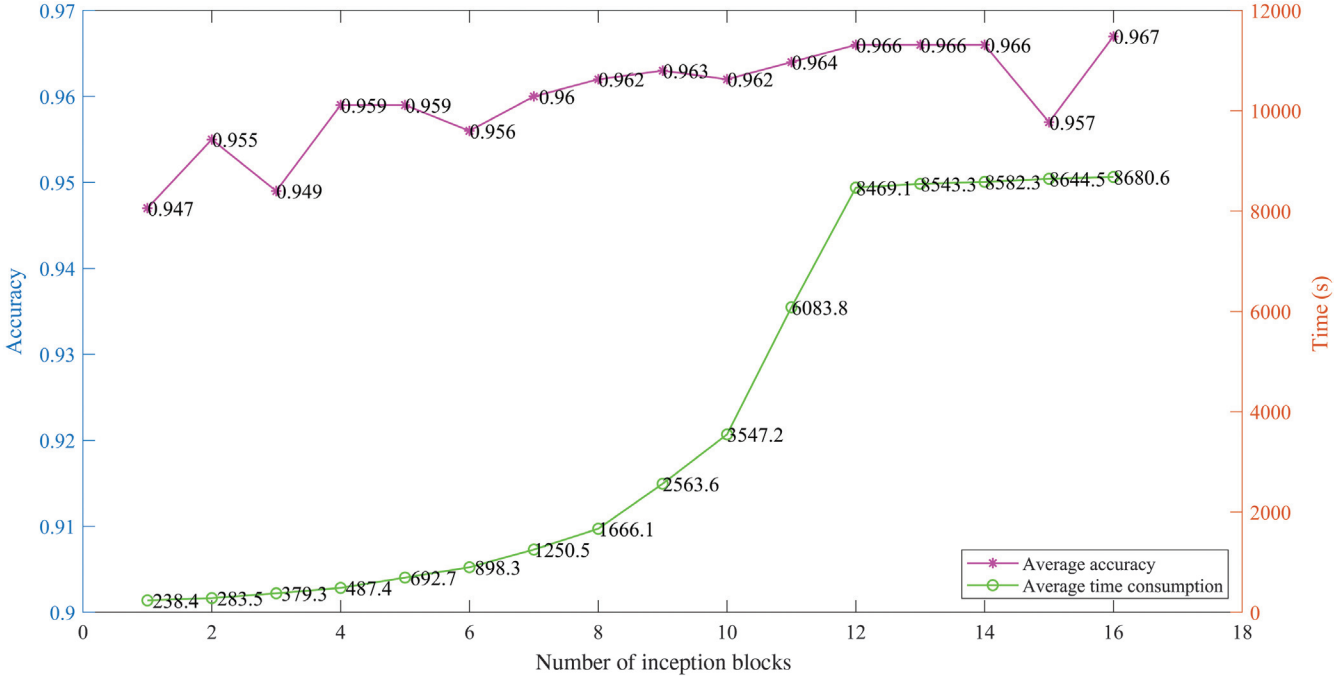

Figure 7. Curves of average accuracy (purple line, left scale) and time consumption (green line, right scale) when the number of inception modules increases.

DOI:10.1190/geo2022-0317.1

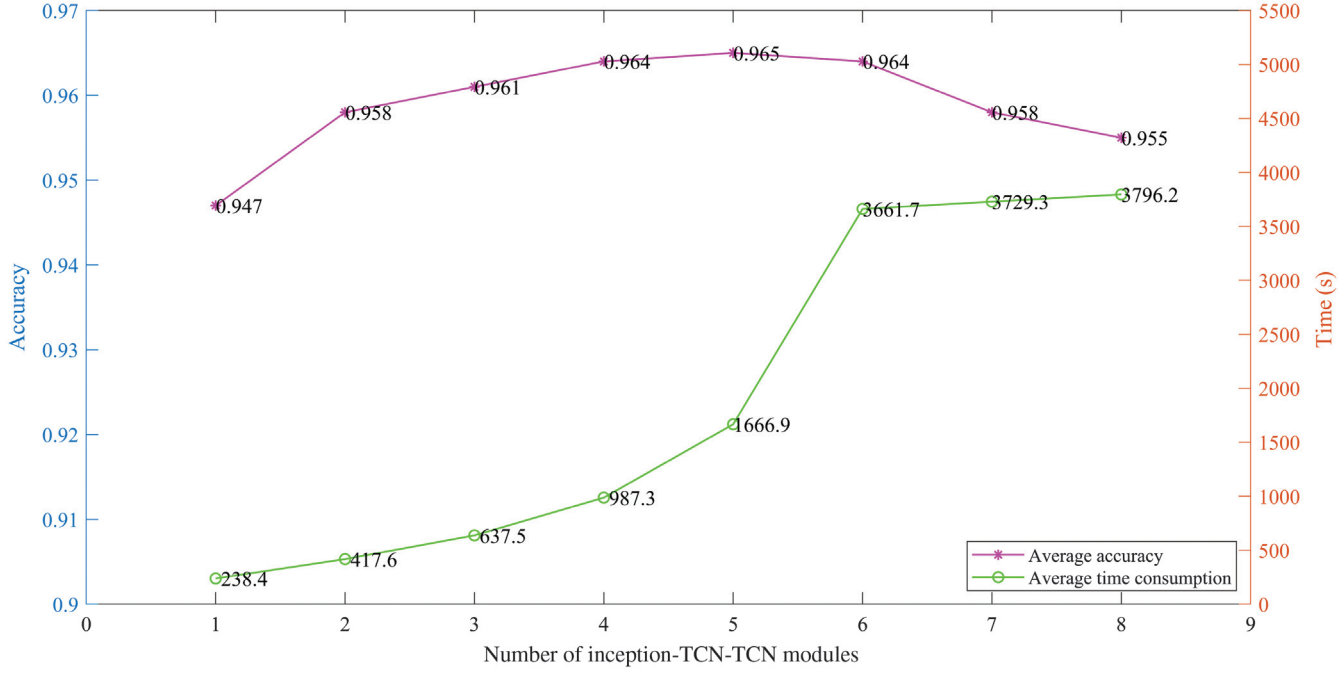

Figure 8. Curves of average accuracy (purple line, left scale) and time consumption (green line, right scale) when the number of inception-TCN-TCN modules increases.

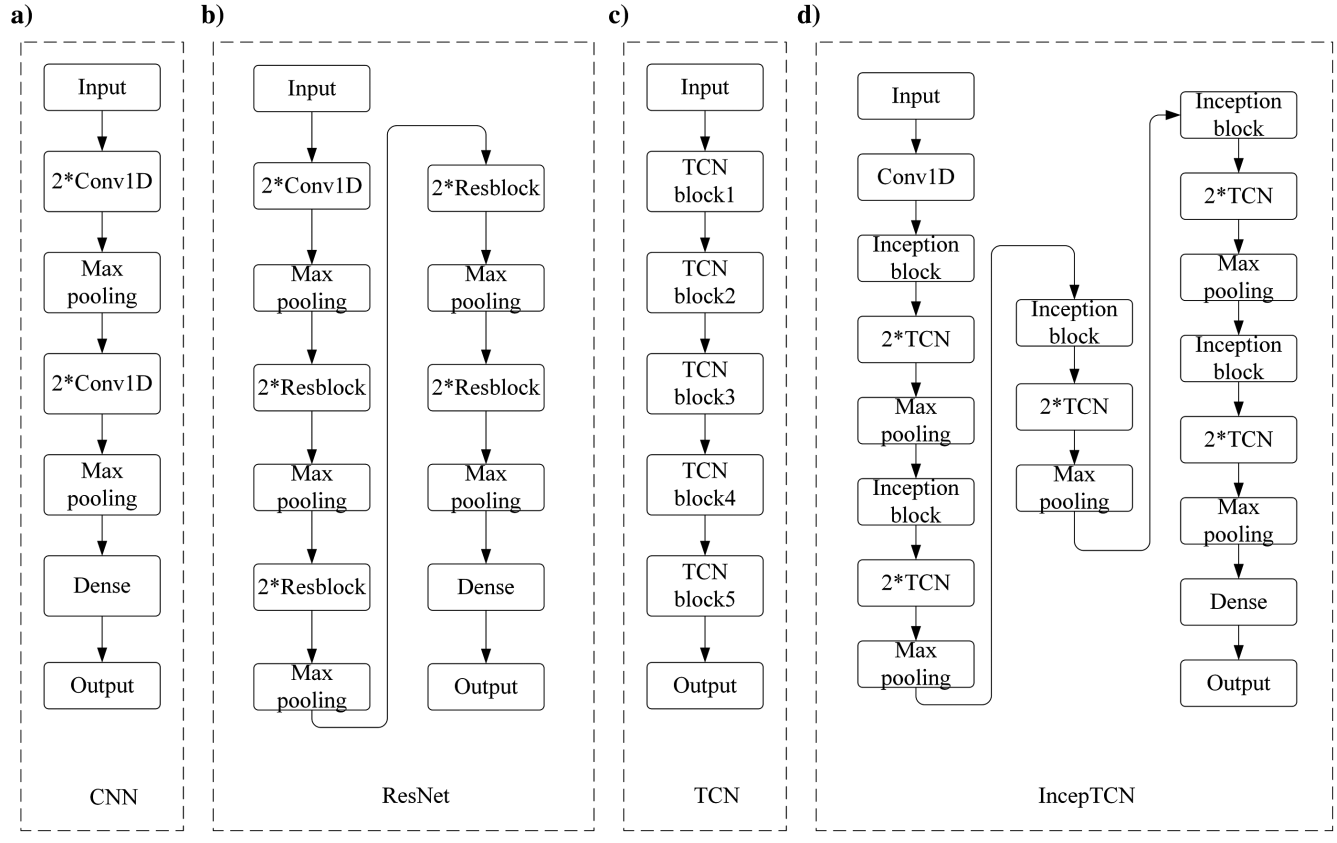

Figure 9. The architectures of four deep-learning networks. (a) CNN, (b) ResNet, (c) TCN, and (d) IncepTCN.

input layer, convolution layer, maximum pooling layer, dense layer, and output layer. It is simple and performs well in earthquake classification, MT data identification, and many other fields (Zhang et al., 2021). ResNet (Figure 9b) is another commonly used deep-learning classification algorithm. The difference between ResNet and CNN is that several residual blocks are added. It uses 1D convolution residual blocks to extract time series features. The number of convolution cores in the network increases gradually, and the size of convolution cores decreases in turn. It is one of the state-of-the-art models. As shown in Figure 9c, the nonimproved TCN network is composed of five TCN blocks. It gradually reduces the number of convolution cores (64, 32, 32, 16, and 6, respectively) and does not use a pooling operation to minimize the data length. The improved TCN network (Figure 9d) first extracts the spatial features by the inception block and then stacks two TCN blocks to extract the timing features of the signal. The improved TCN is consists of five inception blocks and 10 TCN blocks and we called it IncepTCN network.

Establishment of training sets

The intent of supervised machine learning is to learn features from a given pattern (training set) and then use that to predict an unknown pattern. Therefore, a given pattern largely determines the performance of machine learning. Due to the filtering effect of the earth system, the CSEM data observed at different locations are obviously different, and the cultural noise in different regions has different characteristics. It is a challenging task to accurately classify the observation data collected in different regions using the model trained from limited samples. We divide the observed signal into time series segments with a length of one cycle and label them to make samples. According to the requirements of secondary classification, the sample library contains noisy samples and relatively high-quality samples. The current research results (Li et al., 2021a) show that SISC can probably achieve high-precision signal-to-noise separation when the signal-to-noise ratio (S/N) is higher than 0 dB (i.e., the ratio of noise to signal amplitude is equal to one). In other words, SISC can reconstruct the remaining weak-noise segments with high precision as long as we use deep learning to eliminate the segments with S/N lower than 0 dB. To be conservative, we set this threshold to 0.75. That is, the basis for judging a noisy segment is that the amplitude of the noise reaches 0.75 times or more of the signal. There are essential differences between signal and noise, but the boundary between strong-noise segment and weak-noise segment is relatively fuzzy. Therefore, the classification standard described in this section increases the difficulty of machine-learning recognition. This is particularly difficult for unsupervised machine learning because it cannot learn rules from a given pattern.

To be closer to the actual situation, the samples are mainly selected from the measured data, which are collected in Huidong County, Sichuan Province. In addition, some theoretical samples are generated through numerical simulation, which makes the types of samples more abundant and diverse. The ratio of measured samples to simulated samples is approximately 6:1. To balance the ratio of strong-noise samples to weak-noise samples and compensate for the shortage of training samples, this paper uses the method of data expansion to increase the number of samples. In computer vision, the commonly used data augmentation methods include shift, mirroring, adding Gaussian white noise, and clipping (Zhang et al., 2021). Because the controlled-source signal is a 1D time series, we use the shift method to expand the samples. That is, in the time series, different samples are intercepted by sliding with a certain length of the window, and there is some overlap of time series between adjacent samples. This is similar to SISC dictionary learning. They make use of the invariance of shift. Because the effective signals collected at different sites have different amplitudes, we uniformly normalize the samples. The purpose of normalization is to limit the preprocessed data to a certain range (such as [0, 1] or [−1, 1]), so as to reduce the adverse effects caused by singular sample data. Our normalization method is as follows. Let $X = \{x_1, x_2, \ldots, x_M\}$ be the sample to be normalized, M represents the length of the sample, x_m represents the mth element in the sample, and $m \in [1, M]$. The normalized element $x_{m\text{-norm}}$ is

$$x_{m-\text{norm}} = \frac{x_m - \min(X)}{\max(X) - \min(X)}. \tag{7}$$

The normalized sample is $X_{\text{norm}} = \{x_{1\text{-norm}}, x_{2\text{-norm}}, \ldots, x_{M\text{-norm}}\}$ and $x_{m-\text{norm}} \in [0, 1]$. For signals in different frequency bands, we have established different sample libraries (i.e., training sets). The training set for the 7-2 signal (group 2 of pseudorandom 7 frequency wave) contains 74,882 samples, including 54,305 high-quality samples and 20,577 noisy samples. The training set for the 7-3 signal (group 3 of pseudorandom 7 frequency wave) consists of 37,989 high-quality samples and 36,446 noisy samples, with a total of 74,435 samples. The controlled source used in WFEM is pseudorandom multifrequency signal. The main frequencies of the 7-2 signal are 1 Hz, 2 Hz, 4 Hz, 8 Hz, 16 Hz, 32 Hz, and 64 Hz. The sample length is 1200 sampling points when the sampling rate is 1200 Hz. The main frequencies of the 7-3 signal are 0.75 Hz, 1.5 Hz, 3 Hz, 6 Hz, 12 Hz, 24 Hz, and 48 Hz, and the sample length is 1600 at the sampling rate of 1200 Hz.

Figure 10 shows 16 typical samples in the sample library. The measured sample shown in Figure 10m is a high-quality sample because it is very similar to the simulated noise-free models (see Figure 10k and 10o). We cannot obtain ideal samples such as Figure 10c, 10g, 10k, and 10o without any noise because the earth system is equivalent to a filter (usually a low-pass filter). However, these ideal samples are indispensable because the observed signal may be infinitely close to the ideal situation when the transmitter-receiver distance is very small and there is no cultural noise. In fact, we find through testing that the existence of ideal samples makes the model more flexible when facing unfamiliar data sets. The high-quality samples (or weak-noise samples) described in this paper are relative. As long as the amplitude of the noise is not greater than 0.75 times the effective signal, we will judge it as a relatively high-quality sample. For example, the samples shown in Figure 10e and 10i are polluted by random noise and harmonic noise, respectively. Even so, the noise amplitude has not reached 0.75 times the effective signal, so we classify them into weak-noise samples. In fact, retaining as many samples as possible is beneficial to ensure the accuracy of subsequent dictionary learning. The sample shown in Figure 10d also is polluted by random noise but is defined as a noisy sample because its noise intensity obviously exceeds 0.75 times the effective signal. It can be seen from the typical sample diagram that the noise in the CSEM data includes random noise (such as Figure 10a, 10d, and 10e), harmonic noise (such as Figure 10i and 10j), square-wave noise (such as Figure 10f and 10p),

impulse noise (such as Figure 10b, 10j, 10l, and 10n), baseline-drift noise (Figure 10h), and many other types.

Model training

When the total number of samples in the library is large, the proportion of verification sets can be appropriately reduced so that more samples can be used for training. The total number of samples in our library exceeds 74,000. It is sufficient to use 10% of the samples as the validation set. We randomly divide the samples in the library into 90% for the training set and 10% for the validation set to train CNN, ResNet, TCN, and IncepTCN to obtain high-precision classification models. In classification tasks of deep learning, accuracy and loss often are used to evaluate the training results. The accuracy directly illustrates the proportion of accurate classification, that is, the higher the better. The loss value represents the degree of inconsistency between the predicted category and the true category, that is, the smaller the better. Categorical cross-entropy (CCE) is a loss function commonly used in classification tasks. Suppose the training set consists of N pairs: $\{(X_1, t_1), (X_2, t_2), \ldots, (X_N, t_N)\}$, where X_i denotes the ith input sample, $t_i \in R^C$ is the corresponding category label, and $y_i \in R^C$ is its output. The input sample X_i is classified as one of C categories. CCE is then defined as (Rusiecki, 2019)

$$E_{\mathrm{CC}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}(P_{ic}\ \log(y_{ic})), \tag{8}$$

where P_{ic} is a binary indicator function; it indicates whether the y_i belongs to the cth category. The target P_{ic} can be interpreted as true, and output y_{ic} as predicted probability distribution for y_i belonging to cth category. It can be seen from equation 8 that when $y_{ic} = 1$, then $E_{\mathrm{CC}} = 0$.

As shown in Figure 11, with the increase of epochs, the training accuracy of the four networks rises rapidly, and the loss value drops sharply. The accuracy increase shows that the model learns suitable weight parameters, and the extracted features are closer to the distinguishable time-series features. The decline of loss value implies that the model's adaptability to the data is increasing, and the error of the model is decreasing. Finally, all of them reach the convergence state, but the convergence speed of ResNet and IncepTCN is significantly faster than that of TCN and CNN. This indicates that ResNet and IncepTCN are likely to have better adaptability. The validation accuracy of CNN, ResNet, TCN, and IncepTCN is 0.951, 0.961, 0.955, and 0.965, respectively; the validation loss is 0.147, 0.125, 0.140, and 0.115, respectively. The validation accuracy of IncepTCN is 0.014, 0.004, and 0.010 higher than that of CNN, ResNet, and TCN, respectively; its loss value is 0.032, 0.010, and 0.025 lower than that of CNN, ResNet, and TCN, respectively. It is worth noting that IncepTCN has the highest accuracy and the lowest loss, which implies that it is probably the best model. During training, the batch size of the models is set to 128, the optimizer is Adam, the loss function is CCE, and the initial learning rate is set to 1×10^{-5}. In addition, the learning rate attenuation strategy is adopted: when the validation loss increases, the learning rate becomes half of the current learning rate; when the validation loss continues to decline, the learning rate remains unchanged. This strategy contributes to the convergence of the model.

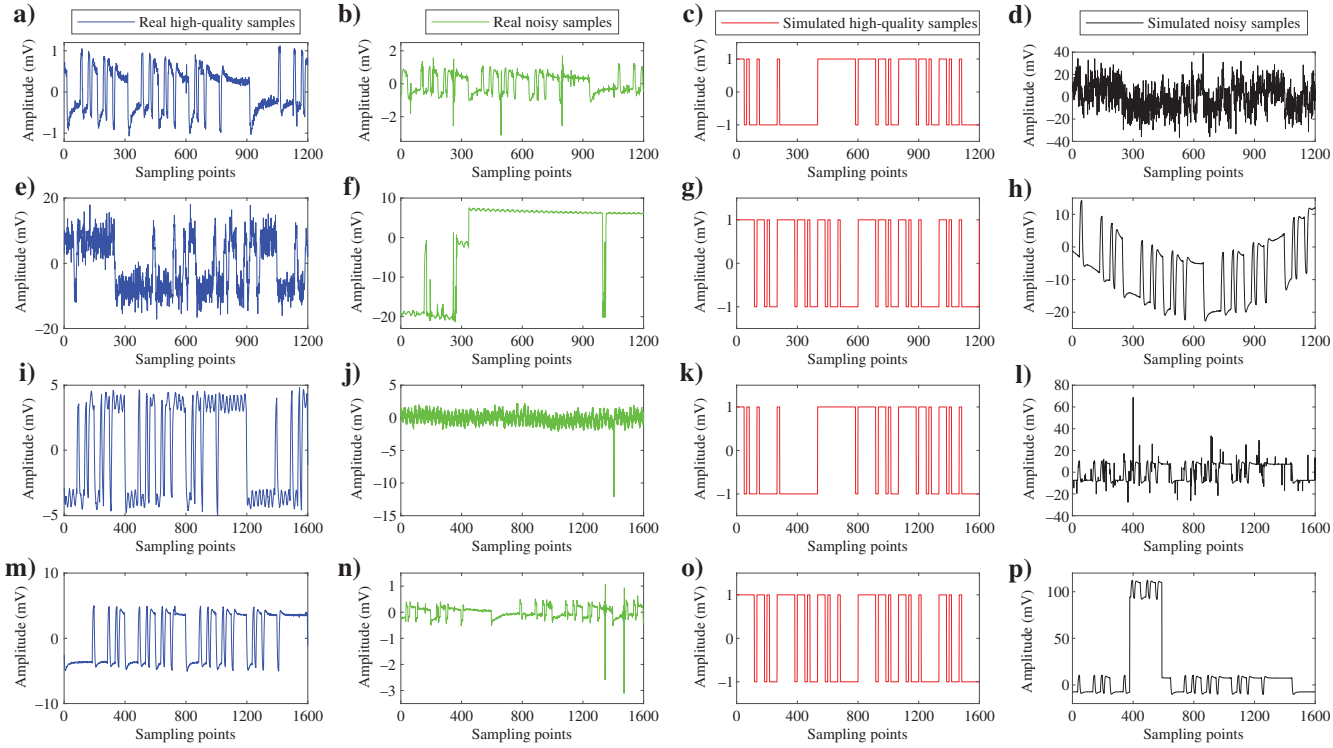

Figure 10. Typical samples in the library. (a, e, i, and m) Measured relatively high-quality samples, (b, f, j, and n) measured noisy samples, (c, g, k, and o) simulated noise-free samples, and (d, h, l, and p) simulated noisy samples, respectively. (a–h) The 7-2 samples, with a length of 1200 sampling points. (i–p) The 7-3 samples, with a length of 1600 sampling points.

Our training platform is an ordinary computer with an i9-10900K CPU, a 32 GB memory, and a 6 GB independent graphics card. The training time of CNN for the 7-2 and 7-3 data sets is 241.5 s and 301.2 s, respectively. The training time of ResNet for the 7-2 and 7-3 data sets is 485.1 s and 594.8 s, respectively. The training time of TCN for the 7-2 and 7-3 data sets is 778.3 s and 964.1 s, respectively. The training time of IncepTCN for the 7-2 and 7-3 data sets is 1648.2 s and 1685.6 s, respectively. Although the training of IncepTCN takes approximately half an hour, in practical application, it only takes less than 2 s to process a data set with a length of 460,800 sampling points, which will be shown subsequently.

Evaluation index of model

For the binary classification problem, there may be four kinds of results. First, noisy samples are mistakenly labeled as high-quality samples, which are called false-positive cases and recorded as P_F; second, noisy samples are identified as noisy samples, which are called true-negative cases and marked with N_T; third, high-quality samples are mistakenly identified as noisy samples, which are called false-negative cases and recorded as N_F; and fourth, high-quality samples are recognized as high-quality samples, which are called true-positive cases and recorded as P_T. Based on the preceding concepts, this paper comprehensively evaluates the model's performance through classification accuracy (A), precision (P), recall rate (R), and $F_{1\text{-SCORE}}$. These evaluation indicators are defined as follows (Huang et al., 2012; Ma et al., 2021):

$$A = \frac{P_T + N_T}{P_T + N_T + P_F + N_F}, \tag{9}$$

$$P = \frac{P_T}{P_T + P_F}, \tag{10}$$

$$R = \frac{P_T}{P_T + N_F}, \tag{11}$$

$$F_{1-\text{score}} = \frac{2 \times P \times R}{P + R}. \tag{12}$$

It can be seen from these equations that the higher the A, the higher the proportion of accurately identified samples. The higher the P, the less noisy samples are included in the high-quality sample set output by the model. If $P = 1$, the identified high-quality sample set does not include noisy samples. Here, R is used to measure the ability of the model to identify true-positive cases. A high recall rate means a low false recognition probability of high-quality fragments. If $R = 1$, all high-quality samples are accurately recognized. The $F_{1\text{-SCORE}}$ is the

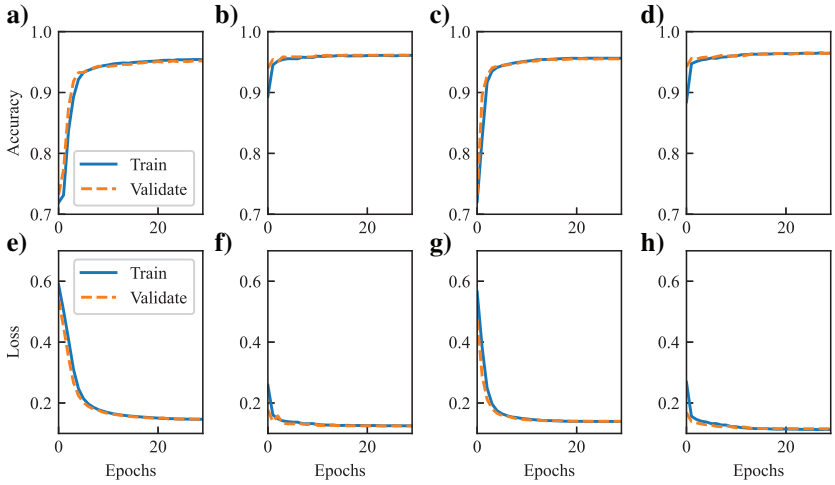

Figure 11. Accuracy curves of (a) CNN, (b) ResNet, (c) TCN, and (d) IncepTCN in training. (e–h) The loss curve corresponding to each model.

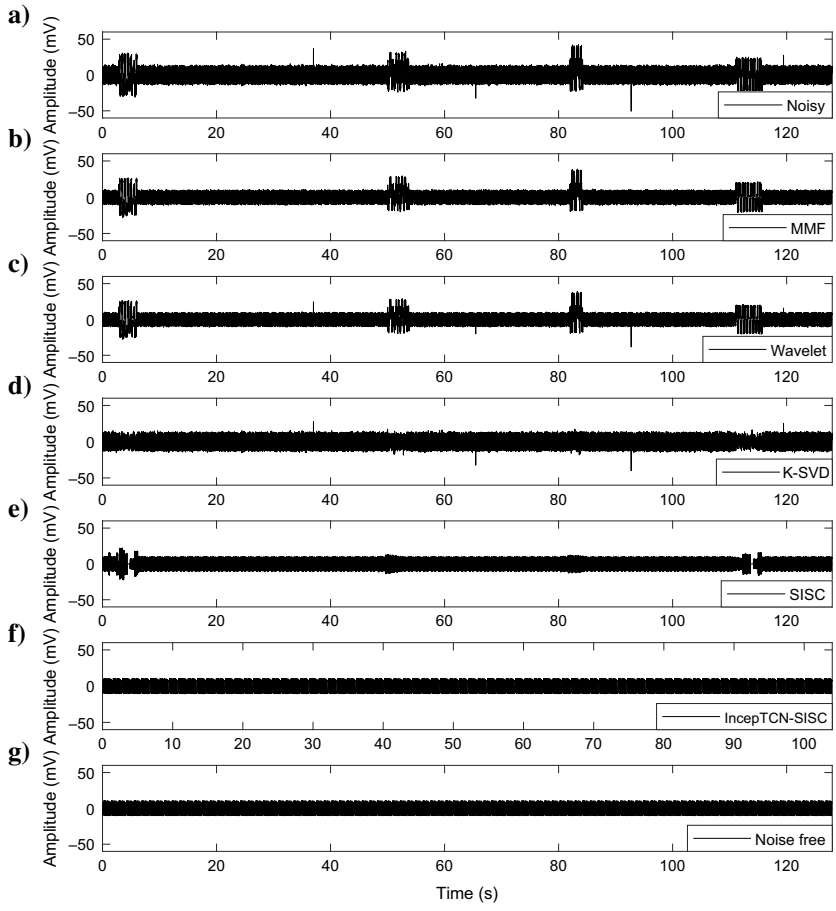

Figure 12. Noise elimination results from different methods. (a) Noisy signal, (b) denoised by MMF, (c) denoised by soft-threshold wavelet filtering (Wavelet), (d) denoised by K-SVD dictionary learning, (e) denoised by SISC dictionary learning, (f) denoised by the proposed IncepTCN-SISC method, and (g) noise-free data.

weighted average of P and R. Ideally, the preceding indicators are equal to one. For the problem described in this paper, our primary goal is to discard the strong-noise segments and second to retain high-quality segments as much as possible. Therefore, P is the most important, followed by R.

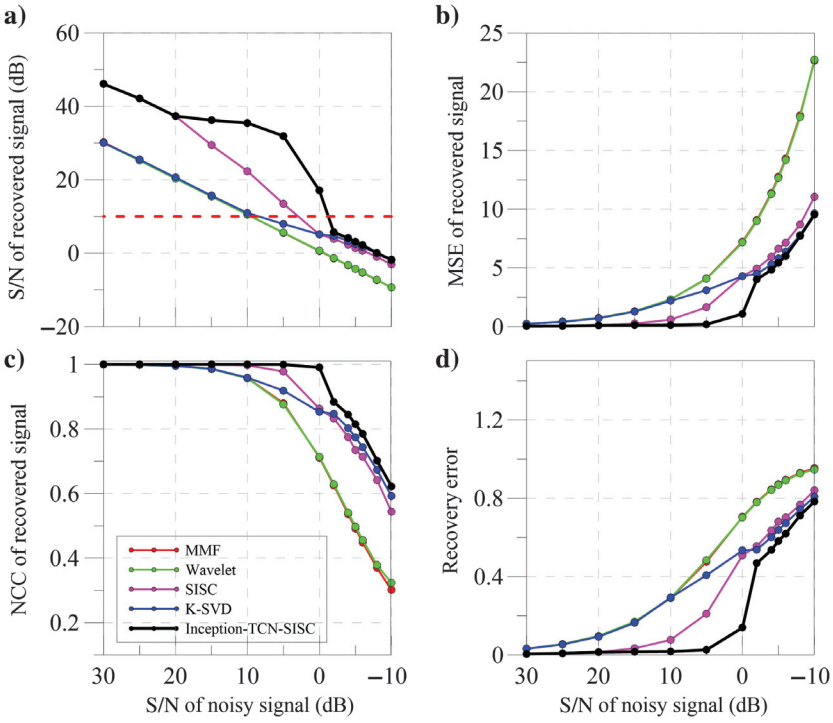

Figure 13. The (a) S/N, (b) MSE, (c) NCC, and (d) recovery error (E) of the signals recovered by different methods at different S/Ns.

SYNTHETIC DATA ANALYSIS

To test the performance of the proposed method, we add square-wave noise, impulse noise, and Gaussian noise to the noise-free data and then try to eliminate the noise via MMF, wavelet, K-SVD dictionary learning, SISC dictionary, and IncepTCN-SISC. Figure 12 shows the denoising effects of different methods when the noisy signal is 5 dB. MMF effectively removes Gaussian noise and impulse noise, but it fails to remove square-wave noise; wavelet only effectively eliminates Gaussian noise; K-SVD eliminates square-wave noise, but the removal effect of Gaussian noise and impulse noise is unsatisfactory; SISC accurately eliminates Gaussian noise and impulse noise, but the square-wave noise is not removed completely; and only the proposed IncepTCN-SISC method accurately eliminates all types of noise and obtains almost perfect recovery signal (see Figure 12g). After calculation with the noise-free signal, the S/N of the preceding five methods is 5.5 dB, 5.6 dB, 7.9 dB, 13.4 dB, and 31.9 dB, respectively. Our new scheme has great advantages.

To test the performance of different methods more comprehensively, we change the amplitude of noise and obtain the noisy signals under different S/Ns. Then, the preceding five methods are

Table 1. Classification accuracy of data sets collected in Huidong County, Sichuan Province.

	Methods				
	FCM	CNN	ResNet	TCN	IncepTCN
1	59.9%	94.8%	96.4%	95.1%	96.1%
2	75.8%	97.4%	98.4%	98.2%	99.2%
3	59.1%	98.4%	98.7%	98.7%	99.0%
4	89.6%	93.8%	96.1%	93.5%	97.4%
5	53.9%	90.2%	96.1%	96.5%	98.4%
6	88.5%	98.2%	98.7%	98.2%	98.2%
7	85.9%	95.8%	98.4%	96.6%	98.4%
8	82.0%	97.4%	98.2%	97.4%	99.2%
9	82.8%	96.9%	99.0%	98.4%	98.4%
10	71.9%	1	1	1	1
11	66.7%	85.4%	92.7%	90.6%	93.8%
12	63.5%	96.9%	99.5%	98.4%	99.0%
13	78.8%	88.2%	97.9%	91.0%	98.3%
14	51.4%	99.3%	99.3%	99.0%	99.3%
15	60.4%	93.8%	98.6%	95.5%	98.3%
16	80.9%	86.1%	98.3%	94.4%	97.2%
Average	71.9%	94.5%	97.9%	96.3%	98.1%

Table 2. Classification accuracy of data sets collected in Qiaojia County, Yunan Province.

	Methods				
	FCM	CNN	ResNet	TCN	IncepTCN
1	68.8%	94.3%	92.7%	96.4%	94.8%
2	83.9%	94.0%	97.1%	94.5%	97.9%
3	75.0%	85.4%	86.5%	91.3%	96.1%
4	79.9%	93.2%	96.3%	94.5%	98.2%
5	70.3%	87.5%	87.0%	88.5%	93.8%
6	75.0%	78.1%	88.0%	90.2%	90.1%
7	55.7%	96.4%	97.9%	94.8%	97.9%
8	61.5%	96.9%	98.4%	96.4%	99.0%
9	58.3%	95.8%	91.7%	87.5%	94.4%
10	70.1%	95.1%	89.6%	91.7%	93.8%
11	73.6%	88.9%	94.4%	93.8%	95.8%
12	69.4%	93.4%	81.9%	89.6%	82.3%
13	51.4%	96.9%	95.5%	94.4%	97.2%
14	68.4%	92.4%	97.2%	95.8%	96.5%
15	79.5%	92.7%	97.6%	93.1%	99.0%
16	79.7%	92.7%	93.2%	90.6%	91.7%
Average	70.0%	92.1%	92.8%	92.7%	94.9%

used to denoise it, and the results are evaluated by four indexes (Li et al., 2021a): S/N, mean-square error (MSE), normalized crosscorrelation (NCC), and recovery error (E). As shown in Figure 13, with the increase of noise amplitude, the denoising effect of all methods becomes worse as the S/N and NCC of their reconstructed signal gradually decrease, and the MSE and E increase progressively. However, the performance of IncepTCN-SISC decreases more slowly. When the S/N of the noisy signal is no less than 20 dB, SISC and IncepTCN-SISC perform almost the same and are significantly better than the other methods. Between 20 dB and 0 dB, IncepTCN-SISC has apparent advantages over SISC, but the effect of SISC is still considerably better than the other methods. When the S/N of the noisy signal is lower than 0 dB, the result of SISC is slightly worse than that of K-SVD, but IncepTCN-SISC is still the best among all methods. When the S/N of the noisy signal is −5.0 dB, the S/N of the signal reconstructed by IncepTCN-SISC still reaches 3.1 dB, and the NCC between the recovered signal and noise-free signal is as high as 0.81. As shown in Figure 13c and 13d, IncepTCN-SISC has obvious advantages in NCC and recovery error, indicating that its results are more reliable. If we regard the S/N of the reconstructed signal higher than 10 dB as the basis for successful reconstruction, the breakdown point of SISC is approximately 3 dB (see the dashed red line in Figure 13a). When the S/N of the noisy signal is lower than 3 dB, SISC reconstruction will fail. The breakdown point of IncepTCN-SISC is below 0 dB, which indicates that the introduction of IncepTCN reduces the lower limit of SISC denoising and solves the problem of poor performance of existing methods under low S/N to a certain extent.

It should be noted that different data have different breakdown points, which are related to the components of noise. Nevertheless, the conclusion that IncepTCN improves the adaptability of SISC will not change. Next, we will illustrate the advantages of the proposed method through practical cases.

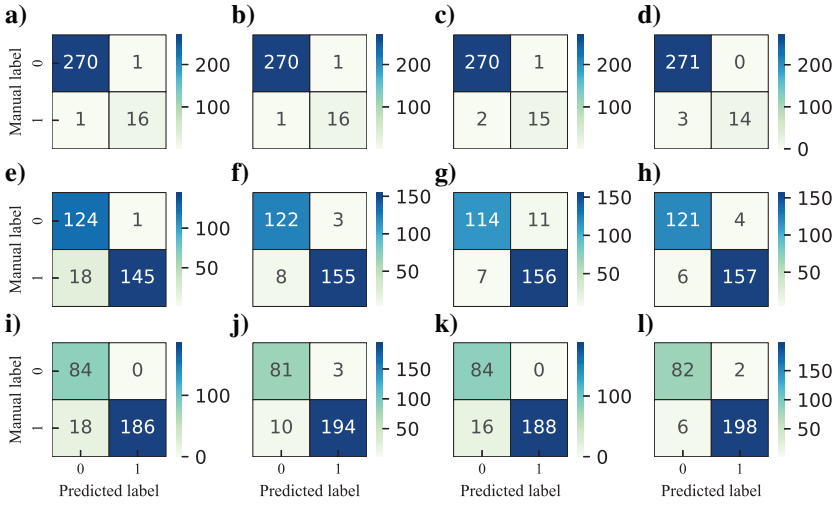

Figure 14. Confusion matrix of different models. The abscissa is the result of deep-learning prediction, and the ordinate is the result of manual marking. A label equal to zero indicates high quality and a label equal to one indicates noisy. Then, the upper-left, upper-right, lower-left, and lower-right parts of each subgraph represent the number of true-positive (P_T) samples, false-negative (N_F) samples, false-positive (P_F) samples, and true-negative (N_T) samples, respectively. The darker the color, the larger the value. The darker the color of the main diagonal, the higher the classification accuracy of the model. The cases are (a–d) signal dominated, (e–h) signal-noise balanced, and (i–l) noise dominated, respectively. The classification results are for (a, e, and i) CNN, (b, f, and j) ResNet, (c, g, and k) TCN, and (d, h, and i) IncepTCN, respectively.

REAL CASE STUDIES

Classification experiments with different methods

To test the performance of IncepTCN in practical application, we first apply the trained CNN, ResNet, TCN, and IncepTCN models and FCM-based method (Li et al., 2021b) to the CSEM data collected in Huidong County, Sichuan Province. To ensure the random performance of the models, a total of 16 data sets collected in different stations and dates are tested in this experiment, including eight 7-2 data sets and eight 7-3 data sets. As shown in Table 1, FCM has the lowest average classification accuracy. CNN, ResNet, TCN, and IncepTCN all achieve good signal-to-noise recognition performance, and their average classification accuracy of the 16 data sets is 94.5%, 97.9%, 96.3%, and 98.1%, respectively. Obviously, IncepTCN has the highest accuracy.

To test the generalization ability of the models, we apply the preceding five methods to the CSEM data collected in Qiaojia County, Yunnan Province, and show the results of 16 data sets. Because there are no samples from Qiaojia in the training set, and the data collected in Qiaojia are more seriously polluted by cultural noise, this is a challenging task for the models. As shown in Table 2, in the face of data sets in unfamiliar areas, the accuracy of all methods has decreased. The average accuracy of FCM, CNN, ResNet, TCN, and IncepTCN is 70.0%, 92.1%, 92.8%, 92.7%, and 94.9%, respectively. IncepTCN is still the model with the highest accuracy.

Table 3. Statistics of classification results of each model under different noise conditions.

Methods	Signal dominated			Balanced			Noise dominated		
	Precision	Recall	$F_{1\text{-score}}$	Precision	Recall	$F_{1\text{-score}}$	Precision	Recall	$F_{1\text{-score}}$
CNN	99.6%	99.6%	99.6%	87.3%	99.2%	92.9%	82.4%	1	90.4%
ResNet	99.6%	99.6%	99.6%	93.8%	97.6%	95.7%	89.0%	96.4%	92.6%
TCN	99.3%	99.6%	99.4%	94.2%	91.2%	92.7%	84.0%	1	91.3%
IncepTCN	98.9%	1	99.4%	95.3%	96.8%	96.0%	93.2%	97.6%	95.3%

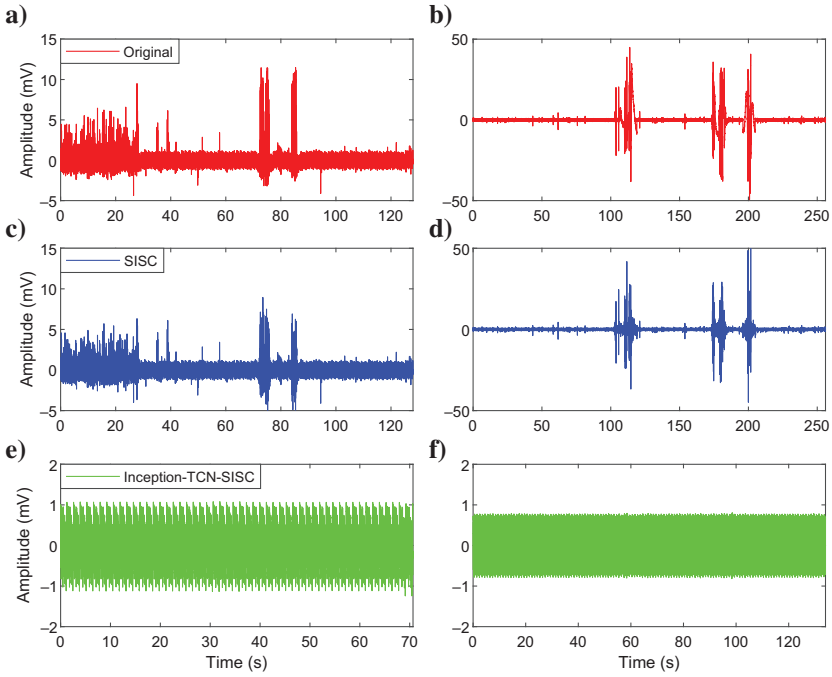

Figure 15. Field data sets of HD L1-3 (Ey) and HD L1-5 (Ey) collected in Huidong County, Sichuan Province. (a and b) The original time series, (c and d) the time series processed by SISC, and (e and f) the time series processed by IncepTCN-SISC, respectively. (a, c, and e) Data sets from station HD L1-3 and (b, d, and f) data sets from station HD L1-5.

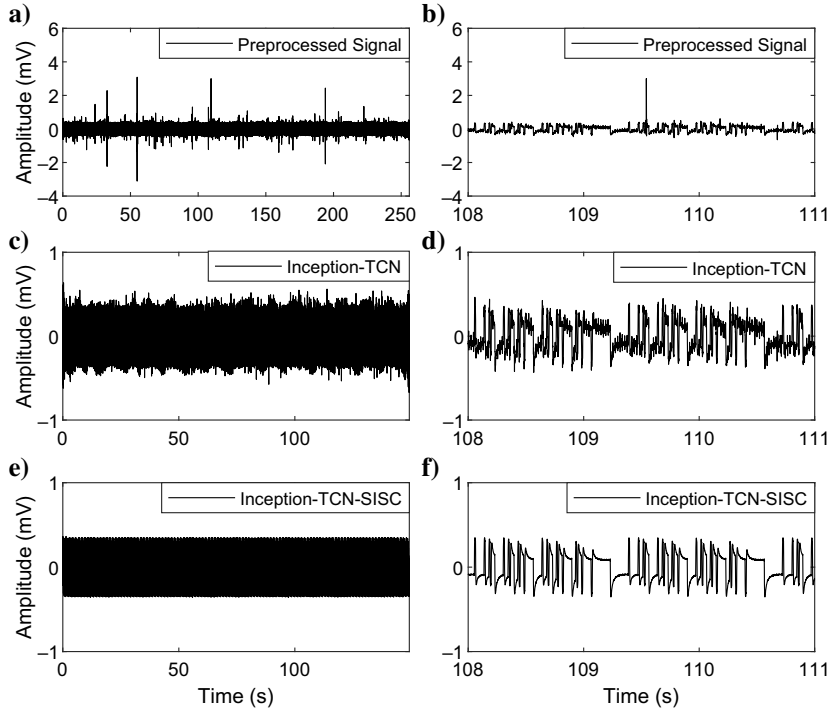

Figure 16. Field data sets of HD L1-6. (a and b) The preprocessed signal, (c and d) the signal recognized by IncepTCN, and (e and f) the signal obtained by IncepTCN-SISC, respectively. (a, c, and e) Complete time series and (b, d, and f) local fragments.

Among the 32 groups of data tested, the average classification accuracy of FCM, CNN, ResNet, TCN, and IncepTCN is 71.0%, 93.3%, 95.4%, 94.5%, and 96.5%, respectively. The accuracy of IncepTCN is 25.5%, 3.2%, 1.1%, and 2.0% higher than that of FCM, CNN, ResNet, and TCN, respectively. Although the average accuracy of IncepTCN is only 1.1% higher than that of ResNet, its performance in unfamiliar data sets is obviously better than ResNet. The advantage over ResNet in dealing with unfamiliar data sets has expanded from 0.2% of familiar data (data of Huidong) to 2.1% of unfamiliar data (data of Qiaojia). In practical application, an advantage of 2.1% may have a great impact. The experiment shows that IncepTCN has the best generalization ability and can calmly deal with fresh data sets. It should be noted that FCM is an unsupervised machine-learning algorithm. It has no training set and just aggregates the data into two categories, and the subsequent judgment is completed manually. In addition, some weak-noise samples are labeled as high quality in our training set, partly accounting for the low accuracy of FCM.

In the experiment, the length of a single data set is 460,800 sampling points (384 s). The time consumption of a single 7-2 and 7-3 data set in FCM processing is approximately 40 s and 74 s, respectively. TCN and IncepTCN are batch processing of 16 data sets. The time consumption of CNN in processing 7-2 and 7-3 data sets is 9.3 s and 11.8 s, respectively; the time consumption of ResNet is 12.1 s and 16.0 s, respectively; the time consumption of TCN is 11.9 s and 15.6 s, respectively; and the time consumption of IncepTCN is 23.0 s and 27.8 s, respectively. The average time consumption of IncepTCN for a single data set is no more than 2 s, which can significantly promote the practical application of the method.

To more accurately evaluate the performance of CNN, ResNet, TCN, and IncepTCN, we test the models with three measured data sets with different degrees of noise pollution and evaluate the results with precision, recall rate, $F_{1\text{-SCORE}}$, and confusion matrix. The ratios of noisy samples to weak-noise samples in the selected three data sets are 17/271 (signal dominated), 163/125 (signal-noise balanced), and 204/84 (noise dominated), respectively. A confusion matrix, also known as an error matrix, is a visual classification performance display tool, which can display the distribution of P_T, P_F, N_T, and N_F. The darker color of the main diagonal of the confusion matrix, the higher the accuracy (A) of the model. In addition, the accuracy, precision, recall rate, and $F_{1\text{-SCORE}}$ can be quickly calculated according to the confusion matrix. According to the confusion

matrix shown in Figure 14, we obtain the quantitative statistical results shown in Table 3.

For the case in which high-quality samples are dominant, all four models have achieved excellent performance, with the $F_{1\text{-SCORE}}$ surpassing 0.99. When the number of high-quality samples and noisy samples is balanced, IncepTCN and ResNet are the two best models because their $F_{1\text{-SCORE}}$ is obviously higher than the other two models. Although the recall rate of IncepTCN is slightly lower than ResNet, when determining precision, which is the most important indicator, IncepTCN is better than ResNet. When the noisy samples are dominant, the precision obtained by CNN and TCN is no more than 84.0%. This shows that the high-quality sample sets identified by CNN or TCN contain many noisy segments, which is unacceptable, especially when the observation data are polluted by intense noise. In contrast, IncepTCN still performs well in this case and achieves the highest precision of 93.2%. The $F_{1\text{-SCORE}}$ of IncepTCN is obviously higher than the other three models. The experiment shows that the IncepTCN has good robustness and obtains high precision and recall rate in different cases. Therefore, it once again validates that the IncepTCN network is an excellent model to solve the problems described in this paper.

Time series analysis

In the previous section, we experimentally demonstrate the excellent performance of the IncepTCN model. This section will combine the IncepTCN model and SISC to achieve a complete denoising process. As shown in Figure 15, these data sets, HD L1-3 (Ey) and HD L1-5 (Ey), suffer from very strong cultural noise. The amplitude of some noise reaches dozens of times that of the effective signal. Obviously, the S/N of these two data sets is lower than 0 dB. In these cases, it is difficult to obtain good results by wavelet, MMF, K-SVD, SISC, or other sparse decomposition methods. As shown in Figure 15c and 15d, SISC only attenuates some strong noise, which is not representative of high-precision noise separation. Usually, SISC should extract the effective signal first. In this case, SISC reconstruction fails due to the large-amplitude noises. After the processing of the proposed IncepTCN-SISC method (see Figure 15e and 15f), strong noise and weak noise are removed with high precision. Although the data length becomes shorter, a high-quality effective signal is finally obtained. The experimental results show that when the S/N of the observed data is significantly lower than 0 dB, the proposed IncepTCN-SISC method can still achieve high-precision signal-to-noise separation, which solves the problem that the existing method has difficulty with low S/N data below 0 dB.

As shown in Figure 16, the preprocessed signal is still polluted by strong spikes, regular impulsive noise, and random noise. After IncepTCN selection, the large spikes disappear, but the steady impulsive noise and random noise are still evident. After further processing with SISC, regular impulsive noise and random noise are accurately removed.

The data set QJ L1-12 (Ex) (see Figure 17) is polluted by a large number of strong impulse noise, harmonic noise, and some other random noise. After IncepTCN treatment, the fragments corrupted by strong impulse noise are discarded. Although there is still random noise and harmonic-like noise in the residual signal, their amplitude is relatively weak, and SISC can accurately remove these noises. As shown in Figure 17f, the resulting signal is very similar to the sample shown in Figure 10m. It shows that the data quality has been dramatically improved after IncepTCN-SISC processing.

Analysis of sounding curves

We calculate the WFEM apparent resistivity (He, 2018) and U/I curve with the raw time series, the time series processed by SISC, and the time series processed by IncepTCN-SISC and display the results of some data sets with serious interference. As shown in Figure 18, the WFEM apparent resistivity and U/I curves calculated by the raw data are seriously distorted, especially in the low-frequency part below 10 Hz. After being processed by SISC or IncepTCN-SISC, all the curves have improved greatly. However, it can be found that the curves of station HD L1-6 (Ey) still have a noticeable distortion after being processed by SISC, whereas the curves obtained by our new scheme are continuous and smooth. In the curves below 10 Hz of station HD L1-5 (Ey), SISC also is not as good as our method, especially approximately 8 Hz and 1 Hz. At 1 Hz of the station HD L1-6 (Ex), our process also shows apparent advantages. This proves that the introduction of deep learning dramatically improves the adaptability and noise suppression effect of SISC.

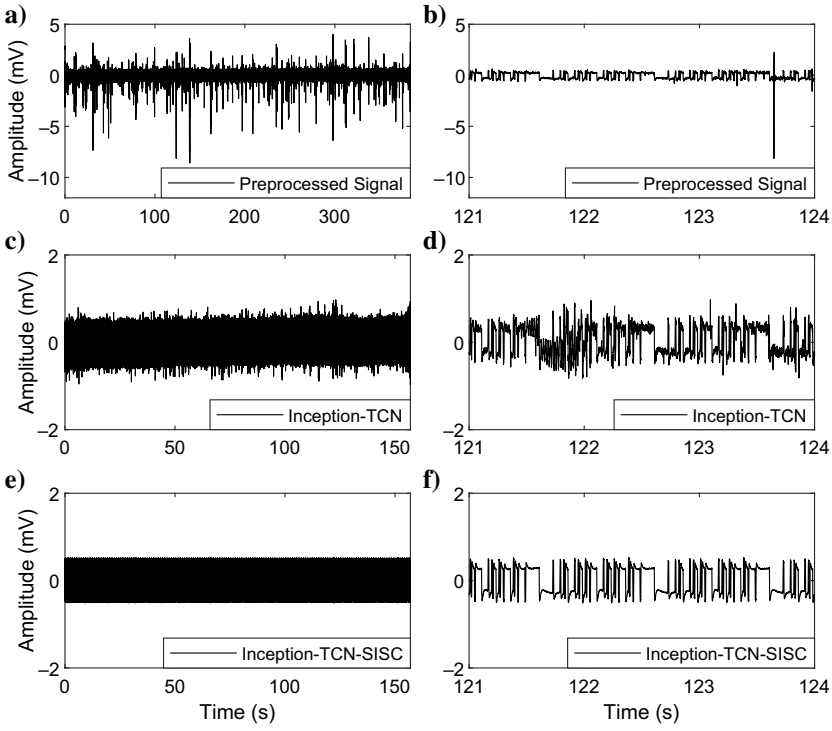

Figure 17. The same as Figure 16 but for field data set QJ L1-12 (Ex), collected in Qiaojia County, Yunnan Province.

DOI:10.1190/geo2022-0317.1

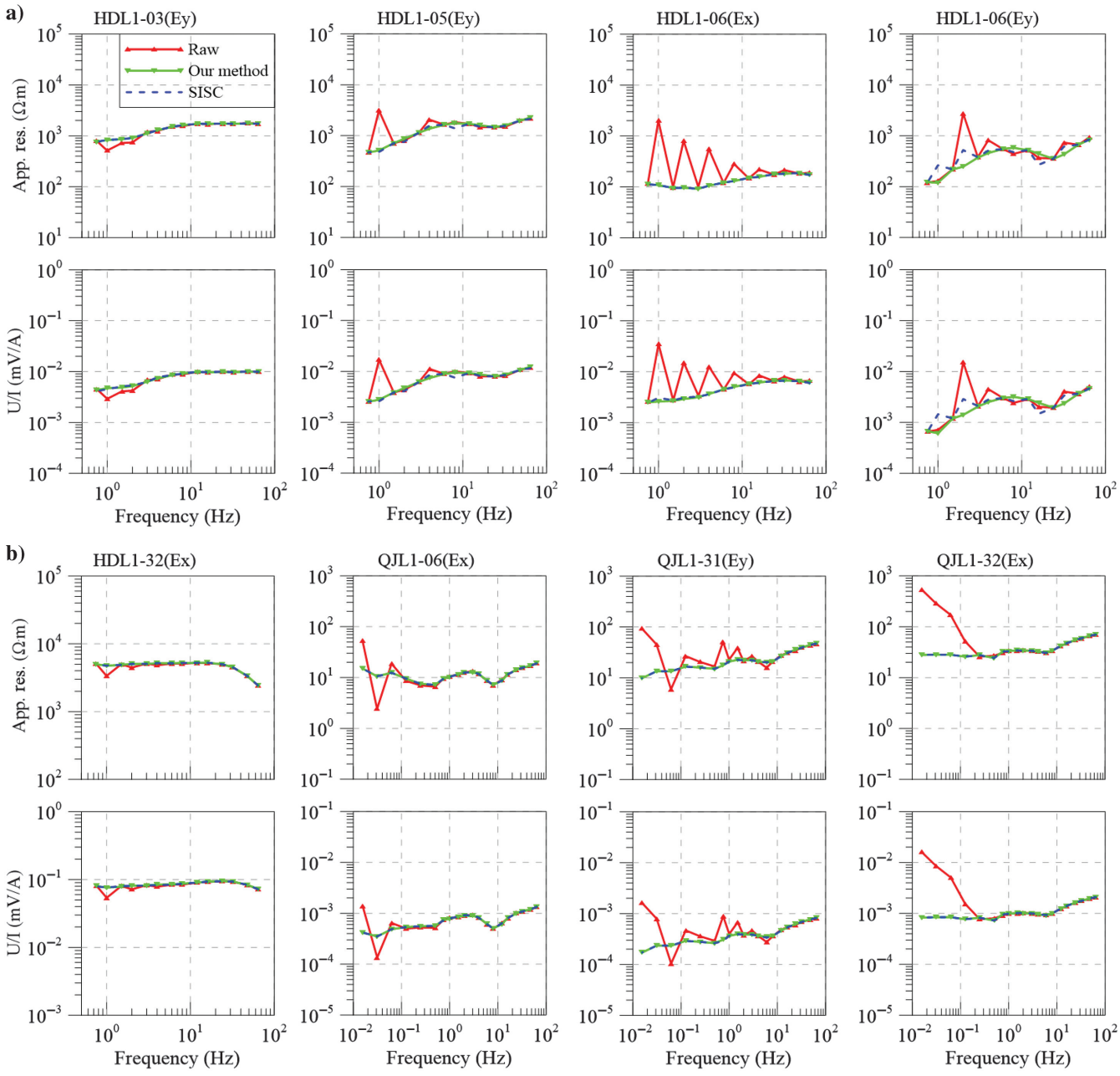

Figure 18. Apparent resistivity and U/I curves. The red curves with the upward-pointing triangle represent the results calculated from the noisy raw data, the green curves with the downward-pointing triangle represent the results obtained by our method, and the dotted blue lines indicate the results processed by the comparative SISC method.

CONCLUSION AND SUGGESTIONS

Deep learning has outstanding performance in identification, and dictionary learning has unparalleled advantages in sparse signal denoising. Based on the periodicity of CSEM data and the ingenious combination of deep learning and dictionary learning, this paper proposes a new noise elimination method called IncepTCN-SISC. In addition, we use the inception block to optimize TCN and propose a novel deep-learning network called IncepTCN. After a series of synthetic and field data processing experiments, the following understanding is obtained.

The proposed IncepTCN-SISC method makes full use of the respective advantages of deep learning and dictionary learning and solves the problem that the existing methods are difficult to use for noisy signals below 0 dB. It can improve the S/N of CSEM data from −5.0 dB to 3.1 dB or from 5.0 dB to 31.9 dB. Its performance is significantly better than soft-threshold wavelet filtering, MMF, K-SVD dictionary learning, and SISC dictionary learning. After IncepTCN-SISC processing, the initially distorted apparent resistivity curves will be corrected, become smooth and continuous, and the processing effect will be better than that of SISC. The proposed method is an intelligent method without any manual intervention

and is suitable for batch processing of CSEM data. This method can be directly transferred to any other periodic signal processing. In addition, with a little modification, it can be used to process other regular signals. For example, if IncepTCN is used to identify noisy signals, then SISC is used to improve the S/N of noisy segments instead of discarding them. Our method can be applied to GPR data denoising and even MT data denoising. Compared to shallow machine learning, deep learning has apparent advantages in accuracy and efficiency. The average recognition accuracy of IncepTCN is 25.5% higher than that of FCM, and the time consumption decreases from no less than tens of seconds for FCM to only a few seconds for IncepTCN. The average accuracy of IncepTCN is 3.2%, 1.1%, and 2.0% higher than that of the CNN, ResNet, and TCN network, and its comparative advantage is more apparent when dealing with unfamiliar data. This shows that the introduction of the inception block significantly improves the accuracy, robustness, and generalization ability of TCN. Our training set does not include samples of Qiaojia County, but the proposed method has achieved excellent results in Huidong and Qiaojia data sets, which shows that our method has good universality. It is foreseeable that as we continue to expand the sample library, the accuracy and adaptability of our method will continue to increase.

The length of data is shorter than that before IncepTCN-SISC processing. Usually, this will not cause adverse consequences because CSEM data are periodic. However, reducing the amount of geophysical data is a relative loss. Deep learning also has a good performance in signal-to-noise separation. Next, we will study the signal-to-noise separation of CSEM data based on deep learning, that is, to realize the mapping from noisy time series to high-quality time series. In addition, we will attempt to reconstruct the missing data using the deep-learning network.

Similar to LSTM, TCN is a recently proposed DNN for sequence modeling, including time series identification and prediction, but its application in geophysics is rarely reported. In addition, the results of this paper have proven that the introduction of the Inception module can significantly improve the accuracy, robustness, and generalization ability of TCN. Therefore, the algorithms presented in this paper have potential significance for other related applications, such as geophysics and biomedicine.

ACKNOWLEDGMENTS

We appreciate the editor-in-chief A. Cheng, associate editor A. Abubakar, assistant editor Y. Liu, reviewer K. MacLennan, and two anonymous reviewers for their valuable comments that greatly improved the paper. We thank W. Tang, C. Chen, and J. Fang for their language editing and suggestions. We also appreciate L. Zhang, S. Hu, X. Huang, and other field colleagues for their contributions in data acquisition. The research is financially supported by the National Natural Science Foundation of China (nos. 41904076, 42130811, and 42274085), the National Key R&D Program of China (no. 2018YFC0603202), China Postdoctoral Science Foundation (no. 2021M692987), Open Fund from Jiangxi Engineering Technology Research Center of Nuclear Geoscience Data Science and System (no. JETRCNGDSS202201), Open Fund from Badong National Observation and Research Station of Geohazards (BNORSG202208), and the Postgraduate Innovation Fund from East China University of Technology (grant no. DHYC-202230).

DATA AND MATERIALS AVAILABILITY

Data and code associated with this research are available at https://github.com/liguangCSU/Code-of-IncepTCN-SISC and can be obtained by contacting the corresponding author.

NOMENCLATURE

DNN = Deep neural network
CNN = Convolutional neural network
LSTM = Long short-term memory network
BP = Back propagation
DAE = Deep denoising autoencoder
RNN = Recurrent neural network
TCN = Temporal convolutional network
ResNet = Residual network
DCC = Dilated causal convolution
CCE = Categorical cross-entropy
MMF = Mathematical morphological filtering
FCM = Fuzzy C-means clustering
SVD = Singular-value decomposition
K-SVD = K-singular-value decomposition
SISC = Shift-invariant sparse coding
FFT = Fast Fourier transform
EMD = Empirical-mode decomposition
CEEMD = Complementary ensemble empirical-mode decomposition
CSEM = Controlled-source electromagnetic method
MCSEM = Marine controlled-source electromagnetic method
WFEM = Wide-field electromagnetic method
CSAMT = Controlled-source audio magnetotelluric method
TEM = Transient electromagnetic method
ATEM = Airborne transient electromagnetic method
IP = Induced polarization method
GPR = Ground-penetrating radar
S/N = Signal-to-noise ratio
MSE = Mean-square error
NCC = Normalized crosscorrelation

REFERENCES

Adrian, S., L. Léa, and J. Jakob, 2021, Automatic processing of time domain induced polarization data using supervised artificial neural networks: Geophysical Journal International, **224**, 312–325, doi: 10.1093/gji/ggaa460.

Bai, S., J. Z. Kolter, and V. Koltun, 2018, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling: arXiv preprint, doi: 10.48550/arXiv.1803.01271 [cs.LG].

Bang, M., S. Oh, K. Noh, S. J. Seol, and J. Byun, 2021, Imaging subsurface orebodies with airborne electromagnetic data using a recurrent neural network: Geophysics, **86**, no. 6, E407–E419, doi: 10.1190/geo2020-0871.1.

Blumensath, T., and M. Davies, 2006, Sparse and shift-invariant representations of music: IEEE Transactions on Audio, Speech, and Language Processing, **14**, 50–57, doi: 10.1109/TSA.2005.860346.

Candès, E. J., and M. B. Wakin, 2008, An introduction to compressive sampling: IEEE Signal Processing Magazing, **25**, 21–30, doi: 10.1109/MSP.2007.914731.

Cardarilli, G. C., L. D. Nunzio, R. Fazzolari, D. Giardino, A. Nannarelli, M. Re, and S. Spanò, 2021, A pseudo-softmax function for hardware-based high speed image classification: Scientific Reports, **11**, 15307, doi: 10.1038/s41598-021-94691-7.

Deo, R. N., and J. P. Cull, 2016, Denoising time-domain induced polarisation data using wavelet techniques: Exploration Geophysics, **47**, 108–114, doi: 10.1071/EG13077.

Fawaz, H. I., B. Lucas, G. Forestier, C. Pelletier, D. F. Schmidt, J. Weber, G. I. Webb, L. Idoumghar, P. Muller, and F. Petitjean, 2020, Inceptiontime:

DOI:10.1190/geo2022-0317.1

Finding AlexNet for time series classification: Data Mining and Knowledge Discovery, **34**, 1936–1962, doi: 10.1007/s10618-020-00710-y.
He, J. S., 2018, Combined application of wide-field electromagnetic method and flow field fitting method for high-resolution exploration: A case study of the Anjialing No. 1 Coal Mine: Engineering, **4**, 667–675, doi: 10.1016/j.eng.2018.09.006.
He, K., X. Zhang, S. Ren, and J. Sun, 2016, Deep residual learning for image recognition: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778, doi: 10.1109/CVPR.2016.90.
He, S., H. Cai, S. Liu, J. Xie, and X. Hu, 2021, Recovering 3D basement relief using gravity data through convolutional neural networks: Journal of Geophysical Research, **126**, e2021JB022611, doi: 10.1029/2021JB022611.
Hewage, P., A. Behera, M. Trovati, E. Pereira, and Y. H. Liu, 2020, Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station: Soft Computing, **24**, 16453–16482, doi: 10.1007/s00500-020-04954-0.
Hu, Y., D. K. Yang, Y. C. Li, Z. G. Wang, and Y. Lu, 2022, 3-D numerical study on controlled source electromagnetic monitoring of hydraulic fracturing fluid with the effect of steel-cased wells: IEEE Transactions on Geoscience and Remote Sensing, **60**, 4504210, doi: 10.1109/TGRS.2021.3100774.
Huang, H., J. Wang, and H. Abudureyimu, 2012, Maximum F1-score discriminative training for automatic mispronunciation detection in computer-assisted language learning: Interspeech, 815–818, doi: 10.21437/Interspeech.2012-248.
Jafari, M. G., and M. D. Plumbley, 2011, Fast dictionary learning for sparse representations of speech signals: IEEE Journal of Selected Topics in Signal Processing, **5**, 1025–1031, doi: 10.1109/JSTSP.2011.2157892.
Kang, Y., B. Yang, H. Li, T. Chen, and Y. Zhang, 2020, Deep spatio-temporal modified-inception with dilated convolution networks for citywide crowd flows prediction: International Journal of Pattern Recognition and Artificial Intelligence, **34**, 2052003, doi: 10.1142/S0218001420520035.
Kaur, H., S. Fomel, and N. Pham, 2021, A fast algorithm for elastic wave-mode separation using deep learning with Generative Adversarial Networks (GANs): Journal of Geophysical Research, **126**, e2020JB021123, doi: 10.1029/2020JB021123.
Lea, C., M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, 2017, Temporal convolutional networks for action segmentation and detection: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1003–1012, doi: 10.1109/CVPR.2017.113.
LeCun, Y., Y. Bengio, and G. Hinton, 2015, Deep learning: Nature, **521**, 436–444, doi: 10.1038/nature14539.
Li, D., F. Jiang, M. Chen, and T. Qian, 2022a, Multi-step-ahead wind speed forecasting based on a hybrid decomposition method and temporal convolutional networks: Energy, **238**, 121981, doi: 10.1016/j.energy.2021.121981.
Li, G., X. Gu, Z. Ren, Q. Wu, X. Liu, L. Zhang, D. Xiao, and C. Zhou, 2022b, Deep learning optimized dictionary learning and its application in eliminating strong magnetotelluric noise: Minerals, **12**, 1012, doi: 10.3390/min12081012.
Li, G., Z. S. He, J. Z. Deng, J. T. Tang, Y. Y. Fu, X. Q. Liu, and C. M. Shen, 2021b, Robust CSEM data processing by unsupervised machine learning: Journal of Applied Geophysics, **186**, 104262, doi: 10.1016/j.jappgeo.2021.104262.
Li, G., Z. S. He, J. T. Tang, J. Z. Deng, X. Q. Liu, and H. J. Zhu, 2021a, Dictionary learning and shift-invariant sparse coding denoising for controlled-source electromagnetic data combined with complementary ensemble empirical mode decomposition: Geophysics, **86**, no. 3, E185–E198, doi: 10.1190/geo.2020-0246.1.
Li, G., X. Xiao, J. T. Tang, J. Li, H. J. Zhu, C. Zhou, and F. B. Yan, 2017, Near-source noise suppression of AMT by compressive sensing and mathematical morphology filtering: Applied Geophysics, **14**, 581–589, doi: 10.1007/s11770-017-0645-6.
Li, J., Y. Liu, J. Tang, Y. Peng, X. Zhang, and Y. Li, 2023, Magnetotelluric data denoising method combining two deep-learning-based models: Geophysics, **88**, no. 1, E13–E28, doi: 10.1190/geo2021-0449.12022.
Li, J., Y. Liu, C. Yin, X. Ren, and Y. Su, 2020, Fast imaging of time-domain airborne EM data using deep learning technology: Geophysics, **85**, no. 5, E163–E170, doi: 10.1190/GEO2019-0015.1.
Liu, W. Q., R. J. Chen, H. Z. Cai, W. B. Luo, and A. Revil, 2017, Correlation analysis for spread spectrum induced polarization signal processing in electromagnetically noisy environments: Geophysics, **82**, no. 5, E243–E256, doi: 10.1190/GEO2016-0109.1.
Liu, W. Q., Q. T. Lü, R. J. Chen, P. R. Lin, C. J. Chen, L. Y. Yang, and H. Z. Cai, 2019, A modified empirical mode decomposition method for multi-period time-series detrending and the application in full-waveform induced polarization data: Geophysical Journal International, **217**, 1058–1079, doi: 10.1093/gji/ggz067.
Ma, H., C. Chen, Q. Zhu, H. T. Yuan, L. M. Chen, and M. L. Shu, 2021, An ECG signal classification method based on dilated causal convolution: Computational and Mathematical Methods in Medicine, **2021**, 6627939, doi: 10.1155/2021/6627939.
Maclennan, K., and Y. Li, 2013, Denoising multicomponent CSEM data with equivalent source processing techniques: Geophysics, **78**, no. 3, E125–E135, doi: 10.1190/geo2012-0226.1.
Mo, D., Q. Y. Jiang, D. Q. Li, C. J. Chen, B. M. Zhang, and J. W. Liu, 2017, Controlled-source electromagnetic data processing based on gray system theory and robust estimation: Applied Geophysics, **14**, 570–580, doi: 10.1007/s11770-017-0646-5.
Moghadas, D., 2020, One-dimensional deep learning inversion of electromagnetic induction data using convolutional neural network: Geophysical Journal International, **222**, 247–259, doi: 10.1093/gji/ggaa161.
Myer, D., S. Constable, and K. Key, 2011, Broad-band waveforms and robust processing for marine CSEM surveys: Geophysical Journal International, **184**, 689–698, doi: 10.1111/j.1365-246X.2010.04887.x.
Nittinger, C. G., and M. Becken, 2018, Compressive sensing approach for two-dimensional magnetotelluric inversion using wavelet dictionaries: Geophysical Prospecting, **66**, 664–672, doi: 10.1111/1365-2478.12605.
Reninger, P. A., G. Martelet, J. Deparis, J. Perrin, and Y. Chen, 2011, Singular value decomposition as a denoising tool for airborne time domain electromagnetic data: Journal of Applied Geophysics, **75**, 264–276, doi: 10.1016/j.jappgeo.2011.06.034.
Rita, S., B. Michael, and R. Oliver, 2013, Robust processing of noisy land-based controlled-source electromagnetic data: Geophysics, **78**, no. 5, E237–E247, doi: 10.1190/geo2013-0026.1.
Rusiecki, A., 2019, Trimmed categorical cross-entropy for deep learning with label noise: Electronics Letters, **55**, 319–320, doi: 10.1049/el.2018.7980.
Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, 2015, Going deeper with convolutions: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1–9, doi: 10.1109/CVPR.2015.7298594.
Wu, S. H., Q. H. Huang, and L. Zhao, 2021, De-noising of transient electromagnetic data based on the long short-term memory-autoencoder: Geophysical Journal International, **224**, 669–681, doi: 10.1093/gji/ggaa424.
Wu, X., G. Q. Xue, Y. M. He, and J. J. Xue, 2020, Removal of the multi-source noise in airborne electromagnetic data based on deep learning: Geophysics, **85**, no. 6, B207–B222, doi: 10.1190/geo2019-0555.1.
Wu, X., G. Q. Xue, P. Xiao, J. T. Li, L. H. Liu, and G. Y. Fang, 2019, The removal of the high-frequency motion-induced noise in helicopter-borne transient electromagnetic data based on wavelet neural network: Geophysics, **84**, no. 1, K1–K9, doi: 10.1190/geo2018-0120.1.
Xue, S. Y., C. C. Yin, Y. Su, Y. H. Liu, and H. F. Sun, 2020, Airborne electromagnetic data denoising based on dictionary learning: Applied Geophysics, **17**, 306–313, doi: 10.1007/s11770-020-0810-1.
Yang, Y., D. Q. Li, T. G. Tong, D. Zhang, Y. Zhou, and Y. Chen, 2018, Denoising controlled-source electromagnetic data using least-squares inversion: Geophysics, **83**, no. 4, E229–E244, doi: 10.1190/geo2016-0659.1.
Yang, Z., and J. Li, 2020, A short-term hybrid wind power prediction model based on singular spectrum analysis and temporal convolutional networks: Journal of Renewable and Sustainable Energy, **12**, 056101, doi: 10.1063/5.0007003.
Yu, S. W., and J. W. Ma, 2021, Deep learning for geophysics: Current and future trends: Reviews of Geophysics, **59**, e2021RG000742, doi: 10.1029/2021RG000742.
Zhang, L., Z. Ren, X. Xiao, J. Tang, and G. Li, 2022b, Identification and suppression of magnetotelluric noise via a deep residual network: Minerals, **12**, 766, doi: 10.3390/min12060766.
Zhang, P., X. Pan, and J. Liu, 2022a, Denoising marine controlled source electromagnetic data based on dictionary learning: Minerals, **12**, 682, doi: 10.3390/min12060682.
Zhang, P. F., M. Deng, J. N. Jing, and K. Chen, 2020, Marine controlled-source electromagnetic method data denoising based on compressive sensing: Journal of Applied Geophysics, **177**, 104011, doi: 10.1016/j.jappgeo.2020.104011.
Zhang, X., M. Zhang, and X. Tian, 2021, Real-time earthquake early warning with deep learning: Application to the 2016 M 6.0 central Apennines, Italy earthquake: Geophysical Research Letters, **48**, 2020GL089394, doi: 10.1029/2020GL089394.

Biographies and photographs of the authors are not available.