A denoising stacked autoencoders for transient electromagnetic signal denoising

Fanqiang Lin [1], Kecheng Chen [1], Xuben Wang [2,3], Hui Cao [2], Danlei Chen [1], and Fanzeng Chen [1]

[1]School of Information Science and Technology, Chengdu University of Technology, Chengdu,610059,China
[2]College of Geophysics, Chengdu University of Technology, Chengdu 610059,China
[3]Key Lab of Geo-Detection and Information Techniques of Ministry of Education, Chengdu 610059,China

Correspondence to: Kecheng Chen(15775118240@163.com)

Abstract.

Transient electromagnetic method (TEM) is extremely important in geophysics. However, the secondary field signal(SFS) in TEM received by coil is easily disturbed by random noise, sensor noise and man-made noise, which results in the difficulty in detecting deep geological information. To reduce the noise interference and detect deep geological information, we apply
5 autoencoders,an unsupervised learning model in deep learning, on the basis of analyzing the characteristic of SFS, to denoise SFS. We introduce SFSDSA, a Secondary Field Signal Denoising Stacked Autoencoders,based on deep neural networks of feature extraction and denoising.SFSDSA maps the signal points of the noise interference to the high probability points with clean signal as reference according to the deep characteristics of the signal, so as to realize the signal denoising and reduce noise interference.The method is validated by the measured data comparison, and the comparison results show that the noise
10 reduction method can effectively reduce the noise of SFS, in contrast with the Kalman and wavelet transform methods, and strongly support the speculation of deeper underground features.

1 Introduction

Through the analysis of SFS in TEM, the information of underground geological composition can be obtained and has been widely used in mineral exploration, oil and gas exploration and other fields (Danielsen et al.,2003, Haroon et al., 2015) Due
15 to the small amplitude of the late field data in the secondary field , it may be disturbed by random noise, sensor noise, human noise and other interference(Rasmussen et al., 2017) which leads to data singularities or interference points, and thus the deep geological information can not be reflected well.Therefore, it is necessary to make full use of the characteristics of the secondary field data to reduce the noise in the data and increase the effective range of the data.

Many methods have been developed for noise reduction of transient electromagnetic method. These methods can be broadly
20 categorised into three groups:

(1)Kalman filter algorithm(Ji et al.,2017)

(2)Wavelet transform algorithm(Ji et al.,2016, Li et al.,2017)

(3)Principal component analysis(PCA) (Wu et al.,2014)

Kalman filtering is an effective method in linear systems, but it has little effect in nonlinear fields such as transient electro-magnetic signals.The acquisition of wavelet threshold is cumbersome, and wavelet base selection is very difficult. In order to achieve the desired separation effect, to design an adaptive wavelet base is necessary.Likewise, the PCA algorithm is cumbersome too.

5 However,deep learning has been used to reduce noise from images, speech, and even gravitational waves(Jifara et al.,2017, Grais et al.,2017, Shen et al.,2017). Meanwhile,the representative model of deep learning Autoencoder(AE)(Bengio et al.,2006)has been successfully applied in many fields (Hwang et al.,2016).AE with noise reduction capability(Denoising Autoencoders,DAE)(Vincent et al.,2008) has been widely used in image denoising (Zhao et al.,2014), audio noise reduction (Dai et al.,2015), the reconstruction of holographic image denoising(Shimobaba et al.,2017) and other fields.

10 Nevertheless,in the field of geophysics, the application of deep learning model is limited (Chen et al., 2014) The use of deep learning model to reduce the noise of geophysical signals has not been applied.

Therefore, in this paper, the Secondary Field Signal Denoising Stacked Autoencoders (SFSDSA) is proposed to reduce noise, based on a deep neural network with SFS feature extraction.

SFSDSA will be affected by noise interference signal point according to the deep characteristics of the signal mapping to
15 the high probability of points by reference to SFS geophysical inversion signal, so as to realize the signal denoise and reduce noise interference.

2   Related Work

The aforementioned kalman filter and wavelet transform are universal traditional filtering methods,and have their own defects.

However, the SFS itself has distribution characteristics,and the distortion of the waveform generated by the noise causes
20 deviation from the signal point of the distribution.

The theoretical research indicates that (Bengio et al.,2006), the incomplete representation of autoencoders will be forced to capture the most prominent features of the training data, and the high order feature of data is extracted,so autoencoders can be applied to the feature extraction and abstract representation of SFS.

Theoretical research also shows that (Vincent et al.,2008)), Denoising Autoencoders (DAE) can map the damaged data
25 points to the estimated high probability points according to the data characteristics, and achieve the target of repairing the damaged data. Therefore, DAE can be appplied to map the SFS data points that will be disturbed by noise to the estimated high probability points, to achieve the purpose of SFS noise reduction.

Studies have found (Vincent et al.,2010)the stacked DAEs (SDAE) have a strong feature extraction capability, and can improve the feature extraction and enhance the ability of calibrating the deviation points disturbed by noise. SDAE is also
30 commonly used in the compression encoding of the pre-processing height of complex images (Ali et al.,2017).

Therefore, based on the study of the distribution characteristics of the Secondary Field data and autoencoder denoising method, we propose SFSDSA, a Secondary Field Signal Denoising Stacked Autoencoders, which is a deep learning model of transient electromagnetic signal denoising

(1)SFSDSA will be stacked by multiple AEs to form a deep neural network of multilayer owe complete coding, and multiple AEs are used as a higher-order feature extraction part, which can utilize its deep structure to maximize the characteristics of secondary field data.

(2)Based on the principle of DAE, SFSDSA will set the secondary field measured data (received data)as the input data, and geophysical inversion method is used to process the measured data of the secondary field to obtain the inversion signal as the clean signal data. SFSDSA maps the signal points of the noise interference to the high probability points with clean signal as reference according to the deep characteristics of the signal. Because maintaining the original data dimension is especially important for the undistorted and post-processing of the signal, it is necessary to set the original dimension after the last coding as the output layer dimension. Although the output method may produce the decoding loss, it can have high abstract retention of the secondary field data characteristics, and map the affected signal points to the high probability position points.

(3)Adam algorithm(Kingma et sl., 2014), introducing regularized loss and scaled exponential linear units(SELU) (Klambauer et al., 2017)as activation function, are applied to SFSDSA model, to solve the problem of over-fitting due to increased depth and the problem of too many dying nodes causing the signal to be unusable, and the problem of SFSDSA only learning an identity function.

3 Mathematical Derivation of SFSDSA

Firstly, the secondary field data(actual detection signal)are treated as a noisy input. Since the secondary field data are mainly a time-amplitude value, we can sample the signal as a point-amplitude value, in the form of matrix A, the dimensions are $1 \times N$:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n-1} & a_{1n} \end{bmatrix} \tag{1}$$

Secondly, the geophysical inversion method is used to obtain the theoretical signal, which can be used as a clean signal, then the theoretical signal is sampled as point-amplitude value, in the form of matrix $\tilde{A}$, the dimensions are $1 \times N$:

$$\tilde{A} = \begin{bmatrix} \tilde{a}_{11} & \tilde{a}_{12} & \cdots & \tilde{a}_{1n-1} & \tilde{a}_{1n} \end{bmatrix} \tag{2}$$

Thirdly, SFSDSA training model can be built, and Adam, which is a stochastic gradient descent (SGD) method, is applied to prevent gradient disappearance, and regularization loss is used to prevent over-fitting and SELU activation function is utilized to prevent too many points of depth.

$$g_\theta(a_{1n}) = f_{SELU}(Wa_{ln} + b) \tag{3}$$

$$g_\theta(a_{1n}) = \lambda \begin{cases} Wa_{1n} + b & a_{1n} > 0 \\ \alpha e^{Wa_{1n}+b} - \alpha & a_{ln} <= 0 \end{cases} \tag{4}$$

Where $\theta = (w, b)$, w denotes the $N \times N'$ parameter matrix $(N' < N)$, b denotes the offset of the N' dimensions. After the first compression coding layer, the signal is extracted features to $1 \times N'$. In order to extract high-level features while removing as much noise as possible and other factors, we can compress again.

$$5 \quad g_{\theta'}(a'_{1N'}) = \lambda \begin{cases} W' a'_{1N'} + b' & a'_{1N'} > 0 \\ \alpha e^{W' a'_{1N'} + b'} - \alpha & a'_{1N'} <= 0 \end{cases} \tag{5}$$

w denotes the $N' \times N''$ parameter matrix $(N'' < N')$, and b denotes the offset of the N'' dimensions, and features of actual detection signal is extracted again, after more feature extraction layers can be stacked. For the secondary field signal, it is necessary to maintain the same input and output dimensions to ensure that the signal is not distorted and later processed. When feature extraction reaches to a certain extent, it is necessary to reconstruct back to input dimensions .

10   Reconstruction can be regarded as the process that the noisy signal points map back to the original dimensions after features being highly extracted .At the same time, reconstruction is the process of signal characteristic amplification. Finally output matrix $\bar{A}$ with the same dimensions as the inputs can be got :

$$\bar{A} = \begin{bmatrix} \bar{a}_{11} & \bar{a}_{12} & \cdots & \bar{a}_{1n-1} & \bar{a}_{1n} \end{bmatrix} \tag{6}$$

The output $\bar{A}$ we obtained can be used to get the loss from the clean signal $\tilde{A}$ using the loss function. The general loss function
15 has square loss, which is mostly used in the linear regression problem. However, the secondary field data are mostly non-linear, and absolute loss is used in this paper:

$$L(\bar{A}, \tilde{A}) = |\bar{A} - \tilde{A}| \tag{7}$$

In the meantime, regularization loss optimization is used in this paper, in order to avoid over-fitting model, then:

$$loss = \theta^{*}, \theta'^{*} = arg_{\theta,\theta'} min \frac{1}{n} \sum_{i=1}^{n} L(x^{i}, g_{\theta'}(f_{\theta}(x^{i}))) + \lambda R(w) \tag{8}$$

20 After the loss is calculated, Adam algorithm is used to reverse optimization of parameters.

    Figure 1 is the algorithm structure diagram of SFSDSA.With reference to the theory of DAE, SFSDSA maps the signal points of the noise interference to the high probability points with clean signal as reference according to the deep characteristics of the signal, so as to realize the signal noise and reduce noise interference. This high probability position is determined by the theoretical clean signal and the multi-layer model and the feature extraction ability.The multi-layer feature extraction makes
25 the deep feature of secondary field data be preserved, and the effect of noise is reduced.

4   Experiment and Analysis

In this paper, the secondary field signal of a certain place is used as the experimental analysis signal. Usually, the secondary field signals can be obtained continuously for a period of time, so a large number of signals can be extracted conveniently as the training samples.
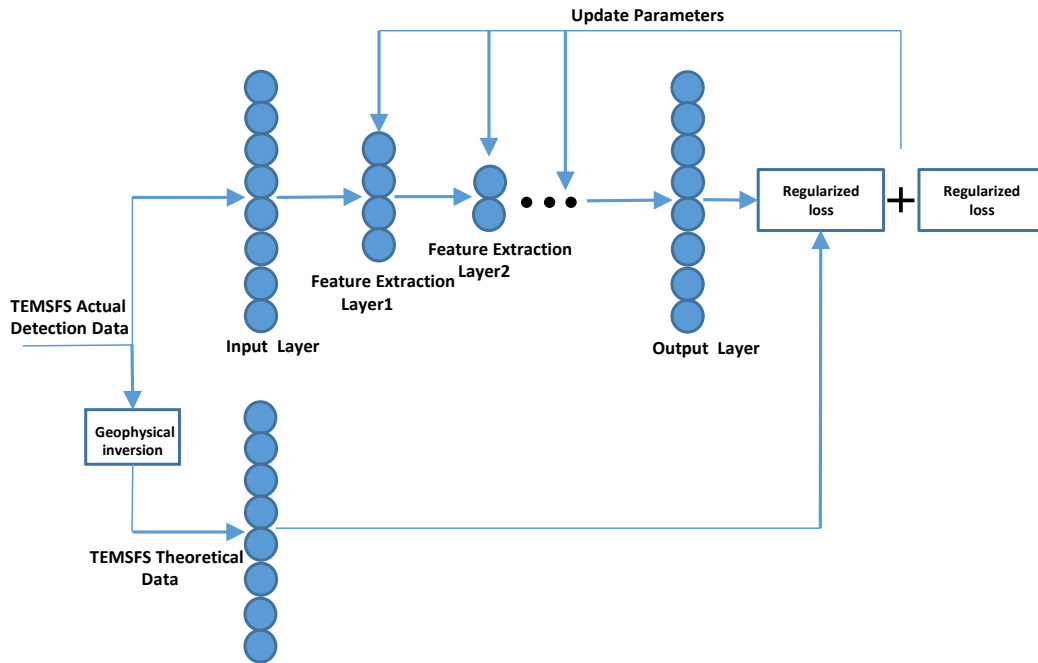
Figure 1. The flow chart.

The secondary field actual signals are extracted as 1×434 as input signals of noise pollution, as is shown in the Figure 2(a).At the same time, based on the secondary field actual signals, the geophysical inversion method is used to obtain the theoretical detection signal as clean signal uncontaminated by noise, as is shown in the Figure 2(b).

In order to be able to highlight the differences between the data, data are expressed in a double logarithmic form(loglog), as
5 is shown in the Figure 3(a) and Figure 3(b).

The deep features of original data are abstracted by features extraction layers(compression coding layers). As the number of layers increases, SFSDSA can be a more complex abstract model with limited neural units, (to get higher-order features for this small-scale input in this paper), and the more features extraction layers will inevitably lead to over-fitting. Moreover, the reconstruction effect can be affected by the number of features extraction layer nodes. If SFSDSA model has too few nodes,
10 the characteristics of the data can not be learned well. However, if the number of features extraction layer nodes are too large, the designed lossy compression noise reduction can not be achieved well and the learning burden is increased.

Therefore, based on the aforementioned questions, we design the SFSDSA model(Figure 1), and the number of nodes in the latter features extraction layer is half the number of nodes in the previous features extraction layer, until finally reconstructed
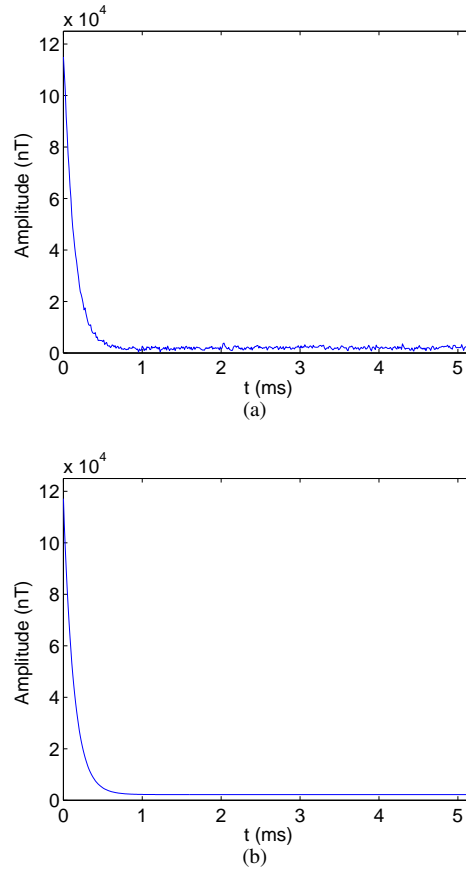
Figure 2. (a)Actual signal curves. (b) Inversion of theoretical signal curves.

back to the original dimension. SFSDSA model is a layer-by-layer features extraction, which can be regarded as a stacked AE process. Low dimensions are represented by the high-dimensional data features, which can learn the input features.

At the same time, since the reconstruction loss is the loss of the output related with the clean signal, it can also be said that the input signal can be regarded as a clean signal based on the noise, the training measure of DAE model increases the

5 robustness of the model and reconstructs the lossy signal, and mapping the signal point to its high probability location can be viewed as a noise reduction process.

The evaluation index is the mean absolute error(MAE) of output reconstruction data and clean input data.The smaller the MAE, the closer the output reconstruction data is to the theoretical data.The model also performs better in noise reduction.

$$MAE(X,h) = \frac{1}{m}\sum_{i=1}^{m}|h(x^{(i)} - y^{(i)}| \tag{9}$$

10 In this paper, we use Google's deep learning framework, Tensorflow.By analyzing Figure 4, the relationship between MAE and the number of hidden layers, we found that MAE was the smallest when the hidden layer was 2.
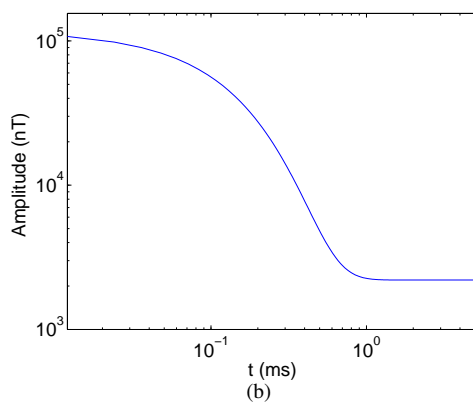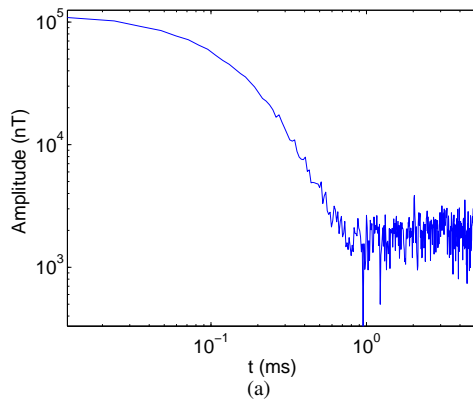
Figure 3. (a)Actual detection signal. (b) Inversion of theoretical detection signal.
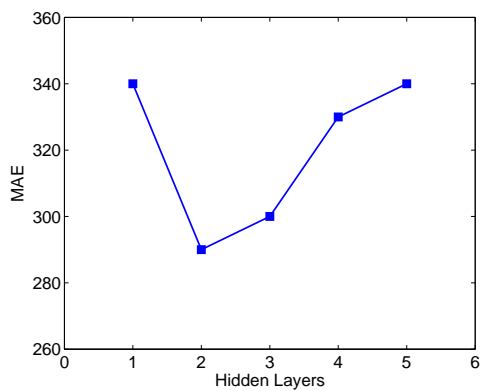
Figure 4. SFSDSA hidden layer number and MAE values.

Small-scale deep learning model, and the training times can be less.By analyzing Figure 2(a), we found that because the amplitude of the tail of the actual signal is small, and the influence of the noise is significant, so the tail of the signal oscillates violently. As is shown in Figure 5, we set each batch to 10 and every 10 training steps output MAE value as a sampling point. We found that the MAE values near the 150th sampling point start to oscillate. Based on these two oscillation relationships, we
5   can make some conclusions that training and reverse update parameters, and reconstruction loss is oscillating, so learning will fall into the bottleneck.Meanwhile, after the feature extraction and noise reduction to a certain extent, the noise interference can not be completely removed, and the reconstruction can not completely present the clean signal, and it is only possible to map the signal points as high probability points as possible to reduce reconstruction loss.
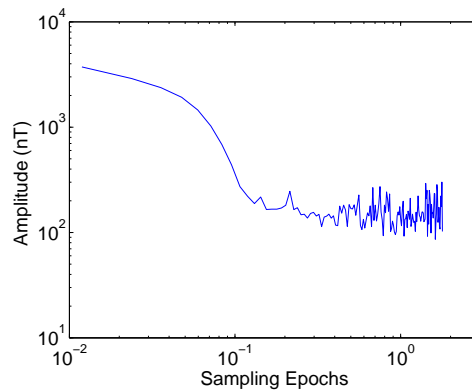
Figure 5. Error sampling point.

4.1   Training results
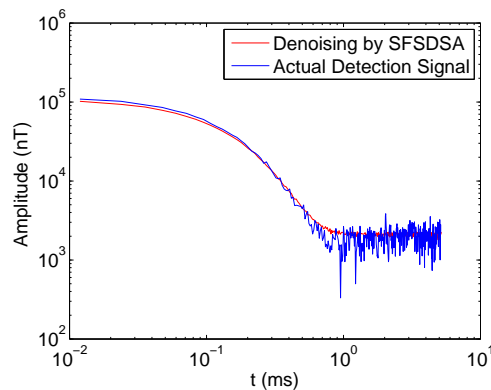
Figure 6. Actual secondary field data after SFSDSA model noise reduction

After several experiments, the MAE of actual signals fell from 534.5 to about 215.Compared with the secondary field actual signals and signals denoised by SFSDSA model, the noise reduction effect of SFSDSA is obvious in Figure 6.
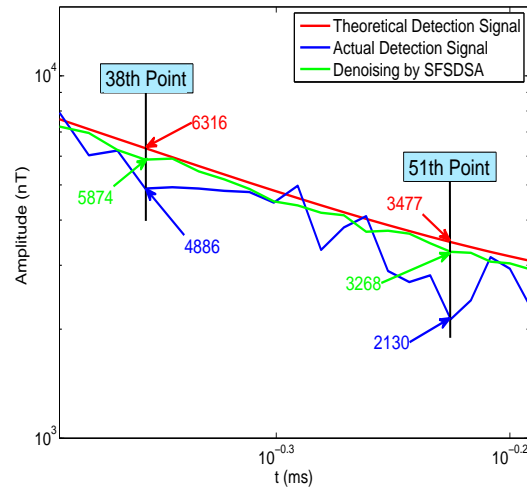
Figure 7. Specific points analysis.

The 35th to 55th points are selected for specific analysis in Figure 7 . Through noise reduction of training good SFSDSA model, the singular points(large amplitude deviation from theoretical signal) affected by the noise map to the high probability positions(e.g., no. 38, no. 51 points). This process is the process of damage reconstruction that the DAE model has verified. At the same time, our stacked AEs model also keeps on extracting the features, and the singular points are restored to the corresponding points according to the characteristics of the data. The whole process realizes the noise reduction of the secondary field actual signal based on the secondary field theoretical signal, and the model maps the singular points to locations where there is a high probability of occurrence, which is also similar to the most estimative method based on observations and model predictions by Kalman filtering.

5   Comparison with traditional noise reduction methods

We also conducted wavelet transform and Kalman filter experiments, in which the number of layers of the wavelet transform is three, DWT () and construction function IDWT () is called in Matlab.

Kalman filtering is implemented in Python, where the system noise Q is set to 1e-4 and the measurement noise R is set to 1e-3. The Figure 8 shows the absolute error distribution for that method. We can find from the figure model of noise reduction based on SFSDSA of secondary field data, SFSDSA is better than kalman filter and wavelet transform. At the same time as the kalman filter is a linear filter, its noise reduction effect is so poor in this paper.
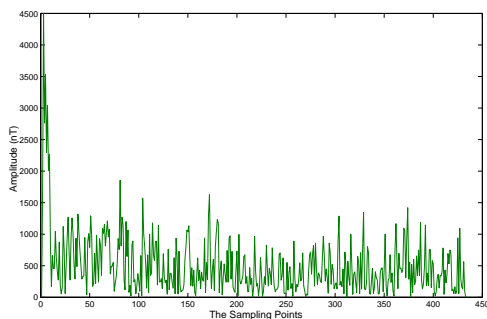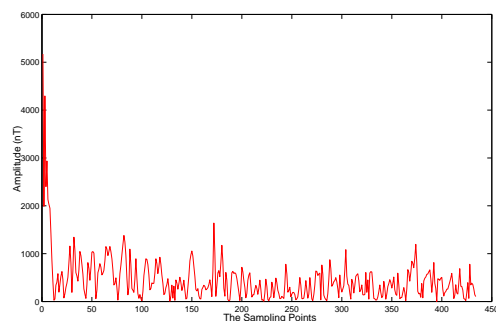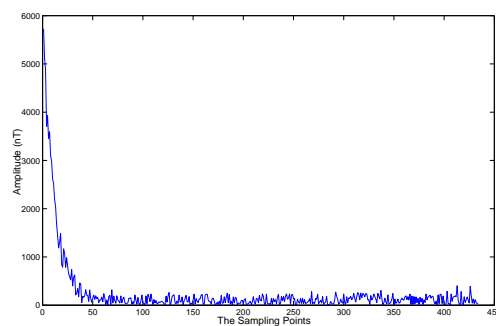
Figure 8. (a)Kalman filter. (b) Wavelet transform filter. (c) SFSDSA denoising.

Table 1. Comparison of MAE models

Model Name	Parameter Setting	MAE
SFSDSA+SELU+REGULARAZTION+ADAM	Learning_rate_base:0.00103	150.36
SFSDSA+RELU+REGULARAZTION+ADAM	Learning_rate_base:0.00103	1500.20
SFSDSA+SELU+ADAM	Learning_rate_base:0.00103	164.30
SFSDSA+SELU+REGULARAZTION	Learning_rate_base:0.00103	5112.30
Wave Transform	Three layers of wavelet transform	451.20
Kalman Filter	Q=e-4 R=e-3	503.20

At the same time, we compared the optimization results of various models using the traditional method with those of the SFSDSA model in table 1.
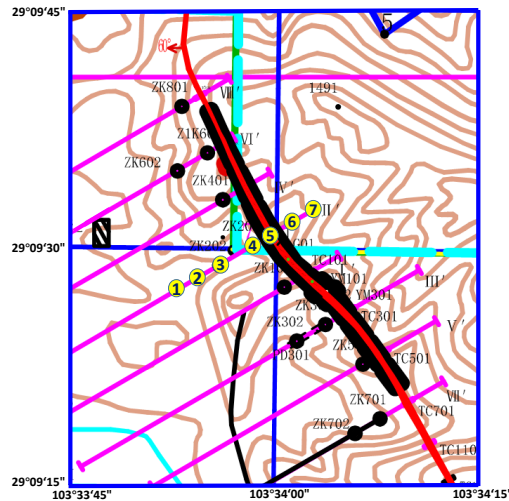
Figure 9. The geographic distribution of the collection points(1th to 7th).

Figure 9 is the diagram of the mine where the exploration experiment was conducted. The red thick curve is the actual mine vein curve. A data collection survey line, which is the southwest-northeast pink curve shown in the figure, is designed with seven points marked as number 1 to 7 along it, and the distance between each point is 50 meters.

In the data analysis, we analyzed the first 50 points in the second field which collected in actual mine. The early signal of the secondary field is stronger than later, and it is not easy be disturbed by the noises. So in the Figure 10, we take out the later 21 points in each collection point, which is used for further analysis. Figure 10 (a) is extracted time-domain order waveforms formed by the actual data acquired at the seven collection points at the same time.Figure 10 (b) extracted time-domain order waveforms formed by the data denoised by SFSDSA model. By comparing the two images in Figure 10, it can be clearly seen that the curves in Figure 10(a) have obvious intersections, and the intersections in Figure 10(b) can't be seen almost.In

(a)

(b)
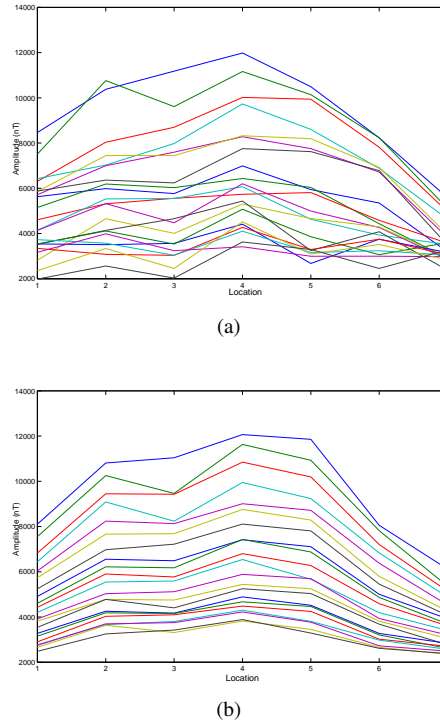
Figure 10. (a)The Original 30th to 50th points from seven actual detecting locations. (b)The denoising 30th to 50th points from seven actual detecting locations.

transient electromagnetic method, the intersected curve can't indicates the information of the deeper underground geological information.It can be explained that the curve after the de-noising model can reflect the deep geological information.

6 Conclusions

Based on the transient electromagnetic method, deep-level surveys are conducted. The deep-seated information is reflected in
5 the late-stage data of the second field. The signals are very weak and easily contaminated by noise. The measured data are modeled to obtain the theoretical model. By introducing the deep learning algorithm integrated with the characteristics of the secondary field data, it can map the contaminated data in late track data to a high probability position. By comparing several filtering algorithms, in which the same sample data are used, the stack noise reduction from the encoder method can reduce the MAE, thereby reducing the noise, and it is conducive to the subsequent pumping processing to further improve the effective
10 detection depth.

Code availability. The code of the SFSDSA are available upon request(15775118240@163.com or linfq@cdut.edu.cn)

Data availability. The data used in this paper are available upon request(15775118240@163.com or linfq@cdut.edu.cn)

Appendix A

A1

Competing interests. The authors declare that they have no conflict of interest.

5 *Acknowledgements.* This paper is supported by National Key R&D Program of China (NO.2018YFC0603300).

References

Danielsen J E, Auken E, Jørgensen F, et al. The application of the transient electromagnetic method in hydrogeophysical surveys, Journal of Applied Geophysics., 53(4),181-198,DOI: 10.1016/j.jappgeo.2003.08.004,2003.

Haroon, A., Adrian,J.,Bergers,R.,Gurk,M.,Tezkan,B., Mammadov,A.L.,Novruzov A. L.: Joint inversion of long-offset and central-loop transient electromagnetic data: Application to a mud volcano exploration in Perekishkul, Azerbaijan, Geophysical Prospecting.,63.2,478-494, DOI: 10.1111/1365-2478.12157,2014.

Rasmussen, S.,Nyboe, N. S.,Larsen J. J.:Extraction and Use of Noise Models from Production-Mode Transient Electromagnetic Data,Geophysics.,83(1),1-40,DOI: 10.1190/geo2017-0299.1,2017.

Ji, Y., Wu, Q., Wang, Y.,Lin,J.,Li,D.,Du,S.,Yu,S.,Guan,S.: Noise reduction of grounded electrical source airborne transient electromagnetic data using an exponential fitting-adaptive Kalman filter, Exploration Geophysics., DOI: 10.1071/EG16046 ,2017.

Ji, Y., Li, D., Yu, M.,Wang,Y.,Wu,Q.,Lin,J. :A de-noising algorithm based on wavelet threshold-exponential adaptive window width-fitting for ground electrical source airborne transient electromagnetic signal, Journal of Applied Geophysics., 128,1-7,DOI: 10.1016/j.jappgeo.2016.03.001,2016.

Li, D., Wang, Y., Lin, J.,Yu,S.,Ji,Y.: Electromagnetic noise reduction in grounded electrical-source airborne transient electromagnetic signal using a stationarywavelet- based denoising algorithm, Near Surface Geophysics.,1(-9),DOI: 10.3997/1873-0604.2017003 2017.

Wu, Y., Lu, C. D., Du, X. Z.,Yu,X.D.: A denoising method based on principal component analysis for airborne transient electromagnetic data, Computing Techniques for Geophysical and Geochemical Exploration.,36(2),170-176,DOI: 10.3969/j.issn.1001-1749.2014.02.08,2014.

Jifara, W., Jiang, F., Rho, S.,Liu,S.: Medical image denoising using convolutional neural network: a residual learning approach, Journal of Supercomputing.,The Journal of Supercomputing,6,1-15,DOI:10.1007/s11227-017-2080-0,2017.

Grais, E. M., Plumbley, M. D: Single Channel Audio Source Separation using Convolutional Denoising Autoencoders,Proceedings of the IEEE GlobalSIP Symposium on Sparse Signal Processing and Deep Learning, 5th IEEE Global Conference on Signal and Information Processing,14-16,DOI:10.1109/GlobalSIP.2017.8309164 ,2017.

Shen, H., George, D., Huerta, E. A.,Zhao,Z.: Denoising Gravitational Waves using Deep Learning with Recurrent Denoising Autoencoders,arXiv:1711.09919v1,2017.

Hwang, Y., Tong, A., Choi, J.:Automatic construction of nonparametric relational regression models for multiple time series,International Conference on International Conference on Machine Learning.3030-3039,2016.

Vincent, P., Larochelle, H., Bengio, Y.,Manzagol,P.A.: Extracting and composing robust features with denoising autoencoders, International Conference on Machine Learning., ACM, 1096-1103,DOI: 10.1145/1390156.1390294,2008.

Zhao, Mingbo.,Chow,T.W.S.,Zhang,Z.,Li,B.:Automatic image annotation via compact graph based semi-supervised learning,Knowledge-Based Systems., 76,148-165,DOI: 10.1016/j.knosys.2014.12.014,2015.

Dai,W.,Brisimi,T.S.,Adams,W.G.,Mela,T.,Saligrama,V.,Paschalidis,I.C.:Prediction of hospitalization due to heart diseases by supervised learning methods, International Journal of Medical Informatics.,84,189-197,2015.

Shimobaba T, Endo Y, Hirayama R, et al. Autoencoder-based holographic image restoration, Applied Optics.,56(13),F27,DOI: 10.1364/AO.56.000F27,2017.

Chen, B., Lu, C.D.,Liu, G.D.:A denoising method based on kernel principal component analysis for airborne time domain electro-magnetic data,Chinese Journal of Geophysics., 57(1), 295–302,DOI: 10.1002/cjg2.20087,2014.

Bengio, Y., Lamblin, P., Dan, P.,Larochelle,H.: Greedy layer-wise training of deep networks, International Conference on Neural Information Processing Systems., MIT Press,153-160,2006.

Vincent, P., Larochelle, H., Lajoie, I.,Manzagol,P.A.: Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion, Journal of Machine Learning Research.,11(12),3371-3408,2010.

5 Ali, A., Fan, Y., Shu, L.:Automatic modulation classification of digital modulation signals with stacked autoencoders, Digital Signal Processing., 71,108-116,DOI: 10.1016/j.dsp.2017.09.005,2017.

Kingma, D. P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint,arXiv:1412.6980,2014.

Klambauer,   G.,   Unterthiner,   T.,   Mayr,   A.,Hochreiter,   S.:   Self-Normalizing   Neural   Networks,   arXiv preprint,arXiv:1706.02515:1706.0251,2014.