



Original Paper

SegNet-based first-break picking via seismic waveform classification directly from shot gathers with sparsely distributed traces

San-Yi Yuan ^a, Yue Zhao ^a, Tao Xie ^b, Jie Qi ^c, Shang-Xu Wang ^{a,*}^a State Key Laboratory of Petroleum Resources and Prospecting, CNPC Key Laboratory of Geophysical Exploration, China University of Petroleum, Beijing 102249, China^b China Oilfield Services Ltd., Tanggu, Tianjin 300452, China^c University of Oklahoma, ConocoPhillips School of Geology and Geophysics, Norman, OK, USA

ARTICLE INFO

Article history:

Received 30 May 2020

Accepted 15 July 2021

Available online 23 October 2021

Edited by Jie Hao and Teng Zhu

Keywords:

First-break picking

Deep learning

Irregular seismic data

Waveform classification

ABSTRACT

Manually picking regularly and densely distributed first breaks (FBs) are critical for shallow velocity-model building in seismic data processing. However, it is time consuming. We employ the fully-convolutional SegNet to address this issue and present a fast automatic seismic waveform classification method to pick densely-sampled FBs directly from common-shot gathers with sparsely distributed traces. Through feeding a large number of representative shot gathers with missing traces and the corresponding binary labels segmented by manually interpreted fully-sampled FBs, we can obtain a well-trained SegNet model. When any unseen gather including the one with irregular trace spacing is inputted, the SegNet can output the probability distribution of different categories for waveform classification. Then FBs can be picked by locating the boundaries between one class on post-FBs data and the other on pre-FBs background. Two land datasets with each over 2000 shots are adopted to illustrate that one well-trained 25-layer SegNet can favorably classify waveform and further pick fully-sampled FBs verified by the manually-derived ones, even when the proportion of randomly missing traces reaches 50%, 21 traces are missing consecutively, or traces are missing regularly.

© 2022 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Seismic data have been commonly used to understand the stratigraphic structures and to predict the distribution of oil and gas in the depth of thousands of meters. The first arrivals correspond to first breaks (FBs) contain important feature information. For instance, they can be adopted to remove the effect of the shallow fluctuating weather layer, which is the cornerstone of the subsequent seismic data processing and interpretation (Yilmaz, 2001). Moreover, they can be utilized to estimate the source wavelet, which is the basis of the forward modelling and seismic inverse problem, and to build the shallow velocity model, which is useful for seismic migration and full waveform inversion (Yilmaz, 2001; Li et al., 2019; Liu et al., 2020; Yao et al., 2020). Consequently, to detect the first arriving waves or to pick FBs is helpful for geophysicists.

Manual picking is the most straightforward method. In addition,

it can introduce any prior knowledge, such as the spatial continuity of first arriving waves. Even for low-quality traces or missing traces, the brain networks can still interpret FBs by making use of the information at neighborhood traces. However, due to increasingly seismic traces, especially for wide-azimuth and high-density acquisition, manual picking is time-consuming and expensive. Various (semi-)automated first-break picking methods have been proposed and developed to improve the picking efficiency and to alleviate the pressure of interpreters. Each method has its own advantages and disadvantages. One traditionally automatically first-break picking technique is based on single-trace time- or frequency-domain methods operating on single- or multi-component recordings from an individual receiver level, such as energy based methods (Coppens, 1985; Sabbione and Velis, 2010), entropy based methods (Sabbione and Velis, 2010), fractal dimension based methods (Boschetti et al., 1996; Jiao and Moon, 2000), and higher-order statistics based methods (Yung and Ikelle, 1997; Saragiotis et al., 2004; Tselentis et al., 2012). In general, such methods can quickly and stably pick FBs on data with high signal-to-noise ratio (SNR). Compared with the time- or frequency-

* Corresponding author.

E-mail address: wangsx@cup.edu.cn (S.-X. Wang).

domain methods, time-frequency domain methods are developed to increase one time dimension or one frequency dimension, and have thus more potential to stably pick FBs in the case of relatively low SNR and to even detect weak signals or weak first-break waves (Galiana-Merino et al., 2008; Mousavi et al., 2016).

In contrast to single-trace (semi-)automated methods, multi-trace ones, such as cross-correlation based methods (Gelchinsky and Shtivelman, 1983; Molyneux and Schmitt, 1999) or template matching methods (Plenkers et al., 2013; Caffagni et al., 2016), make use of information on multiple receivers within the array simultaneously. In essence, multi-trace methods can pick FBs by taking maximum values of the cross-correlation or convolution results from trace(s) to trace(s). Therefore, the original multi-trace time-space amplitude information are generally utilized to pick FBs. Due to the usage of simultaneously multiple traces or high-dimension (2D or 3D) information, the multi-trace cross-correlation or template matching based methods can identify weak signals or pick FBs at low SNR (Gibbons and Ringdal, 2006; Caffagni et al., 2016) to some extent. However, they often do not adapt to the situations that the waveforms change with the trace and there exist missing traces or bad traces.

Most single- or multi-trace methods compute only one attribute for each time sampling point and subsequently select the locations with maximum- or minimum-value attributes as FBs. To improve the picking stability, some attributes including those in the time domain, frequency domain, time-frequency domain and/or time-space domain are considered to combine together to classify waveforms and to further pick FBs by artificial rules (Gelchinsky and Shtivelman, 1983; Akram and Eaton, 2016; Khalaf et al., 2018) or the traditional fully-connected artificial neural networks (ANNs) (McCormack et al., 1993; Gentili and Michelini, 2006; Maity et al., 2014). Multi-attribute first-break picking approaches based on artificial rules are straightforward and explicable. However, they increase the number of parameters that need to be tuned, when multi-attribute generators are chosen. For multi-attribute approaches based on the fully-connected ANNs, they can automatically analyze and combine multiple attributes extracted from the data at the cost of learning a large number of network parameters. Nevertheless, they usually require extra attribute extraction and attribute optimization.

Convolutional neural networks (CNNs), which is one developed type of ANNs, have become popular in recent years by breaking previous records for various classification tasks especially in image and speech processing (LeCun et al., 2015; Russakovsky et al., 2015; Goodfellow et al., 2016; Silver et al., 2016; Esteva et al., 2017; Perol et al., 2018). At the cost of training big data, CNNs have the advantages to automatically extract features or attributes and meanwhile classify data. Moreover, they can be designed deep, due to the properties of local perception and weight sharing. In recent several years, CNNs were successfully introduced into the field of solid Earth geoscience including waveform classification (Bergen et al., 2019; Dokht et al., 2019; Pham et al., 2019; Wu et al., 2019). For instance, several 1D-CNNs are developed to effectively detect or classify earthquake phases or micro-seismic events, and to further pick FBs trace by trace or receiver by receiver (Chen et al., 2019; Wang et al., 2019; Wu et al., 2019; Zhu and Beroza, 2019). Several 2D-CNNs are presented to effectively detect or classify seismic waveforms and to further pick FBs without involving the cases of missing traces or poor-quality traces (Yuan et al., 2018; Hu et al., 2019; Xie et al., 2019; Zhao et al., 2019; Liao et al., 2021).

In this paper, we apply the deep fully-convolutional SegNet, consisting of an end-to-end encoder-decoder automatic spatio-temporal feature extractor and a following pixel-wise classifier (Badrinarayanan et al., 2017), to automatically segment post-FBs data and pre-FBs background in common-shot gathers. At the

cost of learning various representative shot gathers with sparsely distributed traces and the manually interpreted dense FBs, one well-trained deep SegNet can be yielded to pick FBs of any unseen gather having differently distributed missing traces or regularly distributed traces with a rough trace interval. The network architecture is designed deeply to contain a set of convolutional layers and pooling layers, in order to see more information of adjacent seismic traces and to further pick FBs especially in the positions of missing traces. The presented approach does not need pre-extracted and pre-chosen attribute steps, as well as an additional interpolation step. The interpolation is often required to reconstruct sparsely distributed traces and useful for subsequent seismic data processing and inversion (Jia and Ma, 2017; Sun et al., 2019; Wang et al., 2019).

We begin this paper with the introduction of the motivation and general framework, as well as the description of data preparation, network architecture design, network model evaluation, model update and the optimum model application. One seismic dataset including 2251 shot gathers and the other dataset including 2273 gathers are then adopted to illustrate the effectiveness of the proposed 25-layer SegNet for classifying waveform and further picking FBs directly from shot gathers with different distributions of missing traces. Finally, a discussion and some conclusions are given.

2. Methodology

2.1. Motivation and general framework

First breaks (FBs) refer to the earliest arrival travelling time of seismic wave to receivers in the field of P-wave exploration. Therefore, there is only one FB in each seismic trace. Because some reflected waveforms below FBs, such as some separated waves, are similar to first-break ones, and there exists the unbalanced label data challenge, direct classification of FBs and other non-FBs categories may give rise to some false FBs (Yuan et al., 2018). As everyone sees, the waveform features above and below FBs are typically different in seismic common-shot gathers. There is only background or noise above FBs, which are usually spatially uncorrelated, while there are dominant signals below FBs, which are more spatially correlated. Consequently, we can divide shot gathers into two classes including pre-FBs background and post-FBs data. In this way, the boundaries between two classes are FBs.

It is important to obtain regularly and densely distributed FBs from sparsely distributed traces. One challenge of first-break picking is that trace intervals in each shot gather are often different, or trace intervals are regularly large. The quality of first-break picking for these sparsely distributed traces can decrease the quality of the shallow velocity-model building. Herein, we desire to predict the regularly dense waveform classification from shot gathers with randomly or regularly missing traces and to further pick FBs directly. More concretely, we want to find an optimal function to convert shot gathers with missing traces into mask data, where each trace is an idea step function. Due to missing traces, the optimal function should have the ability to capture information of adjacent seismic traces. The convolution-based networks have been demonstrated to be suitable to classify 2D or 3D images and easy to achieve more translation invariance for robust classification via series of high-dimensional convolution layers and pooling layers. Recently, the fully convolutional networks (FCNs) without fully-connected layers (FCLs), proposed by Long et al. (2015) in the context of image and semantic segmentation, was one popular architecture including at seismic exploration field. The FCNs have the advantages of achieving end-to-end pixel-wise classification and designing the network architectures flexibly, in contrast to the typical sliding-window CNNs involving one or

several FCLs before the classification. Several variants of FCNs including U-net (Ronneberger et al., 2015), DeepLab (Chen et al., 2018), SegNet (Badrinarayanan et al., 2017) and DeconvNet (Noh et al., 2015), have been recently widely developed. Taking missing traces into account, the network should be designed relatively deep to see larger context and to utilize more information of adjacent seismic traces. From the viewpoint of the memory, we choose the SegNet architecture, which does not transfer the entire feature maps extracted in the encoder to the corresponding decoder but instead reuse pooling indices in the corresponding decoder, compared with U-net.

Taking SegNet as the function set, the regularly dense pixel-wise classification directly from a certain shot gather \mathbf{X} with missing traces can be expressed as

$$\mathbf{P} = \text{SegNet}(\mathbf{X}, \mathbf{m}), \quad (1)$$

where \mathbf{P} represents the predicted probability with two channels indicating the background class and the signal-dominant class, and \mathbf{m} represents the network parameters or variables in the function. SegNet represents a mapping function from \mathbf{X} to \mathbf{P} , involving some hyper-parameters such as the number of the convolution layers, the number of the (un)pooling layers, and the size of the filters in the convolution layers. As denoted in the SegNet architecture for waveform classification and first-break picking (Fig. 1), SegNet consists of an end-to-end encoder-decoder feature extractor with trainable parameters and a following classifier without any learned parameters. The SegNet input is the preprocessed one-channel shot gather with missing traces, while the output is the two-channel map that is supervised by the given labels with one pre-FBs background class and the other post-FBs data class.

At the cost of training a large amount of labeled data, which can be obtained by manually interpreting dense FBs, SegNet can automatically extract statistically spatio-temporal features or attributes rather than pre-computing several attributes sensitive to FBs. The optimal extractor learned from various labeled data can yield the high-level features condensed by the previous series of layers with the property that there is large difference between the high-level feature values above FBs and those below FBs, but there is small difference among feature values at the same category including those in missing traces. As a result, these high-level features can be fed into a classifier to convert into the pixel-wise probability, which can be used to interpret each pixel as either pre-FBs background class or post-FBs signal-dominant class.

Our general framework for waveform classification and first-break picking from shot gathers with sparsely distributed traces is summarized as follows:

- 1) Data preparation. Divide data into three subsets including the training set, validation set and test set, and preprocess them with the same processing flow.
- 2) Network architecture design. Design the network architecture including the number of the (de)convolution layers, the number of the (un)pooling layers, the size of the filters, and the combination mode of different layers.
- 3) Network model evaluation. Define a loss function and several performance metrics to quantitatively evaluate the quality of a large number of models.
- 4) The optimum model choice. Find an optimum model by using the gradient descent and an early stopping strategy according to both the training set and the validation set.
- 5) The optimum model application. Apply the optimum model to unseen gathers in the test set to automatically extract the features, to meanwhile classify waveforms and to further pick FBs.

2.2. Data preparation

To obtain an effectively well-trained generalizable model or SegNet classifier and picker, three subsets including the training, validation and test sets are prepared. The training set is used for training the model or optimizing network parameters, as well as designing the learning rate, which is a hyper-parameter in the model update. The validation set is adopted to evaluate whether the learnt network is overfitting, and to determine the network hyper-parameters. The test set is considered to investigate the generalization ability of the well-trained model to unseen shot gathers. In both the training and validation sets, various typical diverse samples including shot gathers with different kinds of noise, different geometries of FBs, as well as different distributions of missing traces should be prepared as soon as possible and are cropped into the same size of both time dimension and space dimension. Accordingly, the labels used for supervising the network model are acquired by manually picking dense FBs and then segmenting gathers into one pre-FBs background class and the other post-FBs data class. Due to only two classes, the labels are quantified by binary maps. The ideal binary vectors (1 0) and (0 1) are designed to denote post-FBs data class and pre-FBs background class, respectively. An example of the input shot gather and the corresponding output label can be seen in Fig. 1. To learn the model stably, fast and effectively, all samples in different sets should be further preprocessed mainly including outliers or heavy noise elimination, surface wave attenuation, filtering, cropping and normalization.

2.3. Network architecture design

There exists only one boundary line in each shot gather, but real missing traces can complicate seismic waveform classification and first-break picking. The shallow-layer model can mainly learn the special commonness of waveforms, but it is not easy to accurately classify some complex waveforms with strong lateral variation and those at the locations of relatively dense missing traces. Therefore, the SegNet model is still required to be relatively deep. On the whole, several down-sampling stages in the encoder and up-sampling stages with the same number in the decoder, as well as a final pixel-wise classification layer, are designed, as shown in Fig. 1. Inside each stage, the (de)convolution layer consists of convolution, batch normalization (BN) and rectified linear unit (ReLU) activation, and (un)pooling is placed between several (de)convolution layers.

For each convolution layer, as denoted in the blue boxes of Fig. 1, its output can be expressed as

$$\mathbf{C}_l = \max\{0, \text{BN}(\mathbf{X}_l * \mathbf{W}_l + \mathbf{B}_l)\}, \quad (2)$$

where the symbol $*$ is a 2D convolution or deconvolution operator, \mathbf{W}_l is the 2D convolutional filter or the weights at the l -th layer, \mathbf{B}_l is an aggregation of the bias that is broadcasted to each neuron in the feature or attribute map at the l -th layer, \mathbf{X}_l is the input of the convolution layer, and \mathbf{C}_l is the output of the convolution layer. When l is 1, \mathbf{X}_l is the input shot gathers or \mathbf{X} in Eq. (1). BN represents a differentiable batch-normalization transformation that normalizes the convolution output across a mini-batch (Ioffe and Szegedy, 2015), and thus has the advantage of less internal covariate shift from shallow layers to deep layers. The activation function max or the state-of-the-art ReLU represents an element-wise maximization operation, which can enable the SegNet model to be a universal nonlinear function approximator. It has been demonstrated that ReLU has the advantages of avoiding the notorious vanishing gradient problem and promoting the model sparsity (Glorot et al., 2011).

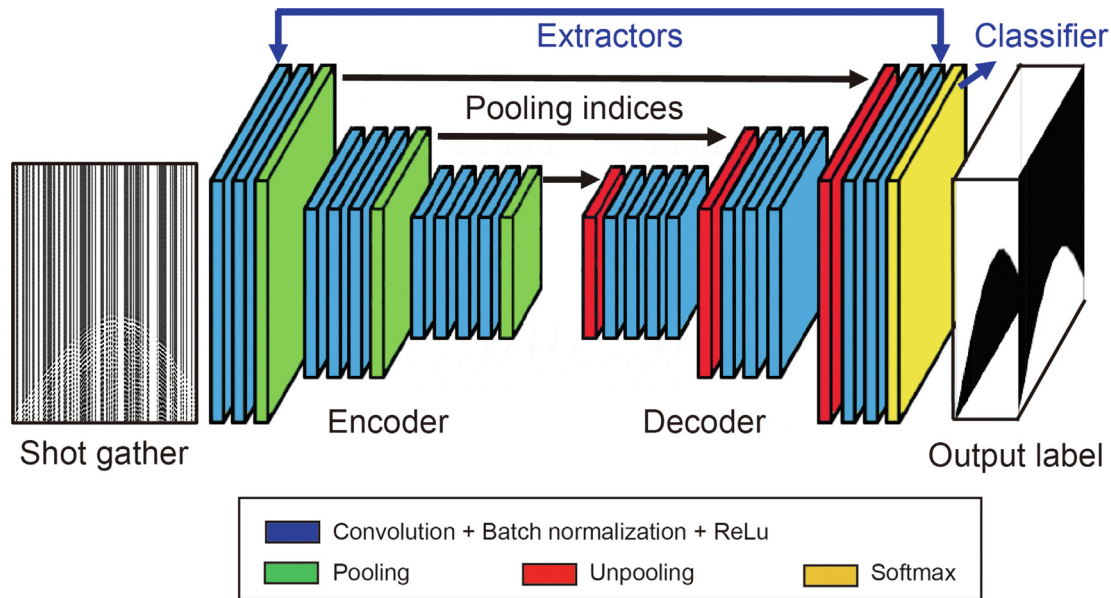


Fig. 1. The SegNet architecture for seismic waveform classification and first-break picking directly from shot gathers with sparsely distributed traces. Blue boxes represent the convolution layers containing a convolution operation, a batch normalization operation and an activation operation, green boxes represent the pooling operators, red rectangles represent the unpooling operators implemented by pooling indices computed in the pooling step, and the yellow box represents the Softmax classifier.

For the pooling layer, as denoted in the green boxes of Fig. 1, its output can be expressed as

$$\mathbf{D}_l = \text{POOL}(\mathbf{C}_l), \quad (3)$$

where the symbol POOL represents a downsampling or pooling operator, and \mathbf{D}_l represents the pooling result at the l -th layer, whose the size is smaller than that of the feature map \mathbf{C}_l . We use max-pooling as a POOL operator to yield the maximum values by comparing the neighborhood of the feature map in a sliding window way, and record the locations of the maximum values with indices at the same time. The max-pooling layer can be used to typically compress the feature maps along all dimensions while retaining the most important information. Several stacked max-pooling layers can gradually reduce data dimension, and thus act as a strong regularization for the network (Goodfellow et al., 2016) to control overfitting. Furthermore, more pooling layers can achieve more translation invariance, which is useful for effectively classifying translational, rotational and deformed spatiotemporal waveform. It is also significant for adopting more pooling layers to see the larger input image context (spatial window), which is useful for the classification of shot gathers with sparsely distributed traces due to the usage of more spatial neighborhood of statistical information. However, the pooling operator can correspondingly lead to a loss of spatial resolution of the feature maps. The increasingly lossy of boundary details is not beneficial for spatiotemporal waveform segmentation where boundary delineation corresponding to FBs is vital. Here, the decoder uses pooling indices computed in the max-pooling step of the corresponding encoder to perform non-linear unpooling. This eliminates the need for learning to upsample.

At last, a pixel-wise Softmax classifier is adopted to convert the output of the attribute extractor into probability distribution over K different possible classes:

$$\mathbf{P}_{u,v,n} = \exp(\mathbf{Y}_{u,v,n}) / \sum_{k=1}^K \exp(\mathbf{Y}_{k,v,n}), \quad u = 1, 2, \dots, K, \quad (4)$$

where $\mathbf{P}_{u,v,n}$ represents the normalized probability that the v -th data point in the n -th shot is classified into the u -th category, $\mathbf{Y}_{u,v,n}$ represents the predicted feature representation using an encoder-decoder extractor composed of convolution layers and pool layers, and K represents the number of classes. Because the Softmax classifies each data point independently, the predicted \mathbf{P} is also a K channel. In our case, there are only two classes including noise-dominant background and signal-dominant waveforms, so K equals to 2. The predicted segmentation corresponds to the class with the maximum probability at each pixel.

2.4. Network model evaluation

Because the quantity of the pre-FBs background class is comparable to that of the post-FBs data class, the binary cross-entropy criterion in an original form is chosen as the model evaluator or the loss function and is given as follows:

$$\begin{aligned} O(\mathbf{m}) &= -\frac{1}{N \times V} \sum_{n=1}^N \sum_{v=1}^V \sum_{k=1}^2 \mathbf{Q}_{k,v,n} \log(\mathbf{P}_{k,v,n}) \\ &= -\frac{1}{N \times V} \sum_{n=1}^N \sum_{v=1}^V [\mathbf{Q}_{1,v,n} \log(\mathbf{P}_{1,v,n}) + (1 - \mathbf{Q}_{1,v,n}) \log(1 - \mathbf{P}_{1,v,n})], \end{aligned} \quad (5)$$

where N represents the total shot number, V represents the total sampling point or pixel number in a shot gather, \mathbf{m} represents the model or network parameters, $\mathbf{Q}_{k,v,n}$ ($k = 1, 2$) represents the desired or true binary probability at the v -th data point of the n -th

shot, and $\mathbf{P}_{k,v,n}$ represents the predicted binary probability via Eq. (1). The closer the predicted probability distribution to the ideal one, the smaller is the cross-entropy value or $O(\mathbf{m})$, and therefore the better the built model mapping \mathbf{X} to \mathbf{P} , as denoted in Eq. (1).

To quantitatively evaluate the well-trained model obtained by minimizing Eq. (5) or the cross entropy as well as its generalization performance, the accuracy between the predicted classification results and the manually interpreted ones or the consistency between these two results are evaluated by using the following expressions:

$$\mathbf{A}_n = (\text{TP}_n + \text{TN}_n)/V, \quad (6a)$$

$$a = \frac{1}{N} \sum_{n=1}^N \mathbf{A}_n, \quad (6b)$$

$$\mathbf{R}_n = \text{TP}_n / (\text{TP}_n + \text{FN}_n), \quad (6c)$$

$$r = \frac{1}{N} \sum_{n=1}^N \mathbf{R}_n, \quad (6d)$$

where true positive TP_n ($n = 1, 2, \dots, N$) represents the number of the sampling points in the n -th shot that correctly predict signal-dominant class below FBs, true negative TN_n represents the number of the sampling points in the n -th shot that successfully predict background above FBs, and false negative FN_n represents the number of the sampling points in the n -th shot which fail to predict signal-dominant class. The accuracy \mathbf{A}_n and a give the percentage of correctly classified sampling points in the n -th shot gather and all shot gathers in the training, validation or test sets, respectively. The recall rate \mathbf{R}_n and r give the percentage of correctly classified signal-dominant sampling points in the whole signal-dominant sampling points of the n -th shot gather and all shot gathers in the training, validation or test sets, respectively (Powers, 2011).

2.5. Model update

A linear combination of the model update $\Delta \mathbf{m}_{t-1}$ from the previous iteration $t-1$ and the negative gradient of the loss function is used to update the model \mathbf{m}_t at the current iteration t as follows:

$$\begin{cases} \Delta \mathbf{m}_t = \mu \Delta \mathbf{m}_{t-1} - \eta_t \partial O(\mathbf{m}_{t-1}) / \partial \mathbf{m}_{t-1}, \\ \mathbf{m}_t = \mathbf{m}_{t-1} + \Delta \mathbf{m}_t \end{cases} \quad (7)$$

where μ represents the momentum coefficient or the weight of the last update $\Delta \mathbf{m}_{t-1}$, η_t represents the learning rate or the weight of the negative gradient at the iteration t , $\frac{\partial O(\mathbf{m}_t)}{\partial \mathbf{m}_t} = \frac{\partial O(\mathbf{m}_t)}{\partial \mathbf{Y}_t} \frac{\partial \mathbf{Y}_t}{\partial \mathbf{m}_t} = (\mathbf{P} - \mathbf{Q}) \frac{\partial \mathbf{Y}_t}{\partial \mathbf{m}_t}$ represents the gradient of the loss function with respect to the model or network parameters, $\partial O(\mathbf{m}_t) / \partial \mathbf{Y}_t$ represents the gradient of the loss function with respect to the input of the Softmax classifier or the output of the encoder-decoder feature extractor, $\partial O(\mathbf{m}_t) / \partial \mathbf{Y}_t$ is equal to the difference between the predicted probability \mathbf{P} and the true probability \mathbf{Q} , $\partial \mathbf{Y}_t / \partial \mathbf{m}_t$ represents the gradient of the output of the encoder-decoder feature extractor with respect to the model parameters, and $\partial \mathbf{Y}_t / \partial \mathbf{m}_t$ can be explicitly calculated by using the back-propagation algorithm (Rumelhart et al., 1986). Herein, the gradient term $\partial O(\mathbf{m}_t) / \partial \mathbf{m}_t$ is approximately calculated by randomly selecting only a small subset (mini-batch) of $\{\mathbf{X}_n, \mathbf{P}_n\}_{n=1}^N$ with improvements on efficiency and convergence, which is the idea of a mini-batch stochastic gradient descent (SGD) algorithm (LeCun et al., 1998).

When the loss for the training set decreases, whereas the validation loss increases (known as overfitting) or remains stable for a

certain period of iterations, the training or the update is stopped. The finally well-trained model *SegNet* with the optimal network parameters \mathbf{m}_{opt} is saved as the automatic classifier.

2.6. The optimum model application

The saved optimum model can be used to convert any unseen shot gather \mathbf{X}_{test} in the test set into a two-channel normalized probability map with the following formula:

$$\mathbf{P} = \text{SegNet}(\mathbf{X}_{\text{test}}, \mathbf{m}_{\text{opt}}). \quad (8)$$

When the predicted binary probability in a certain pixel of \mathbf{P} is closer to (1 0), we classify the corresponding input sampling point in \mathbf{X} into the post-FBs signal-dominant class. On the contrary, when the predicted probability is closer to (0 1), we classify the input sampling point into the pre-FBs background class. In an ideal case, the sampling points above FBs should be one class, while those below FBs are the other. Consequently, the boundary between the two classes or the segmentation line in the classification map can be located, i.e., FBs are picked. Furthermore, we take the maximum of each position in the two-channel probability map to obtain a single-channel probability map. It can be utilized to statistically and directly evaluate the quality for waveform classification along with first-break picking to some extent. The smaller the probability in the single-channel probability map, the lower the accuracy of waveform classification, or the more unreliable waveform classification. In an ideal case, the single-channel probability curve for each trace including missing traces should be one impulse function, where there is only one minimum. The position of the minimum value in each trace can be regarded as the FB. Consequently, we can also pick FBs by locating the minimum values of the single-channel probability map one trace by one trace.

3. Examples

The proposed method is tested on two land real datasets. The two adjacent datasets are from two monitoring receiver lines approximately 500 m apart. The topography of the real working area is complicated with a surface elevation difference up to 500 m. One dataset including 2251 common-shot gathers is used for training and validation, where 1751 gathers are randomly chosen for training while the rest for validation. The other containing 2273 common-shot gathers is adopted for test. Each common-shot gather with the sampling time of 4 ms consists of 168 traces and is cropped into 550 time samples. For the two datasets, all the sources are the same type of explosives, and the distances among all receivers in each monitoring line are approximately the same. All common-shot gathers are preprocessed by implementing outliers or heavy noise elimination, surface wave attenuation, filtering, and normalization. For each processed gather with 92400 sampling points, the labels are made by manually picking FBs and then classifying pre-FBs background and post-FBs data.

For our examples, 25 layers including 18 convolution layers, 3 pooling layers, 3 unpooling layers and 1 Softmax layer are combined into our network architecture by trial and error, as shown in Fig. 1. Except the number of convolution filters in the final convolution layer of the decoder is 2, the number of convolution filters in the other convolution layers is 64. For each convolution kernel, the size is fixed to 3×3 with stride of 1 (i.e., overlapping window). Following the convolution layer, the max-pooling operator is performed over a 2×2 pixel window, with stride of 2 (i.e., non-overlapping window), and thus the resulting output through the pooling operator indicated by the green box is sub-sampled by a factor of 2 along both the time direction and the space direction.

In all experiments, the 25-layer network is trained by using SGD with a momentum coefficient of 0.9 and a mini-batch size of 5 shot gathers. The learning rate starts from 0.01, and is reduced by a factor of 0.2 every 5 epochs. The full pass of the training algorithm over the entire training set using mini-batches is defined as an epoch. The maximum number of epochs for training is assigned to 200, and the cross-entropy loss error is set to 0.001. For all experiments, the biases are initialized as zero, and the weights are initialized as noise in the normal distribution with a mean of zero. The experimental machine configuration includes an Intel core i9-9900k CPU (Central Processing Unit) and an Nvidia GeForce RTX 2080Ti GPU (Graphics Processing Unit).

To investigate the generalization performances of the well-trained SegNet classifier and picker from the training and validation sets without missing traces or regularly dense distributed traces, we apply it to three kinds of unseen testing data sets including 2273 common-shot gathers with different ratios of missing traces. For the well-trained network, the total classification accuracy of both the 1751 testing gathers and the rest 500 validation gathers are over 99%. It may be noticed that it takes about 170 s to classify 2273 gathers without parallel computing. Fig. 2a–c shows the accuracy distributions of 2273 testing common-shot gathers with 0% missing traces (or without missing traces), with 10% randomly missing traces of each gather and with 50% randomly missing traces of each gather (i.e., 84 traces from 168 traces are randomly selected to fill in zero-valued amplitude) versus the source location. It can be observed from Fig. 2a that the classification accuracy of most gathers in the original 2273 testing gathers without missing traces reach 98%, while the accuracy of only about

50 gathers is less than 98% and their corresponding sources are generally close to the black monitoring line. As denoted in Fig. 2b and c, the well-trained network from the training and validation sets without missing traces presents the poor prediction effect on 2273 testing gathers with 10% and 50% randomly missing traces, having the accuracy generally below 80%. It is obvious that the well-trained network from shot gathers without missing traces or with regularly sampled traces cannot be generalized to classify shot gathers with sparsely randomly distributed traces well. It can also be seen that the gathers far from the black monitoring line have the lower accuracy than those close to the monitoring line. In addition, the more missing traces, the lower the accuracy.

In order to generalize to classify shot gathers with missing traces well, we generate an improved well-trained SegNet classifier and picker from the training and validation sets with 50% randomly missing traces, and apply it to various testing common-shot gathers with different distributions of missing traces including 0% missing traces to test its performances. For the new well-trained model, the total classification accuracy of 1751 testing gathers and the rest 500 validation gathers are 99.68% and 99.60%, respectively. Fig. 3a–c shows the accuracy distributions of 2273 testing gathers without missing traces, with 50% randomly missing traces of each gather and with 50% regularly missing traces of each gather versus the source location. The testing sets of Fig. 3a and b are the same as those of Fig. 2a and c, respectively. It can be observed from Fig. 3 that the new well-trained model from shot gathers with missing traces has the excellent generalization ability with the accuracy of almost all gathers over 98% even in the case of regularly missing traces. The accuracy of only several gathers with sources located

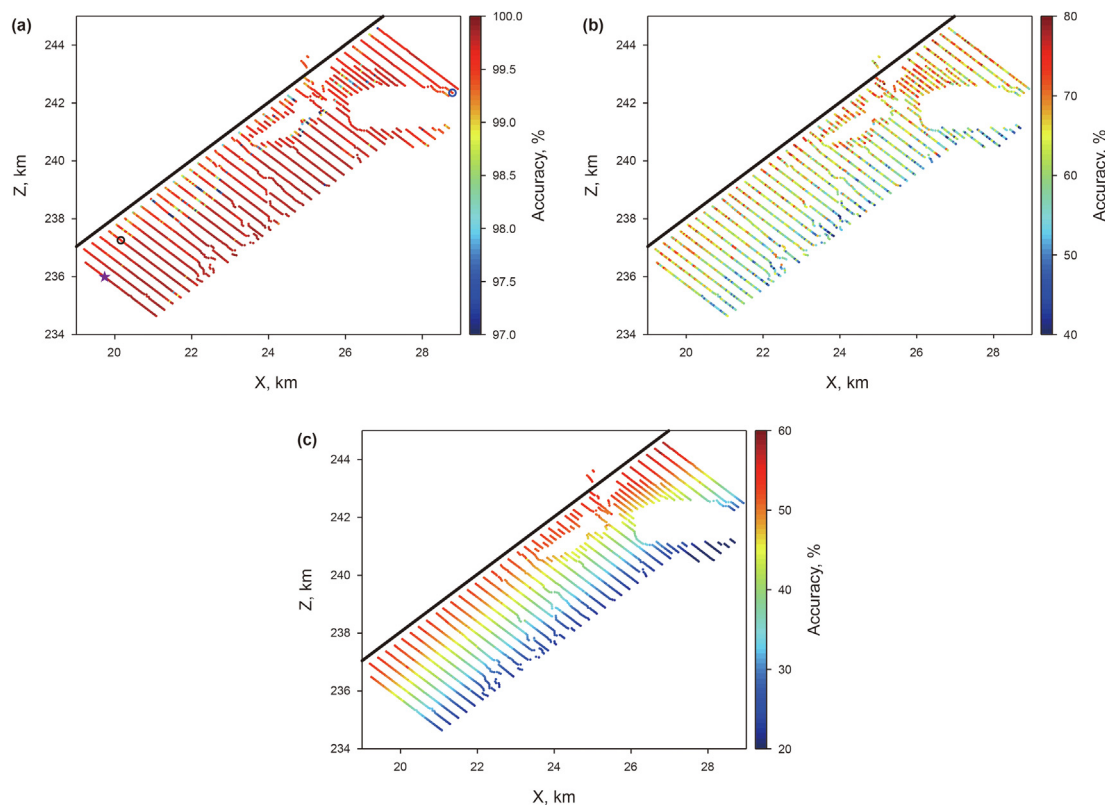


Fig. 2. The accuracy distributions of 2273 testing common-shot gathers with 0% (a), 10% (b) and 50% randomly missing traces of each gather (c) based on the well-trained SegNet from the training and validation sets without missing traces versus the explosive source location. Colored dots denote source locations with the classification accuracy of single gather, and all receivers are located on the black monitoring line. In (a), the hollow pentagram and black hollow circle in the lower left corner, and the blue hollow circle in the upper right corner denote the locations of the 15th shot, 243rd shot and 2213rd shot, respectively. The well-trained model from shot gathers without missing traces cannot be generalized to classify shot gathers with missing traces well.

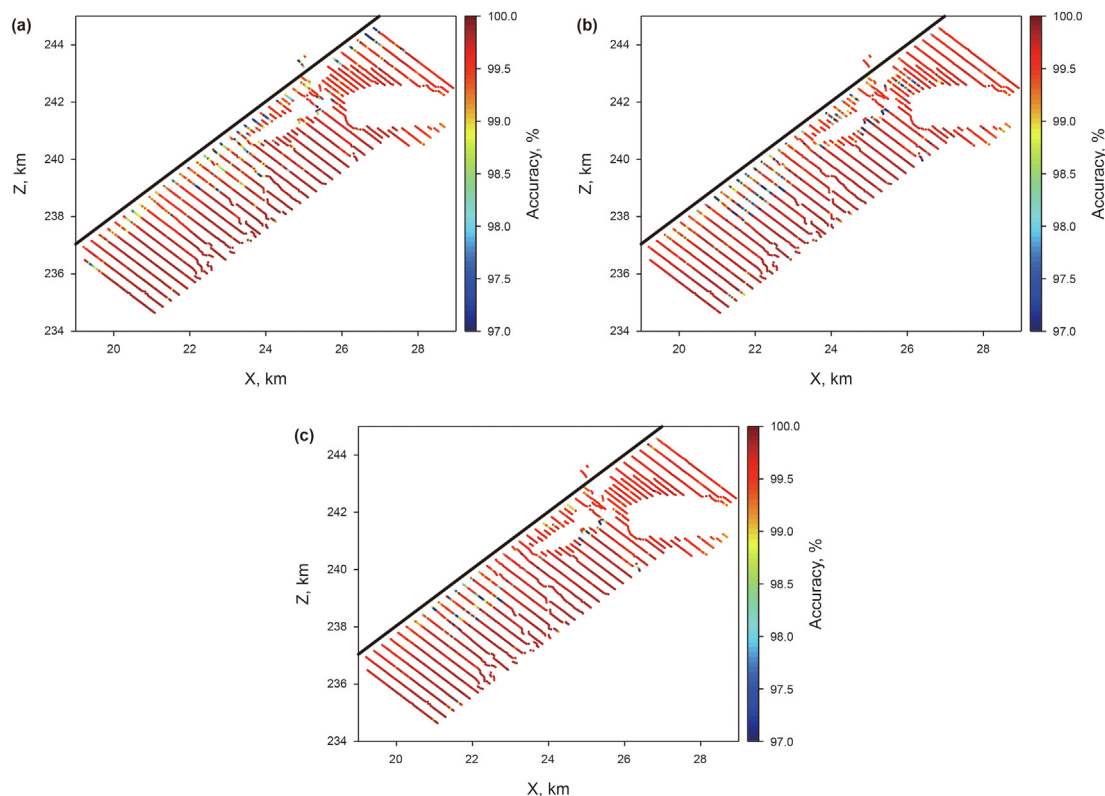


Fig. 3. The accuracy distributions of 2273 testing common-shot gathers with 0% missing traces (a), with 50% randomly missing traces of each gather (b) and with 50% regularly missing traces (c) based on the new well-trained SegNet from the training and validation sets with 50% randomly missing traces of each gather versus the explosive source location. The well-trained model has favorable generalization ability with the accuracy of almost all gathers over 98%.

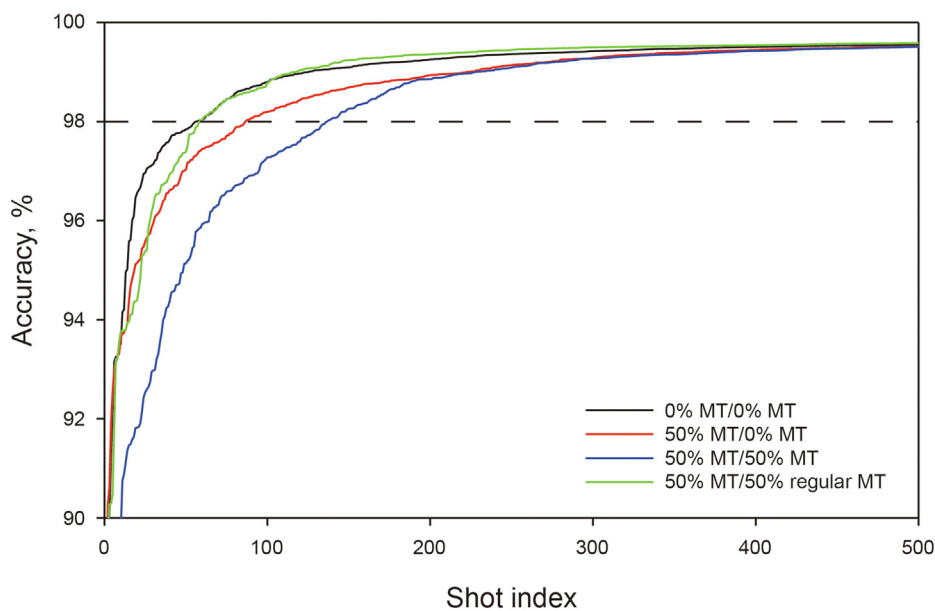


Fig. 4. Comparisons among the classification accuracy of the first 500 common-shot gathers arranged from the smallest value to the largest value corresponding to 2273 testing shot gathers in Figs. 2a and 3a–c. The characters 0% MT/0% MT, 50% MT/0% MT, 50% MT/50% MT and 50% MT/50% regular MT in legend represents four dataset cases of Figs. 2a and 3a–c, respectively. The accuracy of gathers more than 95% in each testing set is over 98%.

near the black monitoring line is relatively low. Fig. 4 displays careful comparisons among the classification accuracy of 500 shot gathers in each testing dataset of Figs. 2a and Fig. 3a–c. For all these four dataset cases, gathers more than 95% in each testing set of

2273 gathers have the accuracy over 98%.

To focus on investigating the classification ability of post-FBs signal-dominant waveforms, we calculate the recall of 2273 testing common-shot gathers or four well-generalized cases of

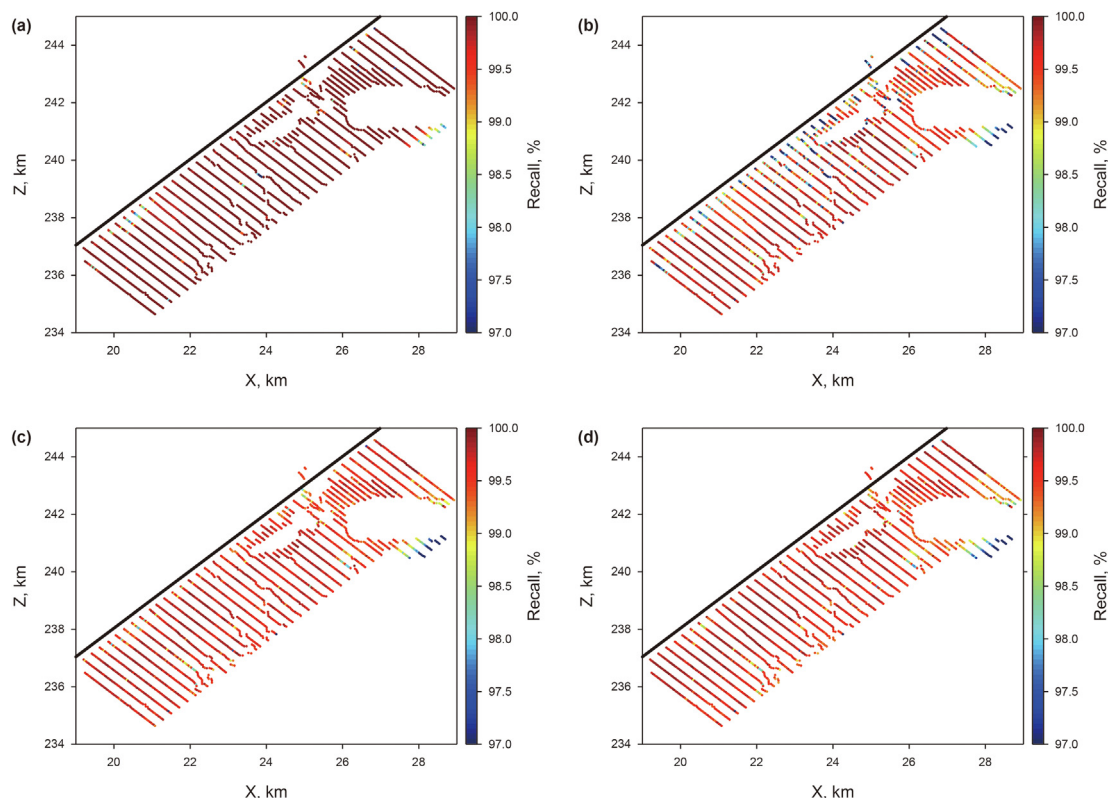


Fig. 5. The recall distributions of 2273 testing common-shot gathers versus the explosive source location corresponding to four dataset cases of Figs. 2a and 3a–3c, respectively. The recall of almost all gathers in each testing set reach 97%.

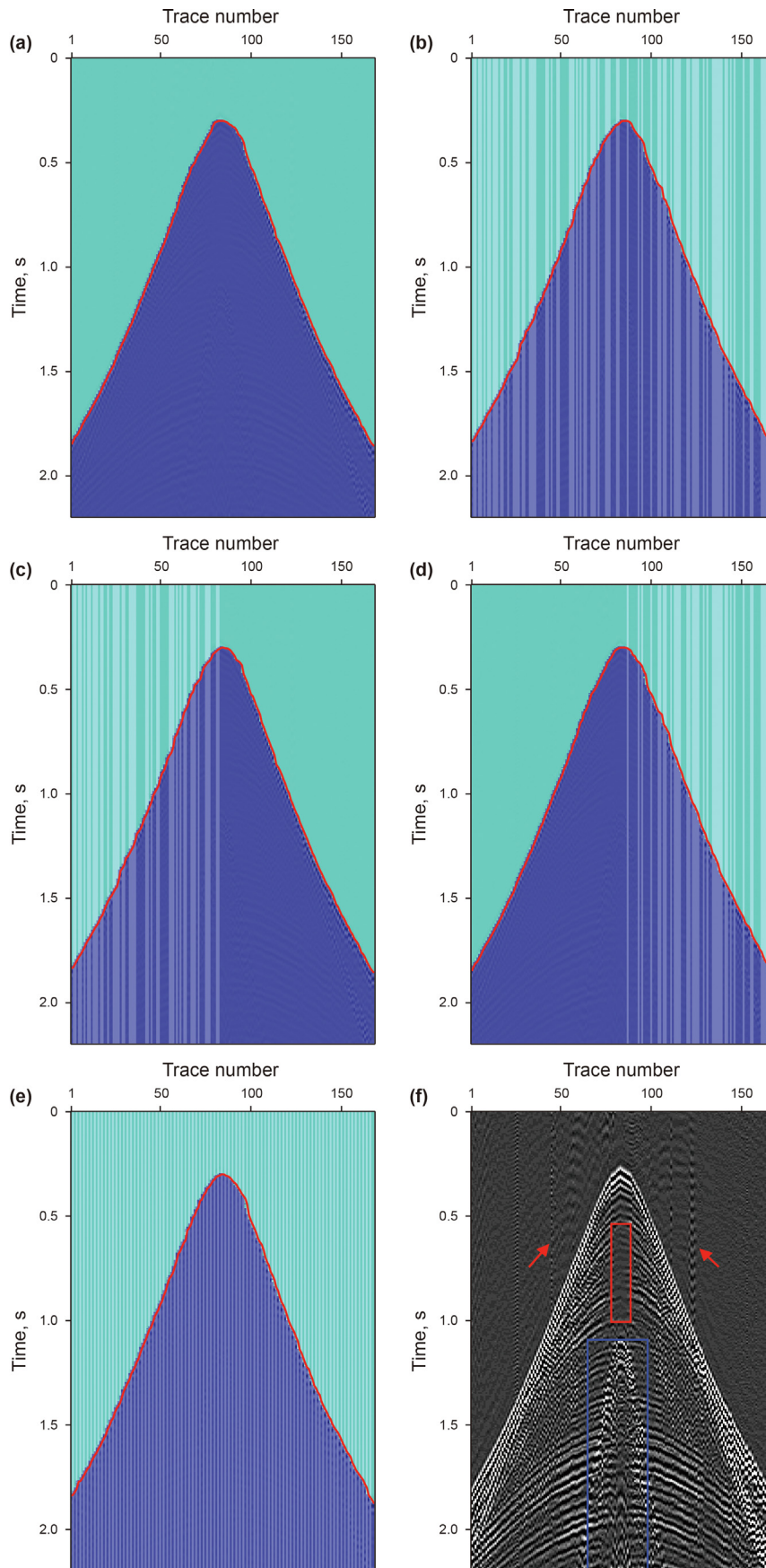
Figs. 2a and Fig. 3a–c, as shown in Fig. 5a–d. For these four dataset cases, the recall of almost all gathers reach 97%, even in the cases of missing traces with different distributions. Compared with three other cases, the generalization accuracy of the new well-trained SegNet classifier and picker from samples with 50% randomly missing traces to 2273 testing shot gathers without missing traces are relatively most dependent on geometry related to the locations of sources. The lower the recall, the higher detection error the signal-dominant samples. In general, the wrong detection happens near sharp top FBs and weak reflection areas.

To illustrate the effect of seismic waveform classification and first-break picking based on the new well-trained SegNet model from the training and validation sets with 50% randomly missing traces more clearly and intuitively, we select two representatively original common-shot gathers including the 243rd shot gather and the 2213rd one. These two chosen original gathers share different first-break shapes, different types of first-break waves, different distances from the source to the monitoring receiver line, different degrees of weak reflection below FBs, and different kinds of noise. Furthermore, we focus on investigating the generalization ability to the gathers with randomly, regularly and several special distributions of missing traces generated from these two representatively original common-shot gathers.

Fig. 6 shows comparisons among the classification results and first-break picking results of the gathers with five distributions of missing traces made from the 243rd original shot gather (Fig. 6f). It can be clearly observed that the classification results of the gathers without missing traces (Fig. 6a), with 50% randomly missing traces (Fig. 6b), with only missing traces on the left (Fig. 6c), with only missing traces on the right (Fig. 6d) and with 50% regularly missing traces (Fig. 6e) are in good accordance each other, even for the locations of six consecutive missing traces (Fig. 6b and d) between

the 135th trace and the 140th trace. In addition, five segment lines (red lateral lines) between the pre-FBs background class (light blue part) and the post-FBs signal-dominant class (dark blue part) or the corresponding picked FBs are basically consistent and continuous along the space direction. The single-channel probability maps corresponding to Fig. 6a–e are shown in Fig. 7a–e, and local zoomed plots of low-value probability zones of the middle 84th traces in Fig. 7a–e are shown in Fig. 8. It can be seen that the maximum probability curve at every trace approximates a pulse-like function with a small width of low-probability values, which are related to the locations of FBs, as well as the uncertainty of waveform classification and first-break picking. Fig. 9 shows the differences between the automatically picked FBs and the manually picked FBs versus the trace number. The average error between the automatically picked FBs from 168 traces in each gather case and the manually picked ones is less than two samples, suggesting that the automatically picked FBs are close to the manually picked ones. It can be inferred from Fig. 9 that the differences of FBs do not so strongly depend on the distributions and the locations of missing traces.

Fig. 10 shows comparisons among the classification results and first-break picking results of the gathers with five distributions of missing traces made from the other 2213rd original common-shot gather (Fig. 10f) based on the new well-trained SegNet from the training and validation sets with 50% randomly missing traces of each gather. The corresponding source is far from the monitoring receiver line, as denoted in the position of the blue hollow circle in Fig. 2a. One can clearly see that the classification results of five gathers with different distributions are good. Moreover, the corresponding picked FBs are visually consistent and spatially continuous, even for the locations at 13 consecutive missing traces (Fig. 10b and d) between the 30th trace and the 44th trace except for



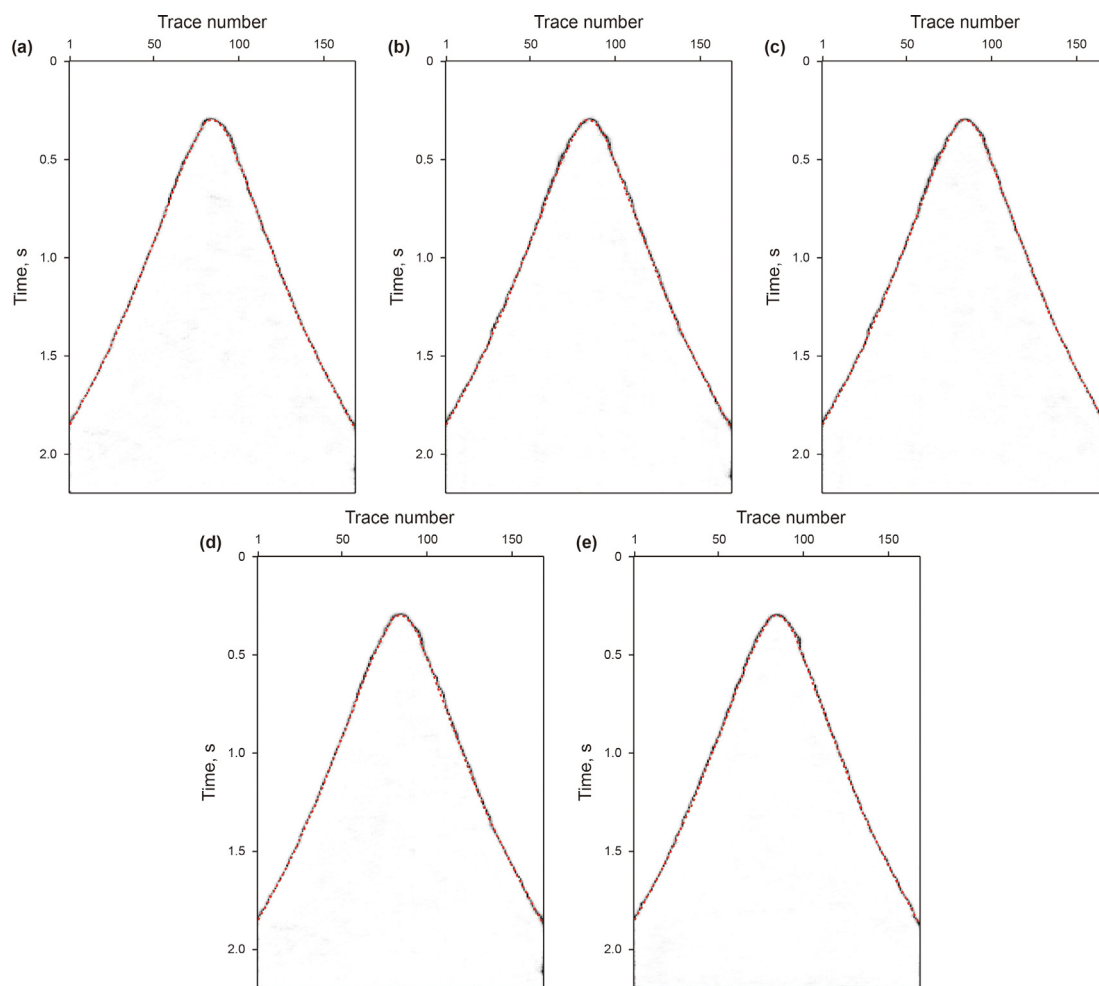


Fig. 7. Comparisons among the single-channel probability maps corresponding to Fig. 6a–e. The black and white represent the low- and high-value probability, respectively. Red lateral dashed lines represent the manually picked FBs from the common-shot gather of Fig. 6f, which are regarded as the reference. It can be observed that almost all referenced FBs are located in low-value zones of five probability maps.

the 35th trace and the 37th trace. Fig. 11a–e displays the single-channel probability maps derived by predicting the original gather and the other four generated gathers, and Fig. 12 displays their corresponding local zoomed plots of low-value probability zones in the middle traces. It can be observed that the probability curve at each trace looks like a pulse function with several continuous low-value probabilities. The differences between the automatically picked FBs and the manually picked FBs versus the trace number are shown in Fig. 13. The average error between the automatically picked FBs from 168 traces in each gather case and the manually picked ones is not more than two samples. In contrast to five cases derived from the 243rd original common-shot gather, the first-break picking of five distributions gathers generated from the 2213rd original gather are relatively unstable and more dependent on the distributions of missing traces, as indicated in Figs. 9 and 13.

To illustrate why the new well-trained SegNet classifier and picker can directly classify seismic waveform and subsequently pick FBs at the positions of missing traces, and to further test its

generalization ability, we choose another representative common-shot gather in the test set and make a special gather obtained by replacing 21 consecutive traces of the original gather from the 74th to 94th traces with zero. The chosen representative original gather (Fig. 14c) is the 15th shot located at the location of the hollow pentagram in Fig. 2a. As the classification results (Fig. 14a and b) of the 15th original shot gather and its corresponding modified gather (Fig. 14d) indicate, the classification result predicted by applying the well-trained SegNet model to a shot gather with 21 consecutive missing traces is similar to that of the original fully-sampled gather, and the classification accuracy for the pair of gathers are 99.78% and 99.52%, respectively. As well, the picked FBs are suitable agreement even in the locations of missing traces. Furthermore, we extract 64 features in the 17th convolution layer and 2 features in the 18th convolution layer obtained by applying the new well-trained SegNet to two given gathers of Fig. 14c and d, and implement careful comparisons, as shown in Figs. 15 and 16. It can be observed that 64 deep-layer features automatically extracted from the gather with 21 consecutive missing traces are similar to the corresponding ones

Fig. 6. Comparisons among the classification results and first-break picking results of the gathers with five distributions of missing traces (a–e) generated from the 243rd original common-shot gather (f) based on the new well-trained SegNet. The representatively cropped original gather in (f) contains background, industrial electrical interference (red arrows), one very weak post-FBs reflection area (red rectangle) similar to the pre-FBs background, the surface wave (blue rectangle), near-offset direct waves and far-offset refracted waves. The classification results are overlaid by the picked FBs (red lateral lines) and the seismic amplitude in (a–e), where white vertical zones denote the positions of missing traces. The classification accuracy for five cases is over 99%. Furthermore, the picked FBs are basically consistent each other and spatially continuous.

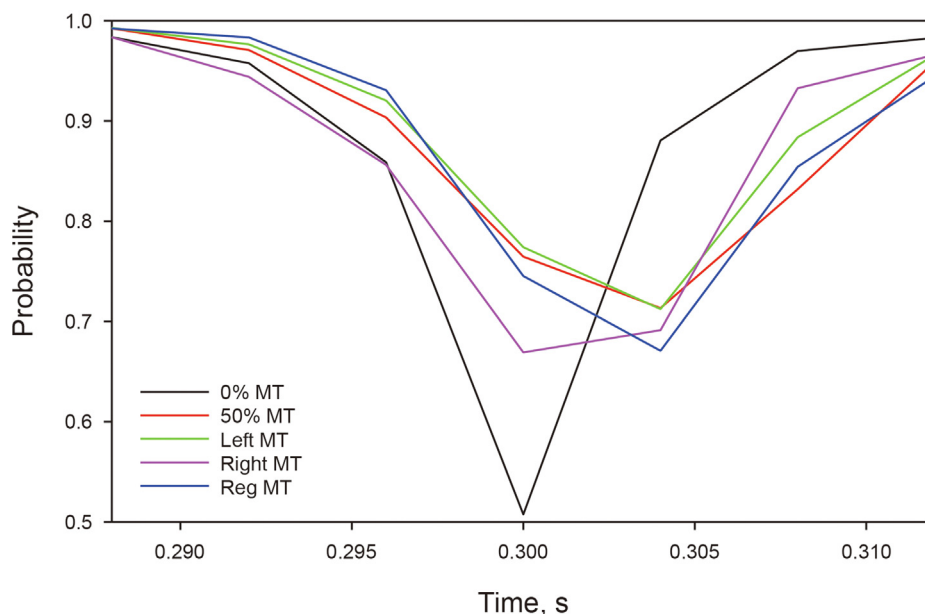


Fig. 8. Local zoomed plots of low-value probability zones of the middle 84th traces in Fig. 7a–e. The characters 0% MT, 50% MT, Left MT, Right MT and Reg MT in legend represent five cases of single-shot gathers with 0%, 50% randomly, only the left, only the right and regularly missing traces, as indicated in Fig. 6a–e, respectively. There are only about three continuous time samples with the classification probability less than 0.9, regardless of the distributions of missing traces. In addition, the differences among the minimum probability for five cases are obviously not more than one time sample.

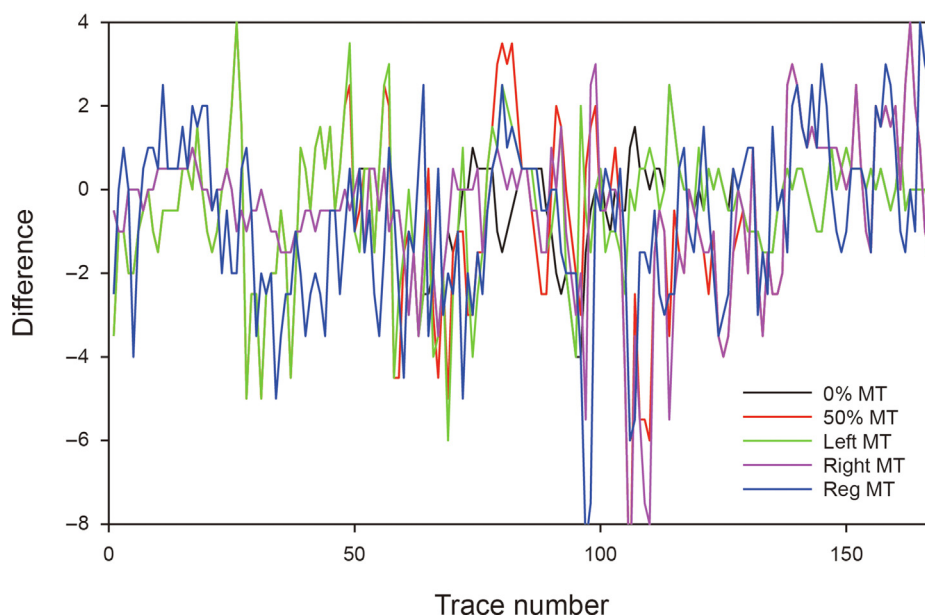


Fig. 9. The differences between the SegNet-based automatically picked FBs or the red lines of Fig. 6a–e and the manually picked FBs versus the trace number. The errors of the automatically picked FBs from 168 traces including the locations of missing traces even six consecutive missing traces are commonly less than four time samples, and the average errors corresponding to five cases are 0.7411, 1.7589, 1.2054, 1.3185 and 1.6756 time samples, respectively.

extracted from the original gather, even in the corresponding positions of consecutive missing traces. Moreover, as the depth of the convolution layer increases from the 17th convolution layer to the 18th convolution layer, the features automatically extracted from the gather with missing traces become closer to the corresponding ones extracted from the original gather. It is noticeable that the new well-trained SegNet model can extract features with obvious boundaries between the shallow background part and the deep signal-dominant part after both the 17th and 18th convolution layers, even in the presence of 21 consecutive missing traces.

4. Discussion

We adopt a large number of preprocessed seismic common-shot gathers as the network input, and use the corresponding pre-FBs background and post-FBs data labels to supervise the network output to gain a well-trained end-to-end encoder-decoder extractor. The extractor can automatically extract features instead of pre-computed ones. The automated extractor constructed by stacking a series of artificial neurons enables FBs to be picked in a manner analogous to how a human would pick them. Although the

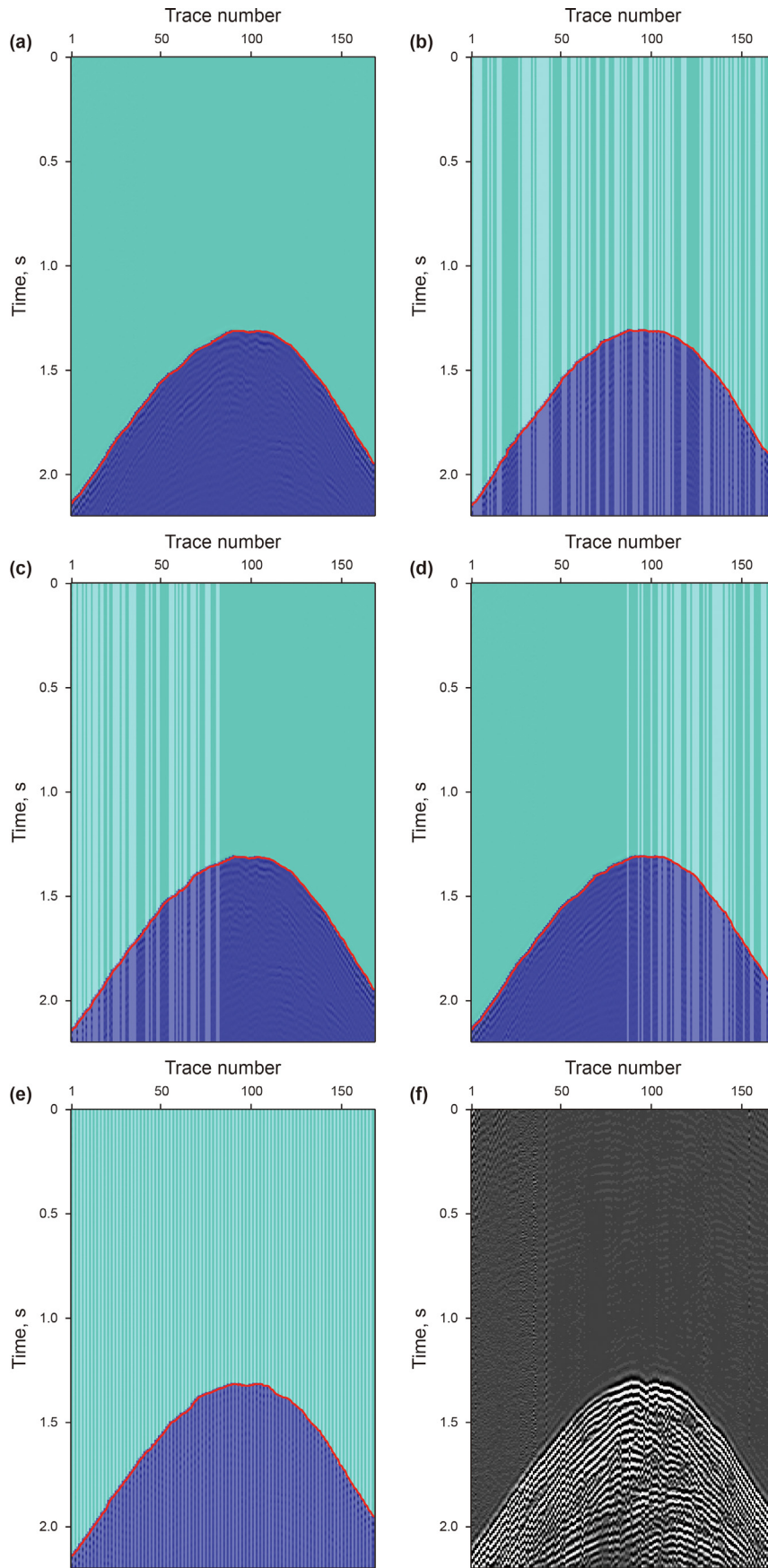


Fig. 10. Comparisons among the classification results and first-break picking results of the gathers with five distributions of missing traces (a–e) generated from the 2213rd original common-shot gather (f) based on the new well-trained SegNet. The representatively cropped original common-shot gather in (f) contains background, industrial electrical interference and asymmetric fluctuant direct waves, which are obviously different from the characteristics of Fig. 6f. The classification accuracy for five cases reaches 99%. Moreover, the corresponding picked FBs are basically consistent and continuous along the space direction.

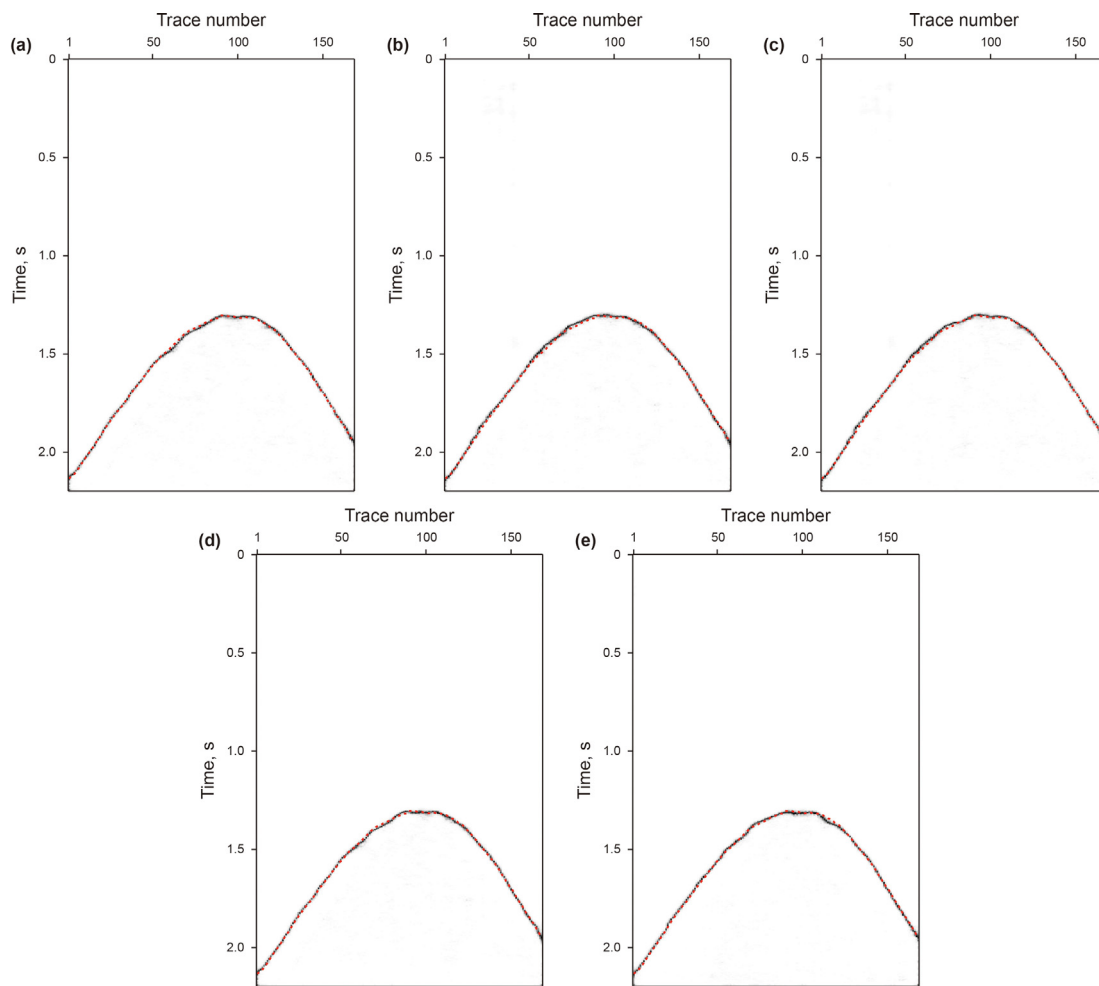


Fig. 11. The single-channel probability maps corresponding to Fig. 10a–e. The black represents low-value probability, and red lateral dashed lines represent the manually picked FBs from the common-shot gather of Fig. 10f. It can be seen that the manually picked FBs are located in or near low-value zones of all probability maps.

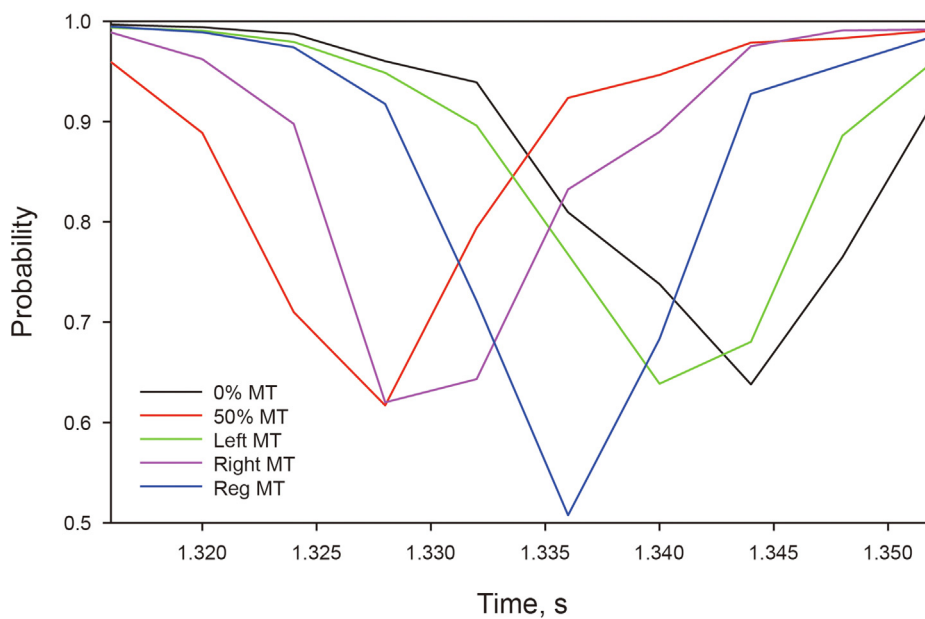


Fig. 12. Local zoomed plots of low-value probability zones of the middle 84th traces in Fig. 11a–e. There are about four continuous time samples with the classification probability less than 0.9. In addition, the differences among the minimum probability for five cases are from 0 to 5 time samples. In contrast to Fig. 8, there exists relatively large uncertainty of low-value probability zones.

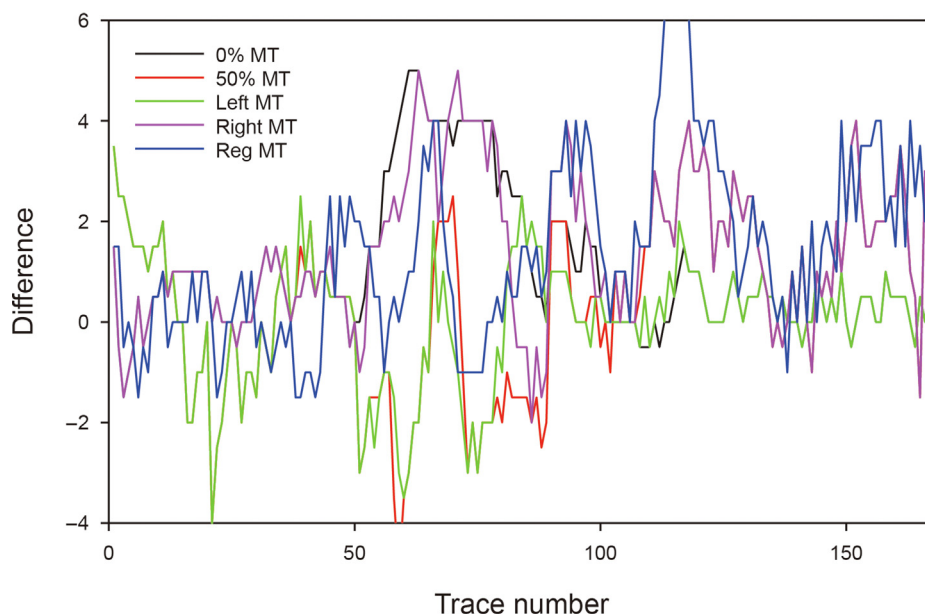


Fig. 13. The differences between the automatically picked FBs (or red lines of Fig. 10a–e) and the manually picked FBs versus the trace number. The errors of the automatically picked FBs from 168 traces including the locations of missing traces are commonly less than 4 samples, and the average errors corresponding to five cases are 1.1429, 1.5536, 0.9345, 1.6548 and 1.7083 samples, respectively.

well-trained automated extractor is built at the cost of learning big data, it is easy to prepare abundant seismic shot gathers and the corresponding labels by segmenting FBs, which can be picked manually. Nevertheless, it is important to learn representative and diverse shot gathers especially including those with different distributions of missing traces. Under the premise that the basic condition is met, our well-trained fully-convolutional SegNet consisting of an end-to-end encoder-decoder feature extractor and a following pixel-wise classifier adapts to automatically pick FBs of various unseen shot gathers fast even acquired in different geometries, such as different trace intervals, different distributions of receivers, or different number of receivers.

We choose the data within the same source type of explosives in our examples in order to focus on investigating the effectiveness of SegNet-based automated first-break picking directly from shot gathers with sparsely distributed traces including randomly missing traces, consecutively missing traces and regularly missing traces. Furthermore, we attempt to account for the reason why our presented SegNet-based model can classify waveform and pick FBs in the locations of missing traces. It is also worth investigating that simultaneously learning a large number of seismic data from various sources, such as explosives, vibroseis or air gun may enable the well-trained SegNet to cause a broader generalization.

The design of labels is not unimportant for learning one data-driven waveform classifier and first-break picker. From the viewpoint of balancing the number of different classes of labels, it should be appropriate to label seismic data with one class on post-FBs data and the other on pre-FBs background, which are just segmented by FBs. The design of network architecture and the performance of the network learning are dependent on the choice of the input size. We choose various cropped 2D seismic data with relatively large size of both time and space dimensions, which can be roughly determined by balancing post-FBs signal-dominant samples and pre-FBs background samples as well as the computer memory. And we input them into the end-to-end SegNet architecture to implement pixel-wise classification. The SegNet architecture can be designed deeper and more flexibly to see more adjacent information. It is useful for a deep SegNet to classify

waveforms and pick FBs of shot gathers with consecutive missing traces. However, the end-to-end SegNet architecture accompanied by the cropped gather input with relatively large-size time-space dimension requires more preprocessing steps, such as the removal of some local noise (e.g., outliers or 50-Hz industrial electrical interference). The space dimension of the input determines the maximum space limitation that the network can see. The space size of the reception field, which can be calculated by using the number and size of pooling layers and convolution layers, determines the real space range that the network can see. The larger the reception field, the more spatiotemporal information the network can capture. Besides directly learning spatiotemporal amplitude information, some other attributes extracted from seismic data can be considered as the input independently or simultaneously to learn other well-trained models for classifying waveform and further picking FBs. Furthermore, the ensemble learning for first-break picking is also worth investigating.

We should consider the distribution of missing traces to design the number of pooling layers, especially for the number of consecutive missing traces. In contrast to convolution layers, the use of more pooling layers can increase the receptive field more rapidly and can help see a larger range of waveform, so more local statistics can be utilized to perform waveform classification and first-break picking. Moreover, more pooling layers along with convolution layers can achieve more translation invariance, which is useful for effectively classifying translational, rotational and deformed spatiotemporal waveform and further picking FBs. However, the usage of more pooling layers may blur some details related to FBs. It may be one main reason why there is the low-value probability of a certain width near the FB in each trace of the one-channel probability map, which is derived by the SegNet-based two-channel probability output map. This paper adopts only 2D pooling and 2D convolution filters to gradually include as much spatiotemporal statistics of neighborhood as possible in the input preprocessed common-shot gathers. However, it can be readily extended to a higher dimension to utilize more information. For instance, 2D pooling and 2D convolutional filters can be directly replaced by 3D or higher dimensional those to implement

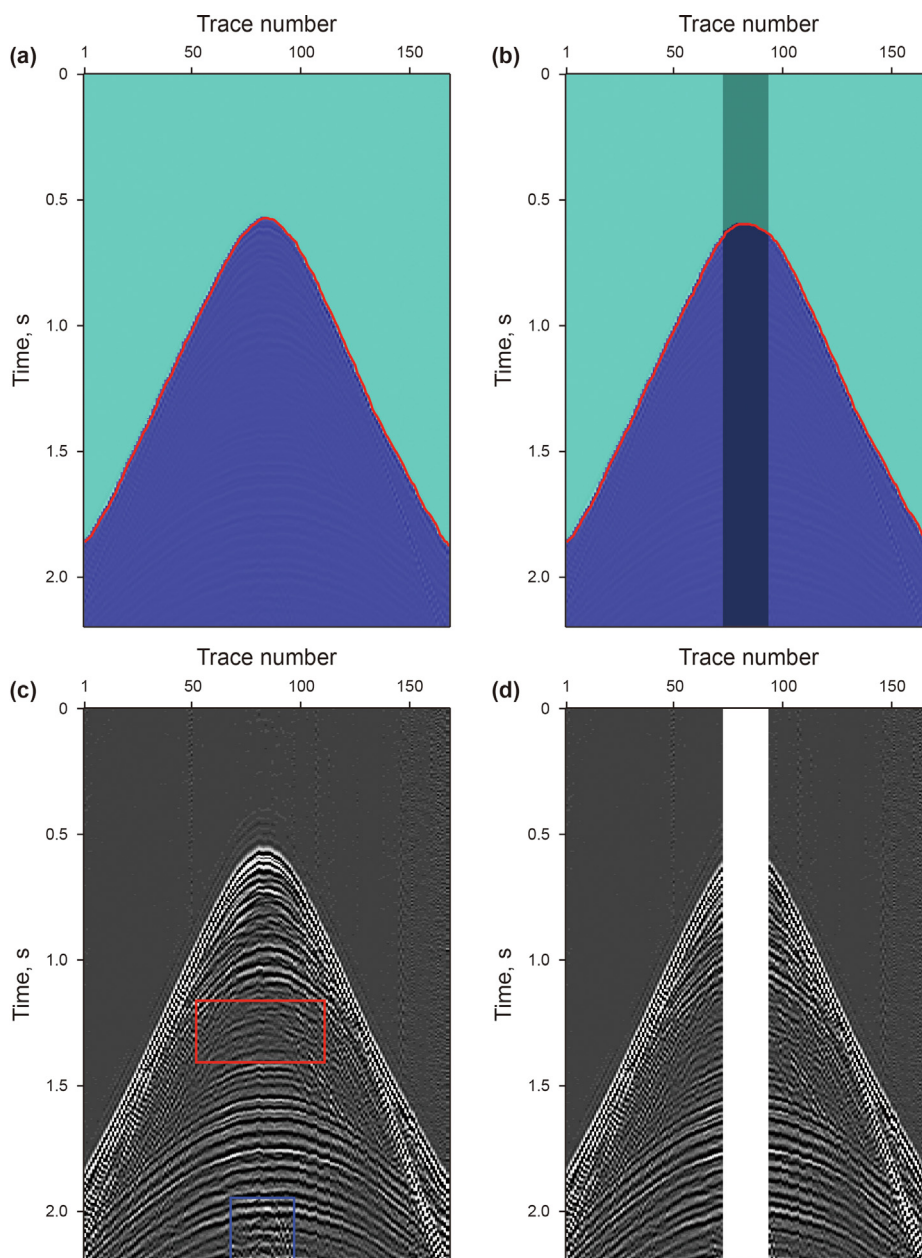


Fig. 14. Comparisons between the classification results (a–b) of the 15th original common-shot gather (c) and its corresponding modified gather (d) involving 21 consecutive missing traces based on the new well-trained SegNet. The original cropped gather in (c) contains weak background, weak industrial electrical interference, the relatively weak post-FBs reflection area (red rectangle), the relatively weak residual surface wave (blue rectangle), near-offset direct waves and far-offset refracted waves. The pixel-wise classification of the shot gather with 21 consecutive missing traces is similar to that of the original gather without missing traces, and the picked FBs are also consistent even in the locations of missing traces.

waveform classification and first-break picking in multiple shots and multiple domains or by increasing a frequency or scale dimension.

We do not impose any constraint of some common-sense factors related to spatio-temporal relationship of FBs on learning SegNet, such as small time delays between adjacent traces and time increase with offset distance, but the picked FBs can visually meet these factors, as illustrated in our examples. It suggests that the proposed well-trained 25-layer SegNet model has the ability to learn these common-sense factors from various seismic gathers with the corresponding labels. In turn, whether some common-sense factors or some experience of interpreters can be considered as a constraint or a guide to integrate into the process of

network learning and what are its roles in classifying waveform and picking FBs is also worth studying.

We derive a single-channel probability map by taking the maximum of each binary probability in the output two-channel normalized probability map. For inputting any unseen shot gather, the resulting single-channel probability curve at each trace including the locations of missing traces is one impulse-like function, where there are the low-value probability of a certain width near FBs. The locations of the minimum values within the low-value probability zone can be interpreted as FBs, which correspond to the locations of the segmentation lines interpreted by directly using the output two-channel normalized probability map. There are properties for the single-channel probability map that

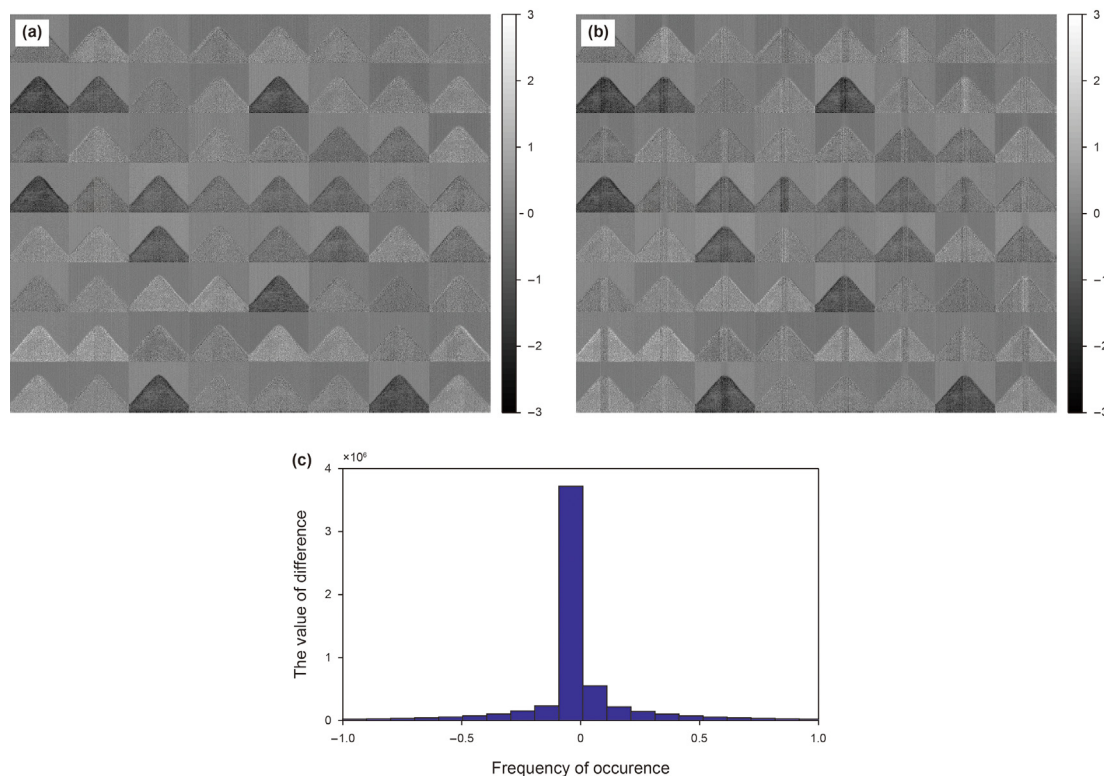


Fig. 15. Comparisons between 64 features (a–b) in the 17th convolution layer automatically extracted by applying the new well-trained 25-layer SegNet to two given gathers of Fig. 14c and d, as well as the histogram of the absolute difference between b and a. Although there are differences between the corresponding features near the missing traces, these corresponding features appear similar. Furthermore, there are obvious boundaries between the shallow background part and the deep signal-dominant part in each extracted feature even in the presence of 21 consecutive missing traces.

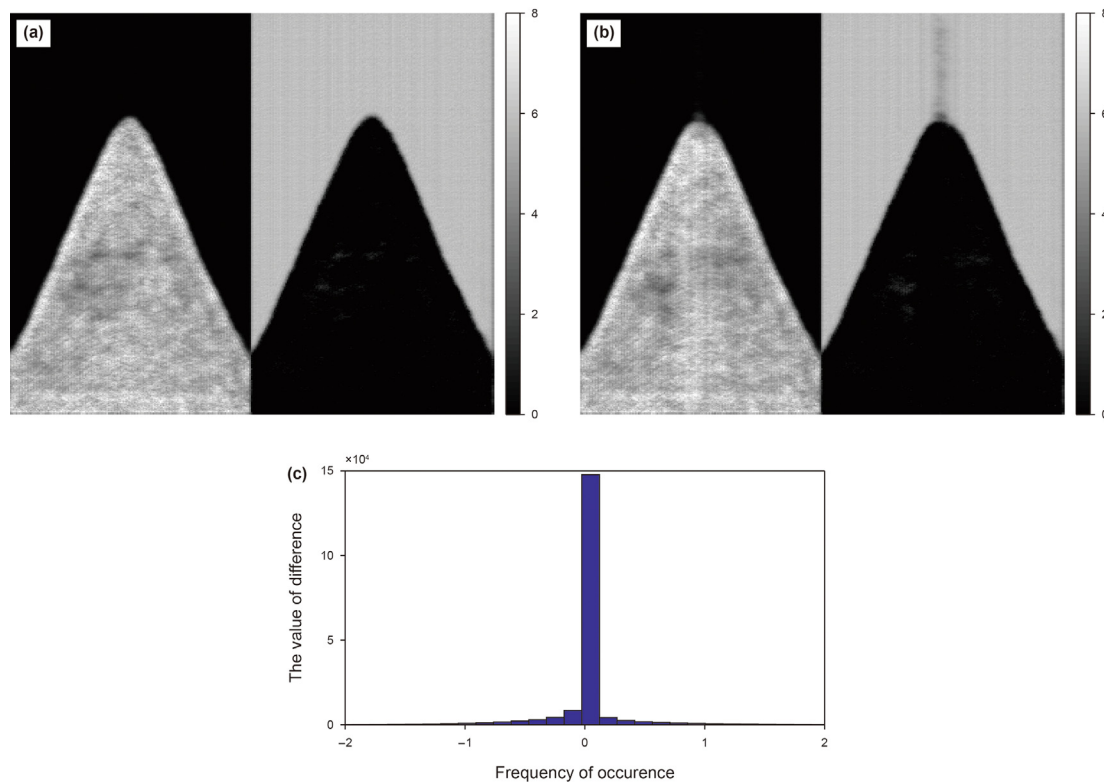


Fig. 16. Comparisons between 2 features (a–b) in the 18th (or final) convolution layer automatically extracted by applying the new well-trained 25-layer SegNet to two given gathers of Fig. 14c and d, as well as the histogram of their absolute difference between b and a. There are relatively weak differences between the corresponding features near the missing traces. In addition, there are clear boundaries between the shallow background part and the deep signal-dominant part in all extracted features.

continuous small-value probability always appears near FBs, and the minimum probability values are related to the locations of FBs. It can probably be inferred that the smaller the minimum probability value in the derived single-channel probability map, and the wider the range of low-probability values, the more uncertain the first-break picking. This may be adopted or further quantified as an additional quality-control factor to evaluate the performance of the already-trained network model from a statistical point of view, and in turn to optimize the network architecture. Furthermore, some continuously low-value probability in the derived single-channel probability map indicates one time range for picking FBs at least. It may provide a new way to combine the SegNet-based first-break picker with a certain traditional picker.

5. Conclusions

We present a SegNet-based waveform classification and first-break picking method. The method can pick FBs from shot gathers with sparsely distributed traces or missing traces directly. At the cost of learning various shot gathers with randomly missing traces, our well-trained fully-convolutional SegNet model can automatically extract spatio-temporal features and classify seismic waveform at the same time. Due to the learning of samples with missing traces and usage of several stacked pooling layers and convolution layers, the well-trained 25-layer SegNet model can adapt to automatically pick FBs of various unseen shot gathers acquired in different geometries, even when the proportion of randomly missing traces reaches 50%, 21 traces are missing consecutively, or traces are missing regularly. As our examples show, even when the proportion of randomly or regularly missing traces of each gather in the testing set reaches up to 50%, the classification accuracy of 2273 gathers is almost over 97%, and the average error between the SegNet-based automatically picked FBs and the manually picked ones is about two time samples. Although we do not impose any constraint of some common-sense factors related to spatio-temporal relationship of FBs on learning SegNet, FBs including at the locations of missing traces can be picked with good lateral continuities. Our work can be readily extended to simultaneous multi-shot or multi-domain learning and first-break picking, 3D amplitude learning and first-break picking by increasing a frequency or scale dimension, as well as simultaneous multiple-attribute learning and first-break picking, which are also our future work.

Acknowledgements

This work was financially supported by the National Key R&D Program of China (2018YFA0702504), the Fundamental Research Funds for the Central Universities (2462019QNXZ03), the National Natural Science Foundation of China (42174152 and 41974140), and the Strategic Cooperation Technology Projects of CNPC and CUPB (ZLZX 2020-03).

References

Akram, J., Eaton, D.W., 2016. A review and appraisal of arrival-time picking methods for downhole microseismic data. *Geophysics* 81 (2), 71–91. <https://doi.org/10.1190/geo2014-0500.1>

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE T Pattern Anal* 39 (12), 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>

Bergen, K.J., Jognson, P.A., De Hoop, M.V., et al., 2019. Machine learning for data-driven discovery in solid Earth geoscience. *Science* 363 (6433). <https://doi.org/10.1126/science.aau0323>

Boschetti, F., Dentith, M.D., List, R., 1996. A fractal-based algorithm for detecting first arrivals on seismic traces. *Geophysics* 61, 1095–1102. <https://doi.org/10.1190/1.1444030>

Caffagni, E., Eaton, D.W., Jones, J.P., et al., 2016. Detection and analysis of

microseismic events using a Matched Filtering Algorithm (MFA). *Geophys. J. Int.* 206 (1), 644–658. <https://doi.org/10.1093/gji/ggx182>

Chen, L.C., Papandreou, G., Kokkinos, I., et al., 2018. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE T Pattern Anal* 40 (4), 834–848. <https://doi.org/10.1109/TPAMI.2017.2699184>

Chen, Y., Zhang, G., Bai, M., et al., 2019. Automatic waveform classification and arrival picking based on convolutional neural network. *Earth Space Sci* 6 (7), 1244–1261. <https://doi.org/10.1109/TPAMI.2017.2699184>

Coppens, F., 1985. First arrival picking on common-offset trace collections for automatic estimation of static corrections. *Geophys. Prospect.* 33, 1212–1231. <https://doi.org/10.1111/j.1365-2478.1985.tb01360.x>

Dokht, R.M.H., Kao, H., Visser, R., et al., 2019. Seismic event and phase detection using time-frequency representation and convolutional neural networks. *Seismol Res. Lett.* 90 (2), 481–490. <https://doi.org/10.1785/0220180308>

Esteva, A., Kuprel, B., Novoa, R.A., et al., 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542 (7639), 115–118. <https://doi.org/10.1038/nature21056>

Galiana-Merino, J.J., Rosa-Herranz, J.L., Parolai, S., 2008. Seismic P phase picking using a kurtosis-based criterion in the stationary wavelet domain. *IEEE Trans. Geosci. Rem. Sens.* 46 (11), 3815–3826. <https://doi.org/10.1109/TGRS.2008.2002647>

Gelchinsky, B., Shivelman, V., 1983. Automatic picking of the first arrival and parameterization of traveltimes curves. *Geophys. Prospect.* 31, 915–928. <https://doi.org/10.1111/j.1365-2478.1983.tb01097.x>

Gentili, S., Michellini, A., 2006. Automatic picking of P and S phases using a neural tree. *J. Seismol.* 10 (1), 39–63. <https://doi.org/10.1007/s10950-006-2296-6>

Gibbons, S.J., Ringdal, F., 2006. The detection of low magnitude seismic events using array-based waveform correlation. *Geophys. J. Int.* 165 (1), 149–166. <https://doi.org/10.1111/j.1365-246X.2006.02865.x>

Glorot, X., Bordes, A., Bengio, Y., 2011. In: *Deep Sparse Rectifier Neural Networks*. AISTATS, pp. 315–323.

Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep Learning*. The MIT Press. <https://doi.org/10.1007/978-1-0710-017-9314-z>

Hu, L., Zheng, X., Duan, Y., et al., 2019. First-arrival picking with a U-net convolutional network. *Geophysics* 84 (6), 45–57. <https://doi.org/10.1190/geo2018-0688.1>

Ioffe, S., Szegedy, C., 2015. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, vol. 1502, 03167v3 arXiv preprint arXiv.

Jia, Y., Ma, J., 2017. What can machine learning do for seismic data processing? An interpolation application. *Geophysics* 82 (3), 163–177. <https://doi.org/10.1190/geo2016-0300.1>

Jiao, L.X., Moon, W.M., 2000. Detection of seismic refraction signals using a variance fractal dimension technique. *Geophysics* 65, 286–292. <https://doi.org/10.1190/1.1444719>

Khalaf, A., Camerlynck, C., Florsch, N., et al., 2018. Development of an adaptive multi-method algorithm for automatic picking of first arrival times: application to near surface seismic data. *Near Surf. Geophys.* 16 (5), 507–526. <https://doi.org/10.1002/nsg.12014>

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521 (7553), 436–444. <https://doi.org/10.1038/nature14539>

LeCun, Y., Bottou, L., Orr, G.B., et al., 1998. Efficient Backprop: Neural Networks. *Tricks of the Trade*, pp. 9–50. https://doi.org/10.1007/3-540-49430-8_2

Li, Z., Jia, X., Zhang, J., 2019. Deep learning guiding first-arrival travel time tomography. In: *SEG Technical Program Expanded Abstracts*, pp. 2513–2517. <https://doi.org/10.1190/segam2019-3215632.1>

Liao, X., Cao, J., Hu, J., et al., 2021. First arrival time identification using transfer learning with continuous wavelet transform feature images. *IEEE Geosci Remote Sens Lett* 17 (11), 2002–2006. <https://doi.org/10.1109/LGRS.2019.2955950>

Liu, D., Huang, J., Wang, Z., 2020. Convolution-based multi-scale envelope inversion. *Petrol. Sci.* 17 (2), 352–362. <https://doi.org/10.1007/s12182-019-00419-8>

Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation. *CVPR*, pp. 3431–3440. <https://doi.org/10.1109/TPAMI.2016.2572683>

Maity, D., Aminzadeh, F., Karrenbach, M., 2014. Novel hybrid artificial neural network based autopicking workflow for passive seismic data. *Geophys. Prospect.* 62 (4), 834–847. <https://doi.org/10.1111/1365-2478.12125>

McCormack, M.D., Zauha, D.E., Dushek, D.W., 1993. First-break refraction event picking and seismic data trace editing using neural networks. *Geophysics* 58, 67–78. <https://doi.org/10.1190/1.1443352>

Molyneux, J.B., Schmitt, D.R., 1999. First-break timing: arrival onset times by direct correlation. *Geophysics* 64, 1492–1501. <https://doi.org/10.1190/1.1444653>

Mousavi, S.M., Langston, C.A., Horton, S.P., 2016. Automatic microseismic denoising and onset detection using the synchrosqueezed continuous wavelet transform. *Geophysics* 81 (4), 341–355. <https://doi.org/10.1190/geo2015-0598.1>

Noh, H., Hong, S., Han, B., 2015. In: *Learning Deconvolution Network for Semantic Segmentation*. ICCV, pp. 1520–1528. <https://doi.org/10.1109/ICCV.2015.178>

Perol, T., Gharbi, M., Denolle, M., 2018. Convolutional neural network for earthquake detection and location. *Sci Adv* 4 (2), e1700578. <https://doi.org/10.1126/sciadv.1700578>

Pham, N., Fomel, S., Dunlap, D., 2019. Automatic channel detection using deep learning. *Interpretation* 7 (3), 43–50. <https://doi.org/10.1190/int-2018-0202.1>

Plenkers, K.J., Ritter, R.R., Schindler, M., 2013. Low signal-to-noise event detection

- based on waveform stacking and cross-correlation: application to a stimulation experiment. *J. Seismol.* 17 (1), 27–49. <https://doi.org/10.1007/s10950-012-9284-9>.
- Powers, D.M.W., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *J. Mach. Learn. Technol.* 2 (1), 37–63. <https://doi.org/10.9735/2229-3981>.
- Ronneberger, O., Fischer, P., Brox, T., 2015. In: U-net: Convolutional Networks for Biomedical Image Segmentation. https://doi.org/10.1007/978-3-319-24574-4_28 arXiv.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature* 323 (6088), 533–536. <https://doi.org/10.1038/323533a0>.
- Russakovsky, O., Deng, J., Su, H., et al., 2015. ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* 115 (3), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>.
- Sabbione, J.L., Velis, D., 2010. Automatic first-breaks picking: new strategies and algorithms. *Geophysics* 75 (4), 67–76. <https://doi.org/10.1190/1.3463703>.
- Saragiotis, C.D., Hadjileontiadis, L.J., Rekanos, I.T., et al., 2004. Automatic P phase picking using maximum kurtosis and k-statistics criteria. *Geosci. Rem. Sens. Lett. IEEE* 1, 147–151. <https://doi.org/10.1109/LGRS.2004.828915>.
- Silver, D., 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529 (7587), 484–489. <https://doi.org/10.1038/nature16961>.
- Sun, M., Li, Z., Li, Z., et al., 2019. Reconstruction of seismic data with weighted MCA based on compressed sensing. *J. Chin Geophys* 62 (3), 1007–1021. <https://doi.org/10.6038/cjg2019L0336> (in Chinese).
- Tselentis, G.A., Martakis, N., Paraskevopoulos, P., et al., 2012. Strategy for automated analysis of passive microseismic data based on S-transform, Otsu's thresholding, and higher order statistics. *Geophysics* 77 (6), 43–54. <https://doi.org/10.1190/GEO2011-0301.1>.
- Wang, B.F., Zhang, N., Lu, W.K., et al., 2019a. Deep-learning-based seismic data interpolation: a preliminary result. *Geophysics* 84 (1), 11–20. <https://doi.org/10.1190/geo2017-0495.1>.
- Wang, J., Xiao, Z., Liu, C., et al., 2019b. Deep learning for picking seismic arrival times. *J. Geophys Res-Sol Ea.* 124 (7), 6612–6624. <https://doi.org/10.1029/2019JB017536>.
- Wu, H., Zhang, B., Li, F., et al., 2019a. Semiautomatic first-arrival picking of microseismic events by using the pixel-wise convolutional image segmentation method. *Geophysics* 84 (3), 143–155. <https://doi.org/10.1190/geo2018-0389.1>.
- Wu, X., Liang, L., Shi, Y., et al., 2019b. FaultSeg3D: using synthetic datasets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. *Geophysics* 84 (3), 35–45. <https://doi.org/10.1190/geo2018-0646.1>.
- Xie, T., Zhao, Y., Jiao, X., et al., 2019. First-break automatic picking with fully convolutional networks and transfer learning. In: SEG Technical Program Expanded Abstracts, pp. 4972–4976. <https://doi.org/10.1190/segam2019-3215277.1>.
- Yao, G., Wu, D., Wang, S., 2020. A review on reflection-waveform inversion. *Petrol. Sci.* 17 (2), 334–351. <https://doi.org/10.1007/s12182-020-00431-3>.
- Yilmaz, O., 2001. Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data. Society of Exploration Geophysicists. <https://doi.org/10.1190/1.9781560801580>.
- Yuan, S., Liu, J., Wang, S., et al., 2018. Seismic waveform classification and first-break picking using convolution neural networks. *Geosci. Rem. Sens. Lett. IEEE* 15 (2), 272–276. <https://doi.org/10.1109/LGRS.2017.2785834>.
- Yung, S.K., Ikelle, L.T., 1997. An example of seismic time-picking by third-order bicoherence. *Geophysics* 62, 1947–1952. <https://doi.org/10.1190/1.1444295>.
- Zhao, Y., Yuan, S., Xie, T., et al., 2019. Direct seismic waveform classification and first-break picking based on fully convolutional networks from shot gathers with missing traces. In: SEG 2019 Workshop: Mathematical Geophysics. Traditional vs. Learning, pp. 26–30. https://doi.org/10.1190/iwmg2019_071.
- Zhu, W.Q., Beroza, G.C., 2019. PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophys. J. Int.* 216 (1), 261–273. <https://doi.org/10.1093/gji/ggy423>.