

JGR Solid Earth

RESEARCH ARTICLE

10.1029/2022JB023944

Special Section:

Machine learning for Solid Earth observation, modeling and understanding

Key Points:

- We propose a new method to estimate crustal H and κ from receiver functions based deep learning
- Systematic synthetic tests show that this method is powerful and more accurate than H - κ and H - κ - c methods
- Applications to real data show that the proposed method is robust and reliable in a wide range of geological settings

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

X. Song,
xiao.d.song@gmail.com

Citation:

Wang, F., Song, X., & Li, J. (2022). Deep learning-based H - κ method (HkNet) for estimating crustal thickness and V_p/V_s ratio from receiver functions. *Journal of Geophysical Research: Solid Earth*, 127, e2022JB023944. <https://doi.org/10.1029/2022JB023944>

Received 2 JAN 2022

Accepted 24 MAY 2022

Deep Learning-Based H - κ Method (HkNet) for Estimating Crustal Thickness and V_p/V_s Ratio From Receiver Functions

Feiyi Wang¹ , Xiaodong Song^{1,2} , and Jiangtao Li³ 

¹Institute of Theoretical and Applied Geophysics, School of Earth and Space Sciences, Peking University, Beijing, China,

²Hebei Hongshan National Observatory on Thick Sediments and Seismic Hazards, Beijing, China, ³Department of Geophysics, School of Geodesy and Geomatics, Wuhan University, Wuhan, China

Abstract The teleseismic receiver function (RF) is commonly used to determine major interfaces of the Earth. If the crustal V_p is known approximately, the Moho converted P_s phase and crustal multiple reverberations can be used to determine the thickness (H) and average V_p/V_s ratio (κ) of the crust. A widely used method for this is H - κ stacking (Zhu & Kanamori, 2000, <https://doi.org/10.1029/1999JB900322>), which uses grid search and superposition to find the maximum coherent energy of the Moho P_s and its reverberated multiples phases. However, this method assumes a homogeneous isotropic crust and a flat Moho. Furthermore, it is affected by the reference crustal V_p . Improved methods, such as the H - κ - c method for anisotropic media and inclined interfaces (J. Li et al., 2019, <https://doi.org/10.1029/2018JB016356>), help alleviate the problem. In this paper, we propose a new method that uses deep learning to estimate H and κ . Our method is divided into two steps. The first step employs a denoise architecture (the DenoiseNet) to reduce the noise level of the RFs and restore missing back-azimuthal information. In the second step, our new deep learning network (the HkNet) is used to estimate H and κ . Deep learning has the inherent ability to automatically extract complex features from RFs, which allows us to estimate the parameters in complex media with different crustal V_p . Synthetic data tests show that the proposed method achieves better accuracy than the H - κ and H - κ - c methods. Applications to real data show that the proposed method is robust and reliable in a wide range of geological settings.

Plain Language Summary Crustal thickness (H) and V_p/V_s ratio (κ) are fundamental parameters of the Earth. Here, we propose a deep learning method to extract the information from receiver functions (RFs). We designed two convolutional neural networks: (a) the DenoiseNet to suppress the noise and re-construct the missing back-azimuthal information of the RFs and (b) the HkNet to estimate H and κ from the processed RFs. Systematic synthetic tests show that this method is powerful and more accurate than existing methods. It yields reliable results with real data and is generally applicable to a wide range of geological settings.

1. Introduction

A receiver function (RF) is the response of the Earth's structure below a seismometer to an incident teleseismic wave and consists of a series of P -to- S (P_s) or S -to- P (S_p) converted waves generated at structural interfaces (Langston, 1979), mainly those generated by the velocity discontinuities in the crust and upper mantle below the seismometer. Crustal thickness (H) and P -wave and S -wave velocity ratio (κ) are important parameters reflecting the crustal structure and internal material composition and can provide an important basis for regional tectonics and dynamics. Zhu and Kanamori (2000) noted that when the Moho converted phase P_s and crustal multiple waves are considered at the same time, H and κ can be estimated under the assumption of the average crustal P wave velocity. This is currently the most commonly used H - κ method. The H - κ algorithm stacks the amplitudes of RFs at predicted arrival times for P_s and crustal multiples ($PpPs$ and $PpSs + PsPs$; here referred to as $M1$ and $M2$, respectively, for convenience, as in J. Li et al., 2019) for different values of H and κ in a grid search.

The assumptions of the H - κ method are very simple, which leads to its limitations. It assumes that the crustal medium is isotropic, the Moho surface is horizontal, and the P -wave velocity of the crust is fixed. However, the above assumptions cannot be satisfied under complex geological settings, so the results of the H - κ method may be inaccurate. The Chinese continent is known to have highly heterogeneous crustal structure and extremely variable Moho depth, as revealed for example, by seismic tomography (Bao et al., 2015; Han et al., 2021; M. Li

et al., 2022; Shen et al., 2016; X. Zhang et al., 2022) and receiver functions (C. Li et al., 2014; J. Li et al., 2017; Y. Li et al., 2014; X. Wang et al., 2017).

Thus, for Chinese continent, because of the complex crustal structure and moho variation, the results obtained by the H - κ method may be less reliable. In view of the above shortcomings, several studies have proposed improved methods that adapt to more complicated crustal structures (Frederiksen & Bostock, 2000; Frederiksen et al., 2003; Hammond, 2014; Kaviani & Rumpker, 2015; Levin & Park, 1997a, 1997b; Liu & Park, 2017; Park & Levin, 2016; Rumpker et al., 2014; Savage, 1998; H. Zhang et al., 2009; Zhu et al., 1995), even though the assumptions of these methods are still relatively simple. J. Li et al. (2019) proposed a more generalized improved method called H - κ - c , which considers anisotropy with a plunge and dipping Moho. The improved method has been quickly applied to a wide range of geological settings (Cui et al., 2021; J. Li et al., 2019; W. Li et al., 2021; Tan & Nie, 2021; B. Zhang et al., 2020). However, the method also has some limitations. The method needs to assume the average crustal V_p (e.g., often at 6.3 km/s) in the application of the method but is in fact not well known. It also depends on data quality; it is not suitable for data in which multiple phases are not obvious.

Deep learning has been rapidly and widely applied to geophysical investigations (Bergen et al., 2019; Yu & Ma, 2021). In some studies of exploration geophysics, it has produced results comparable to and in some cases surpassing expert human performance (Bi et al., 2021; Liang et al., 2019; Reading et al., 2015). In seismology, a successful example is the use of deep learning for earthquake detection and phase picking (Jiang et al., 2021; Z. Li, 2021; Mousavi et al., 2020; J. Wang et al., 2019; Wong et al., 2021; Yu & Ma, 2021; L. Zhang et al., 2020; P. C. Zhou et al., 2021; Y. Zhou et al., 2021). However, applications of deep learning methods in seismic structure inversions have been limited thus far, including super resolution images from low resolution by a CycleGAN and crustal thickness estimated from Rayleigh surface wave based on Deep Neural Networks (Cheng et al., 2019; Niu et al., 2020). A convolutional neural network (CNN) can automatically extract the internal features of an image and construct complex non-mapping relationships. In our case, RFs can essentially be considered as time series. Here, we turn the problem of time series into an image recognition problem, use convolution kernels to identify the waveform information at different positions, and find the correlation between seismic phases. The essence of the classic H - κ method is to estimate H and κ by enhancing the energy of the three phases of P_s , $M1$, and $M2$. In other words, the two parameters H and κ are another way to characterize seismic phase information. The deep learning model works by learning the general characteristics of seismic waveforms and phases from high-level representations. Therefore, if the seismic phase information can be identified through deep learning, the results of H and κ can also be obtained. Fortunately, the seismic phase detection ability of deep learning has been verified extensively.

In this paper, we propose a H and κ estimation method based on deep learning. Relative to prior methods, it can be applied to more generalized situations, including plunging anisotropy and dipping interfaces of different V_p in the crust; it can also be applied to poorer-quality data. After the deep learning model is trained, its inference time is only 1 s. Synthetic data tests of model with different complexity levels show that this method has higher accuracy than the traditional H - κ and H - κ - c methods. We applied our method to the real data of complex settings from the Chinese National Digital Seismic Network (CNSN) and obtained reliable results, which demonstrates the potential of the method for broad applications.

2. Methods

The CNN is an important branch of deep learning, which can automatically extract deep features of images. It has achieved excellent results in computer vision, natural language processing, recommendation systems, and many more fields. Compared with synthetic ideal data without noise, the poor quality of real data is mainly reflected in the loss of certain azimuth information, large amounts of noise, and ambiguous multiple phases. Therefore, we first use a CNN to distinguish noise from effective information and to construct a connection between different back-azimuthal RFs. In the second step, we use a deep learning method called HkNet to predict H and κ from the denoised RFs. These two main steps of our proposed method are illustrated in Figure 1.

2.1. Architectures of the Networks

Figure 2 shows the architectures of the CNN networks. To achieve automatic estimates of H and κ from RFs, we adopted and modified the U-Net and ResNet architectures (He et al., 2016; Ronneberger et al., 2015). The first

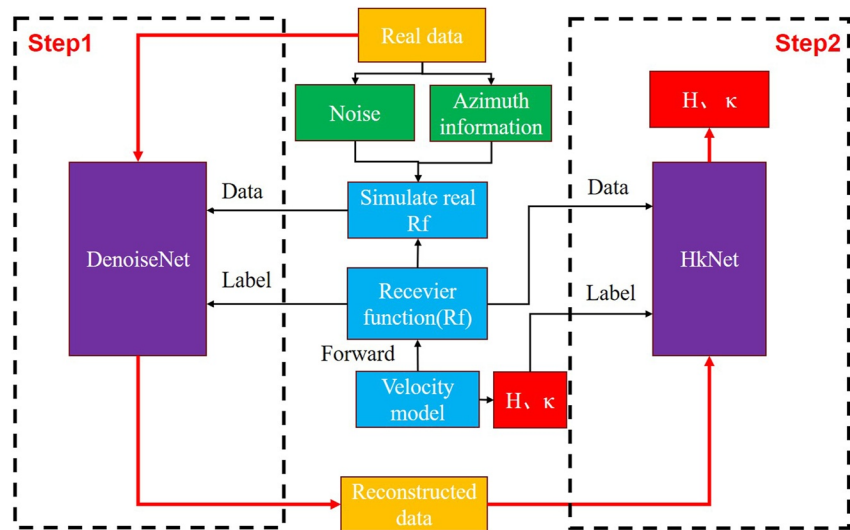


Figure 1. Workflow of the proposed method. The first step (DenoiseNet) is on the left, whose main purpose is to denoise and re-construct the missing lacking back-azimuthal information. The second step (HkNet) is on the right, whose main purpose is to predict H and κ based on the output from the first step.

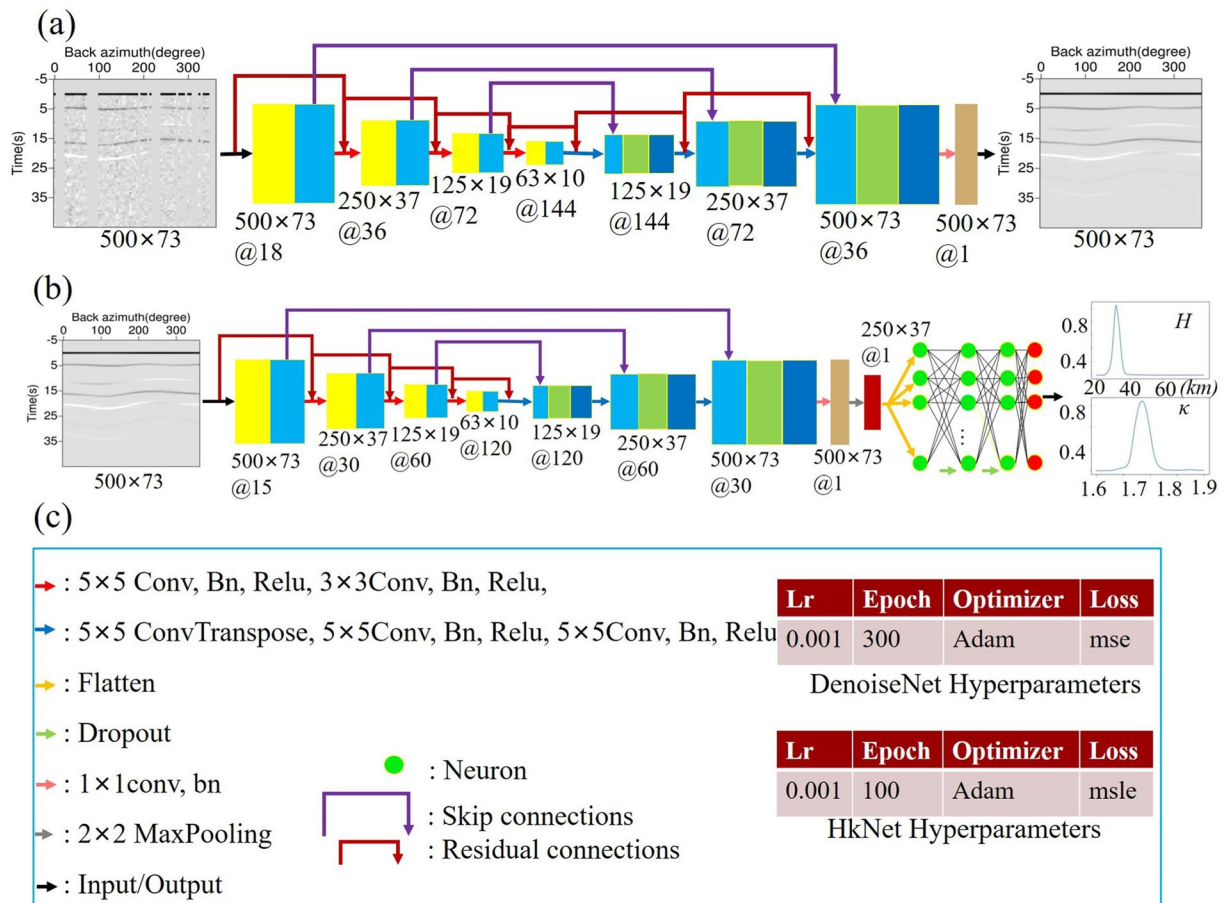


Figure 2. Convolutional neural network networks used in this study. The networks include two parts: (a) DenoiseNet, (b) HkNet, with symbols and hyperparameters of the networks listed in panel (c).

Table 1
Parameter Ranges of Training Models

Parameter ^a	Range
Thickness	25–75 km
V_p	5.0–7.0 km/s
κ	1.6–1.9
S%	0–10
Trend of fast axis	0–180°
Plunge of fast axis	0–60°
Moho direction	0–360°
Moho dip	0–15°

^aAnisotropy amplitude S% indicates the difference between fast and slow speeds relative to the average speed as a percentage.

step is to preprocess the RFs by denoising them with the DenoiseNet and to re-construct the missing RFs of some back-azimuthal directions. When constructing the DenoiseNet, we first gradually extract the deep information of the RFs through convolution with a step size of 2 while using long and short connections to merge shallow information, taking both global (full-time window information) and local (phase information) features into account while preventing the gradient from disappearing or exploding when the error is propagated backwards. The structure of the DenoiseNet is shown in Figure 2a. The second step is to use a fully-connected CNN, called HkNet, to predict H and κ . The network structure is shown in Figure 2b. This fully-connected CNN is introduced to deepen the structure and enhance the learning and prediction capabilities. We use the dropout layer and dynamic learning rate strategy to prevent network overfitting. We summarize the definitions of the different operations in Figure 2c. The acronyms and systems are conventional to the CNN (Raschka, 2017; Ronneberger et al., 2015). Please refer to the Supporting Information S1 for details.

2.2. Training Data Sets

To train an efficient network, a suitable large training set is generally needed. In this study, we used simulated data and information from the real data to generate the training data sets.

First, we generated random 1-D velocity models with different parameters listed in Table 1. The velocity models have only one layer in the crust. We used the Raysum algorithm (Frederiksen & Bostock, 2000) to calculate the synthetic ideal RFs with a ray parameter of 0.06 s/km and Gaussian parameter of 3.0 for each velocity model. We set the size of each RF image to 73×500 , where 73 is the number of stacked 5°-back-azimuthal bins, and 500 is the number of time samples. The sampling interval is 0.1 s and the time window of the RFs is -5 to 45 s. Second, we added noise from real stations and random back-azimuthal distributions to ideal RFs to generate simulated realistic RFs data set.

The input data of DenoiseNet are simulated realistic RFs and the labels are synthetic ideal RFs (in deep learning, a label means the result we want to identify the input data from Net). For the HkNet, the input data are synthetic ideal RFs and the labels are H and κ from velocity models. At the beginning we tried different strategies to design the network, including directly training a single network without DenoiseNet and using simply two variables for labels, but the result was very poor, especially in the application to the real data (Figures S2–S7 in Supporting Information S1). Therefore, we introduce DenoiseNet separately and inspired by the study of Kuang et al. (2021), we give these two values (H and κ) in the form of a Gaussian distribution. The H curve contains 650 points, and the κ curve contains 400 points. The peaks (location parameter μ in Gaussian distribution) are real H and κ values.

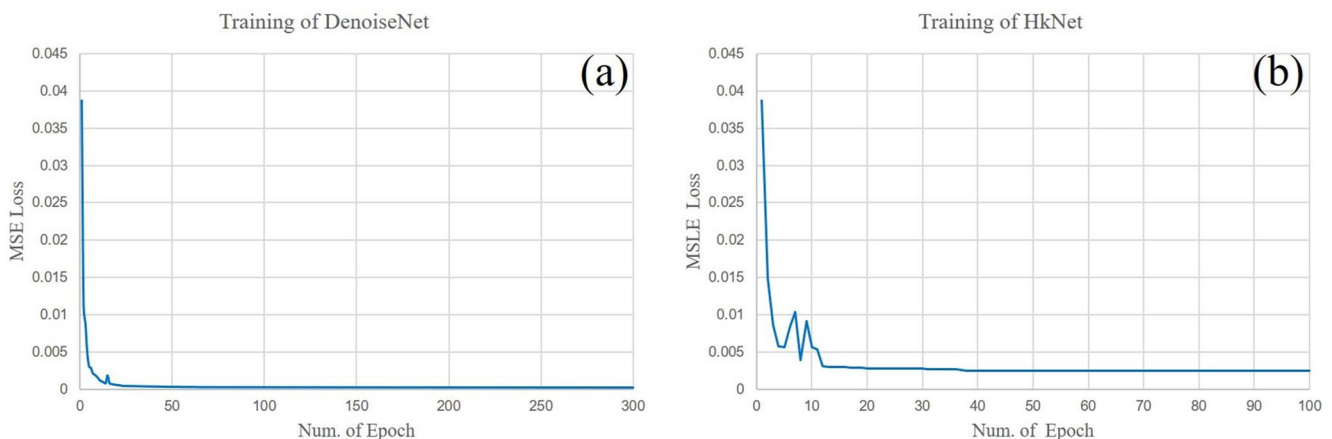


Figure 3. Loss versus epoch during the training process. (a) Mean-squared-error for DenoiseNet; (b) mean-squared-logarithmic-error for HkNet.

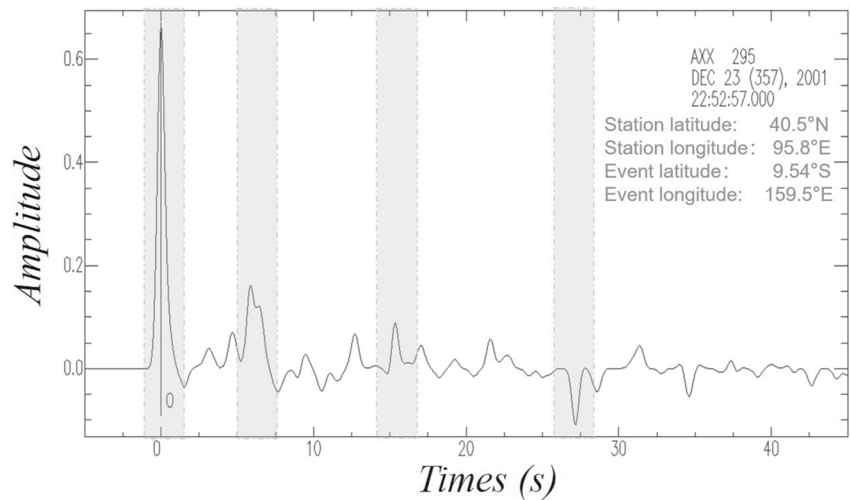


Figure 4. Example of how noise is extracted from the real data. The larger absolute amplitudes in the shaded parts are removed and the remainder are considered to be noise.

The standard deviation (size parameter σ in Gaussian distribution) is 50 points. If the standard deviation is too small, it is difficult to fit labels by deep learning. If the standard deviation is too large, the label resolution will be insufficient, which will significantly affect the prediction accuracy. In our work, the training data sets contained 60,000 input-output pairs and the test data sets contained 10,000. The test data and the corresponding models were not used in the training process. Loss functions during the training process are shown in Figure 3.

2.3. Real Data and Processing

To facilitate comparison, we applied our method to the real data from 38 broadband stations of CNDSN that J. Li et al. (2019) used previously (Figure 8). The construction of RFs followed a typical procedure (e.g., Huang et al., 2014, 2015; Zhu & Kanamori, 2000) and we used the RFs from J. Li et al. (2019) directly for better compar-

Table 2
Descriptions of Synthetic Models (Modified From J. Li et al., 2019)

SYN#	Model name	Model description ^a							
		Thickness (km)	Vs (km/s)	K	S%	Trend	Plunge	Direction	Dip
0	REF	40	3.6	1.75	0	0°	0°	0°	0°
1	DIP8	40	3.6	1.75	0	0°	0°	90°	8°
2	ANI8	40	3.6	1.75	8	0°	0°	0°	0°
3	PLG60	40	3.6	1.75	8	30°	60°	0°	0°
4	S_DIP_ANI	40	3.6	1.75	5	45°	0°	90°	5°
5	L_DIP_ANI	40	3.6	1.75	8	30°	0°	90°	10°
6	DIP_PLG	40	3.6	1.75	8	30°	45°	90°	10°
7	CPLX	20	3.6	1.75	8	15°	10°	240°	15°
		20	3.6	1.75	10	30°	20°	290°	10°
8	RDSTR	Same as SYN7 but using real RF distribution from station SZN.							

^aAnisotropy amplitude S% indicates the difference between fast and slow speeds relative to the average speed as a percentage; trend of the fast axis is measured from the north on the horizontal plane (0–180°); plunge of the fast axis is measured from the horizontal plane (–90–90°; positive and negative when the fast axis rotates clockwise and counter-clockwise, respectively); dip of an interface is the angle of the plane with the horizontal plane (0–90° under the “right hand rule”); dip direction is horizontal projection of the down-dip direction measured from the north, which is also equal to the strike of the plane plus 90° (0–360°). All mantle Versus = 4.5 km/s, mantle V_p = 8.1 km/s.

Table 3
Errors of H - κ , H - κ - c , and Proposed Methods From Synthetic Tests^a

SYN#	Model name	Traditional H - κ		H - κ - c		HkNet		SSIM	PSNR (dB)
		dH (km)	$d\kappa$	dH (km)	$d\kappa$	dH (km)	$d\kappa$		
1	DIP8	-2.10	0.040	-0.90	0.010	-0.10	0.008	/	
2	ANI8	0.00	0.060	-0.10	0.010	-0.70	-0.007	/	
3	PLG60	0.00	-0.050	0.00	-0.050	-0.50	-0.048	/	
4	S_DIP_ANI	-0.60	0.020	-0.50	0.010	0.70	-0.006	/	
5	L_DIP_ANI	-3.10	0.110	-1.50	0.030	0.70	-0.013	/	
6	DIP_PLG	-2.50	0.020	-1.50	0.000	0.10	-0.028	/	
7	CPLX	-3.50	0.110	-1.60	0.030	0.10	0.017	/	
8	RDSTR	-3.90	0.120	-1.50	0.020	-1.50	0.007	0.971	36.2
AVG (abs)	/	1.96	0.066	0.95	0.020	0.55	0.017	0.971	36.2

^aAll results are shown by the differences (dH and $d\kappa$) from the input values of 40 km and 1.75, respectively. The minimum errors of the three methods are in bold font.

isons. Earthquakes of $M \geq 5$ at epicentral distances from 30 to 95° were selected. All the vertical-component waveforms were visually inspected for consistency by discarding noisy events. After that, we removed the mean and linear trend in the waveform data and down-sampled them to 10 Hz. A band-pass filter from 0.05 to 2 Hz was applied. We used three-component P -wave waveforms (30 s before direct P and 90 s after) to calculate the RFs at each station using time-domain iterative deconvolution method (Juan Pablo & Charles, 1999). A Gaussian low-pass filter (Gaussian parameter of 2.5) was applied to reduce the noise above ~ 1.2 Hz. For each station, all RFs were sorted by BAZ and inspected to identify and remove poor RFs (with inconsistent waveforms or obviously wrong negative direct P pulses).

2.4. Simulated Realistic RFs

The real data contains noise and back-azimuthal distribution is uneven with gaps. Thus, the training data set has to consider these factors. Otherwise, the trained network will not perform well for the real data. In the 5° back-azimuthal bins, the 38 stations have 24 to 48 missing data gaps with the mean of 35 gaps in each station among the total 73 bins, that is, the data gaps are about 33%–66% with the mean of 48%. Thus, in generating

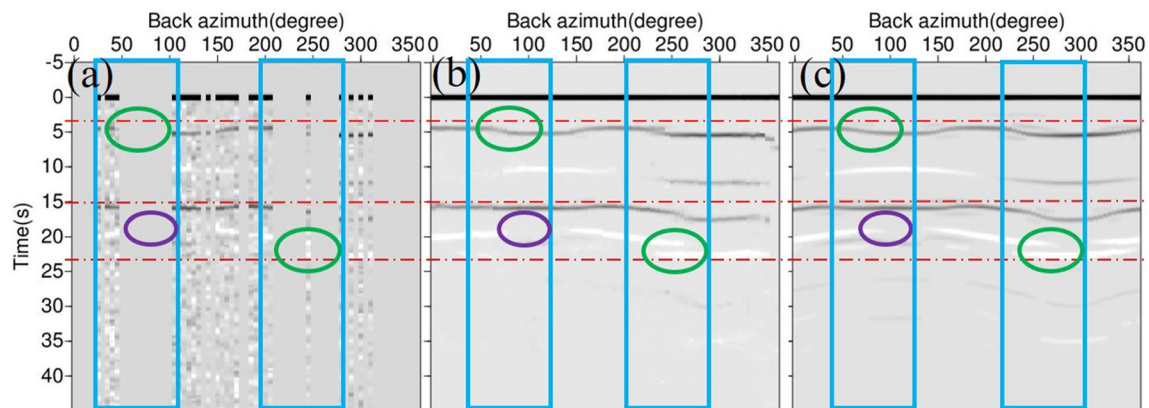


Figure 5. A demonstration of the application of the DenoiseNet with synthetic data. The example is for SYN8 (Table 2). (a) Simulated realistic receiver functions (RFs) (with added noise and realistic azimuthal distribution), (b) denoised RFs, and (c) synthetic ideal RFs (with full and uniform azimuthal distribution and without noise). Inside the blue rectangles indicate regions where the back-azimuthal information is missing from the simulated realistic RFs. The red dashed lines mark the main phases (directed P , P_s , $M1$, and $M2$) for references. Inside the green ellipses are the phases that should have existed, and these phases were indeed restored in the denoised RFs. The purple ellipses are also missing parts of the simulated realistic RFs, but there is no phase after denoising because there was no seismic phase in the synthetic ideal RFs.

Table 4
Descriptions of Synthetic Models With Different Crustal Vp

SYN#	Model name	Model description							
		Thickness (km)	Vp (km/s)	κ	S%	Trend	Plunge	Direction	Dip
9	S_Vp	40	5.5	1.75	5	45°	0°	90°	5°
10	F_Vp	40	6.8	1.75	5	45°	0°	90°	5°
11	S_RDSTR	Same as SYN9, but real RF distribution and noise from station AXX are used.							
12	F_RDSTR	Same as SYN11, but real RF distribution and noise from station AXX are used.							

the simulated training set, we considered the real back-azimuthal distributions in two ways. One way is that set a probability to remove the RF in any given bin. Another is set a probability to remove consecutive bins (3–6 bins) at a random location. As a result, the missing rate of RFs is 30%–70% with the mean of 50% in our training set. The noise we added to synthetic data were from the real data. First, within the time window of –5 to 45 s of each RF, we removed the largest parts (first three parts with the positive and first two with the negative) of the absolute amplitudes (shadowed parts in Figure 4), and consider the remainder to be noise. By performing this operation on the RFs of different stations and different back-azimuthal angles, we obtained a noise data set. Then, these noise points are randomly selected and added to the synthetic data points. Therefore, if there are small phases in a single one noise data, for the 73 RFs of the entire station, their noise is basically irrelevant and has little effect on the result.

3. Results

3.1. Synthetic Tests

3.1.1. Velocity Models of Different Complexities

We first tested models that have been studied before for H - κ - c method (J. Li et al., 2019) (Table 2). These models include one or two layers in the crust, different forms of anisotropy, different dipping interfaces, as well as ideal and realistic back-azimuthal distributions.

To ensure a fair comparison with the H - κ - c method, we skipped the first step of our method and directly used the HkNet to predict H and κ on the synthetic ideal RFs of models SYN1–7. For SYN8, which includes a real RF back-azimuthal distribution, we first used the DenoiseNet to preprocess the simulated realistic RFs and then input the denoised RFs into the HkNet to predict H and κ . The results of the tests are listed in Table 3.

As shown in Table 3, the error of H is reduced by 42.1% on average relative to the H - κ - c method, and the error of κ is reduced by 15.0%. Note that for the H - κ and H - κ - c methods, the dH values are systematically all negative for all models with dipping interface(s), as explained by J. Li et al. (2019). However, in our method, dH and $d\kappa$ are both random, which is due to the method of deep learning itself. It is interesting that our results are also better than those of the H - κ and H - κ - c methods for SYN7 and SYN8 because these two velocity models

Table 5
Errors of Traditional H - κ , H - κ - c , and Proposed Methods From Synthetic Tests of Different Crustal Vp^a

SYN#	Model name	Traditional H - κ		H - κ - c		HkNet		SSIM	PSNR (dB)
		dH (km)	$d\kappa$	dH (km)	$d\kappa$	dH (km)	$d\kappa$		
9	S_Vp	6.00	0.000	6.0	–0.010	0.10	–0.001	/	
10	F_Vp	–6.10	0.080	–6.50	0.120	0.10	–0.001	/	
11	S_RDSTR	6.0	0.000	6.0	–0.010	0.70	0.000	0.991	40.3
12	F_RDSTR	–4.00	0.040	–4.00	0.030	0.70	–0.007	0.997	45.7
AVG (abs)	/	5.50	0.033	5.63	0.043	0.40	0.002	0.994	43.0

^aAll results are given as differences (dH and $d\kappa$) from the input values of 40 km and 1.75, respectively. The minimum errors of the three methods are in bold font.

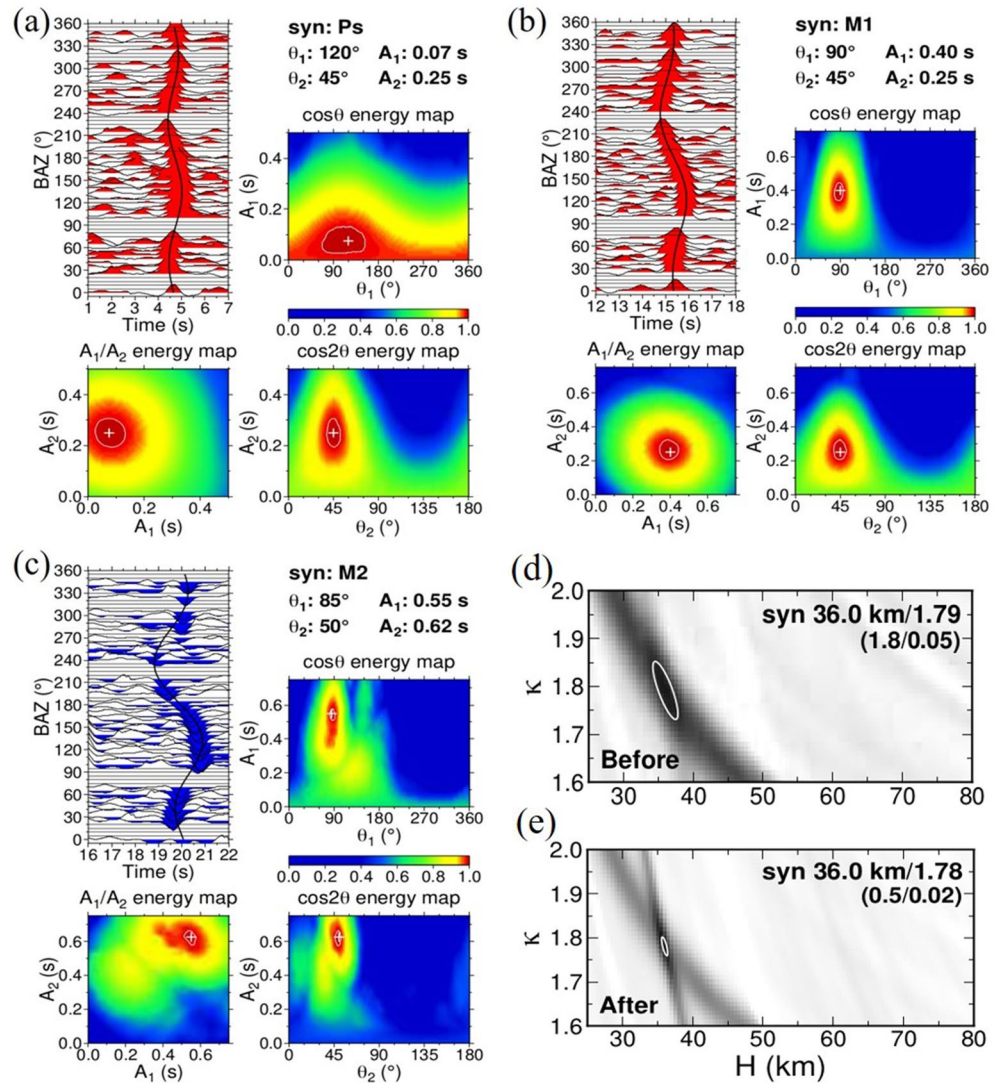


Figure 6. Results from H - κ and H - κ - c methods for a synthetic model, using the H - κ - c package (J. Li et al., 2019). This is a synthetic test example for the synthetic model with $V_p = 6.8$ km/s, $H = 40.0$ km, $\kappa = 1.75$ (Table 4, SYN12). Panels indicate harmonic fitting to the arrival times of (a) P_s , (b) $M1$, and (c) $M2$ in RFs and H - κ stacking (d) before and (e) after harmonic corrections. Panels (a–c) show the fitting curves with $\cos\theta$ and $\cos2\theta$ functions, energy stacking maps of the grid search for harmonic parameters, and the search results. The 99% contour is plotted in each energy map to show the empirical uncertainty. The weightings of P_s , $M1$, and $M2$ for the H - κ stacking are 0.7, 0.2, and 0.1 in the traditional H - κ in panel (d) and 0.5, 0.4, and 0.1 in the H - κ - c after the harmonic corrections in panel (e), respectively. The obtained H and κ values, their errors, the error ellipse, and the input (crosses) are shown in the stack diagrams.

are different from our models in our training set, which have only one layer in the crust. On the other hand, in SYN7 and SYN8, the anisotropy of 20 km in the upper half is different from the anisotropy of 20 km in the bottom half, and the intra-crustal interface has different dip angles and directions. This shows that our deep learning method has indeed learned to some extent the complex mapping relationship between the RFs and predicted parameters.

We used subjective and objective methods to evaluate the results of the DenoiseNet. The subjective evaluation used visual differences (such as main phases locations) to the naked eye in the image to analyze the effect of the DenoiseNet. The objective evaluation used the structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR) (indices that are commonly used in image processing) to evaluate image quality (Z. Wang et al., 2004; Welstead., 1999). The range of SSIM is -1 to 1 . When two images are exactly the same, SSIM is equal to 1 .

We can use this to evaluate the recovery of the main phases in the denoised RFs. The PSNR is used to evaluate the effect of noise reduction. The larger the PSNR (ideally >30 dB), the better the effect of the noise reduction.

Examining the preprocessing results of SYN8 (Figure 4), we can see that the missing data (in the blue rectangular boxes; back-azimuthal angles of 50–100° and 210–280°) are re-constructed and the noise has been suppressed. First, the main phases in the denoised RFs and synthetic ideal RFs at different back-azimuthal angles are essentially the same (marked by the red dashed line), which indicates that the most important RF information is recognized. Second, the denoised RFs contain the real phase information that appears in the synthetic ideal RFs but is lacking in the simulated realistic RFs (green ellipses, Figure 4). Third, there are regions (purple ellipses) where no seismic phase is observed in the simulated realistic RFs or the denoised data. This is because in the synthetic ideal RFs, there is no seismic phase in the corresponding parts. The SSIM and PSNR values, which were 0.155 and 22.5 before the DenoiseNet, reached 0.971 and 36.2, respectively indicating that the DenoiseNet indeed denoised the simulated realistic RFs and recovered the main information of the synthetic ideal RFs very well. Therefore, the above results show that the DenoiseNet can not only reduce noise, but more importantly, it can determine the correlation between the RFs of different back-azimuthal angles and correctly restore the missing information based on existing information. This will greatly facilitate the subsequent processing of the RFs. Comparing the final H and κ results, we can see that our method is more stable than previous methods.

3.1.2. Models With Different P -Wave Velocities

In the H - κ and H - κ - c methods, the average V_p in the crust is set to 6.3 km/s. However, the real V_p in the crust varies; its range is approximately 5.5–6.8 km/s, for example, from the global crustal reference model Crust1.0 (Laske et al., 2012). In our training set, the V_p in the crust of the velocity models is set at a wide range of 5.0–7.0 km/s, so our method can theoretically be adapted to different crustal V_p .

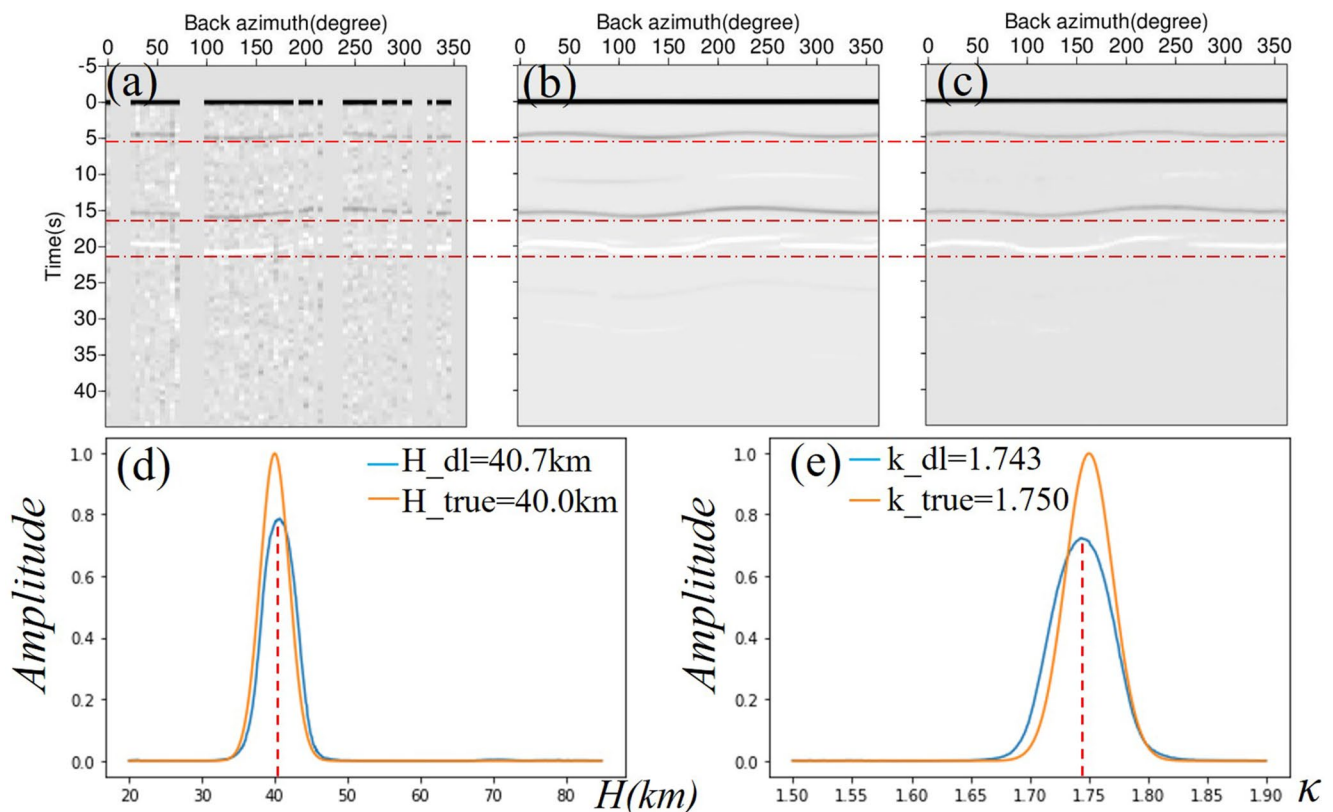


Figure 7. Results from the proposed method in this study for the same synthetic model as in Figure 5. (a) Simulated realistic receiver functions (RFs), (b) denoised RFs, and (c) synthetic ideal RFs from the SYN12 model. The red dashed lines mark the main phases. The results of our method (dl: deep learning) are shown in panels (d and e). The red dashed lines, which mark the peaks of blue curves, represent the predicted H and κ values. The blue curves are the output of HkNet, and their peak values are the predicted H and κ .

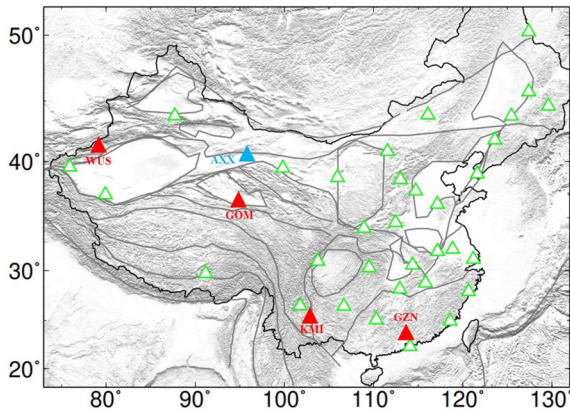


Figure 8. Map of major blocks (block boundaries indicated by gray lines), and distribution of the 38 Chinese National Digital Seismic Network stations (triangles) in this study. The marked stations (red) are examples highlighted in the main figures. Station AXX marked in blue is the one recording the noise shown in Figure 4.

We tested models of slow (SYN9, S_Vp) and fast (SYN10, F_Vp) crustal P -wave velocities and with realistic noise and back-azimuthal distributions (SYN11, S_RDSTR; SYN12, F_RDSTR) (Table 4). The results are listed in Table 5. As an example, Figures 6 and 7 show the results obtained by the H - κ - c and our methods for the SYN12 model, respectively. The H - κ - c method fits the trends of P_s , $M1$, and $M2$ with back-azimuthal angle well (Figure 6). Compared with the traditional H - κ method, the energy is much more focused after the harmonic fitting, and the standard deviations are also reduced greatly. However, the final results have large errors compared with the true values (errors in H and κ are 4 km and 0.030, respectively), similar to the errors from the traditional H - κ method. On the other hand, our method has much smaller errors of 0.70 km in H and -0.007 (Figure 7). This is because the H - κ or H - κ - c method assumes crustal $V_p = 6.3$ km/s, which differs greatly from the V_p of the SYN9–12 models so that the wrong results are obtained when calculating the theoretical phase arrival times. On average, the HkNet reduces the error of H by 92.9% and the error of κ by 95.3% over the H - κ - c method. The tests clearly demonstrate that our method has strong adaptability to different crustal V_p , even in extreme cases.

3.2. Application to Real Data

Figures 9 and 10 show the results of two representative stations at different locations (GZH and KMI, respectively). Overall, the noise of the RFs was significantly suppressed, and the phase information was also significantly enhanced. For the GZH station (Figure 9), the P_s , $M1$, and $M2$ phases in the original RFs are relatively continuous and clear, which can be used to verify the validity of the re-constructed information. It can be seen that the P_s , $M1$, and $M2$ phases in the denoised RFs agree very well with the original RFs, and the continuity is greatly enhanced. The H and κ predicted by the HkNet are 32.40 and 1.747, respectively, which are close to the results of the H - κ - c method ($h = 30.40$, $\kappa = 1.710$). The KMI station has more complicated RFs (Figure 10). There is almost no obvious phase in the original RFs. Therefore, in this case, if the H - κ method is applied, the accuracy of the results will be greatly reduced. When applying the H - κ - c method, the search parameters must be very carefully selected; the results can only be considered reliable when the correct time window is selected. After processing by the DenoiseNet, the P_s phase was recovered well, but it seems that the other two phases could not be clearly distinguished. This may be due to the complexity of the original data itself. The H and κ predicted by the HkNet are 42.60 and 1.820, respectively, which are close

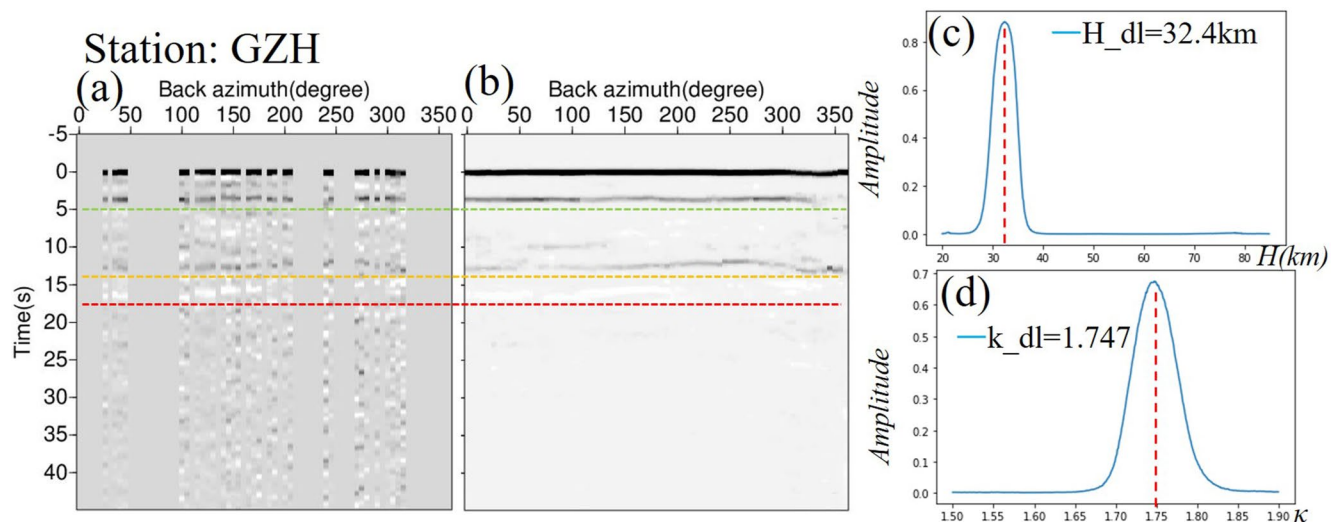


Figure 9. Results of GZH station: (a) real receiver functions (RFs), (b) denoised RFs, and (c) H and (d) κ predicted by the HkNet. The dashed lines mark the main phases (green: P_s ; orange: $M1$; red: $M2$). Obvious phases (P_s , $M1$, $M2$) in the denoised RFs can be seen at this station.

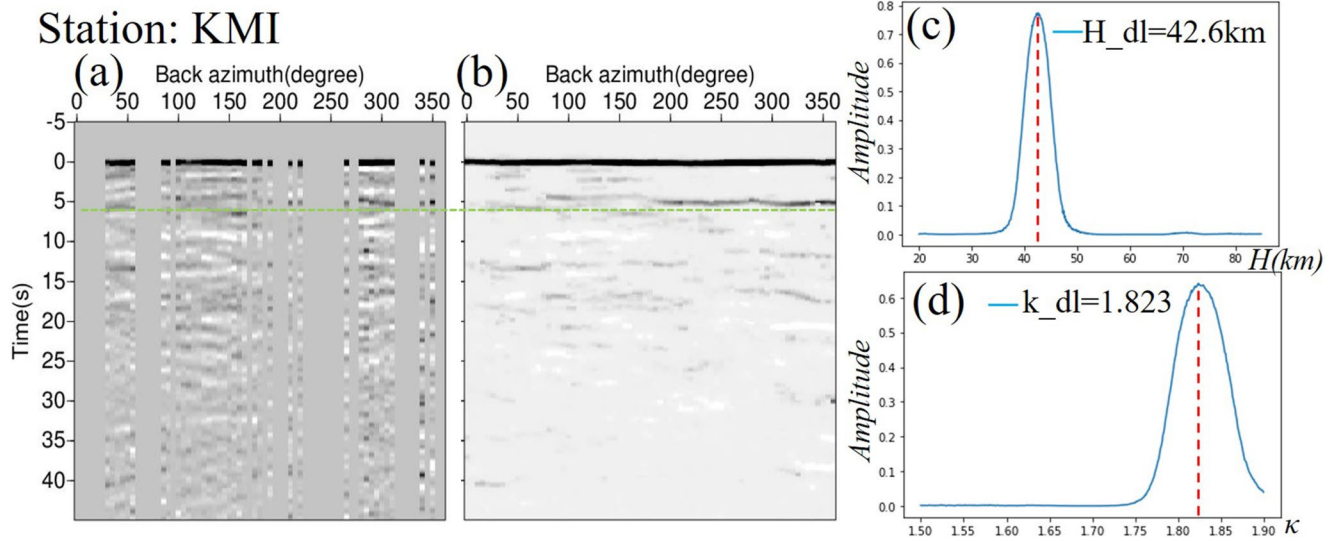


Figure 10. Same as Figure 9 but for the KMI station. The green dashed line marks the P_s phases. Because the original receiver functions (RFs) of the KMI station are too complicated, the phases cannot be clearly distinguished from the denoised RFs.

to the values obtained by J. Li et al. (2017) through joint inversion ($H = 40.00$, $\kappa = 1.790$). The HkNet result agrees better than the result from the H - κ method (45.50 km, 1.680) and is comparable to the result from the H - κ - c method (40.50 km, 1.740). It should be noted that the joint inversion mentioned above assumes isotropic media, so the small discrepancy between the results from the joint inversion and the new method is quite normal.

Figures 11 and 12 show the results of H and κ at 38 CNDN stations. The new results show general agreements with the H - κ - c results in both H and κ . The crustal thickness H follows a well-known trend of gradual thinning from west to east. The Tibetan Plateau, Tien Shan, Qilian Orogenic Belt, and Qinling–Dabie Orogenic Belt have thicker crust and higher κ , and the stable craton areas in eastern China have thinner crust but lower κ . However, there are also clear differences between the results of the two methods. Most of these 38 stations are concentrated in eastern China. Our H is generally higher in these places. Part of the reason for this systematic deviation of H may be that the average V_p of the crust is different. The average crustal V_p used in the H - κ - c method is fixed at 6.3 km/s, but the real V_p is higher in the east. This will be discussed in detail in Section 4. There are also large difference the κ values at some stations. Compared with the H - κ - c results, the new κ values are more coherent within the same tectonic blocks.

As an example, the biggest difference in H is the GOM station, where our result (73.9 km) is quite different from that of the H - κ - c method (62.0 km) by more than 10 km. Examining the original data carefully, we can see only a small part of a vaguely continuous P_s phase in the RFs of the GOM station, and the other phases are completely

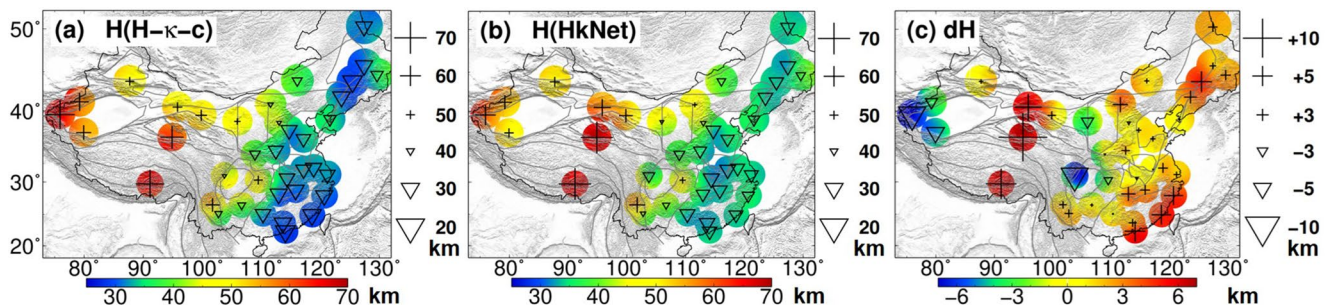


Figure 11. Summary of HkNet results of crustal thickness (H) for the 38 Chinese National Digital Seismic Network stations, where the results for each station are color coded within 2 degrees of the station: (a) H - κ - c method, (b) HkNet, and (c) difference (HkNet results minus the H - κ - c method's) in H (dH).

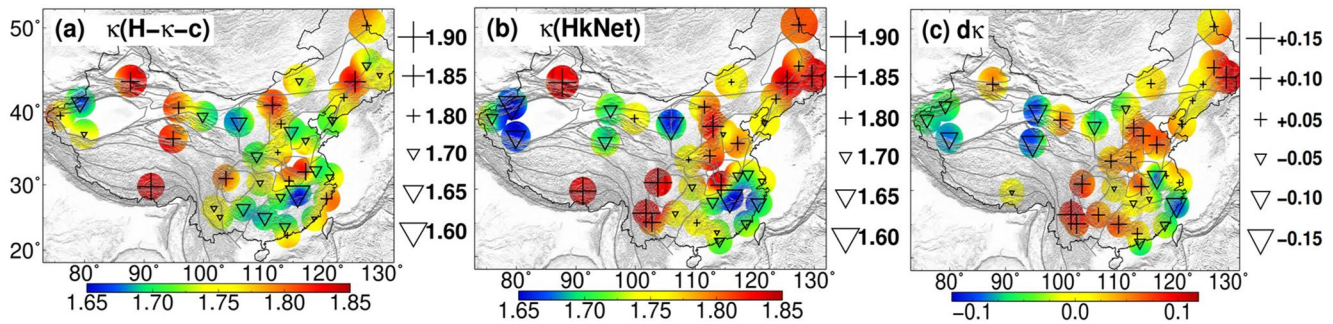


Figure 12. Summary of HkNet results of Crustal V_p/V_s (κ) for the 38 Chinese National Digital Seismic Network stations, where the results for each station are color coded within 2 degrees of the station: (a) H - κ - c method, (b) HkNet, and (c) difference (HkNet results minus the H - κ - c method's) in κ ($d\kappa$).

submerged in noise (Figure 13). After the DenoiseNet, the P_s phase is more clearly observable at approximately 9 s, which corresponds to H of approximately 70 km. Thus, the new result is likely more reliable.

The discrepancies above may be due to the complex geological environment in these places, which does not match the assumptions of the H - κ - c or our method. In such a complicated situation, for example, in cases of P anisotropy and internal interfaces in crust, the synthetics discussed in the following section demonstrate that our method is more reliable.

4. Discussion

The generalization of a deep learning method, that is, how widely the method can be used, is a major challenge that restricts its application. To improve generalization, the preparation of the training set is a key step. For our problem, if the training set is highly representative and the real situations can be completely covered by the parameters of our velocity models, the generalization ability of our method will be greatly enhanced. In our training set, the variation range of each parameter is listed in Table 1. Note if the complex crustal structure is too different from the model in our training set (e.g., sedimentary basin of low-velocity, gradational Moho), our proposed method may not be able to find the reliable results, which we have not tested out. However, examination of a few stations at the edges of major basins shows reasonable agreements between the results from the H - κ - c method and the HkNet (Figure S8 in Supporting Information S1). The following subsections discuss the factors that may affect the final result.

Station: GOM

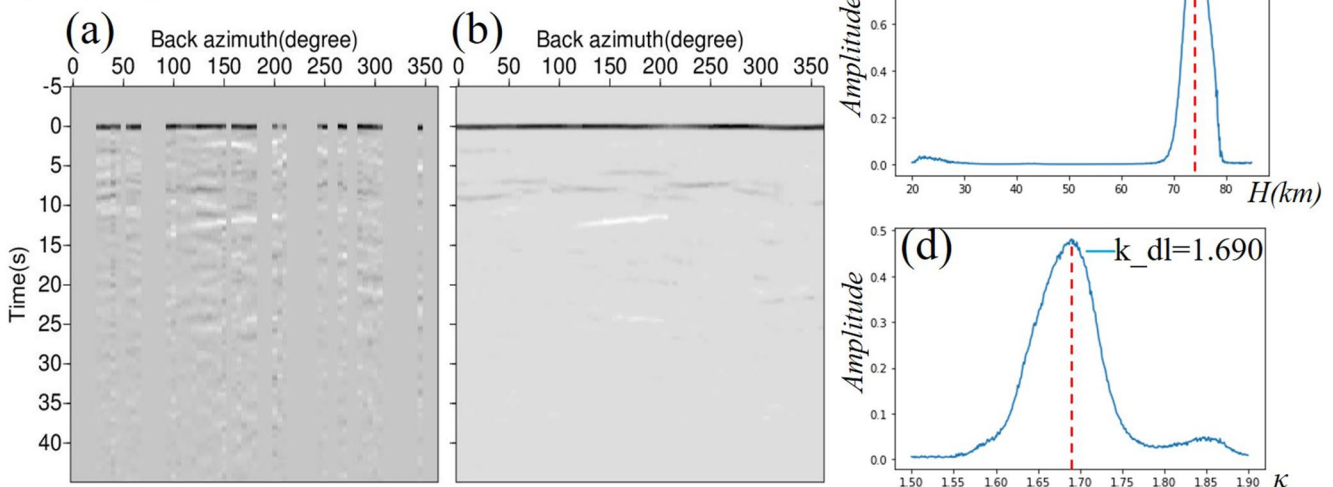


Figure 13. Same as Figure 9 but for the GOM station: (a) real receiver functions (RFs), (b) denoised RFs, and (c) H and (d) κ predicted by the HkNet.

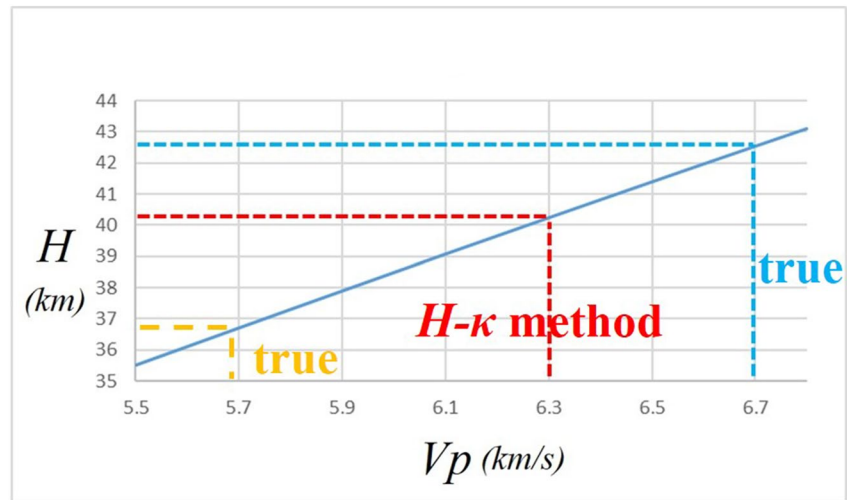


Figure 14. Trade-off between H and average V_p given the same P_s arrival time and κ . Any pair of H and V_p along the dark blue curve (including the pairs highlighted by the yellow, red, and blue dashed lines) can result in the same P_s phase arrival time of 5 s for the same κ (1.750). Thus, if the real V_p is higher than that used in the H - κ method, the obtained H from the H - κ method will be too low; otherwise, H will be too high.

4.1. Average V_p in Crust

As shown in Table 1, the crustal V_p varies randomly within 5.0–7.0 km/s, and κ varies randomly within 1.6–1.9. In the traditional H - κ and H - κ - c methods, V_p needs to be given in advance and can be adjusted, but since we don't know the real value of V_p , we usually assume the default average V_p is 6.3 km/s. When the real V_p does not match this, the search results will be affected. Let us take the theoretical P_s time as an example to analyze the error caused by the H - κ method for different V_p . Figure 14 shows the calculated relationship between V_p and H when the fixed theoretical P_s phase arrival time is 5 s, κ is 1.75, and the ray parameter is 0.06. It can be seen that when the H - κ method is used for the grid search and when the average V_p changes from 5.5 to 6.8 km/s, the error between the searched H and correct value reaches almost 8 km.

The simple calculations above show only the influence on the P_s phase. If multiple phases are considered, the results are more complicated but are in the same trend as shown in the synthetic tests (Tables 4 and 5). For fast V_p (SYN10 and SYN12, Table 4) (relative to the default value of 6.3 km/s in the H - κ - c method), the H result from the H - κ or the H - κ - c method is indeed too low from the true value (Table 5). Our conclusion is also consistent with previous studies (C. Li et al., 2014; Y. Li et al., 2014; Nair et al., 2006). The results from the new method are more accurate in both H and κ .

4.2. P Azimuthal Anisotropy

In the training set of our method, P anisotropy is not considered. The existence of P anisotropy will have a significant impact on the arrival times of different seismic phases. To examine the influence of P anisotropy on our method, we added different degrees of P anisotropy (5% P anisotropy to SYN13 and 10% P anisotropy to

Table 6
Errors of Synthetic Tests With P Azimuthal Anisotropy^a

SYN#	Model name	H - κ - c		HkNet		SSIM	PSNR (dB)
		dH (km)	$d\kappa$	dH (km)	$d\kappa$		
13	P5	1.00	−0.050	−0.60	−0.047	0.9970	45.70
14	P10	2.50	−0.110	0.30	−0.095	0.9966	45.00

^aModel name indicates percentage of P azimuthal anisotropy (e.g., 5% in P5). All results are shown by the differences (dH and $d\kappa$) from the input values of 40 km and 1.75, respectively. The minimum errors of the three methods are in bold font.

Table 7
Descriptions of the Multi-Layered Synthetic Models^a

SYN#	Model name	Model description							
		Thickness (km)	Vp (km/s)	κ	S%	Trend	Plunge	Direction	Dip
15	2_ANI	20	6.3	1.75	5	0°	0°	0°	0°
		20	6.5	1.75	8	0°	0°	0°	0°
16	2_ANI_PLG	20	6.3	1.75	5	0°	10°	0°	0°
		20	6.5	1.75	8	0°	20°	0°	0°
17	2_ANI_PLG_ DIP	20	6.3	1.75	5	0°	10°	90°	6°
		20	6.5	1.75	8	0°	20°	90°	10°
18	3_ANI_PLG_ DIP	20	6.3	1.75	5	0°	10°	90°	6°
		20	6.5	1.75	8	0°	20°	90°	8°
		20	6.8	1.75	8	0°	40°	90°	10°

^aAll models use real receiver function distribution from station AXX.

SYN14) to the basic REF model (SYN 0 in Table 2). The specific results are listed in Table 6. Once P anisotropy exists, the overall image characteristics and time series of the RFs will change, causing the knowledge learned by the network to be less applicable. It can be seen that as the P anisotropy increases, the errors of H and κ gradually increase. However, our errors are still smaller than those from the H - κ - c method, especially for H . Note that the presence of P azimuthal anisotropy reduces the κ value, because the teleseismic wavefield arrives at the seismic station at steep angle thus samples the relatively slow direction of the P wave (J. Li et al., 2019). If the distribution of the P anisotropy is known for a specific area, we believe we can make a targeted training set, which will improve the accuracy of the HkNet results.

4.3. Multi-Layered Velocity Model

We assume that there is no internal structure in the crust of our training set. We set up several two-layered velocity models of different complexities to test the prediction results of this method. The specific model parameters are listed in Table 7, and the results are listed in Table 8. It can be seen that our method is much more stable than the H - κ - c method for these models. The errors of H and κ were reduced by 63.3% and 90.0%, respectively. It can be seen that as the complexity of the velocity models increases, the errors of the HkNet and H - κ - c results gradually increase, but the HkNet is obviously more stable than the H - κ - c method. Especially for κ , its error is reduced by an order of magnitude compared with the H - κ - c method. Therefore, part of the reason for the relatively large difference between the results of κ in real data applications (above) may be that the crust stratification clearly exists in the real earth. If models with different layers are added in generating the training set, the accuracy of HkNet will be correspondingly improved.

Table 8
Errors of Multi-Layered Synthetic Test Results^a

SYN#	Model name	H - κ - c		HkNet		SSIM	PSNR (dB)
		dH (km)	$d\kappa$	dH (km)	$d\kappa$		
15	2_ANI	-0.9	0.020	0.70	0.000	0.993	42.5
16	2_ANI_PLG	-1.00	0.010	0.70	0.000	0.993	42.1
17	2_ANI_PLG_DIP	-2.40	0.040	0.70	-0.006	0.993	42.6
18	3_ANI_PLG_DIP	-5.50	0.050	-1.5	0.006	0.985	39.7
AVG (abs)	/	2.45	0.030	0.90	0.003	0.991	41.7

^aAll results are shown by the differences (dH and $d\kappa$) from the input values of 40 km and 1.75, respectively. The minimum errors of the three methods are in bold font.

Table 9
Descriptions of Synthetic Models With New Noise and Back-Azimuthal Distribution

SYN#	Model description							
	Thickness (km)	V _p (km/s)	κ	S%	Trend	Plunge	Direction	Dip
REF	40	6.3	1.75	5	30°	0°	90°	8°
19		Same as REF, but use real RF distribution from station WMQ.						
20		Same as REF, but use real RF distribution from station WUS.						
21		Same as REF, but use real RF distribution from station WZH.						
22		Same as REF, but use real RF distribution from station XAN.						
23		Same as REF, but use real RF distribution from station XLT.						
24		Same as REF, but use real RF distribution from station YCH.						

4.4. Noise and Back-Azimuthal Distribution

Following the method of Section 2.4 to generate the noise data set, we used only 32 of the 38 stations, leaving out 6 stations (Table 9). The purpose is to test the generalization capability of our trained networks by excluding the information from the data to be tested. These 6 stations were selected only because they are the last stations when the station codes are ordered alphabetically. The results on these 6 stations are still very accurate (Table 10). The average error of H is 0.97 km, that of κ is 0.005, SSIM is 0.989, and PSNR is 40.8, indicating that our method has excellent generalization with regards to the noise and back-azimuthal distribution. These results are similar to the stations whose noise are used in generating noise data set.

We also tested the reliability of the results from our proposed method when the RFs are missing to various degrees. We used synthetic models SYN1–7 except SYN3 in Table 2 for the tests. The reason for excluding model SYN3 is that its κ error is still very large even when all RFs are not missing (Table 3). We designed two missing-coverage cases under the same level of SNR. The first is uniformly missing (missing ratios were 25%, 33%, 50%, 67%, 75%, 80%), and the second is consecutively missing (missing back-azimuthal RF coverages of 40°, 60°, 100°, 140°, 180°, and 240°). The examples of RFs in two cases are shown in Figure S9 in Supporting Information S1. The average error of SYN models is shown in Figure S10 in Supporting Information S1. The details of the errors are shown in Table S1 and S2 in Supporting Information S1.

Our proposed method can achieve stable results for RFs with varying degrees of uniform missing information, even when the missing information reaches 80% (Figure S10a in Supporting Information S1). For the RFs with consecutively missing information, the errors increase with the gap and increase rapidly when the gap is greater

Table 10
Errors of the Synthetic Tests in Table 9^a

SYN#	dH (km)	$d\kappa$	SSIM		PSNR (dB)	
			Before	After	Before	After
REF	0.70	−0.006	/	/	/	/
19	1.40	−0.006	0.377	0.992	23.2	41.8
20	1.30	−0.003	0.276	0.988	22.5	40.3
21	0.70	−0.006	0.244	0.982	22.6	38.5
22	1.30	−0.003	0.361	0.990	23.2	41.2
23	0.70	−0.006	0.417	0.990	23.6	40.8
24	0.70	−0.006	0.494	0.992	23.6	42.1
AVG (abs)	0.97	0.005	0.361	0.989	23.1	40.8

^aAll results are shown by the differences (dH and $d\kappa$) from the input values of 40 km and 1.75, respectively. “Before” and “After” represent before and after DenoiseNet processing, respectively.

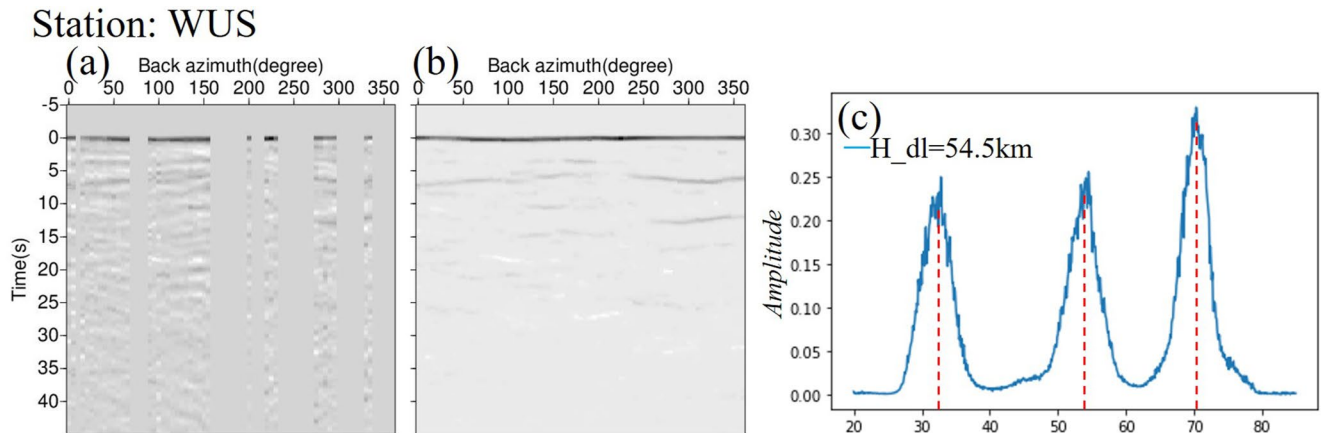


Figure 15. Crust thickness (H) predicted by HkNet for the WUS station: (a) real receiver functions (RFs), (b) denoised RFs, and (c) H predicted by HkNet. Obviously, H has three peaks (red dashed lines: 32.80 km, 54.50 km, 70.30 km). Based on prior knowledge, we are more inclined to select 54.50 km as the final predicted value.

than 180° (Figure S10b in Supporting Information S1). Thus, limiting the back-azimuthal gaps in RFs to less than 90° as suggested by Rumpker et al. (2014) is a good rule of thumb in obtaining a result.

4.5. Prior Knowledge Constraints

Due to the “black box” nature of machine learning, we cannot quantitatively explain why the network obtained certain results and cannot determine the errors of the results. The results of the HkNet are outputs in the form of Gaussian distributions only because our labels are Gaussian, so the width of the Gaussian distribution does not strictly represent the deviation, and its amplitude does not strictly represent the probability. However, combining the data-driven approach and constraints from prior knowledge will significantly improve the accuracy of predictions. For example, in the results of the WUS station, the crustal thickness predicted by the HkNet obviously has three peaks (Figure 15): 32.8 km, 54.5 km, 70.3 km. This station is located at the edge of the Qaidam Basin, and its crustal thickness should be approximately 50 km (Bao et al., 2015). In addition, the result from the H - κ - c method is 57 km, so we prefer choosing 54.5 km as the final result.

5. Conclusions

This paper proposes a method for automatically estimating H and κ from RFs based on deep learning. The method uses the powerful feature-mining capabilities of deep learning to automatically extract effective information from complex RFs to achieve its goals. Our method considers different complex geological settings, including different average crustal V_p , S anisotropy, and dipping interface(s). The method can not only identify the correct phase information, but also find the correlations of RFs at different back-azimuths and restore certain missing back-azimuthal information. An RF is essentially a time series consisting of different seismic phases, from which a CNN network can mine high-dimensional information in the time domain and analyze the relationship between seismic phases at different time points. As a by-product of this method, the DenoiseNet can be used for other data preprocessing and the denoised receiver functions can be used for further structural inversions.

Compared with the traditional H - κ or the recently developed H - κ - c methods, the proposed method has higher accuracy and more stable results, as demonstrated through extensive and systematic synthetic tests. Through the verification of real data, we obtained results that are consistent with those of previous studies and appear more coherent with geological settings. In addition, the method seems to have a strong generalization capability in that it can be widely used in various geological settings to obtain reliable results. Furthermore, the trained model can be used to obtain the estimated value within one second and does not need to adjust any parameters, so it is extremely efficient and flexible in processing a large amount of RF data. The H - κ - c method, on the other hand, may need to adjust parameters carefully.

Data Availability Statement

The waveform data for the Chinese National Digital Seismic Network receiver functions used in this study were from J. Li et al. (2019), <https://doi.org/10.1029/2018JB016356>. Some figures were plotted with GMT 4.5.18 software (<https://www.generic-mapping-tools.org/download/> (Wessel & Smith, 1998)). We make available the HkNet deep learning software developed in this study, including the trained deep learning model, the calling program, and all the real receiver-function data used in this study. In the application, the user only needs to preprocess the new data as described in Section 2.3 and feed it into the model. The new software can be obtained upon request from the authors or directly from Zenodo: <https://doi.org/10.5281/zenodo.6360874>.

Acknowledgments

We thank constructive and detailed comments from two anonymous reviewers, which helped improved the manuscript. This research was supported by the Special Fund of the Institute of Geophysics, China Earthquake Administration (Grant No. DQJB21B32) and National Natural Science Foundation of China (Grant No. U1939204).

References

- Bao, X., Song, X., & Li, J. (2015). High-resolution lithospheric structure beneath Mainland China from ambient noise and earthquake surface-wave tomography. *Earth and Planetary Science Letters*, 417, 132–141. <https://doi.org/10.1016/j.epsl.2015.02.024>
- Bergen, K. J., Johnson, P. A., Hoop, M. D., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, 363(6433), eaau0323. <https://doi.org/10.1126/science.aau0323>
- Bi, Z., Wu, X., Geng, Z., & Li, H. (2021). Deep relative geologic time: A deep learning method for simultaneously interpreting 3-D seismic horizons and faults. *Journal of Geophysical Research: Solid Earth*, 126(9), e2021JB021882. <https://doi.org/10.1029/2021JB021882>
- Cheng, X., Liu, Q., Li, P., & Liu, Y. (2019). Inverting Rayleigh surface wave velocities for crustal thickness in eastern Tibet and the western Yangtze craton based on deep learning neural networks. *Nonlinear Processes in Geophysics*, 26(2), 61–71. <https://doi.org/10.5194/npg-26-61-2019>
- Cui, Q., Zhou, Y., Li, J., Song, X., Gao, Y., & Cui, R. (2021). Crustal thickness (H) and V_p/V_s ratio (κ) images beneath the central Tien Shan revealed by the H - κ - c method. *Tectonophysics*, 822, 229157. <https://doi.org/10.1016/j.tecto.2021.229157>
- Frederiksen, A. W., & Bostock, M. G. (2000). Modelling teleseismic waves in dipping anisotropic structures. *Geophysical Journal International*, 141(2), 401–412. <https://doi.org/10.1046/j.1365-246x.2000.00090.x>
- Frederiksen, A. W., Folsom, H., & Zandt, G. (2003). Neighbourhood inversion of teleseismic P_s conversions for anisotropy and layer dip. *Geophysical Journal International*, 155(1), 200–212. <https://doi.org/10.1046/j.1365-246X.2003.02043.x>
- Hammond, J. O. S. (2014). Constraining melt geometries beneath the Afar Depression, Ethiopia from teleseismic receiver functions: The anisotropic H - κ stacking technique. *Geochemistry, Geophysics, Geosystems*, 15(4), 1316–1332. <https://doi.org/10.1002/2013GC005186>
- Han, S., Zhang, H., Xin, H., Shen, W., & Yao, H. (2021). USTClitho2.0: Updated unified seismic tomography models for continental China lithosphere from joint inversion of body-wave arrival times and surface-wave dispersion data. *Seismological Research Letters*, 93(1), 201–215. <https://doi.org/10.1785/0220210122>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity mappings in deep residual networks. In *European Conference on computer vision*. Springer International Publishing. <https://doi.org/10.48550/arXiv.1603.05027>
- Huang, R., Xu, Y., Zhu, L., & He, K. (2015). Detailed Moho geometry beneath southeastern China and its implications on thinning of continental crust. *Journal of Asian Earth Science*, 112, 42–48. <https://doi.org/10.1016/j.jseaes.2015.09.002>
- Huang, R., Zhu, L., & Xu, Y. (2014). Crustal structure of Hubei Province of China from teleseismic receiver functions: Evidence for lower crust delamination. *Tectonophysics*, 636, 286–292. <https://doi.org/10.1016/j.tecto.2014.09.001>
- Jiang, C., Fang, L., Fan, L., & Li, B. (2021). Comparison of the earthquake detection abilities of PhaseNet and EQTransformer with the Yangbi and Maduo earthquakes. *Earthquake Science*, 34(5), 425–435. <https://doi.org/10.29382/eqs-2021-0038>
- Juan Pablo, L., & Charles, J. A. (1999). Iterative deconvolution and receiver-function estimation. *Bulletin of the Seismological Society of America*, 89(5), 1395–1400. <https://doi.org/10.1785/BSSA0890051395>
- Kaviani, A., & Rümpker, G. (2015). *Generalization of the H- κ stacking method to anisotropic media*. EGU General Assembly Conference Abstracts.
- Kuang, W., Yuan, C., & Zhang, J. (2021). Real-time determination of earthquake focal mechanism via deep learning. *Nature Communications*, 12(1), 1432. <https://doi.org/10.1038/s41467-021-21670-x>
- Langston, C. A. (1979). Structure under Mount Rainier, Washington, inferred from teleseismic body waves. *Journal of Geophysical Research*, 84(B9), 4749–4762. <https://doi.org/10.1029/JB084iB09p04749>
- Laske, G., Masters, G., Ma, Z., & Pasyanos, M. E. (2012). *CRUST1.0: An updated global model of Earth's crust*. EGU General Assembly Conference Abstracts.
- Levin, V., & Park, J. (1997a). Crustal anisotropy in the Ural Mountains foredeep from teleseismic receiver functions. *Geophysical Research Letters*, 24(11), 1283–1286. <https://doi.org/10.1029/97GL51321>
- Levin, V., & Park, J. (1997b). P - SH conversions in a flat-layered medium with anisotropy of arbitrary orientation. *Geophysical Journal International*, 131(2), 253–266. <https://doi.org/10.1111/j.1365-246X.1997.tb01220.x>
- Li, C., Shen, X., & Qin, M. (2014). The effect of P wave velocity on the stacking results of receiver functions. *Acta Seismologica Sinica*, 36(3), 480–490. <https://doi.org/10.3969/j.issn.0253-3782.2014.03.013>
- Li, J., Song, X., Wang, P., & Zhu, L. (2019). A generalized H - κ method with harmonic corrections on P_s and its crustal multiples in receiver functions. *Journal of Geophysical Research: Solid Earth*, 124(16), 3782–3801. <https://doi.org/10.1029/2018JB016356>
- Li, J., Song, X., Zhu, L., & Deng, Y. (2017). Joint inversion of surface wave dispersions and receiver functions with P velocity constraints: Application to Southeastern Tibet. *Journal of Geophysical Research: Solid Earth*, 122, 7291–7310. <https://doi.org/10.1002/2017JB014135>
- Li, M., Song, X., Li, J., & Bao, X. (2022). Lithospheric structure of East Asia from ambient-noise and earthquake surface-wave tomography. *Earthquake Science*. <https://doi.org/10.1016/j.eq.s.2022.05.004>
- Li, W., Chen, Y., & Xu, Y. (2021). Crustal SiO_2 content of the Emeishan Large Igneous Province and its implications for magma volume and plumbing system. *Geochemistry, Geophysics, Geosystems*, 22(8), e2021GC009783. <https://doi.org/10.1029/2021GC009783>
- Li, Y., Gao, M., & Wu, Q. (2014). Crustal thickness map of the Chinese mainland from teleseismic receiver functions. *Tectonophysics*, 611, 51–60. <https://doi.org/10.1016/j.tecto.2013.11.019>
- Li, Z. (2021). Recent advances in earthquake monitoring II: Emergence of next-generation intelligent systems. *Earthquake Science*, 34(6), 531–540. <https://doi.org/10.29382/eqs-2021-0054>
- Liang, L., Deng, S., Gueguen, L., Wei, M., Wu, X., & Qin, J. (2019). *Convolutional neural network with median layers for denoising salt-and-pepper contaminations*. <https://doi.org/10.48550/arXiv.1908.06452>

- Liu, Z., & Park, J. (2017). Seismic receiver function interpretation: *Ps* splitting or anisotropic underplating. *Geophysical Journal International*, 208, 1332–1341. <https://doi.org/10.1093/gji/ggw455>
- Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11(1), 3952. <https://doi.org/10.1038/s41467-020-17591-w>
- Nair, S., Gao, S., Liu, K., & Silver, P. (2006). Southern african crustal evolution and composition: Constraints from receiver function studies. *Journal of Geophysical Research*, 111(B2), B02304. <https://doi.org/10.1029/2005jb003802>
- Niu, Y., Wang, Y. D., Mostaghimi, P., Swietojanski, P., & Armstrong, R. T. (2020). An innovative generative adversarial network application for physically accurate rock images with an unprecedented field of view. *Geophysical Research Letters*. <https://doi.org/10.1029/2020GL089029>
- Park, J., & Levin, V. (2016). Anisotropic shear zones revealed by back-azimuthal harmonics of teleseismic receiver functions. *Geophysical Journal International*, 207, 1216–1243. <https://doi.org/10.1093/gji/ggw323>
- Raschka, S. (2017). *Python machine learning : Machine learning and deep learning with Python, scikit-learn, and TensorFlow*.
- Reading, A. M., Cracknell, M. J., Bombardieri, D. J., & Chalke, T. (2015). Combining machine learning and geophysical inversion for applied geophysics. *ASEG Extended Abstracts*, 2015(1), 1–5. <https://doi.org/10.1071/aseg2015ab070>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-Assisted Intervention*. Springer International Publishing. <https://doi.org/10.48550/arXiv.1505.04597>
- Rümpker, G., Kaviani, A., & Latifi, K. (2014). *Ps*-splitting analysis for multilayered anisotropic media by azimuthal stacking and layer stripping. *Geophysical Journal International*, 199(1), 146–163. <https://doi.org/10.1093/gji/ggu154>
- Savage, M. K. (1998). Lower crustal anisotropy or dipping boundaries? Effects on receiver functions and a case study in New Zealand. *Journal of Geophysical Research*, 103, 15069–15087. <https://doi.org/10.1029/98JB00795>
- Shen, W., Ritzwoller, M., Kang, D., Kim, Y., Lin, F., Ning, J., et al. (2016). A seismic reference model for the crust and uppermost mantle beneath China from surface wave dispersion. *Geophysical Journal International*, 206(2), 954–979. <https://doi.org/10.1093/gji/ggw175>
- Tan, P., & Nie, S. (2021). Crustal deformation in eastern margin of Tibetan Plateau from a dense linear seismic array. *Physics of the Earth and Planetary Interiors*, 321, 106801. <https://doi.org/10.1016/j.pepi.2021.106801>
- Wang, J., Xiao, Z., Liu, C., Zhao, D., & Yao, Z. (2019). Deep-learning for picking seismic arrival times. *Journal of Geophysical Research: Solid Earth*, 124, 6612–6624. <https://doi.org/10.1029/2019JB017536>
- Wang, X., Li, Y., Ding, Z., Zhu, L., Wang, C., Bao, X., & Wu, Y. (2017). Three-dimensional lithospheric S wave velocity model of the ne Tibetan plateau and Western north China craton. *Journal of Geophysical Research: Solid Earth*, 122, 6703–6720. <https://doi.org/10.1002/2017JB014203>
- Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612. <https://doi.org/10.1016/j.jvcir.2019.102655>
- Welstead, S. (1999). Fractal and wavelet image compression techniques. In *Storage & retrieval for image & video databases*. Society of Photo-Optical Instrumentation Engineers. (SPIE).
- Wessel, P., & Smith, W. H. F. (1998). New, improved version of generic mapping tools released. *Eos, Transactions American Geophysical Union*, 79(47), 579. <https://doi.org/10.1029/98eo00426>
- Wong, W. C. J., Zi, J. P., Yang, H. F., & Su, J. R. (2021). Spatial-temporal evolution of injection-induced earthquakes in the Weiyuan Area determined by machine-learning phase picker and waveform cross-correlation. *Earth and Planetary Physics*, 5(6), 485–500. <https://doi.org/10.26464/epp2021055>
- Yu, S., & Ma, J. (2021). Deep learning for geophysics: Current and future trends. *Reviews of Geophysics*, 59(3), e2021RG000742. <https://doi.org/10.1029/2021RG000742>
- Zhang, B., Bao, X., & Xu, Y. (2020). Distinct orogenic processes in the south-and north-central Tien Shan from receiver functions. *Geophysical Research Letters*, 47(6), e2019GL086941. <https://doi.org/10.1029/2019GL086941>
- Zhang, H., Tian, X., & Teng, J. (2009). Estimation of crustal Vp/Vs with dipping Moho from receiver functions. *Chinese Journal of Geophysics*, 52(5), 1243–1252. (in Chinese). <https://doi.org/10.3969/j.issn.0001-5733.2009.05.013>
- Zhang, L., Fang, L., Wang, W., & Lv, Z. (2020). Seismic phase picking in China Seismic Array using a deep convolutional neuron network. *Earthquake Science*, 34(5), 425–435. <https://doi.org/10.29382/eqs-2020-0072-03>
- Zhang, X., Song, X., & Li, J. (2022). A comparative study of seismic tomography models of the Chinese continental lithosphere. *Earthquake Science*. <https://doi.org/10.1016/j.eqs.2022.05.005>
- Zhou, P. C., Ellsworth, W. L., Yang, H. F., Tan, Y. J., Beroza, G. C., Sheng, M. H., & Chu, R. S. (2021). Machine-learning-facilitated earthquake and anthropogenic source detections near the Weiyuan Shale Gas Blocks, Sichuan, China. *Earth and Planetary Physics*, 5(6), 501–519. <https://doi.org/10.26464/epp2021053>
- Zhou, Y., Yue, H., Fang, L., Zhou, S., Zhao, L., & Ghosh, A. (2021). An earthquake detection and location architecture for continuous seismograms: Phase picking, association, location, and matched filter (PALM). *Seismological Research Letters*, 93(1), 413–425. <https://doi.org/10.1785/0220210111>
- Zhu, L., & Kanamori, H. (2000). Moho depth variation in southern California from teleseismic receiver functions. *Journal of Geophysical Research*, 105(B2), 2969–2980. <https://doi.org/10.1029/1999JB900322>
- Zhu, L., Owens, T. J., & Randall, G. E. (1995). Lateral variation in crustal structure of the northern Tibetan Plateau inferred from teleseismic receiver functions. *Bulletin of the Seismological Society of America*, 85(6), 1531–1540. <https://doi.org/10.1785/BSSA0850061531>

Reference From the Supporting Information

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning* (Vol. 1). MIT press.