



Intelligent AVA Inversion Using a Convolution Neural Network Trained with Pseudo-Well Datasets

Jiaxing Sun¹ · Jidong Yang¹ · Zhenchun Li¹ · Jianping Huang¹ · Xin Luo² · Jie Xu¹

Received: 18 April 2022 / Accepted: 27 December 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract

The amplitude-variation-with-angle (AVA) inversion for seismic data has been widely used for hydrocarbon detection in exploration seismology. Traditional AVA inversion quantitatively estimates high-resolution elastic parameters, i.e., P-wave velocity, S-wave velocity and density, from migrated seismic gathers by solving either a linear or nonlinear inverse problem. It is commonly an ill-posed problem and the inversion accuracy depends on initial models. Recently, deep learning has been introduced into the AVA inversion by building a complicated nonlinear relation between seismic data and elastic parameters based on the training on a large amount of labeled data. The performance of the deep-learning-based inversion is determined by the diversity of training datasets. Because of sparse well locations, the application of deep-learning-based AVA inversion is limited by well-log label sets in production. To mitigate this problem, we present an intelligent AVA inversion method using a convolutional neural network trained by realistic pseudo-well logs. By considering spatial and inter-parameter correlation of elastic parameters, we first generate a large number of realistic pseudo-well logs based on Monte Carlo simulation. Then, angle-domain common-image gathers are computed by convolving a source wavelet with angle-dependent reflectivity series, which are used to train a convolutional neural network (CNN) to predict elastic parameters. In this study, we introduce two CNN frameworks to investigate the feasibility of the proposed pseudo-well-based CNN AVA inversion method using both synthetic and field data. We also compare the proposed CNN-based AVA inversion method with traditional linear and nonlinear inversion methods constrained by prior knowledge in terms of efficiency and accuracy. The results of synthetic data show that the pseudo-well-based CNN AVA inversion method can accurately and efficiently estimate P-wave velocity, S-wave velocity and density, and has a potential to reduce inter-parameter crosstalk artifacts. In the tests of field data, because of inaccurate background velocity models and noisy angle-domain gathers, the accuracy of CNN prediction results is not as high as in synthetic example. However, the pseudo-well-based CNN AVA inversion method still has better performance to reduce inter-parameter crosstalk artifacts and requires less computing time than traditional AVA inversion method.

✉ Jidong Yang
jidong.yang@upc.edu.cn

Extended author information available on the last page of the article

Keywords Pseudo-well logs · Convolutional neural network · AVA inversion · Computational seismology

Article Highlights

- Using Monte Carlo simulation and considering inter-parameter correlation of elastic parameters, we create a large number of realistic pseudo-well logs as the training data for the deep-learning AVA inversion
- With angle-domain common-image gathers as the input, we apply two convolutional neural network frameworks to predict P-wave velocity and S-wave velocity and density
- Numerical examples for synthetic and field data demonstrate that the proposed method produces accurate AVA inversion results with less inter-parameter crosstalk than traditional linear and nonlinear inversion methods

1 Introduction

With the rapid development of oil and gas industry, a large amount of accumulated seismic data provide a potential to predict hydrocarbon information using statistical deep learning methods. The stacked seismic images can reveal subsurface impedance interfaces, and prestack common-image gathers are related to angle-dependent reflection coefficients. With prestack seismic image gathers, amplitude-variation-with-angle (AVA) analysis can produce reservoir-related elastic parameters and pore fluid indicators, which plays an important role in reservoir characterization. Koefoed (1955) analyzed the angle-dependent plane-wave reflection coefficients for models with different Poisson's ratios and concluded that the amplitude gradients are sensitive to Poisson's ratio. Ostrander (1984) applied the AVA analysis for identifying the “bright spot” zone in the sandstone to detect oil and gas reservoir in the Gulf of Mexico.

Over the past decades, the AVA inversion methods have been made great progress. The Zoeppritz equation is the theoretical basis of AVA inversion. But the high nonlinearity of the Zoeppritz equation limits its direct application. Based on the weak reflection assumption, many approximate expressions for the Zoeppritz equation have been proposed, including classical Aki-Richard (1980), Shuey (1985) and Fatti (1994) approximations. These approximations make it possible to estimate subsurface elastic parameters by solving a linear inverse problem. However, the linear AVA inversion methods are difficult to accurately estimate seismic velocity and density models with large contrast and for data at large incident angles. Mallick (1995) proposed a genetic algorithm, and Kuzma and Rector (2004) presented a support vector machine method to solve the nonlinear AVA inversion problem based on the Zoeppritz equation. However, these intelligent optimization algorithms need large computational cost. Chen et al. (2009) and Wei and Chen (2011) reformulated the Zoeppritz equation and introduced a generalized method to solve the nonlinear equation. Stewart (1990) constructed an elastic equation for joint inversion and proposed an AVA inversion framework for both P-P and P-SV seismic data. The multiwave AVA analysis improves the inversion accuracy for large incident angles. Downton and Lines (2001) proposed a constrained AVA inversion method by introducing prior information, which can effectively reduce the non-uniqueness and improve stability. Based on Aki-Richard approximation, Biot and Gassmann equation, Zong et al. (2012)

directly estimated the P-wave modulus, S-wave modulus and fluid factor from prestack seismic data. Zhou et al. (2017) derived the Zoeppritz equivalent equation using Young's modulus and Poisson's ratio.

The convolutional neural network (CNN) has the advantages of weight sharing and regional locality perception, which enables it to describe sophisticated nonlinear relations for feature recognition and extraction. These advantages make it popular in many geoscience studies. For instance, Lewis and Vigh (2017) added the CNN predicted prior information to the objective function of full waveform inversion as a regularization term and reconstructed the salt body which is difficult to recover in conventional workflow. Zhao (2018) used the encoder–decoder CNN to classify seismic facies, and obtained better results than traditional classification method. Shi et al. (2018) applied the CNN to capture the patterns of target salt body and conducted salt deposit analysis on seismic images. Wu et al. (2019) used synthetic data to train an end-to-end CNN to detect faults for reservoir characterization and structural interpretation. Kaur et al. (2020) approximated the inverse Hessian in the least-squares reverse time migration with generative adversarial networks, and which improves the resolution of migrated image. Kaur et al. (2021) proposed a CNN-based interpolation method to compensate for missed common-shot data. In addition, CNN can also be applied to first break picking (Yuan et al. 2018; Hollander et al. 2018), wavefield separation (Bauer et al. 2021; Serfaty et al. 2017; Wang and Ma 2019) and noise suppression (Jin et al. 2018; Zhang et al. 2018b; Li et al. 2018; Siahkoohi et al. 2019).

In seismic inversion, the neural network can be interpreted as an inverse problem, while the network learning process can be considered as a solution to solve the inverse problem. The network is often thought of as a “black box”, which can generate a complicated nonlinear mapping function between seismic data and subsurface model parameters by updating the internal parameters of the network in training. Therefore, the CNN-based learning methods can be used to predict rock parameters and reservoir properties from seismic data. Biswas et al. (2019) applied the physics-guided CNN to predict P-wave velocity, S-wave velocity and density. Du et al. (2019) applied SeisInv-ResNet with deeper network structure to fit the nonlinear relationship between impedance parameters and prestack seismic records from the azimuthal anisotropic medium. Das et al. (2019) obtained P-impedance model from normal-incident seismic data using the CNN trained on a large amount of augmented training data. Alfarraj and AlRegib (2019) proposed a semisupervised sequence modeling framework based on a recurrent neural network, which can accurately predict elastic impedance from prestack angle-domain gathers. Das and Mukerji (2020) predicted porosity and shale content from prestack seismic gathers using an end-to-end neural network. Aleardi and Salusti (2021) combined discrete cosine transform with CNN to effectively approximate the complex nonlinear solver, and reconstructed subsurface elastic parameter models from prestack angle-domain gathers. Wang et al. (2021) proposed a novel cycle-consistent generative adversarial network architecture to extract impedance information from the seismic profile, and produced accurate estimation results. Kaur et al. (2022) introduced an effective CNN-based inversion method to predict dynamic reservoir properties and the CNN constrained by acquisition geometry and frequency content produces stable and accurate estimation results in both synthetic and field data.

As a data-driven methodology, one key factor limiting the application of deep learning is whether there are enough datasets to train the neural network. Few training data will produce biased fitting results. The large labeled data are commonly difficult to obtain in seismic interpolation as discussed by Bergen et al. (2019). When CNN is applied to seismic inversion, the labeled datasets are usually from drilling. However, due to the high cost of drilling and well-logging, the number of labeled data used for network training is often limited. The lack of labeled

data can severely influence the performance of deep learning methods. In order to improve the performance of CNN with limited labeled datasets, many strategies have been developed for data augmentation. Wang et al. (2019) introduced a stationary random matrix to increase the training data without using additional information. Das et al. (2019) used physics and geology to generate synthetic labeled datasets at the well location for training the CNN model. Wu et al. (2020) developed a workflow to build diverse models with realistic fold and fault features. Downton et al. (2020) simulated many pseudo-well logs based on the well statistics in the exploration area, and explored a hybrid theory-guided method to predict reservoir properties.

In order to mitigate the issue of limited labeled datasets, we propose a novel pseudo-well-based CNN AVA inversion method. First, we use a geostatistical operator to calculate spatial and inter-parameter correlation of the well data within target reservoirs. We randomly perturb the P-wave velocity, S-wave velocity and density to simulate the high-frequency pseudo-well logs with the calculated spatial and inter-parameter correlation values. Then, the high-frequency variations are added to the low-frequency background models to generate a set of geologically realistic pseudo-well logs. Synthetic prestack angle-domain gathers are calculated by convolving angle-dependent reflectivities computed from pseudo-well logs with source wavelet. We calculate the scale between real and synthetic seismic angle gathers and use a scaled source wavelet to generate realistic prestack angle gathers. Next, the pseudo-well logs and corresponding scaled synthetic seismic data are used to train neural network. At last, the trained network is applied to migrated prestack seismic angle-domain gathers to predict P-wave velocity, S-wave velocity and density simultaneously.

The rest of the paper is organized as follows. We first give an overview of traditional linear and nonlinear inversion for AVA analysis. Then, we present the detailed workflow of the proposed pseudo-well based CNN AVA inversion method. Next, we adopt two CNN models to test the feasibility and practicability of the proposed seismic inversion method for both synthetic and field data. Finally, we compare CNN-based and traditional AVA seismic inversion methods in terms of accuracy and efficiency.

2 Overview of Traditional AVA Analysis

2.1 Linear and Nonlinear AVA Inversion

In seismic AVA inversion, the expected subsurface model parameter is \mathbf{m} , and the input is migrated prestack angle-domain gathers \mathbf{d} . In general, the relationship between \mathbf{m} and \mathbf{d} can be written as

$$\mathbf{d} = \mathbf{G}\mathbf{m} + \mathbf{n}, \quad (1)$$

where \mathbf{G} is a forward operator that links the model \mathbf{m} to data \mathbf{d} , and \mathbf{n} is the noise associated with observed seismic records.

The linearized AVA inversion methods are based on the convolutional theory and approximated Zoeppritz equation. Based on the classical Aki–Richard linear Eq. (Aki and Richards 1980)

$$r(\theta) = A \frac{\Delta v_p}{v_p} + B \frac{\Delta v_s}{v_s} + C \frac{\Delta \rho}{\rho} \quad (2)$$

and if there are m incident angles (θ) and n , reflection interfaces, Eq. (1) can be extended as

$$\begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \\ \vdots \\ \mathbf{d}_m \end{bmatrix} = \begin{bmatrix} \mathbf{WA}_1 & \mathbf{WB}_1 & \mathbf{WC}_1 \\ \mathbf{WA}_2 & \mathbf{WB}_2 & \mathbf{WC}_2 \\ \vdots & \vdots & \vdots \\ \mathbf{WA}_m & \mathbf{WB}_m & \mathbf{WC}_m \end{bmatrix} \begin{bmatrix} \mathbf{R}_p \\ \mathbf{R}_s \\ \mathbf{R}_\rho \end{bmatrix} + \mathbf{n}, \quad (3)$$

where $A = \frac{\sec^2 \theta}{2}$, $B = -4 \frac{v_s^2}{v_p^2} \sin^2 \theta$ and $C = 1 - 4 \frac{v_s^2}{v_p^2} \sin^2 \theta$. \bar{v}_p , \bar{v}_s and $\bar{\rho}$ are averages over the reflecting interface. Δv_p , Δv_s and $\Delta \rho$ are corresponding differences. \mathbf{W} is a wavelet matrix, \mathbf{A}_i , \mathbf{B}_i , \mathbf{C}_i are the diagonal coefficient matrices corresponding to the i -th incident angle with the size of $n_t \times n_r$. $\mathbf{R}_p = [r_{p_1} \dots r_{p_{n_r-1}} r_{p_{n_r}}]$, $\mathbf{R}_s = [r_{s_1} \dots r_{s_{n_r-1}} r_{s_{n_r}}]$ and $\mathbf{R}_\rho = [r_{\rho_1} \dots r_{\rho_{n_r-1}} r_{\rho_{n_r}}]$ are the relative perturbation vectors with components as $r_{p_j} = 2 \frac{v_{p(j+1)} - v_{pj}}{v_{p(j+1)} + v_{pj}}$, $r_{s_j} = 2 \frac{v_{s(j+1)} - v_{sj}}{v_{s(j+1)} + v_{sj}}$, and $r_{\rho_j} = 2 \frac{\rho_{(j+1)} - \rho_j}{\rho_{(j+1)} + \rho_j}$.

However, there are accumulated errors when calculating P-wave velocity, S-wave velocity and density using the inverted relative perturbations. This can be mitigated by choosing a logarithmic parameterization (Peterson and Coker 1995), and then the classical Aki–Richard equation can be rewritten as (Buland and Omre 2003)

$$\begin{aligned} r(\theta) = & A(\ln(v_{p(j+1)}) - \ln(v_{pj})) + B(\ln(v_{s(j+1)}) - \ln(v_{sj})) \\ & + C(\ln(\rho_{(j+1)}) - \ln(\rho_j)) \end{aligned} \quad (4)$$

and Eq. (3) becomes

$$\begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \\ \vdots \\ \mathbf{d}_m \end{bmatrix} = \begin{bmatrix} \mathbf{WA}_1 & \mathbf{WB}_1 & \mathbf{WC}_1 \\ \mathbf{WA}_2 & \mathbf{WB}_2 & \mathbf{WC}_2 \\ \vdots & \vdots & \vdots \\ \mathbf{WA}_m & \mathbf{WB}_m & \mathbf{WC}_m \end{bmatrix} \mathbf{L} \begin{bmatrix} \mathbf{In}_p \\ \mathbf{In}_s \\ \mathbf{In}_\rho \end{bmatrix} + \mathbf{n} \quad (5)$$

or in a matrix form as $\mathbf{d} = \mathbf{GLIn} + \mathbf{n}$, where \mathbf{In}_p , \mathbf{In}_s and \mathbf{In}_ρ denote the logarithms of elastic parameter, and the matrix \mathbf{L} has the following form

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}_{3n_r \times 3(n_r+1)}, \quad \mathbf{I} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}_{n_r \times (n_r+1)}. \quad (6)$$

Under the Bayesian framework, the objective function (the details are given in Appendix A) of seismic inversion based on Cauchy distribution can be written as

$$J(\mathbf{In}) = \frac{1}{2} (\mathbf{d} - \mathbf{GLIn})^T \mathbf{C}_D^{-1} (\mathbf{d} - \mathbf{GLIn}) + 2 \sum_{i=1}^N \ln(1 + (\mathbf{In} - \boldsymbol{\mu})^T \boldsymbol{\Phi}^i (\mathbf{In} - \boldsymbol{\mu})), \quad (7)$$

where \mathbf{C}_D^{-1} is covariance matrix of noise, $\boldsymbol{\mu}$ is the mean of the logarithmic model parameter and $\boldsymbol{\Phi}^i$ is the statistical information obtained from well data.

The solution of Eq. (7) can be calculated by setting its derivative with respect to the logarithmic model parameter as zero. This yields

$$\mathbf{m} = \mathbf{e}^{[(\mathbf{GL})^T \mathbf{GL} + \sigma_{n_r}^2 \mathbf{Q}]^{-1} [(\mathbf{GL})^T \mathbf{d} + \sigma_{n_r}^2 \mathbf{Q} \boldsymbol{\mu}]}, \quad (8)$$

where \mathbf{Q} has the following expression

$$\mathbf{Q} = \sum_{i=1}^N \frac{4\boldsymbol{\Phi}^i}{1 + (\mathbf{In} - \boldsymbol{\mu})^T \boldsymbol{\Phi}^i (\mathbf{In} - \boldsymbol{\mu})}, \quad (9)$$

and $\sigma_{n_i}^2$ is the variance of noise. Equation (8) is a nonlinear problem, and it can be solved by iterative reweighted least squares algorithm (Alemie and Sacchi 2011).

The exact Zoeppritz equation can improve the accuracy of seismic prestack inversion compared with that using approximate equations. Assuming that the noises obey a Gaussian distribution, and the model parameters obey a Cauchy distribution, the objective function of nonlinear AVA inversion can be written as

$$J(\mathbf{m}) = \frac{1}{2}(\mathbf{d} - \mathbf{G}(\mathbf{m}))^T(\mathbf{d} - \mathbf{G}(\mathbf{m})) + 2\sigma_{n_i}^2 \sum_{i=1}^N \ln(1 + (\mathbf{m} - \boldsymbol{\mu})^T \boldsymbol{\Phi}^i (\mathbf{m} - \boldsymbol{\mu})). \quad (10)$$

Nonlinear inversion problems can be solved by the generalized Gauss–Newton optimization algorithm (Nocedal and Wright 1999) (the detailed derivation is given in Appendix B). The model perturbation $\Delta\mathbf{m}$ can be written as

$$\Delta\mathbf{m} = -\mathbf{H}(\mathbf{m})^{-1} \boldsymbol{\gamma}(\mathbf{m}), \quad (11)$$

and the update of model parameters has the form

$$\mathbf{m}^{k+1} = \mathbf{m}^k - \eta^k \Delta\mathbf{m}^k, \quad (12)$$

where $\boldsymbol{\gamma}(\mathbf{m})$ and $\mathbf{H}(\mathbf{m})$ are the gradient and Hessian, respectively, and η^k is the step length of k-th iteration.

3 The Pseudo-Well-Based CNN AVA Inversion

3.1 Construction of Realistic Pseudo-Well Log

Deep learning algorithms build a complex mapping relationship through the training process. In supervised learning task, the training needs sufficiently large and diverse data to update the network weights. In CNN-based AVA inversion, how to obtain a large number of labeled data and corresponding prestack angle-domain gathers is the key to obtain accurate solution. The Monte Carlo statistical simulation technology can produce many geologically realistic well data. A simple Monte Carlo simulation can be implemented using the following steps. First, it is assumed that the parameters obey Gaussian or uniform distribution. Then, the mean and variance of parameters are estimated from available well data. Finally, a set of pseudo-well logs are simulated by random sampling from the distribution function. However, the random sampling in the simulated pseudo-well logs might have large contrasts between neighboring samples, which are geologically meaningless. In other words, the geological depositional law is not taken into account. In addition, elastic parameters are spatially correlated because of layered sediments. Therefore, the pseudo-well logs simulated by the simple Monte Carlo method cannot be used as labeled data of CNN for AVA inversion. In order to simulate geologically realistic pseudo-well logs, we first apply a detailed statistical analysis of available well data around target reservoir, then add the spatial correlation to the pseudo-well logs simulated using the Monte Carlo method, and at last introduce a cross-correlation matrix to describe inter-parameter correlation (Downton

et al. 2020; Dvorkin et al. 2014). In this way, a set of geologically realistic pseudo-well logs can be built to mitigate the problem of limited labeled data. The detailed methodology for simulating geologically realistic labeled datasets can be summarized as the following steps:

- 1 Obtaining input data including available well data \mathbf{x} , low-frequency models \mathbf{m}_0 and seismic wavelet \mathbf{w} .
- 2 Performing statistical analysis on well data to obtain the variance c and maximum length d_{\max} of spatial correlation. Using analytical Gaussian variogram function γ instead of the correlation function to represent spatial correlation. The analytical Gaussian variogram is a function of length h between two sampling points. The analytical Gaussian variogram can be written as

$$\gamma(h) = c \left(1 - e^{-\frac{h^2}{d_{\max}^2}} \right). \quad (13)$$

- 3 Constructing spatial symmetrical covariance matrix using variogram function

$$\mathbf{C} = \begin{bmatrix} \gamma(d_{\max}) & \dots & \gamma(0) \\ \vdots & \ddots & \vdots \\ \gamma(0) & \dots & \gamma(d_{\max}) \end{bmatrix}. \quad (14)$$

- 4 When simultaneously simulating multiple pseudo-well logs, in addition to considering the spatial vertical correlation in a single log, we also need to consider the inter-parameter correlation among multiple pseudo-well logs.
 - (a) Introducing covariance matrix \mathbf{S} to characterize the inter-parameter correlation and the three-parameter covariance matrix for P-wave velocity, S-wave velocity and density can be written as

$$\mathbf{S} = \begin{bmatrix} \sigma_p^2 & \sigma_{p,s}^2 & \sigma_{p,\rho}^2 \\ \sigma_{s,p}^2 & \sigma_s^2 & \sigma_{s,\rho}^2 \\ \sigma_{\rho,p}^2 & \sigma_{\rho,s}^2 & \sigma_\rho^2 \end{bmatrix}, \quad (15)$$

where $\sigma_p^2 = \sum_{i=1}^n (p_i - \mu_p)^2$, $\sigma_s^2 = \sum_{i=1}^n (s_i - \mu_s)^2$, $\sigma_\rho^2 = \sum_{i=1}^n (\rho_i - \mu_\rho)^2$ are the residual variances for P-wave velocity, S-wave velocity and density, respectively. $\sigma_{p,s} = \sum_{i=1}^n (p_i - \mu_p)(s_i - \mu_s)$, $\sigma_{p,\rho} = \sum_{i=1}^n (p_i - \mu_p)(\rho_i - \mu_\rho)$, $\sigma_{s,\rho} = \sum_{i=1}^n (\rho_i - \mu_\rho)(s_i - \mu_s)$ are the residual covariances between two different variables. n is the number of sampling points. p , s and ρ are the residual P-wave velocity, S-wave velocity and density, respectively, and μ is the mean value.

- (b) Calculating the Kronecker product: $\mathbf{K} = \mathbf{S} \otimes \mathbf{C}$.
 - (c) Applying the Cholesky decomposition: $\mathbf{K} = \mathbf{R}\mathbf{R}^T$,
- where the sizes of \mathbf{K} and \mathbf{R} are $3n \times 3n$, \mathbf{R} is a lower triangular matrix and characterizes spatial correlation and cross-correlation, and \mathbf{R}^T is its transpose.

- 5 Multiplying the random signal \mathbf{u} with \mathbf{R} to compute residual vector as

$$\mathbf{R}_t = \mathbf{R}\mathbf{u}. \quad (16)$$

- 6 Adding the residual pseudo-well logs \mathbf{R}_t to low-frequency models \mathbf{m}_0 to generate multiple geologically realistic pseudo-well logs as

$$\begin{bmatrix} \mathbf{m}_p \\ \mathbf{m}_s \\ \mathbf{m}_\rho \end{bmatrix} = \begin{bmatrix} \mathbf{m}_{p0} \\ \mathbf{m}_{s0} \\ \mathbf{m}_{\rho_0} \end{bmatrix} + \mathbf{R}_t. \quad (17)$$

3.2 Computing Synthetic Prestack Angle-Domain Gathers for CNN Training

The proposed pseudo-well-based CNN AVA inversion is a supervised learning task, in which the input that corresponds to the labeled data is the prestack angle-domain gathers. To accurately describe the AVA response, we use the exact Zoeppritz equation to calculate angle-dependent reflectivities, which has the following form:

$$\begin{bmatrix} \sin \alpha & \cos \beta & -\sin \alpha' & \cos \beta' \\ \cos \alpha & -\sin \beta & \cos \alpha' & \sin \beta' \\ \cos 2\beta & -\frac{V_{S_1}}{V_{P_1}} \sin 2\beta & -\frac{\rho_2 V_{P_2}}{\rho_1 V_{P_1}} \cos 2\beta' & -\frac{\rho_2 V_{S_2}}{\rho_1 V_{P_1}} \sin 2\beta' \\ \sin 2\alpha & \frac{V_{P_1}}{V_{S_1}} \cos 2\beta & \frac{\rho_2}{\rho_1} \frac{V_{S_2}^2}{V_{S_1}^2} \frac{V_{P_1}}{V_{P_2}} \sin 2\alpha' & -\frac{\rho_2}{\rho_1} \frac{V_{S_2} V_{P_1}}{V_{S_1}^2} \cos 2\beta' \end{bmatrix} \begin{bmatrix} R_{PP} \\ R_{PS} \\ T_{PP} \\ T_{PS} \end{bmatrix} = \begin{bmatrix} -\sin \alpha \\ \cos \alpha \\ -\cos 2\beta \\ \sin 2\alpha \end{bmatrix}, \quad (18)$$

where V_P , V_S and ρ are P-wave velocity, S-wave velocity and density, subscript 1 and 2 denote the upper and lower layers across a interface. α and β are the incidence angles of P-wave and S-wave, respectively. α' and β' are the transmission angles of the P-wave and S-wave. R_{PP} and T_{PP} are reflection and transmission coefficients of P-wave, R_{PS} and T_{PS} are for PS-wave.

The convolutional theory is usually used to compute band-limited angle-domain gathers with the full-band reflection coefficients R_{PP} . The seismic forward modeling can be written as

$$\mathbf{d}(\theta) = \mathbf{W}(t) * \mathbf{R}(\mathbf{m}, \theta) + \mathbf{n}(\theta), \quad (19)$$

where \mathbf{m} is subsurface model, $\mathbf{d}(\theta)$ is synthetic prestack angle-domain gathers at the incidence angle θ , $\mathbf{W}(t)$ is source wavelet matrix, $\mathbf{R}(\mathbf{m}, \theta)$ is the angle-dependent reflectivity calculated using the simulated pseudo-well logs and exact Zoeppritz Eq. (18), $\mathbf{n}(\theta)$ denotes the random noise.

3.3 Deep-Learning-Based AVA Inversion

With prestack angle gather \mathbf{d} , the solution \mathbf{m} of AVA inversion can be expressed as

$$\mathbf{m} = \arg \min_{\mathbf{m}} = \frac{1}{2} \sum_{i=1}^p \|\mathbf{G}(\mathbf{m}) - \mathbf{d}\|^2, \quad (20)$$

where p denotes the number of angle gathers, \mathbf{G} is the forward operator. In traditional deterministic inversion, it starts from initial model and uses gradient-based method to estimate elastic parameters by minimizing the difference between synthetic and observed data. Because of limited angle information and noise, it is an ill-posed problem and the inversion accuracy highly depends on the initial model.

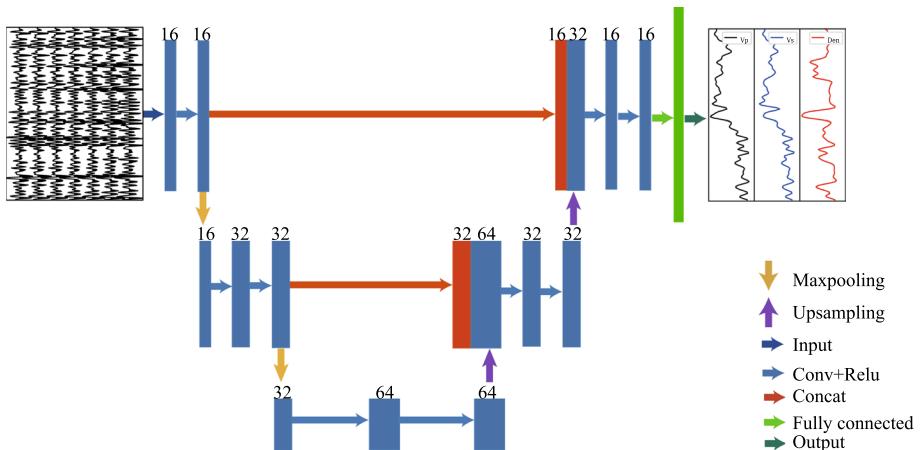


Fig. 1 A simplified Unet architecture for elastic parameters prediction. The number of feature channels is denoted on top of the box. Different operators are shown in different colors

The deep learning algorithms have a remarkable learning ability to describe the nonlinear relation between network input and output by training on a large dataset, which can partially reduce the influence of inaccurate initial model. In addition, the deep learning method can predict model parameters \mathbf{m} from prestack seismic angle gathers \mathbf{d} very quickly once the network is trained. The mapping relationship of deep learning framework for AVA analysis can be expressed as

$$\Omega(\mathbf{d}; \Theta) \rightarrow \mathbf{m}, \quad (21)$$

where Ω is an approximate inverse of the forward operator \mathbf{G} and defined by trainable weights Θ . The mapping function from prestack angle gathers \mathbf{d} to elastic parameters \mathbf{m} can be determined by iteratively refining and updating the network parameters in the training.

Note that the deep-learning-based inversion approach is different from conventional method in terms of objective function. In deep-learning-based method, the objective function is based on the misfit between network prediction results and labeled data, while it is computed in conventional inversion methods as the misfit between simulated and migrated prestack gathers. The objective function of deep-learning-based inversion method can be written as

$$L_{\text{learn}} = \frac{1}{n} \sum_{i=0}^n \left\| \mathbf{m}_i - \Omega(\mathbf{d}_i; \Theta) \right\|^2, \quad (22)$$

where n is the number of paired samples in the training datasets.

As a data-driven method, the CNN with sophisticated structures has a stronger mapping ability compared to an artificial neural network. In this study, we consider two CNN architectures to predict 1D P-wave velocity, S-wave velocity and density logs simultaneously from prestack angle gathers at a single Common Depth Point (CDP), and 2D profile is calculated by assembling single-trace prediction results. The first network (Fig. 1) is a simplified Unet proposed by Ronneberger et al. (2015). This U-shape network contains a contracting path (left side) and a symmetric expanding path (right side). There exists a total of ten 2D convolutional layers, and each convolutional layer is followed by a rectified linear unit

(ReLU) activation function. The large number of extracted feature channels at the front of the network can enrich the basic feature information recognized from the input data to facilitate the deepening of the network. The contracting path consists of the pooling operation for downsampling, and the expansive path includes the up-convolution for upsampling. In addition, the network has a concatenate operation that links the feature maps of the expansive and contracting paths. The second network architecture is shown in Fig. 2. This network consists of four blocks and one FC layer. Each block contains the convolution, Batch Normalization, ReLU and Dropout operators. The Dropout operator enhances the ability of the network to generalize and prevent overfitting by stopping the work of a neuron in the network with a lower limit probability p (Baldi and Sadowski 2013). Batch Normalization standardizes the input data and speeds up the network training.

In the proposed CNN-based AVA inversion method, the input of both networks are prestack angle-domain gathers, and the corresponding labeled data are P-wave velocity, S-wave velocity and density logs. The convolution operator in the first network has a fixed size of 3×3 . In the second network, the size of the kernel varies with the depth of the convolutional layer. During training, the convolution kernels slide on the input data to capture regional features. Both the networks have full connected (FC) structures in the last part, which helps to reshape the extracted feature maps in the convolutional layer to match the dimension of output. The number of weights and extracted feature maps of both networks are shown in Table 1 and Table 2, respectively. After training on the number of training datasets, the network learned to transform prestack angle gathers from the data domain to elastic parameters in the model domain.

In the training, if training data are totally inputted into the network to calculate an average gradient of the objective function, it might be difficult to find a global learning rate and needs massive computational cost. An improved strategy is first to randomly divide the entire training datasets into several bundles and then calculate the average gradient of each bundle to update network weights

$$\hat{\Theta} = \arg \min_{\Theta} \frac{1}{k} \sum_{i=1}^k \left\| \mathbf{m}_i - \mathcal{Q}(\mathbf{d}_i; \Theta) \right\|^2, \quad (23)$$

where k is the number of paired samples in each bundle.

When the network completely passes through the entire training datasets, the network weights have been updated several times. This strategy not only ensures an

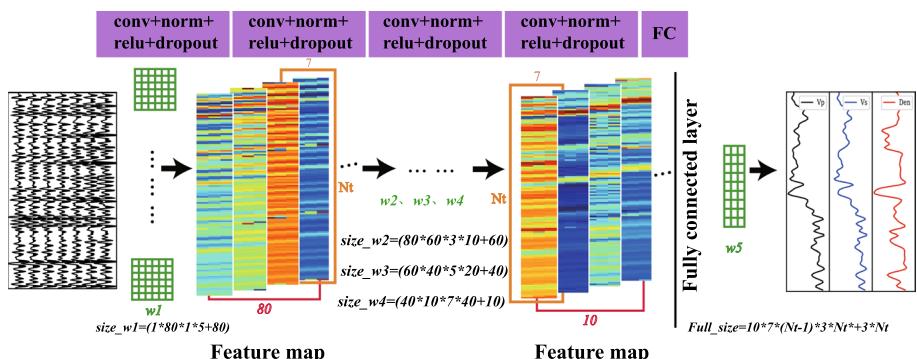


Fig. 2 Second convolutional neural network used to predict elastic parameters in this study. It consists of five blocks, the first four blocks have the same structure, and the last layer is a fully connected layer

Table 1 Detail information of the first CNN architecture

Network architecture	Kernel size	Number of filters	Trainable parameters
conv1	(3,3)	16	$1 * 16 * 3 * 3 + 16 = 160$
conv2	(3,3)	16	$16 * 16 * 3 * 3 + 16 = 2320$
conv3	(3,3)	32	$16 * 32 * 3 * 3 + 32 = 4640$
conv4	(3,3)	32	$32 * 32 * 3 * 3 + 32 = 9248$
conv5	(3,3)	64	$32 * 64 * 3 * 3 + 64 = 18496$
conv6	(3,3)	64	$64 * 64 * 3 * 3 + 64 = 36928$
conv7	(3,3)	32	$96 * 32 * 3 * 3 + 32 = 27680$
conv8	(3,3)	32	$32 * 32 * 3 * 3 + 32 = 9248$
conv9	(3,3)	16	$48 * 16 * 3 * 3 + 16 = 6928$
conv10	(3,3)	16	$16 * 16 * 3 * 3 + 16 = 2320$
conv11	(3,3)	1	$16 * 1 * 3 * 3 + 1 = 145$
FC			$1 * 7 * 296 * 3 * 297 + 3 * 297 = 1847043$

Table 2 Detail information of the second CNN architecture

Network architecture	Kernel size	Number of filters	Trainable parameters
conv1	(1,5)	80	480
conv2	(3,10)	60	144,060
conv3	(5,20)	40	240,040
conv4	(7,40)	10	112,010
FC			18,462,411

accurate gradient descent direction and can produce stable prediction results (Bottou 2012), but also improves the efficiency of network training.

4 Examples

In this section, we use synthetic and field data examples to verify the feasibility of the proposed pseudo-well-based CNN AVA inversion method. The networks are trained using simulated pseudo-well logs as the output and prestack seismic gathers as the input. We perform a comprehensive analysis to quantitatively evaluate the results of CNN-based and traditional theory-based AVA inversion methods. The background models in the synthetic example are computed by applying a Gaussian filter to the true models. The low-frequency models in the field example are obtained by interpolating the well data. The synthetic prestack seismic gathers are calculated by convolving seismic wavelet with reflectivity series. The source time function in the synthetic example is a Ricker wavelet, and in the field example it is extracted from prestack seismic gathers.

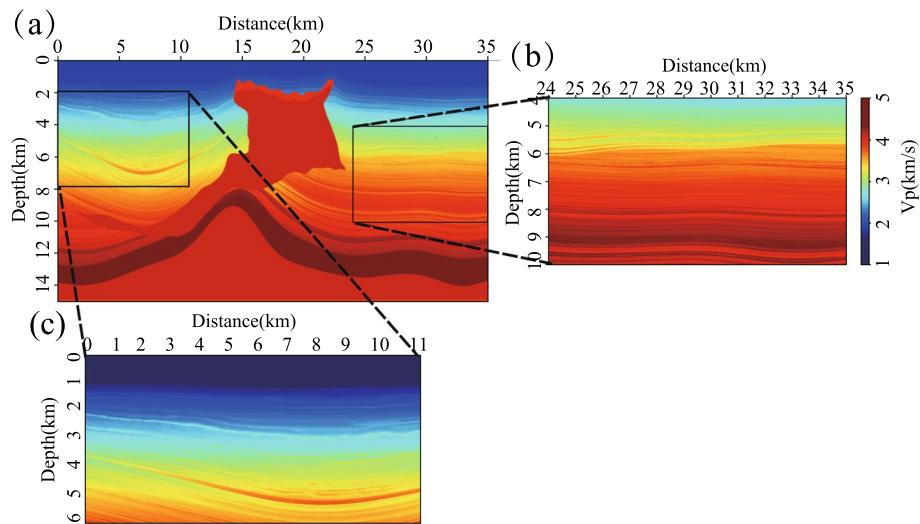


Fig. 3 SEAM model (a) and the portions we used to test the performance (b) and generalization ability (c) of the proposed CNN-based AVA inversion method

4.1 SEG Advanced Modeling (SEAM) Model

We select a portion of the SEAM model (Fig. 3b) to test the performance of intelligent AVA inversion trained with pseudo-well logs. The horizontal and vertical size of this SEAM model are 11 km and 6 km, respectively. The model has 300-time-sampling points and the time interval is 8 ms. The well logs extracted from 2 km, 6 km and 10 km locations capture the variability of the reservoir. The logs at other locations are used as blind wells to test the performance of seismic AVA inversion method. By performing geostatistical analysis on three extracted well logs, we first calculate the spatial correlation and cross-correlation of three elastic parameters. Then, we simulate realistic pseudo-well logs at all

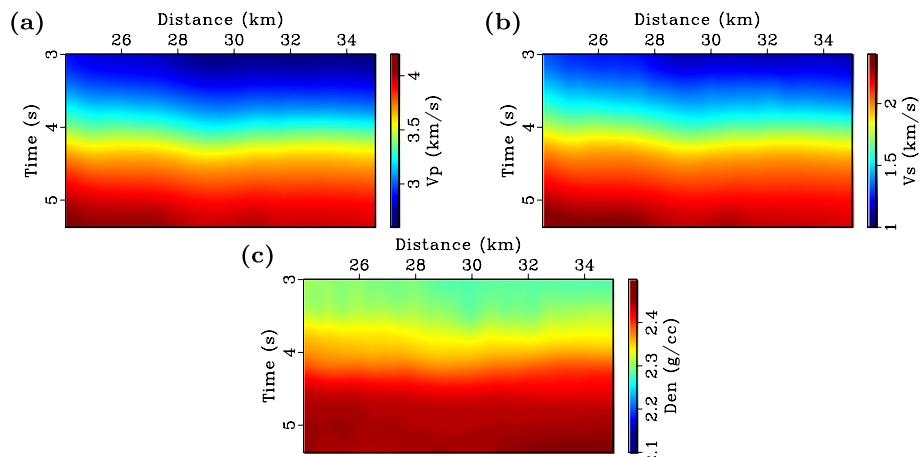
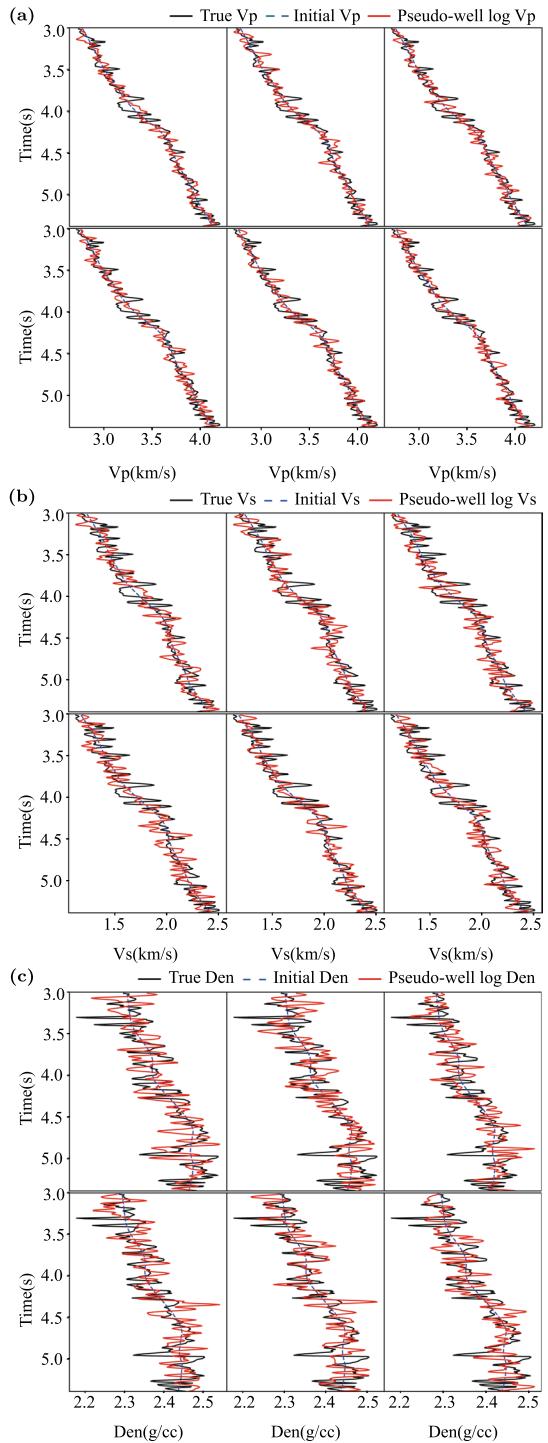


Fig. 4 Panels a–c are initial P-wave velocity, S-wave velocity and density of the local SEAM model (Fig. 3b)

Fig. 5 Pseudo-well logs (red) of P-wave velocity (**a**), S-wave velocity (**b**) and density (**c**). The true curves are plotted in black solid lines and the smoothed curves are plotted in blue dotted lines



CDP locations following the background model trends (Fig. 4a–c) which are calculated by smoothing the true model with a 150 m × 150-m Gaussian filter. The background models only contain the trend and all of the details for the strata are removed. Six examples of simulated pseudo-well logs at one Common Depth Point (CDP) are shown in Fig. 5. Last, the synthetic prestack angle-domain gathers (Fig. 6a) are computed by convolving a 30-Hz Ricker wavelet with reflectivity series, and Fig. 6b shows the resulting prestack seismic gathers. We simulate 22000 synthetic prestack seismic gathers and then randomly divide the training set into two sets: 5% for validation and 95% for training. The proportion of training and validation datasets is set based on the performance of CNN-based AVA inversion method.

The two CNN architectures in Figs. 1 and 2 are used to carry out AVA inversion on two Quadro RTX 5000nvidia graphical processing units. The initial weights of network are set using the Glorot Uniform (Glorot and Bengio 2010). The mean-squared error is used as the loss function to calculate the misfit between the network output and labeled data. The neural network weights are updated in back-propagation with the Adam optimizer (Kingma and Ba 2014). The batch sizes of input data for the first and second networks are 128 and 64, respectively. Their learning rates are set to 0.001 and 0.0001. These hyperparameters are tuned according to the accuracy of CNN prediction results. After training the two CNN architectures with selected optimized hyperparameters, we apply the trained CNN to prestack seismic gathers to predict P-wave velocity, S-wave velocity and density. Compared to the first network estimates (Fig. 8g–i), the prediction results of the second network (Fig. 8j–l) have higher signal-to-noise ratio and better lateral continuity. The second network training is stopped after 50 epochs to avoid the biased fitting issue (Fig. 7), and the loss curves of training and validation converge quickly with the training epoch. The results show that the second network can produce accurate subsurface elastic parameters and perform better than the first network. In the rest tests, we only utilize the second network architecture for CNN-based AVA inversion.

To illustrate the advantages of CNN-based AVA inversion method, we compare it with traditional deterministic linear and nonlinear inversion methods. Traditional inversion

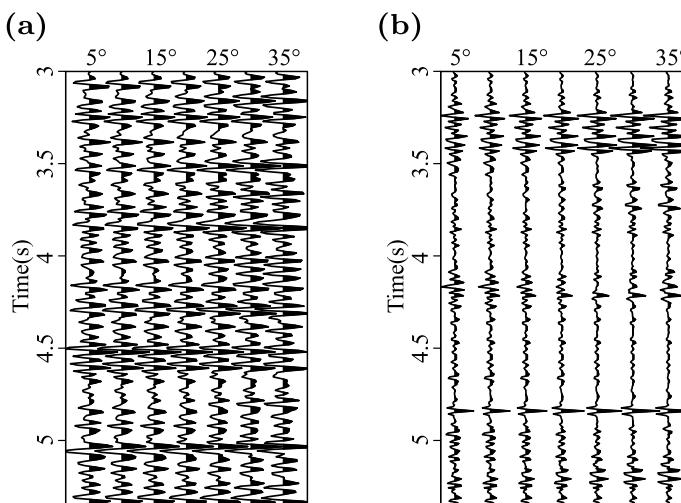


Fig. 6 Synthetic angle-dependent gathers (a) calculated from pseudo-well logs, and real seismic gathers (b). Incident angle varies from 5° to 35°

Fig. 7 Loss curves of the second CNN over 50 iterations

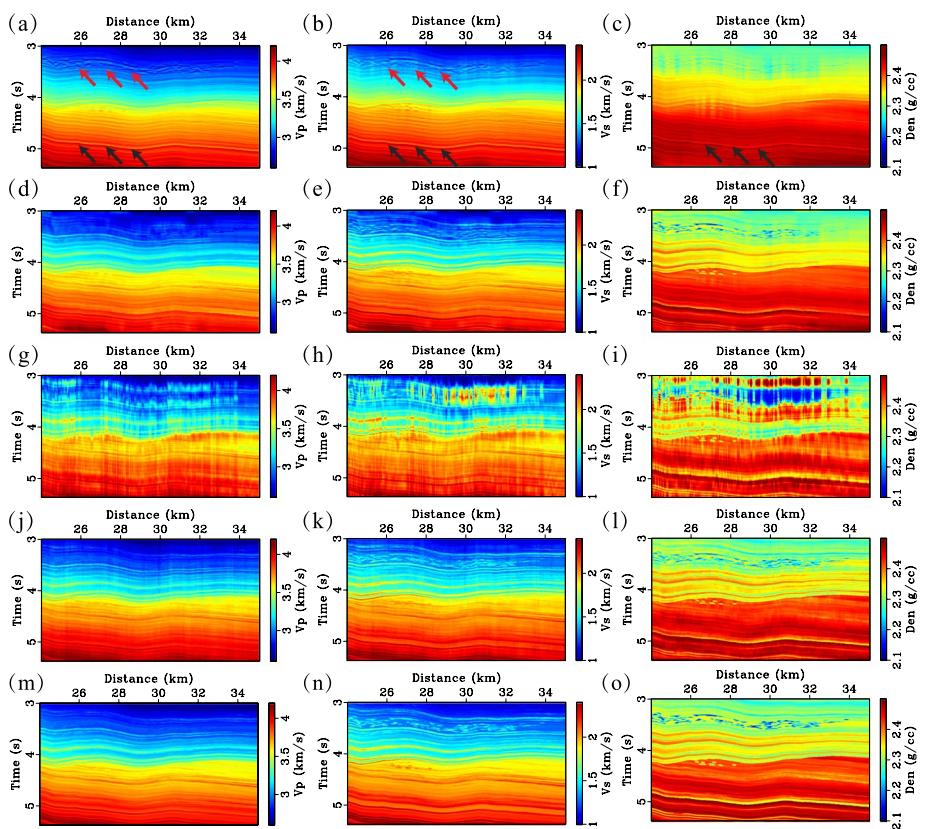
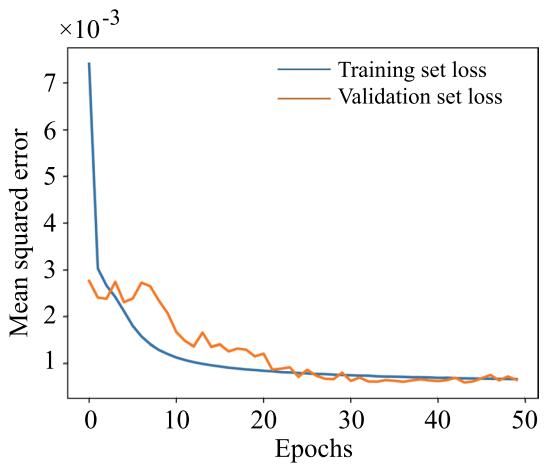


Fig. 8 Prediction results of linear (a–c), nonlinear (d–f), Unet-based (g–i) and the second CNN-based (j–l) AVA inversion methods. The local SEAM model (m–o). The three columns from left to right are for P-wave velocity, S-wave velocity and density, respectively

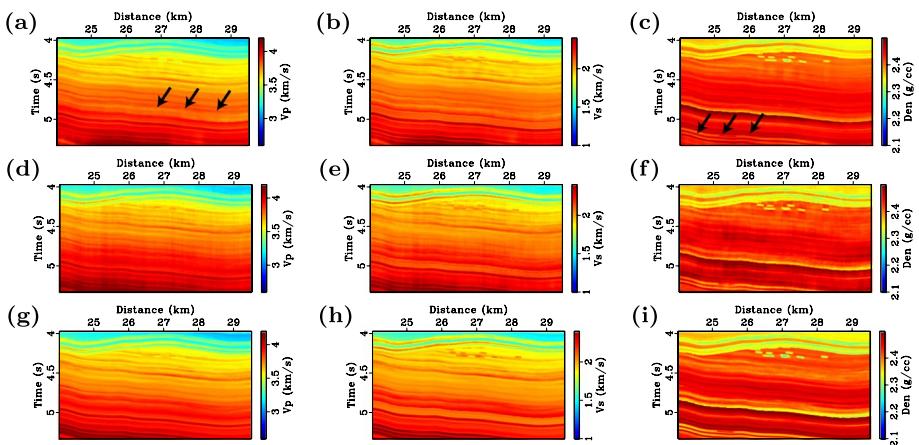


Fig. 9 Enlarged views for the local SEAM model at the region of $x = [24.02 \text{ km}, 29.515 \text{ km}]$ and $z = [3.975 \text{ s}, 5.327 \text{ s}]$. The prediction results of nonlinear (**a–c**) and the second CNN-based (**d–f**) inversion methods. The local SEAM model (**g–i**)

methods start from the initial models and use optimization algorithm to achieve the best fit of synthetic and observed data. Figure 8a–c shows the linear inversion results, and Fig. 8d–f shows the nonlinear inversion results. Linear inversion results have crosstalk interference among P-wave velocity S-wave velocity and density at the depths of $z = [3.1\text{s}, 3.4\text{s}]$ and $z = [4.7\text{s}, 5\text{s}]$ (the red arrows in Fig. 8a, b and the black arrows in Fig. 8a–c). Compared to the linear inversion method, the nonlinear inversion method produces more high-frequency

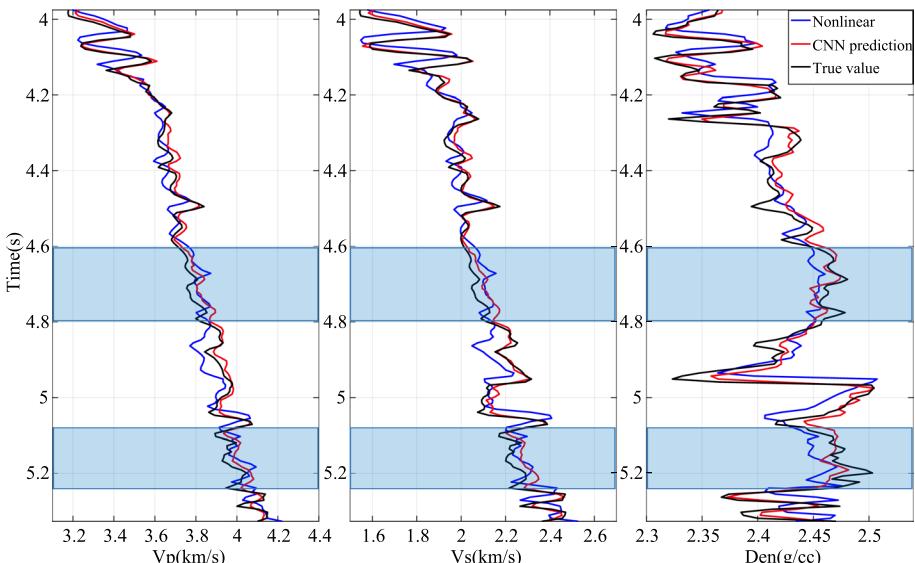


Fig. 10 A comparison of prediction results of nonlinear (blue) and CNN-based (red) AVA inversion methods in the enlarged region (Fig. 9) with true values (black). The blue boxes indicate the crosstalk interference areas. The three columns from left to right are for P-wave velocity, S-wave velocity and density, respectively

Table 3 Average correlation coefficients between true values and the first CNN predictions of training, validation and prediction datasets

	Average correlation coefficient		
	P-wave velocity	S-wave velocity	Density
Training set	0.99816	0.99556	0.97847
Validation set	0.99799	0.99481	0.97541
Prediction set	0.99847	0.99681	0.97898

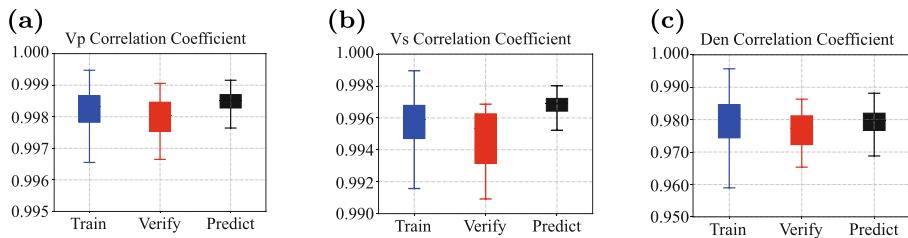


Fig. 11 Box plots of average correlation coefficients between true values and CNN predictions of P-wave velocity (a), S-wave velocity (b) and density (c) in training, validation and prediction datasets

components. However, there are still some crosstalk artifacts among three inverted elastic parameters at the region within $x = [24.02 \text{ km}, 29.515 \text{ km}]$ and $z = [3.975 \text{ s}, 5.327 \text{ s}]$, and the corresponding enlarged views and single-trace fitting are shown in Figs. 9 and 10. The black arrows in Fig. 9 denote the crosstalk artifacts and the density in the areas indicated by blue boxes is influenced by velocity estimates (Fig. 10). The second CNN prediction results (Fig. 8j–l) show that the proposed method can produce stable and accurate results, which match well with the true models (Fig. 8m–o).

We also conduct quantitative analyses for all prediction results. The average correlation coefficients of training, validation and prediction datasets for P-wave velocity, S-wave velocity and density are shown in Table 3 and the corresponding box plots are presented in Fig. 11. The average correlation coefficients of the prediction datasets are larger than that of training datasets, which may be because there are a large number of samples in training datasets. Table 3 indicates that the CNN is stable in three stages and has learned the mapping function between model parameters and prestack seismic data. The comparison of three methods in terms of time consumption and average

Table 4 Comparison of different methods in average correlation coefficient and running time of SEAM example

Method contrast	CNN	Nonlinear inversion	Linear inversion
P-velocity average correlation	0.99847	0.99847	0.92475
S-velocity average correlation	0.99681	0.97478	0.90123
Density average correlation	0.97894	0.86392	0.79483
Time consuming	Training time≈ 0.38h Application time=6s	≈1.0h	≈0.6h

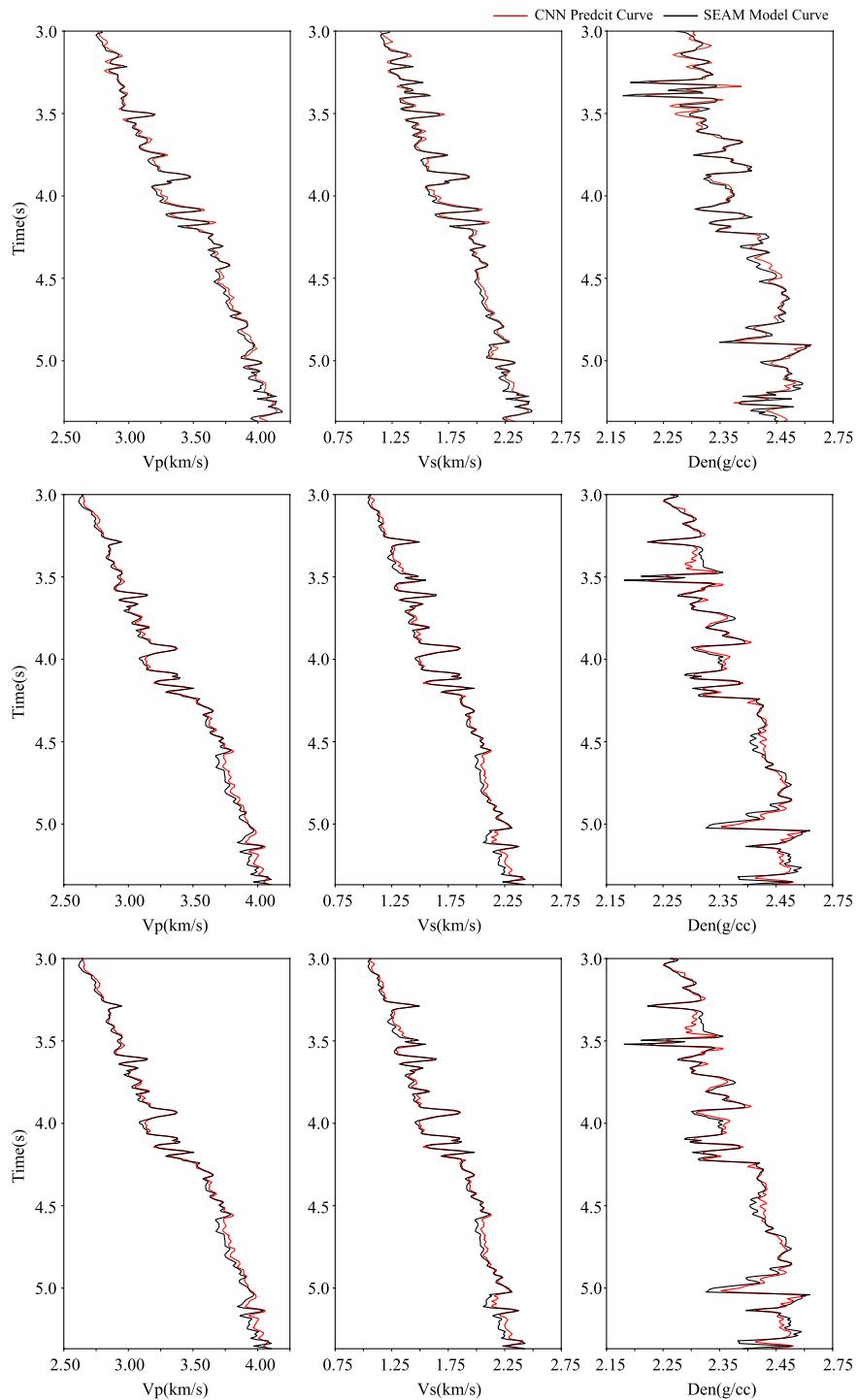
Fig. 12 Comparison of CNN prediction results (red) with true values (black) at three different positions. ▶ Each row corresponds to one location. The three columns from left to right are for P-wave velocity, S-wave velocity and density, respectively

correlation coefficients is presented in Table 4. The CNN-based and nonlinear AVA inversion methods provide similar correlation coefficients for P- and S-wave velocity, and the former has better performance for the density prediction. The traditional inversion methods are more time-consuming compared to CNN-based method, and the time of nonlinear method is about three times of the CNN training time. Figure 12 shows more details of CNN prediction results, which shows that the CNN estimates match well with ground truth. Figure 13 shows the details of prediction results of three AVA inversion methods. The CNN prediction results, especially the density parameter, fit better with ground truth. The plots of correlation coefficients between CNN prediction results and ground truth are presented in Fig. 14. The minimum correlation coefficients of P-wave velocity, S-wave velocity and density are greater than 99.65%, 99.40% and 95.50%, respectively. From the correlation coefficient, we can see that most correlation coefficients are close to 1. Figure 15 shows the histograms of root-mean-square errors (RMSEs) of prediction results of three inversion methods. The prediction results of linear and nonlinear methods have similar RMSEs, which are larger than that of CNN prediction results.

At last, we use the portion of the SEAM model (Fig. 3c) at the depths 0–6 km and distance 0–10 km to test the generalization ability of the trained network (Fig. 3b). The model (Fig. 17a, c, e) has 300-time-sampling points and the time interval is 8 ms. The synthetic prestack angle-domain gathers (Fig. 16b) are computed by convolving a 30-Hz Ricker wavelet (Fig. 16a) with the reflectivity series, which are used as the input for the trained CNN to predict P-wave velocity, S-wave velocity and density (Fig. 17b, d, f). The prediction results produce detailed structures with good lateral continuity and match with the true model well.

4.2 Field Data

We use the field data to test the proposed pseudo-well-based CNN AVA inversion method. The initial models (Fig. 21a, d, g) are calculated by interpolation using available well logs and are used to simulate pseudo-well logs. The horizontal length of the low-frequency model is 2.75 km. The model has 133-time-sampling points and the time interval is 3 ms. The simulated pseudo-well logs at one CDP location are shown in Fig. 18. The synthetic prestack seismic gather (Fig. 20a) is computed by convolving the source wavelet (Fig. 19a) with reflectivity series, in which the source wavelet is extracted from real seismic data (Fig. 20b) and its frequency amplitude spectrum is displayed in Fig. 19b. Figure 21b, e, h and c, f, i shows the prediction results of nonlinear and CNN-based AVA inversion methods, respectively. The stratigraphic structures and location of target reservoir are accurately estimated by CNN-based inversion method. Figure 22 shows a comparison of nonlinear and CNN-based AVA inversion results with well data at 52.85 km distance in Fig. 21, and the corresponding correlation coefficients are shown in Table 5. The prediction results of CNN-based inversion method have similar correlation coefficients to those of nonlinear method, but the CNN-based method is much more efficient. This example demonstrates that the CNN



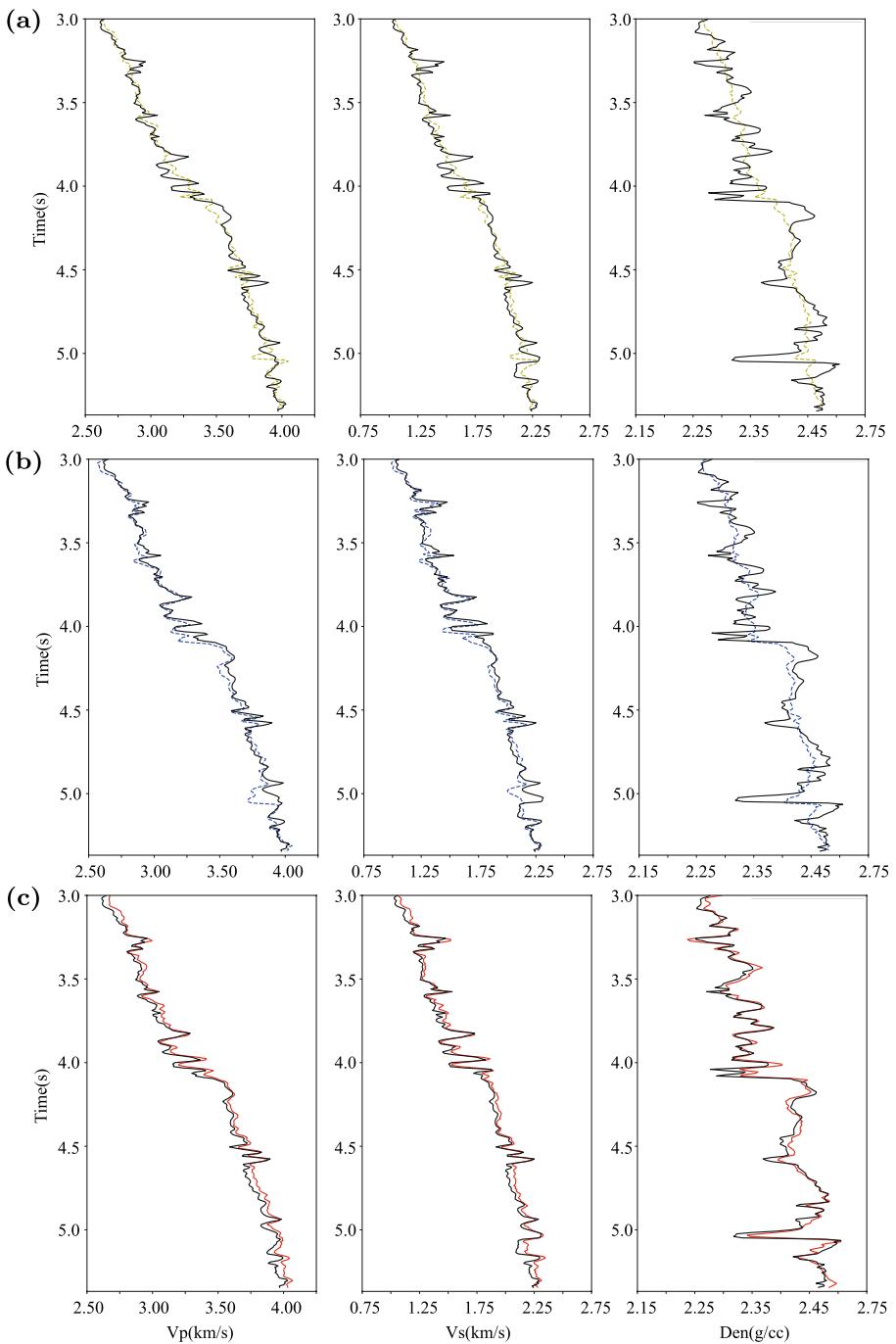


Fig. 13 A comparison of prediction results of linear (a), nonlinear (b) and CNN-based (c) AVA inversion methods with true values (black). The three columns from left to right are for P-wave velocity, S-wave velocity and density, respectively

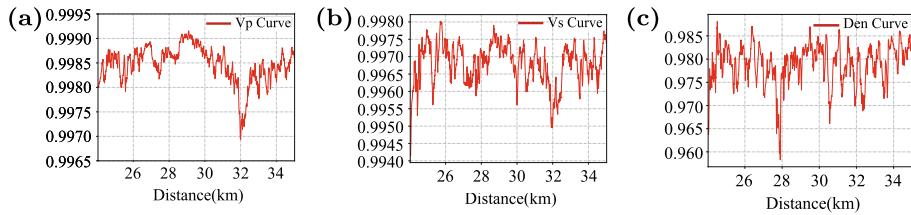


Fig. 14 Average correlation coefficients between CNN predictions and true values of P-wave velocity (a), S-wave velocity (b) and density (c)

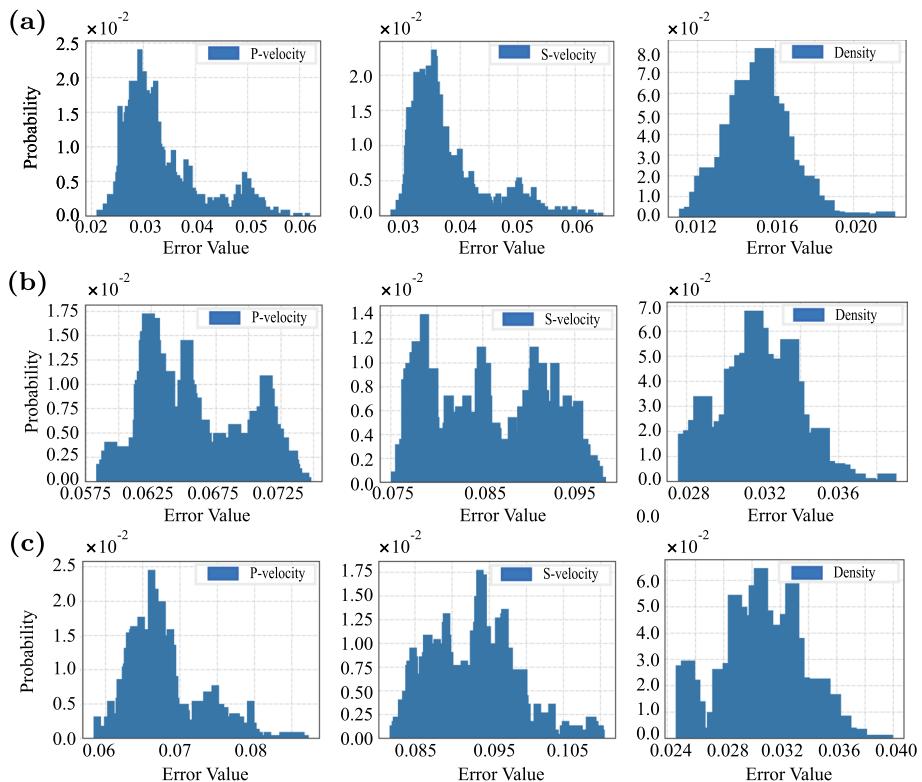


Fig. 15 Histograms of root-mean-square errors between prediction results of the CNN-based (a), nonlinear (b), linear (c) inversion methods and true values

trained with pseudo-well logs can produce accurate results with less cost, which provides an alternative for traditional linear and nonlinear AVA inversion methods.

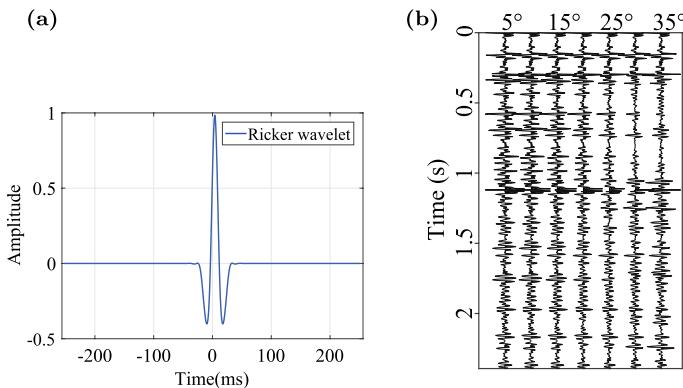


Fig. 16 Ricker wavelet (a) with a peak frequency of 30 Hz and the synthetic angle-dependent gathers (b) corresponding to local SEAM model in Fig. 3c. Incident angle varies from 5° to 35°

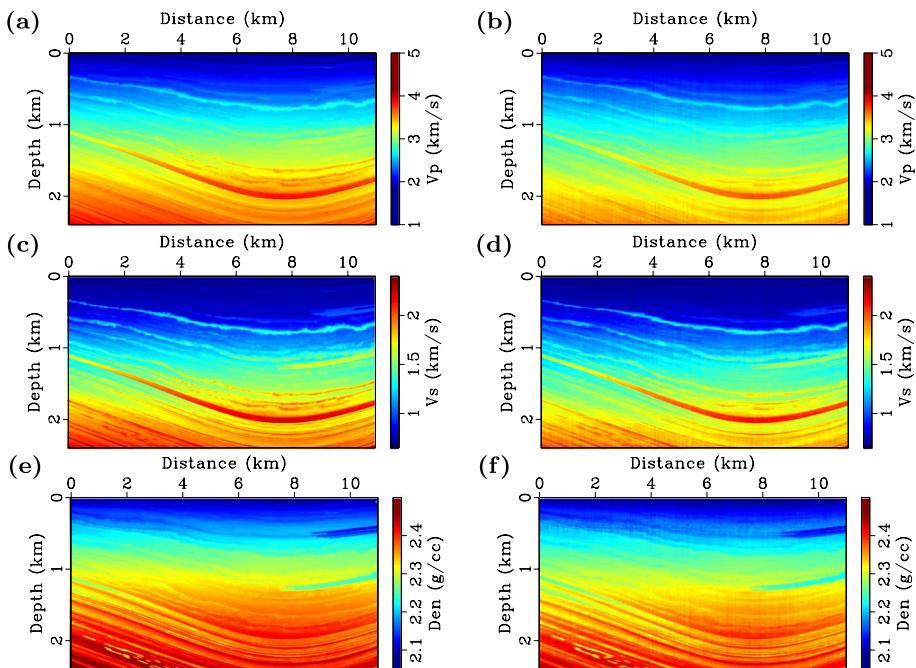
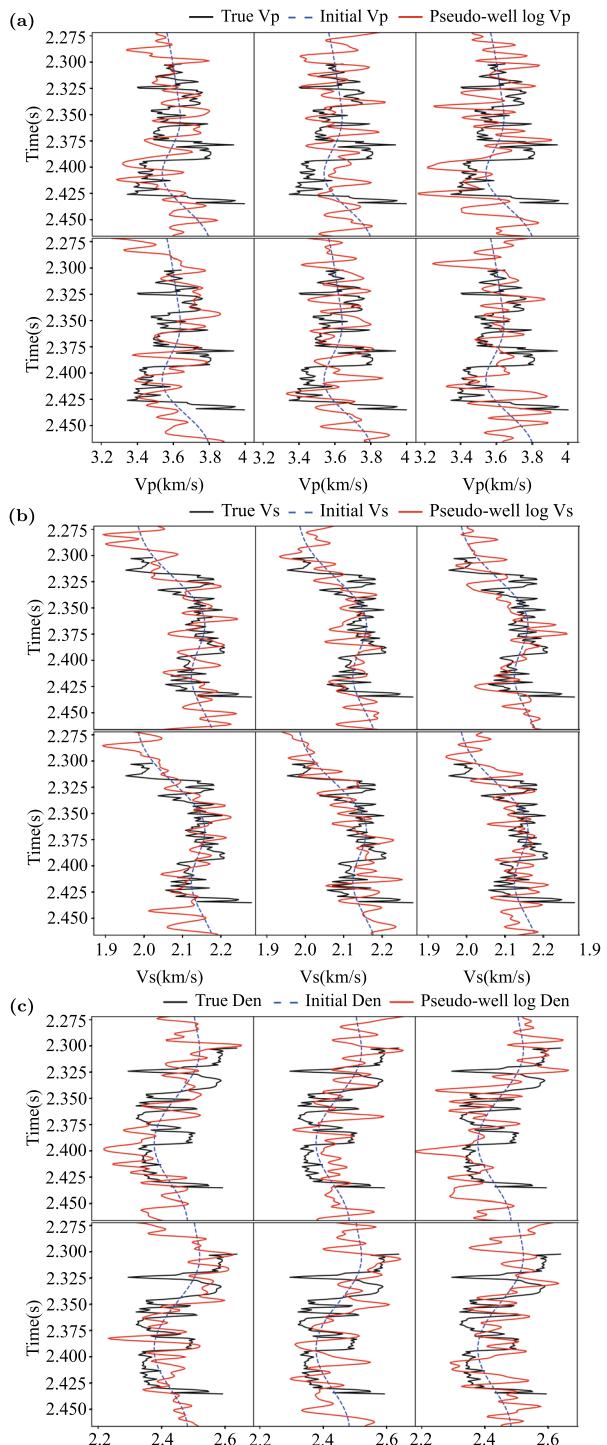


Fig. 17 Panels a, c and e are P-wave velocity, S-wave velocity and density of the local SEAM model (Fig. 3c). Panels b, d and f are prediction results estimated by trained CNN using prestack angle-domain gathers independent of training datasets

5 Discussion

The CNN framework obtains good generalization ability based on the assumption that the feature distribution of training datasets is similar to those of the prediction datasets.

Fig. 18 Pseudo-well logs (red) of P-wave velocity (**a**), S-wave velocity (**b**) and density (**c**). The actual well logs are plotted in black solid lines and the initial curves are plotted in blue dotted lines



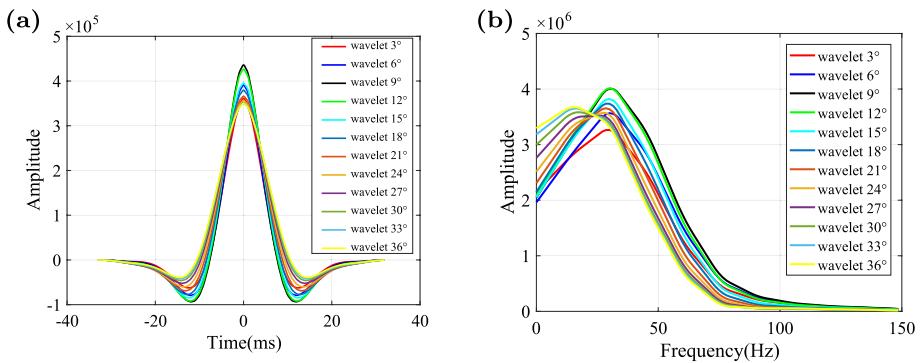


Fig. 19 Panel **a** is the source wavelet extracted from seismic angle gathers and Panel **b** is the corresponding amplitude spectrum

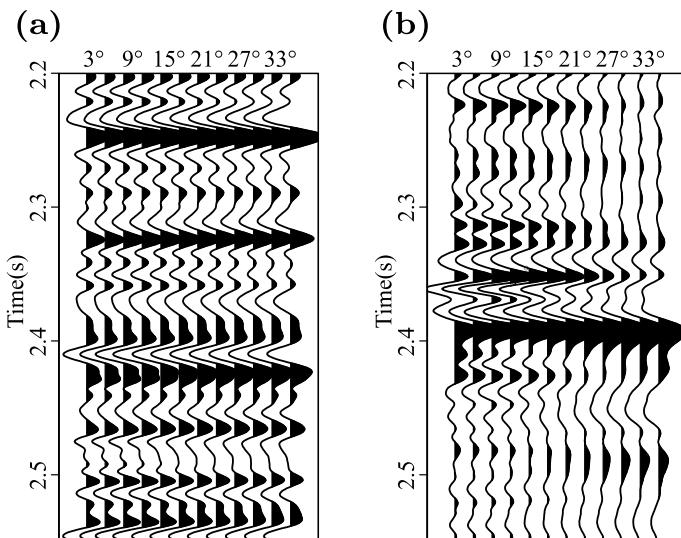


Fig. 20 Synthetic angle-dependent gathers (**a**) calculated from pseudo-well logs, and real seismic gathers (**b**). Incident angle varies from 3° to 36°

However, the statistical information of one example may be not like those in another example. This indicates that for each application we may need to retrain the CNN model using the new training datasets of seismic angle gathers and pseudo-well logs. Although we need to train a new model in different examples, it is still relatively cheap because the training takes only several hours and prediction takes a couple of seconds.

In synthetic examples, the prediction results of traditional linear inversion method have a strong crosstalk. The CNN-based AVA inversion method has an obvious advantage in reducing inter-parameter crosstalk artifacts. In addition, two-dimensional profiles inverted by CNN-based method show clear stratigraphic structure and target reservoir location. The computational time of network training is about one-third of traditional AVA inversion, and the network prediction only takes a few seconds. Quantitative analyses indicate the network

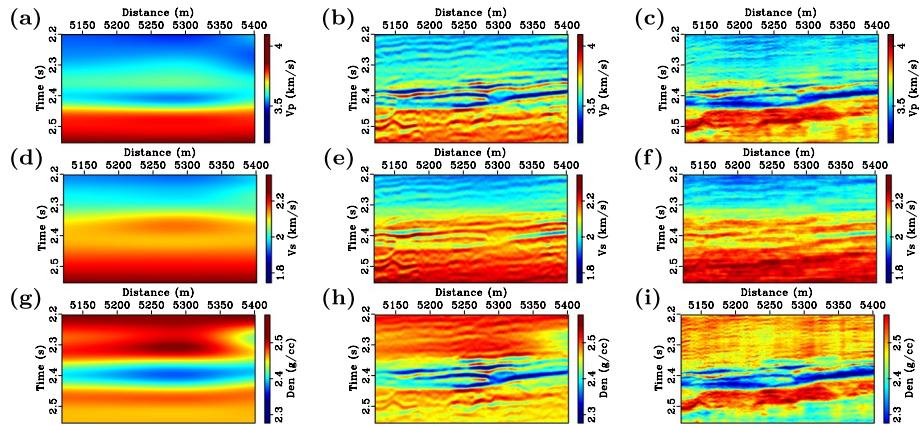


Fig. 21 Initial models (a, d, g), and the prediction results of nonlinear (b, e, h) and CNN-based (c, f, i) AVA inversion methods

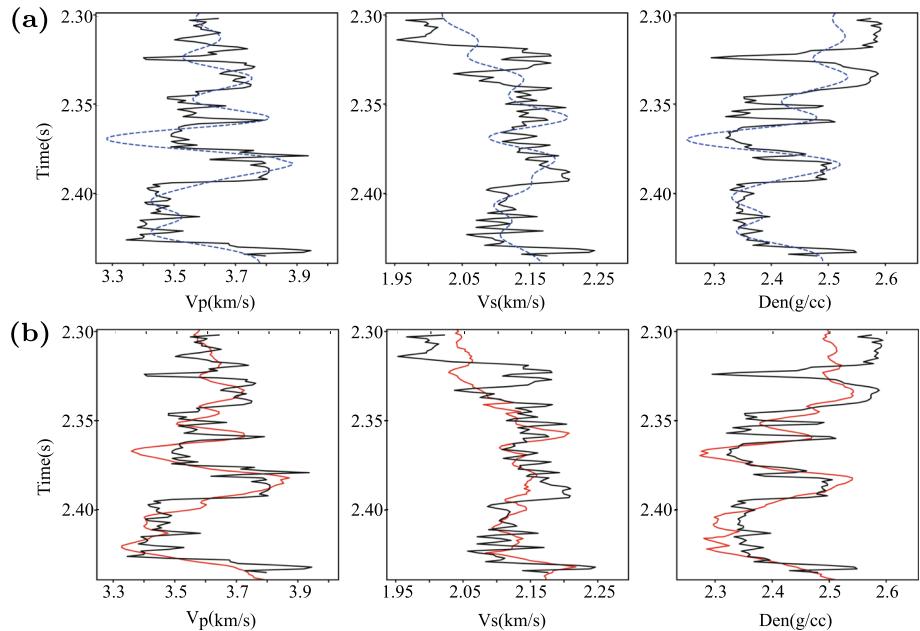


Fig. 22 A comparison of prediction results of nonlinear (a) and CNN-based (b) AVA inversion methods with well logs (black). The three columns from left to right are for P-wave velocity, S-wave velocity and density, respectively

prediction results have better accuracy than traditional inversion methods. The CNN results have a good match with well data and correspond to high correlation coefficients. The analysis not only reveals the accuracy of CNN prediction results, but also determines whether the labeled data (pseudo-well logs) have captured the characteristics of well data. In field data example, the CNN-based AVA inversion method only inverts the structures around the target reservoir, and the signal-to-noise ratio and lateral continuity of prediction results are

Table 5 Comparison of different methods in average correlation coefficient and running time of field example

Correlation coefficient				Time consuming
	P-wave velocity	S-wave velocity	Density	
CNN	0.7436	0.5474	0.7762	Training time \approx 1.01h Application time \approx 4.9s
Nonlinear inversion	0.6942	0.5915	0.8093	\approx 1.35h

slightly lower than that of synthetic results. There are three possible reasons. First, the initial model computed by well interpolation is not as accurate as the smoothed background model in synthetic tests. Second, the angle-domain gathers of field data have some noise which is not considered in the network training. Third, we only calculate the spatial correlation and cross-correlation parameters for reservoir strata in this experiment. It might be different from that of layers far away from reservoirs, which could lead to inaccurate prediction results.

Although it takes only a few seconds when applying the trained network to seismic data for elastic parameters prediction, it needs a long time to update CNN weights in training. There are many hyperparameters that need to be adjusted repeatedly in the training stage, such as the learning rate, network update epoch, division proportion of training and verification set, the size of the mini-batch and the ratio of the numbers between training data samples and network weights, etc. If one of the hyperparameters is adjusted, the network needs to be retrained. In addition, to achieve good learning ability, the synthetic prestack angle-domain gathers and real seismic gathers should be preprocessed manually to ensure that they have the same attributes. The generation of training samples, the design of network architecture, the adjustment of hyperparameters and data preprocessing all need large time consumption and human experience. Therefore, when using CNN for AVA inversion, we should try to strike a balance between efficiency and accuracy.

6 Conclusions

Due to the advances in deep learning and computer technology, the applications of CNN have made great progress in various geoscience problems. The accuracy of parameter prediction using CNN from seismic data largely depends on the size of training datasets. However, the labeled data in seismic inversion are limited by sparse well locations. To mitigate this drawback, we propose an intelligent AVA inversion using CNN trained with a pseudo-well datasets technique. We first compute geologically realistic pseudo-well logs based on the statistical analysis. Then, prestack angle-domain seismic gathers are calculated using pseudo-well logs. Next, the synthetic pseudo-well logs and seismic gathers are used to update the CNN weights. At last, the trained CNN is applied to seismic gathers for elastic parameter prediction. We use the SEAM and field data examples to test the feasibility and adaptability of the proposed CNN-based AVA inversion method. The CNN-based method not only produces more high-frequency components and avoids the crosstalk interference, but also improves the efficiency of seismic inversion.

Appendix A

Bayesian theory is a probability statistical method to calculate conditional probability, which can easily introduce prior information of model parameters in the objective function. Thus, it is often used in the inverse problem (Buland and Omre 2003). In seismic inversion, the Bayesian expression can be written as

$$P(\mathbf{m} | \mathbf{d}) = \frac{P(\mathbf{d} | \mathbf{m})P(\mathbf{m})}{P(\mathbf{d})}, \quad (\text{A.1})$$

where $P(\mathbf{m} | \mathbf{d})$ represents the posterior probability distribution of \mathbf{m} to the \mathbf{d} , $P(\mathbf{d} | \mathbf{m})$ is the likelihood function that describes how likely the observed data are for elastic parameters, $P(\mathbf{m})$ is the prior probability distribution of \mathbf{m} , and $P(\mathbf{d})$ is the marginal probability distribution of the \mathbf{d} . Because $P(\mathbf{d})$ is regarded as a constant value and can be ignored, equation (A.1) can be simplified as

$$P(\mathbf{m} | \mathbf{d}) \propto P(\mathbf{d} | \mathbf{m})P(\mathbf{m}). \quad (\text{A.2})$$

In most cases, the noise of the recorded seismic is assumed to satisfy Gaussian distribution, and the likelihood function has the following expression

$$P(\mathbf{d} | \mathbf{m}) = [(2\pi)^{n_t} |\mathbf{C}_D|]^{1/2} \exp \left[-\frac{1}{2} (\mathbf{d} - \mathbf{GLm})^T \mathbf{C}_D^{-1} (\mathbf{d} - \mathbf{GLm}) \right], \quad (\text{A.3})$$

where \mathbf{C}_D^{-1} is the covariance matrix of noise. Assuming the noise values in seismic data are independent, the covariance matrix can be simplified to a diagonal matrix $\mathbf{C}_D = \sigma_{n_t}^2 \mathbf{I}$, where $\sigma_{n_t}^2$ is the variance of noise, \mathbf{I} is an identity matrix of $n_t \times n_t$.

The heavy-tailed distribution of Cauchy can enforce sparsity characterization and shows great advantages in improving the stability and accuracy of the inversion results (Alemie and Sacchi 2011; Ma and Geng 2013). The prior probability distribution of model parameters can be expressed as

$$P(\mathbf{m}) = \frac{1}{\pi^{(2N)} |\boldsymbol{\Psi}|^{N/2}} \exp \left(-2 \sum_{i=1}^N \ln (1 + (\mathbf{m} - \boldsymbol{\mu})^T \boldsymbol{\Phi}^i (\mathbf{m} - \boldsymbol{\mu})) \right), \quad (\text{A.4})$$

where $\boldsymbol{\Phi}^i = (\mathbf{D}^i)^T \boldsymbol{\Psi}^{-1} \mathbf{D}^i$, $\boldsymbol{\Psi}$ is a 3×3 covariance matrix, which contains the statistical information obtained from well data. N and $\boldsymbol{\mu}$ represent the length and mean values of the model parameters, respectively. \mathbf{D}^i is a $3 \times 3N$ matrix, the specific form can be written as

$$[\mathbf{D}_{nl}^i] = \begin{cases} 1, & \text{if } n = 1 \text{ and } l = i \\ 1, & \text{if } n = 2 \text{ and } l = i + N \\ 1, & \text{if } n = 3 \text{ and } l = i + 2N \\ 0, & \text{other} \end{cases}. \quad (\text{A.5})$$

The maximum posteriori probability solution of equation (A.2) is equivalent to the solution of the objective function $J(\mathbf{m})$, and the objective function has the following form

$$J(\mathbf{m}) = \frac{1}{2} (\mathbf{d} - \mathbf{GLm})^T \mathbf{C}_D^{-1} (\mathbf{d} - \mathbf{GLm}) + 2 \sum_{i=1}^N \ln (1 + (\mathbf{m} - \boldsymbol{\mu})^T \boldsymbol{\Phi}^i (\mathbf{m} - \boldsymbol{\mu})). \quad (\text{A.6})$$

Appendix B

The model parameter perturbation can be computed by Gaussian–Newton optimization algorithm. The second-order Taylor series expansion of the objective function at \mathbf{m}^0 can be written as

$$J(\mathbf{m}) \approx J(\mathbf{m}^0) + \frac{\partial J(\mathbf{m}^0)}{\partial \mathbf{m}} \Delta \mathbf{m} + \frac{1}{2} \frac{\partial^2 J(\mathbf{m}^0)}{\partial \mathbf{m}^2} \Delta \mathbf{m}^2, \quad (\text{B.1})$$

where $\mathbf{m} = \mathbf{m}^0 + \Delta \mathbf{m}$, \mathbf{m}^0 represents the initial model and $\Delta \mathbf{m}$ is the model perturbation.

The first derivative of the objective function with respect to the model parameters can be calculated as

$$\frac{\partial J(\mathbf{m})}{\partial \mathbf{m}} = \left(\frac{\partial \mathbf{G}(\mathbf{m})}{\partial \mathbf{m}} \right)^T (\mathbf{G}(\mathbf{m}) - \mathbf{d}) + \sigma_{n_i}^2 \frac{\partial R(\mathbf{m})}{\partial \mathbf{m}} = \boldsymbol{\gamma}(\mathbf{m}), \quad (\text{B.2})$$

where $R(\mathbf{m}) = 2 \sum_{i=1}^N \ln (1 + (\mathbf{m} - \boldsymbol{\mu})^T \boldsymbol{\Phi}^i (\mathbf{m} - \boldsymbol{\mu}))$.

The second derivative of the objective function with respect to the model parameters is computed as

$$\begin{aligned} \frac{\partial^2 J(\mathbf{m})}{\partial \mathbf{m}^2} &= \left(\frac{\partial^2 \mathbf{G}(\mathbf{m})}{\partial \mathbf{m}^2} \right)^T (\mathbf{G}(\mathbf{m}) - \mathbf{d}) + \left(\frac{\partial \mathbf{G}(\mathbf{m})}{\partial \mathbf{m}} \right)^T \left(\frac{\partial \mathbf{G}(\mathbf{m})}{\partial \mathbf{m}} \right) \\ &\quad + \sigma_{n_i}^2 \frac{\partial^2 R(\mathbf{m})}{\partial \mathbf{m}^2}. \end{aligned} \quad (\text{B.3})$$

Because the second derivative of the nonlinear forward operator with respect to the model parameters too expensive to compute, it is often ignored in the actual application. A simplified form is

$$\frac{\partial^2 J(\mathbf{m})}{\partial \mathbf{m}^2} \approx \left(\frac{\partial \mathbf{G}(\mathbf{m})}{\partial \mathbf{m}} \right)^T \left(\frac{\partial \mathbf{G}(\mathbf{m})}{\partial \mathbf{m}} \right) + \sigma_{n_i}^2 \frac{\partial^2 R(\mathbf{m})}{\partial \mathbf{m}^2} = \mathbf{H}(\mathbf{m}). \quad (\text{B.4})$$

Calculating the first derivative of the objective function in equation (B.1) with respect to the model parameter' perturbations $\Delta \mathbf{m}$ and setting the derivative value equals to zero, the specific form of perturbations can be written as

$$\Delta \mathbf{m}^0 = -\mathbf{H}(\mathbf{m}^0)^{-1} \boldsymbol{\gamma}(\mathbf{m}^0). \quad (\text{B.5})$$

Acknowledgements This research is supported by the startup funding (no. 20CX06069A) of Guanghua Scholar in Geophysics Department, at China University of Petroleum (East China). We thank the support from the National Outstanding Youth Science Foundation (no. 41922028), the National Natural Science Foundation of China (General Program) (no. 41874149), the Strategic Priority Research Program of the Chinese Academy of Sciences (no. XDA14010303), the Key Program for International Cooperation Projects of China (no. 41720104006), the Natonal Key R &D Program of China (no. 2019YFC0605503), the Major Scientific and Technological Projects of CNPC (no. ZD2019183003), the Key technique of full node seismic processing (no. 30200020-21-ZC0607-0021), and the Funds for Creative Research Groups of China (no. 41821002).

References

- Aki K, Richards P (1980) Quantitative seismology: theory and methods. Freeman, San Francisco, p 859
- Aleardi M, Salusti A (2021) Elastic prestack seismic inversion through discrete cosine transform reparameterization and convolutional neural networks. *Geophysics* 86(1):R129–R146
- Alemie W, Sacchi MD (2011) High-resolution three-term avo inversion by means of a trivariate cauchy probability distribution. *Geophysics* 76(3):R43–R55
- Alfarraj M, AlRegib G (2019) Semisupervised sequence modeling for elastic impedance inversion. *Interpretation* 7(3):SE237-se249
- Baldi P, Sadowski PJ (2013) Understanding dropout. *Adva Neural Inform Process Syst* 26:2826–284
- Bauer A, Walda J, Gajewski D (2021) Wavefield decomposition for diffraction separation with convolutional neural networks. In: SEG technical program expanded abstracts, pp 2874–2878
- Bergen KJ, Johnson PA, Maarten V, Beroza GC (2019) Machine learning for data-driven discovery in solid Earth geoscience. *Science* 363:6433
- Biswas R, Sen MK, Das V, Mukerji T (2019) Pre-stack inversion using a physics-guided convolutional neural network. In: SEG technical program expanded abstracts, pp 4967–4971
- Bottou L (2012) Stochastic gradient descent tricks. In: *Tricks of the trade, neural networks*, pp 421–436
- Buland A, Omre H (2003) Bayesian linearized AVO inversion. *Geophysics* 68:185–198
- Chen T, Wei S, Jia L (2009) Nonlinear simultaneous PP and PS AVO inversion based on Zeeppritz equations. In: Beijing international geophysical conference and exposition 2009: Beijing 2009 international geophysical conference and exposition, Beijing, China, 24–27 April 2009, pp 248–248
- Chua KK, Tay YH (2013) Enhanced image super-resolution technique using convolutional neural network. In: International visual informatics conference, pp 157–164
- Das V, Mukerji T (2020) Petrophysical properties prediction from prestack seismic data using convolutional neural networks. *Geophysics* 85(5):N41–N55
- Das V, Pollack A, Wollner U, Mukerji T (2019) Convolutional neural network for seismic impedance inversion. *Geophysics* 84(6):R869–R880
- Downton JE, Lines LR (2001) Constrained three parameter AVO inversion and uncertainty analysis, pp 251–254
- Downton JE, Collet O, Hampson DP, Colwell T (2020) Theory-guided data science-based reservoir prediction of a North Sea oil field. *Lead Edge* 39(10):742–750
- Du J, Liu J, Zhang G, Han L, Li N (2019) Pre-stack seismic inversion using seisinv-resnet, pp 2338–2342
- Dvorkin J, Gutierrez MA, Grana D (2014) Seismic reflections of rock properties. Cambridge University Press, Cambridge
- Fatti J, Smith G, Vail P, Strauss P, Levitt PR (1994) Detection of gas in sandstone reservoirs using AVO analysis: a 3-D seismic case history using the Geostack technique. *Geophysics* 59(9):1362–1376
- Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics, pp 249–256
- Hollander Y, Merouane A, Yilmaz O (2018) Using a deep convolutional neural network to enhance the accuracy of first-break picking. In: SEG technical program expanded abstracts, pp 4628–4632
- Jha D, Riegler MA, Johansen D, Halvorsen P, Johansen HD (2020) Doubleu-net: a deep convolutional neural network for medical image segmentation. In: 2020 IEEE 33rd international symposium on computer-based medical systems (CBMS), pp 558–564
- Jin Y, Wu X, Chen J, Han Z, Hu W (2018) Seismic data denoising by deep-residual networks. In: SEG technical program expanded abstracts, pp 4593–4597
- Kaur H, Pham N, Fomel S (2020) Improving the resolution of migrated images by approximating the inverse Hessian using deep learning. *Geophysics* 85(4):WA173–WA183
- Kaur H, Pham N, Fomel S (2021) Seismic data interpolation using deep learning with generative adversarial networks. *Geophys Prospect* 69(2):307–326
- Kaur H, Zhong Z, Sun A, Fomel S (2022) Time-lapse seismic data inversion for estimating reservoir parameters using deep learning. *Interpretation* 1:10
- Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
- Koefoed O (1955) On the effect of Poisson's ratios of rock strata on the reflection coefficients of plane waves. *Geophys Prospect* 3(4):381–387
- Kuzma HA, Rector JW (2004) Non-linear AVO inversion using support vector machines. In: SEG Technical program expanded abstracts, pp 203–206
- Lewis W, Vigh D (2017) Deep learning prior models from seismic images for full-waveform inversion, pp 1512–1517

- Li H, Yang W, Yong X (2018) Deep learning for ground-roll noise attenuation. In: SEG Technical program expanded abstracts, pp 1981–1985
- Ma JQ, Geng JH (2013) Cauchy prior distribution-based AVO elastic parameter estimation via weakly nonlinear waveform inversion. *Appl Geophys* 10(4):442–452
- Mallick S (1995) Model-based inversion of amplitude-variations-with-offset data using a genetic algorithm. *Geophysics* 60(4):939–954
- Nocedal J, Wright SJ (1999) Numerical optimization. New York, NY: Springer New York
- Ostrander W (1984) Plane-wave reflection coefficients for gas sands at nonnormal angles of incidence. *Geophysics* 49(10):1637–1648
- Raymond PWRF, Alfred CFB (1995) The synthesis of seismograms from well log data. *Geophysics* 20(3):516–538
- Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional Networks for Biomedical Image Segmentation. MICCAI, pp 234–241
- Serfaty Y, Itan L, Chase D, Koren Z (2017) Wavefield separation via principle component analysis and deep learning in the local angle domain. In: SEG technical program expanded abstracts, pp 991–995
- Shi Y, Wu X, Fomel S (2018) Automatic salt-body classification using a deep convolutional neural network, pp 1971–1975
- Shuey R (1985) A simplification of the Zoeppritz equations. *Geophysics* 50(4):609–614
- Siahkoohi A, Verschuur DJ, Herrmann FJ (2019) Surface-related multiple elimination with deep learning. In: SEG technical program expanded abstracts, pp 4629–4634
- Stewart RR (1990) Joint P and P-SV inversion. CREWES Project Res Rep 2:112–115
- Wang W, Ma J (2019) PS decomposition of isotropic elastic wavefields using CNN-learned filters. In: 81st EAGE conference and exhibition 2019 2019, vol 1, pp 1–5
- Wang K, Bandura L, Bevc D, Cheng S, DiSiena J, Halpert A, Osypov K, Power B, Xu E (2019) End-to-end deep neural network for seismic inversion. In: SEG technical program expanded abstracts, pp 4982–4985
- Wang Y, Wang Q, Lu W, Ge Q, Yan X (2021) Seismic impedance inversion based on cycle-consistent generative adversarial network. Pet Sci
- Wei X, Chen T (2011) Joint PP and PS AVO inversion based on Zoeppritz equations. *Earthq Sci* 24(4):329–334
- Wu X, Liang L, Shi Y, Fomel S (2019) FaultSeg3D: using neural network for 3D seismic fault segmentation. *Geophysicssynthetic* data sets to train an end-to-end convolutional 84(3):WA35–WA45
- Wu X, Geng Z, Shi Y, Pham N, Fomel S, Caumon G (2020) Building realistic structure models to train convolutional neural networks for seismic structural interpretation. *Geophysics* 85(4):WA27–WA39
- Yuan S, Liu J, Wang S, Wang T, Shi P (2018) Seismic waveform classification and first-break picking using convolution neural networks. *IEEE Geosci Remote Sens Lett* 15(2):272–276
- Zhang K, Zuo W, Chen Y, Meng D, Zhang L (2017) Beyond a gaussian denoiser: residual learning of deep cnn for image denoising. *IEEE Trans Image Process* 26(7):3142–3155
- Zhang Y, Lin H, Li Y (2018b) Noise attenuation for seismic image using a deep residual learning. In: SEG technical program expanded abstracts, pp 2176–2180
- Zhao T (2018) Seismic facies classification using different deep convolutional neural networks. In: SEG technical program expanded abstracts, pp 2046–2050
- Zhou L, Li J, Chen X, Liu X, Chen L (2017) Prestack amplitude versus angle inversion for Young's modulus and Poisson's ratio based on the exact Zoeppritz equations. *Geophys Prospect* 65(6):1462–1476
- Zong Z, Yin X, Wu G (2012) AVO inversion and poroelasticity with P-and S-wave moduli. *Geophysics* 77(6):N17–N24

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Jiaxing Sun¹ · Jidong Yang¹  · Zhenchun Li¹ · Jianping Huang¹ · Xin Luo² · Jie Xu¹

Jiaxing Sun
b20010035@s.upc.edu.cn

Zhenchun Li
19830010@upc.edu.cn

Jianping Huang
jphuang@upc.edu.cn

Xin Luo
luoxin21@cdut.edu.cn

Jie Xu
1701040128@s.upc.edu.cn

¹ School of Geosciences, Laboratory for Marine Mineral Resources, and National Laboratory for Marine Science and Technology, Key Laboratory of Deep Oil and Gas, China University of Petroleum, Qingdao 266580, Shandong, China

² Institute of sedimentary Geology, Chengdu University of Technology, Chengdu 610059, Sichuan, China