Near-Field Geoacoustic Inversion Using Bottom Reflection Signals via Self-Attention Mechanism

Yuxuan Ma, Xiaobo Zhang, Fukun Jiang, Zhengrong Wei, and Chenguang Liu

***Abstract*—Geoacoustic inversion typically involves the collection of far-field underwater acoustic data to obtain seabed geoacoustic parameters using empirical formulas and matched field inversion (MFI) techniques. However, acoustic data propagated over long distances can introduce inevitable errors in inversion results, and traditional MFI techniques suffer from low computational efficiency. Although deep learning technologies have been applied to geoacoustic inversion, conventional deep neural network (DNN) models struggle to capture the long- and short-term dependencies in bottom reflection data, leading to suboptimal inversion accuracy. These issues present challenges in rapidly and accurately acquiring geoacoustic parameters over large areas. To address this, we propose a near-field bottom reflection signal collection method, collecting bottom reflection signals over a wide range of grazing angles by drifting. Utilizing the characteristics of near-field sound propagation, we constructed the bottom reflection coefficient sequence dataset using the wavenumber integration method. We then introduce a novel deep learning model, self-attention geoacoustics, based on multihead self-attention mechanisms, which improves inversion accuracy. In addition, we propose an adaptive-weight multitask learning training strategy, significantly enhancing the prediction accuracy of sound attenuation. Experimental results demonstrate that our method outperforms conventional geoacoustic inversion methods based on MFI and DNNs in terms of efficiency and accuracy, proving the superiority of our approach.**

***Index Terms*—Geoacoustic inversion, near-field bottom reflection signal, self-attention mechanism, wavenumber integration.**

I. Introduction

TWO primary methods exist for acquiring seabed geoacoustic parameters (typically refers to sound speed, density, and sound attenuation coefficient): experimental measurement (including in-situ measurements and sample measurements) and geoacoustic inversion. Experimental measurement methods are costly and are not suitable for obtaining geoacoustic parameters over vast areas, and they encounter challenges when acquiring low-frequency acoustic properties [1]. In contrast, geoacoustic inversion infers the geoacoustic parameters of a target marine area from acoustic signals, offering a rapid and economical method of acquisition.

Geoacoustic inversion typically involves receiving bottom reflection signals at considerable distances (ranging from a few miles to several tens of miles). While the far-field approach facilitates the acquisition of geoacoustic parameters over a large area, the signals are susceptible to the complex marine environment, resulting in data that contain a significant amount of noise. Moreover, obtaining bottom reflection data at larger grazing angles becomes challenging at long distances, negatively impacting inversion accuracy. Therefore, collecting acoustic data from the far field is not conducive to improving the accuracy of geoacoustic inversion, making inversion based on near-field acoustic data a potential solution.

After collecting actual bottom reflection data, empirical formulas and matched field inversion (MFI) are generally used to estimate the geoacoustic parameters of the target sea area [2], [3]. The empirical formulas are derived from a plethora of measured data, encapsulating statistical relationships among different sediment property parameters. However, empirical formulas lack generalization ability, only applicable to limited maritime areas. MFI iteratively optimizes based on a forward model of the sound field, seeking geoacoustic parameters that best match the measured acoustic signals. But MFI can invert geoacoustic parameters from only one set of bottom reflection data at a time, preventing rapid acquisition of geoacoustic parameters over a large area and failing to account for the influence of noise.

With the rapid development of deep learning, it has also yielded promising results in the field of ocean acoustics [4], [5]. Geoacoustic inversion based on deep learning [5], [6], [7] can be conceptualized as a multiparameter nonlinear regression problem. The goal of training deep neural networks (DNNs) is for the model to learn the mapping relationship from acoustic signals to geoacoustic parameters, thereby enabling the prediction of geoacoustic parameters based on the input of measured acoustic signals. Compared to typical approaches, DNN models possess superior efficiency and generalization capabilities, paving the way for rapid acquisition of geoacoustic parameters across extensive regions. However, we found that conventional DNNs cannot effectively capture the long- and short-term dependencies in bottom reflection data, resulting in limitations in accuracy. In addition, when predicting multiple geoacoustic parameters

Manuscript received 1 March 2024; revised 18 April 2024; accepted 21 May 2024. Date of publication 24 May 2024; date of current version 5 May 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42106072 and in part by the Shandong Provincial Natural Science Foundation under Grant ZR2020QD071. *(Corresponding author: Xiaobo Zhang.)*

Yuxuan Ma, Fukun Jiang, and Zhengrong Wei are with the College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China (e-mail: myx@sdust.edu.cn; jfk@sdust.edu.cn; wzr@sdust.edu.cn).

Xiaobo Zhang is with the College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China, and also with the Laboratory for Marine Geology, Qingdao Marine Science and Technology Center, Qingdao 266237, China (e-mail: zxb@sdust.edu.cn).

Chenguang Liu is with the First Institute of Oceanography, Ministry of Natural Resources, Qingdao 266061, China (e-mail: lcg@fio.org.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3405202

simultaneously, the significant difference in sensitivity of different geoacoustic parameters to bottom reflection data may lead to substantial errors in the inversion results for specific parameters.

To address the aforementioned challenges in the field of geoacoustic inversion, it is necessary to tackle several key issues, including optimizing seabed reflection data collection methods, constructing simulation datasets that accurately reflect geoacoustic features, developing more effective DNN model architectures, and devising training strategies that efficiently balance the variability among different parameters.

Effective data collection methods are foundational for precise geoacoustic inversion. Geoacoustic inversion research typically collects far-field geoacoustic data, which introduces obvious errors and is not conducive to accurately obtaining local geoacoustic parameters [8], [9], [10]. Moreover, geoacoustic inversion studies usually employ fixed-position sound sources and receiver arrays, making it difficult to receive bottom reflection signals across a wide range of grazing angles. To solve these data collection issues, we propose a method for drifting collection of near-field bottom reflection data (NBRC). It utilizes drifting sound sources and fixed receiver arrays to acquire bottom reflection data. By continuously changing the horizontal distance between the sound source and receiver elements, we collected bottom reflection data at a wide range of grazing angles, simultaneously improving the signal-to-noise ratio of the measurement signals.

To enhance the accuracy of geoacoustic inversion, it is necessary to establish a dataset that accurately reflects the characteristics of geoacoustic features. Due to constraints in real marine experiments, geoacoustic inversion typically relies on a forward model for dataset construction rather than extensive marine data collection. Geoacoustic inversion research commonly utilizes the normal mode method [2], [3], [5] for forward modeling, but this approach neglects the effect of tangential integration, leading to errors when calculating the sound field at close distances. To capture precise near-field geoacoustic features, we employ the wavenumber integration [11], [12] method to compute bottom reflection coefficient (BRC) at varying grazing angles. In shallow water environments, sound waves undergo multiple reflections off the seabed, carrying substantial information about the geoacoustic parameters of the seabed [2]. Thus, a certain relationship exists between the BRC sequences and the geoacoustic parameters of the seabed. From these BRC sequences derived from different geoacoustic parameters, we create the BRC sequence (BRCS) dataset, which is the foundation for training a DNN model.

Besides dataset construction, it is also crucial to choose an appropriate DNN model. For geoacoustic inversion problems, multilayer perceptron (MLP) is a common choice [5]. It updates model weight parameters using the backpropagation (BP) algorithm [13], enabling the DNN model to learn the representation of transforming acoustic signals into geoacoustic parameters. However, due to the relatively simple structure of MLP, there is potential to lose contextual information within the BRC sequence, diminishing the accuracy of geoacoustic inversion. Convolutional neural networks (CNNs) are also employed in geoacoustic inversion [6]. Their convolutional kernels and pooling layers can reduce data dimensionality and extract features from complex data such as images. However, as the BRC sequences in this study, their inherent data structure is relatively simple, making CNNs not ideal for processing them. Therefore, we introduce a novel geoacoustic inversion model named self-attention geoacoustics (SAG). SAG utilizes the transformer architecture [14], which has attracted a lot of attention in recent years for its excellent performance in various fields [15], [16], [17]. The multihead self-attention (MHSA) mechanism in SAG captures both short-term and long-term dependencies among the elements in the input sequence [14], enhancing the prediction accuracy of geoacoustic inversion.

BRCs exhibit significant variations in sensitivity to different geoacoustic parameters [18]. Through numerical simulations, we observed that the influence of sound attenuation coefficients on BRCs is relatively minimal, resulting in a noticeably lower prediction accuracy for sound attenuation than that for sound speed and density. Therefore, we adopted an adaptive-weight multitask learning (AW-MTL) strategy, which notably enhances the prediction accuracy for sound attenuation coefficients. The AW-MTL strategy treats the prediction of different geoacoustic parameters as distinct tasks, assigning unique weight coefficients to each task. During the training process, these adaptive weights, along with model weights, are optimized via the BP algorithm, dynamically adjusting the significance of individual tasks.

In this study, we propose a near-field geoacoustic inversion approach based on self-attention mechanism, which includes the following procedures. We first introduce the NBRC method, which acquires high signal-to-noise ratio bottom reflection data across a wide range of grazing angles through drifting methods. Then, based on the wavenumber integration method, we construct the BRCS dataset that accurately reflects near-field geoacoustic characteristics, supporting the training of DNN models. Since the inversion accuracy of conventional DNN models is limited, especially the precision of sound attenuation coefficients. Thus, we propose the SAG model based on self-attention mechanisms, capable of capturing global and local features within the BRC sequence, significantly improving the prediction accuracy of geoacoustic parameters. In addition, considering the sensitivity differences among various parameters, we introduce an AW-MTL strategy, significantly enhancing the prediction accuracy of sound attenuation coefficients.

The rest of this article is organized as follows. First, we introduced our NBRC method for collecting near-field bottom reflection data. Then, we elucidated the forward modeling method employed and the establishment of the BRCS dataset.

Following this, we provided a detailed description of the SAG model structure and the AW-MTL training strategy. In numerical experiments, we compared the performance of the SAG with MLP and MFI on the BRCS dataset and demonstrated the robustness of SAG against noisy data. Finally, we employed the SAG model to predict the geoacoustic parameters of the actual marine area and discussed some results from visualizing the self-attention weights of SAG.

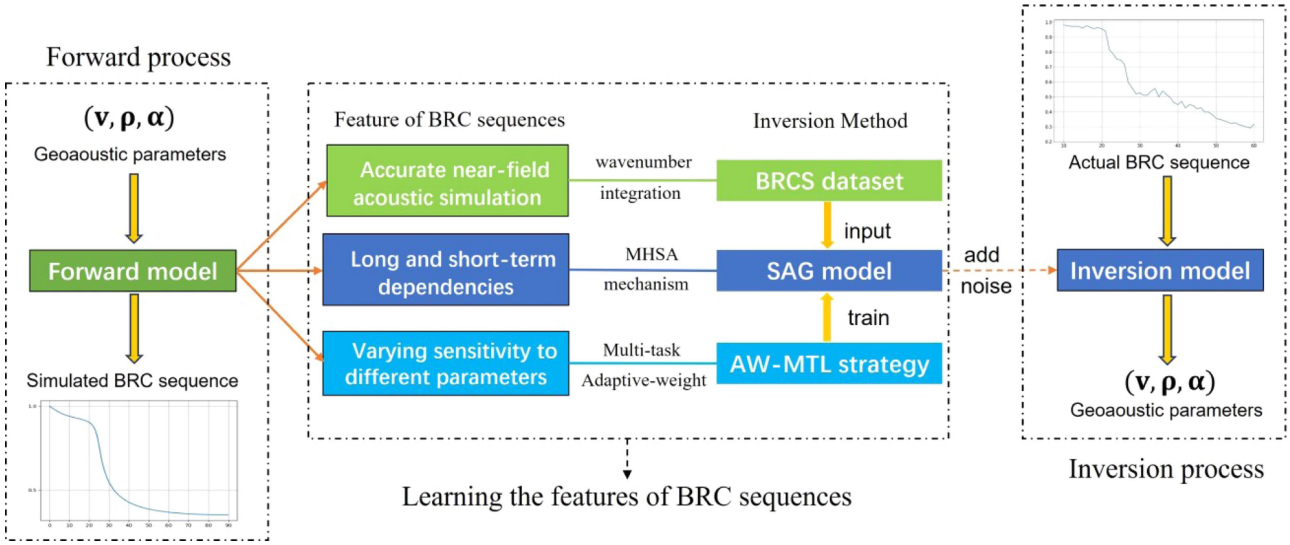

Fig. 1. Near-flied geoacoustic inversion using BRC sequences based on deep learning.

The visualization of self-attention weights demonstrates that our inversion method can capture long- and short-term dependencies in bottom reflection data, showcasing its advantages over traditional inversion methods and DNN models. Therefore, our approach offers the possibility of rapidly and accurately obtaining geoacoustic parameters over large areas based on near-field bottom reflection data.

II. Method

Geoacoustic inversion infers geoacoustic parameters from observed acoustic signals, essentially solving for the inverse function of the forward geoacoustic model. For deep learning-based geoacoustic inversion, the crux lies in iteratively optimizing the parameters of DNN models using the BP algorithm, eventually employing the DNN to substitute the actual geoacoustic model's inverse function, that is, the inversion model. Specifically, the objective is to identify an optimal inversion model that minimizes the discrepancy between the actual values and the predicted values of the geoacoustic parameters. The optimization of the DNN model parameters can be expressed as

$$\arg_{\boldsymbol{g}}\min \boldsymbol{J}\left(\mathrm{FM}\left(\mathbf{g},\mathbf{s};\mathbf{x}_{\boldsymbol{r}}\right),\mathbf{d}^{\mathrm{obs}}\left(\mathbf{x}_{\boldsymbol{r}}\right)\right) \tag{1}$$

where J is commonly referred to as the objective function or loss function, FM represents the forward modeling operator, $\mathbf{g}$ is the predicted value of geoacoustic parameters, $\mathbf{s}$ is the source function, $\mathbf{x}_r$ is the positions of the receivers, and $\mathbf{d}^{\mathrm{obs}}$ is the observed acoustic signals [19].

The essence of geoacoustic inversion based on deep learning is to train a DNN model to recognize geoacoustic features within acoustic signals. To achieve this, we need to accomplish three tasks. First, we need to collect actual bottom reflection signals that have a high signal-to-noise ratio and contain sufficient geoacoustic features. Second, we establish a dataset rich in geoacoustic features. Third, we identify an effective DNN model and train it to extensively learn these geoacoustic features. Fig. 1 illustrates our approach for near-field geoacoustic inversion based on deep learning.

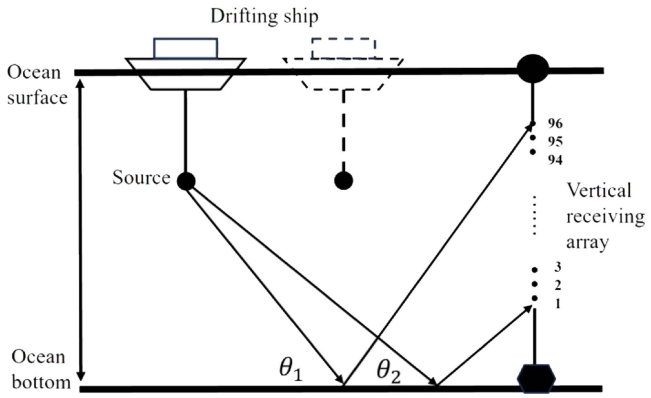

Fig. 2. Schematic diagram of NBRC.

Furthermore, the actual marine environment is complex and noisy. There is a notable difference between simulated and actual BRC sequences. Training the DNN model solely on simulated data results in errors in predicting actual geoacoustic parameters. Hence, we also conduct noise tests to assess the model's performance on noisy BRC data.

A. NBRC Method

Our near-field BRC acquisition method employs a vertical receiver array consisting of 96 units, along with a low-frequency sound source mounted beneath the transmission ship, as illustrated in Fig. 2, where $\boldsymbol{\theta}_1$ is the upper limit of grazing angles and $\boldsymbol{\theta}_1$ is the lower limit of grazing angles. The bottom end of the vertical receiving array is anchored to the seabed, whereas the top is subjected to upward pulling forces by floats, thereby maintaining an approximately vertical orientation. Prior to the collection of bottom reflection data, the engine is first started to

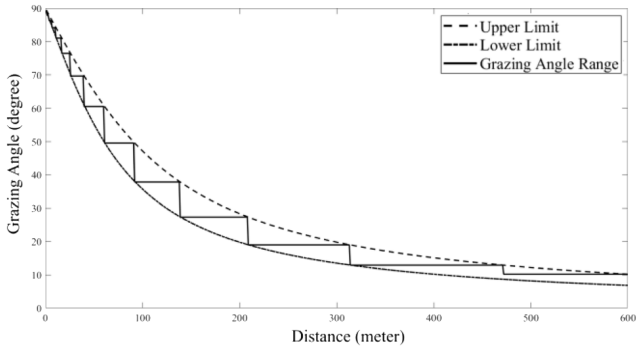

Fig. 3. Grazing angle range of the bottom reflected signal.

move the vessel near the top of the vertical receiving array. Then, the engine is shut down, and as the ship begins to drift, the sound source is activated, and the vertical receiving array collects the bottom reflection signals. Shutting OFF the engine avoids noise generation, thus enhancing the quality of the collected data. As the ship drifts, the horizontal distance between the sound source and the elements of the vertical receiving array changes, leading to variations in the grazing angle of the bottom reflection signals. For instance, at a water depth of 50 m, the range of grazing angles for the bottom reflection signals changes with the horizontal distance between the sound source and the receiving array, as shown in Fig. 3. Consequently, our data collection method allows for the simple and efficient observation of a wide range of grazing angles for bottom reflection signals, while also improving the signal-to-noise ratio.

An example of the raw acoustic signals collected during the marine experiment is shown in Fig. 4(a), which displays the multipath arrival peaks of a single pulse emitted by the sound source as received by receiver number 19 at a certain moment. Fig. 4(b) presents the actual signal after high-frequency noise has been removed via wavelet transformation.

According to the experimental records of the signals in Fig. 4, the depth of the sound source $\mathbf{S_d}$ was approximately 30 meters, the depth of receiver $\mathbf{R_d}$ was 32.5 meters, and the horizontal distance between the receiver and the sound source $\mathbf{R_r}$ was 23.5 meters. The multipath arrival structure can be computed and simulated using the ray-based Bellhop program [20], as shown in Fig. 5. By comparing Figs. 4 and 5, the different arrival paths in the actual acoustic signal can be distinctly identified, allowing for the extraction of the amplitudes of the direct and bottom-reflected waves.

The actual BRCs were calculated as follows:

$$R(\boldsymbol{\theta}) = \frac{\mathbf{P_r D_r}}{\mathbf{P_d D_d}} \tag{2}$$

where $\mathbf{P_r}$ is the intensity of the bottom reflection wave, $\mathbf{P_d}$ is the intensity of the direct wave, $\mathbf{D_r}$ represents the path length of the bottom reflection wave, and $\mathbf{D_d}$ represents the path length of the direct wave. $\mathbf{P_r}$ and $\mathbf{P_d}$ can be represented by the amplitudes of the actual signal, whereas $\mathbf{D_r}$, $\mathbf{D_d}$ and the grazing angle $\boldsymbol{\theta}$ can be calculated using simple geometric relationships. Based on the

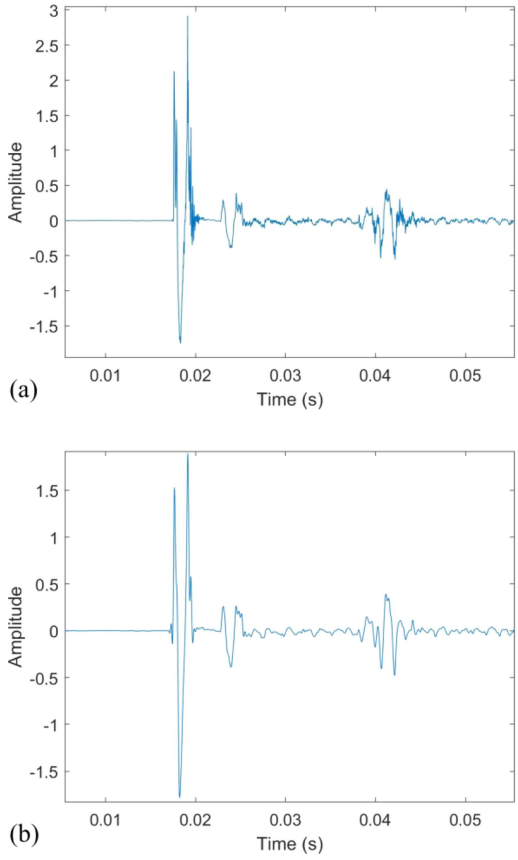

Fig. 4. Example of actual acoustic signals collected in marine experiments. (a) Raw signal. (b) Low-pass-filtered signal by wavelet transform.

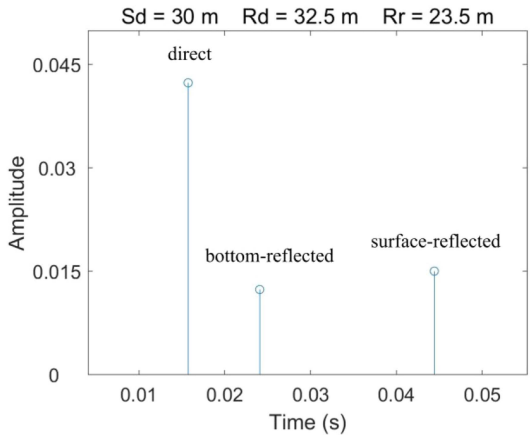

Fig. 5. Multiway arrival structure simulated by ray model. The horizontal axis represents the arrival time of sound waves in each path. The vertical axis represents the relative amplitude calculated by Bellhop.

BRCs at different grazing angles, we obtained the actual bottom reflection curve of the experimental area, as shown in Fig. 6.

B. BRCS Dataset

The wavenumber integration method is a numerical computation approach designed for horizontally layered media models,

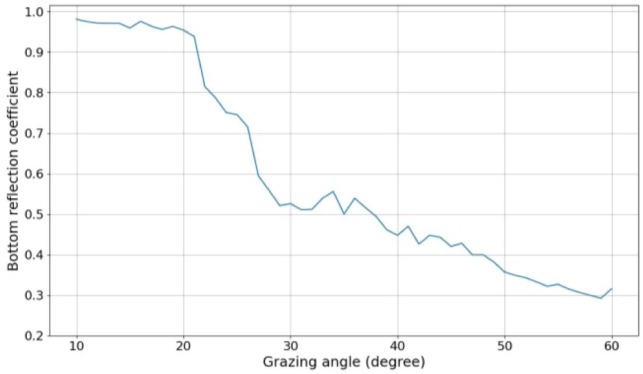

Fig. 6. Actual BRC curve of a certain test area.

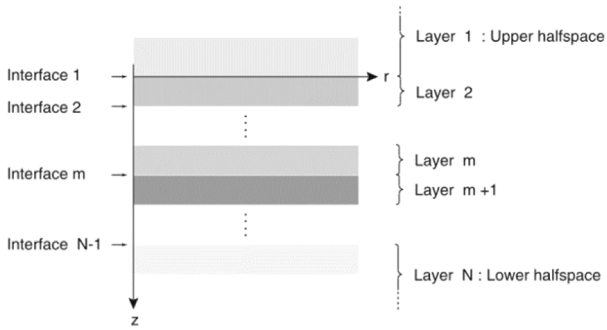

Fig. 7. Horizontally layered media model.

applicable in environments that are invariant with distance, with layer properties solely dependent on depth, as depicted in Fig. 7.

In cylindrical coordinates, if horizontal layering is disregarded, the sound field satisfies the Helmholtz equation

$$\left[\nabla^2 + k_m^2\left(\mathbf{z}\right)\right]\psi_m\left(\mathbf{r},\mathbf{z}\right) = f_s\left(\mathbf{z},\omega\right)\frac{\delta\left(\mathbf{r}\right)}{2\pi\mathbf{r}} \tag{3}$$

where $f_s(\mathbf{z},\omega)$ represents the source function, and $k_m(\mathbf{z})$ denotes the wavenumber of the medium in the mth layer

$$k_m\left(\mathbf{z}\right) = \frac{\omega}{c\left(\mathbf{z}\right)}. \tag{4}$$

For layers devoid of sound sources, $f_s(\mathbf{z},\omega)$ equals zero. The essence of the wavenumber integration method is to apply the Hankel transform to (3), transforming the original four-dimensional partial differential equation into a series of ordinary differential equations in terms of depth coordinates

$$\left[\frac{d^2}{d\mathbf{z}^2} - \left[k_r^2 - k_m^2\left(\mathbf{z}\right)\right]\right]\psi_m\left(k_r,\mathbf{z}\right) = \frac{f_s\left(\mathbf{z}\right)}{2\pi} \tag{5}$$

where the horizontal wavenumber k_r represents the component of the wavenumber $k_m(\mathbf{z})$ in the horizontal direction within the mth layer. Equation (5) is then analytically solved within each layer, followed by the computation of the inverse integral transform to determine displacements and stresses.

Unlike the normal mode method [3] typically employed in forward modeling, the wavenumber integration method directly performs numerical integration to solve for the spectral integral.

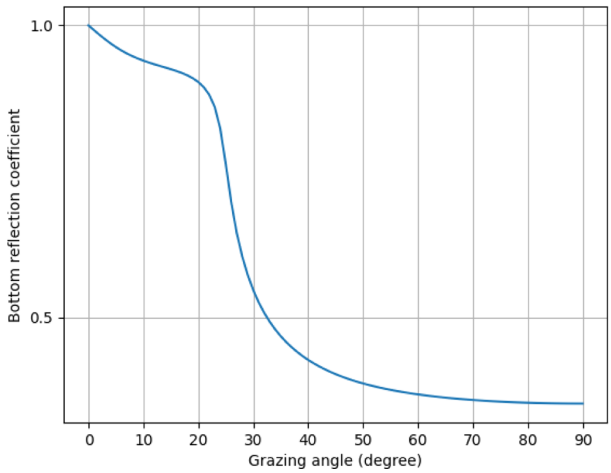

Fig. 8. BRC curve in numerical simulation.

This approach has the advantage of not neglecting the influence of branch cut integrals, thereby yielding more accurate calculations of the near-field sound field compared to the normal mode method.

In this study, we treat sound waves as plane waves and obtain equations related to depth for several discrete horizontal wavenumbers at selected receiver depths. Numerical calculations are then performed on the sound field integral expressions to derive the relationships between sound field, distance, and depth. For sound waves incident at different grazing angles $\boldsymbol{\theta}$, the complex amplitude of the reflected waves can be calculated, yielding the BRC for that grazing angle as

$$R\left(\boldsymbol{\theta}\right) = \frac{A^-\left(k_m\cos\boldsymbol{\theta}\right)}{A^+\left(k_m\cos\boldsymbol{\theta}\right)} \tag{6}$$

where $A^+(k_m\cos\boldsymbol{\theta})$ and $A^-(k_m\cos\boldsymbol{\theta})$ respectively represent the complex amplitudes of the incident and reflected waves at the interface of the mth layer.

In actual marine experiments, it is typically possible to acquire BRCs only within a certain range of grazing angles $[\theta_1, \theta_2]$. By uniformly selecting several angles within $[\theta_1, \theta_2]$, and subsequently arranging the reflection coefficients of all angles in order, a BRC sequence is obtained. For each combination of geoacoustic parameters including sound speed($\mathbf{v}$), density($\boldsymbol{\rho}$), and attenuation coefficient ($\boldsymbol{\alpha}$), the corresponding BRC sequences can be calculated, with Fig. 8 displaying a BRC curve obtained through numerical simulation.

Deep learning-based geoacoustic inversion essentially involves teaching the model the relationship between input vector $X(x_1, x_2, \ldots, x_i)$ and output vector $Y(y_1, y_2, \ldots, y_j)$. Consequently, we treat BRC sequences as the *X* vector and geoacoustic parameters as the *Y* vector. The dimensions of *X* and the parameter range of *Y* are then determined based on the needs of the specific problem. All corresponding [*X*, *Y*] pairs together constitute the BRCS dataset.

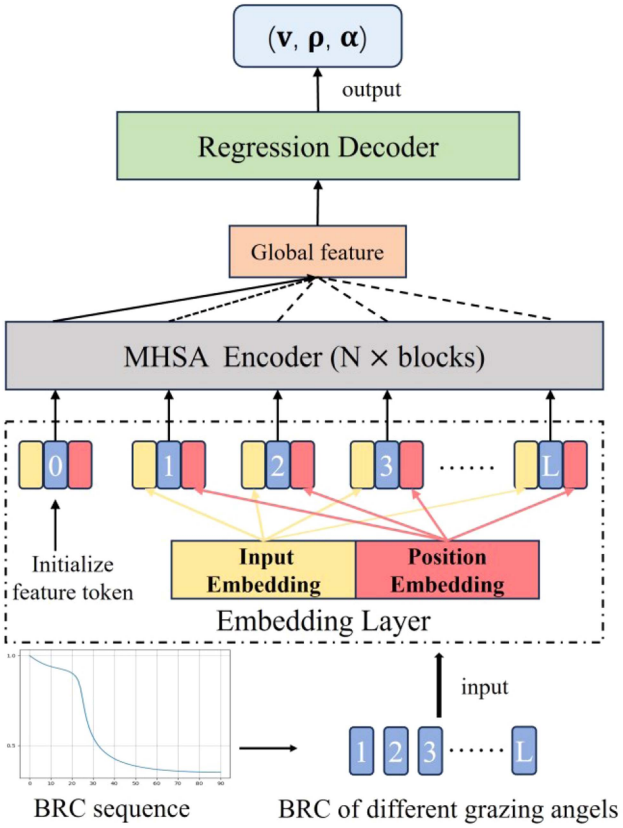

Fig. 9. Architecture of SAG.

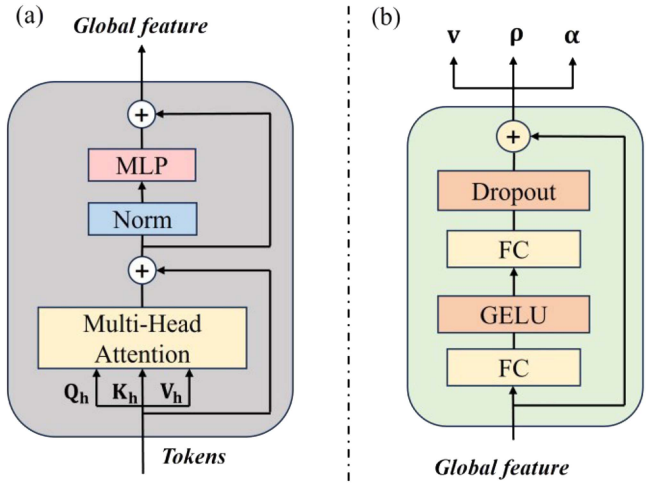

Fig. 10. (a) MHSA encoder block structure. (b) Regression decoder structure.

C. SAG Model Architecture

Our proposed DNN model SAG consists of three main components: the embedding layer, MHSA encoder, and regression decoder, as illustrated in Fig. 9. The embedding layer is designed to map each element of the BRC sequence to a token vector with a dimension of $\dim_E$. The MHSA encoder computes attention scores between all token vectors, capturing a global feature vector that represents the entire sequence. This feature vector is then fed into the regression decoder, which maps the global feature to the space of geoacoustic parameters. Finally, the output of the regression decoder is the predicted values of the three geoacoustic parameters.

In the embedding layer, a fully connected (FC) layer maps each element of the input BRC sequence to an embedding vector of length $\dim_E$, a process referred to as input embedding. Given that the MHSA mechanism inherently does not consider the order within the sequence, it is imperative to incorporate position embeddings within the embedding layer. This enables the model to capture the positional information of elements in the BRC sequence, which corresponds to the grazing angle of each reflection coefficient. The position embeddings are generated using a sine and cosine embedding function [14]

$$\mathrm{PE}\ (\mathbf{pos}, 2\mathbf{i}) = \sin\left(\frac{\mathbf{pos}}{10000^{\frac{2\mathbf{i}}{\dim_E}}}\right) \tag{7}$$

$$\mathrm{PE}\ (\mathbf{pos}, 2\mathbf{i}+1) = \cos\left(\frac{\mathbf{pos}}{10000^{\frac{2\mathbf{i}}{\dim_E}}}\right) \tag{8}$$

where $\mathbf{pos}$ is the position of a specific element within the BRC sequence and $\mathbf{i}$ is the dimension of the embedding vector. Each element thus obtains a position vector of length $\dim_E$. For every element in the BRC sequence, its embedding vector is summed with its position vector, resulting in a token vector of length $\dim_E$. Through this process, each element of the BRC sequence is transformed into a token vector, encapsulating information about both the reflection coefficient and the grazing angle. The original BRC sequence of dimension L, upon processing through the embedding layer, yields L token vectors each of $\dim_E$ dimensions. We initialize a feature token vector of the same $\dim_E$ dimensions randomly, analogous to the CLS token in Bert [16].

The MHSA encoder is comprised of multiple identical MHSA encoder blocks stacked together. The structure of each MHSA encoder block is depicted in Fig. 10(a). The MHSA mechanism multiplies all token vectors by the learnable transformation matrices $\mathbf{W_Q}$, $\mathbf{W_K}$, and $\mathbf{W_V}$ to produce the $\mathbf{Q}$ (query), $\mathbf{K}$ (key), and $\mathbf{V}$ (value) matrices, respectively. The attention weights are then computed using the Softmax function, and this calculation process can be described as

$$\mathrm{Attention}\ (\mathbf{Q_h}, \mathbf{K_h}, \mathbf{V_h}) = \mathrm{Softmax}\left(\frac{\mathbf{Q_h}\mathbf{K_h^T}}{\sqrt{\dim_E}}\right)\mathbf{V} \tag{9}$$

where variable $\mathbf{h}$ denotes the attention head, which serves to divide the token vector into different parts, each undergoing a distinct transformation with its own set of $\mathbf{W_Q^h}$, $\mathbf{W_K^h}$, and $\mathbf{W_V^h}$ matrices. This arrangement enables the application of diverse linear transformations to the token vectors, yielding richer representations. To prevent issues of gradient vanishing and exploding gradients, residual connections are employed within the MHSA encoder blocks. In addition, dropout layers are utilized in the MLP to mitigate the risk of overfitting.

After processing through multiple MHSA encoder blocks, all token vectors encapsulate rich dependency relations with

other tokens. The feature token, having computed attention weights with every element of the input sequence, incorporates information from all elements of the original BRC sequence, as well as their interdependencies. Consequently, we use the feature token as the global feature vector of the input BRC sequence for subsequent geoacoustic parameter prediction.

The regression decoder is tasked with learning the nonlinear mapping relationship between the feature token vector and the three geoacoustic parameters. Its architecture comprises two FC layers, an activation function layer, and a dropout layer, as shown in Fig. 10(b). We utilize the Gaussian error linear unit [21] as the activation function to enhance the model's nonlinear representational capacity and employ dropout to prevent overfitting. The input to the regression decoder is the global feature vector, and its output is the predicted values for the corresponding geoacoustic parameters.

D. AW-MTL Strategy

In addition to the dataset and model, setting an appropriate training objective, often called a loss function, is also crucial. Geoacoustic inversion requires the model to predict multiple parameters simultaneously. During the training process, it is common practice to add the errors of different parameters together to form the total loss, with mean squared error (MSE), root mean squared error, and mean absolute error (MAE) being commonly used loss functions. In this study, we adopt MSE to calculate the loss as

$$\text{Loss}_{\text{MSE}} = \frac{1}{k}\sum_{i=1}^{k}\left(\mathbf{y}^{\mathbf{i}}_{\mathbf{pred}} - y^{i}_{\text{true}}\right)^2 \tag{10}$$

where k is the number of geoacoustic parameters, $\mathbf{y}^{\mathbf{i}}_{\mathbf{pred}}$ is the predict value of the ith geoacoustic parameter, and y^{i}_{true} is the true value of the ith geoacoustic parameter.

However, different geoacoustic parameters usually exhibit varying influence on the same acoustic signal, resulting in noticeable discrepancies in the model's predictive accuracy for different parameters. To address this variability, we treat the prediction of multiple parameters as a multitask learning problem, considering the prediction of each parameter as a subtask. By introducing a weight coefficient for each subtask, the model's gradient updates will be more inclined towards tasks with larger weights [22]. The weight coefficients for all tasks can be set as fixed values or variable parameters

$$\text{Loss}_{\text{vari}} = \frac{1}{k}\sum_{i=1}^{k}\mathbf{w}_i\left(\mathbf{y}^{\mathbf{i}}_{\mathbf{pred}} - y^{i}_{\text{true}}\right)^2 + \lambda \cdot R(\mathbf{w}) \tag{11}$$

where $\mathbf{w}_i$ is the variable weight of the kth geoacoustic parameter and $R(\mathbf{w})$ serves as a penalty term for the task weights, utilized for regularization purposes to prevent the task weights from diminishing too small. λ is the regularization parameter, dictating the strength of regularization. The task weights $\mathbf{w}_i$ in (11) are learnable parameters, capable of evolving throughout the training process.

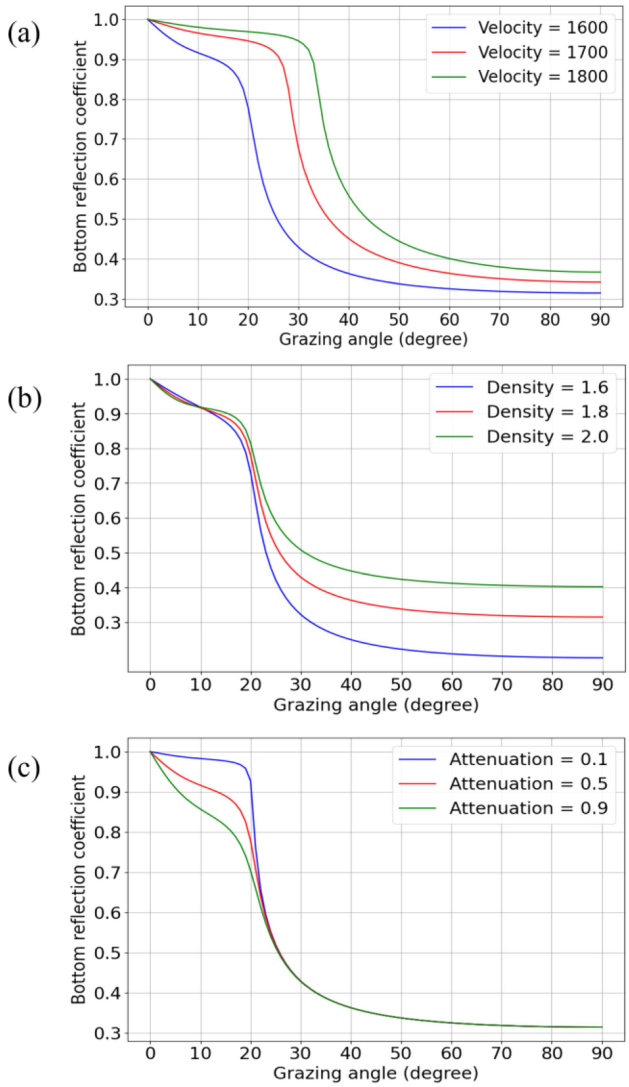

Fig. 11. Discrepancies exist in the sensitivity of different geoacoustic parameters to the simulated BRC curve. (a) $\rho = 1.8$ g/cm^2 and $\alpha = 0.5$ dB/λ. (b) $v = 1600$ m/s and $\alpha = 0.5$ dB/λ. (c) $v = 1600$ m/s and $\rho = 1.8$ g/cm^2.

In our study, numerical simulations have revealed that sound attenuation has a noticeably weaker influence on BRCs compared to sound speed and density, as depicted in Fig. 11. This disparity between parameters poses challenges for predicting sound attenuation. To significantly enhance the model's focus on sound attenuation during the training process, we propose a multitask loss function incorporating adaptive weights

$$\text{Loss}_{\text{AW-MTL}} = \sum_{i=1}^{k}\boldsymbol{\mu}_{\mathbf{i}}^{\boldsymbol{\eta}_{\mathbf{i}}}\left(\mathbf{y}^{\mathbf{i}}_{\mathbf{pred}} - y^{i}_{\text{true}}\right)^2 + \lambda \cdot R(\boldsymbol{\mu}, \boldsymbol{\eta}) \tag{12}$$

where the uncertainty parameter $\boldsymbol{\mu}$ and the exponent parameter $\boldsymbol{\eta}$ are adaptive parameters that can be dynamically adjusted, and the equation also includes a regularization term. Throughout the training process, all adaptive parameters are updated iteratively alongside the model parameters.

To ensure the stability of the training process, we have also employed the Softplus activation function to constrain the values

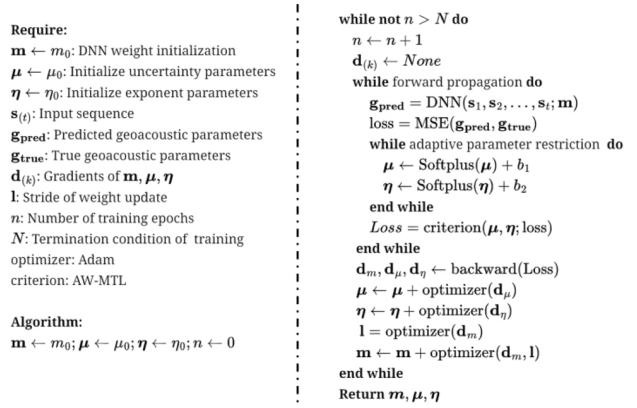

Fig. 12. Pseudo-code of the AW-MTL training strategy.

of the adaptive parameters as

$$\begin{cases} \boldsymbol{\mu}_{\mathbf{out}} = Softplus\left(\boldsymbol{\mu}_{\mathbf{in}}\right) + b1 \\ \boldsymbol{\eta}_{\mathbf{out}} = Softplus\left(\boldsymbol{\eta}_{\mathbf{in}}\right) + b2 \\ Softplus\ \left(\mathbf{x}_{\mathbf{out}}\right) = \log\left(1 + e^{\mathbf{x}_{\mathbf{in}}}\right) \end{cases} \tag{13}$$

where $\boldsymbol{\mu}_{\mathbf{in}}$ and $\boldsymbol{\eta}_{\mathbf{in}}$ are the original adaptive parameters, $\boldsymbol{\mu}_{\mathbf{out}}$ and $\boldsymbol{\eta}_{\mathbf{out}}$ are the activated adaptive parameters, and $b1$ and $b2$ are bias terms. The Softplus function yields outputs close to zero for negative inputs, providing a smoother transition compared to the commonly used ReLU activation function. By applying (13), we guarantee that the values of the adaptive parameters $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$ remain consistently greater than b_1 and b_2. This ensures that the optimization of the adaptive parameters stays within a controllable range during the training process. We adopt an adaptive moment estimation (Adam) optimizer to update both model weights and adaptive weights. Fig. 12 illustrates the pseudo-code for our AW-MTL training strategy.

III. Numerical Examples

In this study, actual near-flied bottom reflection data was collected in the South Yellow Sea. The test area is located roughly 210 km southeast of Qingdao, in the central part of the Yellow Sea continental shelf. The seabed of our data collection site is composed primarily of clayey sand. The water depth in the experimental area ranges from approximately 40–50 m, and the seabed terrain is relatively flat.

In our numerical examples, the BRC sequences in the BRCS dataset also encompass the BRCs for 51 grazing angles ranging from 10° to 60°, consistent with the actual measured data. The source frequency is set to 500 Hz, with the sound speed and seawater density in the water being 1500 m/s and 1000 kg/cm^3, respectively. Subsequently, we determined the approximate range of geoacoustic parameters within the BRCS dataset by referring to the results of MFI and experimental measurement data from the same sea area [23]. In addition, based on empirical formulas, a positive correlation exists among the three geoacoustic parameters [24].

To ensure that the training data reflect this pattern, we divided the BRCS dataset into three parts: high sound speed zone, medium sound speed zone, and low sound speed zone, as shown in Table I. Within each zone, we uniformly sampled the geoacoustic parameters and calculated the corresponding BRC sequences using the wavenumber integration method, followed by the removal of some duplicate data. In this manner, we obtained a total of 24 400 samples. Each sample [X, Y] includes a BRC sequence $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{51})$ and the corresponding three geoacoustic parameters $(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3)$. All of the [X, Y] samples constitute the BRCS dataset used in this example. The numerical experiments in this study were conducted under the same hardware and software configurations, as illustrated in Table II.

TABLE I
BRCS Dataset Settings

Zone	Sound speed(m/s)	Density(g/cm^3)	Attenuation coefficient(dB/λ)
A	1400–1600	1.7–1.9	0.3–0.7
B	1500–1700	1.8–2.0	0.6–0.8
C	1600–1800	1.9–2.1	0.5–0.9

TABLE II
Numerical Experiments Configurations

CPU	Intel i9-13900K
GPU	Nvidia RTX4090
Python version	3.7
Cuda version	11.7
Pytorch version	1.13

A. Inversion Results of SAG

We utilize the AW-MTL strategy to train SAG on the BRCS dataset. Initially, all the geoacoustic parameters in the dataset were subjected to min-max normalization, ensuring that their values ranged between 0 and 1. This normalization process aimed to prevent the disparities in the numerical values of geoacoustic parameters from impacting the training process. During the training process, we observed that a larger batch size could improve hardware utilization efficiency, but it makes gradient updates more challenging, leading to a slower model convergence rate. Conversely, a smaller batch size facilitated quicker gradient updates but failed to fully utilize the computational resources. After testing, we set the batch size to 300 and the number of training epochs to 500. The initial learning rate is set at 0.001, and the Adam optimizer is used for optimization.

The BRCS dataset is randomly divided into training, validation, and test sets in certain proportions. In each epoch, the model is trained on the training set, followed by validation on the validation set, where the validation loss is calculated. After training, the model weight corresponding to the minimum validation loss is selected as the final model weight, and the performance of the final model is subsequently tested on the test set. This approach helps to prevent overfitting and enhances the model's generalization ability.

We test the performance of SAG models with varying numbers of MHSA blocks (N), MSHA heads (H), and embedding vector dimensions ($\dim_E$) across the entire test set. Experimental

TABLE III
INVERSION PERFORMANCE OF SAG VARYING WITH THE STRUCTURAL HYPERPARAMETERS

(a)

Number of MHSA blocks N	MAE_ v(m/s)	MAE_ ρ(g/cm^3)	MAE_ α(dB/λ)
3	1.805	0.003	0.018
4	1.738	0.004	0.019
5	1.439	0.003	0.008
6	2.031	0.005	0.022

(b)

Number of MHSA heads H	MAE_ v(m/s)	MAE_ ρ(g/cm^3)	MAE_ α(dB/λ)
6	2.087	0.008	0.027
8	1.862	0.004	0.012
12	1.439	0.003	0.008
16	1.482	0.003	0.011

(c)

Dimension of MHSA embedding vector dim_E	MAE_ v(m/s)	MAE_ ρ(g/cm^3)	MAE_ α(dB/λ)
48	3.574	0.008	0.039
96	1.817	0.004	0.025
192	1.439	0.003	0.008
288	1.956	0.005	0.014

results reveal that the structural hyperparameters of the SAG model significantly influence the inversion accuracy. The MAE between the predicted and true geoacoustic parameters for all samples of SAG with different structural hyperparameters is presented in Table III. The inversion accuracy first increased and then decreased with the rise of N, peaking at $N = 5$. An increase in H improved the model's prediction accuracy for different geoacoustic parameters. However, when H increased from 12 to 16, the model's performance barely improved and even slightly declined, indicating that more MSHA heads do not always result in a better model. In terms of $\dim_E$, the inversion accuracy was highest when it was set to 192; both smaller and larger $\dim_E$ values were detrimental to the model's performance.

B. Comparison of Inversion Models

Unlike DNN models, MFI requires the constant invocation of the forward model to assess the fitness of the inversion result, performing the inversion on one BRC sequence at a time. In this study, we adopt the genetic algorithm (GA) for MFI. GA is a search heuristic that simulates the process of natural selection. Drawing inspiration from the mechanics of natural genetics and Darwin's theory of evolution, the algorithm is designed to find solutions to optimization and search problems. Traditional optimization techniques can get trapped in local minima, but GAs, with their inherent randomness, can search the space more thoroughly and find global or near-global optima [23].

TABLE IV
COMPARISON OF DIFFERENT INVERSION MODELS

Inversion models	ARE_ v (%)	ARE_ ρ (%)	ARE_ α (%)
SAG	0.10	0.14	1.53
MLP	0.23	0.42	9.24
MFI	1.78	2.62	6.38

TABLE V
ADAPTIVE PARAMETERS INITIALIZATION

Geoacoustic parameters	Sound speed	Density	Sound attenuation	Bias
$\boldsymbol{\mu}$	0.5	0.5	1	1
$\boldsymbol{\eta}$	0.5	0.5	1	0.1

TABLE VI
COMPARISON OF DIFFERENT TRAINING STRATEGY

Training strategy	MAE_ v(m/s)	MAE_ ρ(g/cm^3)	MAE_ α(dB/λ)
AW-MTL	1.439	0.003	0.008
MSE	2.461	0.007	0.092
VMSE	2.593	0.005	0.034

To compare the inversion accuracy of different inversion models, we randomly selected a subset of samples from the test set and conducted inversions using MFI, MLP, and SAG. Subsequently, we calculated the average relative error (ARE) for each parameter, with the results presented in Table IV. The results clearly demonstrate that the inversions from DNN models outperform those from MFI. Specifically, our proposed SAG model exhibits higher inversion accuracy than MLP, showcasing the superiority of the MSHA mechanism.

C. Superiority of AW-MTL Strategy

We employ the AW-MTL strategy to train the SAG model, aiming to balance the variability in prediction difficulty across different parameters. Before the commencement of training, the uncertainty parameter $\boldsymbol{\mu}$, exponent parameter $\boldsymbol{\eta}$, and their bias terms in (13) are initialized as shown in Table V. During the training process, the Adam optimizer is utilized to simultaneously optimize the adaptive parameters $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$, ensuring that the value of $\boldsymbol{\mu} > 1$, and the value of $\boldsymbol{\eta} > 0.1$.

In the early stages of training, the prediction accuracy for sound speed and density rapidly improved, leading to a decrease in their adaptive parameters. Consequently, the direction of the total loss gradient descent gradually shifted toward minimizing the error in sound attenuation. We trained the same SAG model using both the AW-MTL strategy, direct MSE loss, and the variable MSE loss (VMSE) in (11), the inversion results are presented in Table VI. The AW-MTL strategy, with its adaptive parameters dynamically adjusting the weights of different tasks,

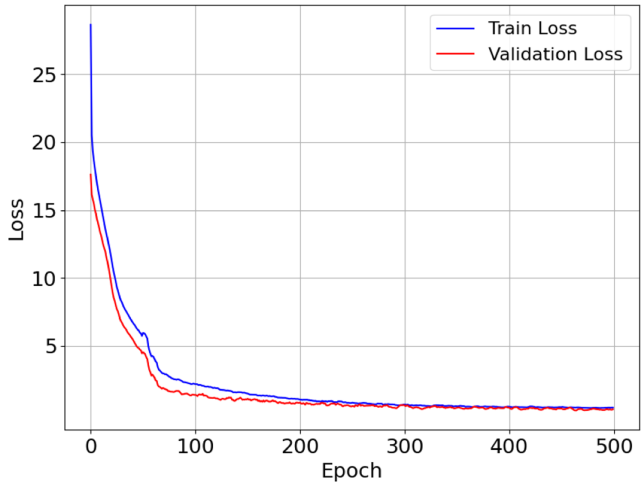

Fig. 13. Loss curve of SAG trained with AW-MTL strategy.

made the training smoother and significantly enhanced the inversion accuracy for sound attenuation. We showcased the loss curve of the model trained with the AW-MTL strategy when $N = 5$, $H = 12$, and $di\ m_E = 192$ in Fig. 13.

D. Robustness Against Noisy Data

The marine environment is replete with substantial noise. Even with denoising procedures applied to the original observational data, it is impossible to fully eliminate the effects of noise. Thus, it is necessary to introduce noise into the dataset used for training the SAG model to examine the model's robustness to noisy data. The most natural approach would be to inject noise directly into the raw acoustic signals collected by the receivers, as this carries clear physical implications. However, training DNN models requires extensive datasets; in this study, we utilized a dataset of 24 400 samples, each containing an independent BRC curve. Given the limitations of data collection in marine experiments and the data processing workload involved in converting raw acoustic signals to BRC curves, it is quite challenging to directly add noise to the raw acoustic data in this research. In Section II-B, we introduced the wavenumber integration method used for forward modeling, which can rapidly compute simulated BRC sequences for a set marine bottom environment, making the construction of a BRCS dataset feasible. Therefore, we opted to introduce Gaussian noise at varying signal-to-noise ratio (SNR) into the BRC sequences within the BRCS dataset. We regard reflection coefficients as the signal, and the SNR can be represented as follows:

$$\mathrm{SNR} = 10\log_{10}\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}} = 10\log_{10}\frac{\frac{1}{L}\sum_{i=1}^{L}R(i)^2}{P_{\mathrm{noise}}} \quad (14)$$

where P_{signal} and P_{noise} respectively denote the average power of the signal and noise. Here, $R(i)$ represents the ith element in the BRC sequence, and L is the length of the sequence. Therefore, the introduction of noise into the $\mathbf{n}$th element $R(\mathbf{n})$ of the BRC sequence can be expressed as

$$R_{\mathrm{noise}}(\mathbf{n}) = R(\mathbf{n}) + \mathbf{Noise}\cdot\sqrt{P_{\mathrm{noise}}} = R(\mathbf{n}) + \mathbf{Noise}\cdot\sqrt{\frac{\frac{1}{L}\sum_{i=1}^{L}R(i)^2}{10^{\frac{\mathrm{SNR}}{10}}}}. \quad (15)$$

$R_{\mathrm{noise}}(\mathbf{n})$ is the element after adding noise, $\mathbf{Noise}$ is a random variable with $\mathbf{Noise} \sim \mathcal{N}(0,1)$. The BRC sequence after adding noise was used for training and testing of the SAG model. We demonstrated the MAE of the SAG model's inversion results under various SNR conditions in Table VII. It can be observed that even under low to medium SNR conditions, the results from the SAG model remain relatively accurate. This underscores the robustness of our model, highlighting its applicability to real BRC sequences containing a certain level of noise.

TABLE VII
SAG PERFORMANCE ON NOISY DATA WITH DIFFERENT SNR

SNR	MAE_ v(m/s)	MAE_ ρ(g/cm^3)	MAE_ α(dB/λ)
20	4.080	0.023	0.068
10	10.07	0.053	0.106
5	11.48	0.059	0.127
0	25.46	0.081	0.136

TABLE VIII
ACTUAL INVERSION RESULTS OF DIFFERENT SAG MODELS

SNR (dB)	Sound speed (m/s)	Density (g/cm^3)	Sound attenuation (dB/λ)
20	1638	1.910	0.473
10	1654	1.936	0.685
5	1643	1.923	0.593

E. Actual Inversion Results

In this study, although the actual reflection data has been stripped of high-frequency noise, it may still contain a small amount of low-frequency noise. Therefore, we conducted inversions using models with SNR = 20, and SNR = 10, SNR = 5, respectively, and the prediction results are displayed in Table VIII.

It can be observed that the results for sound speed and density are relatively close, while there is a slightly larger difference in sound attenuation. Referring to the experimental measurement and MFI results in the same sea area [23], the SAG model provides fairly accurate results for sound speed and density. In addition, considering that in the actual marine environment, a difference of about 0.1 in sound attenuation has a negligible impact on the BRC. Therefore, our proposed SAG model performs excellently not only on simulated datasets but is also effective for geoacoustic inversion of actual measurement data. In Fig. 14, we present the BRC curves calculated using the inversion results and compare them to the actual BRC curves.

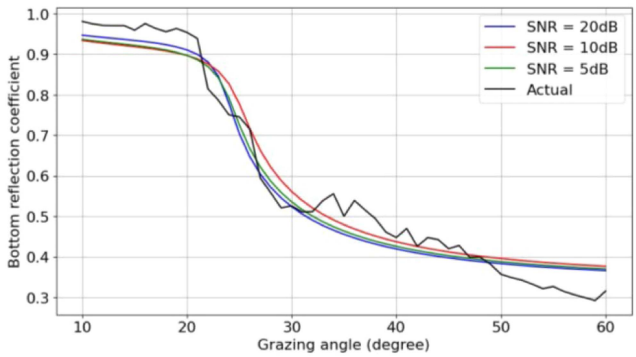

Fig. 14. BRC curve of inversion results by SAG trained with different SNR.

IV. Discussion

A. Performance of SAG and AW-MTL

Through numerical experiments, we have confirmed the superiority of the SAG model over MLP. In addition, the AW-MTL strategy facilitates the model in dynamically adjusting the importance of different tasks, notably enhancing the prediction accuracy for sound attenuation compared to other loss functions. We will now delve into the reasons for the excellent performance of SAG and AW-MTL in near-field geoacoustic inversion using BRC sequences.

During geoacoustic inversion, the MLP model directly learns the mapping relationship between BRC sequences and geoacoustic parameters through its multilayered network structure. In contrast, the SAG model initially utilizes the embedding layer and MHSA encoder depicted in Fig. 9 to extract the global feature representing the original BRC, subsequently learning the relationship between this global feature and the geoacoustic parameters. This distinction represents the core difference between MLP and SAG. Initially, the SAG model transforms each element in the original BRC sequence into a token vector of length $\dim_{\mathrm{E}}$ through the embedding layer. This process elevates the data's dimensionality, granting the model access to a richer set of information. We then input the sequence of all token vectors into the MHSA encoder. Unlike the convolution window in CNN, the MHSA mechanism can compute relationships between any two elements in a sequence while simultaneously extracting global and local features.

After processing through the MHSA encoder, the feature token encapsulates information from all elements in the original BRC sequence and their interrelationships. The commendable performance of SAG validates that performing geoacoustic inversion using global features extracted from the original BRC sequences through the MHSA mechanism can significantly improve prediction accuracy. This outcome underscores the efficacy of the MHSA mechanism in extracting sequential features.

Compared to MSE and VMSE loss functions, our proposed AW-MTL views the prediction of each geoacoustic parameter as a distinct task, flexibly adjusting weights for different parameters during training. As depicted in Fig. 12, the AW-MTL strategy considers both adaptive parameters and DNN model parameters as aspects of optimization updates, distinguishing it entirely from common loss functions. For the uncertainty parameter $\boldsymbol{\mu}$ and the exponent parameter $\boldsymbol{\eta}$ of sound attenuation, we assign relatively large initial values. At the beginning of training, the model is more likely to focus on the errors of sound speed and density, rapidly improving the predictive accuracy for these two parameters. As training progresses, the direction of gradient updates gradually shifts towards minimizing the regularization term for sound attenuation, resulting in an increased weight for sound attenuation and thereby directing more of the model's attention to minimizing its error. By adopting the AW-MTL strategy, the model can focus on different tasks at various stages of training, ultimately achieving high accuracy for all tasks. In addition, the use of the uncertainty parameter $\boldsymbol{\mu}$ and the exponent parameter $\boldsymbol{\eta}$ allows for more flexibility in the direction of gradient updates, smoothing the training process.

B. Analysis of Self-Attention Weights

The core of MSHA is to calculate the self-attention weights of the input sequence, that is, the $\mathrm{Attention}(\mathbf{Q_h}, \mathbf{K_h}, \mathbf{V_h})$ in (9). The self-attention weights of different elements in the sequence reflect the interdependencies among elements within the sequence. During training, different Heads in MSHA have different $\mathbf{W_Q}$, $\mathbf{W_K}$, and $\mathbf{W_V}$ transformation representations, resulting in different self-attention weights. Concatenating the outputs of each head together yields the output of the MSHA block, allowing the model to learn richer attention representations.

We input actual BRC sequences into SAG and showcase the self-attention weight maps of the SAG model after Softmax function processing in Fig. 15. Each pixel represents the attention weight of one element in the sequence to another, with larger values indicating more concentrated attention. In SAG, each MSHA block at different layers generated different attention weights, indicating that different MSHA blocks in the model focused on different interdependencies among elements, extracting both global and local features from the sequence. Specifically, the fifth MSHA block is the last layer of the MSHA encoder, and its 0th row corresponds to the visualization of the feature token vector in Section II-B. It can be seen that as the global feature vector of SAG, its attention weights are mainly distributed over the 0th–20th tokens, suggesting that this part of the BRC sequence is especially important for geoacoustic inversion. This portion of the sequence roughly corresponds to the BRCs at grazing angles of 10°–30°. Observing the simulated BRC curve in Fig. 8 and the actual BRC curve in Fig. 6, the BRC curve changes most dramatically around grazing angles of 10°–30°. The results of the BRC curve and attention visualization corroborate each other. Therefore, we believe that for near-field geoacoustic inversion in this study, the BRCs at grazing angles of 10°–30° are relatively important. Similarly, this also demonstrates MSHA's ability to identify the most critical parts of the input sequence for the task.

In addition, each MHSA block may represent a different level of processing strategy within the model. For instance, the initial blocks may focus more on the local features of the sequence,

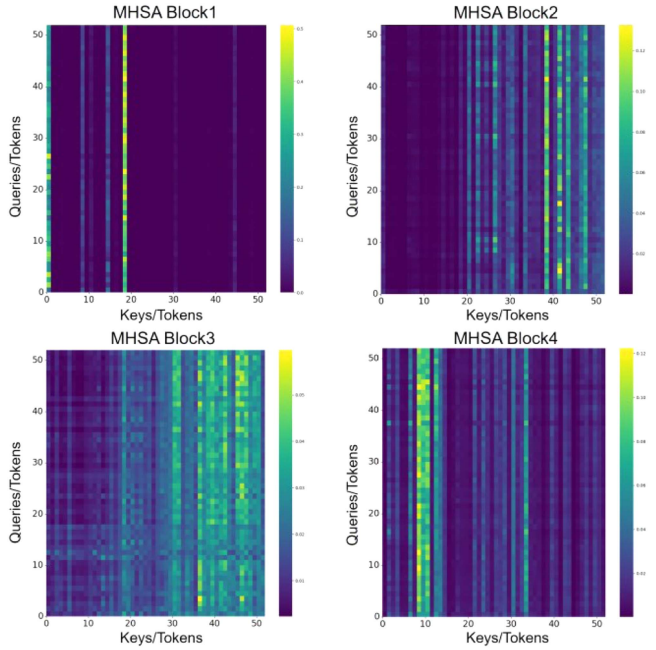

Fig. 15. Self-attention weight visualization of different MSHA blocks in the same SAG model. The brightness of individual pixels illustrates the distribution of attention weights across elements in the BRC sequence, with the ith row representing the ith query and the jth column representing the jth key, as mentioned in (9). A pixel located at coordinate (x, y) with higher brightness indicates a greater attention weight between the yth element (query) of the BRC sequence and the xth element (key), and vice versa.

whereas the deeper blocks might learn more abstract global features of the sequence. This is evidenced in Fig. 13 by the more dispersed attention distribution in the initial blocks and the more concentrated attention in specific regions in the deeper blocks. The differences between blocks reveal the SAG model's stratified information processing strategy, demonstrating how the model incrementally abstracts and integrates the input data.

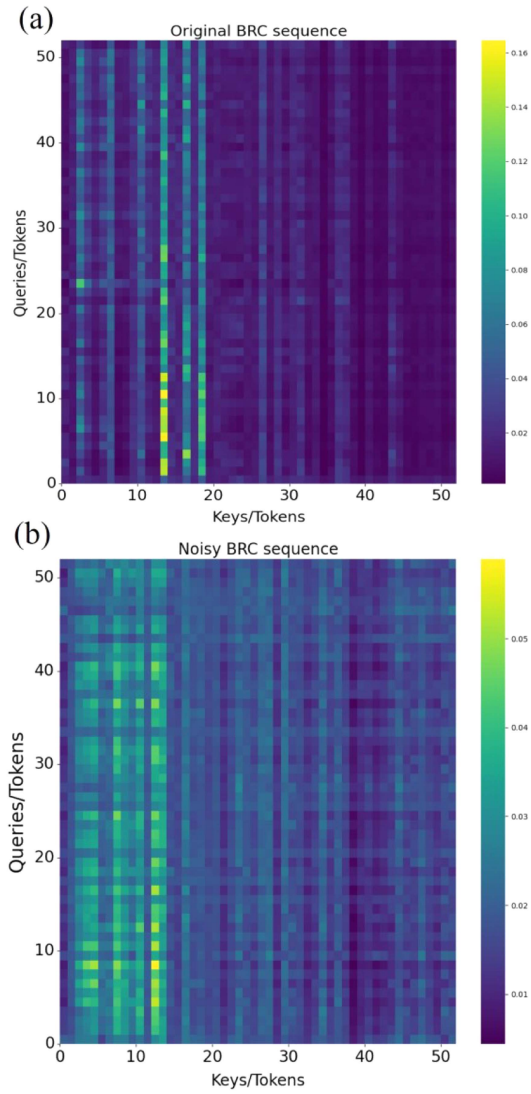

Fig. 16. Self-attention weight visualization of SAG models trained by original BRC sequences and noisy BRC sequences.

C. Noise Patterns in Self-Attention Weight Maps

The self-attention weight map can reflect the strength of interrelationships between different elements in the sequence, allowing us to identify elements that are more critical for the inversion task. Besides, another interesting observation on the self-attention visualization is that SAG trained with noisy data shows noticeable differences from that obtained with noise-free data, as shown in Fig. 16. In the noise-free attention map Fig. 16(a), we can clearly see that the attention distribution forms clear stripes. This indicates that for noise-free data, certain specific elements in the sequence are particularly important for geoacoustic inversion. On the other hand, in the noisy attention map Fig. 16(b), the distribution of attention is more chaotic, making it difficult to discern the importance of specific elements. This suggests that the self-attention mechanism model can detect anomalies in seabed observation data, potentially offering new approaches for denoising ocean exploration data and monitoring the marine environment.

As we introduced in Section II-C, it is very challenging to invert sound attenuation using BRC sequences. We believe that sound attenuation is sensitive to changes in BRC sequences, and even slight variations in the BRC sequence could lead to significant changes in the predicted results for sound attenuation. Therefore, when the BRC sequence contains noise, the impact of those elements that are relatively less important for geoacoustic inversion can no longer be ignored, which may be the reason for the chaotic distribution in the attention map. However, on the other hand, the impact of sound attenuation on the near-field BRC sequence in this study is relatively minor. Even if the error in sound attenuation appears to be relatively large, it is acceptable because it does not have a significant impact on the BRC sequence.

Comparing the attention visualization between noisy and noise-free data can provide references for analyzing noise in BRC sequences. In future geoacoustic inversion research, we hope to obtain more actual data and train the SAG model. Then,

through the self-attention weight map, we can judge the extent of noise impact and even analyze the patterns of noise if possible.

V. Conclusion

This study conducts geoacoustic inversion based on near-field bottom reflection signals, utilizing self-attention mechanisms. Employing the NBRC method, we collected bottom reflection signals at a wide range of grazing angles, providing high-quality observation data for inversion. By employing the wavenumber integration method, the BRCS dataset accurately reflects near-field geoacoustic features. The SAG model, via the MHSA mechanism, captures both global and local features in the BRC sequences, thus enhancing inversion accuracy. The AW-MTL strategy allows for adaptive adjustment of the importance of individual geoacoustic parameters during training, significantly improving sound attenuation prediction accuracy. Experiments show that our approach remains effective for simulated data with random noise, making it applicable to actual noisy data.

However, the noise in real marine environments differs from random noise, exhibiting both randomness and patterns. In Fig. 14, we observe that within the grazing angle range of 10°–20° and 50°–60°, there is a significant discrepancy between the actual measured BRC sequence and the simulated BRC sequence. This mismatch may be attributed to environmental noise during the data collection process. At small angles, the distance between the sound source and the receiver is greater, and the complex shallow marine environment introduces additional noise. Conversely, at large angles, the proximity of the sound source and receiver means that vessel noise can also affect the signal. Furthermore, as observed in Fig. 11, the critical angle of the BRC curve is between 20° and 30°, and the reflection coefficients at small and large angles are relatively insensitive to geoacoustic parameters, which may also contribute to the mismatch. In future work, we will acquire more actual measurement data and explore better signal-denoising methods. In addition, we will also focus on applying a self-attention mechanism to delve deeper into the disparities between actual and simulated data.

References

[1] C. Huang and W. S. Hodgkiss, "Matched-field geoacoustic inversion of low-frequency source tow data from the ASIAEX East China Sea experiment," *IEEE J. Ocean. Eng.*, vol. 29, no. 4, pp. 952–963, Oct. 2004, doi: 10.1109/JOE.2004.836989.

[2] N. R. Chapman and C. E. Lindsay, "Matched-field inversion for geoacoustic model parameters in shallow water," *IEEE J. Ocean. Eng.*, vol. 21, no. 4, pp. 347–354, Oct. 1996, doi: 10.1109/48.544045.

[3] D. Tollefsen, S. E. Dosso, and M. J. Wilmut, "Matched-field geoacoustic inversion with a horizontal array and low-level source," *J. Acoustical Soc. Amer.*, vol. 120, no. 1, pp. 221–230, Jul. 2006, doi: 10.1121/1.2205132.

[4] E. L. Ferguson, R. Ramakrishnan, S. B. Williams, and C. T. Jin, "Deep learning approach to passive monitoring of the underwater acoustic environment," *J. Acoustical Soc. Amer.*, vol. 140, no. 4, pp. 3351–3351, Oct. 2016, doi: 10.1121/1.4970708.

[5] Y. Shen, X. Pan, Z. Zheng, and P. Gerstoft, "Matched-field geoacoustic inversion based on radial basis function neural network," *J. Acoustical Soc. Amer.*, vol. 148, no. 5, pp. 3279–3290, Nov. 2020, doi: 10.1121/10.0002656.

[6] M. Liu, H. Niu, Z. Li, Y. Liu, and Q. Zhang, "Deep-learning geoacoustic inversion using multi-range vertical array data in shallow water," *J. Acoustical Soc. Amer.*, vol. 151, no. 3, pp. 2101–2116, Mar. 2022, doi: 10.1121/10.0009850.

[7] C. D. Escobar-Amado et al., "Seabed classification from merchant ship-radiated noise using a physics-based ensemble of deep learning algorithms," *J. Acoustical Soc. Amer.*, vol. 150, no. 2, pp. 1434–1447, Aug. 2021, doi: 10.1121/10.0005936.

[8] M. S. Ballard, K. M. Becker, and J. A. Goff, "Geoacoustic inversion for the New Jersey shelf: 3-D sediment model," *IEEE J. Ocean. Eng.*, vol. 35, no. 1, pp. 28–42, Jan. 2010, doi: 10.1109/JOE.2009.2034490.

[9] G. R. Potty et al., "Geoacoustic inversion results from the ASIAEX East China Sea experiment," *IEEE J. Ocean. Eng.*, vol. 29, no. 4, pp. 1000–1010, Oct. 2004, doi: 10.1109/JOE.2004.833224.

[10] Y. M. Jiang and N. R. Chapman, "Bayesian geoacoustic inversion in a dynamic shallow water environment," *J. Acoustical Soc. Amer.*, vol. 123, no. 6, pp. EL155–EL161, May 2008, doi: 10.1121/1.2908405.

[11] J. T. Goh and H. Schmidt, "A hybrid coupled wave-number integration approach to range-dependent seismoacoustic modeling," *J. Acoustical Soc. Amer.*, vol. 100, no. 3, pp. 1409–1420, Sep. 1996, doi: 10.1121/1.415988.

[12] S. Durand, S. Gaffet, and J. Virieux, "Seismic diffracted waves from topography using 3-D discrete wavenumber-boundary integral equation simulation," *Geophysics*, vol. 64, no. 2, pp. 572–578, Apr. 1999, doi: 10.1190/1.1444564.

[13] X. Yu, M. O. Efe, and O. Kaynak, "A general backpropagation algorithm for feedforward neural networks learning," *IEEE Trans. Neural Netw.*, vol. 13, no. 1, pp. 251–254, Jan. 2002, doi: 10.1109/72.977323.

[14] A. Vaswani et al., "Attention is all you need," 2017, *arXiv:1706.03762*.

[15] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[16] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[17] D. Hong et al., "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518615, doi: 10.1109/TGRS.2021.3130716.

[18] Z. Li, R. Zhang, J. Yan, F. Li, and J. Liu, "Geoacoustic inversion by matched-field processing combined with vertical reflection coefficients and vertical correlation," *IEEE J. Ocean. Eng.*, vol. 29, no. 4, pp. 973–979, Oct. 2004, doi: 10.1109/JOE.2004.834172.

[19] S. Wang, Y. Jiang, P. Song, J. Tan, Z. Liu, and B. He, "Memory optimization in RNN-based full waveform inversion using boundary saving wavefield reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5919212, doi: 10.1109/TGRS.2023.3317529.

[20] M. B. Porter, "The bellhop manual and user's guide: Preliminary draft," Heat, Light, and Sound Research, La Jolla, CA, USA, Tech. Rep. 260, 2011.

[21] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," 2016, *arXiv:1606.08415*.

[22] M. Liu, H. Niu, Z. Li, and Y. Guo, "A case study of geoacoustic inversion based on convolutional neural network using vertical array data," in *Proc. 5th Int. Conf. Inf. Commun. Signal Process.*, 2022, pp. 1–6, doi: 10.1109/ICICSP55539.2022.10050595.

[23] Z. Wang, Y. Ma, G. Kan, B. Liu, X. Zhou, and X. Zhang, "An inversion method for geoacoustic parameters in shallow water based on bottom reflection signals," *Remote Sens.*, vol. 15, no. 13, Jun. 2023, Art. no. 3237, doi: 10.3390/rs15133237.

[24] J. Wang et al., "Physical properties and in situ geoacoustic properties of seafloor surface sediments in the East China Sea," *Front. Mar. Sci.*, vol. 10, May 2023, Art. no. 1195651, doi: 10.3389/fmars.2023.1195651.

Yuxuan Ma received the B.S. degree in physics from Nanjing University, Nanjing, China, in 2020. He is currently currently working toward the M.E. degree in naval architecture and ocean engineering with Shandong University of Science and Technology, Qingdao, China.

His research interests include ocean acoustic field modeling, geoacoustic inversion, and deep learning.

Xiaobo Zhang received the Bachelor's degree in earth information science and technology and Ph.D. degree in marine geophysics from the Ocean University of China, Qingdao, China, in 2012 and 2017, respectively.

He is an Associate Professor with the College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao, China. He is majoring in marine geophysics and artificial intelligence.

Fukun Jiang is currently working toward the B.E. degree with Shandong University of Science and Technology, Qingdao, China.

His research interests include forward modeling, inversion methods of ocean acoustics, and deep learning.

Zhengrong Wei received the Ph.D. degree in geodetection and information technology from Chang'an University, Xi'an, China, in 2021.

He is a Lecturer with the College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao, China. His research interests include marine geophysics, ocean detection, and observation technology.

Chenguang Liu received the Master's degree in marine geology from the First Institute of Oceanography, Ministry of Natural Resources, Beijing, China, in 2003.

He is currently a Senior Engineer with the First Institute of Oceanography, Ministry of Natural Resources. He is mainly engaged in marine geophysical investigation and research work.