Seismic Stratigraphic Interpretation Based on Unsupervised Validation and Spectral Clustering Sampling

Jianyu Luo, Xiaofeng Gu, Wenkai Lu, *Member, IEEE*, and Yinshuo Li

***Abstract*—With the continuous development of computer technology and significant improvements in computing power, deep learning has found increasing applications in seismic stratigraphy interpretation, showcasing notable advancements over traditional methods. However, due to the unique characteristics of seismic data, labeling such data has become extremely challenging and time-consuming, necessitating the involvement of professional geologists. Consequently, few-shot learning has garnered considerable attention for seismic image segmentation. Nevertheless, two key challenges remain in few-shot learning: selecting more representative samples and validating the model during training. As the availability of labeled samples decreases, we are left with inadequate data for the validation set. In this article, instead of solely focusing on enhancing the network structure, we propose the utilization of spectral clustering sampling (SCS) methods for training data selection. In addition, we introduce a metric called sum of differences (SD), which can be computed without the need for labeled data, to replace the conventional validation set loss employed in traditional validation approaches. Notably, by employing SCS methods for training data selection and introducing the SD metric to replace traditional validation set loss in F3 dataset, we have achieved remarkable outcomes.**

***Index Terms*—Deep learning, image segmentation, sampling method, seismic interpretation.**

I. Introduction

GEOLOGY plays a vital role in our understanding of the Earth's composition [1]. Seismic exploration specifically utilizes the principles of wave propagation to study subsurface structures and material properties. By studying the intricate nature of seismic wave propagation, we can effectively analyze and interpret the distribution of valuable energy resources, such as petroleum and natural gas [2].

Seismic stratigraphy interpretation is a critical aspect of seismic exploration; however, manual interpretation of seismic stratigraphy is known to be inefficient and time-consuming. This is primarily due to the fact that accurate interpretations require specialized knowledge and expertise from geological experts. In recent years, with the advancements in deep learning techniques [3], an increasing number of researchers have started integrating these techniques into the interpretation of seismic images [4], [5], [6]. By leveraging deep learning methods, researchers aim to improve the efficiency and accuracy of seismic stratigraphy interpretation, allowing for more reliable and expedited analysis of subsurface structures.

Segmentation is a common problem in computer vision, and seismic stratigraphic interpretation can be viewed as a segmentation problem as well. The use of a convolutional neural network-based encoder allows for automatic feature extraction and facilitates the segmentation of seismic images [7]. One popular network architecture for segmentation tasks is U-Net [8], which has been successfully applied in various domains, including cell segmentation. In the field of seismic data segmentation, researchers have extensively utilized U-Net to achieve remarkable results [2], [9], [10]. To address the challenge of limited labeled seismic data, Ferreira et al. [11] incorporated generative adversarial networks to generate synthetic seismic images. This approach helps augment the available data and improves the performance of segmentation models.

In response to the limited availability of labeled seismic data, researchers have increasingly focused on few-shot seismic image segmentation in recent years. Alaudah et al. [12] proposed an automated strategy for generating weak labels to train the model effectively. Babakhin et al. [13] introduced an iterative loop that continuously incorporates high-confidence unlabeled samples during training, enabling few-shot learning. Seismic feature self-learning [14] and semisupervised learning [15] techniques leverage abundant unlabeled data by utilizing pretext tasks, which improve the efficiency of training models with limited labeled data. In addition, Gu et al. [16] incorporated spatial structural constraints of seismic volumes by introducing a structural loss function alongside the conventional loss function.

It is true that geological formations often exhibit spatial continuity, and seismic data collected from these formations demonstrate strong lateral similarity [17]. As a result, seismic datasets tend to have a higher degree of similarity among adjacent images, leading to the presence of redundant information. To address this issue, one effective approach is to select

Manuscript received 6 December 2023; revised 18 March 2024; accepted 20 March 2024. Date of publication 25 March 2024; date of current version 10 April 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFA0702501, in part by the National Natural Science Foundation of China under Grant 41974126 and Grant 41674116, and in part by the Key Technologies and Prototype of Synchronous Seabed Seismic and EM Exploration Systems under Grant 42327901. *(Corresponding author: Wenkai Lu.)*

The authors are with the Beijing National Research Center for Information Science and Technology (BNRist), Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: luo-jy22@mails.tsinghua.edu.cn; fengxiaogu@qq.com; lwkmf@mail.tsinghua.edu.cn; ys-li22@mails.tsinghua.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3381454

representative samples from the dataset. Uniform sampling at equal intervals is a commonly used method in seismic data segmentation [10], [18], [19], [20]. However, itis important to consider that seismic data often possesses distinct structural patterns that should be taken into account during sample selection. Recent research by Chen et al. [21] has proposed the use of principal component analysis-based dimensionality reduction sampling and spatial pyramid sampling (SPS) methods to obtain more representative samples for training segmentation network models. These methods help capture the structural characteristics of the seismic data and improve sample selection. When dealing with a limited sample size, active learning (AL) has been shown to be more effective in iteratively selecting representative samples throughout the training process [22], [23]. AL allows for the targeted selection of informative samples, but it does increase training time significantly due to the iterative process of sample selection.

Clustering is a commonly employed approach for selecting representative samples. However, conventional clustering methods often fail to produce satisfactory results when applied to seismic data. This phenomenon arises from the inherent spatial structure present in seismic data [17], which is often overlooked by conventional clustering methods, leading to unexpected outcomes. Therefore, this article introduces a seismic image sampling method based on spectral clustering [24]. Spectral clustering is a leading and popular technique in unsupervised data analysis [25]. It leverages the spectral properties of the Laplacian matrix to partition the data into meaningful clusters, even in cases where the clusters are not linearly separable in the original feature space [26], [27]. This method takes into account prior information about the structural characteristics of seismic images. Compared with AL and equal interval sampling, it demonstrates improved performance during training.

In addition to data selection, model selection is a crucial issue in few-shot learning. The scarcity of labeled data implies the unavailability of a validation set, which is a common problem in seismic image segmentation with limited samples. Previous studies [19], [23], [28] have also lacked a validation set for model selection during the training process. The absence of a validation set remains a concerning matter, despite the availability of various techniques to mitigate overfitting. To address this issue, we propose a novel approach in this study. We introduce an unsupervised validation method for model selection based on the continuity characteristics of seismic images. The main contributions of this article are summarized as follows.

1) We applied a clustering approach specifically tailored to the structure of seismic data to select training samples. This approach outperforms interval sampling and AL methods, leading to improved model training.
2) To address the model selection problem in few-shot learning, we introduce an unsupervised validation method. This method utilizes the continuity of seismic data to overcome the challenge of lacking a partition for a validation set due to limited labeled data.

The rest of this article is organized as follows. In Section II, We present theoretical approaches to spectral clustering sampling (SCS) and unsupervised validation. In Section III, we conduct experiments and do analyses. In Section IV, We discuss the significance of the proposed methods in seismic image segmentation. Finally, Section V concludes this article.

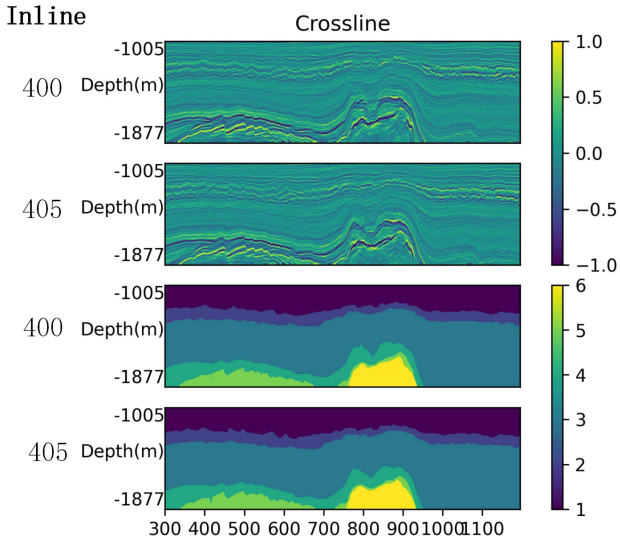


Fig. 1. First image represents the 3-D seismic data of the inline 400 profile, while the second image corresponds to the inline 405 profile. Subsequently, the labels corresponding to each profile are provided.

II. Theory

In this chapter, for the sake of facilitating subsequent explanations, we will begin by providing a detailed introduction to the dataset used in our study. Subsequently, we will elaborate on the effectiveness of SCS in seismic data and discuss the fundamental principles of unsupervised validation.

A. Datasets

The proposed method will be implemented on the F3 dataset [29], which comprises six distinct seismic phases as follows:

1) the upper N.S. group;
2) the middle N.S. group;
3) the lower N.S. group;
4) the Rijnland/Chalk group;
5) the Scruff group;
6) the Zechstein group.

To facilitate model training, each profile in this dataset has been resized to dimensions of 256×896. The dataset consists of a total of 601 profiles spanning the inline range of 100–700. The F3 dataset is characterized by two key aspects that differentiate it from typical natural image datasets.

Initially, as demonstrated in Fig. 1, the seismic profiles exhibit robust spatial continuity between adjacent profiles. A cursory observation of the figure reveals that the inline 400 and 405 profiles in this dataset look nearly indistinguishable. Here, the similarity implies that if a model performs well in segmenting the 400 profile, it is likely to perform well in segmenting the 405 profile as well.

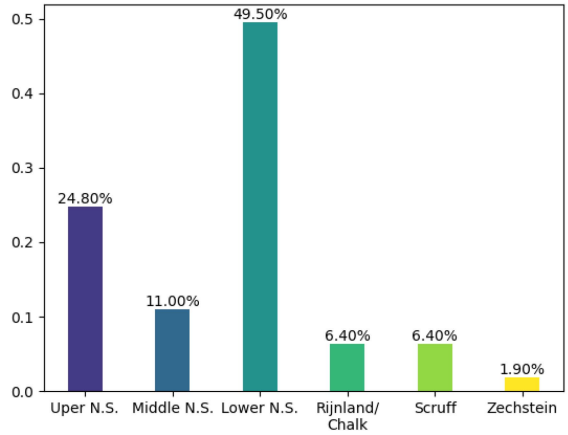


Fig. 2. Proportions of the six seismic facies in the overall F3 dataset.

Another notable characteristic of the F3 dataset is the imbalanced distribution of seismic facies classes. The lower N.S. group, comprising nearly 50% of the dataset, exhibits the largest proportion, while the Zechstein group is limited to a mere 1.9% of the dataset. Fig. 2 provides a comprehensive visualization of the category distribution across the six geological layers present in the F3 dataset.

B. Spectral Clustering Sampling

In our previous discussion, we noted the strong similarities in seismic profiles within 3-D seismic data cubes. When choosing seismic profiles to predict the entire 3-D data cube, there may be redundant information present within the profiles. Clustering offers an unsupervised means to categorize the data, partitioning it into different classes, each exhibiting unique features. Sampling profiles through clustering can help circumvent the selection of profiles containing redundant information. However, in seismic data, clustering methods often produce results that conflict with geological principles. The Euclidean distance is defined by the following formula:

$$D(X,Y)=\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-x_{ij})^2} \tag{1}$$

where X and Y represent two distinct seismic profiles. As illustrated in Fig. 3(a) represents the European distance from inline 100–700 profiles to inline 400 profile. The three red dots correspond to (b) inline 360, 400, and 655 profiles. It is counterintuitive to observe that, although inline 360 and inline 400 appear more similar, they are farther apart in Euclidean distance. Using this distance metric for clustering might lead to clustering inline 400 and inline 655 together. In such a scenario, if the inline 400 profile is chosen for training the model, it may not predict the inline 655 profile accurately. The authors give a more detailed explanation of how spectral clustering solves this problem next.

Spectral clustering is a traditional clustering algorithm that follows a two-step process. First, it constructs a similarity matrix,

Algorithm 1: SCS.

Input: Data $X \in \mathbf{R}^{N\times d}$, Similarity matrix $W \in \mathbf{R}^{N\times N}$, Sample size k

Output: $x_1^*,\ldots,x_k^*$

1: Based on W using Spectral Clustering algorithm to obtain k sets $A_1,\ldots,A_k$[30].

2: **for** each $i \in [1,k]$

$$\overline{a_i}=\frac{1}{|A_i|}\sum_{x_j\in A_i}x_j$$

3: **end for**

4: **for** each $i \in [1,k]$ **do**

$$x_i^*=\underset{x_j\in A_i}{\operatorname{argmin}}\,\|\overline{a_i}-x_j\|_2^2$$

5: **end for**

6: **return** $x_1^*,\ldots,x_k^*$

denoted as W

$$W=\begin{bmatrix} w_{11} & w_{12} & \cdots \\ w_{21} & w_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}_{N\times N} \tag{2}$$

where N is the total number of data points and w_{ij} is the similarity between the ith data point and the jth data point. This similarity matrix captures the pairwise similarities between all data points. Subsequently, spectral clustering performs clustering based on this similarity matrix, leading to the formation of distinct clusters. When generating the similarity matrix W, there are three available methods [30], and in this case, we choose to establish the similarity matrix based on k-nearest neighbor (KNN) graphs. For each data point x, if it is among the KNN of another point, we compute the similarity value w_{ij} between them, which is defined as follows:

$$w_{ij}=\begin{cases} 0 & x_i\notin K(x_j) \quad \text{or} \quad x_j\notin K(x_i) \\ 1 & x_i\in K(x_j) \quad \text{and} \quad x_j\in K(x_i) \end{cases} \tag{3}$$

where $K(x_i)$ is defined as the set of M data closest to x_i in the entire dataset X. As depicted in Fig. 3(a), it aligns with the observation that, as the spatial distance increases, the Euclidean distance also increases within a specific range around the 400 profile. Notably, from profiles 380 to 420, representing the nearest 40 points, satisfy this criterion. To ensure all profiles adhere to this condition, we set the value of M for nearest neighbors to $M=30$ (a little smaller than 40). As long as M is less than 30, we can then guarantee that the M nearest points to x_i under the Euclidean distance are also around x_i in the actual spatial sequence. After calculating the similarity matrix W, the detailed procedure of the SCS algorithm is as follows.

When dealing with seismic image profiles that exhibit spatial continuity, using the nearest-neighbor method to construct the similarity matrix in spectral clustering ensures that the clustering results will also display continuity. In other words, the results obtained from such spectral clustering will not exhibit abrupt

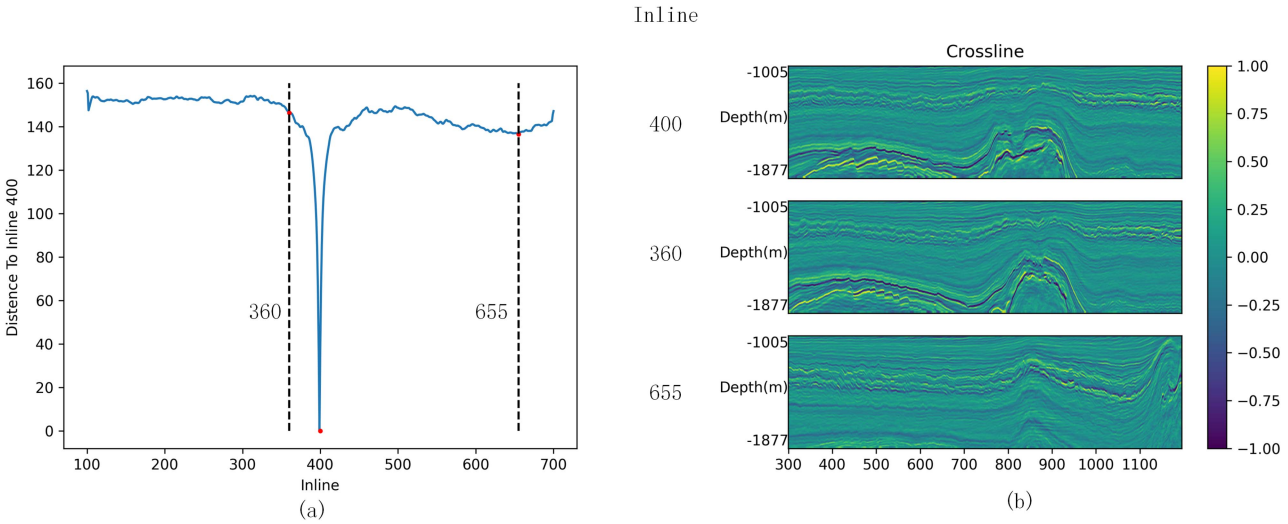


Fig. 3. (a) European distance from inline 100–700 profiles to inline 400 profile. The three red dots correspond to (b) inline 360, 400, and 655 profile. We find that the inline 400 profile is further away from the inline 360 profile and closer to the inline 655 profile in terms of Euclidean distance, but intuitively the 360 and 400 profiles are more similar.

changes or jumps. In Fig. 4, we depict the clustering results of three methods—K-means, spectral clustering, and SPS [21]—applied to the F3 dataset. A comparison of these methods reveals that both SPS sampling and K-means clustering results exhibit jumps or discontinuities. In Fig. 4(a), the portion highlighted by the red line illustrates that the inline 400 and 600 profiles are clustered together, despite the presence of an intermediate cluster separating them. This clustering result clearly violates the assumption of spatial continuity inherent in seismic data.

Spectral clustering not only produces superior results but also demonstrates enhanced stability, thereby being less susceptible to variations caused by random seed initialization. In contrast, both k-means and SPS are inclined to yield varying outcomes when the random seed is altered. Unstable clustering results can be especially problematic when training a model since it becomes uncertain whether the obtained results will prove satisfactory or unsatisfactory.

C. Unsupervised Validation

In the field of few-shot learning, specifically in seismic segmentation, the scarcity of labeled samples poses challenges in establishing a separate validation set. The concept of unsupervised validation is based on the recognition that, due to the spatial continuity of adjacent seismic profiles, the performance of a network model may be inferred from the continuity of the predicted labels of adjacent profiles. Simply put, as a network model's performance improves, the predicted labels should exhibit a more pronounced continuity. Fig. 5 presents the prediction results of two models possessing mean intersection over union (MIoU) values of 67% and 86%, respectively, for inline 650 and 649 profiles. The figures illustrate that the model with poorer performance displays substantial differences in the predicted results for adjacent profiles due to its unpredictable nature. Conversely, the model with higher performance exhibits minimal variations in the predicted results for neighboring profiles, which can be discerned visually. We define the distance $D(X, Y)$ between two processed labels as follows:

$$\delta(x_{ij}, y_{ij}) = \begin{cases} 0, & x_{ij} = y_{ij} \\ 1, & x_{ij} \neq y_{ij} \end{cases} \tag{4}$$

$$D(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} \delta(x_{ij}, y_{ij}) \tag{5}$$

where X and Y is the input matrices, while n and m are their respective sizes. To increase the stability and reliability of the sum of differences (SD) indicator, we employ the following procedure to handle the labels generated by the model: if the probability value of the predicted class for a pixel falls below a threshold, denoted as T, we assign a class label of 0 to that particular pixel. This approach ensures that any pixel whose predicted class probability is below the threshold is effectively classified as 0. This is mathematically expressed as follows:

$$\text{Softmax}(x): \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_c \end{bmatrix} \rightarrow \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_c \end{bmatrix} \tag{6}$$

$$\text{Threshold}(p) = \begin{cases} \text{argmax}(p_i) & \max(p_i) >= T \\ 0 & \max(p_i) < T. \end{cases} \tag{7}$$

Processing all the pixel points p_{ij} of a profile in this way, we can obtain Y. The new metric for unsupervised validation SD is the sum of the differences in the high-confidence regions of all profiles, can be expressed as follows:

$$\text{SD} = \sum_{i=2}^{N-1} D(Y_i, Y_{i-1}) + D(Y_i, Y_{i+1}) \tag{8}$$

TABLE I
RELATIONSHIP BETWEEN THE VALUE OF SD AND MIOU AND MFIOU

SD$\times 10^6$	3.86	3.66	3.46	3.26	3.06	2.86	2.66	2.46	2.35	1.50
FWIoU(%)	81.27	83.99	85.41	85.63	86.55	86.51	87.96	88.20	88.61	100.00
MIoU(%)	94.53	94.81	94.97	95.10	95.21	95.24	95.4	95.51	95.59	100.00

The relationship between the SD values predicted by the different models and the two metrics MFIoU and MIoU for these models. Note that the last column of models with 100% correct is the SD value calculated with real labels.

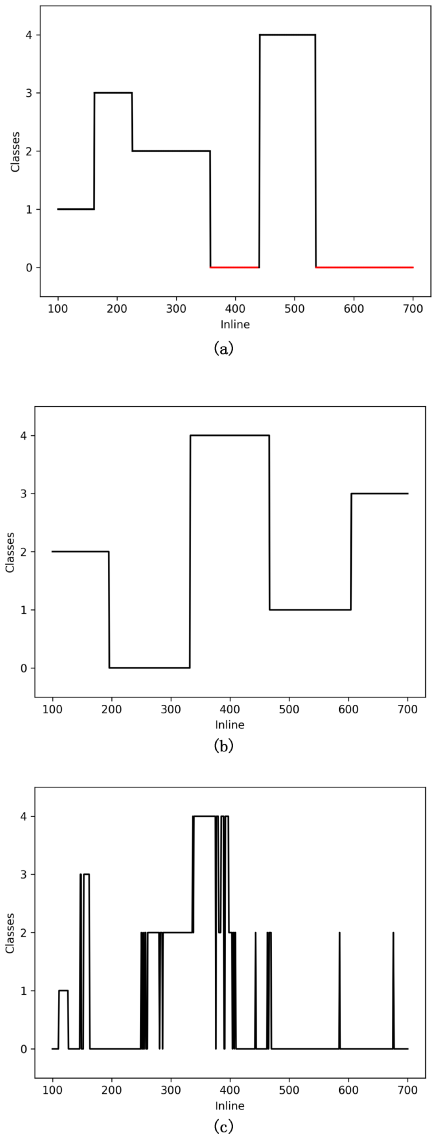


Fig. 4. Clustering results of the three different methods on the inline 100–700 profiles. (a) K-means clustering. (b) Spectral clustering. (c) SPS. In the K-means method, the red line indicates that the data points have been clustered into one group.

where N is the size of the data. The relationship between the metric SD and the model metrics MIoU and frequency-weighted intersection over union (FWIoU) is given in Table I. It is evident that as the metrics FWIoU and MIoU improve, the value of SD gradually decreases, indicating a strong correlation between them. When calculating the SD using the ground truth labels (i.e., with a 100% accurate model), the SD value is minimal.

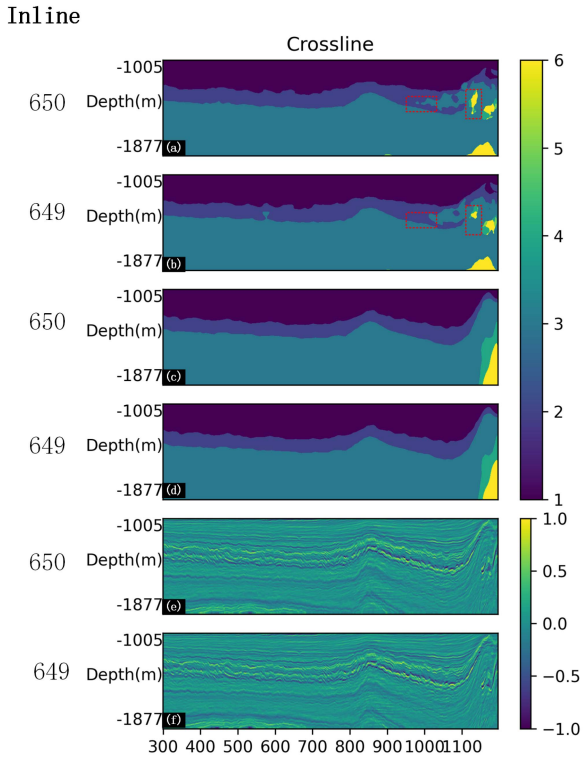


Fig. 5. (a) and (b) is the prediction results of the model with an MIoU of 67%, (c) and (d) is the prediction results of the model with an MIoU of 86%, and (e) and (f) are seismic data profiles of inline 650 and 649, respectively.

In the context of few-shot learning for seismic segmentation, conventional validation methods suffer from the limitation of utilizing only a small number of seismic profiles as the validation set. Due to this limited data, the validation set fails to provide a comprehensive representation of the overall data distribution, resulting in substantial bias during model selection. To address this issue, we propose the use of the SD value, which calculates the differences in spatial continuity between each seismic profile and its neighboring profiles. By considering the entire dataset, the SD value enables us to select a more optimal model that better captures the underlying patterns and structures in the seismic data. This approach helps mitigate the bias introduced by using a small validation set and enhances the overall robustness of the model selection process.

III. EXPERIMENTS

A. Implementation Details

In our experiments, we utilize the U-Net++ architecture [31] as the backbone network, which is a state-of-the-art deep

TABLE II
PERFORMANCE OF DIFFERENT METHODS FOR SELECTING TRAINING DATASETS

Sample size	Method	MPA(%)	MIoU (%)	FWIoU (%)	IoU (%)					
					Upper N.S.	Middle N.S.	Lower N.S.	Rijnland/ Chalk	Scruff	Zechstein
2	Interval	95.21	75.88	91.56	96.84	87.82	95.98	71.59	58.37	30.39
	SCS	96.42	79.24	93.56	97.41	87.97	97.30	77.83	59.23	50.53
	AL	95.49	78.11	91.94	97.40	87.29	96.28	76.11	59.81	43.66
3	Interval	96.07	81.01	92.77	**97.84**	87.58	95.82	76.98	66.14	53.43
	SCS	96.40	81.59	93.45	97.54	88.32	97.31	78.07	57.33	59.74
	AL	96.76	81.56	94.06	97.56	88.46	97.30	80.58	65.07	54.91
4	Interval	96.92	82.76	94.23	97.63	88.43	96.99	78.58	71.47	55.06
	SCS	97.07	83.95	94.48	97.50	88.32	97.66	84.09	65.98	61.01
	AL	97.17	83.56	94.77	97.63	88.72	97.71	82.39	64.75	65.29
5	Interval	96.89	83.70	94.16	97.63	88.82	96.71	79.95	70.17	61.29
	SCS	**97.69**	**88.01**	**95.59**	97.71	88.67	97.83	**87.04**	**77.53**	**75.66**
	AL	97.57	86.18	95.45	97.63	**88.97**	**97.84**	85.81	74.05	69.10

The table shows the performance of different methods for selecting training datasets, including interval, SCS, and AL [23]. In the same approach, we have extracted different quantities of samples for training. MIoU, FWIoU (Frequency-Weighted Intersection Over Union), IoU, and MPA (mean average precision) were displayed in the table as metrics used to evaluate the performance of the model. The best results for each metric are highlighted in bold.

learning architecture that enhances skip connections compared with U-Net and demonstrates superior performance in segmentation tasks. Furthermore, we employ ResNet34 [32] as the encoder while keeping the decoder part unchanged. The model is trained using the Adam optimization method [33], with a learning rate of 0.0002. Detailed information about the dataset can be found in Section II-A. After selecting the training samples, the remaining samples are designated as the testset for evaluation. To augment the data, in addition to traditional techniques, such as flipping and Gaussian noise, we also employ the TPS method [23]. The model follows a section-based paradigm with a batch size of 16. The spectral clustering algorithm is implemented using Scikit-learn, while the segmentation task is implemented using the PyTorch framework [34] on an NVIDIA GeForce GTX 4090 GPU.

B. Metrics

MIoU is one of the most commonly used metrics in segmentation tasks, and its calculation can be expressed as follows:

$$\text{MIoU} = \frac{1}{k}\sum_{i=1}^{k}\frac{p_{ii}}{\sum_{j=1}^{k}p_{ij}+\sum_{j=1}^{k}p_{ji}-p_{ii}}. \tag{9}$$

Here, k represents the number of categories or classes. p_{ij} denotes the classification outcome where a pixel originally belonging to category i is predicted to fall under category j. This metric provides a visual representation of the ratio between the predicted segmentation and the ground truth.

Mean pixel accuracy (MPA) is considered one of the fundamental evaluation metrics in segmentation tasks, which is defined as follows:

$$MPA = \frac{1}{k}\sum_{i=1}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}}. \tag{10}$$

Given the imbalanced nature of class distribution in the F3 dataset, it is crucial to incorporate the frequency-weighted intersection over union (FWIoU) metric. FWIoU is defined as follows:

$$S = \frac{1}{\sum_{i=1}^{k}\sum_{j=1}^{k}p_{ij}} \tag{11}$$

$$\text{FWIoU} = \frac{1}{S}\sum_{i=1}^{k}\frac{\sum_{j=1}^{k}p_{ij}p_{ii}}{\sum_{j=1}^{k}p_{ij}+\sum_{j=1}^{k}p_{ji}-p_{ii}}. \tag{12}$$

In our experiments, it can be observed that the MIoU metric displays larger fluctuations, whereas FWIoU demonstrates greater stability.

One important point to note is that all the aforementioned metrics are calculated based on the predicted outcomes for a single profile. However, in later sections of the article, these metrics are averaged over all profiles when evaluating a model. If a category does not appear in both the predicted and true labels of a profile, that particular profile is excluded when calculating the model's single IoU for that specific category.

C. Experiments of SCS

To validate the effectiveness of SCS in seismic data analysis, we compared it with interval sampling and AL. In addition, we evaluated the performance across different sample sizes ranging from 2 to 5. The results are presented in Table II. From the table, we observe a significantly superior performance of SCS sampling when compared to interval sampling. Particularly, for sample sizes of 2 and 5, the MIoU metric improved by 3.36% and 4.31%, respectively. However, for sample sizes 3 and 4, the SCS method did not exhibit a notable improvement over the interval sampling. This can be attributed to the fact that interval sampling does not incorporate the inherent structural information of the data, rendering it less stable. Consequently, while interval

sampling may occasionally yield favorable results, it can also lead to highly unfavorable outcomes in other cases.

In terms of trained model performance, the SCS method exhibits slightly superior outcomes to AL, although the difference is not statistically significant. Nonetheless, SCS provides two additional advantages over AL. First, as illustrated in Table II, SCS demonstrates remarkable stability due to its clustering mechanism based on the structure of seismic data. It is highly robust and immune to any fluctuations resulting from random seed initialization. Given the same dataset, SCS produces identical results consistently. On the other hand, the AL selects varied samples each time, leading to significant fluctuations in the corresponding outcomes, which may not always be favorable. Second, the AL algorithm is inherently complex, requiring consideration of initial sample selection during the iterative process. Conversely, the SCS algorithm significantly reduces the computational time. For instance, when selecting 5 samples, AL takes approximately 1 h to complete, whereas the SCS method only requires 4 s.

From the table, we can observe a gradual improvement in the performance of the SCS method compared with the AL sampling method as the number of samples increases. This can be easily understood as the AL method selects certain samples and then fixes them, subsequently adding new samples on top of those fixed ones. In this process, the AL method may obtain a suboptimal solution. This can be analogous to feature selection, where it is known that a single optimal combination of a few features may not necessarily yield superior performance, as interference between features can arise.

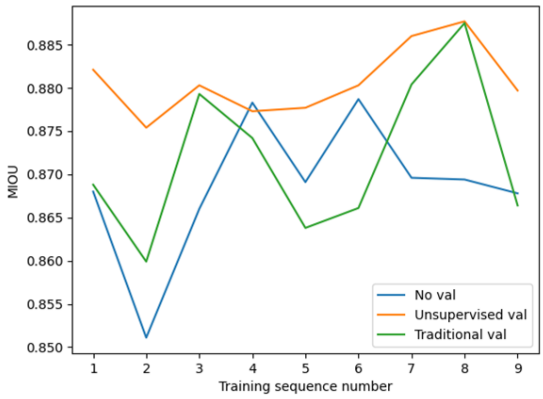
D. Experiments of Unsupervised Validation

In Table II, we can see that regions upper N.S., middle N.S., and lower N.S. have the highest IOU scores. With a small size of samples in the training set, we can achieve excellent results in the test set, and itit is stable and difficult to further improve their accuracy. Due to their smaller proportion in the overall data, Scruff and Zechstein exhibit greater variability in their predictions. Selecting a model using unsupervised validation means choosing a better model among these variations. As shown in Fig. 6, a total of nine experiments were conducted, and the significant variations in the results between experiments can be attributed to the stochastic nature of data augmentation and the limited number of samples in classes Zechstein and Scruff. In these experiments, the indexes of the profiles used in training set and in the validation set of the traditional validation method are, respectively

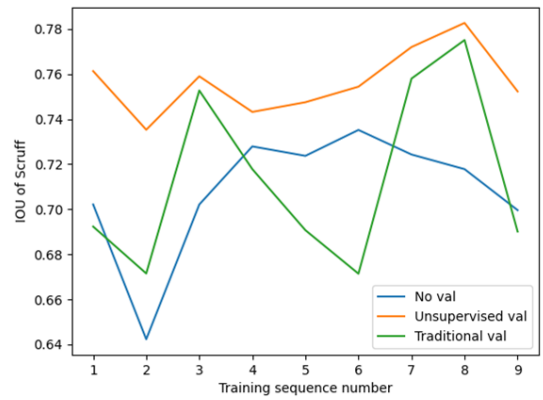
$$\textbf{Train Index} = \{156, 243, 382, 550, 651\} \tag{13}$$

$$\textbf{Validation Index} = \{450\}. \tag{14}$$

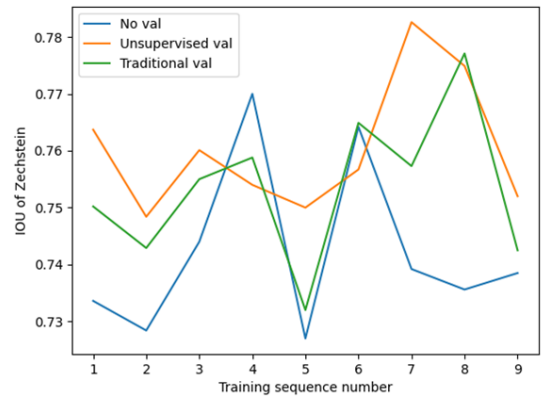
The profile with index 450 was chosen as the validation set because it is relatively distant from all five profiles in the training set. In the "no validation" method, we selected the model from the final training iteration for evaluation. The experimental results clearly demonstrate that in all nine trials, the worst results consistently occurred in the absence of a validation set. This observation indirectly highlights the necessity of a validation set in networks where overfitting is not severe. Futhermore,



(a)



(b)



(c)

Fig. 6. Comparison of three methods, the method of not using the validation set, the traditional validation method and the unsupervised validation in nine training sessions. (a) MIoU. (b) IoU of Scruff. (c) IoU of Zechstein.

there is generally minimal difference between the traditional validation method and the "no validation" method. This can be attributed to the limited size of the validation set, which fails to fully represent the overall data distribution, thereby limiting its effectiveness in selecting models during the training process. To choose a superior model, it becomes crucial to increase the sample size of the validation set. However, this approach contradicts the essence of few-shot learning, which aims to train better models with a small number of training examples. The requirement for a larger labeled validation dataset goes against

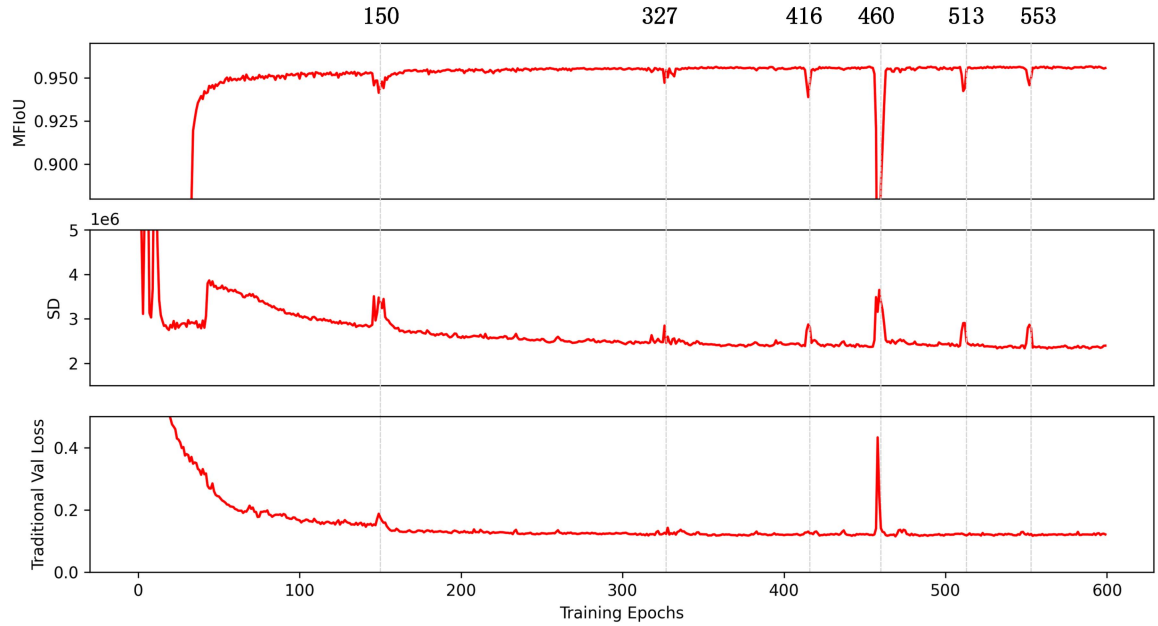


Fig. 7. Displays the values of FWIoU, SD, and traditional val loss for the trained model across 600 rounds of the training process. The gray dotted lines indicate points where the FWIoU experienced sudden fluctuations during the model training, specifically indicating deteriorations. The numbers located at the top indicate the corresponding epochs for each dotted line.

this principle and becomes counterproductive. By employing the unsupervised validation method, we utilized the entire dataset and leveraged the full information available, resulting in superior performance compared to traditional methods. This approach consistently identifies better models, as it takes advantage of the complete dataset and its inherent properties.

E. More Experiments

In order to get a more intuitive feel for how the SD value screening model works, as shown in Fig. 7 we then evaluated it once more with the test set after each epoch of one training was completed. There were 600 epochs in total.

The training set and validation set used in the traditional method remained the same as previously established. The overall trend of the SD value curve closely resembles that of the traditional validation method, gradually decreasing from a higher value. The SD values in the initial 100 rounds exhibit instability, which can be attributed to the model's instability during this stage, causing greater fluctuations in the probability values for each pixel point. Consequently, pixel points with probability values below the threshold of T may have been set to 0, leading to a decrease in the SD value. After training the model for 200 epochs, the FWIoU stabilizes, but there are sporadic instances where the model's performance deteriorates and jitter occurs. Notably, such an occurrence is observed at 460 epochs, with the FWIoU abruptly dropping from 95% to below 88%. This anomaly was effectively identified using both the traditional validation method and the method proposed in this chapter, resulting in a sudden increase in both the SD and traditional val loss values.

But in some instances, the degradation of model performance goes unnoticed by traditional validation methods. At epochs 416, 513, and 553, the FWIoU experienced a slight drop, albeit not as significant as the earlier mentioned one. During these epochs, the SD values exhibited pronounced spikes in the corresponding positions. However, the traditional method failed to detect this information, resulting in a flat curve at those points. The one-to-one correspondence between FWIoU and SD values in the model training process serves as compelling evidence for the importance of SD values in model selection. Furthermore, the fact that the SD increases when the model performance worsens further validates the effectiveness of the method in the opposite direction.

IV. Discussion

SCS algorithms and unsupervised validation are relevant for the task of segmentation of seismic image. Instead of focusing on improving the deep network structure, the application scope of our proposed methods will be broader. Regardless of the training method used for the network, the selection of samples and model selection are necessary. In Fig. 6, it appears that including or excluding the validation set does not significantly affect the final model results. This observation suggests that overfitting is not prominent in this network. However, it is important to note that we cannot guarantee this is the case for every network. Undoubtedly, processing the SD values after each training epoch would entail a substantial computational overhead. For instance, in the case of 600 total training epochs, calculating the SD value for each round in the F3 dataset alone would incur an additional hour of computational time.

Another important issue in few-shot learning for seismic images is determining the appropriate sample size, i.e., the number of clusters in spectral clustering. If the distribution of profiles within a 3-D seismic volume exhibits greater variability and

complexity, it is likely to contain less redundant information. Consequently, it may be necessary to sample a larger number of instances from this volume to train a more effective model. This is a very challenging problem, selecting samples with different sampling methods and then training with the samples is a two-stage problem. Whether the selected samples are good or not is determined by the training results, but you have to know the labels of the samples to get the training results. This creates a contradiction, which will increase the amount of labelling of seismic data, contrary to our original intention.

V. Conclusion

In this article, we propose methods for data sampling and model selection in seismic image segmentation using deep learning models based on the spatial structure of seismic data. Our data sampling method, called SCS, constructs a similarity matrix for spectral clustering using the nearest-neighbor method, ensuring that the clustering results align with the spatial characteristics of the seismic data. We conducted experiments on the F3 dataset and compared SCS with AL [16]. The results demonstrate that SCS selects superior samples within seconds compared with the hours taken by AL, making it a simple yet effective approach for sampling seismic data.

Given the scarcity of labeled data in few-shot learning scenarios, we propose an unsupervised validation approach. After each round of training, we generate processed labels for all the data based on the model. Subsequently, we designed a metric to quantify the differences between these labels, which is the value of sum of difference (for). This unsupervised validation method outperforms traditional supervised validation approaches, enabling the selection of better models during the training process. Moreover, the unsupervised validation approach eliminates the need to allocate a portion of labeled data as a separate validation set, allowing us to utilize the entire dataset for training purposes.

References

[1] S. Gandhi and B. Sarkar, *Essentials of Mineral Exploration and Evaluation*. Amsterdam, The Netherlands: Elsevier, 2016.

[2] E. Tolstaya and A. Egorov, "Segmentation of seismic images," in *Proc. Conf. Comput. Graph. Vis.*, 2021, pp. 564–570.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[4] T. Zhao, "Seismic facies classification using different deep convolutional neural networks," in *Proc. SEG Int. Expo. Annu. Meeting*, 2018, pp. 2046–2050.

[5] H. Di, Z. Wang, and G. AlRegib, "Real-time seismic-image interpretation via deconvolutional neural network," in *Proc. SEG Int. Expo. Annu. Meeting*, 2018, pp. 2051–2055.

[6] S. Yu and J. Ma, "Deep learning for geophysics: Current and future trends," *Rev. Geophys.*, vol. 59, no. 3, 2021, Art. no. e2021RG000742.

[7] F. Qian, M. Yin, X.-Y. Liu, Y.-J. Wang, C. Lu, and G.-M. Hu, "Unsupervised seismic facies analysis via deep convolutional autoencoders," *Geophysics*, vol. 83, no. 3, pp. A39–A43, 2018.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.

[9] H. Su-Mei, S. Zhao-Hui, Z. Meng-Ke, Y. San-Yi, and W. Shang-Xu, "Incremental semi-supervised learning for intelligent seismic facies identification," *Appl. Geophys.*, vol. 19, no. 1, pp. 41–52, 2022.

[10] D. Wang and G. Chen, "Seismic stratum segmentation using an encoder–decoder convolutional neural network," *Math. Geosci.*, vol. 53, no. 6, pp. 1355–1374, 2021.

[11] R. S. Ferreira, J. Noce, D. A. B. Oliveira, and E. V. Brazil, "Generating sketch-based synthetic seismic images with generative adversarial networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1460–1464, Aug. 2020.

[12] Y. Alaudah, S. Gao, and G. AlRegib, "Learning to label seismic structures with deconvolution networks and weak labels," in *Proc. SEG Int. Expo. Annu. Meeting*, 2018, pp. 2121–2125.

[13] Y. Babakhin, A. Sanakoyeu, and H. Kitamura, "Semi-supervised segmentation of salt bodies in seismic images using an ensemble of convolutional neural networks," in *Proc. 41st DAGM German Conf. Pattern Recognit.*, 2019, pp. 218–231.

[14] H. Di, Z. Li, H. Maniar, and A. Abubakar, "Seismic stratigraphy interpretation by deep convolutional neural networks: A semisupervised workflow," *Geophysics*, vol. 85, no. 4, pp. WA77–WA86, 2020.

[15] B. A. A. Monteiro, H. Oliveira, and J. A. D. Santos, "Self-supervised learning for seismic image segmentation from few-labeled samples," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8028805.

[16] X. Gu, W. Lu, Y. Li, and Y. Wang, "Semi-supervised seismic stratigraphic interpretation constrained by spatial structure," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5912710.

[17] A. M. Figueiredo, M. Gattass, and F. Szenberg, "Seismic horizon mapping across faults with growing neural gas," in *Proc. 10th Int. Congr. Braz. Geophysical Soc.*, 2007, pp. 1476–1481.

[18] Y. Zhao, B. Chai, L. Shuo, Z. Li, H. Wu, and T. Wang, "Few-shot learning for seismic facies segmentation via prototype learning," *Geophysics*, vol. 88, no. 3, pp. IM41–IM49, 2023.

[19] K. Li, W. Liu, Y. Dou, Z. Xu, H. Duan, and R. Jing, "Contrastive learning approach for semi-supervised seismic facies identification using high-confidence representations," 2022, *arXiv:2210.04776*.

[20] F. Li, H. Zhou, Z. Wang, and X. Wu, "ADDCNN: An attention-based deep dilated convolutional neural network for seismic facies analysis with interpretable spatial– spectral maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1733–1744, Feb. 2021.

[21] X. Chen, Q. Zou, X. Xu, and N. Wang, "A stronger baseline for seismic facies classification with less data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5914910.

[22] L. Wang et al., "Semisupervised semantic segmentation for seismic interpretation," *Geophysics*, vol. 88, no. 3, pp. IM61–IM76, 2023.

[23] X. Gu, W. Lu, Y. Ao, Y. Li, and C. Song, "Seismic stratigraphic interpretation based on deep active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5912511.

[24] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 849–856.

[25] U. Shaham et al., "SpectralNet: Spectral clustering using deep neural networks," in *Proc. 6th Int. Conf. Learn. Representations*, 2018.

[26] Y. Ogino and M. Yukawa, "Spectral clustering with automatic cluster-number identification via finding sparse eigenvectors," in *Proc. 26th Eur. Signal Process. Conf.*, 2018, pp. 1187–1191.

[27] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[28] M. A. Shafiq, M. Prabhushankar, H. Di, and G. AlRegib, "Towards understanding common features between natural and seismic images," in *Proc. SEG Int. Expo. Annu. Meeting*, 2018, pp. 2076–2080.

[29] Y. Alaudah, P. Michałowicz, M. Alfarraj, and G. AlRegib, "A machine-learning benchmark for facies classification," *Interpretation*, vol. 7, no. 3, pp. SE175–SE187, 2019.

[30] U. V. Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, pp. 395–416, 2007.

[31] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet: A nested U-Net architecture for medical image segmentation," in *Proc. 4th Int. Workshop Deep Learn. Med. Image Anal., 8th Int. Workshop Multimodal Learn. Clin. Decis. Support, Held Conjunction*, 2018, pp. 3–11.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[34] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8026–8037.

Jianyu Luo received the bachelor degree in mechanical engineering from the College of Engineering, Southwest Jiaotong University, Chengdu, China, in 2021. He is currently working toward the master's degree in automation with Tsinghua University, Beijing, China.

His current research interests include seismic stratigraphy interpretation, fault interpretation, deep learning, and its applications.

Xiaofeng Gu received the master's degree in control science and engineering from the Department of Automation, Tsinghua University, Beijing, China, in 2023.

His research interests include image processing, pattern recognition, and artificial intelligence.

Wenkai Lu (Member, IEEE) received the bachelor degree in automation control from Tsinghua University, Beijing, China, in 1991, and the Ph.D. degree in geophysics from the Petroleum University, Beijing, in 1996.

He is currently a Full Professor with the Department of Automation, Tsinghua University. His research interests include seismic signal processing, image processing, pattern recognition, machine learning, and artificial intelligence.

Yinshuo Li received the bachelor's degree in automation from the College of Engineering, Ocean University of China, Qingdao, China, in 2019, and the M.E. degree in control science and engineering in 2022 from the Department of Automation, Tsinghua University, Beijing, China, where he is currently working toward the Ph.D. degree in control engineering with the Department of Automation.

His current research interests include image super-resolution, seismic signal processing, deep learning, and its applications.