

# Unsupervised contrastive learning for seismic facies characterization

Jintao Li<sup>1</sup>, Xinming Wu<sup>1</sup>, Yueming Ye<sup>2</sup>, Cun Yang<sup>2</sup>, Zhanxuan Hu<sup>3</sup>, Xiaoming Sun<sup>1</sup>, and Tao Zhao<sup>4</sup>

## ABSTRACT

Seismic facies characterization plays a key role in hydrocarbon exploration and development. The existing unsupervised methods are mostly waveform-based and involve multiple steps. We have developed a method to leverage unsupervised contrastive learning to automatically analyze seismic facies. To obtain a stable result, we use 3D seismic cubes instead of seismic traces or their variants as inputs of networks to improve lateral consistency. In addition, we treat seismic attributes as geologic constraints and feed them into the network along with the seismic cubes. These different seismic and multiattribute cubes from the same position are

regarded as positive pairs and the cubes from a different position are treated as negative pairs. A contrastive learning framework is used to maximize the similarities of positive pairs and minimize the similarities of negative pairs. In this way, we can enforce the samples with similar features to get close while pushing the samples with different features to be separated in the space where we make the seismic facies clustering. This contrastive learning framework is a one-stage, end-to-end, and unsupervised fashion without any manual labels. We have determined the effectiveness of this method by using it to a turbidite channel system in the Canterbury Basin, offshore New Zealand. The obtained facies map is continuous, resulting in a stable and reliable classification.

## INTRODUCTION

Seismic facies can be defined as groups of seismic reflections whose patterns (such as amplitude, frequency, and geometry) are different from those of adjacent groups (West et al., 2002), and their characterization plays a key role in hydrocarbon exploration and development. However, the characterization of seismic facies still heavily relies on experienced interpreters and needs laborious manual work. Moreover, the growth of seismic data significantly increases the difficulty of manual interpretation. It is necessary to develop some automatic seismic interpretation tools to address these issues.

During the past few decades, many machine-learning algorithms have been proposed to help interpreters automatically analyze seismic facies, such as *k*-means (Barnes and Laughlin, 2002; Sabeti and Javaherian, 2009), the self-organizing map (SOM) (de Matos et al., 2007; Zhao et al., 2015, 2016), generative topographic mapping

(Wallet et al., 2009; Roy et al., 2014), and support vector machine (Kuzma and Rector, 2005; Zhao et al., 2005; Zhang et al., 2015). Although these conventional machine-learning methods reduce the heavy workload, intervention from interpreters is still necessary to obtain an acceptable result (Zhang et al., 2021).

In recent years, deep learning has achieved large success in many computer vision tasks including pattern recognition (Ronneberger et al., 2015; Chen et al., 2017) and deep clustering (Chang et al., 2017; Caron et al., 2018; Li et al., 2021b). Due to the powerful ability for learning features and representation, it soon attracted great attention from seismic interpreters, who have proposed many methods based on deep learning to automatically characterize seismic facies. These methods can generally be divided into three categories: supervised learning, semisupervised learning, and unsupervised learning.

Manuscript received by the Editor 15 March 2022; revised manuscript received 11 May 2022; published ahead of production 1 August 2022; published online 3 October 2022.

<sup>1</sup>University of Science and Technology of China, School of Earth and Space Sciences, Laboratory of Seismology and Physics of Earth's Interior, Hefei, China and University of Science and Technology of China, Mengcheng National Geophysical Observatory, Hefei, China. E-mail: lijintao@mail.ustc.edu.cn; xinmwu@ustc.edu.cn (corresponding author); sunxiaoming\_cupb@sina.com.

<sup>2</sup>PetroChina Hangzhou Research Institute of Geology, Hangzhou, China. E-mail: yeym\_hz@163.com; yangc\_hz@petrochina.com.cn.

<sup>3</sup>Xi'an University of Posts and Telecommunications, Xi'an, China. E-mail: huzhanxuan@mail.nwpu.edu.cn.

<sup>4</sup>Schlumberger, Houston, Texas, USA. E-mail: tao.zhao@alumni.ou.edu.

© 2023 Society of Exploration Geophysicists. All rights reserved.

The first group of methods (Wrona et al., 2018; Zhang et al., 2019; 2021; Liu et al., 2020; Kaur et al., 2021; Li et al., 2021a), supervised learning, usually need large amounts of training seismic data and corresponding exact seismic facies labels, and then feed them into a neural network to train until the network converges. These supervised methods can achieve high accuracy; however, they still face two serious challenges. First, the well-interpreted seismic facies labels are insufficient, or even lacking, because it can be highly time-consuming and subjective to manually prepare labels. Second, most of the supervised methods are typically trained with data sets from a single survey and, therefore, may not generalize well when applied to data in different regions. Some researchers proposed semisupervised learning methods (Qi et al., 2016), which deal with the large amounts of unlabeled seismic data in combination with typically smaller sets of labeled data (van Engelen and Hoos, 2020).

Unsupervised methods attract significant amounts of attention when dealing with the problem of seismic facies interpretation because they do not need any manual labels. Some researchers use an autoencoder network to extract discriminant and invariant features from unlabeled data, and then leverage conventional machine learning methods (such as *k*-means and SOM) to classify the extracted features (Qian et al., 2018; Li et al., 2019; Puzyrev and Elders, 2020). Some other methods combine the feature extraction and clustering to simultaneously optimize the two steps of feature extraction and clustering (Duan et al., 2019; Zhu et al., 2022). However, these methods still require multiple steps, which often include a pretrain step. In addition, most of these methods are based on seismic traces or their variants which lack the horizontal consistency constraints, resulting in a discontinuous and noise classification.

Recently, contrastive learning has emerged as a promising paradigm of unsupervised learning to achieve state-of-the-art performance

in some computer vision tasks. In contrastive learning, positive pairs from the same images with different augmentations would be enforced closer, whereas negative pairs from the different images are supposed to be pushed away (Chen et al., 2020; He et al., 2020; Wang et al., 2021). In this paper, we leverage a contrastive learning framework based on Li et al. (2021b) to automatically analyze seismic facies. This method is a one-stage and end-to-end single process fashion without any manual label. Unlike the previously mentioned methods, we extract 3D seismic cubes instead of seismic traces or their variants to form training data set, which can impose lateral consistency and help facies map to be more stable and continuous. In addition, as one of the important conventional interpretation methods, seismic attributes are regarded as geologic constraints and fed into the network along with seismic data (Zhao et al., 2017). By replacing data augmentations with seismic attributes, the two types of inputs can be properly processed by the contrastive learning framework, which maximizes the similarities of the seismic and multiattribute cubes from the same position and minimizes the similarities of the cubes from different positions. We apply the proposed method to seismic data that contain a turbidite system with a special focus on a channel feature, acquired over the Canterbury Basin, offshore New Zealand. We obtain a continuous and stable result that is consistent with previous interpretations by Zhao et al. (2016). To summarize, the major point of our work is as follows:

- 1) We leverage a one-stage and end-to-end contrastive learning fashion without any manual labels to solve seismic facies interpretation problems.
- 2) To improve lateral consistency, we extract 3D seismic cubes as training data set, which can avoid a discontinuous and noise result.
- 3) We use seismic attributes as geologic constraints in the proposed method.

We begin our paper by discussing the geologic setting and the workflow to prepare training data sets. Second, we introduce the contrastive learning framework used in this work. Then, we apply the network to the turbidite system. Finally, we conclude our work.

## GEOLOGIC SETTING AND TRAINING DATA

We first introduce the basic geologic setting, i.e., the turbidite system within a field seismic volume used in this paper. Based on this turbidite system, we generate suitable training data sets for our unsupervised method.

### Geologic setting

The study area lies in the Canterbury Basin, offshore New Zealand (Figure 1). This turbidite system is situated on the transition zone of continental slope and rise and filled with plentiful paleocanyons and turbidite deposits of Cretaceous and Tertiary ages. The sedimentary process is controlled by a single tectonically driven transgressive-regressive cycle (Zhao et al., 2015, 2016). A 3D seismic survey called Waka-3D was

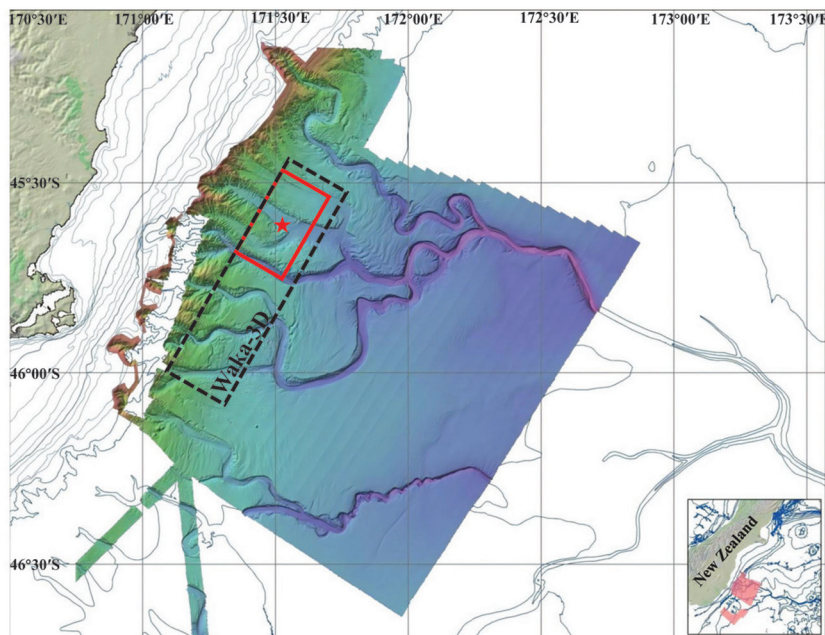


Figure 1. This map shows the location of the turbidite channel system. The black rectangle denotes the boundary of the Waka 3D data set wherein the red rectangle shows the area used in this study. The color means the elevation of the ocean basin (modified from Mitchell and Neil, 2012; Zhao et al., 2015).

acquired over this area and was made public by New Zealand Petroleum and Minerals. The black rectangle shown in Figure 1 denotes the whole survey of the data set. Several studies have explored the survey with abundant channels (Zhao et al., 2015, 2016; Wallet and Hardisty, 2019), while they only focus on one part of the survey (denotes by the red rectangle in Figure 1).

Zhao et al. (2016) interpret the Miocene turbidite system through the phantom horizon of the seismic volume shown in Figure 2. A 2D view of this horizon is shown in Figure 3. The white arrows indicate main slope multistory channels that converge and form a lobate feature downstream. The black arrow indicates several sinuous channel complexes positioned around the main channels, and some channels are difficult to identify on the seismic amplitude horizon slice. These channels (in small and large scales) are probably mud-filled. The blue arrows are identified as some older, probably sand-filled channels that show high amplitude in seismic data and developed earlier than the mud-filled channels. Therefore, these earlier sand-filled channels are cut through and covered by the later deposited mud-filled channels. Then, the red arrows indicate possible slope fans, which are widely distributed in this turbidite system, and the orange arrow is identified as some messy turbidite current or slump deposits.

### Training data

Most of the existing unsupervised methods for automatic seismic facies analysis are only based on seismic waveforms or their variants and abandon seismic attributes. However, the characteristics of an individual seismic trace are too simple for networks to extract effective features. Meanwhile, single seismic trace-based operations lack lateral consistency. Both of these facts bring challenges to subsequent clustering. To tackle this issue, we first prepare training data sets by extracting 3D cubes (centered at the positions to be classified) of seismic amplitude and other attributes. We then use the seismic amplitude cubes and the corresponding attribute cubes as the inputs of the network in which the strategy of contrastive learning is used to analyze the abundant information of these

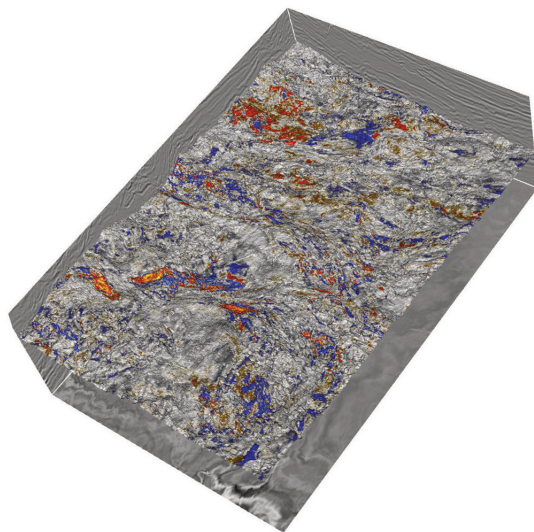


Figure 2. A 3D view of the seismic horizon used by Zhao et al. (2016).

two types of inputs simultaneously (for the details of contrastive learning, see the next section).

Seismic attributes are calculated from seismic data, and they have been widely used in seismic facies analysis for the past few decades (La Marca and Bedle, 2022). A good seismic attribute is directly sensitive to the desired geologic feature of interest, which helps us to interpret seismic facies (Chopra and Marfurt, 2007). In other words, by properly choosing seismic attributes as inputs of neural networks, we are able to introduce prior geologic constraints (those highlighted by the attributes) into the networks.

To discriminate the facies of interest, we must choose suitable and independent seismic attributes without redundancy (Zhao et al., 2015). Here, we follow Zhao et al. (2016) and Wallet and Hardisty (2019) and choose four types of attributes including coherent energy (Figure 4a), peak spectral magnitude (Figure 4b), peak spectral frequency (Figure 4c), and curvedness (Figure 4d). The coherent energy is sensitive to the amplitude response. Peak spectral frequency and peak spectral magnitude highlight the variation of the seismic response that can be used to distinguish thick channels from thin ones and overbank deposits. Curvedness defines the magnitude of reflector structural or stratigraphic deformation (Zhao et al., 2016; Wallet and Hardisty, 2019).

In preparing training data sets, we first normalize the seismic data and the four seismic attributes by the mean and standard deviation of each data as follows:

$$x_{\text{norm}} = \frac{x - x_{\text{mean}}}{x_{\text{std}}}. \quad (1)$$

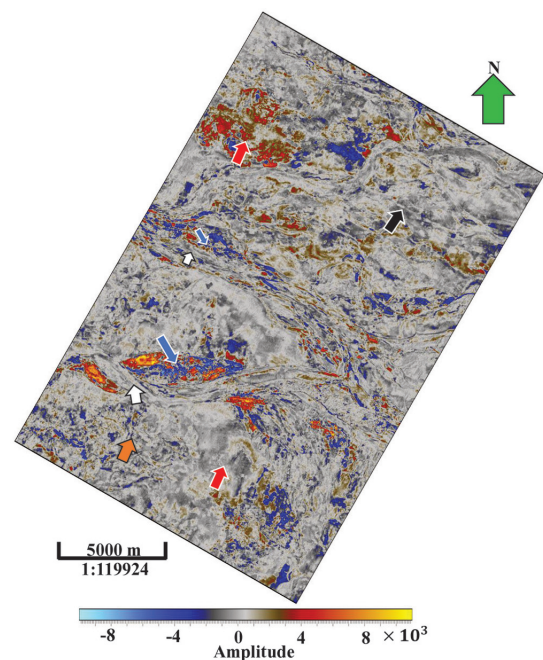


Figure 3. The seismic amplitude horizon slice. Interpreted by Zhao et al. (2016), the white arrows indicate two main stacked channels and the black arrow indicates several sinuous channel complexes. The blue arrows are identified as older sand-filled channels that cut through by the stacked channels. The red arrows indicate widely distributed slope fans and the orange arrow indicates turbidite current or slump deposits.

This normalization would be helpful for the training and prediction because the range of values varies greatly from one data set to another. Assuming that the size of the target slice is  $n_1 \times n_2$  (i.e., inline  $\times$  crossline), we extract small 3D seismic and attributes centered at the samples on this slice to form training data set, following the three steps, as shown in Figure 5. (1) Vertically selecting a suitable time window (window size is  $t$ ) centered on the target horizon seismic slice, then, we can obtain a volume of size  $n_1 \times n_2 \times t$ . This step is to flatten the target horizon slice. (2) We cut a cube of size  $w \times w \times t$  by using a  $w \times w$  horizontal square window to the volume obtained in step 1 and the window centers at each point of the target slice. Note that we discard the points near the boundaries and can obtain  $(n_1 - w) \times (n_2 - w)$  cubes. (3) A seismic cube and an attribute cube extracted from the same position form a training sample pair, and the attribute cube can be any of four seismic attributes. Thus, we can obtain  $4 \times (n_1 - w) \times (n_2 - w)$  pairs of cubes to generate training data sets altogether.

In this work, the size of the target slice is  $576 \times 1767$ , the time window size is nine, and the horizontal square window size is  $17 \times 17$ . In fact, we do not need to feed all training pairs to the network, and only using 1/16 of all sample pairs can not only achieve almost the same performance as using all the data but also

greatly reduce graphics processing unit memory and computational costs during the training.

## CONTRASTIVE LEARNING FOR SEISMIC FACIES CHARACTERIZATION

Based on the prepared training data sets without any labeling, we present a workflow of seismic facies characterization by using an unsupervised learning method inspired by the idea of contrastive learning (Li et al., 2021b).

The basic idea of contrastive learning is to learn an embedding space where the similarities of different augmentations from the same image are maximized, and the similarities of different images are minimized (Wang et al., 2021). They usually use a siamese network to learn the feature matrix of data pairs constructed through a variety of data augmentations (Li et al., 2021b), and then update the network by minimizing contrastive loss functions. However, using data augmentations to construct pairs shows poor performance in seismic facies analysis because the seismic cubes are too small compared with nature images. Fortunately, a seismic attribute can be regarded as a special augmentation that also

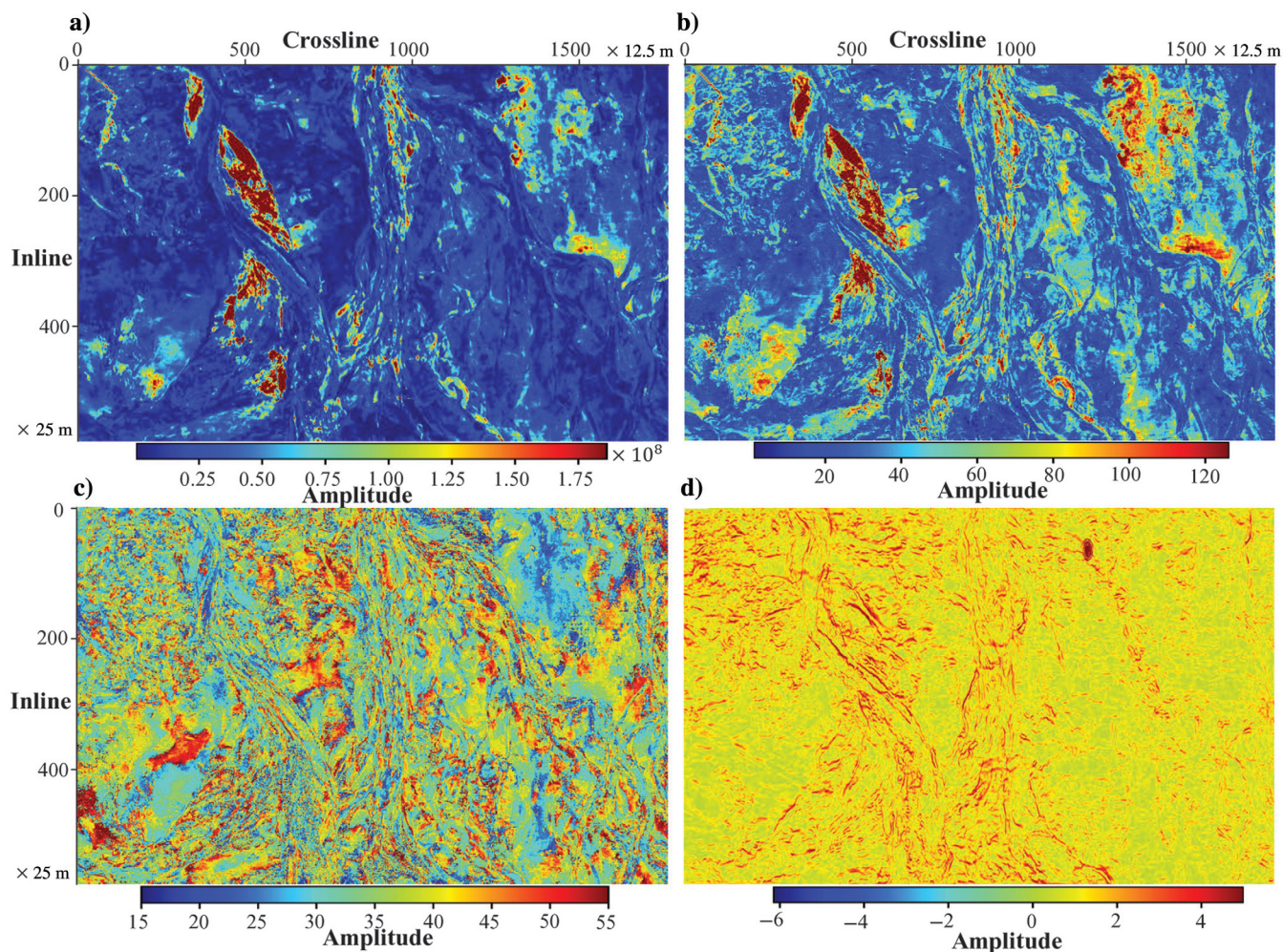


Figure 4. The four seismic attributes used in this study. (a) Coherent energy, (b) peak spectral magnitude, (c) peak spectral frequency, and (d) curvedness.

belongs to a transformation of seismic data, and we can replace data augmentations with different attributes.

**Network**

The architecture used in our work is shown in Figure 6, which consists of two parts: a shared encoder followed by two separate projection heads. Given the two inputs of seismic cubes  $I^a$  and multiattribute cubes  $I^b$ , the encoder network  $f$  first maps them to a common feature space, i.e.,  $h_a = f(I_a)$  and  $h_b = f(I_b)$ . Then, the two multilayer perceptron (MLP) heads  $\sigma_1$  and  $\sigma_2$  are used to, respectively, project the extracted features  $h^a$  and  $h^b$  into two different embedding space. Specifically,  $\sigma_1$  is a cluster contrastive head, which projects  $\{h_a, h_b\}$  as two probability matrices  $\{Y_a, Y_b\} \in \mathbb{R}^{N \times C}$  that directly reveal the clustering result of inputs, where  $N$  denotes the batch size and  $C$  denotes the number of clus-

ters;  $\sigma_2$  is an instance contrastive head, which projects  $\{h_a, h_b\}$  as two new feature matrices  $\{Z_a, Z_b\} \in \mathbb{R}^{N \times M}$  and plays a role of promoting the network to learn better features, where  $M$  is the feature length of each instance. We experimentally validate it in the ‘‘Ablation study’’ section.

In this work, we adopt 2D ResNet18 (He et al., 2016) as our encoder network. The MLPs consist of a fully connected layer and a rectified linear unit activation followed by another fully connected layer. The MLP of the cluster contrastive head is followed by an additional softmax function to produce the output whose  $i$ th element represents the probability belonging to cluster  $i$ . We consider the time slices of seismic cubes as  $t$  channels that are input to the 2D convolutional network. In the application step, we input only the seismic cubes to the trained network and apply an argmax operation to the feature matrix of the cluster contrastive head to compute clustering results of seismic facies.

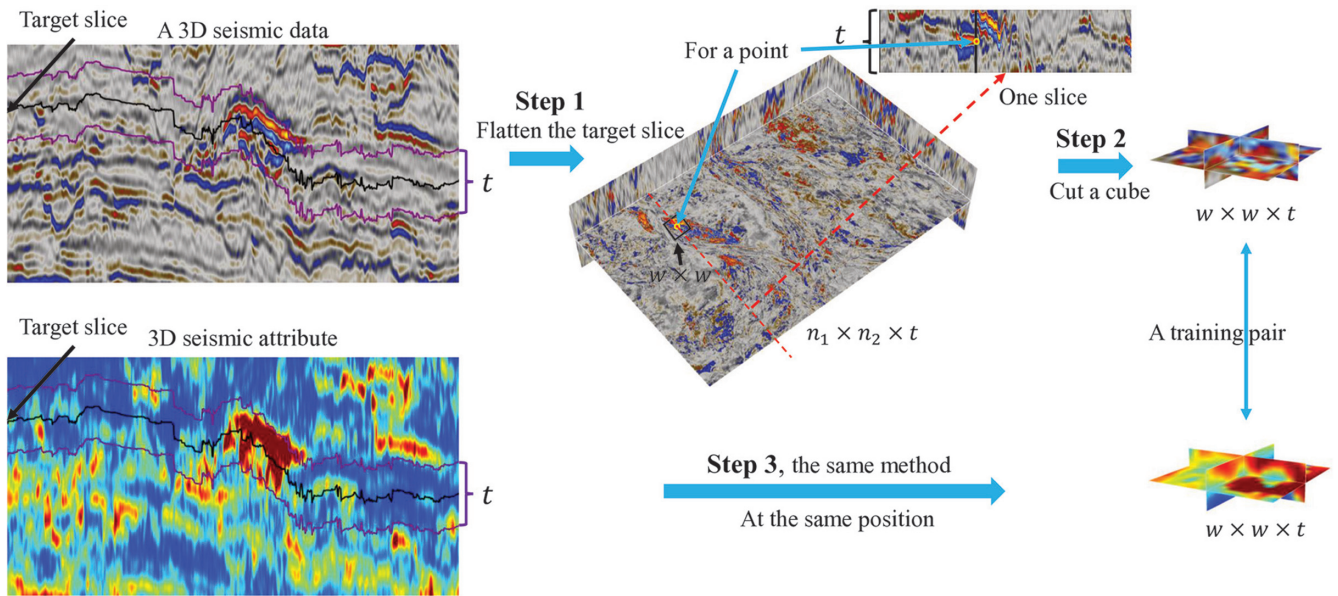


Figure 5. The workflow to generate training data.

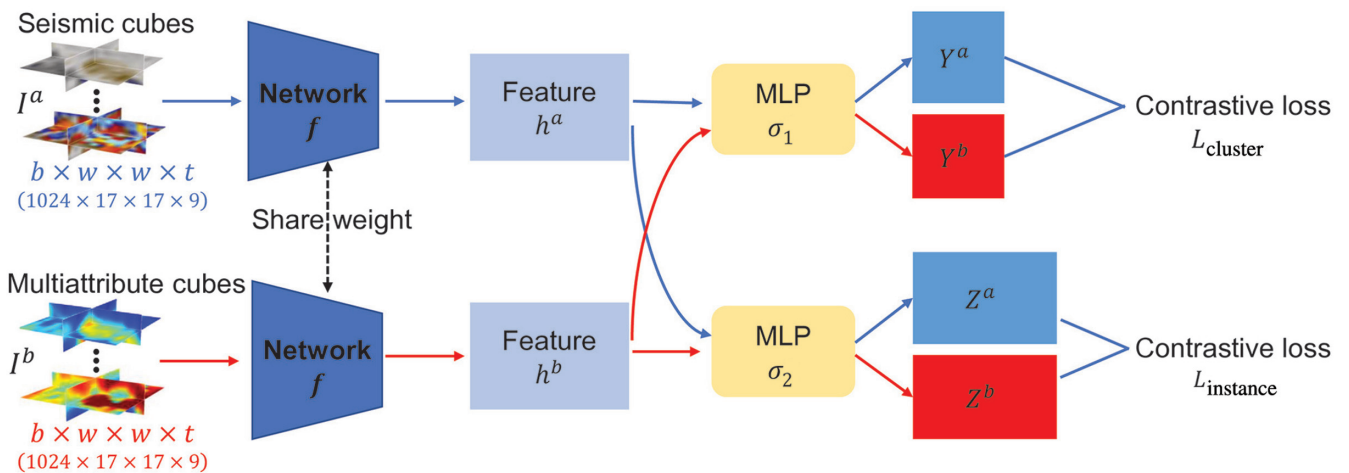


Figure 6. The framework used in this work.

**Loss function**

For a feature matrix extracted from the network, its rows can be regarded as the cluster assigned probabilities of instances, and its columns can be interpreted as the cluster distributions over instances. We use two contrastive loss functions to maximize the similarities of positive pairs while minimizing the similarities of negative pairs over the row and the column space, respectively. We treat the cubes from the same position as positive pairs and the cubes from a different position as negative pairs. Therefore, given a batch of size  $N$ , the total number of samples is  $2N$ . Each sample can form one positive pair with the corresponding seismic or attributes cube, and  $2N - 2$  negative pairs with other cubes. We use cosine distance to measure the pair-wise similarity.

For the instance contrastive head, we have its feature matrices  $Z^a, Z^b \in \mathbb{R}^{N \times M}$ , and we use  $z_i^a$  to denote the  $i$ th row of feature matrix  $Z^a$ . We can define the loss of a seismic cube as

$$\ell_i^a = -\log \frac{\exp(s(z_i^a, z_i^b)/\tau_I)}{\sum_{j=1}^N [\exp(s(z_i^a, z_j^a)/\tau_I) + \exp(s(z_i^a, z_j^b)/\tau_I)]}, \quad (2)$$

where  $\tau_I$  is the temperature parameter,  $s(z_i^a, z_j^b)$  means the cosine similarity between  $z_i^a$  and  $z_j^b$ , and the loss of an attribute cube  $\ell_i^b$  is defined in the same way. In this formula, the numerator denotes the

similarity of the positive pair, and the denominator denotes the total similarities of all pairs which could be approximated as the similarities of negative pairs when the batch size is large. This loss will decrease when the positive pairs are attracted and the negative pairs are separated. Thus, the instance contrastive loss over every instance can be written as

$$\mathcal{L}_{\text{instance}} = \frac{1}{2N} \sum_{i=1}^N (\ell_i^a + \ell_i^b). \quad (3)$$

Note that the feature matrices produced in the cluster contrastive head  $Y^a, Y^b \in \mathbb{R}^{N \times C}$ , each row of  $Y^a, Y^b$  should tend to be one-hot, and the columns can be interpreted as the cluster distributions over instances. Therefore, all columns of each matrix should differ from each other and the similarity of the columns  $Y_i^a$  and  $Y_i^b$  should be maximized. In other words, given a column  $Y_i^a$ , it can form one positive pair with  $Y_i^b$  and  $2C - 2$  negative pairs with other columns. In this sense, similar to  $\ell_i^a$ , the cluster contrastive loss for a seismic cube can be defined as

$$\hat{\ell}_i^a = -\log \frac{\exp(s(y_i^a, y_i^b)/\tau_C)}{\sum_{j=1}^C [\exp(s(y_i^a, y_j^a)/\tau_C) + \exp(s(y_i^a, y_j^b)/\tau_C)]}, \quad (4)$$

Figure 7. Training record of (a) the instance contrastive loss, (b) the cluster contrastive loss, and (c) the total loss.

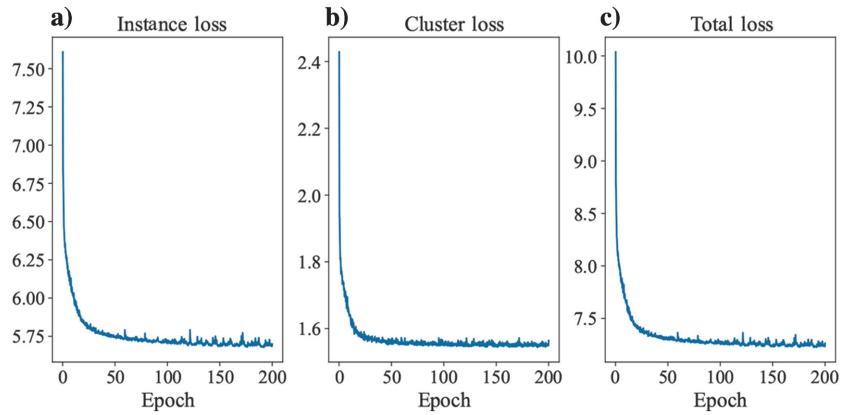
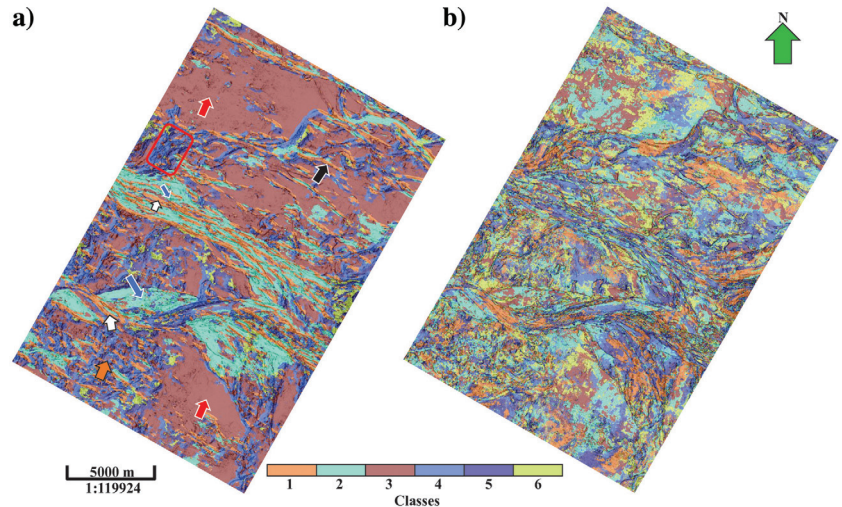


Figure 8. Classification of the turbidite channel system based on contrastive learning. (a) The result of our method, each cluster is assigned a color and (b) the result produced by a VAE network followed by  $k$ -means. The sobel-filter similarity (Chopra et al., 2014; Phillips and Fomel, 2017) is overlaid in black to show the boundaries.



where  $\tau_C$  is a temperature parameter. The cluster contrastive loss over all columns is defined as

$$\mathcal{L}_{\text{cluster}} = \frac{1}{2C} \sum_{i=1}^C (\hat{\epsilon}_i^a + \hat{\epsilon}_i^b) - H(Y), \quad (5)$$

where  $H(Y)$  is the entropy of cluster assignment probabilities within a batch to avoid the trivial solution.

Therefore, the total loss is defined as

$$\mathcal{L} = \mathcal{L}_{\text{instance}} + \mathcal{L}_{\text{cluster}}. \quad (6)$$

These two parts of loss simultaneously optimize the network in training.

## RESULT AND ABLATION STUDY

After constructing the whole workflow, we use the training pairs to train the network until it converges. Then, we compute a seismic facies map by feeding all of the seismic cubes, one for each pixel on the horizon slice, into the trained network. In addition, we discuss some ablation experiments to further understand our work.

### Result

We train our model with an Adam optimizer (Kingma and Ba, 2014) over 200 epochs. The learning rate is initialized to  $3 \times 10^{-4}$  without any decay. Contrastive learning usually requires a large batch size to cover a rich set of negative samples to achieve good performance. Benefiting from the small size of cubes, we can set the batch size to 1024. As mentioned previously, we use only approximately 250,000 cube pairs (i.e., 1/16 of all possible pairs) to train our network. The number of seismic facies classes is set to six, which is larger than the five classes shown in Figure 2. In fact, the exact number of facies classes is unknown because some facies are few in the turbidite system but they do exist. Thus, we only focus on the main seismic facies, especially channel facies and set a slightly larger number. The inability to determine the number of facies classes is a limitation of our method and most of the unsupervised learning methods. Finally, we recorded the values of the loss functions at each epoch shown in Figure 7. The two loss function curves converge quickly at approximately 100 epochs. The whole training process took up only 3G memory of NVIDIA Tesla V100 and was completed in 4 h.

The classification result of the turbidite channel system is shown in Figure 8a where the sobel-filter similarity is overlaid in black to highlight the boundaries. In Figure 8, each pixel in the seismic horizon slice is painted with the same color as the class to which it belongs. The first class marked in shallow orange is identified as the mud-filled channel. Two main slope channels (white arrows) are classified as the first class that converge downstream. The second class (light green) denotes the sand-filled channels. The upper blue arrow denotes the older and probably sand-filled channels within the multistoried stacked channel. The lower blue arrow also indicates an older, high amplitude, sand-filled channel that developed earlier than the mud-filled cutting through it. These two classes show clear lineaments along the sinuous channels. Most of the sinuous channel complexes are classified as mud-filled channels (black arrow), but some are still marked in light green, i.e., sand-filled. The third class characterized by reddish-brown denotes the slope fans, which is continuous and wide-

spread in the northeast and southwest (red arrows). The fourth and fifth classes are discontinuous and widespread in the northwest and the area between the two main slope channels. We identify them as some messy turbidite current or slump deposits. Finally, the sixth class characterized by lemon green is few but widely distributed in the whole turbidite system. This class also is probably turbidite current or some unknown facies.

We then compared the autoencoder-based method. We first follow the operation of Zhu et al. (2022) and exact a seismic trace and copy it 17 times to form a small 2D vertical seismic section where we apply 2D convolution layers. Then, we feed the seismic sections to a variational autoencoder (VAE) network and set the number of latent features as six. Finally, we obtain the facies map shown in Figure 8b by inputting the latent features into  $k$ -means. The facies map is noise and discontinuous, and many points are assigned incorrect clusters.

Figure 9 shows the classification results of two previous works by Zhao et al. (2016) and Wallet and Hardisty (2019). Zhao et al. (2016) leverage SOM to obtain a seismic facies map shown in Figure 9a where they paint each point with a color in a 2D colorbar. Thus, this method does not explicitly specify the exact number of facies classes, and it implicitly determines whether two points are in the same class by color difference. Wallet and Hardisty (2019) feed the two SOM latent axes produced by Zhao et al. (2016) to a Gaussian mixture model (GMM) to achieve a seismic facies map (Figure 9b), and they set the number of seismic facies as seven. Compared with the two previous works, our result is generally consistent with their results. In addition, our result is more continuous and stable, especially in the slope fans (the two red arrows). However, some areas still fall into the incorrect classes in our result; for instance, the sinuous channel complexes positioned near the right main channel (the red rectangle) are assigned turbidite current or slump deposits (blue) instead of the two kinds of the channel (shallow orange or light green).

### Ablation study

Three ablation experiments are carried out to better understand the importance of using 3D seismic cubes, seismic attributes, and the instance contrastive head.

We first evaluate the importance of 3D seismic cubes by replacing them with waveform-based sections (including seismic and corresponding attributes) and 2D seismic slices to train our network.

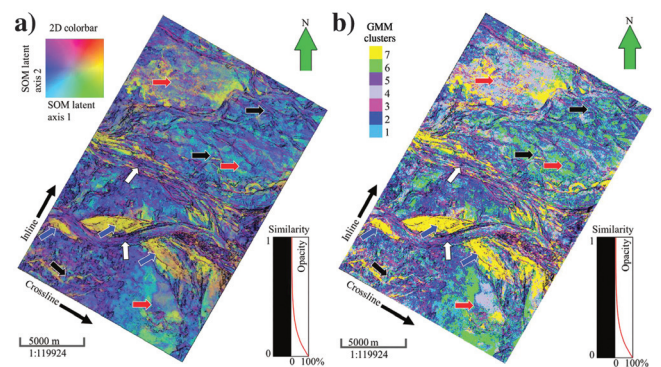


Figure 9. Previous works using (a) SOM (Zhao et al., 2016) and (b) GMM (Wallet and Hardisty, 2019). This figure is from Wallet and Hardisty (2019).

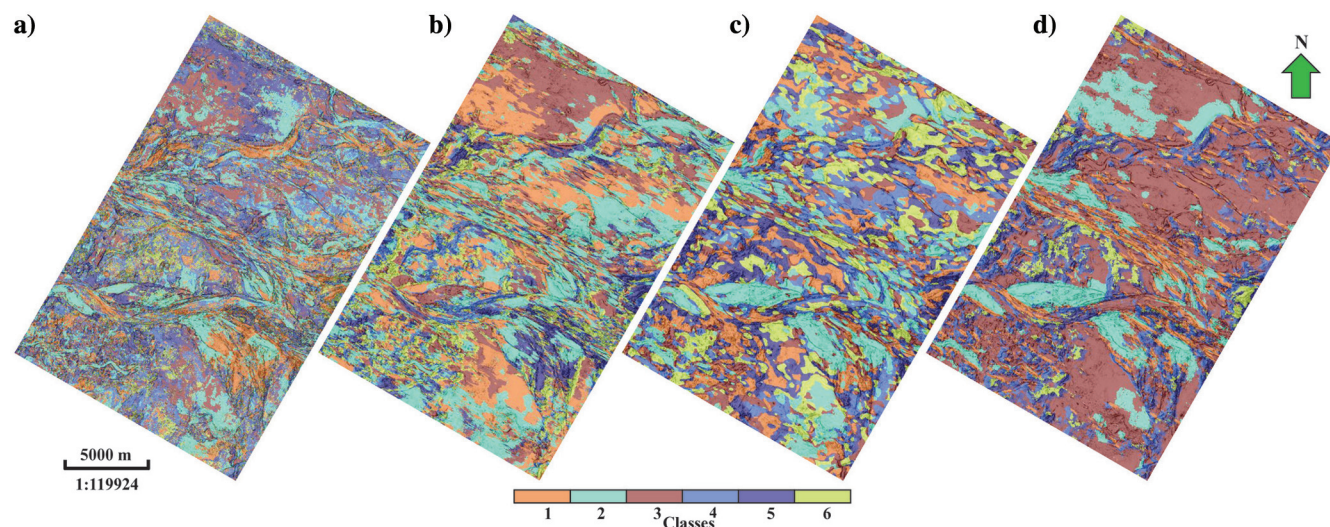


Figure 10. The result of ablation experiments (a) using seismic traces instead of 3D seismic cubes, (b) using 2D slices instead of 3D seismic cubes, (c) using image augmentations instead of seismic attributes, and (d) discard instance contrastive head. The sobel-filter similarity is overlaid in black to show the boundaries.

We follow the aforementioned VAE-based method and extract trace-based sections as inputs, i.e., the small 2D vertical seismic (or attributes) sections. The result is shown in Figure 10a. It is almost as discontinuous and noisy as the VAE-based method (Figure 8b), but more reasonable such as the sinuous channel complexes and the slope fans in the northeast. Then, we extract 2D slices with the size of  $17 \times 17$  as inputs, and the result is shown in Figure 10b. This facies map is more continuous than Figure 10a, while the classification result is still inaccurate.

Seismic attributes also play a key role in our method. To demonstrate this point, we replace the seismic attributes with sample augmentations to construct training pairs in the ablation experiment. Due to the size of cubes being too small, we only implement some simple data augmentations including flipping and rotating. Figure 10c exhibits the result of this experiment. This result is more continuous and stable than Figure 10a because we use 3D cubes instead of traces, while it still shows more incorrect classifications than the result (Figure 8a) with attributes.

Finally, we remove the instance contrastive head and only keep the cluster contrastive head. The result is shown in Figure 10d. This result is more continuous than Figure 10a and more accurate than Figure 10c. However, many areas are still assigned the incorrect clusters such as the slope fans in the northeast, the sinuous channel complexes in the southeast, and the turbidite current in the northwest.

## CONCLUSION

We have discussed a workflow based on contrastive learning for automatic seismic facies characterization. This method is a one-stage, end-to-end, and unsupervised fashion without any manual label. Unlike the mostly existing unsupervised works, we exact 3D seismic cubes instead of seismic traces or their variants as the network inputs to impose lateral consistency and avoid discontinuous and noisy results. In addition, we feed both seismic and seismic attributes into the network rather than only one of them and the seismic attributes can be regarded as geologic constraints for networks. The two types of inputs are processed by a contrastive

learning framework, which maximizes the similarities of different seismic and attributes cubes from the same position and minimizes the cubes from a different position. This framework contains two contrastive heads including the basic cluster contrastive head and the other instance contrastive head which are trained by two contrastive loss functions, respectively. Finally, we have demonstrated the effectiveness of the method by applying it to a turbidite system in Canterbury Basin, offshore New Zealand, and the three ablation studies help us to understand the effectiveness of our method.

Some limitations remain in our method. One is that the cluster number cannot be automatically determined and must be manually specified. Moreover, the result leaves room for improvement, such as the sinuous channel complexes are unreasonably classified near the main channel. For future work, the result may be improved by leveraging transformer instead of convolutional neural network-based contrastive learning methods. In addition, we plan to take location information into the network because two cubes should be similar and belong to the same class if they are only one or two pixels shifted.

## ACKNOWLEDGMENTS

This research was supported by the National Science Foundation of China under grant nos. 42050104 and 41874164 and CNPC Innovation Found under grant no. 2020D-5007-0303. We used the Attribute-Assisted Seismic Processing & Interpretation software to compute the seismic attributes used in this work. We would like to thank New Zealand Petroleum and Minerals for generously providing the Waka-3D seismic survey to the public which can be accessed at <https://wiki.seg.org/wiki/Waka-3D>.

## DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be obtained by contacting the corresponding author.



## REFERENCES

- Barnes, A. E., and K. J. Laughlin, 2002, Investigation of methods for unsupervised classification of seismic data: 72nd Annual International Meeting, SEG, Expanded Abstracts, 2221–2224, doi: [10.1190/1.1817152](https://doi.org/10.1190/1.1817152).
- Caron, M., P. Bojanowski, A. Joulin, and M. Douze, 2018, Deep clustering for unsupervised learning of visual features: Proceedings of the European Conference on Computer Vision, 132–149.
- Chang, J., L. Wang, G. Meng, S. Xiang, and C. Pan, 2017, Deep adaptive image clustering: Proceedings of the IEEE International Conference on Computer Vision, 5879–5887.
- Chen, L.-C., G. Papandreou, F. Schroff, and H. Adam, 2017, Rethinking atrous convolution for semantic image segmentation: arXiv preprint, arXiv:1706.05587.
- Chen, T., S. Kornblith, M. Norouzi, and G. Hinton, 2020, A simple framework for contrastive learning of visual representations: International Conference on Machine Learning, PMLR, 1597–1607.
- Chopra, S., R. Kumar, and K. J. Marfurt, 2014, Seismic discontinuity attributes and Sobel filtering: 84th Annual International Meeting, SEG, Expanded Abstracts, 1624–1628, doi: [10.1190/segam2014-0465.1](https://doi.org/10.1190/segam2014-0465.1).
- Chopra, S., and K. J. Marfurt, 2007, Seismic attributes for prospect identification and reservoir characterization: SEG and EAGE.
- de Matos, M. C., P. L. Osorio, and P. R. Johann, 2007, Unsupervised seismic facies analysis using wavelet transform and self-organizing maps: Geophysics, **72**, no. 1, P9–P21, doi: [10.1190/1.2392789](https://doi.org/10.1190/1.2392789).
- Duan, Y., X. Zheng, L. Hu, and L. Sun, 2019, Seismic facies analysis based on deep convolutional embedded clustering: Geophysics, **84**, no. 6, IM87–IM97, doi: [10.1190/geo2018-0789.1](https://doi.org/10.1190/geo2018-0789.1).
- He, K., H. Fan, Y. Wu, S. Xie, and R. Girshick, 2020, Momentum contrast for unsupervised visual representation learning: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9729–9738.
- He, K., X. Zhang, S. Ren, and J. Sun, 2016, Deep residual learning for image recognition: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778.
- Kaur, H., N. Pham, S. Fomel, Z. Geng, L. Decker, B. Gremillion, M. Jervis, R. Abma, and S. Gao, 2021, A deep learning framework for seismic facies classification: First International Meeting for Applied Geoscience & Energy, SEG, Expanded Abstracts, 1420–1424, doi: [10.1190/segam2021-3583072.1](https://doi.org/10.1190/segam2021-3583072.1).
- Kingma, D. P., and J. Ba, 2014, Adam: A method for stochastic optimization: arXiv preprint, arXiv:1412.6980.
- Kuzma, H. A., and J. W. Rector, 2005, The Zoeppritz equations, information theory, and support vector machines: 75th Annual International Meeting, SEG, Expanded Abstracts, 1701–1704, doi: [10.1190/1.2148025](https://doi.org/10.1190/1.2148025).
- La Marca, K., and H. Bedle, 2022, Deepwater seismic facies and architectural element interpretation aided with unsupervised machine learning techniques: Taranaki Basin, New Zealand: Marine and Petroleum Geology, **136**, 105427, doi: [10.1016/j.marpetgeo.2021.105427](https://doi.org/10.1016/j.marpetgeo.2021.105427).
- Li, F., H. Zhou, Z. Wang, and X. Wu, 2021a, ADDCNN: An attention-based deep dilated convolutional neural network for seismic facies analysis with interpretable spatial-spectral maps: IEEE Transactions on Geoscience and Remote Sensing, **59**, 1733–1744, doi: [10.1109/TGRS.2020.2999365](https://doi.org/10.1109/TGRS.2020.2999365).
- Li, K., Z. Liu, B. She, G. Hu, and C. Song, 2019, Orthogonal deep autoencoders for unsupervised seismic facies analysis: 89th Annual International Meeting, SEG, Expanded Abstracts, 2023–2028, doi: [10.1190/segam2019-3213841.1](https://doi.org/10.1190/segam2019-3213841.1).
- Li, Y., P. Hu, Z. Liu, D. Peng, J. T. Zhou, and X. Peng, 2021b, Contrastive clustering: AAAI Conference on Artificial Intelligence.
- Liu, M., M. Jervis, W. Li, and P. Nivlet, 2020, Seismic facies classification using supervised convolutional neural networks and semisupervised generative adversarial networks: Geophysics, **85**, no. 4, O47–O58, doi: [10.1190/geo2019-0627.1](https://doi.org/10.1190/geo2019-0627.1).
- Mitchell, J., and H. Neil, 2012, Os20/20 Canterbury — Great South Basin TAN1209 voyage report: National Institute of Water and Atmospheric Research Ltd.
- Phillips, M., and S. Fomel, 2017, Plane-wave sobel attribute for discontinuity enhancement in seismic images: Geophysics, **82**, no. 6, WB63–WB69, doi: [10.1190/geo2017-0233.1](https://doi.org/10.1190/geo2017-0233.1).
- Puzyrev, V., and C. Elders, 2020, Unsupervised seismic facies classification using deep convolutional autoencoder: arXiv preprint, arXiv:2008.01995.
- Qi, J., T. Lin, T. Zhao, F. Li, and K. Marfurt, 2016, Semisupervised multi-attribute seismic facies analysis: Interpretation, **4**, no. 1, SB91–SB106, doi: [10.1190/INT-2015-0098.1](https://doi.org/10.1190/INT-2015-0098.1).
- Qian, F., M. Yin, X.-Y. Liu, Y.-J. Wang, C. Lu, and G.-M. Hu, 2018, Unsupervised seismic facies analysis via deep convolutional autoencoders: Geophysics, **83**, no. 3, A39–A43, doi: [10.1190/geo2017-0524.1](https://doi.org/10.1190/geo2017-0524.1).
- Ronneberger, O., P. Fischer, and T. Brox, 2015, U-net: Convolutional networks for biomedical image segmentation: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 234–241.
- Roy, A., A. S. Romero-Peláez, T. J. Kwiatkowski, and K. J. Marfurt, 2014, Generative topographic mapping for seismic facies estimation of a carbonate wash, Veracruz Basin, southern Mexico: Interpretation, **2**, no. 1, SA31–SA47, doi: [10.1190/INT-2013-0077.1](https://doi.org/10.1190/INT-2013-0077.1).
- Sabeti, H., and A. Javaherian, 2009, Seismic facies analysis based on *k*-means clustering algorithm using 3D seismic attributes: Shiraz 2009-1st EAGE International Petroleum Conference and Exhibition, cp-125, doi: [10.3997/2214-4609.20145876](https://doi.org/10.3997/2214-4609.20145876).
- van Engelen, J. E., and H. H. Hoos, 2020, A survey on semi-supervised learning: Machine Learning, **109**, 373–440, doi: [10.1007/s10994-019-05855-6](https://doi.org/10.1007/s10994-019-05855-6).
- Wallet, B. C., M. C. de Matos, J. T. Kwiatkowski, and Y. Suarez, 2009, Latent space modeling of seismic data: An overview: The Leading Edge, **28**, 1454–1459, doi: [10.1190/1.3272700](https://doi.org/10.1190/1.3272700).
- Wallet, B. C., and R. Hardisty, 2019, Unsupervised seismic facies using Gaussian mixture models: Interpretation, **7**, no. 3, SE93–SE111, doi: [10.1190/INT-2018-0119.1](https://doi.org/10.1190/INT-2018-0119.1).
- Wang, F., T. Kong, R. Zhang, H. Liu, and H. Li, 2021, Self-supervised learning by estimating twin class distributions: arXiv preprint, arXiv:2110.07402.
- West, B. P., S. R. May, J. E. Eastwood, and C. Rossen, 2002, Interactive seismic facies classification using textural attributes and neural networks: The Leading Edge, **21**, 1042–1049, doi: [10.1190/1.1518444](https://doi.org/10.1190/1.1518444).
- Wrona, T., I. Pan, R. L. Gawthorpe, and H. Fossen, 2018, Seismic facies analysis using machine learning: Geophysics, **83**, no. 5, O83–O95, doi: [10.1190/geo2017-0595.1](https://doi.org/10.1190/geo2017-0595.1).
- Zhang, B., T. Zhao, X. Jin, and K. J. Marfurt, 2015, Brittleness evaluation of resource plays by integrating petrophysical and seismic data analysis: Interpretation, **3**, no. 2, T81–T92, doi: [10.1190/INT-2014-0144.1](https://doi.org/10.1190/INT-2014-0144.1).
- Zhang, H., T. Chen, Y. Liu, Y. Zhang, and J. Liu, 2021, Automatic seismic facies interpretation using supervised deep learning: Geophysics, **86**, no. 1, IM15–IM33, doi: [10.1190/geo2019-0425.1](https://doi.org/10.1190/geo2019-0425.1).
- Zhang, Y., Y. Liu, H. Zhang, and H. Xue, 2019, Seismic facies analysis based on deep learning: IEEE Geoscience and Remote Sensing Letters, **17**, 1119–1123, doi: [10.1109/LGRS.2019.2941166](https://doi.org/10.1109/LGRS.2019.2941166).
- Zhao, B., H.-W. Zhou, and F. Hilterman, 2005, Fizz and gas separation with SVM classification: 75th Annual International Meeting, SEG, Expanded Abstracts, 297–300, doi: [10.1190/1.2144325](https://doi.org/10.1190/1.2144325).
- Zhao, T., V. Jayaram, A. Roy, and K. J. Marfurt, 2015, A comparison of classification techniques for seismic facies recognition: Interpretation, **3**, no. 4, SAE29–SAE58, doi: [10.1190/INT-2015-0044.1](https://doi.org/10.1190/INT-2015-0044.1).
- Zhao, T., F. Li, and K. J. Marfurt, 2017, Constraining self-organizing map facies analysis with stratigraphy: An approach to increase the credibility in automatic seismic facies classification: Interpretation, **5**, no. 2, T163–T171, doi: [10.1190/INT-2016-0132.1](https://doi.org/10.1190/INT-2016-0132.1).
- Zhao, T., J. Zhang, F. Li, and K. J. Marfurt, 2016, Characterizing a turbidite system in Canterbury Basin, New Zealand, using seismic attributes and distance-preserving self-organizing maps: Interpretation, **4**, no. 1, SB79–SB89, doi: [10.1190/INT-2015-0094.1](https://doi.org/10.1190/INT-2015-0094.1).
- Zhu, D., J. Cui, Y. Li, Z. Wan, and L. Li, 2022, Adaptive Gaussian mixture model and convolution autoencoder clustering for unsupervised seismic waveform analysis: Interpretation, **10**, no. 1, T181–T193, doi: [10.1190/INT-2021-0087.1](https://doi.org/10.1190/INT-2021-0087.1).

Biographies and photographs of the authors are not available.