Petroleum Science 21 (2024) 1569–1583

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

An improved deep dilated convolutional neural network for seismic facies interpretation

Na-Xia Yang [a], Guo-Fa Li [a, *], Ting-Hui Li [b], Dong-Feng Zhao [a], Wei-Wei Gu [a]

[a] State Key Laboratory of Petroleum Resources and Engineering, CNPC Key Lab of Geophysical Exploration, China University of Petroleum (Beijing), Beijing, 102249, China
[b] Dagang Branch, Research Institute, BGP, Tianjin, 300280, China

ARTICLE INFO

Article history:
Received 23 March 2023
Received in revised form
17 October 2023
Accepted 28 November 2023
Available online 2 December 2023

Edited by Jie Hao and Meng-Jiao Zhou

Keywords:
Seismic facies interpretation
Dilated convolution
Spatial pyramid pooling
Internal feature maps
Compound loss function

ABSTRACT

With the successful application and breakthrough of deep learning technology in image segmentation, there has been continuous development in the field of seismic facies interpretation using convolutional neural networks. These intelligent and automated methods significantly reduce manual labor, particularly in the laborious task of manually labeling seismic facies. However, the extensive demand for training data imposes limitations on their wider application. To overcome this challenge, we adopt the UNet architecture as the foundational network structure for seismic facies classification, which has demonstrated effective segmentation results even with small-sample training data. Additionally, we integrate spatial pyramid pooling and dilated convolution modules into the network architecture to enhance the perception of spatial information across a broader range. The seismic facies classification test on the public data from the F3 block verifies the superior performance of our proposed improved network structure in delineating seismic facies boundaries. Comparative analysis against the traditional UNet model reveals that our method achieves more accurate predictive classification results, as evidenced by various evaluation metrics for image segmentation. Obviously, the classification accuracy reaches an impressive 96%. Furthermore, the results of seismic facies classification in the seismic slice dimension provide further confirmation of the superior performance of our proposed method, which accurately defines the range of different seismic facies. This approach holds significant potential for analyzing geological patterns and extracting valuable depositional information.

1. Introduction

Seismic facies serve as comprehensive representations of sedimentary characteristics in seismic data. In terms of understanding the underground depositional environment and lithofacies distribution, seismic facies interpretation plays a crucial role in the analysis of geological conditions and prediction for oil and gas distribution. Manual seismic facies analysis heavily relies on interpreters, which is time-consuming and subjective and inevitably susceptible to non-objective factors. Hence, it is necessary to develop fast and effective automatic seismic facies interpretation schemes. In recent years, with the significant progress and breakthrough of deep learning technology in the field of computer vision, seismic facies classification based on deep learning approaches has greatly attracted the attention of researchers and has achieved substantial developments and applications.

The highly nonlinear nature of geophysical problems fits the application characteristics of deep learning technology, which can mine high-dimensional nonlinear relationships from data itself when specific mathematical expressions cannot be established. Taking advantage of this feature, deep learning approaches have found widespread applications across various disciplines (Saxena et al., 2021; Wang et al., 2021; Liu et al., 2021a). Particularly, convolutional neural networks (CNNs) have been demonstrated with good application effects in pattern recognition, target detection and image processing. Due on the similarity between seismic data interpretation and computer image processing, CNNs have been extensively utilized in seismic data interpretation. Compared to seismic data processing based on the deep learning technology, the combination of seismic data interpretation with deep learning

* Corresponding author.
E-mail address: lgfseismic@126.com (G.-F. Li).

https://doi.org/10.1016/j.petsci.2023.11.027

approaches shows stronger stability and migration generalization ability. Consequently, in recent few years, including fault detection (Liu et al., 2020b; Hu et al., 2022), salt dome identification (Amin and Deriche, 2016; Alfarhan et al., 2022; Chung et al., 2022), horizon tracking (Wang et al., 2022; Birnie and Hansteen, 2022; Shi et al., 2022), first-break picking (Duan and Zhang, 2020; Yuan et al., 2018; Han et al., 2022) and seismic facies classification (Puzyrev and Elders, 2021; Nasim et al., 2022; Chai et al., 2022; Liu et al., 2020a) have been combined and applied with deep learning methods. Among them, as the major goals of seismic stratigraphy, seismic facies classification is increasingly integrated with deep learning techniques.

Seismic facies analysis approaches can be broadly divided into two major categories: one is seismic-attributes-based analysis methods (Liu et al., 2019; Mirzakhanian and Hashemi, 2022; Qi et al., 2020), and the other is waveform-clustering-based classification approaches (Duan et al., 2019; Zhang et al., 2021; Feng et al., 2021). Compared with manual seismic facies interpretation, the introduction and application of various different seismic attributes in deep neural network models have significantly improved the efficiency of seismic facies classification. However, due to the numerous seismic attributes, interpretation work is heavy. In addition, the relationship between various seismic attributes and the target geological task is complicated. Consequently, Selecting the appropriate attributes and evaluating the classification results are very difficult. Although deep learning alleviates the difficulty in attributes selection and improves the reliability of analysis results to some extent. However, the main challenge for deep-learning-based seismic facies classification remains the selection of appropriate characterization. Waveform clustering-based methods offer an alternative to the limitations of seismic attribute analysis and open up new ideas for seismic facies classification. The most commonly used waveform clustering algorithms include K-means clustering (Song et al., 2021), principal component analysis (PCA) (Babikir et al., 2022) and the self-organizing map (SOM) (Liu et al., 2021b). These methods are unsupervised machine learning techniques that do not rely on the labeled data, but they usually come at the cost of lower classification accuracy and more difficult optimization process. For example, it is difficult to determine the most reasonable initial clustering center for the K-means clustering algorithm, which requires multiple iterations and the classification results are often the local optimal solution. With the advancement of image segmentation algorithm based on deep feature extraction (Yin et al., 2020; Liu et al., 2018; Xie et al., 2016; Zhao et al., 2015), the seismic facies recognition method of waveform classification based on the deep CNNs has become a research hot-spot. In contrast to unsupervised clustering algorithms, the deep CNNs architectures also obtain high-precision classification results even with small training samples.

Deep CNNs are widely used in image segmentation. Its characteristics, such as weight sharing and local connections, result in the network having a small number of parameters and being easy to train. The fully convolutional network (FCN) replaces the fully connected layers in the classical CNNs structure with convolutional layers, enabling it to handle any size of inputs and achieve pixel-level image segmentation (Shelhamer et al., 2017). FCN realizes the end-to-end image segmentation, but the details of its segmentation results still need improvement. To address this issue, the UNet structure is developed. UNet is a variant network of FCN, which follows the structural features of encoders and decoders. The entire network structure consists of three main parts: down-sampling, up-sampling, and skip connections. During down-sampling, the image is compressed, the receptive field continues to grow, and the convolutional kernel perceives more low-frequency information. In the up-sampling process, the information from each stage of down-sampling is integrated through skip connections. The image resolution is restored in combination with the information from each level. UNet excels at solving cell-level segmentation tasks and has been demonstrated to perform well on small training samples initially (Ronneberger et al., 2015). Different from FCN, which integrates features by adding the corresponding pixel values of feature maps during up-sampling. While the UNet integrates features by concatenation, which retains more dimensional and position information. Consequently, the UNet architecture can make full use of shallow and deep features, which is beneficial to image segmentation (Noh et al., 2015). In the geophysical application fields, UNet has been successfully used for fault detection (Wu et al., 2019) and seismic facies classification (Qian et al., 2018; Tolstaya and Egorov, 2022; Zhang et al., 2020).

To combine the advantages of UNet model for training on small samples with the potentials of supervised learning, we propose an improved deep dilated convolutional neural network with spatial pyramid pooling module (DCSPP) for seismic facies classification. Taking the UNet structures as the basic network configurations, the spatial pyramid pooling module (SPPM) is introduced into the network structure to improve the ability to obtain the global information of the model. Additionally, during the down-sampling process of the network, we strategically employ dilated convolutional layers to expand the receptive field. The pooling of different scales can perceive input information of different sizes, while the expansion of the receptive field allows to detect and segment the larger targets. By combining these two modules with UNet, the classification accuracy is further improved. The paper is organized as follows. First, we sort out the problem and explain the proposed network architectures. Then, we provide specific and detailed descriptions of the network's structure blocks, the features and principles of spatial pyramid pooling and dilated convolution modules. Next, the training process and performance of the proposed network are evaluated. We illustrate the advantages of the DCSPP network model by testing it on the public F3 seismic data. Subsequently, through the visualization of internal feature maps and training accuracy, we evaluate the network architectures using various image segmentation evaluation indicators, which objectively demonstrates the effectiveness and reliability of the proposed method. Finally, we conduct the tentative experiments with field data. The test results show that the proposed network structure achieves better classification effect and has certain practicality and generalization ability.

2. Methodology

Seismic facies classification can be regarded as a special application of image segmentation problems, which is essentially to take advantage of the differences between seismic waveforms. By dividing seismic profiles into multiple single-channel images and feeding them into the network structures, pixel-level classification prediction are achieved. The abundance of seismic data provides the prerequisites for the application of deep learning techniques. However, many redundant information contained in seismic data also weakens the application performance. Thus, this requires the network to extract useful information related to the classification target during the training process and abandon the interference of useless information. Deep CNNs meet this requirement through multiple layers of convolution and sampling. Taking the UNet architecture as an example, the two ends of the structure capture the detailed information of the image. The image is large at this time and many details are preserved. The middle part of the structure obtains the low-frequency information, which extracts more boundary information as the receptive field increases continuously

during down-sampling. During the up-sampling process to restore the image resolution, the skip connections retain the image information at all levels. Consequently, the network can remember all the information contained in the image. Therefore, UNet model achieves good classification results even with limited training samples.

The proposed DCSPP network not only inherits the advantages of UNet in small sample training, but also better handles the global and detailed information. The workflow of the DCSPP network structure for seismic facies classification includes the following steps: (1) Network structure construction. Firstly, seismic profiles are divided to make training samples. Then, the DCSPP network architecture is built by incorporating spatial pyramid pooling and dilated convolution modules into the UNet framework. (2) Network training and optimization. The training and validation set are split to train the network structure. Comparative tests of different network structures are conducted to verify the effectiveness of the proposed method. By deploying the hybrid dilated convolution (HDC) principle, the classification effect is further improved. During training, the network learns high-dimensional features associated with seismic facies and expands the receptive field. (3) Model visualization and result evaluation. From the two aspects of classification accuracy and the change of internal feature maps to visualize and understand the training process of the network intuitively. Furthermore, various indicators are used to evaluate the classification results of the network model. The evaluation parameters objectively show that the DCSPP network structure has better migration generalization ability. (4) Field data tests. Based on data augmentation and transfer learning technologies, the experimental test of seismic facies classification with field data is carried out. Additionally, the experimental results are analyzed to illustrate the effectiveness and applicability of the proposed network architectures.

3. Neural network architectures

3.1. UNet

The UNet model is a segmentation network derived from the FCN. As the pioneering work of semantic segmentation network, FCN has the following three main characteristics. (1) Replacing the fully connected layers at the end of the FCN model with convolutional layers allows the network to adapt to any size of input. Benefiting from the characteristics of weight sharing and local connection of convolutional operation, the computational complexity and model size are greatly reduced. (2) FCN employs the bilinear interpolation for up-sampling. The resolution of the image is restored and the pixel-wise prediction is provided. (3) If up-sampling operates directly based on the down-sampling feature maps, it will lose a lot of detailed information and blur image edges. Therefore, FCN introduces skip connections into the network architecture to integrate the shallow information through pixel addition, which helps to restore image details.

Due to the characteristics of small datasets and large single images in medical images, the segmentation effect of FCN is not satisfactory. However, the UNet structure has achieved remarkable results with only a few dozen medical image datasets. Fig. 1 displays a schematic of the UNet structure for seismic facies classification. It consists of two parts: encoding and decoding. Each layer of the encoding part contains two convolutional layers with a convolutional kernel size of 3×3 and a maximum pooling layer with a pooling kernel of 2×2. After the input seismic section is convoluted and sampled layer by layer in the encoding blocks, the image size is halved layer by layer while the number of channels is doubled layer by layer. The receptive field gradually increases and the number of channels is expanded into 1024. Consequently, the network can perceive and recognize the boundary and low-frequency information of the image, which is crucial and valuable for seismic facies analysis. Each block of the decoding part contains a transposed convolutional layer with a convolution kernel size of 2×2, a dimensional concatenation layer, and two convolutional layers with a convolutional kernel size of 3×3. During the up-sampling process, the size of the feature map gradually expands to the original size, the number of channels is compressed to 64. The network perceives more high-frequency and texture information. In addition, the concatenation layer in each block of up-sampling receives feature maps with different levels from the down-sampling process. This skip connection allows the network to utilize both shallow and deep features, which alleviates the problem of insufficient and incomplete information during up-sampling. Meanwhile, the feature map retains more dimensional and positional information through feature integration by concatenation approach, which is beneficial for pixel-level prediction and segmentation tasks.

In addition to the interpolation method used for up-sampling, up-pooling and transposed convolution are also common methods for up-sampling. Fig. 2 illustrates the up-sampling process of up-pooling after maximum pooling and the up-sampling process of transposed convolution. The up-pooling process restores the image size by filling with zero values, which preserves the position information of the maximum value within the pooling kernel range. However, it is evident that the up-pooling cannot restore the original data, the pooling process retains the main information while discarding some of it. As a result, there is a lack of information during the up-sampling process. The maximum information completion can only be achieved by filling in zero values to ensure the integrity of the information.

Transposed convolution, also known as deconvolution, is a special convolution used for up-sampling to expand the feature map. The up-sampling process is completed by flipping the convolution kernel and then convolution with the feature map. In addition to the same function of restoring image resolution as up-pooling, the biggest difference between transposed convolution and up-pooling is that the transposed convolution process has parameters to learn, so that an optimal up-sampling method can be learned. Theoretically, transposed convolution can realize up-pooling operations, as long as the parameters of the convolutional kernel are set appropriately. Moreover, transposed convolution itself is a convolution operation. Therefore, it combines all the advantages of convolution operations, which significantly benefits in reducing the model's complexity. Additionally, transposed convolution enables feature visualization during the image processing in the neural network, which facilitates the reasonable adjustment and optimization of the network structure according to the visualization results. Hence, we apply the transposed convolution operation to restore the resolution of the input seismic profiles and maximize the preservation and recovery of image information to provide complete image information for seismic facies analysis.

3.2. Spatial pyramid pooling

In the UNet structure, the fixed convolutional kernel size of 3×3 allows the network to extract target features with only a single receptive size. This limitation hinders the network from fully capturing important spatial information. The acquisition of global information is crucial for image segmentation problems. Therefore, we introduce the SPPM into the network structure, which aims to preserve the global information through pooling at different scales. The network considers the characteristics of multiple receiving

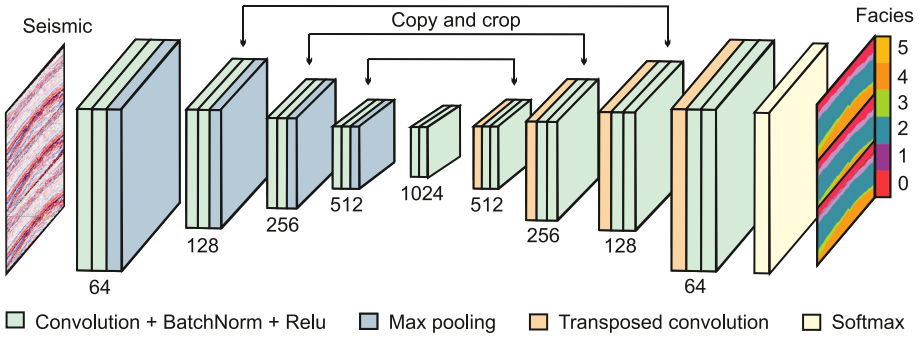

Fig. 1. Architectures of the UNet model. Rectangles of different colors denote the feature map corresponding to different operations. The number of channels is displayed at the bottom for each feature map. The network consists of four down-sampling convolutional blocks, an intermediate feature layer, four up-sampling transposed convolutional blocks, and a softmax probability calculation layer. The network maps the input seismic data to the seismic facies.
Note: BatchNorm = batch normalization; Relu = linear rectification function; Conv = convolution.

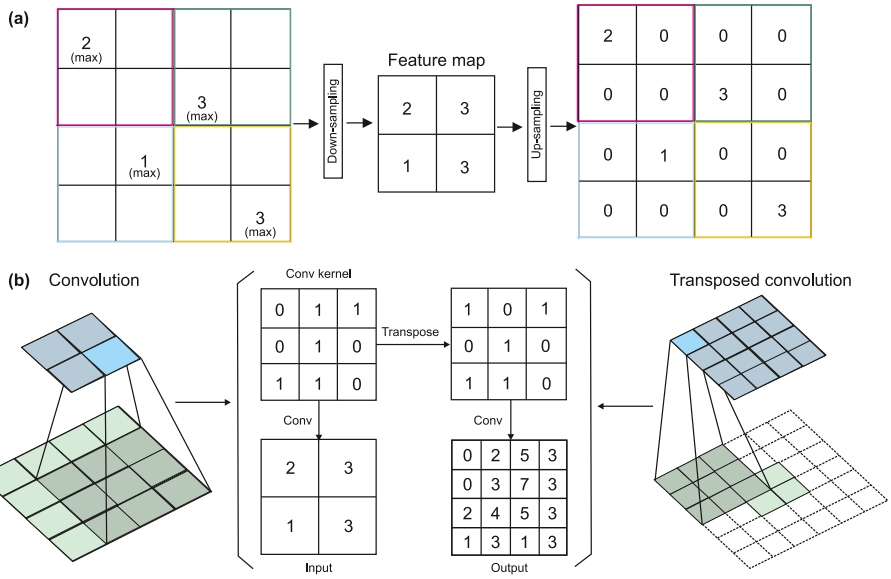

Fig. 2. The schematic of up-sampling operation. **(a)** Up-pooling and **(b)** transposed convolution. The same input feature map correspond to the different up-sampling results after two different up-sampling approaches. The pool size is 2 × 2 for the max-pooling and the kernel size is 3 × 3 for the transposed convolution. The up-pooling preserves the position information of the input data. The image recovery is achieved by filling in zero values within the feature map. While the transposed convolution is an up-sampling method achieved by flipping the convolution kernel, which is consistent with the convolution operation. Therefore, it inherits all the characteristics of convolution. In addition, the entire up-sampling is a learning and optimization process.

zones in parallel. Meanwhile, the information of different scales and different regions is integrated for target segmentation. This enables the network to extract and obtain comprehensive global information. Fig. 3 shows the UNet model structure with the SPPM, which adds four pooling operations of different sizes for 1 × 1, 2 × 2, 3 × 3, and 6 × 6 after the UNet model structure. First, the four

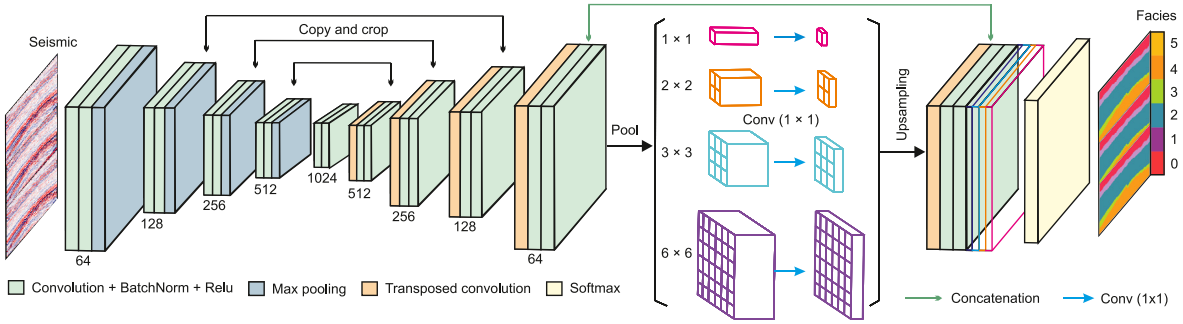

Fig. 3. Architectures of the Unet model with the spatial pyramid pooling module (SPPM). The information of different ranges and scales is extracted through pooling of four different scales of 1 × 1, 2 × 2, 3 × 3, and 6 × 6, and after convolution and up-sampling, it is merged with the output of UNet model. Finally, send the merged result to the softmax layer to output the classification probability of each seismic facies.

pooling feature maps will be individually convoluted by convolutional kernels with size of 1 × 1, the output channels are reduced to 1/4 of the input image. Then, after up-sampling, the original image information is restored. Finally, it is merged with the UNet's output as the input of the softmax layer to yield the classification probability.

The introduction of the SPPM allows the network to effectively capture global features. Information from various levels and regions is extracted and integrated, which is beneficial for improving recognition accuracy. In the SPPM, the size of the pooling kernel can be adjusted according to the application scenario and specific problem. The 1 × 1 pooling kernel represents global-scale pooling, while pooling at other levels produces different degrees of semantic and edge information. The combination of different pooling scales results in a more complete representation of spatial information. The features extracted by the UNet model and the features integrated by the SPPM are combined through concatenation. Consequently, the model extracts more precise seismic facies texture features and spatial boundary features simultaneously, which yields the high-precision seismic facies classification results.

3.3. Dilated convolution

The dilated convolution, originally proposed and applied in signal processing for wavelet decomposition, is now used in image semantic segmentation. It addresses the issue of reduced image resolution and information loss during down-sampling. When computational resources are too limited to increase the number and size of convolutional kernels, or when excessive convolution layers are stacked, it may lead to gradient problems. The dilated convolution expands the receptive field without adding additional computational parameters. Therefore, the output of each convolution contains a larger range of information. In real-time image segmentation problems like seismic facies classification, the network often requires a larger receptive field to perceive the edge and contour information. The dilated convolution introduces a new parameter to the convolutional layer called the dilation ratio, which defines the spacing of values when the convolutional kernel processes the data. By filling the hole points with a value of 0 and then performing convolution. The dilation ratio for the ordinary convolution is 1. Furthermore, the dilated convolution can replace the pooling module to capture a larger range of information without losing image size. Benefitting from above advantages, the dilated convolution is widely used in image semantic segmentation.

Fig. 4 displays the specific process of the ordinary convolution and the dilated convolution. An image of size 7 × 7 is operated with a normal and dilated convolution with kernel size of 3 × 3 and a dilated coefficient of 2. The output image sizes are 3 × 3 and 5 × 5, respectively. For the dilated convolution with a dilated coefficient of 2, which is similar to the ordinary convolution and same 9 values actually participate in the convolution operation, the remaining points have a weight of 0. Even though the size of the convolution kernel is 3 × 3, the effective receptive field of the dilated convolution is equivalent to a 5 × 5 convolutional kernel. This is the basic principle of the dilated convolution to expand the receptive field. With the same parameters and calculations, a 3 × 3 convolution kernel has a 5 × 5 receptive field. After two consecutive layers of 3 × 3 dilated convolution with a dilation ratio of 2, the receptive field reaches 13 × 13, which increases exponentially compared to the 5 × 5 receptive field size of ordinary convolution.

Indeed, the receptive field of dilated convolution can be increased exponentially. Does this mean that superimposing more dilated convolution layers can obtain a better segmentation effect? Actually, when using consecutive dilated convolutional layers with same dilation ratio, the shallow data used by a deep pixel is not continuous. There is an interval between each non-zero element, which is the gridding effect phenomenon. In other words, the deep pixels do not utilize all the pixel values within the receptive field and lose some detailed information. To address this issue, Wang et al. (2018) proposes the solution — HDC to deal with the lack of information caused by the gridding effect. The goal of HDC is to completely cover the square region of the underlying feature layer after a series of dilated convolutions. Meanwhile, there are no holes or missing edges inside this square region. By setting the dilation ratio to a jagged structure, HDC can capture both near and far information simultaneously. Following this principle, we design the dilated convolutional layers with dilation coefficients of {1, 2, 5} to expand the convolutional receptive fields. Fig. 5 shows the dilated convolution that satisfies the HDC condition. The kernel size of convolution is 3 × 3 and the dilation ratio is {1, 2, 5}.

3.4. DCSPP structures

Table 1 presents the structural composition of the proposed improved deep dilated convolutional neural network for seismic facies classification. It comprises five down-sampling convolutional blocks, four up-sampling transposed convolutional blocks, one SPPM and a softmax probability layer. During down-sampling, we incorporate the dilated convolution into two convolutional blocks to expand the receptive field and capture more boundary

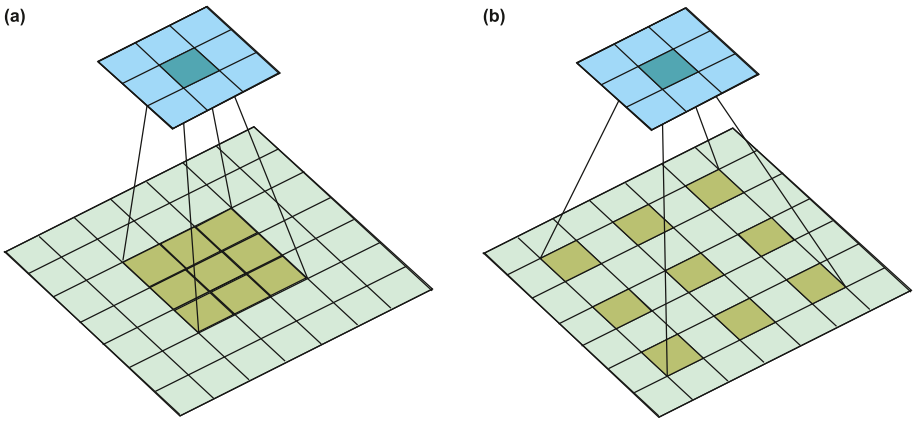

Fig. 4. The schematic of **(a)** regular convolution and **(b)** the dilated convolution. The size of the input image and convolutional kernel size are 7 × 7 and 3 × 3, respectively. The dilation ratio of the dilated convolution is 2. The effective receptive field of the convolution is expanded by supplementing zero elements around each convolution kernel parameter.

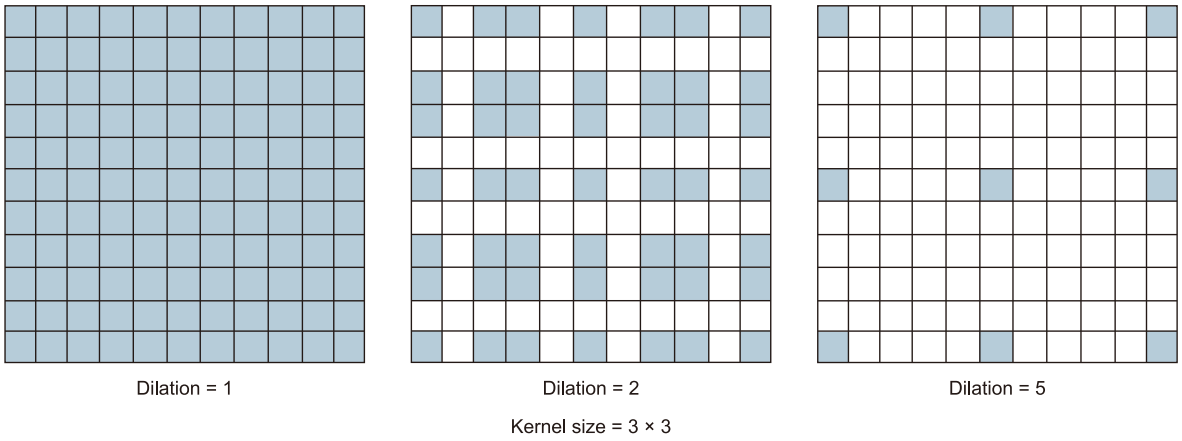

Fig. 5. Rules for the use and design of the dilated convolution: hybird dilated convolution (HDC). A dilated convolution block consists of a convolution kernel of size 3 × 3, and the design of parameter combinations with dilation ratio of {1, 2, 5} can effectively avoid the griding effect caused by pixel loss.

information. The specific operations contained in each block, the structural parameters, and the size of the feature map are all listed in the table. The whole model is built upon the UNet architectures. The PPM is introduced to obtain global information, and the dilated convolution is used to extract seismic facies boundaries in down-sampling process.

4. Model experiments

4.1. Data description

We use the F3 3D open seismic data in the Netherlands for model tests, it is a block in the Dutch sector of the North Sea. Alaudah et al. (2019) provides 401 training profiles, each line consists of 701 seismic traces and 255 sampling points (401*701*255), which includes all the data in the range of inlines [300, 700] and crosslines [300, 1000]. The test set contains 200 seismic profiles (200*701*255), which includes all the data in the range of inlines [100, 299] and crosslines [300, 1000]. Within the seismic data we used, 6 groups of lithofacies units are divided according to lithology: (1) Upper North Sea group: claystones and sandstones from Miocene to Quaternary; (2–3) Middle North Sea group/Lower North Sea group: sands, sand-stones, and claystones from Paleocene to Miocene; (4) Scruff group: clay-stones of Upper Jurassic and Lower Cretaceous; (5) Rijnland/Chalk: clay formations with sandstones of Upper Cretaceous./carbonates of Upper Cretaceous and Paleocene; (6) Zechstein group: evaporites and carbonates of Zechstein (Yan et al., 2020). Fig. 6 shows the training seismic data and the corresponding seismic facies distribution. These are used to evaluate the classification effect of the proposed network configurations.

4.2. Experimental process

We use the inline number as the index to convert the seismic data into a two-dimensional datasets. Then, we intercept the seismic data with a window size of 256 × 256. To maintain the spatial distribution of seismic facies while increasing the diversity of training samples, the overlapping step size is set as 0. In order to speed up the convergence speed and ensure the stability of training process, we normalize the amplitude of the seismic data to $[-1, 1]$. Fig. 7 displays the corresponding training samples. In addition to the inline section, we also construct training samples of the same size with CDP and time indexes respectively. Subsequently, we randomly select 320 samples as the training set and 80 samples as the validation set from the divided samples. The label contains a total of 6 different seismic facies and its one-hot encoding is {[1, 0, 0, 0, 0, 0], [0, 1, 0, 0, 0, 0], [0, 0, 1, 0, 0, 0], [0, 0, 0, 1, 0, 0], [0, 0, 0, 0, 1, 0], [0, 0, 0, 0, 0, 1]}. Fig. 8 displays the percentage of each seismic facies in training samples. It shows a very obvious unbalanced data distribution with a large difference in the proportion of each class of seismic facies. The percentage of category 2 seismic facies is approximately 45%. It is much higher than that of category 5, which accounts for nearly 3%. For such classification problems with severely imbalanced category under small samples, Wong et al. (2018) pointed that the exponential logarithmic loss (ELL) function has better performance than dice loss and cross-entropy. ELL is a kind of compound loss function, which pay more attention to samples that are not classified accurately under the cross-entropy loss function. It can adapt to highly unbalanced segmentation tasks.

$$L_{ELL} = w_{Dice}E[(-\ln(Dice_i))^{\gamma_{Dice}}] + w_{CE}E[w_l(-\ln(p_l(x)))^{\gamma_{CE}}]$$
$$Dice_i = \frac{2\left(\sum_x \delta_{il}(x)p_i(x)\right) + \epsilon}{\left(\sum_x \delta_{il}(x) + p_i(x)\right) + \epsilon} \qquad w_l = \left(\left(\sum_k f_k\right)/f_l\right)^{0.5} \tag{1}$$

where x is the sample pixel position and i denotes the label. l is the ground-truth label at x. $E[\bullet]$ is the mean value with respect to i and x in L_{Dice} and L_{CE}, respectively. w_{Dice} and w_{CE} are the trade-off parameters. $\delta_{il}(x)$ is the Kronecker delta, which is 1 when $i = l$ and 0 otherwise. $p_i(x)$ is the softmax probability that the x sample is predicted to be i label. The results of the neural network output are normalized by the softmax function and output the classification probability between 0 and 1. ε is a smoothing parameter. f_k is the frequency of label k, which can reduce the weight of categories with higher frequency. γ_{Dice} and γ_{CE} are the nonlinearities parameter. Compared with medical image segmentation, the distinction of seismic waveforms are more complicated. In order to avoid excessive parameter optimization and greatly increase the difficulty of training, we set the nonlinear parameter γ_{Dice} and γ_{CE} to 0.3 in this paper. In this way, only the weight coefficients in the loss function need to be determined experimentally. We test the weight coefficient of the loss function on the DCSPP network structure, and finally determine $w_{Dice} = 0.6$ and $w_{CE} = 0.4$.

To demonstrate the effectiveness of the proposed DCSPP network architecture, we set up a regular UNet and a UNet model with SPPM for comparison. The other structural parameters are consistent with the UNet model. Prepared training samples and

Table 1
The structure composition, operating parameters and feature map size of the proposed DCSPP model.

Modules	Operations	Parameters	Sizes
Input			256 × 256 × 1
Conv block1	Conv + BatchNorm and Relu	Kernel size = 3	256 × 256 × 64
	Conv + BatchNorm and Relu	Filters = 64	
Conv block2	MaxPooling	Pool size = 2	128 × 128 × 128
	Conv + BatchNorm and Relu	Kernel size = 3	
	Conv + BatchNorm and Relu	Filters = 128	
Conv block3	MaxPooling	Pool size = 2	64 × 64 × 256
	Conv + BatchNorm and Relu	Kernel size = 3	
	Conv + BatchNorm and Relu	Filters = 256	
Conv block4	MaxPooling	Pool size = 2	32 × 32 × 512
	Dilated Conv + BatchNorm and Relu	Dilated rate = 1	
	Dilated Conv + BatchNorm and Relu	Dilated rate = 2	
	Dilated Conv + BatchNorm and Relu	Dilated rate = 5	
		Filters = 512	
Conv block5	MaxPooling	Pool size = 2	16 × 16 × 1024
	Dilated Conv + BatchNorm and Relu	Dilated rate = 1	
	Dilated Conv + BatchNorm and Relu	Dilated rate = 2	
	Dilated Conv + BatchNorm and Relu	Dilated rate = 5	
		Filters = 1024	
Transposed Conv block1	Transposed Conv + BatchNorm and Relu	Kernel size = 2	32 × 32 × 512
		Strides = 2	
	Concatenate	Kernel size = 3	
	Conv + BatchNorm and Relu	Filters = 512	
	Conv + BatchNorm and Relu		
Transposed Conv block2	Transposed Conv + BatchNorm and Relu	Kernel size = 2	64 × 64 × 256
		Strides = 2	
	Concatenate	Kernel size = 3	
	Conv + BatchNorm and Relu	Filters = 256	
	Conv + BatchNorm and Relu		
Transposed Conv block3	Transposed Conv + BatchNorm and Relu	Kernel size = 2	128 × 128 × 128
		Strides = 2	
	Concatenate	Kernel size = 3	
	Conv + BatchNorm and Relu	Filters = 128	
	Conv + BatchNorm and Relu		
Transposed Conv block4	Transposed Conv + BatchNorm and Relu	Kernel size = 2	256 × 256 × 64
		Strides = 2	
	Concatenate	Kernel size = 3	
	Conv + BatchNorm and Relu	Filters = 64	
	Conv + BatchNorm and Relu		
SPPM	MaxPooling	Pool size = 1, 2, 3, 6	256 × 256 × 64
	Conv2D	Kernel size = 1	
	UpSampling		
	Concatenate		
	Conv2D		
Output	Conv	Filters = 6	256 × 256 × 6
		Kernel size = 1	
		Other parameters:	
		padding = 'same',	
		activation = 'Relu'	
Prediction	Softmax		256 × 256 × 6

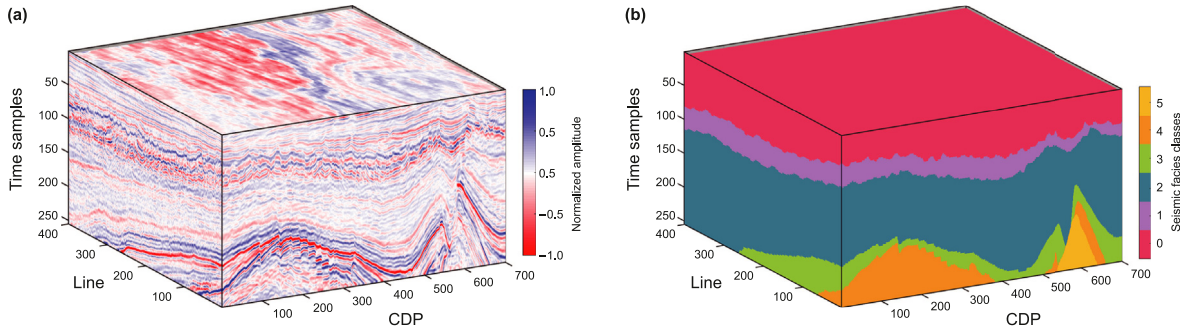

Fig. 6. The F3 public seismic data and corresponding seismic facies label. **(a)** 3D seismic data volume consisting of 601 seismic profiles, 701 CDPs, 255 time sampling points with sampling intervals of 4 ms. **(b)** The seismic facies labels of six categories are interpreted according to the differences in lithology.

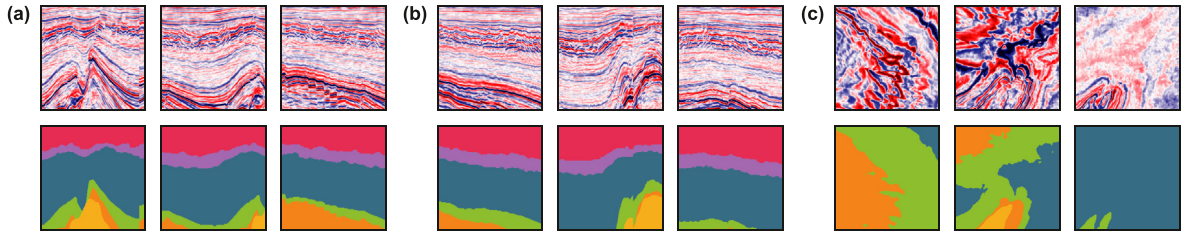

Fig. 7. Training samples of the network, which from the **(a)** inline, **(b)** CDP, **(c)** seismic slices and are intercepted to a size of 256 × 256 to increase the diversity of the samples.

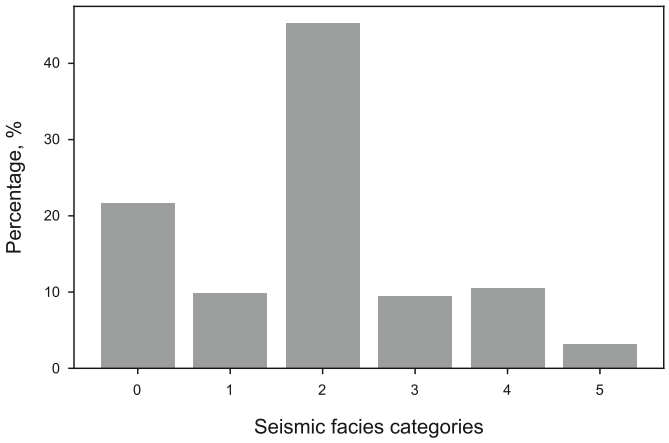

Fig. 8. The percentage of each type of seismic facies in the training samples.

corresponding labels are input into the three different sets of deep convolutional neural networks, and then we adjust and optimize the model's hyperparameters. Through iterative training, the network's loss value continuously decreases and the prediction accuracy is improved. We chose the Adam optimization algorithm, set the learning rate to 1e−4 and batch size to 1, to evaluate the classification accuracy of these three network structures on seismic sections and seismic slices.

4.3. Numerical results

We measure the performance of the proposed DCSPP network configurations from two perspectives, one is the prediction results with different accuracy, and the other is the change of internal feature map during network training. Fig. 9 displays the change in prediction accuracy for one training seismic data as the training epochs increases, the classification accuracy at different epochs is 41.80%, 78.28%, 90.73%, 94.56%, 96.73%, 98.10%, 98.77%, 98.90%, 99.02%, 99.35%, 99.53% respectively. During the training process, the accuracy increases rapidly and exceeds 90%. Once the training accuracy reaches 98.77%, the network accurately predicts the correct seismic facies categories. After the training accuracy exceeds 96.73%, even a slight increase in accuracy leads to a significant improvement in the prediction results. Fig. 9 dynamically and intuitively illustrates the prediction process of the DCSPP neural network, which efficiently and accurately provides seismic facies classification results. Moreover, we also analyze the changes in the internal feature map during the training process of the proposed network. Fig. 10 shows feature maps corresponding to the four down-sampling layers, one intermediate layer, and four up-sampling convolutional blocks. In the shallow convolutional layers, the network extracts more texture features of the seismic facies. Meanwhile, the high-frequency information contained in the seismic data is perceived. As the receptive field continues to increase, the network receives a larger range of information and more accurately perceives the boundary and contour of seismic facies. In the subsequent up-sampling process, these large-scale and detailed information are integrated and the network provides correct seismic facies predictions.

To illustrate the effectiveness of the proposed network structure, we test and compare it with the UNet model and the UNet model with the SPPM. Fig. 11 shows the curves of error and accuracy for different network structures under the same training epochs and training samples. The comparison results show that the DCSPP network has faster convergence speed, lower error and higher accuracy. As the training epochs increase, the difference in error and accuracy between the different network structures gradually decrease. This is because the advantages of small-sample training of the UNet structure. Furthermore, we compare the classification effect of different network architectures on seismic

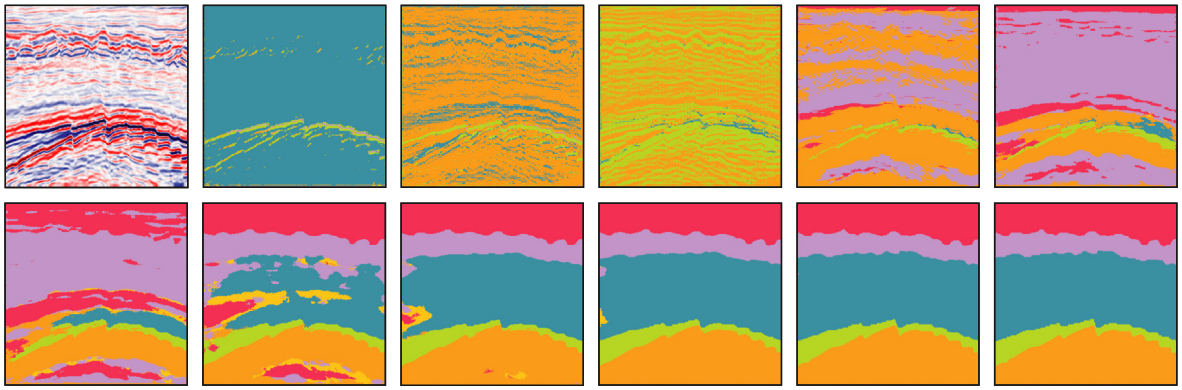

Fig. 9. The seismic facies classification results corresponding to different accuracy during the training process for the proposed DCSPP network model are 41.80%, 78.28%, 90.73%, 94.56%, 96.73%, 98.10%, 98.77%, 98.90%, 99.02%, 99.35%, 99.53%, respectively.

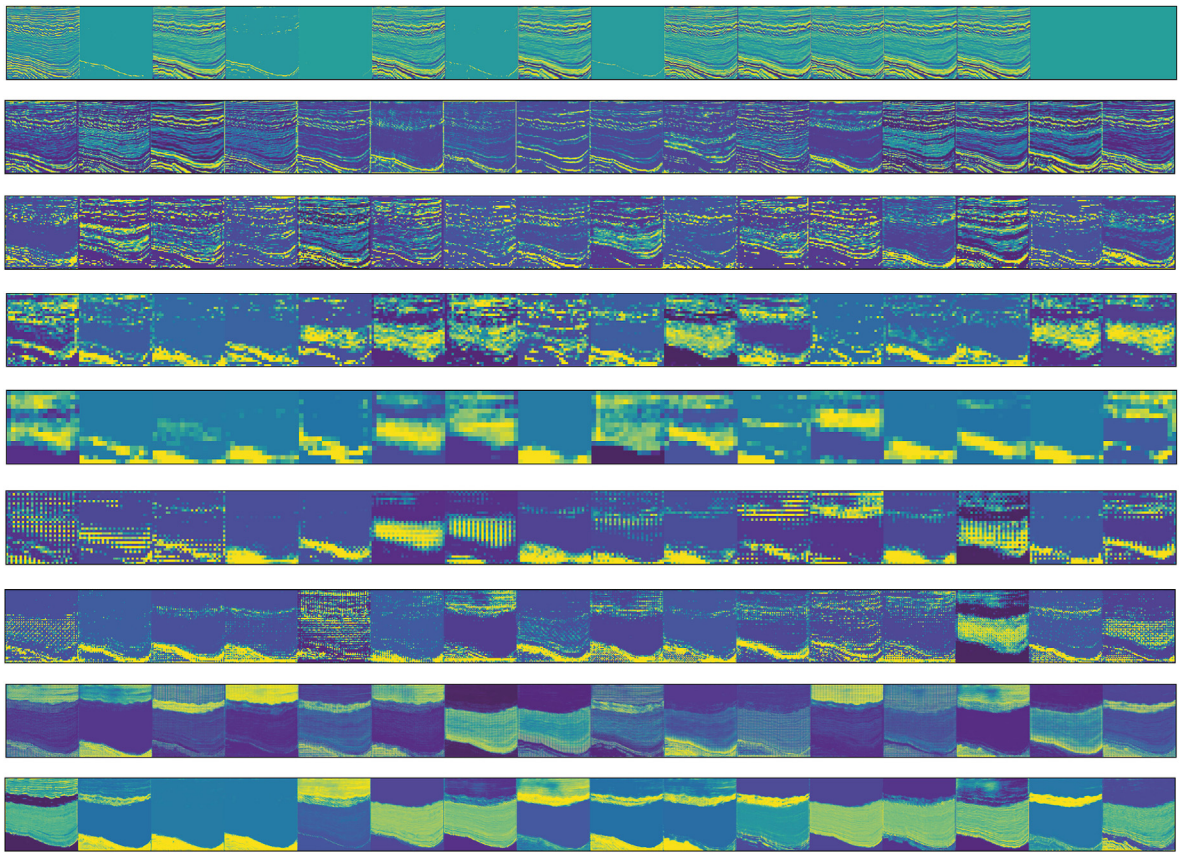

Fig. 10. The internal feature maps during the DCSPP model training, which consist of four down-sampling, one compressed representation, and four up-sampling feature maps.

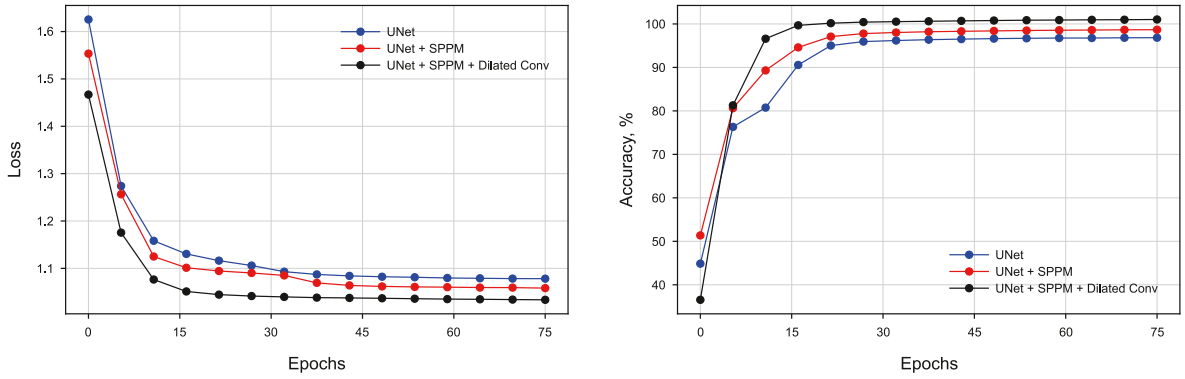

Fig. 11. Comparison of **(a)** training errors and **(b)** prediction accuracy of three different network structures.

profiles and seismic slices. The comparison results are shown in Figs. 12–14.

The UNet model struggles with accurately classifying all the seismic facies as shown in Fig. 12(c), and the boundaries of the seismic facies are blurred, resulting in some incorrect classification results. On the one hand, this may be due to the limited number of training samples. On the other hand, the UNet model is at a disadvantage in obtaining the global perception. Some seismic facies with similar weak amplitude characteristics are misclassified. However, the UNet basically provides correct classification of seismic facies with an accuracy of 91.17%. After introduction the spatial pyramid pooling into the UNet model, the classification accuracy is improved to 93.98% (Fig. 12(d)). The ability to classify similar weak amplitudes is significantly enhanced. Since the spatial pyramid pooling helps the model to obtain the global and spatial information, there are basically no misclassified seismic facies. However, limited by the size of the convolutional receptive field, the prediction results still lack accurate seismic facies boundary characterization. The introduction of the dilated convolution effectively deals with this issue. Fig. 12(e) shows the classification results of the improved DCSPP network model. It not only provides correct seismic facies categories, but also accurately depicts the boundaries of each seismic facies. This indicates that the improved neural network model is more suitable for delineation and analysis of seismic facies under small training samples.

In addition to verifying the effectiveness of the improved network architectures on other sections, we also test and compare the different network structures on CDP profile and seismic slice.

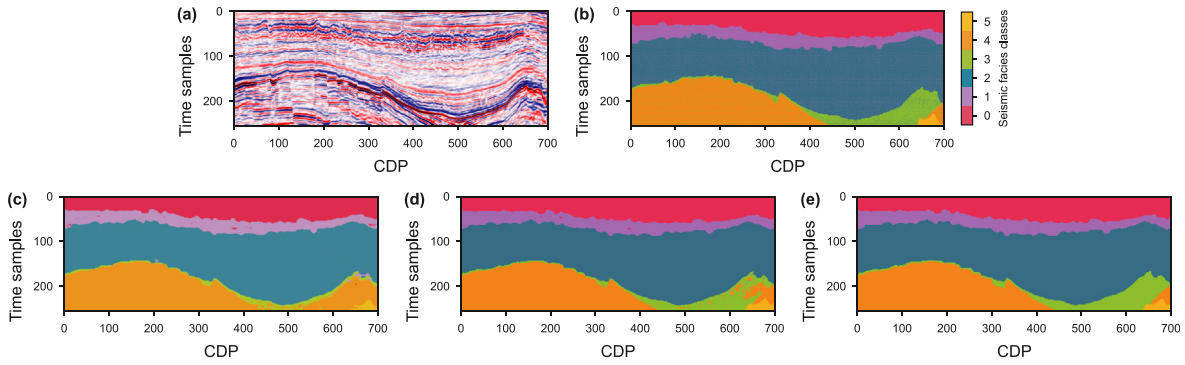

Fig. 12. Results of the seismic facies interpretation of a inline seismic section for three different neural network model: **(a)** seismic profile, **(b)** ground-truth seismic facies, **(c)** classification result from the UNet model, **(d)** interpretation result from the UNet model with the SPPM and **(e)** the classification result of the proposed DCSPP model.

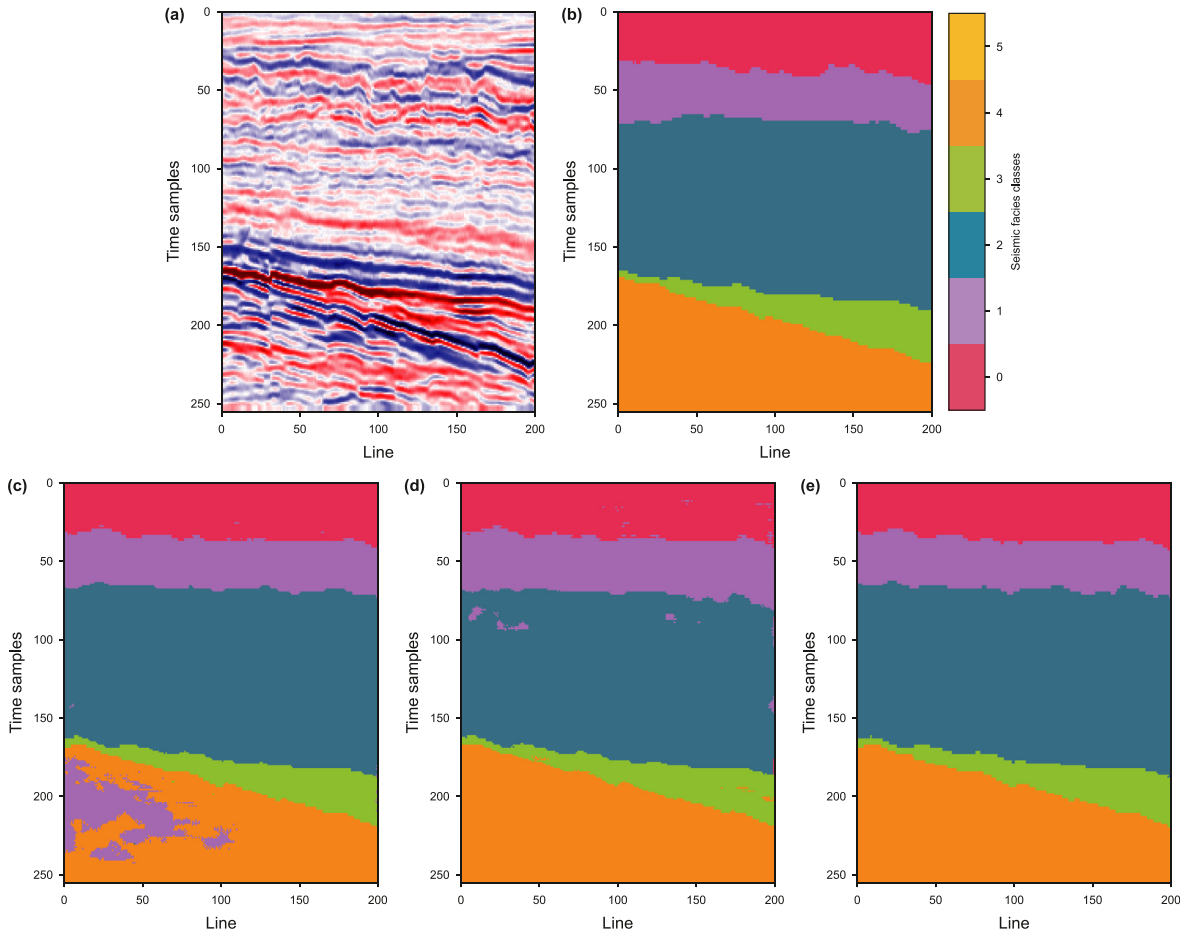

Fig. 13. Results of the seismic facies interpretation of a CDP seismic section for three different neural network model: **(a)** seismic profile, **(b)** ground-truth seismic facies, **(c)** classification result from the UNet model, **(d)** interpretation result from the UNet model with the SPPM and **(e)** the classification result of the proposed DCSPP model.

Figs. 13 and 14 show the seismic facies classification results on CDP profile and seismic slice, respectively. Fig. 13(d) fully illustrates the significance of the introduction of spatial pyramid pooling to improve the accuracy of seismic facies classification. Meanwhile, Fig. 14(e) also adequately demonstrates that the incorporation of dilated convolution greatly enhances the model's ability to delineate seismic facies boundaries. Compared with the UNet model and the UNet model with SPPM, the improved DCSPP model fully integrates the seismic amplitude information of different scales regions, which provides accurate and efficient seismic facies classification results. The results of the F3 model test fully demonstrate and illustrate the effectiveness and reliability of the proposed DCSPP network architecture. The proposed method has certain application prospects and potentials for seismic facies interpretation.

4.4. Evaluation metrics

In order to evaluate the performance of different network structures in seismic facies classification more objectively and

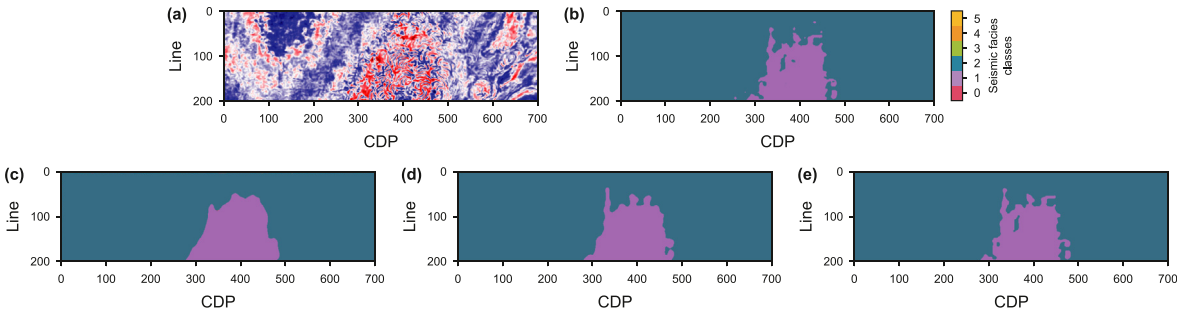

Fig. 14. Results of the seismic facies interpretation of a seismic slice for three different neural network model: **(a)** seismic slice, **(b)** ground-truth seismic facies, **(c)** classification result from the UNet model, **(d)** interpretation result from the UNet model with the SPPM and **(e)** the classification result of the proposed DCSPP model.

Table 2
Evaluation statistics of seismic facies classification results of different network configurations.

Evaluation, %	UNet		UNet + SPPM		UNet + SPPM + Dilated Conv	
	Training data set	Validation data set	Training data set	Validation data set	Training data set	Validation data set
Accuracy	96.80	91.17	97.83	93.98	99.95	97.34
Precision	96.18	72.36	97.32	82.25	99.83	90.03
Recall	96.45	72.33	97.50	83.54	99.77	90.69
F1-score	96.12	79.87	98.30	85.46	99.48	88.27
AUC	95.44	89.50	97.48	93.36	99.87	96.53
mIOU	60.73	60.15	66.30	64.38	75.35	72.41

comprehensively, we use the image segmentation evaluation indicators such as accuracy, precision, and recall to measure the advantages and disadvantages of the network model. Table 2 presents the prediction performance of different network configurations on the training and validation datasets. The six evaluation indicators listed in the table are all derived from the parameters involved in the confusion matrix. The confusion matrix, also known as the error matrix, is a standard format for accuracy evaluation. It consists of four parameters: true positive (*TP*), false positive (*FP*), true negative (*TN*), and false negative (*FN*). In a binary classification problem, *TP* refers to correctly identifying the input image's labeled information as the corresponding label, while *FP* indicates incorrectly identifying the input image's background information as label. *TN* represents correctly identifying the input image's background information as background, while *FN* refers to incorrectly identifying the input image's labeled information as background. In the confusion matrix, *TP* and *TN* correspond to the part where the network predicts correctly, while *FP* and *FN* correspond to the part where the network predicts incorrectly. Therefore, the larger the *TP* and *TN* values, the better model's performance. Conversely, larger *FP* and *FN* values indicate worse performance. For the multi-classification tasks, the confusion matrix becomes more complex, but its fundamental principle remains the same.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

Accuracy is the simplest indicator of image segmentation, which indicates the proportion of correctly classified pixels. In general, the higher the accuracy, the better classification model.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

Precision refers to the proportion of correctly predicted part in the whole prediction part.

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

Recall is a measure of coverage, which measures the proportion of correctly predicted parts in the ground-truth.

$$F1 - score = 2*\frac{Precision*Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN} \tag{5}$$

In image segmentation, *Precision* and *Recall* indicators sometimes have contradictory situations, so they need to be considered together. The most common method is the F-Measure, which also known as the F-Score. The F-Measure is a weighted harmonic average of *Precision* and *Recall*. The most commonly used is the *F1-score*, which combines the results of *Precision* and *Recall*. When the *F1-score* is high, it indicates that the classification model has good validity. Area under the curve (AUC) refers to the area below the receiver operator characteristic (ROC) curve. The ROC curve is a crucial and common statistical analysis method for evaluating the generalization and performance of neural network models. Since the ROC curves of different models may coincide, AUC is introduced to evaluate the generalization ability of the model. In general, the value of AUC is between 0.5 and 1.0, and the larger it is, the better prediction performance and generalization ability of the model has.

$$mIOU = \frac{TP}{TP + FP + FN} \tag{6}$$

The mean intersection over union ratio (*mIOU*) refers to the average intersection over union (IOU) of the real images to the predicted images for each category. IOU indicates the ratio of the intersection of the true value of a pixel to the predicted value and the union of the true value and the predicted value. *mIOU* refers to the mean of IOU for all categories. It is the fundamental indicator for evaluating the effect and performance of image segmentation. Therefore, in addition to the overall *mIOU* index, we also evaluate the *mIOU* of each seismic facies in the training and validation datasets for three different network architectures, as shown in Table 3. The *mIOU*

Table 3
The *mIOU* evaluation indices of each seismic facies of the training and validation data sets with different network architectures.

Seismic facies categories	UNet		UNet + SPPM		UNet + SPPM + Dilated Conv	
	Training data set	Validation data set	Training data set	Validation data set	Training data set	Validation data set
	60.73	60.15	66.30	64.38	75.35	72.41
0	70.53	67.30	78.20	70.35	85.86	82.45
1	58.42	58.46	60.67	54.64	69.25	65.17
2	84.06	82.23	87.64	82.23	94.55	90.12
3	57.48	55.81	62.19	58.60	70.77	68.27
4	60.39	57.38	64.52	59.73	72.90	67.15
5	49.33	48.85	52.92	48.97	65.42	60.31

index is much lower than other evaluation parameters, which is because seismic facies classification results are affected by the uneven data distribution. For category 0 and category 2 seismic facies, the *mIOU* is relatively high. However, for category 5 seismic facies, which has a smaller percentage of training samples, it has the lowest *mIOU* index. The unbalanced distribution of the training samples leads to low values of the *mIOU* indicator.

We use the above six indicators to evaluate the performance of these three different network architectures. In summary, the improved DCSPP neural network model performs the best. After the introduction of SPPM, the classification accuracy and precision of the model have been greatly improved. Moreover, the comparison of the *mIOU* index indicates that the network has good prediction effect and generalization ability. This means that the model can better adapt to changes within the data and has good adaptability in migration applications.

5. Field data example

To further illustrate the practicability and applicability of our proposed method, we conduct a comprehensive test and demonstration of different methods based on the field data and corresponding seismic facies interpretation results. The study area is located to the south of the A Sag and to the north of the B Uplift. The strata develop from the bottom to the top, including the Permian, Triassic, Jurassic and Three series. Among them, the Jurassic, Lower Tertiary X Formation, and Upper Tertiary Y Formation are the main oil-bearing strata in this area. The target layer for this study is the Lower Tertiary Y Formation. The target layer develops light gray, thick layers of gravel-bearing and unequal-grained sandstone, sandy conglomerate, fine sandstone intercalated with thin layers of mudstone, and a set of flint conglomerate layers at the bottom are in angular unconformity contact with the underlying strata. The Tertiary System in this area develops multi-stage and multi-type reservoirs. Among them, the X Formation mainly develop braided river delta and carbonate platform sedimentary systems based on the sources of the B Uplift in the south. During the deposition period of the Y Formation, braided river and meandering river sand bodies are widely developed. These sand bodies that overlap each other vertically and horizontally have good physical properties and create good storage space for oil and gas accumulation in the study area. The Y Formation in the study area is a braided river deposit controlled by the source from the B Uplift. The river flows through the study area from south-east to north-west. During this period, four braided main river channels developed. The river channels sway frequently and cut each other. The rivers are large in size and develop braided bars. The braided bars are mainly developed in the C2, C3 and B4 well area. Fig. 15 displays a seismic amplitude slice and the interpretation result of sedimentary facies provided by geologists from the study area. The original data is digitized and grid-uniformed, respectively, which used as the basic data to facilitate subsequent training.

In order to verify the effectiveness of our proposed method, we supplement the lithofacies classification test experiment with field seismic data. Due to the limited results of seismic facies interpretation of field seismic data, in order to effectively test the different network configurations, we first carry out the basic work of data enhancement on the collected data, so as to make up for the impact of limited labels on the application effect of the method as much as possible. We use data augmentation techniques to increase the diversity and coverage of training samples. Data augmentation includes operations such as rotation, flipping, translation, and scaling to simulate seismic data at different angles and locations. In this way, we can increase the diversity of training samples, enabling the model to generalize better to the task of seismic surface classification over small patches. Fig. 16 shows the raw data, seismic slice and corresponding sedimentary facies after data augmentation, respectively. In addition, we also fine-tune the network structure trained on F3 to adapt to the characteristics of field seismic data through transfer learning technology, so as to improve and enhance the learning performance. Transfer learning can be implemented in many ways, one of the common methods is to use pre-trained models. A pre-trained model is a model trained on a large-scale dataset, usually used to solve a related task. By pre-training on the source domain, the model can learn general features or knowledge. These learned features or knowledge can then be transferred or fine-tuned to tasks on the target domain to improve learning performance. Fig. 17 displays the facies classification result of field seismic slice for three different neural network model. Consistent with the conclusions obtained from the model experiments, the DCSPP network architecture we proposed exhibits the best classification effect. It distinctly delineates the boundaries of different seismic facies and accurately provides the spatial distribution range of each seismic facies. From the comparison of experimental results, it is found that after the introduction of the SPPM, the network can extract and obtain large-scale global information more reasonably and effectively compared to the UNet structure. Furthermore, the dilated convolution and skip connection promotes the DCSPP network structure to better handle the spatial information at different scales in down-sampling and up-sampling, resulting in the optimal classification performance. Seismic facies interpretation of field seismic data based on deep neural networks is a very meaningful work. Here, the preliminary tests have been processed based on technologies such as image augmentation and transfer learning. The current work is still in the experimental stage and there are still several shortcomings, there are many areas that need to be supplemented and improved. We will continue to explore in the future.

6. Discussion

The wide application of deep learning technology in geophysics

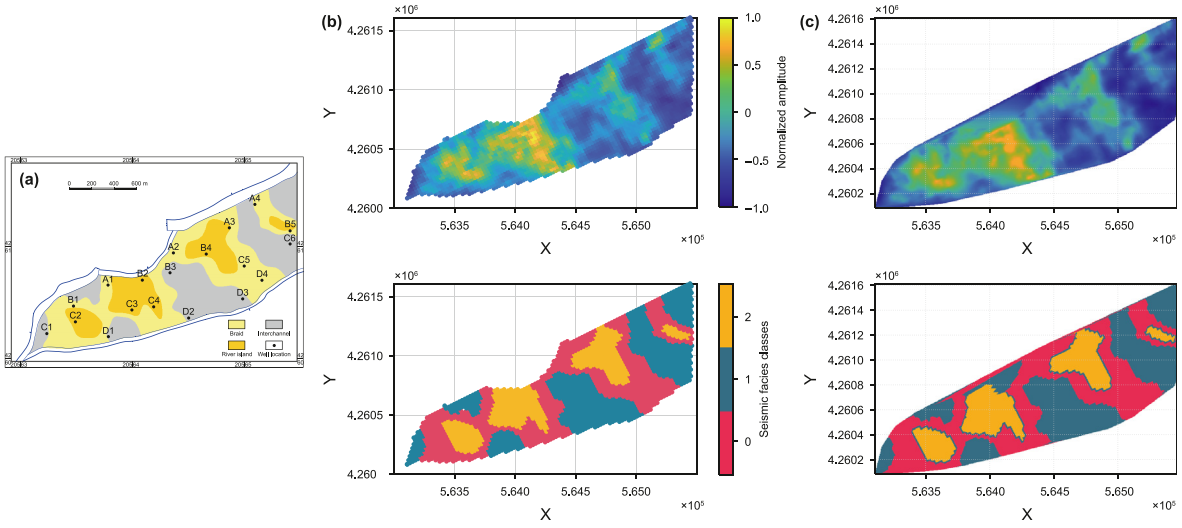

Fig. 15. Field seismic data and corresponding seismic facies interpretation results: **(a)** is the interpretation results of sedimentary facies given by geologists based on the characteristics of well logging curves of well points, the plane distribution of sand bodies and seismic data; **(b)** is the original seismic amplitude slice and the facies distribution map after the digitization of sedimentary facies interpretation results; **(c)** is the basic network training data after interpolation and homogenization processing on the basis of (b).

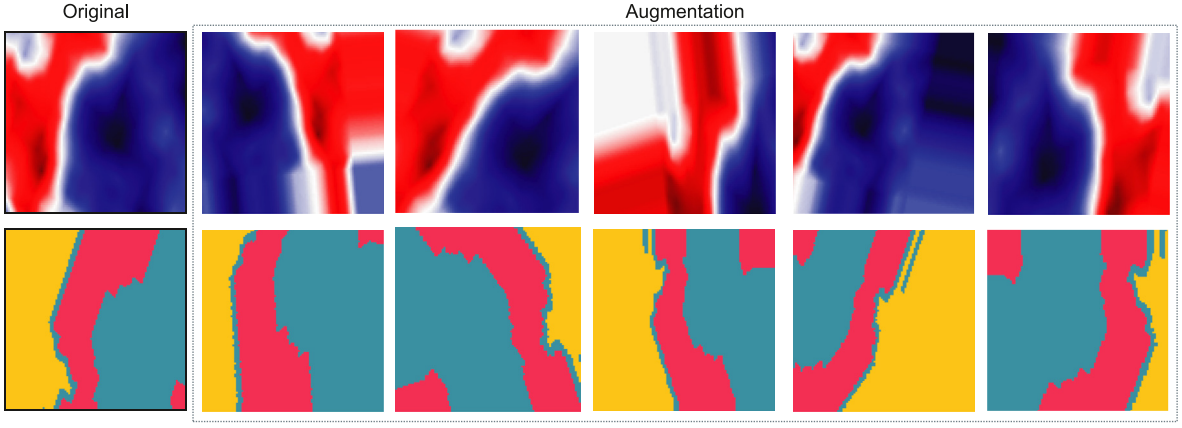

Fig. 16. Seismic slice and sedimentary facies after data augmentation.

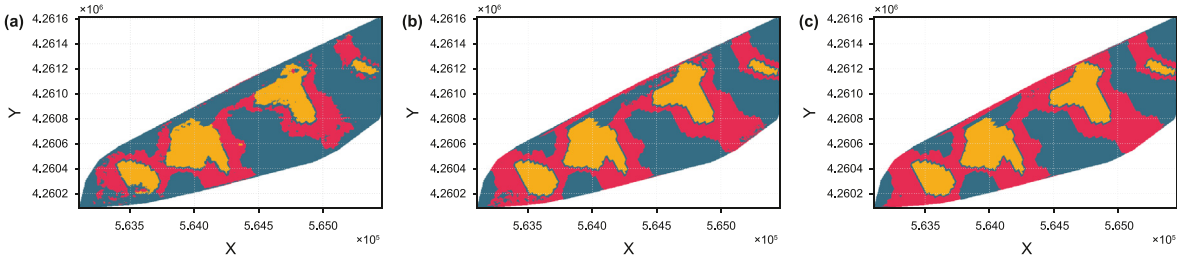

Fig. 17. Facies interpretation results of a field seismic slice for three different neural network model: **(a)** classification result from the UNet model, **(b)** interpretation result from the UNet model with the SPPM and **(c)** the classification result of the proposed DCSPP model.

owes much to its powerful nonlinear mapping relationship. Many geophysical problems are challenging to describe with explicit mathematical expressions, given their highly complex relationships between variables. The nonlinear transformation and high-dimensional mapping of the network have proven to be a valuable solution to these challenges. Moreover, the efficiency and immediacy of deep-learning techniques save time and alleviate manual labor. Therefore, the seismic data processing and interpretation work based on the deep learning technology have been increasing and developing. Supervised learning as the most common and widely used deep learning approach, which is particularly popular nowadays. Although the proposed DCSPP neural network model achieves automatic seismic facies interpretation, it essentially belongs to the supervised learning mechanism. Even though the small-sample training advantage of the UNet model is used to alleviate the dependence of extensive training samples, it still requires labeled data. The diversity of samples and the imbalance in categories moderately affect the classification results. In order to

solve the problem of limited training samples, image augmentation and other technical approaches have been applied to seismic data cutting and division. Image augmentation refers to expanding the datasets through various data enhancement transformations, such as clipping, rotation, flip, scale, translation, contrast, noise disturbance and color transformation. By applying random changes to the training samples, similar but different training samples are generated and the number of training samples is expanded. Furthermore, random changes to the training samples reduce the dependence of the network model on specific attributes, thereby improving the generalization ability of the model. However, whether this application will destroy the original morphology and spatial distribution of seismic data and adversely affect the seismic facies classification remains to be considered.

Another bothering problem in the application of deep learning in geophysics is its transfer and generalization ability. How to successfully and effectively apply the trained model to field seismic data is another crucial aspect to evaluate and investigate the applicability of the proposed method. However, the field seismic data often have very few labeled samples available for network training. Therefore, model pre-training is usually used to test and optimize the network configurations on the model data, and then fine-tune the network structures and neuron weights to adapt to the field seismic data through transfer learning technology. The application of the proposed DCSPP network model to field seismic data is still undergoing testing and trials. Currently, we only use seismic data as input to the network, which makes classifications and predictions by distinguishing differences in seismic waveforms. Whether the single seismic waveform contains sufficient information to distinguish different seismic facies. In fact, many of the early intelligent seismic facies classification work was based on various seismic attributes. Seismic facies classification work combining seismic attributes and seismic waveforms can be considered subsequently. This requires the optimization and analysis of different seismic attributes, so as to make predictions of seismic facies based on the necessary and non-redundant input information.

7. Conclusions

We propose an improved deep dilated convolutional neural network DCSPP to classify different seismic facies based on seismic amplitude differences. The model architecture takes the encoder and decoder structures of UNet as the basic configurations, with the incorporation of the SPPM and the dilated convolution layers. The model combines the advantages of small-sample training of the UNet model, the spatial information acquisition and integration of SPPM, and the strengths of the dilated convolution to expand the receptive field and perceive more spatial information. Besides that, to address the issue of sample category imbalance under small training samples, we introduce a composite loss function to enhance the classification performance. The test results of F3 public data show that the proposed model successfully distinguishes different seismic waveform features and achieves effective classification of seismic facies. Compared to the UNet model, the introduction of SPPM significantly improves the classification accuracy. The model integrates the seismic characteristics of different regions and scales to obtain the global information. However, its depiction of seismic facies boundaries is still not ideal. The introduction of dilated convolution layers allows the model to capture an exponentially increasing receptive field during down-sampling. More low-frequency and contour information is perceived to distinguish the boundaries of different seismic facies. The test results of both the seismic profiles and the seismic slices demonstrate and illustrate the effectiveness and reliability of the proposed network architectures. Moreover, we visualize the changes of the internal feature maps during the network training process and more intuitively understand the process of information extraction, integration and prediction of the model. Among them, the proposed DCSPP network model exhibits the best performance and prediction potential. The application of the method to field seismic data further confirms its effectiveness and reliability. As an intelligent, real-time approach for seismic facies analysis, which greatly saves time and labor.

CRediT authorship contribution statement

Na-Xia Yang: Data curation, Methodology, Writing – original draft. **Guo-Fa Li:** Project administration, Resources, Supervision. **Ting-Hui Li:** Investigation, Validation. **Dong-Feng Zhao:** Investigation, Writing – review & editing. **Wei-Wei Gu:** Conceptualization, Formal analysis.

Declaration of competing interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled "An improved deep dilated convolutional neural network for seismic facies interpretation".

Acknowledgement

This research was funded by the Fundamental Research Project of CNPC Geophysical Key Lab (2022DQ0604-4) and the Strategic Cooperation Technology Projects of China National Petroleum Corporation and China University of Petroleum-Beijing (ZLZX 2020-03).

References

Alaudah, Y., Michalowicz, P., Alfarraj, M., AlRegib, G., 2019. A machine-learning benchmark for facies classification. Interpretation 7, SE175–SE187. https://doi.org/10.1190/int-2018-0249.1.

Alfarhan, M., Deriche, M., Maalej, A., 2022. Robust concurrent detection of salt domes and faults in seismic surveys using an improved UNet architecture. IEEE Access 10, 39424–39435. https://doi.org/10.1109/ACCESS.2020.3043973.

Amin, A., Deriche, M., 2016. Salt-Dome detection using a codebook-based learning model. Geosci. Rem. Sens. Lett. IEEE 13, 1636–1640. https://doi.org/10.1016/j.petrol.2022.110911.

Babikir, I., Elsaadany, M., Sajid, M., Laudon, C., 2022. Evaluation of principal component analysis for reducing seismic attributes dimensions: implication for supervised seismic facies classification of a fluvial reservoir from the Malay Basin, offshore Malaysia. J. Petrol. Sci. Eng. 217, 110911. https://doi.org/10.1016/j.petrol.2022.110911.

Birnie, C., Hansteen, F., 2022. Bidirectional recurrent neural network for seismic event detection. Geophysics 87, KS97–KS111. https://doi.org/10.1190/GEO2020-0806.1.

Chai, X.T., Nie, W.H., Lin, K., Tang, G.Y., Yang, T.H., Yu, J.Y., Cao, W.J., 2022. An open-source package for deep-learning-based seismic facies classification: benchmarking experiments on the SEG 2022 open data. IEEE Trans. Geosci. Rem. Sens. 60, 4507719. https://doi.org/10.1109/TGRS.2022.3144666.

Chung, Y., Lu, W.K., Tian, X.Y., 2022. Data cleansing for salt dome dataset with noise robust network on segmentation task. Geosci. Rem. Sens. Lett. IEEE 19, 8027805. https://doi.org/10.1109/LGRS.2022.3170497.

Duan, X.D., Zhang, J., 2020. Multitrace first-break picking using an intergrated seismic and machine learning method. Geophysics 85, WA269–WA277. https://doi.org/10.1190/GEO2019-0422.1.

Duan, Y.T., Zheng, X.D., Hu, L.L., Sun, L.P., 2019. Seismic facies analysis based on deep convolutional embedded clustering. Geophysics 84, IM87–IM97. https://doi.org/10.1190/geo2018-0789.1.

Feng, R.H., Balling, N., Grana, D., Dramsch, J.S., Hansen, T.M., 2021. Bayesian convolutional neural networks for seismic facies classification. IEEE Trans. Geosci. Rem. Sens. 59, 8933–8940. https://doi.org/10.1109/TGRS.2020.3049012.

Han, S., Liu, Y.J., Li, Y.B., Luo, Y., 2022. First arrival traveltime picking through 3-D U-Net. Geosci. Rem. Sens. Lett. IEEE 19, 8016405. https://doi.org/10.1109/

LGRS.2021.3096572.
Hu, G., Hu, Z.W., Liu, J.P., Cheng, F., Peng, D.C., 2022. Seismic fault interpretation using deep learning-based semantic segmentation method. Geosci. Rem. Sens. Lett. IEEE 19, 7500905. https://doi.org/10.1109/LGRS.2020.3041301.
Liu, B., Yu, A.Z., Yu, X.C., Wang, R.R., Gao, K.L., Guo, W.Y., 2021a. Deep multiview learning for hyperspectral image classification. IEEE Trans. Geosci. Rem. Sens. 59, 7758–7772. https://doi.org/10.1109/TGRS.2020.3034133.
Liu, J.L., Dai, X.F., Gan, L.D., Liu, L., Lu, W.K., 2018. Supervised seismic facies analysis based on image segmentation. Geophysics 83, O25–O30. https://doi.org/10.1190/GEO2015-0539.1.
Liu, M.L., Jervis, M., Li, W.C., Nivlet, P., 2020a. Seismic facies classification using supervised convolutional neural networks and semisupervised generative adversarial networks. Geophysics 85, O47–O58. https://doi.org/10.1190/GEO2019-0627.1.
Liu, N.H., He, T., Tian, Y.J., Wu, B.Y., Gao, J.H., Xu, Z.B., 2020b. Common-azimuth seismic data fault analysis using residual UNet. Interpretation 8, SM25–SM37. https://doi.org/10.1190/int-2019-0173.1.
Liu, Z.G., Cao, J.X., Chen, S.N., Lu, Y.J., Tan, F., 2021b. Visualization analysis of seismic facies based on deep embedded SOM. Geosci. Rem. Sens. Lett. IEEE 18, 1491–1495. https://doi.org/10.1109/LGRS.2020.3003585.
Liu, Z.G., Cao, J.X., Lu, Y.J., Chen, S.N., Liu, J.L., 2019. A seismic facies classification method based on the convolutional neural network and the probabilistic framework for seismic attributes and spatial classification. Interpretation 7, SE225–SE236. https://doi.org/10.1190/INT-2018-0238.1.
Mirzakhanian, M., Hashemi, H., 2022. Semisupervised fuzzy clustering for facies analysis using extended elastic impedance seismic attributes. Geophysics 87, N75CN84. https://doi.org/10.1190/GEO2021-0330.1.
Nasim, M.Q., Maiti, T., Srivastava, A., Singh, T., Mei, J., 2022. Seismic facies analysis: a deep domain adaptation approach. IEEE Trans. Geosci. Rem. Sens. 60, 4508116. https://doi.org/10.1109/TGRS.2022.3151883.
Noh, H., Hong, S., Han, B., 2015. Learning deconvolutional network for semantic segmentation: 2015. IEEE International Conference on Computer Vision (ICCV) 1520–1528. https://doi.org/10.1109/ICCV.2015.178.
Puzyrev, V., Elders, C., 2021. Unsupervised seismic facies classification using deep convolutional autoencoder. Geophysics 87, IM125–IM132. https://doi.org/10.1190/GEO2021-0016.1.
Qi, J., Zhang, B., Lyu, B., Marfurt, K., 2020. Seismic attributes selection for machine-learning-based facies analysis. Geophysics 85, O17–O35. https://doi.org/10.1190/GEO2019-0223.1.
Qian, F., Yin, M., Liu, X.Y., Wang, Y.J., Lu, C., Hu, G.M., 2018. Unsupervised seismic facies analysis via deep convolutional autoencoders. Geophysics 83, A39–A43. https://doi.org/10.1190/GEO2017-0524.1.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation. 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 9351. SPRINGER, pp. 234–241. Proceedings Paper. https://doi.org/10.1007/978-3-319-24574-4_28.
Saxena, A., Tomar, S.S., Jain, G., Gupta, R., 2021. Deep learning based diagnosis of diseases using image classification. 11th International Conference on Cloud Computing, Data Science and Engineering (Confluence). IEEE, pp. 399–404. Proceedings Paper. https://doi.org/10.1109/Confluence51648.2021.9377154.
Shelhamer, E., Long, J., Darrell, T., 2017. Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39, 640–651. https://doi.org/10.1109/TPAMI.2016.2572683.
Shi, Y.Z., Wu, X.M., Fomel, S., 2022. Waveform embedding: automatic horizon picking with unsupervised deep learning. Geophysics 85, WA67–WA76. https://doi.org/10.1190/GEO2019-0438.1.
Song, C.Y., Li, L., Li, L.X., Li, K.H., 2021. Robust K-means algorithm with weighted window for seismic facies analysis. Journal of Geophysics and Enigineering 18, 618–626. https://doi.org/10.1093/jge/gxab039.
Tolstaya, E., Egorov, A., 2022. Deep learning for automated seismic facies classification. Interpretation 10, SC31–SC40. https://doi.org/10.1190/INT-2021-0140.1.
Wang, F., Wu, X.M., Wang, H.Z., 2022. Seismic horizon identification using semsupervised learning with virtual adversarial training. IEEE Trans. Geosci. Rem. Sens. 60, 4508611. https://doi.org/10.1109/TGRS.2022.3154439.
Wang, P.Q., Chen, P.F., Yuan, Y., Liu, D., Huang, Z.H., Hou, X.D., Cottrell, G., 2018. Understanding convolution for semantic segmentation. 18th IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, pp. 1451–1460. Proceedings Paper. https://doi.org/10.1109/WACV.2018.00163.
Wang, W.P., Wang, Z.R., Zhou, Z.F., Deng, H.X., Zhao, W.L., Wang, C.Y., Guo, Y.Z., 2021. Anomaly detection of industrial control systems based on transfer learning. Tsinghua Sci. Technol. 26, 821–832. https://doi.org/10.26599/TST.2020.9010041.
Wong, K.C.L., Moradi, M., Tang, H., Syeda-Mahmood, T., 2018. 3D segmentation with exponential logarithmic loss for highly unbalanced object sizes. Medical Image Computing and Computer Assisted Intervention MICCAI 612–619. https://doi.org/10.1007/978-3-030-00931-1_70.
Wu, X.M., Liang, L.M., Shi, Y.Z., Fomel, S., 2019. FaultSeg3D: using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. Geophysics 84, IM35–IM45. https://doi.org/10.1190/GEO2018-0646.1.
Xie, J.Y., Girshick, R., Farhadi, A., 2016. In: Unsupervised deep embedding for clustering analysis. JMLR 33rd International Conference on Machine Learning, 48. Proceedings Paper. https://doi.org/10.48550/arXiv.1511.06335.
Yan, X.Y., Gu, H.M., Luo, H.M., Yan, Y.P., 2020. Intelligent seismic facies classification based on an improved deep learning method. Oil Geophys. Prospect. 55, 1169–1177. https://doi.org/10.13810/j.cnki.issn.1000-7210.2020.06.001.
Yin, P.S., Yuan, R., Cheng, Y.M., Wu, Q.Y., 2020. Deep guidance netwoek for biomedical image segmentation. IEEE Acess 8, 116106–116116. https://doi.org/10.1109/ACCESS.2020.3002835.
Yuan, S.Y., Liu, J.W., Wang, S.X., Wang, T.Y., Shi, P.D., 2018. Seismic waveform classification and first-break picking using convolution neural networks. Geosci. Rem. Sens. Lett. IEEE 15, 272–276. https://doi.org/10.1109/LGRS.2017.2785834.
Zhang, H.R., Chen, T.S., Liu, Y., Zhang, Y.X., Liu, J., 2021. Automatic seismic facies interpretation using supervised deep learning. Geophysics 86, IM15–IM33. https://doi.org/10.1190/GEO2019-0425.1.
Zhang, Y.X., Liu, Y., Zhang, H.R., Xue, H., 2020. Seismic facies analysis based on deep learning. Geosci. Rem. Sens. Lett. IEEE 17, 1119–1123. https://doi.org/10.1109/LGRS.2019.2941166.
Zhao, T., Jayaram, V., Roy, A., Marfurt, K.J., 2015. A comparision of classification techniques for seismic facies recognition. Interpretation 3, SAE29–SAE58. https://doi.org/10.1190/INT-2015-0044.1.