

# Accelerating Seismic Dip Estimation With Deep Learning

Xiaokai Wang<sup>1</sup>, Member, IEEE, Dawei Liu<sup>1</sup>, and Wenchao Chen<sup>1</sup>

**Abstract**—The seismic volumetric dip is a crucial seismic geometric attribute, which can provide useful information for assisting subsequent processing and interpretation. Waveform similarity scanning-based dip estimation (WSSB) delivers reliable dip estimation but encounters problems of expensive computation. To improve computing efficiency, we use multitask deep learning to simultaneously estimate the inline dip and crossline dip directly from a 3-D field seismic dataset. Our method considers dip estimation as a regression problem and trains a multilayer convolutional neural network with dual-channel output. It aims to output continuous values of seismic apparent dip from two directions simultaneously. To train the network, we propose an effective and efficient workflow to create a training sample dataset, which consists of field seismic cubes and the corresponding dip labels estimated by WSSB. After training, the network automatically learns how to extract rich and proper features that are important for dip estimation. By sliding the extraction window within the full 3-D seismic data, the network can output many overlapping dip cubes that are stacked to get two complete 3-D volumes of seismic dip. The final results of dip estimation by our method are similar to those by WSSB. We further demonstrate the accuracy of our approach by comparing the structural curvature. However, the computation time of our method is much less than that of WSSB. The proposed method can accurately estimate seismic volumetric dips with high computational efficiency.

**Index Terms**—Convolutional neural network (CNN), deep learning, dip estimation, seismic attribute, similarity scanning.

## I. INTRODUCTION

SEISMIC attributes play an important role in seismic structural interpretation, stratigraphic analysis, dynamic reservoir detection, reservoir characterization, and modeling. They can be roughly divided into two categories, i.e., geometric attributes and physical attributes [1]. Geometric attributes are general tools in seismic graphic interpretation, such as identification of fault orientation and channels. As one of the representative geometric attributes, seismic volumetric dip can not only be used to identify subtle structures but also identify valuable geological information for seismic processing and

interpretation, such as seismic curvature and structure-oriented filter. Therefore, estimating the seismic volumetric dip is a subject worthy of study.

Many scholars have proposed different methods to estimate seismic volumetric dip. Cross correlation-based methods are the first category. Bahorich and Farmer [2] utilized cross correlation over a fixed length windowed seismic data to calculate seismic volumetric dip and found it can better image the submarine fan. The complex trace analysis-based methods are the second category. Barnes [3] calculated the temporal and spatial partial derivatives of the instantaneous phase using the complex seismic trace analysis theory and then calculated instantaneous dip angle and instantaneous azimuth angle. To reduce the influence of amplitude lateral variation on the estimation results, Barnes [4] applied a smooth weighted average window to smooth the instantaneous phase. Waveform similarity scanning-based dip estimation (WSSB) methods are the third category. Milkereit [5] divided vertical seismic profiles into localized windows and calculated seismic apparent dip in each window using the localized slant stacks. Marfurt *et al.* [6] measured seismic coherency to estimate seismic dips by calculating semblance between the windowed seismic traces along with a series of preset dips and selected the dip with the highest semblance value as the local reflector's dip. Marfurt *et al.* [7] also applied the eigenstructure coherence algorithm to dip scanning. Gradient structure tensor-based methods are the fourth category. After calculating the 3-D seismic signals' gradient within a fixed-size window and constructing a covariance matrix, seismic volumetric dip can be estimated by the dominant eigenvectors of the covariance matrix [8], [9]. Among the above methods, the WSSB dip estimation can result in one reliable dip estimation if the user-defined increment of discrete candidate dip is small enough. However, computation costs increase with a small dip scanning interval. Therefore, we should consider defining a suitable interval of dips and balancing the accuracy and consuming time for the whole seismic survey for the successful WSSB dip estimation.

Deep learning has a powerful ability to extract complex features and has been successfully applied to solve many image processing problems, such as classification [10], target recognition [11], and human action recognition [12]. Recently, deep learning has been introduced to solve problems in seismic processing [13]–[15] and interpretation [16]. The regression model, patch-based classification methods, and the encoder–decoder segmentation model are the three main methods. The regression model, whose output size is the same

Manuscript received October 8, 2020; revised December 31, 2020 and August 14, 2021; accepted October 12, 2021. Date of publication October 15, 2021; date of current version January 10, 2022. This work was supported by the National Natural Science Foundation of China under Grant 41774135 and Grant 41974131. The work of Dawei Liu was supported by the China Scholarship Council. (Corresponding author: Dawei Liu.)

The authors are with the School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: xkwang@xjtu.edu.cn; liudawei2015@stu.xjtu.edu.cn; wenchao@xjtu.edu.cn).

Digital Object Identifier 10.1109/LGRS.2021.3120315

as the input size, is widely used in seismic noise suppression. However, it was rarely used in dip estimation because its output boundary values are ambiguous due to the lack of sufficient input signals. Patch-based methods, which output the predicted central voxel of the input seismic patch, have been applied to salt interpretation [17] and fault detection [18]. However, their results are suboptimal compared to encoder–decoder segmentation models, especially when the output voxel is at the intersection of the two different categories. Besides, the test stage is computationally intensive. Encoder–decoder segmentation models, which have the same high calculation efficiency as the regression model, are famous for fault segmentation. However, they also encounter some drawbacks. As part of the input information is lost due to downsampling, their output resolution will be affected. Besides, they cannot accurately output a broad number of segmentation categories, thus not meeting the needs of high-resolution dip estimation whose classification categories can reach more than 1000.

In this letter, we propose to apply a deep regression model to estimate 3-D seismic dip automatically. The dip estimation ability of the seismic dip estimation convolutional neural network (SDE-CNN) is learned from training samples. To make the network obtain a high-precision dip extraction ability, we use WSSB with a small scanning interval to obtain high-resolution labels. WSSB estimates each voxel of dip labels from seismic data within a fixed window, allowing us to cut the training data into small subvolumes to facilitate network training and reduce GPU memory usage. Besides, to make the voxels at the output boundary have enough input information, we make the network output size smaller than the input size. We also notice that inline dip and crossline dip are coupled with each other during the scanning processing, and they are both estimated from seismic structural features. Therefore, we adopt multitask learning [19] to simultaneously perform dip estimation tasks in two directions. In this way, we can further reduce the consuming time of dip estimation. In Section II, we introduce the model formulation, training datasets, network architecture, and training strategy. In Section III, we use field seismic data to prove the effectiveness and efficiency of our method. Finally, we conclude in Section IV.

## II. METHOD

### A. Model Formulation and Training Datasets

In order to leverage the promising progress in multitask learning, particularly based on the deep neural networks, we propose to combine the inline dip estimation task and the crossline dip estimation task in the unified multitask framework aiming to jointly learn the shared features and task-specific features to reduce the total testing time. The total loss of multitask learning is the sum of the weighted losses of the multiple tasks as follows:

$$\mathcal{L}(\Theta) = \sum_{i=1}^T w_i \mathcal{L}_i(\Theta', \Theta_i) \quad (1)$$

where  $T$  denotes the number of the tasks; here,  $T = 2$ .  $\Theta_i$  is the parameter corresponding to each special task, and  $\Theta'$  is the sharing parameters between them.  $\Theta = \{\Theta', \Theta_1, \dots, \Theta_T\}$

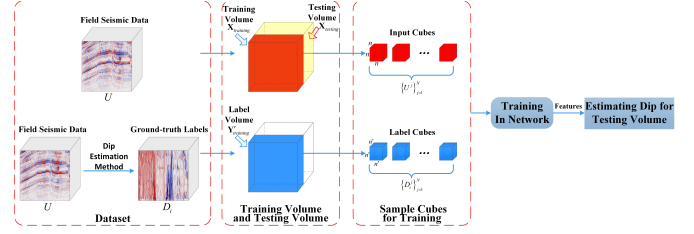


Fig. 1. Procedure for each specific task of dip estimation.

are the set of network parameters to be optimized by minimizing  $\mathcal{L}$ .  $w_i$  is the weight of each task and determined by its importance in the holistic networks. In particular, when  $w_1 = 1$  and  $w_2 = 0$ , this multitask learning method is equal to a single-task method for inline dip estimation, while  $w_1 = 0$  and  $w_2 = 1$  are degraded as a single-task method for crossline dip estimation.

As analyzed in Section I, we treat each SDE problem as a regression problem rather than a classification problem to meet the requirement of accuracy and speed at the same time. We denote the original 3-D seismic data volume as  $U$ , the inline dip volume, and crossline dip volume as  $D_1$  and  $D_2$ , respectively. Since conventional methods always estimate the seismic apparent dip using one sliding window, we do not need the whole seismic volume but local 3-D seismic cubes of training datasets for network training. Therefore, we cut them into overlapping subvolumes and construct the set of training sample pairs  $\{U^j, D_1^j, D_2^j\}_{j=1}^N$ , where  $N$  is the total number of samples,  $U^j \in \mathbb{R}^{n \times n \times n}$ , and  $D_1^j$  and  $D_2^j \in \mathbb{R}^{n' \times n' \times n'}$  ( $n' < n$  to ensure sufficient input information). With respect to each task, we assume a direct mapping function  $h_i(U^j; \Theta', \Theta_i) : \mathcal{U} \rightarrow \mathcal{D}_i$  to map the input original seismic data to seismic apparent dip. The procedure for each task is shown in Fig. 1. The loss function of each task that we used is given as follows:

$$\mathcal{L}_i(\Theta', \Theta_i) = \frac{1}{2N} \sum_{j=1}^N \|h_i(U^j; \Theta', \Theta_i) - D^j\|_F^2. \quad (2)$$

$\mathcal{L}_i(\Theta', \Theta_i)$  is used to measure the mean square error (mse) between the target dip label and the network predicted dip.

To obtain high-quality dip labels, we use the most reliable conventional method, WSSB dip estimation, to compute the seismic apparent dip. The 3-D seismic volume after migration can be indicated by  $u(t, x, y)$ , where  $t$ ,  $x$ , and  $y$  are time, inline, and crossline coordinates, respectively. To get the robust results of dip estimation against noise, we use the Hilbert transform to obtain the image part  $u^H(t, x, y)$ . By sliding the analysis window, we can compute every central point  $(t, x, y)$ 's waveform coherence [20] as (3), shown at the bottom of the next page.

Here,  $M$  denotes the half-window size in the time direction,  $J$  is the number of seismic traces in the analysis window,  $H$  denotes the Hilbert transform, and  $\theta_x$  and  $\theta_y \in [-\theta_{\max}, \theta_{\max}]$  are the preset dip indexes along inline and crossline directions, respectively. We can obtain a series of values of coherence over discrete apparent dip pairs  $(\theta_x, \theta_y)$  and choose the dip pair that maximizes the coherence as the apparent dip of the

analysis point. To further improve the apparent dip, on the one hand, we adopt a small dip scanning interval at the cost of heavy computation. On the other hand, following Marfurt and Kirilin [20], we fit a 2-D paraboloid through the nine discretely sampled points neighboring the point with maximum coherence.

### B. Network Architecture

Both estimation tasks of inline dip and crossline dip involve analyzing geometric features in the 3-D seismic data. We use a fully convolutional network to simultaneously extract features of seismic inline dip  $D_1$  and crossline dip  $D_2$  from the input seismic cube  $U$ . Fig. 2 shows the specific architecture of our network. We adopt a direct mapping architecture rather than residual learning because the magnitude difference between network input and output in our task is enormous, while residual learning is easy to learn small amplitude changes. Our network adopts a hard parameter sharing structure. It consists of four stages: the first two stages are shared layers that extract common features, and the last two stages are two branches of task-specific layers, which computes two outputs of an inline dip and a crossline dip. The first stage is an input layer composed of one convolutional layer and one activation function. Similar to the parameter setting in [21], we have 64 convolutional kernels, and their sizes are  $3 \times 3 \times 3$ . The 3-D convolutional kernels can capture more structural features of seismic data than 2-D kernels. We also delete all pooling layers to keep the output size the same as the input size. The rectified linear unit (ReLU), which helps solve the vanishing gradient problem and prevent overfitting, is set as the activation function. The second stage is composed of seven convolutional layers. Unlike the input layer, each of them is followed by a batch normalization (BN) layer before activation. The BN is used to accelerate training convergence and improve the accuracy of dip estimation. The common features extracted by shared layers are delivered to the two branches to implement specific dip estimation tasks. The third stage has a dual-channel structure and has more convolutional layers up to 10. In the convolution calculation process, padding the boundary is necessary to allow the data on the boundary to participate in the convolution calculation as a center. There are many kinds of padding operations, such as zero padding, symmetric boundary padding, and periodic padding. In this stage, we chose a different padding method from the previous same paddings. Valid padding [22] keeps the filter window staying at a valid position inside the input map and shrinks output size. We adopt valid padding at intervals to reduce the size of network output. Reducing the network output ensures that the voxels on the output boundary also have enough corresponding receptive fields. In other words, the ambiguity of the network output on the boundary can be reduced in this

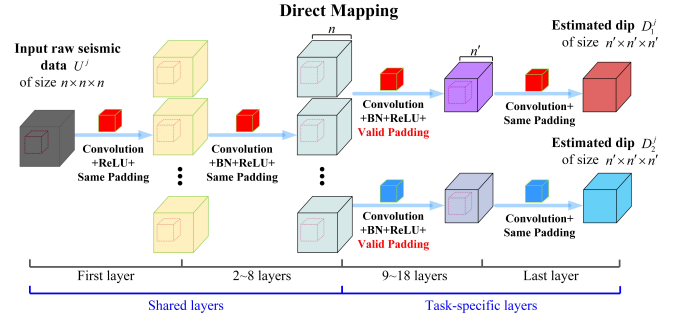


Fig. 2. Deep neural network with shared convolutional layers and task-specific layers.

way. The last stage is the output layer, which only has one convolution layer to output seismic apparent dips' values.

### C. Network Training

We train the network by minimizing the network's total loss in (1) through gradient-based optimization algorithms. Parameters of shared layers are learned jointly together, and parameters of task-specific layers are learned using an alternating minimization scheme. The task-specific layers consisting of two branches are, respectively, dedicated to tasks of inline and crossline dip estimation. The weight of each task is a crucial hyperparameter in multitask learning. We notice that the two branches have almost identical structures, and their output ranges are the same. In other words, their gradients need to be backpropagated with the same impact on the shared layers. Therefore, we adopt the static update method to facilitate the training process, which sets the weights  $w_1 = 0.5$  and  $w_2 = 0.5$  manually before the training of the networks, and they are fixed during the whole training of the network.

## III. DATA EXAMPLES

We evaluate our method on two 3-D field seismic datasets originating from the same survey acquired by Daqing Oilfield. The first dataset has 681 lines with a spacing of 20 m and 401 traces with a spacing of 20 m. The time sampling point is 501 with a 1-ms sampling rate. We first construct the relatively high-quality training dataset and use the WSSB dip estimation to estimate the seismic apparent dips along with the inline and crossline directions. We set the dip searching range  $[-\theta_{max}, \theta_{max}]$  as  $[-4, 4]$  with a small dip searching interval of 0.125 to increase dip resolution. The first 150 lines of the original seismic data are designated as the training volume, while the first 150 lines of the estimated dips are selected as the corresponding ground-truth labels. Then, we use a fixed-size sliding window with  $50 \times 50 \times 50$  voxels and a stride of 20 to sequentially cut the training volumes into overlapping subvolumes, as shown in the smaller boxes of different

$$S(\theta_x, \theta_y) = \frac{\sum_{n=-M}^M \left\{ \left[ \sum_{j=1}^J u(t-n-\theta_x x_j - \theta_y y_j) \right]^2 + \left[ \sum_{j=1}^J u^H(t-n-\theta_x x_j - \theta_y y_j) \right]^2 \right\}}{J \sum_{n=-M}^M \left\{ \sum_{j=1}^J [u(t-n-\theta_x x_j - \theta_y y_j)]^2 + \sum_{j=1}^J [u^H(t-n-\theta_x x_j - \theta_y y_j)]^2 \right\}} \quad (3)$$

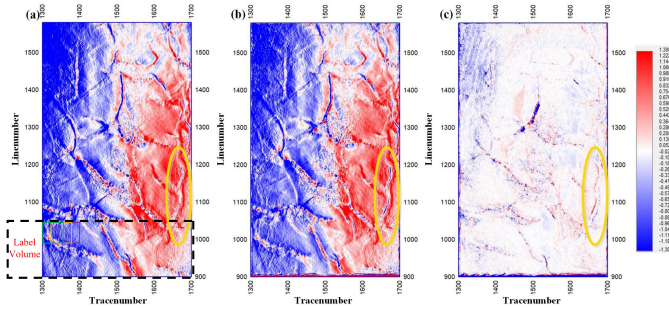


Fig. 3. First dataset's time slices of (a) WSSB dip estimation result, (b) our trained network result, and (c) difference between (a) and (b).

colors in Fig. 3(a). Thus, we constructed 8190 training sample pairs. After feeding these training sample pairs, the network will automatically update the parameters and learn feature representation for estimating seismic apparent dip from the raw seismic dataset. Finally, we use the well-trained network to estimate the apparent dip of the rest 531 lines, and the network simultaneously computes the inline and crossline dip.

First, we compare the calculation efficiency of WSSB dip estimation with our method. When we estimate the seismic apparent dip using the WSSB dip estimation, the parameters' set is consistent with those of constructing training labels, that is, the dip searching range is  $[-4, 4]$  and the dip searching interval is 0.125. The WSSB is computationally intensive. To reduce the calculation time as much as possible, the program is implemented in C++ and run on a workstation with the following specifications: an Intel Xeon E5-1620 v3 @ 3.50 GHz and 32-GB RAM. This program also employs multithreading to realize parallel computing with four cores, and the total consuming time of WSSB dip estimation is 4.7 h. When we use the well-trained network to the same 3-D dataset using one GPU (GeForce GTX 1080Ti), we cut the full 3-D seismic data into overlapping cubes and stack two full 3-D volumes of seismic dips after estimating each cube with the network. The total consuming time is about 6 min, which is greatly shortened. Therefore, our SDE-CNN is more efficient than the commonly used WSSB dip estimation.

Then, we compare the results of dip estimation on a time slice in Fig. 3. For the sake of saving space, only results of inline direction are displayed. Fig. 3(a) and (b) shows two time slices of dip estimation results based on the conventional method and our trained SDE-CNN, respectively. Except for the area near the data boundary, there is no significant difference between the WSSB dip estimation results and the dip estimation result based on our trained network. Moreover, we notice that our method is superior to the conventional method in some regions because Fig. 3(b) (the result of our trained SDE-CNN) has a more accurate and sharper channel edge, as indicated by the orange ellipse. To further evaluate our method's estimation accuracy, we also calculate the difference between the WSSB dip estimation's result and our result. Fig. 3(c) shows the time slice of the difference cube. Although there is some visible difference neighboring some big faults, the rest regions' errors are tiny. We speculate that the limited dip searching range caused the visible difference near the big fault.

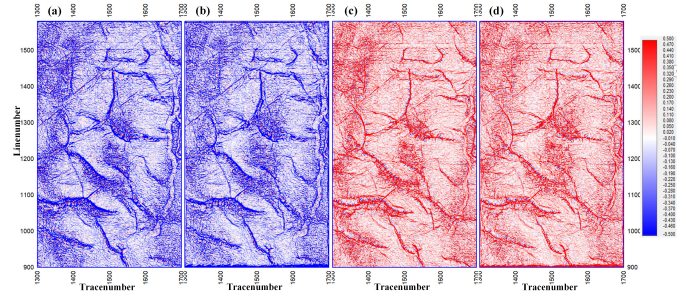


Fig. 4. First dataset's time slices of the most negative curvature based on (a) WSSB dip estimation and (b) our trained network result, and two time slices of the most positive curvature based on (c) WSSB dip estimation and (d) our trained network result.

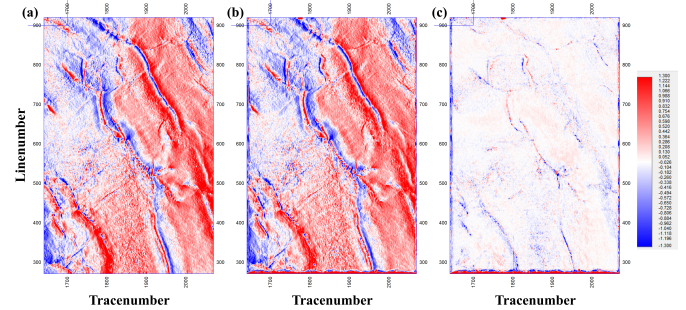


Fig. 5. Second dataset's time slices of (a) WSSB dip estimation result, (b) our trained network result, and (c) difference between (a) and (b).

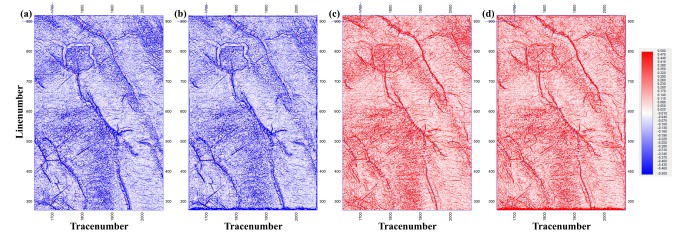


Fig. 6. Second dataset's time slices of the most negative curvature based on (a) WSSB dip estimation and (b) our trained network result, and two time slices of the most positive curvature based on (c) WSSB dip estimation and (d) our trained network result.

The structural curvature, which can be obtained by calculating the partial derivatives of apparent dips, plays a vital role in exploring and developing complex structural reservoirs. To test our network's performance in detail, we calculate the structural curvature based on the estimated dip of WSSB and our proposed method, respectively. The corresponding slices of the most negative curvature and the most positive curvature are shown in Fig. 4. We can also observe that our network can generate curvature results comparable to the curvature results based on WSSB dip estimation except for the region near the boundary.

To further test the network's generalization performance, we estimate dips of another dataset that has 650 lines, 430 traces, and 1001 time samples without retraining the network. Spatial dimensions and the time sampling rate are the same as the previous dataset. The blue lines in the upper left corner of Fig. 5 indicate the location relative to the previous dataset. The results of dip estimation and structural curvature are shown in Figs. 5 and 6, respectively. We can see that our results are very close to the results of WSSB. In other

words, the network is universal for datasets belonging to the same work area. However, the network's consumption time of 12.5 min is far less than 7.5 h of WSSB, which shows that the network can significantly reduce the dip estimation time of the same work area. In summary, our network obtains the ability to output reliable dip estimation results after learning the training samples.

#### IV. CONCLUSION

We propose a seismic apparent dip estimation method based on deep learning. A multitask convolutional regression model is introduced to simultaneously map the poststack seismic dataset to inline dip and crossline dip. We have also carefully designed the network architecture, and the network is trained and tested on the field seismic dataset. Compared with the conventional dip estimation method, our well-trained network can significantly save the computing time and provide satisfactory and comparable results of dip estimations except for the boundary area.

In addition, there are two issues that need further study. One issue is to reduce the training time. We used four GPUs to train the network in parallel for 18.92 h. Although the network test is computationally efficient, only when the seismic survey area is large or the dip resolution is high, the network, including the training time, has obvious advantages over the traditional method. The other issue is that the network generalization between different work areas is poor. We think that transfer learning is a powerful tool to help solve these two issues. Transfer learning refines the network to be adapted to seismic data from other work areas, reducing training time and improving generalization.

#### REFERENCES

- [1] M. T. Taner, J. S. Schuelke, R. O'Doherty, and E. Baysal, "Seismic attributes revisited," in *Proc. SEG Tech. Program Expanded Abstr.*, Oct. 1994, pp. 1104–1106.
- [2] M. Bahorich and S. Farmer, "3-D seismic discontinuity for faults and stratigraphic features: The coherence cube," *Lead. Edge*, vol. 14, no. 10, pp. 1053–1058, 1995.
- [3] A. E. Barnes, "Theory of 2-D complex seismic trace analysis," *Geophysics*, vol. 61, no. 1, pp. 264–272, 1996.
- [4] A. E. Barnes, "Weighted average seismic attributes," *Geophysics*, vol. 65, no. 1, pp. 275–285, 2000.
- [5] B. Milkereit, "Decomposition and inversion of seismic data—An instantaneous slowness approach," *Geophys. Prospecting*, vol. 35, no. 8, pp. 875–894, 1987.
- [6] K. J. Marfurt, R. L. Kirlin, S. L. Farmer, and M. S. Bahorich, "3-D seismic attributes using a semblance-based coherency algorithm," *Geophysics*, vol. 63, no. 4, pp. 1150–1165, 1998.
- [7] K. J. Marfurt, V. Sudhaker, A. Gersztenkorn, K. D. Crawford, and S. E. Nissen, "Coherency calculations in the presence of structural dip," *Geophysics*, vol. 64, no. 1, pp. 104–111, 1999.
- [8] P. Bakker, L. J. van Vliet, and P. W. Verbeek, "Edge preserving orientation adaptive filtering," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 1999, pp. 535–540.
- [9] X. Wang, W. Chen, and Z. Zhu, "Robust seismic volumetric dip estimation combining structure tensor and multiwindow technology," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 395–405, Jan. 2019.
- [10] S. W. Chen and C. S. Tao, "PolSAR image classification using polarimetric-feature-driven deep convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 627–631, Apr. 2018.
- [11] L. Tao, Y. Zhou, X. Jiang, X. Liu, and Z. Zhou, "Convolutional neural network-based dictionary learning for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 10, pp. 1776–1780, Oct. 2021.
- [12] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, "Sequential deep learning for human action recognition," in *Human Behavior Understanding* (Lecture Notes in Computer Science), vol. 7065, A. A. Salah and B. Lepri, Eds. Berlin, Germany: Springer, 2011, pp. 29–39. [Online]. Available: [https://link.springer.com/chapter/10.1007%2F978-3-642-25446-8\\_4](https://link.springer.com/chapter/10.1007%2F978-3-642-25446-8_4), doi: 10.1007/978-3-642-25446-8\_4.
- [13] D. Liu, W. Wang, X. Wang, C. Wang, J. Pei, and W. Chen, "Poststack seismic data denoising based on 3-D convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1598–1629, Mar. 2020.
- [14] H. Zhang, X. Yang, and J. Ma, "Can learning from natural image denoising be used for seismic data interpolation?" *Geophysics*, vol. 85, no. 4, pp. WA115–WA136, 2020.
- [15] H. Ma, H. Yao, Y. Li, and H. Wang, "Deep residual encoder-textendashdecoder networks for desert seismic noise suppression," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 3, pp. 529–533, Mar. 2020.
- [16] X. Wu, Y. Shi, S. Fomel, L. Liang, Q. Zhang, and A. Z. Yusifov, "FaultNet3D: Predicting fault probabilities, strikes, and dips with a single convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9138–9155, Nov. 2019.
- [17] H. Di, Z. Wang, and G. AlRegib, "Deep convolutional neural networks for seismic salt-body delineation," in *Proc. AAPG Annu. Conv. Exhib.*, 2018, pp. 1–12.
- [18] T. Zhao and P. Mukhopadhyay, "A fault-detection workflow using deep learning and image processing," in *Proc. SEG Tech. Program Expanded Abstr.*, Anaheim, CA, USA, Oct. 2018, pp. 1966–1970.
- [19] R. Caruana, "Multitask learning," *Mach. Learn.*, vol. 28, no. 1, pp. 41–75, 1997.
- [20] K. J. Marfurt and R. L. Kirlin, "3-D broad-band estimates of reflector dip and amplitude," *Geophysics*, vol. 65, no. 1, pp. 304–320, 2000.
- [21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [22] A. Wiranata, S. A. Wibowo, R. Patmasari, R. Rahmania, and R. Mayasari, "Investigation of padding schemes for faster R-CNN on vehicle detection," in *Proc. Int. Conf. Control, Electron., Renew. Energy Commun. (ICCEREC)*, 2018, pp. 208–212.