Multitask Training as Regularization Strategy for Seismic Image Segmentation

Surojit Saha, Wasim Gazi, Rehman Mohammed, Thomas Rapstine, Hayden Powers, and Ross Whitaker, *Fellow, IEEE*

***Abstract*— This letter proposes multitask learning as a regularization method for segmentation tasks in seismic images. We examine application-specific auxiliary tasks, such as the estimation/detection of horizons, dip angle, and amplitude that geophysicists consider relevant for identification of channels (a geological feature), which is currently done through painstaking outlining by qualified experts. We show that multitask training helps in better generalization on test datasets with very similar and different structure/statistics. In such settings, we also show that multitask learning performs better on unseen datasets relative to the baseline.**

***Index Terms*— Multitask learning, seismic interpretation.**

I. Introduction

SEISMIC images are one of the principal ways for visualizing the *crust* of the solid Earth, for the identification of resources. Seismic images show sedimentary, metamorphic, and igneous rocks depending on the depth.

Seismic images contain different interesting seismic objects or geobodies such as faults, salt domes, unconformities, channels, and others. A *fault* is a fracture or discontinuity in a volume of rock caused due to plate tectonic forces resulting in large displacement between the layers. A *salt dome* is a dome-like structure formed when evaporite materials such as salt intrude into the overlying layers. An *unconformity* is a surface where part of the section is removed/eroded, followed by a subsequent deposition (at a later geologic period). The sections above and below an unconformity are very different in terms of strata and geometry, and the seismic reflection at the unconformity is often very strong. Seismic channels or simply *channels* are the channels of ancient rivers observed in the seismic images. They are the cardinal part for constructing hydrocarbon reservoir [1], fluvial reservoirs [2] characterization, ancient channel geomorphology [3], and well control [4], and thus are of immense interest to geologists and geophysicists. In this letter, in the interest of brevity, we study the identification/segmentation of seismic channels because of both their importance and difficulty in segmentation.

Manuscript received 7 July 2023; revised 8 October 2023 and 9 October 2023; accepted 17 October 2023. Date of publication 3 November 2023; date of current version 14 November 2023. This work was supported by ExxonMobil Upstream Inc. *(Surojit Saha and Wasim Gazi contributed equally to this work.) (Corresponding author: Surojit Saha.)*

Surojit Saha, Wasim Gazi, Rehman Mohammed, and Ross Whitaker are with the Scientific Computing and Imaging Institute, Kahlert School of Computing, The University of Utah, Salt Lake City, UT 84112 USA (e-mail: surojit@cs.utah.edu; wasim.gazi@utah.edu; rehman.mohammed@utah.edu; whitaker@cs.utah.edu).

Thomas Rapstine and Hayden Powers are with ExxonMobil Upstream Integrated Solutions Company, Houston, TX 77389 USA (e-mail: thomas.d.rapstine@exxonmobil.com; hayden.powers@exxonmobil.com).

Digital Object Identifier 10.1109/LGRS.2023.3328837

A. Related Work

The boundaries of channels often appear as discontinuities with different seismic attributes, such as sweetness [5], semblance [6], [7], and coherence [8], [9], [10]. However, seismic discontinuities are observed for other geobodies, and complex geologies can result in many false positives for simple thresholding or detection algorithms [8], [9], [10], [11]. Thus, researchers have proposed CNNs for channel detection in seismic images. For instance, Pham et al. [12] were able to achieve some success for a dataset with narrow channels using SegNet [13] for segmentation along with Bayesian [14] SegNet for uncertainty measurement. However, their model could not distinguish individual channels well, and prediction uncertainty was high. To address the lack of annotated data methods in [15] and [16] generate synthetic data. However, the statistics learned from the synthetic data by deep neural networks (DNNs) often differ from real-world data.

B. Generalization and Multitask Learning

The complex image statistics and geometries with relatively little training data make the channel segmentation task challenging. Seismic images are expensive, highly variable, and often proprietary, while interpretation requires significant time from highly trained experts. Without sufficient labeled training data, learning-based methods tend to overfit the training data, resulting in poor generalization.

Multitask learning [17], using self-supervised auxiliary tasks, is an effective way to incorporate domain knowledge and unlabeled data into training with limited labels. Self-supervised tasks can be very generic, such as contrastive loss functions or image completion [18], [19], or they can be specific to the application domain, such as the interpretation of biomedical images. In this letter, we consider the application-specific multitask learning, similar to [20] and [21], which proposes self-supervised auxiliary tasks to improve the primary task, such as the classification of histology images in [20]. The auxiliary tasks are referred to as self-supervised, as the labels used in training the auxiliary tasks are produced by processing input data. In this work, we use analytical methods (not annotated by humans nor generated using any automated tool) to produce labels. Cotraining the model to

solve the auxiliary task along with the primary task provides a regularization, and thus, we expect it to lead to better generalization.

The application scenario we consider is the following. There is a moderately small set of interpreted (segmented) seismic slices and a large set of unlabeled data, e.g., from various seismic surveys/volumes for training. A user would like to adapt/apply the previously trained model to a new survey with *very few* labeled examples (e.g., one to five slices) from the new dataset. Thus, the scenario relates to domain adaptation, transfer learning, and few-shot learning. Our experiments have found that conventional domain adaptation techniques (e.g., maximum mean discrepancy (MMD) and adversarial adaptation [22]) perform poorly due to the variations between the datasets and the limited labeled data. Thus, this letter focuses on developing self-supervised multitask training approaches to improve generalization.

The contributions of this letter are given below.

1) Defining/proposing domain-specific, self-supervised auxiliary tasks for regularizing DNNs designed to segment channels in seismic images.
2) Analysis and demonstration of the efficacy of the proposed multitask learning for segmentation of channels.
3) Exhibition of the effectiveness of multitask learning in scenarios with limited annotated data to interpret complex seismic images.

II. Auxiliary Tasks

Channels in seismic images are typically characterized by the changes in the properties of horizons, such as termination and deformation of horizons. In addition, we observe the variations in amplitude for channels relative to the surrounding layers. Motivated by such observations and discussions with domain experts, we devised the following subtasks: *horizon* detection to study the irregularity in horizons, *dip angle* estimation to capture the structural variations in horizons, and *horizon amplitude* estimation to detect channels.

Horizons in seismic images are the visible low- and high-intensity lines that indicate strong reflections associated with material boundaries in the geology (e.g., the layer cake associated with sedimentary processes). For the self-supervised approach, we have formulated the horizon label generation as a high and low ridge detection problem. A ridge in an image is defined as a local minimum or maximum in a specific direction [23]. In seismic images, this direction is orthogonal to the sedimentary layers. Thus, we use a structure tensor (at a given scale) to find the first-order directional derivatives, followed by a zero-crossing algorithm to identify loci of minimum/maximum values along this direction. Additional parameters, such as smoothing and the threshold for minimum/maximum values and curvature, control the number and strength of detected horizons. Fig. 1 shows an example of horizons detected in a seismic image.

The *dip angle* is the difference in orientation of a vector relative to the vectors perpendicular to the prevailing set of horizons (or layer-cake structure). We use the sine of the angle between the principal eigenvector of the structure tensor (used in the processing of horizons) and the vertical direction to measure the dip angle, as shown in the example in Fig. 1.

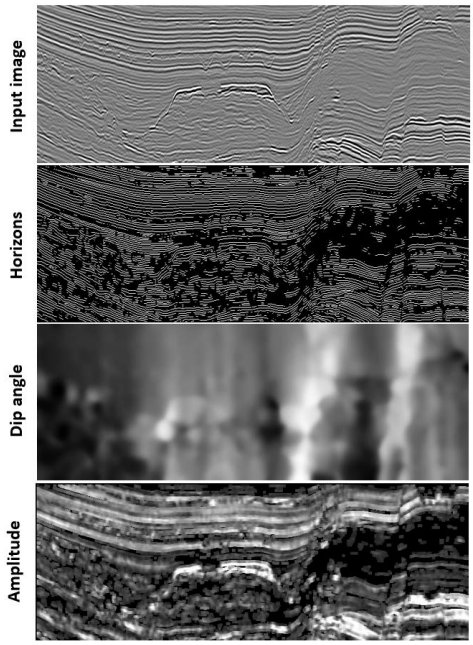

Fig. 1. Examples of seismic image and self-supervised, auxiliary tasks considered in this work.

The *horizon amplitude* is the instantaneous amplitude of the wave packets associated with the reflections in the seismic images (e.g., the amplitude of horizons). For this, we compute the signal's magnitude at all the detected horizons and interpolate this value into the entire image using kernel regression. This is shown in Fig. 1.

III. Experimental Setup

The data used for this work are a set of volumetric regions (subvolumes) from the somewhat large *Parihaka 3-D* dataset, provided by the New Zealand Crown Minerals [24]. We identified two regions of the volume that contain channel systems, one shallow (1000 2-D images of size 151×1151) in the survey and one much deeper (1126 2-D slices of size 268×923), which have significantly different geometries and statistics—in the seismic frequencies, the underlying deformations, and the structure of the channel systems. The channels were segmented in both the shallow (1000 annotations) and deep subvolumes (24 out of 1126 slices, spaced throughout the volume) by a professional seismic interpreter, henceforth referred to as Parihaka-shallow and Parihaka-deep datasets, respectively. Annotation of only 24 slices in the Parihaka-deep dataset presents a practical scenario where we deal with *very few* labels from a new survey. Unlike the analysis of photographic images, seismic interpretation entails a great deal of subjectivity and variability between interpreters. To address this, the interpreter focused only on regions of high confidence for labeling channel and nonchannel areas. This information is used to produce a mask, referred to herein as a confidence mask, used in training (for computing loss) and evaluating models (computing mean intersection over union (mIoU) scores). Fig. 2 shows sections from the Parihaka-shallow and Parihaka-deep datasets with the corresponding segmentation and confidence masks.

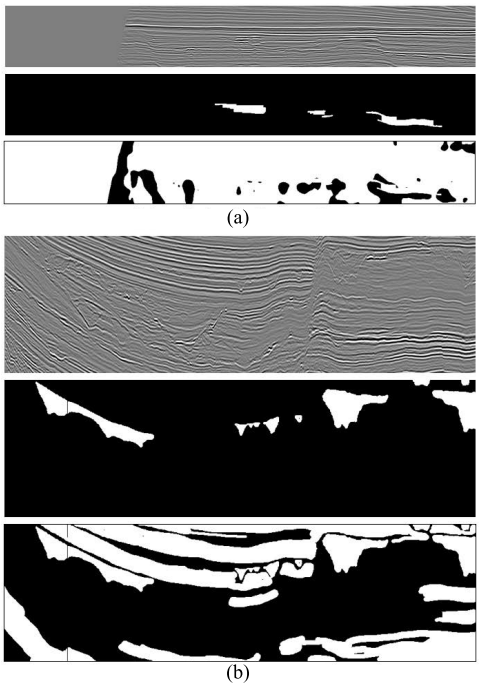

Fig. 2. (top) Seismic image, (middle) channel labels, and (bottom) confidence mask for the experimental datasets (a) Parihaka-shallow and (b) Parihaka-deep, toward the end of both the volumes.

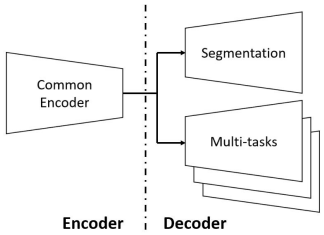

Fig. 3. Multitasking architecture using separate decoders for segmentation and auxiliary tasks with a shared encoder.

A. Computational Approach

1) Architecture: For the neural network, we have used a U-net architecture [25] as in Fig. 3. It comprises a common encoder layer followed by one segmentation decoder layer and an optional decoder layer for each auxiliary task. The details of the encoder and decoder are reported in Table I, where C64 is a convolution layer producing 64 channels. Convolution operations within parentheses, such as {C64}, are followed by the batch normalization (BN) layer and rectified linear unit (ReLU) activation. We use (3×3) convolution filters in all the layers. The M2 is a 2-D max pool operation with a kernel size of 2 that reduces the spatial size of the tensor in the encoder. The TC is a transpose convolution layer in the decoder that doubles the spatial size of the input tensor and reduces the channel depth by half. The skip connection concatenates, CON($N + N$), the output of the transpose convolution of the decoder with the corresponding encoder layer along the feature map depth (aka channels). The sigmoid activation is applied on the decoder output, C1 (without activation), for tasks using the mse loss.

2) Training, Validation, and Testing: The data volume is split into the training (80%) and validation (20%) volume. For training, we sample 1000 (64×64) patches at random from the 3-D training volume in each epoch for both the datasets.

TABLE I
ENCODER AND DECODER ARCHITECTURE USED FOR THE PRIMARY AND AUXILIARY TASKS. THE ARROWS INDICATE FLOW OF DATA IN THE MODEL

Encoder (Input)	Decoder (Output)
{2×C64} → MP2↓	CON(64+64) → {2×C64} → C1
{2×C128} → MP2↓	CON(128+128) → {2×C128} → TC64↑
{2×C256} → MP2↓	CON(256+256) → {2×C256} → TC128↑
{2×C512} → MP2↓	CON(512+512) → {2×C512} → TC256↑
{2×C1024} → Decoder	TC512↑

Validation loss is computed using a fixed set of patches that are extracted from the validation data a priori. We have used a minibatch size of 64 and trained the models for a maximum of 100 epochs. We have used standard data augmentation techniques such as random horizontal flip, random rotation, random rescaling, and noise, which are applied to members of a minibatch at random. We have found a learning rate of 0.01 for the stochastic gradient descent (SGD) (with Nesterov momentum) optimizer to be effective for our experiments. Training of the model (shown in Fig. 3) for the primary segmentation task and different multitasks involves updating the decoder parameters associated with a multitask (regularization strategy) by the corresponding loss. Nevertheless, the encoder parameters are updated by the combined gradient from the segmentation loss and different multitask losses. For testing, the CNN processes whole slices for speed to mimic practical use cases. Moreover, the patchwise evaluation (used in patch stitching) did not provide better results during model evaluation.

3) Loss Function: Segmentation (primary task) and ridge detection are binary classification tasks, and therefore, we use weighted binary cross-entropy (BCE with logits) as the loss for those labels. For nonbinary auxiliary tasks (i.e., dip angle and amplitude), we have used mse loss. When the segmentation network is trained in conjunction with the auxiliary tasks, the auxiliary losses are added to the primary segmentation loss without any scaling factor. This removes the challenge of tuning the task-specific weighing factors (hyperparameters).

4) Evaluation Metric: For evaluating relative performance, we use the Jaccard index or intersection over union (IoU) as the accuracy measurement. For calculating the IoU score, we threshold the decoder output at 0.5. This evaluation has several limitations. The IoU score does change with different choices of the threshold value, but a complete sensitivity study of this parameter is beyond the scope of this letter. The IoU score also penalizes minor differences in boundaries of objects (especially small objects) that are, in this application, nonexact by nature. Thus, a good/acceptable model prediction might have an IoU score of less than 1.0. The ground-truth label is interpreted with the help of the associated confidence mask. Therefore, for our work, due to the lack of a better accuracy measurement, the IoU score serves only as an indicator, and visual feedback from the expert is considered for holistic model performance evaluation.

IV. RESULTS

Here, we compare the proposed multitask training with two baseline approaches: 1) a CNN trained on the Parihaka-shallow and 2) a CNN cotrained on Parihaka-shallow and a small set

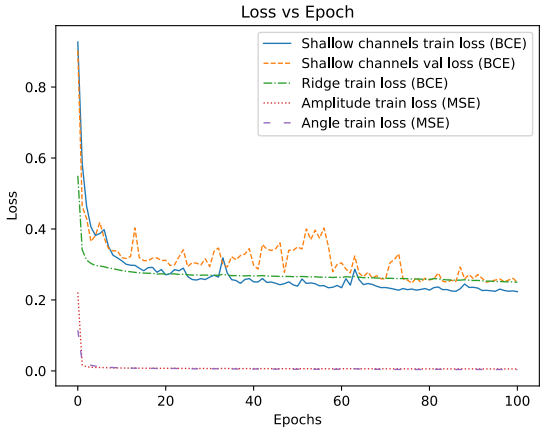

Fig. 4. Training loss versus epochs show that the network trains well for segmentation in conjunction with the auxiliary tasks.

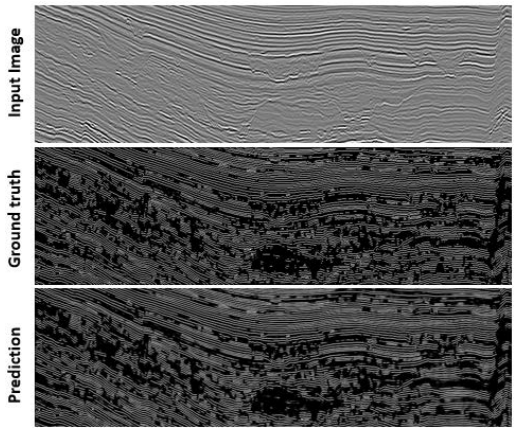

Fig. 5. Prediction of horizons by the model on the P-deep test data when trained on P-shallow and five slices of P-deep.

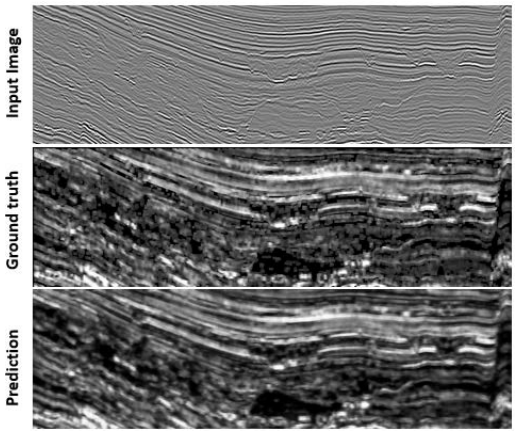

Fig. 6. Prediction of horizon amplitude by the model on the P-deep test data when trained on P-shallow and five slices of P-deep.

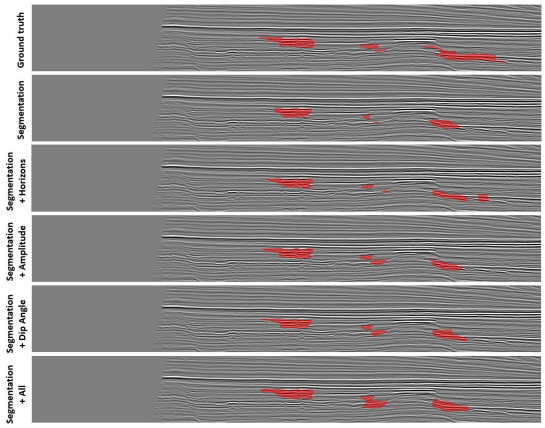

Fig. 7. Segmentation results (in red) for different scenarios on the P-shallow test data when trained on P-shallow only.

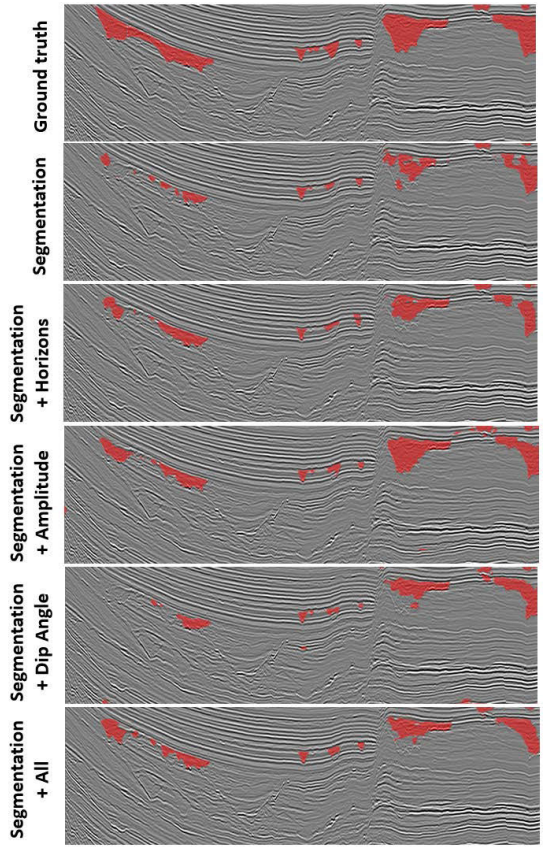

Fig. 8. Segmentation results (in red) on the P-deep test data when trained on P-shallow and five slices of P-deep.

of labeled Parihaka-deep and then tested on both the shallow (38 test slices) and deep data (19 test slices). For held-out data, we always consider slices not adjacent to or near the training data. For the multitask training on the Parihaka-shallow data, we observe that the auxiliary tasks (green, red, and magenta curves in Fig. 4) train well with the primary segmentation task (blue and orange curves), even without using any task-specific hyperparameters. Figs. 5 and 6 show the model predictions for different auxiliary tasks.

Table II shows the mIoU results for both the shallow and deep test data under various training scenarios. The same test dataset (for both the volumes) is used to report the mIoU scores. For training only on the shallow data, we see that the combination of all the multitask regularization achieves the best performance (highlighted in bold), followed by the regularization with amplitude (in italics). A similar trend is observed for Parihaka-shallow data even when it is cotrained with few slices from the deep data, where the best-performing method is produced by amplitude (using three slices) and a combination of all the multitasks (using five slices). For the datasets considered in this work, horizon amplitude is observed to be an important regularization strategy, as it produces the best mIoU score even for the Parihaka-deep data with limited supervision (three and five slices). Other

TABLE II
MIOU SCORES FOR CHANNEL SEGMENTATION ON THE PARIHAKA-SHALLOW AND PARIHAKA-DEEP TEST DATA. BEST MIOU SCORES ARE REPORTED IN **BOLD** AND THE SECOND BEST IN *ITALICS*

	Training Scenario	P-shallow (mIoU over 38 slices)	P-deep (mIoU over 19 slices)
Training on P-shallow only (no P-deep supervision)	Segmentation	0.497	0.039
	+ Horizons	0.553	0.202
	+ Amplitude	*0.557*	0.065
	+ Dip Angle	0.532	**0.222**
	+ All Aux Tasks	**0.572**	*0.219*
Training on P-shallow + 3 slices from P-deep	Segmentation	0.339	0.317
	+ Horizons	0.412	0.409
	+ Amplitude	**0.503**	**0.431**
	+ Dip Angle	0.413	*0.415*
	+ All Aux Tasks	*0.427*	0.327
Training on P-shallow + 5 slices from P-deep	Segmentation	*0.386*	0.475
	+ Horizons	0.277	*0.516*
	+ Amplitude	0.316	**0.587**
	+ Dip Angle	0.271	0.420
	+ All Aux Tasks	**0.478**	0.506

regularization methods proved to be effective under different scenarios, such as the dip angle and horizons producing the second-best score for the Parihaka-deep data using three and five slices, respectively. The combination of all the auxiliary tasks is possibly leading to overregularization for the Parihaka-deep volume, resulting in low mIoU scores for both three and five slices. Drop in the best mIoU scores for the Parihaka-shallow data when cotrained with Parihaka-deep slices could be attributed to the difference in structure/statistics between the volumes that makes the problem harder. Overall improvement in model performance over the baseline using the proposed regularization strategies demonstrates the efficacy of the regularization methods.

The qualitative results also show some improvement from the multitask regularization. Figs. 7 and 8 show segmentations on test data (overlaid on the seismic data in red) under various training scenarios for shallow and deep, respectively. The training with auxiliary tasks shows fewer mispredictions and (in most cases) complete identification of channels.

V. CONCLUSION

This letter demonstrates the advantage of application-specific, self-supervised auxiliary tasks for better generalization in the limited training data scenario in seismic image processing. Under different testing scenarios, some of the regularization strategies are observed to be more effective than others, illustrating their advantage based on the statistics of the seismic features. Among all the proposed regularization strategies, horizon amplitude has consistently produced better results, followed by horizons and dip angle. Empirical evaluation of the proposed auxiliary tasks holds strong motivation for us to devise new self-supervised tasks to improve the interpretation of seismic data and extend this work to detect other seismic features and/or geobodies. Moreover, the rich latent representation built by the encoder using the proposed strategy finds potential application in several downstream tasks with limited supervision, aka few-shot learning.

REFERENCES

[1] T. H. D. Payenberg and S. C. Lang, "Reservoir geometry of fluvial distributary channels—Implications for Nothwest Shelf, Australia, deltaic successions," *APPEA J.*, vol. 43, no. 1, p. 325, 2003.

[2] J. S. Bridge and R. S. Tye, "Interpreting the dimensions of ancient fluvial channel bars, channels, and channel belts from wireline-logs and cores," *AAPG Bull.*, vol. 84, no. 8, pp. 1205–1228, 2000.

[3] R. Pu, L. Zhu, and H. Zhong, "3-D seismic identification and characterization of ancient channel morphology," *J. Earth Sci.*, vol. 20, no. 5, pp. 858–867, Oct. 2009.

[4] Y. Suarez, K. Marfurt, and M. Falk, "Seismic attribute-assisted interpretation of channel geometries and infill lithology: A case study of Anadarko basin red fork channels," *Seg Tech. Program Expanded Abstr.*, vol. 27, pp. 963–967, Jan. 2008.

[5] B. S. Hart, "Channel detection in 3-D seismic data using sweetness," *AAPG Bull.*, vol. 92, no. 6, pp. 733–742, Jun. 2008.

[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, Jul. 1998.

[7] D. Hale, "Structure-oriented smoothing and semblance," Center Wave Phenomena, Colorado School Mines, Golden, CO, USA, Tech. Rep., 2009.

[8] 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, Jan. 1999.

[9] X. Wu, "Directional structure-tensor-based coherence to detect seismic faults and channels," *Geophysics*, vol. 82, no. 2, pp. 13–17, Mar. 2017.

[10] A. Gersztenkorn and K. J. Marfurt, "Eigenstructure-based coherence computations as an aid to 3-D structural and stratigraphic mapping," *Geophysics*, vol. 64, no. 5, pp. 1468–1479, Sep. 1999.

[11] 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, Oct. 1995.

[12] N. Pham, S. Fomel, and D. Dunlap, "Automatic channel detection using deep learning," *Interpretation*, vol. 7, no. 3, pp. 43–50, Aug. 2019.

[13] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," 2015, *arXiv:1511.00561*.

[14] Z. Ghahramani, "Probabilistic machine learning and artificial intelligence," *Nature*, vol. 521, no. 7553, pp. 452–459, May 2015.

[15] H. Gao, X. Wu, and G. Liu, "ChannelSeg3D: Channel simulation and deep learning for channel interpretation in 3D seismic images," *Geophysics*, vol. 86, no. 4, pp. 73–83, Jul. 2021.

[16] X. Wu, L. Liang, Y. Shi, and S. Fomel, "FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation," *Geophysics*, vol. 84, no. 3, pp. 35–45, May 2019.

[17] R. Caruana, "Multitask learning," *Mach. Learn.*, vol. 28, pp. 41–75, Jul. 1997.

[18] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.

[19] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2536–2544.

[20] N. A. Koohbanani, B. Unnikrishnan, S. A. Khurram, P. Krishnaswamy, and N. Rajpoot, "Self-path: Self-supervision for classification of pathology images with limited annotations," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2845–2856, Oct. 2021.

[21] A. Barbalau et al., "SSMTL++: Revisiting self-supervised multi-task learning for video anomaly detection," *Comput. Vis. Image Understand.*, vol. 229, Mar. 2023, Art. no. 103656.

[22] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 723–773, Jan. 2012.

[23] D. Eberly, *Ridges in Image and Data Analysis*. Berlin, Germany: Springer, 1996.

[24] NZ Petroleum and Minerals (NZPM). (Oct. 2022). *Parihaka 3D Data*. [Online]. Available: https://wiki.seg.org/wiki/Parihaka-3D

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," 2015, *arXiv: 1505.04597*.