Transfer Learning Applied to Seismic Images Classification*

Daniel Chevitarese[1], Daniela Szwarcman[1], Reinaldo Mozart D. Silva[1], and Emilio Vital Brazil[1]

Search and Discovery Article #42285 (2018)**
Posted October 1, 2018

*Adapted from extended abstract prepared for poster presentation given at 2018 AAPG Annual Convention & Exhibition, Salt Lake City, Utah, May 20-23, 2018. Please view related article by the same authors in Search and Discovery Article #42286 (2018).
 DOI:10.1306/42285Chevitarese2018

[1]Research, IBM, Rio de Janeiro, Brazil (sallesd@ibm.com)

Abstract

Seismic imaging is an essential tool in oil and gas (O&G) exploration since it provides information about subsurface geometry and structural features. Also, by gathering other sources of data, it is possible to identify relevant localities of hydrocarbon accumulation. Current petroleum exploration demands analysis and interpretation of large volumes of seismic data within in strict deadlines. Therefore, computational systems that assist the expert in classifying subsurface features aiming to speed-up the analysis process are paramount to the industry development.

The growing popularity of deep learning inspired scientists to apply such methods to seismic data in real data sets. Although this technique has shown good results, a well known issue in deep learning systems is the difficulty to find a good starting point to adjust model's parameters. A poor initialization may lead to longer training sessions or to the inability of finding a solution. To address the initialization problem, we propose the use of transfer learning to set a good starting point to the parameters of our model. The idea behind this technique is to use previous knowledge obtained from one task in another. In our approach, to train a convolutional neural network (CNN) for a new data set, we initialize the model using the values of the parameters from a CNN trained with another seismic cube.

We conducted two main experiments using real seismic data sets from Scotia and Central-Graben (North Sea) basins. The first one was designed to verify if it is possible to train a model with a highly limited number of examples using previous knowledge. We proposed the second experiment to check whether training information from one cube would be useful to set a good starting point for the new model. The results showed performance improvements, even using cubes from regions that are not geologically similar.

Introduction

The workflow adopted by the O&G industry to extract subsurface knowledge involves seismic interpretation, which is one of its essential processes. By analyzing seismic images, the geoscientists can infer about subsurface structural settings and strata geometry. In addition, it is possible to locate areas of hydrocarbon accumulation by combining the interpretation results with other sources of information. However, seismic interpretation is a time-consuming and human-intensive task. Furthermore, geologists and geophysicists must deal with a continually increasing amount of data. This scenario has inspired researchers to propose computer systems to assist interpreters, aiming to reduce the time and improve the accuracy of the interpretation task.

Computer vision techniques have been applied to seismic facies analysis in the attempt to automate part of the interpretation process, as demonstrated in (Shafiq et al, 2017; Wang et al, 2016; Mottos et al., 2017; Amin and Deriche , 2016). Moreover, a few authors have presented works in which deep learning techniques are used in the seismic facies analysis context (Liu, 2017; Huang et al., 2017; Chevitarese et al., 2018a, 2018b). Regarding seismic facies classification specifically, in Chevitarese et al. (2018a, 2018b), the system splits seismic images into small patches of 40 x 40 pixels, and a deep convolutional neural network identifies to which facies that patch belongs.

We applied the transfer learning technique in the seismic facies classification context, using the dataset pre-processing presented in Chevitarese et al. (2018a) and the best convolutional network (CNN) introduced in Chevitarese et al. (2018b). We designed experiments to verify if the use of transfer learning can be valuable for our facies classification task in two scenarios. The first one simulates a situation where the available data for training the network is highly limited. The second one represents the cases in which a model could not be trained using a specific dataset. In such cases, we investigate if transfer learning can enable a model to find a reasonable solution.

This article is divided as follows: in Datasets, we briefly describe the datasets used in our experiments; in Transfer Learning Scheme, we define the transfer learning technique and our methodology; in Results and Discussion, we discuss our results; finally, we present our Conclusions.

Datasets

In our experiments, we used two public seismic datasets available at available at Open Seismic Repository (dGB Earth Sciences, 2017): Netherlands dataset, from Central-Graben Basin and Penobscot, from Scotia Basin. Netherlands cube comprises 951 crosslines of dimensions 651 × 462 pixels, and 651 inlines of size 951 × 462 pixels. Interpreters have labeled the cube, resulting in seven different facies or classes. Penobscot cube consists of 481 crosslines, and 601 inlines, with sizes 601 × 1501 pixels and 481 × 1501 pixels, respectively. Our specialists interpreted the cube marking eight different classes.

Notice that we used only inline slices to generate our datasets. Furthermore, we removed corrupted or poor-quality images from the edges of the cube, as indicated by an expert. After the removal, there were 459 inlines available for Penobscot, and 591 inlines for Netherlands data.

To train a CNN to classify seismic facies, we generated datasets for Netherlands and Penobscot following the ideas presented in Chevitarese et al. (2018a, 2018b). In these works, the inputs to the networks are small parts (size 40×40) of seismic images. The authors claim that the seismic images must be divided into small tiles because the input examples to the network must have the majority of its area belonging to only one facies. Otherwise, the model would not be able to learn the differences between facies. By fixing the size of these patches in 40×40 pixels, we would not obtain sufficient examples of the thinner layers, that is, the tiles would not have the majority of its pixels belonging to these layers.

In this case, the solution is to unite some classes in both cubes. Figure 1 shows the interpreted facies in (b) and the result after uniting the thinner layers in (c) for Netherlands cube. Figure 2 depicts the same process for Penobscot.

Regarding the pre-processing procedures, we also adopted the ones presented in Chevitarese et al. (2018a):

- Quantize the seismic images into 256 gray levels;
- Allow 30% of class interference (at most 30% of the pixels in a tile can belong to facies other than the main one in that tile).

It is important to mention that we must generate two tile datasets for each cube: one for training and one for validation. The latter is used to calculate the accuracy of the model in the classification task during training. To accomplish that, we separated the inline slices into train and validation sets so that we could guarantee the tile datasets would be created from different sets of images. For Netherlands and Penobscot, we divided the cube into ten blocks. In each block, the first 70% of inline slices were placed into the training set and the remaining 30% into the validation set. Figure 3 shows an example of these blocks.

Transfer Learning Scheme

The motivation behind transfer learning is the idea of applying knowledge learned earlier to solve new problems faster or with better solutions. More formally speaking, given a source domain *DS* and a task *Ts*, a target domain *DT*, and a task *TT*, transfer learning aims to help improve the learning of the target predictive function *fT* using the knowledge in *DS* + *Ts* (Pan and Yang, 2010). In our context, *DS* is the dataset of one seismic cube, *DT* the dataset from a different cube and the tasks are the classification of seismic facies for each dataset. The predictive function we are trying to learn is represented by the neural network model and its parameters.

Although deep neural networks have provided good results in the facies classification task (Chevitarese et al., 2018a, 2018b), a well-known issue in deep learning systems is the difficulty to find a good starting point to adjust model's parameters. A bad initialization may lead to longer training times or even to the inability of finding a solution. To address the initialization problem, we propose the use of transfer learning to set a good starting point for the parameters of our model. Here, we define as previous knowledge the values of the parameters learned in an earlier training session with a different dataset. In our transfer learning approach, to train the model for a new data set, we initialize it using the values of the parameters from a CNN trained with another seismic cube.

Figure 4 depicts the transfer learning workflow: the data processor generates tiles from the inline slices to feed the network; network parameters are initialized using one of two different methods; we train the model and evaluate it. For regular training, we randomly initialize the parameters. In the transfer learning context, we initialize the parameters with specific values, which represent our previous knowledge. These values were obtained in a regular training session, using a different but similar dataset.

We believe that there are two scenarios that one could benefit from the transfer learning technique:

1. The available data is highly limited and not sufficient to adequately train a deep CNN
2. The dataset can be challenging to learn from scratch, and may not be able to train the model.

In the next session, we detail the experiments we designed to investigate these scenarios.

Results and Discussion

In all experiments, we trained the networks for 200 epochs and used the same configuration for hyperparameters such as learning rate, weight decay, and dropout as described in Chevitarese et al. (2018b). We ran the training sessions on two Intel Xeon E5 CPU and one K40 NVIDIA GPU.

Experiment 1 involved simulating the environment in which only a few labeled slices are available for training. We used the CNN Danet-3 introduced in Chevitarese et al. (2018b), and we generated several tile datasets for Netherlands and Penobscot, by varying the number of blocks in which we split the cubes and the number of inline slices used. To help identify the results on the tables, we named them using a code that represents the number of blocks and the number of slices. For example, *b10-sl20* means that we divided the cube into ten blocks, and 20 slices were used to generate the dataset. Table 1 indicates all the generated datasets for Experiment 1.

The first step in Experiment 1 was to verify if the model can learn to classify facies from both Netherlands and Penobscot. For this verification, we trained Danet-3 using the tile datasets b10-sl414 for Netherlands and b10-sl322 for Penobscot. On these datasets, 70% of all the available slices were used to generate our tiles for training and 30% for validation. The final accuracy results were 99.55% and 99.40%, for Netherlands and Penobscot, respectively. We can then consider that our model is capable of learning the differences between seismic facies for both datasets when trained with a reasonable amount of data.

The second part of Experiment 1 consists of training Danet-3 with all the limited datasets indicated in Table 1 without transfer learning (control) and using previous knowledge (TL).

In all these training procedures, we used a batch size of 32 examples. For all TL rounds of training for Penobscot, we used the previous knowledge from the Netherlands model trained with *b10-sl414*. As for the Netherlands TL, we used the Penobscot model trained with *b10-sl322*.

The results are summarized in Table 2 for Penobscot (control and TL) and in Table 3 for Netherlands cube. We should highlight that we ran all the experiments three times, and we chose the best performing models in the TL scheme for each dataset. In all cases, it is possible to see an accuracy increase if one compares the control with the TL experiments. In the best case, we got 24.29% of accuracy improvement for Netherlands data and in the worst case, 0.48% for Penobscot.

Experiment 1 results show that in the highly limited amount of data for training, the transfer learning technique can be applied to improve model's performance.

As mentioned before, Experiment 2 was designed to evaluate if transfer learning can be useful when we are not able to train a model against a specific dataset. In this experiment, we used two Pre-salt datasets, which we call Pre-salt 1 and Pre-salt 2 to train Danet-1. We were able to train Danet-1 with Pre-salt 2 dataset, and the model presented 98.0% of accuracy. However, for Pre-salt 1, the results were not satisfactory, as demonstrated in Table 4. Pre-salt 1 has four classes, which imply that if we randomly select one we would get an accuracy of 25%; this means that the training was unsuccessful for that model in the control case since 25.7% of accuracy is almost the same as pure chance.

Experiment 2 also demonstrates that transfer learning can be a powerful tool for training deep CNNs for seismic facies classification. We were only able to train Danet-1 using our transfer learning scheme initialization.

Conclusions

In this work, we applied a transfer learning technique to train a deep neural network to classify seismic facies and assist the expert analysis. We show that the knowledge obtained by training a network with a specific seismic data set could be reused in a similar task. This means that once a model is trained, we can use the values of the parameters as the starting point to adjust the same model to another dataset. Our results show that it is possible to improve accuracy results in the seismic facies classification task by using previous knowledge from another similar task. Also, transfer learning demonstrated it could be an essential tool in deep learning for seismic facies classification as it enabled us to train a model against a dataset that could not be trained using random initialization.

References Cited

Amin, A., and M. Deriche, 2016, Salt-dome detection using a codebook-based learning model: IEEE Geoscience and Remote Sensing Letters, v. 13/11, p. 1636-1640.

Chevitarese, D.S., D. Szwarcman, R.M.G. e Silva, and E.V. Brazil, 2018a, Deep learning applied to seismic facies classification: A methodology for training: EAGE Saint Petersburg International Conference, April 9-12, 2018. Accepted for publication.

Chevitarese, D.S., D. Szwarcman, E.V. Brazil, and B. Zadrozny, 2018b, Efficient classification of seismic textures: 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, July 8-13, 2018. Accepted for publication.

dGB Earth Sciences, 2017, Open Seismic Repository. Website accessed September 18, 2018, https://www.opendtect.org/osr/.

Huang, L., X. Dong, and T.E. Clee, 2017, A scalable deep learning platform for identifying geologic features from seismic attributes: Leading Edge, v. 36/3, p. 249-256, 2017.

Liu, Y., 2017, “Application of deep learning for seismic image interpretation: (Presentation): GeoConvention, Calgary May 15-17, 2017.

Mattos, A.B. Mattos, R.S. Ferreira, R.M.D.G. e. Silva, M. Riva, and E. Vital Brazil, 2017, Assessing texture descriptors for seismic image retrieval: 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), p. 292-299.

Pan, S.J., and Q. Yang, 2010, A survey on transfer learning: IEEE Transactions on Knowledge and Data Engineering, v. 22/10, p. 1345-1359.

Shafiq, M. A., Y. Alaudah, G. AlRegib, and M. Deriche, 2017, Phase congruency for image understanding with applications in computational seismic interpretation:2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), p. 1587-1591.

Wang, Z., Z. Long, and G. AlRegib, 2016, Tensor-based subspace learning for tracking salt-dome boundaries constrained by seismic attributes: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), p. 1125-1129.

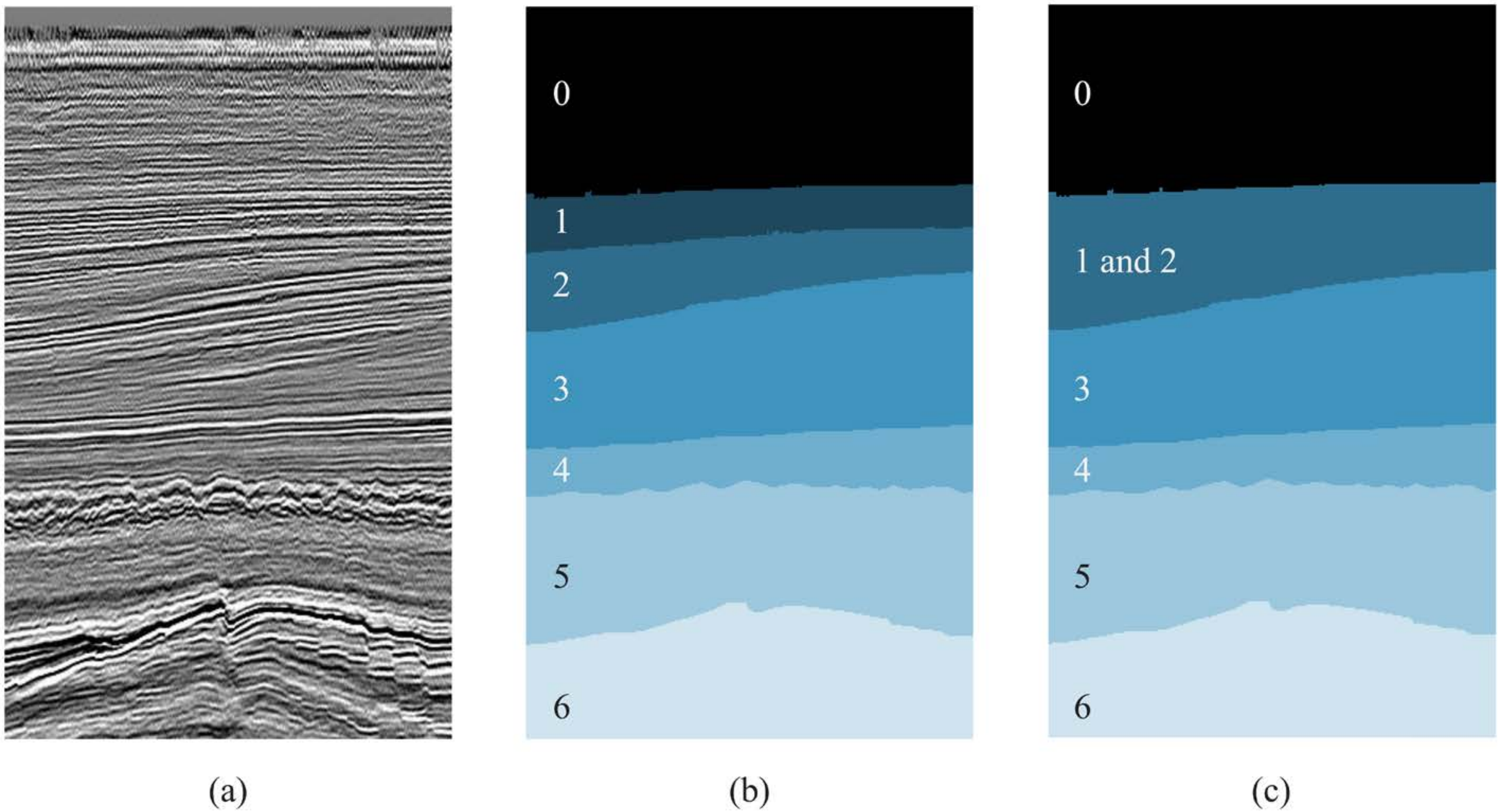

Figure 1. (a) Part of an inline of seismic data, Netherlands Central-Graben Basin (North Sea) and its respective mask with seven classes (b). In (c), the same mask with classes 1 and two united.

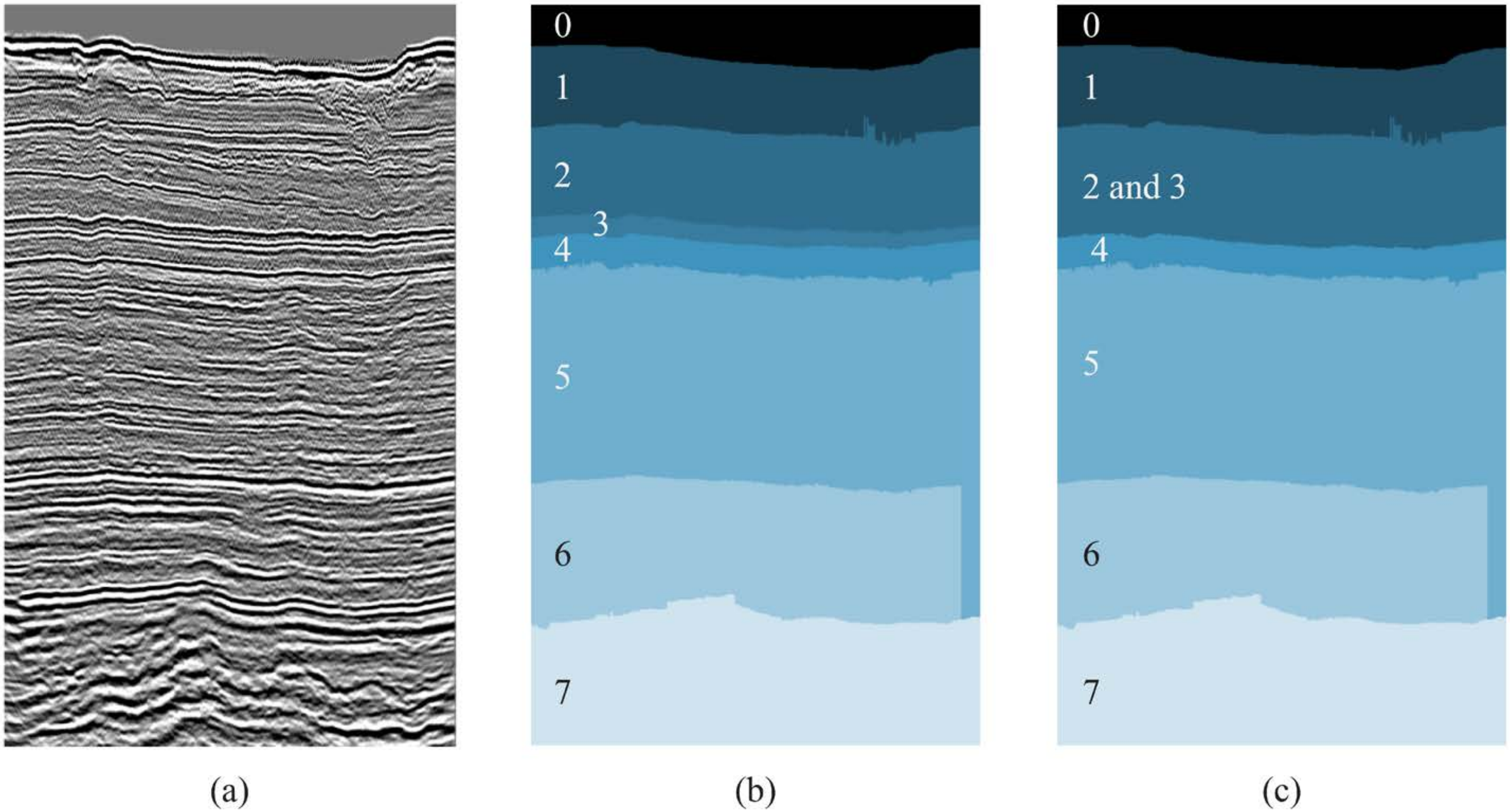

Figure 2. (a) Part of an inline of seismic data, Penobscot (in Scotia Basin) and its respective mask with eight classes (b). In (c), the same mask with classes two and three united.

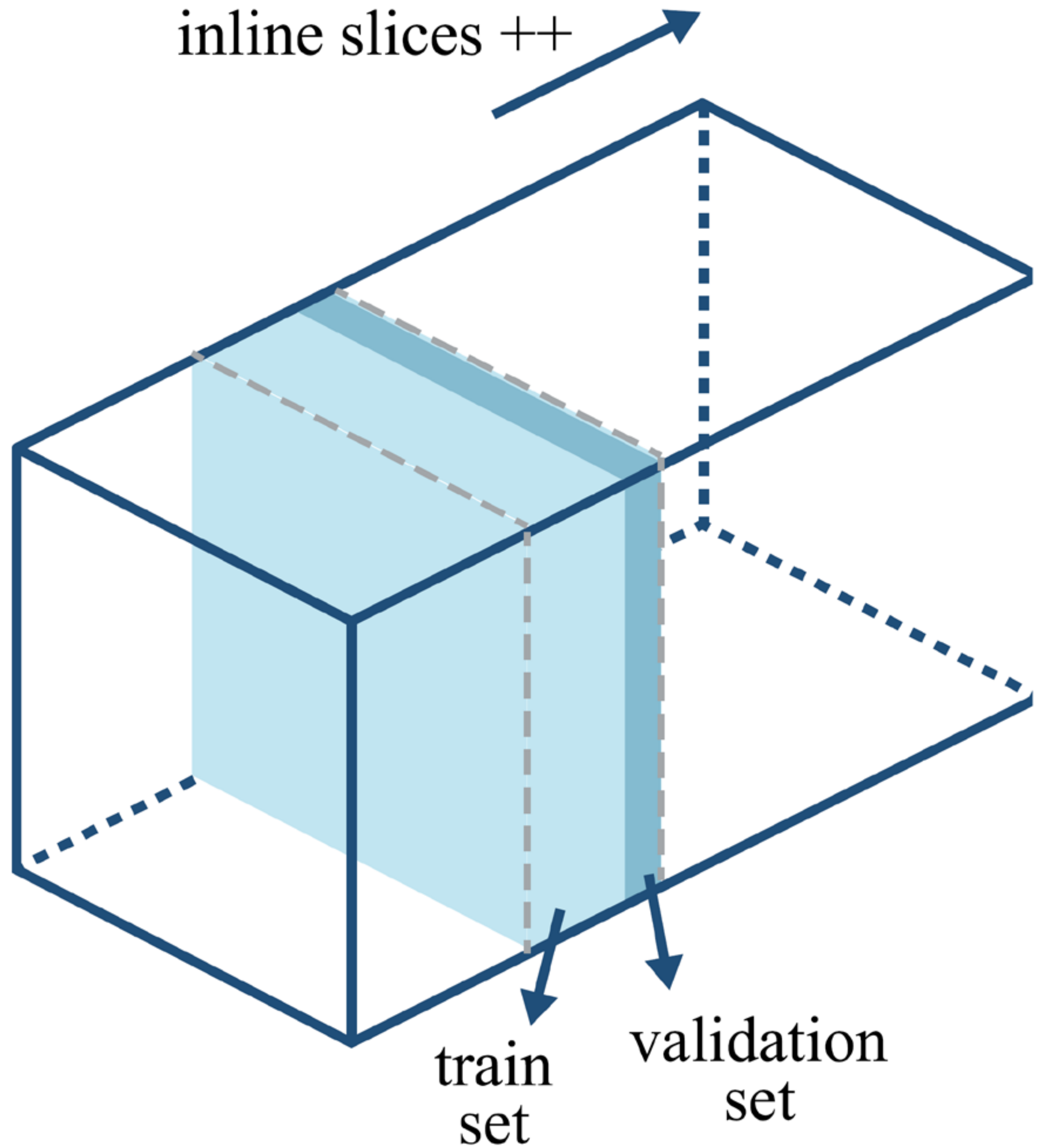

Figure 3. Example of a block of inline slices and its separation in train and validation sets.

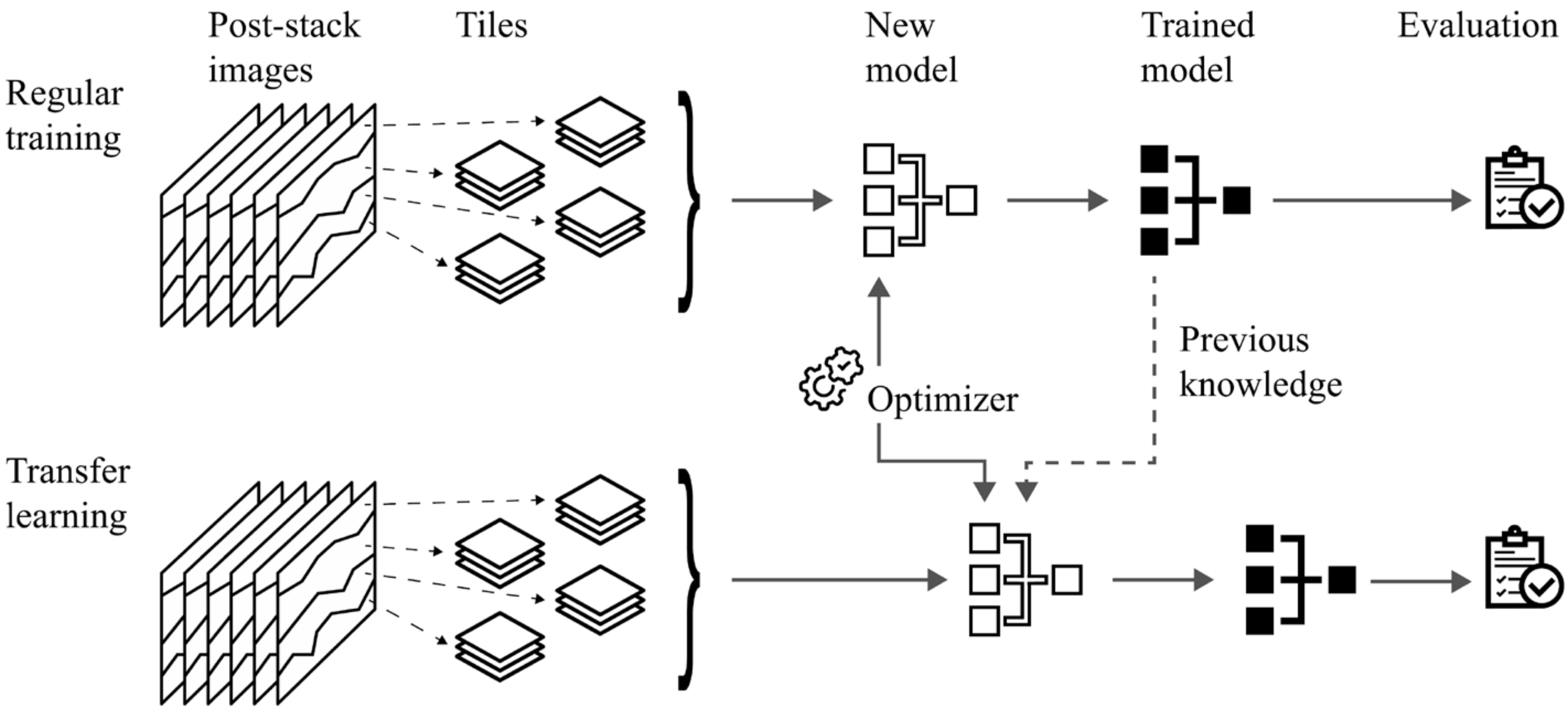

Figure 4. Regular training and Transfer Learning Scheme.

Netherlands	b10-sl414	b10-sl20	b10-sl10	b5-sl5	b3-sl3
Num. of tiles for training	63,456	3,126	1,554	762	486
Penobscot	**b10-sl322**	**b10-sl20**	**b10-sl10**	**b5-sl5**	**b3-sl3**
Num. of tiles for training	18,998	1,204	581	301	189

Table 1. Generated tile datasets for Experiment 1.

	b10-sl20	b10-sl10	b5-sl5	b3-sl3
	Accuracy	Accuracy	Accuracy	Accuracy
control	92.11%	79.19%	77.43%	68.37%
TL	95.74%	93.30%	86.66%	68.70%
Difference	3.94%	17.82%	11.92%	0.48%

Table 2. Results using Penobscot dataset for training and previous knowledge from Netherlands data.

	b10-sl20	**b10-sl10**	**b5-sl5**	**b3-sl3**
	Accuracy	Accuracy	Accuracy	Accuracy
control	93.65%	84.58%	73.23%	68.18%
TL	97.10%	95.88%	91.02%	80.92%
Difference	3.68%	13.36%	24.29%	18.69%

Table 3. Results using Netherlands dataset for training and previous knowledge from Penobscot.

	Pre-salt 1 acc.
control	25.7%
TL from Pre-salt 2	76.7%

Table 4. Results using Pre-salt datasets.