Manual active learning for salt interpretation: an empirical study to avoid forgetting during incremental trainings

Introduction

Salt bodies play a key role in different geoscience applications. They can act as an efficient sealing preventing hydrocarbons to leak from underlying reservoir or be used as caverns for hydrogen, carbon dioxide or methane storage Yet, their complex geometry makes the salt delineation accuracy difficult which increases uncertainties and risks. Deep learning models unleashed significant advances in seismic interpretation, specifically for salt segmentation. However, these models rely on big training sets which is very demanding in terms of labelling effort. In seismic applications, the labelling is a challenging and tedious task due to the broad areas covered by the seismic data and requires expert knowledge. Consequently, finding solutions to limit the labelling effort is a priority to accelerate workflows and to optimize the human resources. The technique of active learning can help in reaching these goals. It consists in selecting the best data to label in order to improve the model performance based on an iterative approach during which, at each step, unlabeled data are chosen to be labelled and used to train the model. This process is repeated until the model reaches acceptable performances.

Objective

In this paper, we focus on defining an active learning approach for the segmentation of salt bodies on seismic data. The aim of this paper is to propose an optimal workflow and set of hyperparameters for which a model can learn incrementally new labels without forgetting the previously learned salt geometries. Ideally, we would like to find a method which enables to (i) incrementally learn without forgetting, (ii) by using a model which can be trained on a mid-range laptop GPU, (iii) with training phases which must be not too long in order to ease the interactivity.

Review of active learning methods

There are different strategies in active learning to choose the data samples to label. They vary based on the level of information the new labels can bring to the model when it is trained on. These strategies can be divided into two categories: a quantitative and a qualitative approach. The first category is based on the quantitative evaluation of the informativeness of the data which allows an automatic selection of the unlabeled data to label. On the other hand, the second category is based on a qualitative evaluation where a human decides which data to label in order to improve the model performance by visually screening the predictions. The first category mainly relies on model uncertainty and has been applied in numerous fields. The aim is to select data samples for which the prediction is the most uncertain. Gal et al. (2017) introduced the Monte Carlo dropout at inference time to build prediction probabilities. Beluch et al. (2018) used an ensemble-based approach to compute the prediction probabilities. For the works mentioned above, the models have been retrained from scratch after each iteration by concatenating the training set from previous iterations with the new labeled data. This implies a longer training time as the number of iterations increases to the expense of the desired interactive workflow. Finally, adopting such a quantitative approach for the seismic segmentation requires the computation of the prediction probability for each tile of the entire seismic volume for each iteration of the active learning process. This might be unpractical in industrial context with big seismic volumes. For the application of seismic segmentation, to the best of our knowledge, only Di et al. (2022) applied a combination of deep learning and active learning with an automatic selection of the data sample to annotate. Their approach does not rely on the evaluation of the prediction probability but on the reconstruction error of a relative geological time model used as input during the training phase.

The second strategy, where a human decides which data to label in order to improve the model performance, has been adopted with success by Tschannen et al. (2020) for the interpretation of horizons on seismic data. Di et al. (2018) applied this qualitative strategy as well for facies classification. One limitation of this approach is the possibility to introduce bias during the process of selecting which data samples to label (Di et al., 2022). Indeed, some patterns might be overlooked by the interpreter during the visual screening of the data. In this paper, we decided to choose this approach to select the data to label as it does not require any additional computation cost.

The catastrophic forgetting issue

The main challenge when incrementally training a neural network is the forgetting of the patterns learned during the previous training iterations. This difficulty had been early highlighted by McCloskey et al. (1989). They named this phenomenon "catastrophic interference". Some attempts had been proposed to overcome this problem. Robins (1995) tackled the forgetting issue by using a rehearsal technique which consists in retraining part of the previously learned labels as new labels are introduced. More recently, Goodfellow et al. (2015) found that dropout can limit the loss of previous knowledge when training a neural network on a new set of data. They stated that with dropout, the model size can be larger giving a protection to forgetting. A complementary explanation might be that the dropout can reduce the co-adaptation of model weights limiting the model to be too specialized to the current training set and giving a better capability to learn new features for the next training iteration. Kirkpatrick et al. (2017) introduced an approach which consists in slowing down the learning for some weights according to their importance related to the previously learned patterns. For the image classification problem, Li & Hoeim (2017) proposed a method inspired from the fine-tuning and the knowledge distillation (Hinton et al., 2015) methods. Firstly, they computed some pseudo-labels, called "soft targets", by evaluating the trained network on the new training data, i.e. the new task to learn. Then they added a new branch to their network which will be specialized for the new task. They performed a first warm-up training of the new branch followed by a second training of the entire network. For both trainings, they used only the new training data. They used a regularizing term in their loss function which aims at penalizing the network when the predictions from the old task layers are too far from the soft targets. This enables the old task accuracy to be preserved to a certain extent. The limitation of this approach is that the network keeps growing as new tasks are learned.

Methodology

To limit the loss of knowledge through the successive training iterations, we followed the rehearsal technique proposed by Robins (1995). We tried to find the best way to combine labels used in the previous trainings with the new labels. To do so we varied, in the training set and the validation set, the proportion of labeled data from the preceding trainings with the new labelled data. In addition to that, we varied the learning rate (LR) as well as the patience of the earlystopping. We expected that a gradual decrease of the LR from one training step to the next one would allow to learn the new labels without dramatically degrading the previously learned labels. Our target approach being interactive, it implies fast trainings, hence we were interested in finding the optimal earlystopping patience value. Decreasing the patience value will reduce the training time to the expense of missing a better set of model parameters to fit the new labels. In addition, lowering the patience value might favor the older over the new labels' performances.

Three trainings have been performed successively. After each training, the model has been evaluated on full seismic data. To evaluate the knowledge loss at the second and third iterations, a collection of monitor lines has been selected. These monitor lines correspond to inlines where the delineation of the salt is acceptable for the previous training iteration. Consequently, the model performance must be preserved for these monitor lines. At the end of the second and third training iterations, the intersection over union (IOU) metric has been computed on their respective monitor set to have a quantitative insight on the level of knowledge loss.

Our experiment has been performed on the Mississippi Valley seismic dataset, offshore northern Gulf of Mexico, made of 1613 inlines, 1213 crosslines and 1500 samples per trace. In total, 3.6% of the data (only inlines) has been labelled trough the three training steps. The number of monitor and interpreted inlines for each iteration is reported on Table 1. In our experiment, we performed a grid search over the parameters mentioned above, the associated values are listed in the Table 1.

	1st training	2nd training	3rd training
Previous / new labels ratio in training set	N/A	0% - 100% ; 12% - 88% ; 34% - 66% ; 50% - 50% ; 66% - 34%	12% - 88% ; 34% - 66% ; 50% - 50% ; 66% - 34%
Previous / new labels ratio in validation set	N/A	0% - 100% ; 12% - 88% ; 34% - 66% ; 50% - 50% ; 66% - 34%	0% - 100% ; 12% - 88% ; 34% - 66% ; 50% - 50% ; 66% - 34%
Learning rate	0.001	0.0001 ; 0.0003 ; 0.0005 ; 0.0007	0.00008 ; 0.0001 ; 0.0003 ; 0.0005
Patience	20	5 ; 10 ; 15 ; 20	5 ; 10 ; 15
Nb. of interpreted inlines	30	11	17
Nb. of monitor inlines	N/A	62	82

***Table 1** Parameters tested at the different training steps.*

Note: For a training step, whatever the previous/new labels ratio, the number of new labels in the training and validation sets is fixed, it is the number of old labels which varies. A 12% - 88% proportion for the 2nd training set means that the new interpreted labels constitute 88% of the training set and the remaining 12% are labels from the 1st training set. For the second training iteration, we chose to start from the model trained with the following set of parameters: {T. set: 34% - 66%; V. set: 12% - 88%; LR: 0.0003; Pat.: 5}.

We used a 5 layers 2D U-Net with relatively small dropout values: 0.1 for shallower layers and 0.2 for deeper layers. The input data are seismic patches of 512x512 in size. We used a binary cross-entropy loss function with the Adam optimizer. The model has been trained on a NVIDIA Quadro P620 GPU.

Results

Our preliminary study has shown that the variation of the proportion of old labels from previous trainings combined with new labels in both the training and validation sets seems to have an influence on the trade-off old/new labels performance. For both iterations, increasing the proportion of old labels in the new training set improves the performance on the previously learned labels (Figure 1c & 1g) while not degrading the performance on the new ones (Figure 1d & 1h). To a lower extent, the performance on the new labels will be degraded when the proportion of old labels in the validation set is increased (Figure 1b & 1f). Meanwhile, the performance on the old labels seems not to be very sensitive to the variation of old/new labels in the validation sets (Figure 1a & 1e).

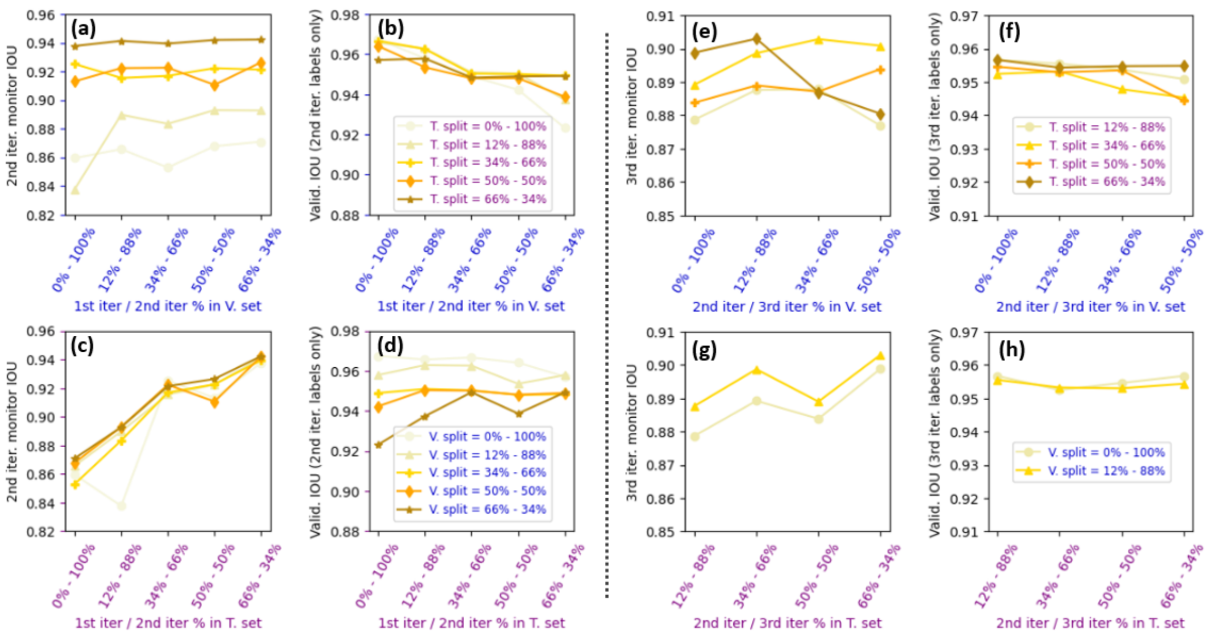

***Figure 1** Impact of old/new labels proportion variation in the valid. set (resp. train set) on the monitor IOU (proxy for old labels performance) for (a) the 2nd and (e) 3rd training iterations (resp. (c), (g)), and on the new labels' validation IOU (resp. train. set) acting as a proxy for new labels performance; for (b) the 2nd and (f) 3rd training iterations (resp. (d), (h)).*

Our preliminary study has shown that reducing the learning rate has a lower impact on the performance on the new labels compared to its influence on the old labels. Finally, a patience value of 5 epochs gave an acceptable trade-off old/new labels performance with fast training times.

For the third training, we selected the model trained with the following sets of parameters {34% - 66%; 12% - 88%; 0.00008; 5}. The model from the first training and the selected models for the second and third trainings has been evaluated on the entire seismic (Figure 2).

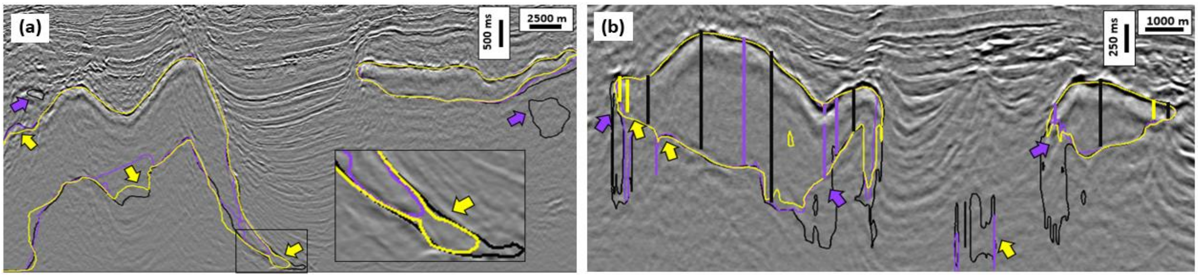

Figure 2 *(a) Unlabeled inline with salt predictions from 1st iteration (black), 2nd iteration (purple), and 3rd iteration (yellow). (b) Crossline with prediction from the three iterations. Bold vertical lines are inlines interpreted during the incremental trainings.*

Conclusion

We showed that the choice of the old/new labels ratio in the training and validations sets, as well as the choice of the LR and the patience can help mitigate the knowledge loss in the case of incremental trainings. This needs to be confirmed on different datasets. Further analysis must be performed, particularly on the impact of the dropout as suggested by Goodfellow et al. (2014). Additionally, it might be interesting to evaluate the IOU on the monitor inlines during the training for an accurate monitoring and building a stop criterion partially based on that metric.

Acknowledgements

The authors would like to thank The National Archive of Marine Seismic Survey (United States of America) for the authorization to publish their data.

References

Beluch, W. H., Genewein, T., Nurnberger, A., Kohler, J. M. [2018] The Power of Ensembles for Active Learning in Image Classification, IEEE/CVF Conference on CV and PR, 9368-9377.

Di, H., Wang, Z., AlRegib, G. [2018] Real-time seismic-image interpretation via deconvolutional neural network, SEG Technical Program Expanded Abstracts, 2051-2055.

Di, H., Truelove, L., Abubakar, A. [2022] Automated active learning for seismic facies classification, SEG Technical Program Expanded Abstracts, 1694-1698.

Gal Y., Islam R., Ghahramani, Z. [2017] Deep bayesian active learning with image data. 34th ICML, Vol. 70, PMLR, 1183– 1192.

Goodfellow, I.J., Mirza, M., Xiao, D., Courville, A., Bengio, Y. [2015] An Empirical Investigation of Catastrophic Forgetting in Gradient-Based Neural Networks, Technical Report.

Hinton, G., Vinyals, O., Dean, J. [2014] Distilling the knowledge in a neural network, NIPS Workshop, 2014

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D. [2017] Overcoming catastrophic forgetting in neural networks, Proceedings of the national academy of sciences, 114, 3521–3526.

Li, Z., Hoiem, D. [2018] Learning without Forgetting, IEEE Transactions on Pattern Analysis and Machine Intelligence, 40 (12), 2935-2947.

McCloskey, M., Cohen., N. J. [1989] Catastrophic interference in connectionist networks: The sequential learning problem, Psychology of learning and motivation, 24, 109–165.

Robins, A.V. [1995] Catastrophic Forgetting, Rehearsal and Pseudorehearsal, Connect. Sci., 7, 123-146.

Tschannen, V., Delescluse, M., Ettrich, N., Keuper, J. [2020] Extracting horizon surfaces from 3D seismic data using deep learning, GEOPHYSICS, 85, N17-N26.