SPE-211654-MS

Uncertainty Analysis for Seismic Salt Interpretation by Convolutional Neural Networks

Fan Jiang, Konstantin Osypov, and Julianna Toms, Halliburton Landmark

This paper was prepared for presentation at the ADIPEC held in Abu Dhabi, UAE, 31 October – 3 November 2022.

Abstract

One of the significant challenges in seismic interpretation is to accurately delineate subsurface features and quantify the uncertainty of the interpretation results due to the non-unique nature of seismic processing and imaging. Salt interpretation usually has limited resolution and relies upon an interpreter's experience with a limited set of geological concepts. In seismic interpretation, especially salt interpretation, researchers have focused on improving the accuracy of pixel predictions by developing various neural network architectures, such as Dense U-Net, Attention U-Net, Residual U-Net, etc. Studying uncertainty quantification of point predictions is important in assessing prediction quality. In this paper, we implemented Monte-Carlo dropout analysis in the variational inference setting with a Bayesian Neural network (BNN) to analyze the aleatoric and epistemic uncertainty of the salt classification. Our approach helps to analyze the posterior distribution from the variational inference and quantitively measure the range of predictive probability distribution.

Introduction

Traditionally, uncertainty is modeled in a probabilistic way by statistics and mathematics. In the context of machine learning, one source of uncertainty occurs when training data and the test data are mismatched. In some cases, the classes overlap due to noise in the data. In the history of deep learning, machine-based image segmentation has beaten human performance in different challenges, e.g., ImageNet Large Scale Visual Recognition Challenge (ILSVRC). The improvement was hit by more advanced architectures. Abdar et al., 2021 review various successful deep learning applications ranging from medical diagnosis to autonomous vehicles, from game-playing to machine translations. However, how to build a deep neural model with the quantification of reliability and robustness is often ignored. Researchers (e.g., Seoh et al., 2020, Jospin et al., 2022) have been striving to comprehend the uncertainty of predictions in deep neural networks.

Bayesian neural networks (BNNs) are stochastic neural networks trained using a Bayesian approach (Lampinen and Vehtari, 2001, Goan and Fookes, 2020). In BNNs, the weights and outputs are treated as the variables, and the goal is to discover the marginal distributions that best fit the data. The distribution will be used to quantify the uncertainty introduced by the data and models to explain the trustworthiness of the prediction. Table 1 compares a standard Point Estimate Neural Network (PENNs) and BNNs.

Table 1—The comparison between PENNs and BNNs for classification.

	PENNs	BNNs
Goal	Optimization/Minimization	Marginalization
Weight	A single set value	Probabilistic distribution
Regularization	Penalization in Loss	Prior beliefs
Training methods	Gradient descent Back-propagation	Markov Chain Monte Carlo (MCMC) Variational Inference (Monte Carlo Dropout) Bayes-by-backprop Normalizing Flows
Estimate	Maximum likelihood estimators	Maximum A Posteriori Predictive distribution

Bayesian methods are mathematically sophisticated and allow rich probability interpretations by obtaining a predictive distribution. Blundell et al. (2015) developed a back-propagation compatible algorithm for learning a probability distribution on the weights of a neural network, called Bayes by Backprop. It regularizes the weights by minimizing a compression cost, known as the variational free energy or the expected lower bound on the marginal likelihood. Gal (2016) showed that a specific optimization of dropout neural networks is equivalent to Bayesian learning via variational inference with a specific variational distribution. He casts the deep learning models as Bayesian models without changing either the models or the optimization. Kwon et al. (2018) proposed a method to quantify uncertainties in classification using BNNs to exploit the relationship between the variance and the mean of a multinomial random variable and avoid estimating extra parameters for the variance.

Kendall and Gal (2017) decomposed the predictive uncertainty into two types of uncertainties, Aleatoric uncertainty, and Epistemic uncertainty, by explicitly modeling the variability of the last layer of neural network outputs. Aleatoric uncertainty captures the inherent randomness of distribution from the natural stochasticity of the data or data uncertainty. In Statistics, it is representative of unknowns that differ each time we train the model, and we could regard it as the confidence level of the prediction. Epistemic uncertainty expresses in the spread of the predicted probabilities of one class or model uncertainty. In deep learning, epistemic uncertainty refers to the uncertainty of the model weights. Every time we train the model, the weights may vary slightly. This variation is called epistemic uncertainty (Figure 1).

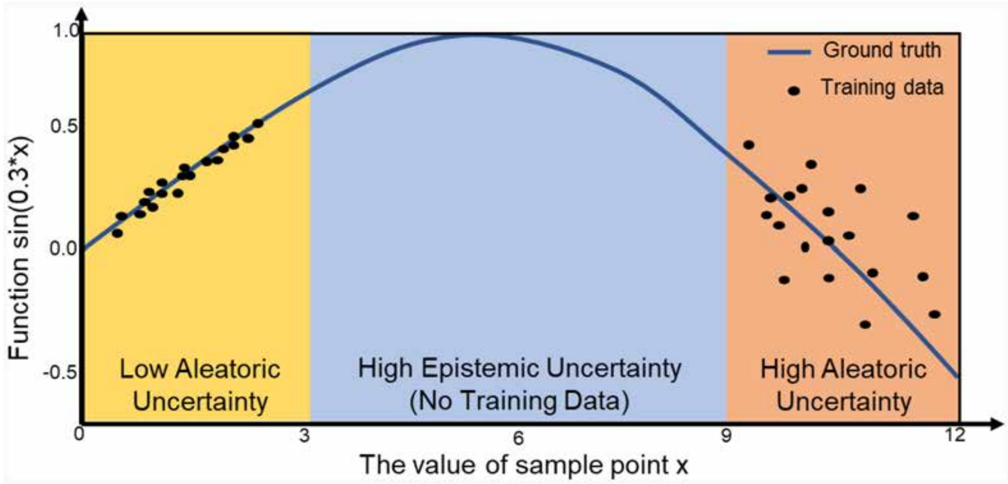

Figure 1—An example of aleatoric and epistemic uncertainty distribution. (After Tuna et al., 2022)

In this paper, we implemented Monte-Carlo dropout variation in our deep learning model to capture the relationship between the variance and the mean of a multinomial random variable. This generates the variability of the predictive probability and captures both uncertainties. The estimated uncertainty could be used to evaluate the prediction result where high uncertainty and low probability exist.

Methods

Point Estimate Neural Network

The base neural network, or point estimate neural network, used in this paper is shown in Figure 2. The architecture of our neural net is based on U-Net (Ronneberger et al., 2015) with additional residual blocks and multiple pooling and dropout layers. The point estimate neural network is used to classify each point, or pixel, in the seismic data to its labeled class, such as falt or non-fault, salt or non-salt. The input is a seismic section where each pixel is assigned an initial weighting coefficient and fed into a neural network. After downsampling convolution, upsampling convolution, and concatenation, the weighting coefficient is updated to optimize by the loss function. The output will be a pixel-wise prediction generated by a softmax layer, which means each pixel represents a probability of "salt" or "non-salt". During training, weighting coefficients are fixed during each epoch, only updating when back-propagating. Point estimate neural network has shown outstanding performance in different geophysical applications but lacks to assess uncertainty for probabilistic interpretations. The bayesian neural network could be an advanced neural network to capture uncertainties in seismic classification problems.

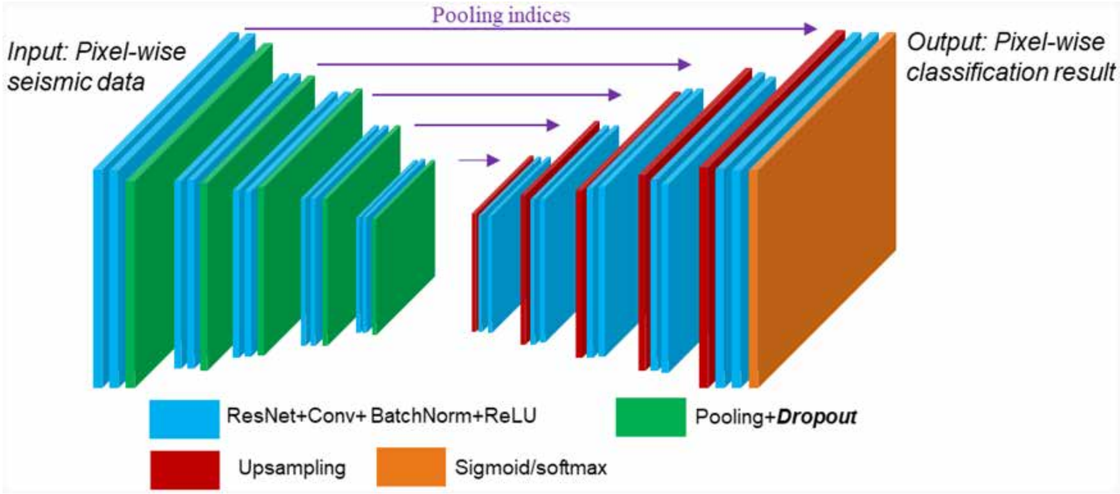

Figure 2—The architecture of Point Estimate Neural Network.

Bayesian Neural Networks

The Bayesian paradigm in Statistics is based on two ideas: The first is that probability is a measure of belief in the occurrence of events, rather than the limit in the frequency of occurrence when the number of samples goes towards infinity. The second idea is that prior beliefs influence posterior beliefs, resulting in posterior distribution:

$$P(\omega|D) = \frac{P(D|\omega)P(\omega)}{P(D)} = \frac{\prod_{i=1}^{N} P(y_i|x_i, \omega)P(\omega)}{P(D)} \tag{1}$$

And predictive distribution:

$$P(y^*|x^*, D) = \int_{\Omega} P(y^*|x^*, \omega)P(\omega|D)d\omega \tag{2}$$

For a new input x^* and a new output y^*. Here $D = \{(x_i, y_i)\}_{i=1}^{N}$ is a realization of independently and identically distributed random variables where x_i are the ith input and y_i are the corresponding output. N denotes the sample size. In the Bayesian neural network model, a prior distribution $P(\omega)$ is placed on a parameter vector $\omega \in \Omega$, which are weights and bias vectors in a neural network (Kwon et al., 2018).

An alternative Bayesian method is a variational inference (Blundell et al., 2015) which approximates the posterior distribution by a tractable variational distribution $q_\theta(\omega)$ indexed by a variational parameter θ. The measure of closeness is commonly used by the Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951). It measures the differences between probability distributions based on Shannon's information theory (Shannon, 1948). The KL-divergence represents the average number of additional bits required to encode a sample from P. For Bayesian inference, it is computed between $q_\theta(\omega)$ and $P(\omega|D)$ as:

$$KL\{q_\theta(\omega)\|P(\omega|D)\} = \int_{\Omega} q_\theta(\omega)\log\frac{q_\theta(\omega)}{P(\omega|D)}d\omega \tag{3}$$

The most popular optimization method is stochastic variational inference (Hoffman et al., 2013), which means the stochastic gradient descent algorithm is applied to variational inference. Variational inference converts standard Bayesian learning from integration to optimization problems, which is suited for batch learning due to MCMC sampling.

Aleatoric uncertainty and Epistemic uncertainty

In exploration geophysics, recent advances in seismic acquisition and imaging algorithms could significantly improve the resolution. However, there is still ambiguity in velocity model building caused by the inherent non-uniqueness of the seismic experiment (Osypov et al., 2013) or misinterpretation due to the complex geological settings (Alcalde et al., 2017). Recent development in computer vision, especially the implementation of deep learning models, can learn powerful representations to map high dimensional data to an array of output. Those deep learning models have been used in many geophysical areas (e.g., Wu et al., 2019, Jiang et al., 2020).

Understanding uncertainty, captured with Bayesian modeling as aleatoric and epistemic uncertainty, is critical, especially in the context of interpreting seismic images. The goal of quantifying uncertainty is primarily divided into regression settings such as depth regression, trace interpolation, and classification settings such as semantic seismic segmentation for interpretation. Those uncertainties are formalized as probability distributions over either model parameters or model outputs. Kendall and Gal (2017) combined aleatoric and epistemic uncertainty in one model by applying Bayesian inferences. At inference, the variational predictive distribution approximating the predictive distribution.

$$q_\theta(y^*|x^*) = \int_\Omega P(y^*|x^*, \omega) q_\theta(\omega) d\omega = \frac{1}{T}\sum_{t=1}^{T} p(y^*|x^*, \hat{w}_t) \tag{4}$$

Where a set of realized vectors $\{\hat{w}_t\}_{t=1}^{T}$ is randomly drawn from the variational distribution $q_\theta\,(\omega)$ with the pre-defined sampling number T. Kendall and Gal (2017) constructed a Bayesian neural network model with the last layer before activation consisting of the mean and variance of logits:

$$\frac{1}{T}\sum_{t=1}^{T} diag(\hat{\sigma}_t^2) + \frac{1}{T}\sum_{t=1}^{T}(\hat{\mu}_t - \overline{\mu})(\hat{\mu}_t - \overline{\mu})^T \tag{5}$$

Where μ and σ^2 represent the mean and the variance, and $\overline{\mu} = \sum_{t=1}^{T}\hat{\mu}_t \big/ T$. The first part of the above equation represents aleatoric uncertainty, and the second part represents epistemic uncertainty. Equation (5) models the variability of the linear predictors, not the predictive probabilities, e.g., salt prediction. Kwon et al. (2018) proposed an estimator for predictive uncertainty by

$$\frac{1}{T}\sum_{t=1}^{T} diag(\hat{p}_t) - \hat{p}_t\hat{p}_t{}^T + \frac{1}{T}\sum_{t=1}^{T}(\hat{p}_t - \overline{p})(\hat{p}_t - \overline{p})^T \tag{6}$$

Where p is the probability output from the network, $\hat{p}_t = p(\hat{w}_t) = \mathrm{Softmax}\{f^{\omega}_{Kendall}(x^*)\}$ where $f^{\omega}_{Kendall}(x^*)$ are the pre-activated linear output of the neural network proposed by Kendall and Gal (2017).

In this paper, we implemented Equation (6) in a deep learning neural network architecture proposed by Jiang et al. (2020) for salt prediction and uncertainty quantification. The task is to classify salt and non-salt regions as a binary classification from seismic data and estimate the aleatoric and epistemic uncertainty. We added a series of dropout layers after convolutional layers with the same kernel in the architecture (Figure 3). Dropout has initially been proposed as a regularization method. It works by applying multiplicative noise to the target layer and is widely used to solve over-fitting problems in deep neural networks. Then we run the prediction process T times to simulate the Monte Carlo dropout (MC-Dropout). MC-Dropout is a variational inference with a distribution defined for each weight matrix. It is straightforward to implement, requires little additional knowledge compared with traditional methods, and can easily transform a model into a BNN. In this paper, we implemented MC-Dropout to analyze the binary classification result and draw the uncertainty maps.

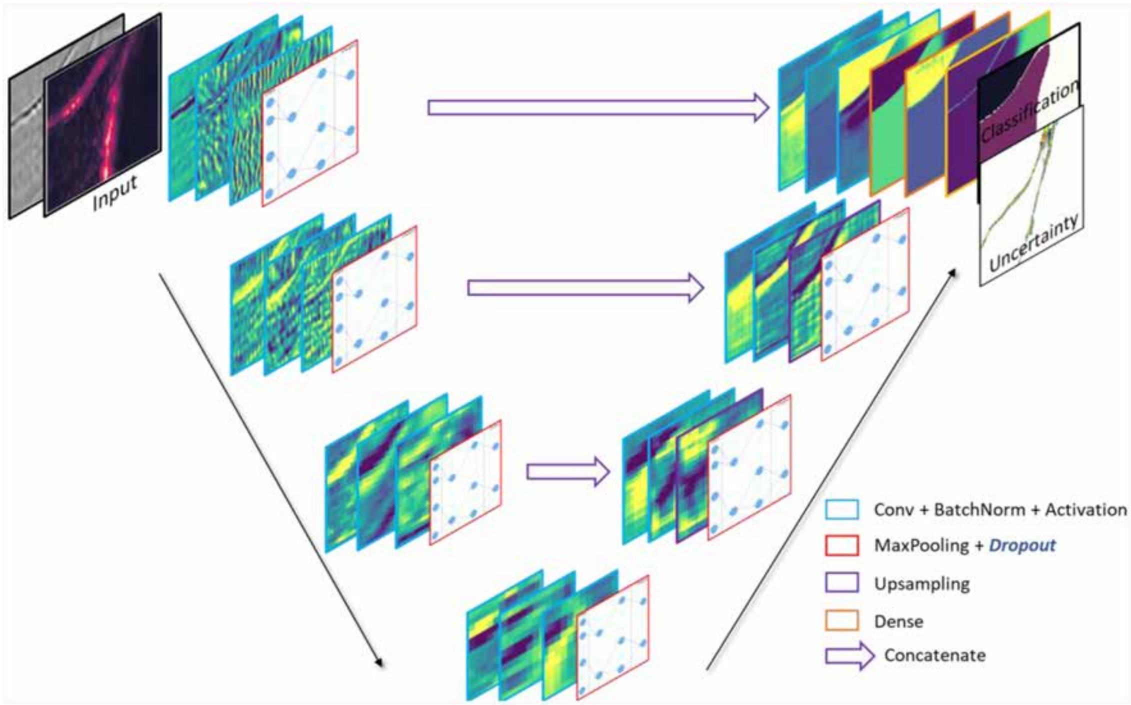

Figure 3—A neural network architecture includes several dropout layers (After Jiang and Norland, 2021).

Examples

We used the synthetic SEG Advanced Modeling (SEAM) dataset as an example to illustrate our method and perform uncertainty analysis. Figure 4(a) shows a seismic section from SEAM data. The goal is to classify the seismic data as salt or non-salt. We randomly picked ~3% data as the training dataset and converted the velocity model to a binary mask. We first trained the neural network with an epoch of 70 and a learning rate of 0.0001. During the prediction process, we literately perform prediction T times (as T = 10 in this case). The prediction from each sampling will be used as input to Equation (6).

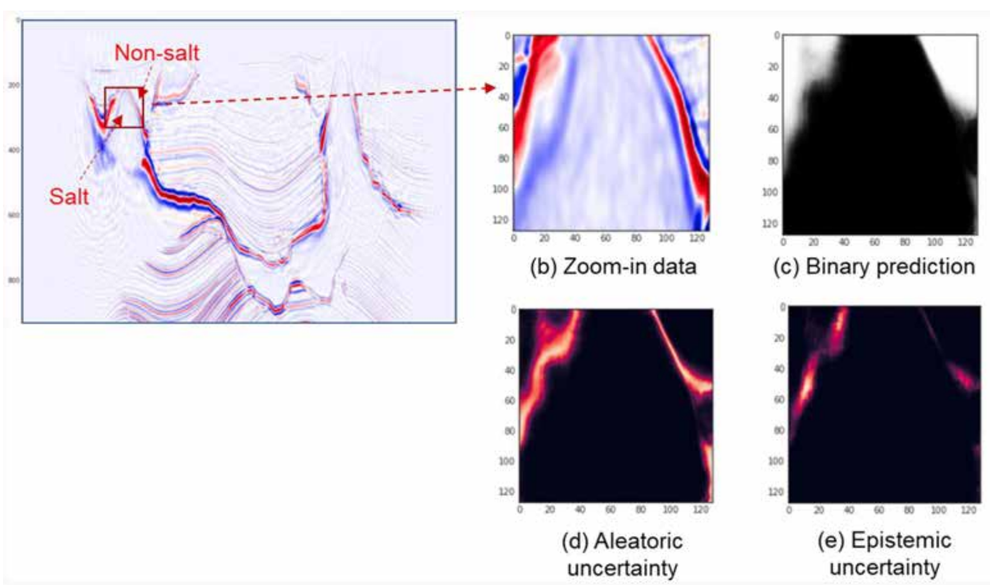

Figure 4—The uncertainty analysis of seismic salt classification by deep learning. (a) Original seismic data; (b) Zoom-in seismic data; (c) Binary prediction, where black color represents the salt body. (d) Aleatoric uncertainty; (e) Epistemic uncertainty.

From Figure 4, we can observe that the primary source of uncertainty comes from the aleatoric part, which indicates our training data may not be sufficient to represent all features in this area, or more training data may be needed. The epistemic part shows relatively low uncertainty, representing that our model parameterization is relatively stable with current training data. The boundaries in the uncertainty maps show more uncertainty than the interior regions. Those uncertainty maps provide extra information in addition to the prediction map. Compared with the ground truth, areas where the prediction map disagrees identify the misclassified region in the uncertainty maps. Those maps also reflect a lack of confidence around the incorrectly identified region. The future work will focus on implementing uncertainty maps to help refine the prediction result and correct where areas are misclassified.

Figure 5 shows the full prediction with its uncertainty quantification maps but applied different colormaps to uncertainty for better visualization on a large scale. In Figure 5(b), the salt body was predicted accurately for both middle salt and salt basement. Figures 5(c) and 5(d) represent aleatoric uncertainty and epistemic uncertainty derived from Equation (6). The total uncertainty for this dataset is the sum of two types of uncertainty. In this process, we adopt the standard classification model to marginalize over intermediate regression uncertainty placed over the logit space and passed through a softmax operation to form a probability vector p. The result shows that aleatoric uncertainty can capture many aspects of salt prediction. For example, Figure 5(c) shows that aleatoric uncertainty is more significant for strong reflective surfaces and patch boundaries. When the prediction map disagrees with the ground truth, the uncertainty maps identify the interior salt regions at the base salt, misclassified by the prediction map. Since aleatoric uncertainty measures how data affects the prediction accuracy, providing more training data to train a deep learning model should improve the prediction accuracy and reduce uncertainty. On the other hand, epistemic uncertainty captures difficulty due to a mismatch of data and model.

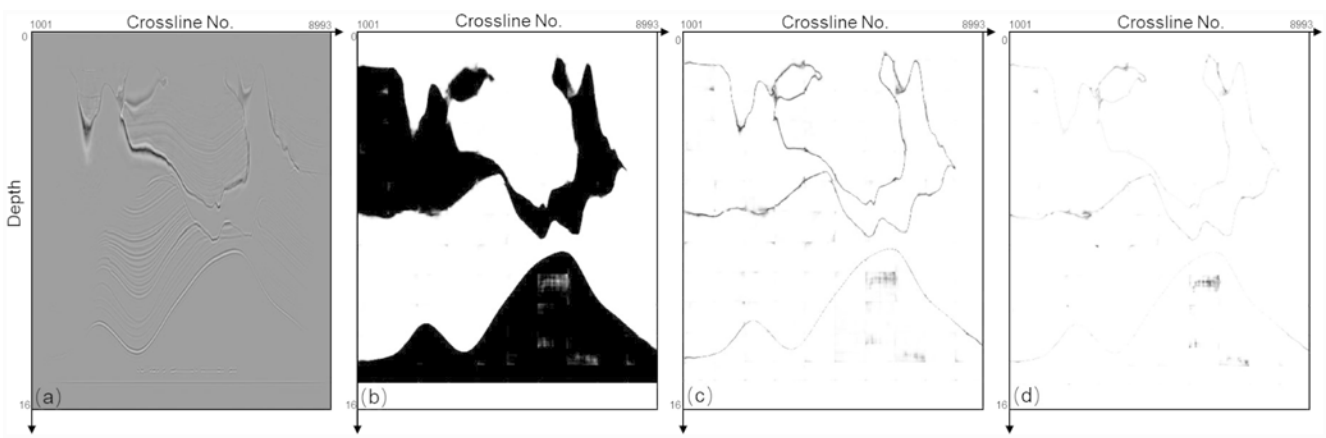

Figure 5—Synthetic test on SEAM dataset. (a) The original seismic data; (b) The binary prediction result. (c) Aleatoric uncertainty; (d) Epistemic uncertainty.

In this case, we consider the total uncertainty map as a proxy to help identify areas where false positives exist. Figure 6(c) is updated by applying the total uncertainty map (Figure 6b) to the initial classification result (Figure 6a). Some false positives in salt and non-salt areas are reduced, which could further improve the accuracy of prediction by deep learning neural network.

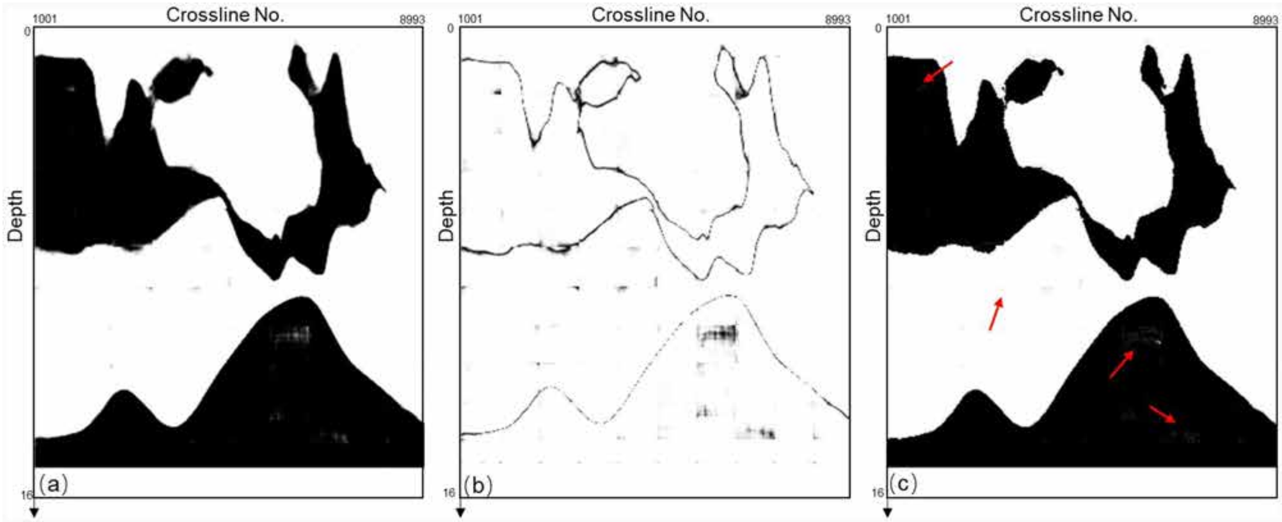

Figure 6—Implementation of uncertainty maps to filter out false positives. (a) The initial classification result; (b) Total uncertainty map; (c) The updated result filtered by total uncertainty map where black color is the salt body regions and white color is non-salt regions.

We assessed the models using MC-dropout with sampling number T = 10 forward passes. Each pass evaluated the model and got a mean and standard deviation for the test error. The aleatoric and epistemic uncertainty is calculated based on the predictive mean and the standard deviation resulting from the multiple repetitions of the experiments to see if the pre-trained model is statistically significant. We observe that the updated classification result by MC-dropout shows significant improvement comparing with the initial classification result.Greater sampling number T will bring more confidence to evaluate uncertainty in this scenario and could be potentially helpful for multi-facies or multi-class classification projects.

A big challenge in seismic interpretation is obtaining labeled data. This could be a laborious and tedious process. Manually labeling data could make the development of an automated seismic interpretation system uneconomical and degrade the effectiveness of machine learning applications. A framework, where a model could learn from small amounts of data and choose by itself where the seismic would like the user to label will be an optimal design in machine learning applications, even for a more comprehensive class of problems, such as seismic processing, imaging, and reservoir simulation. This framework is called deep active learning (Hemmer et al., 2020). It is helpful to train on a small amount of data and an acquisition function, based on the model's uncertainty, to decide what data points are required to have a geologist label. Therefore, a Bayesian-based neural network can work with a small amount of data and possesses uncertain information that can be used with existing acquisition function, a function of inputs that the neural network uses to decide where to query next. A measure of expressivity in characterizes the complexity of functions that can possibly be computed by a parametric function, such as neural network. Deep neural networks are sufficiently expressive for most supervised problems. MC-Dropout method may lack to express complex manifolds that are nearly indistinguishable from real data. The future research will be focused on how to build a generative model to analyze posterior distribution from sampling methods and express complex manifolds in an invertible neural network.

Conclusions

We present an approach to evaluate the uncertainty quantification in a standard neural network and to quantify the machine learning classification in salt prediction by applying a Monte Carlo dropout analysis in the variational inference setting. The advantage of this method in seismic interpretation is that it expresses the inherent variability to understand the distribution of the predictive results and is considered a necessary component for a Bayesian deep neural network. Understanding the underlying principles leading to good models allows us to improve upon them. The performance of dropout in terms of the predictive mean and variance is assessed, implementation of uncertainty maps as a proxy to filter out false positives should be considered as a standard post-processing workflow for seismic interpretation. Those uncertainty maps are a good indicator for an acquisition function for deep active learning and accelerate the automation of the seismic interpretation cycle.

Acknowledgments

The authors would like to thank the SEG for using the SEAM seismic dataset.

References

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U., Makarenkov, V. and Nahavandi, S., 2021, A review of uncertainty quantification in deep learning: techniques, applications and challenges, arXiv:2011.06225

Alcalde, G., Fonseca, A., Boscoa, T., Goncalves, M., Bernardo, G., Pianna, B., Carnavale, B., Gimenes, C., Barrile, S, and Arca, E., 2017, Randomized Controlled Trial, **18**, 317.

Blundell, C., Cornebise, J., Kavukcuoglu, K. and Wierstra, D., 2015, Proceedings of the 32nd International Conference on Machine Learning, *PMLR* **37**:1613–1622.

Gal, Y., 2016, Uncertainty in deep learning, Ph.D. Dissertation, University of Cambridge.

Goan, E. and Fookes, C., 2020, Bayesian neural networks: An introduction and survey, arXiv:2006.12024.

Hemmer, P., Kuhl, N., and Schoffer, J., 2020, DEAL: Deep evidential active learning for image classification, arXiv:2007.11344

Hoffman, M., Blei, D., Wang, C. and Paisley, J., 2013, Stochastic variational inference, *Journal of Machine Learning Research*, **14**, 1303–1347.

Jiang, F., Norlund, P., and Wei, Z., 2020, Analysis of seismic attributes to assist in the classification of salt by multi-channel convolutional neural networks, the First EAGE Digitalization Conference and Exhibition, Extended Abstract.

Jiang, F. and Norlund, P., 2021, Saliency-map guided salt prediction by a multi-channel convolutional neural network, SPE-207244-MS.

Jospin, L., Laga, H., Boussaid, F., Buntine, W. and Bennamoun, M., 2022, Hands-on Bayesian neural networks – A tutorial for deep learning users, arXiv:2007.06823.

Kendall, A. and Gal, Y., 2017, What uncertainties do we need in Bayesian deep learning for computer vision? arXiv:1703.04977.

Kullback, S. and Leibler, R.A., 1951, On information and sufficiency, *Ann. Math. Statist*. **22**, 79–86.

Kwon, Y., Won, J., Kim, B., Paik, M, 2018, Uncertainty quantification using Bayesian neural networks in classification: Application to ischemic stroke lesion segmentation, *Medical Imaging with Deep Learning*, 2018.

Lampinen, J. and Vehtari, A., 2001, Bayesian approach for neural networks – review and case studies, *Neural Network*, **14**, 257–74.

Osypov, K., Yang, Y., Fournier, A., Ivanova, N., Bachrach, R., Yarman, C., You, Y., Nichols, D. and Woodward, M., 2013, Geophysical Prospecting, **61**, 1114–1134.

Ronneberger,O.,, P. Fischer and T. Brox, 2015, U-Net: Convolutional networks for biomedical image segmentation: International conference on medical image computing and computer-assisted intervention, 234–241.

Seoh, R., 2020, Qualitative analysis of Monte Carlo dropout, arXiv:2007.01720.

Shannon, C. E., 1948, A mathematical theory of communication, *Bell System Technical Journal*, **27**, 379–423.

Tuna, O., Catak, F. and Eskil, M., 2022, Exploiting epistemic uncertainty of the deep learning models to generate adversarial samples, *Multimedia Tools and Applications*, 81, 11479–11500.

Wu, X., Liang, L., Shi, Y. and Fomel, S., 2019, FultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation, *Geophysics*, **84**, IM35–IM45.