SPE-205985-MS

Automating Well Log Correlation Workflow Using Soft Attention Convolutional Neural Networks

Aria Abubakar, Schlumberger, USA; Mandar Kulkarni, Former Schlumberger, USA; Anisha Kaul, Schlumberger, USA

This paper was prepared for presentation at the 2021 SPE Annual Technical Conference and Exhibition held in Dubai, UAE, 21 - 23 September 2021.

Abstract

In the process of deriving the reservoir petrophysical properties of a basin, identifying the pay capability of wells by interpreting various geological formations is key. Currently, this process is facilitated and preceded by well log correlation, which involves petrophysicists and geologists examining multiple raw log measurements for the well in question, indicating geological markers of formation changes and correlating them with those of neighboring wells. As it may seem, this activity of picking markers of a well is performed manually and the process of 'examining' may be highly subjective, thus, prone to inconsistencies. In our work, we propose to automate the well correlation workflow by using a Soft- Attention Convolutional Neural Network to predict well markers. The machine learning algorithm is supervised by examples of manual marker picks and their corresponding occurrence in logs such as gamma-ray, resistivity and density. Our experiments have shown that, specifically, the attention mechanism allows the Convolutional Neural Network to look at relevant features or patterns in the log measurements that suggest a change in formation, making the machine learning model highly precise.

Introduction

The journey (see Figure. 1) towards finding the reservoir properties of a basin is aided by understanding the geological formations at a much granular level. In this workflow, various log measurements such as Gamma Ray (GR), Resistivity (RES), Density(DEN), etc. are acquired in a well. All logs are initially checked for presence of washout zones, outliers, missing data and other issues that could be resulting from limitations and errors in acquisition. After processing the data for known imperfections, the next logical step is to perform log correlation to identify geological formations in the field or basin using the logs and align the wells according to the placement of the 'markers' or 'tops' of these formations. It is important to note that not all wells in a given dataset have all log measurements and not all wells have all the markers present because of the geological complexity of that field.

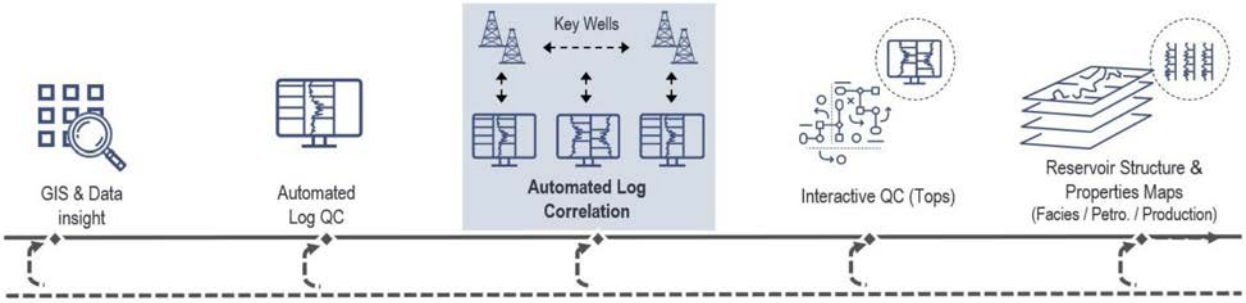

Figure 1—An end to end workflow for deriving reservoir properties from high volume log measurements.

Traditionally, well log correlation is performed by geologists and/or petrophysicists who, in order to pick the markers for a well, study the logs (GR and/or, RES and/or, DEN etc.) that belong to the well and those of the neighboring wells. Essentially, they look for changes in formations that can correspond to the changes in logs. This method of identifying markers is very subjective and can introduce inconsistencies in formation tops throughout the wells of the field. These inconsistencies can propagate throughout the workflow shown in Figure 1, leading to erroneous reservoir properties. Additionally, the effort required to pick the markers for a field with various wells, large number of markers and subtle variations in the logs in response to the changing formations, can be significant.

Therefore, to remedy the complications above, we propose a deep learning algorithm capable of predicting formation tops. This architecture is constructed using soft attention-based Convolutional Neural Network (CNN)(see Figure. 2) with its unique ability to automatically scale down the effect of those portions of logs that contribute less significantly towards predicting a marker. This is important to note especially since the log measurements can be very sensitive to unwanted environment and acquisition-based noise and have nothing to do with a change in formation. This deep learning formulation is supervised in nature, meaning that in order to train the model, we require example of markers 'labels' picked on wells. As we observe the label markers, we notice that the geologist focuses on not only the global trend within the well, but also the local context of the area that indicate changing formation. This information is key, and we employ this concept of local and global 'view' within our machine learning framework as well, as can be seen in Figure 2.

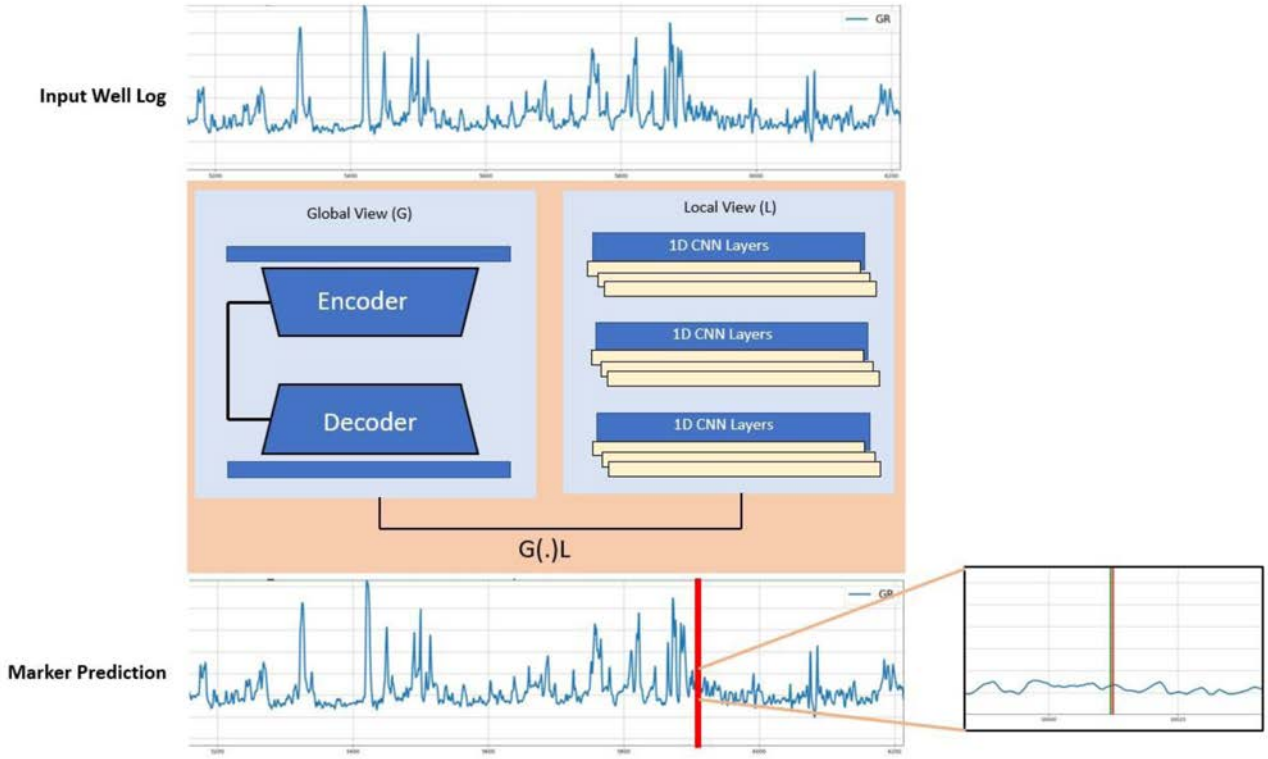

Figure 2—Neural network architecture using soft-attention mechanism.

We remain cautious of the fact that wells belonging to different fields and basins can be geologically different. Additionally, logs acquired at different times, and with different configurations and in different environments can result in variations in the signature of the logs around the markers. So, in order to validate our approach, we apply the machine learning architecture with constant hyperparameters to various basins. This allowed us to reduce the dependence on hyper-parameter tuning for each new basin. The machine learning problem was formulated to keep the manual picks for training purposes to a minimum.

Dataset

For our work, we considered a dataset of wells from the Williston Basin [Gibson (2015)], known as, Bakken. The formations of this basin are in majority present in western North Dakota state of USA. The Williston basin exhibits layer cake geology, meaning that the formations are generally flat with no obvious folding deformations. Additionally, there are some stratigraphic complexities with pinch outs that cause the formations to become thinner and eventually disappear. Hence not all markers will be existed in all wells. This dataset consists of 144 wells with logs such as, Gamma Ray (GR), Resistivity (RES), and Density (DEN). There are 144, 122 and 118 wells with GR; GR & RES; GR & DEN & RES, respectively. Tops of 10 formations have already been picked by geologists. The markers are named as follows: UB000, MB000, LB000, and LB009, which have a very strong corresponding signature, see Figure 3B in the logs indicating their presence and BB000, TF180, TF200, TF300 and TF330, which having only subtle signatures, see Figure 3A.

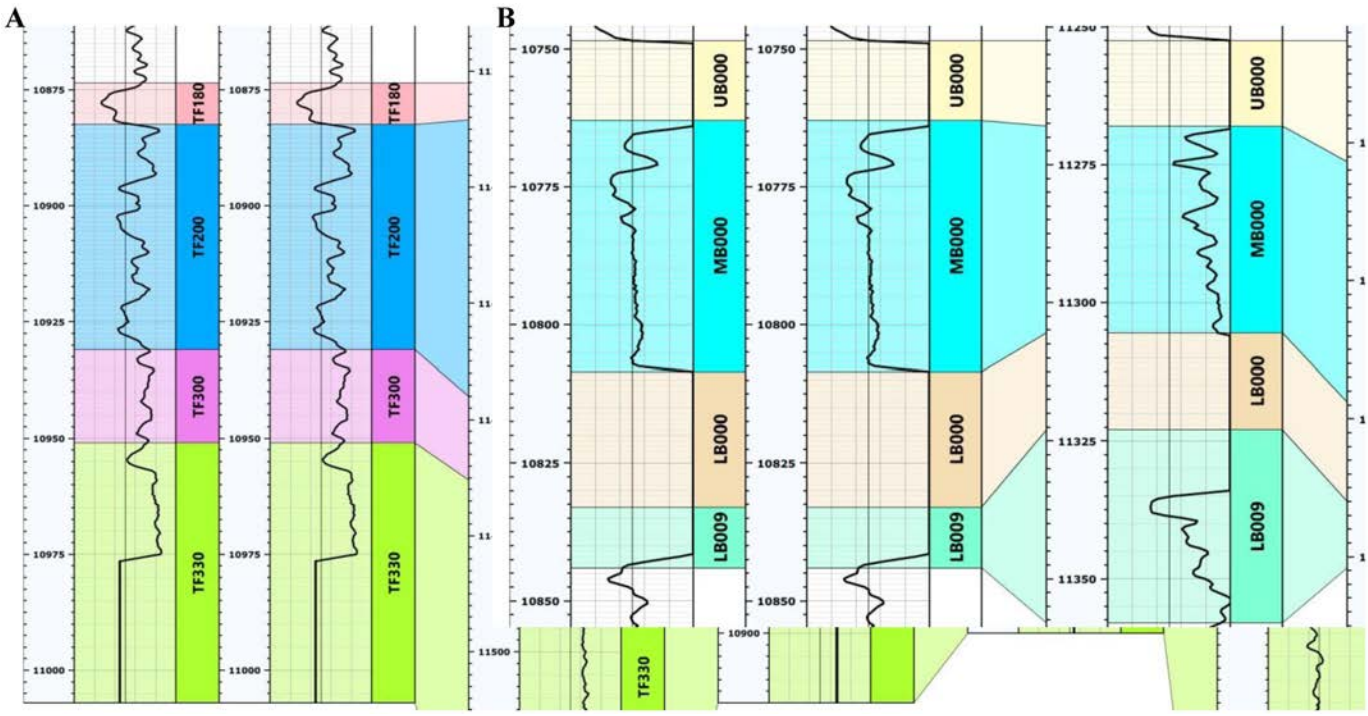

Figure 3—(A) Bakken Gamma Ray logs depicting subtle changes in signature corresponding to TF330, TF300 & TF200 formation tops (B) Bakken Gamma Ray logs with strong changes in the signature in response to UB000, MB000, LB000, LB009 formation tops.

For the purpose of training our machine learning model, we randomly pick training and testing wells from the entire Bakken basin as will be shown in the methodology section. For the purpose of training our machine learning model, we randomly pick training and testing wells from the entire Bakken basin as will be shown in the methodology section.

Workflow

Recently, CNNs with a soft-attention mechanism (Wang et al. (2017), Chen el at. (2016)) have shown considerable improvements in detecting objects and segmenting areas of interest in images. This paper aims to showcase the extension of this application to perform rare event detection in well logs. As can be seen in Figure 2, we propose an architecture that is broken down into three important components: global-view, local-view, and the soft-attention block. The global-view is essentially a U-Net (Ronnenberger et al., 2015) based encode-decoder machine learning architecture with skip connections between the mirroring layers of the encode and the decoder. The skip connections provide a solution for vanishing gradients. Furthermore, the encoding layers consist of one-dimensional (1D) convolution layers followed by a pooling layer, whereas, the decoding layers consist of 1D convolution layers following upsampling (interpolation) layers. Each convolution layer is followed by a normalization and dropout layer, which acts as a form of regularization at the time of training, when the objective is to minimize the loss function. Presence of successive pooling layers in the 'global view' model allows the model to progressively increase the receptive field towards the deeper layers of the model and, thus, capture the long-range patterns in the log. On the other hand, the local-view component of the architecture, consists of several 1D convolutions stacked together. Each of these convolution operations is a local operation, allowing the model to focus on variations in the logs at a local level, especially those areas that bound formations.

The convolutional layers described above, for local and global model, are the inception layers as used by Szegedy et al. (2015). In this setup, the input tensor is convoluted with filters of varying dimensions and each of these resulting feature maps is concatenated and fed to the next layer in the architecture.

During training, input to the model is an entire log in a well and the objective of the machine learning architecture is to detect changes in the log that correspond to the changes in formations. To supervise this learning, for each well, the interpreter provides the depth values at which a marker is present and well as the name of the formation who top this marker represents. This information is converted to a one-hot vector for each marker and this way we train the model independently for each marker. In should be noted that, the architecture was designed to be fully convolutional to allow for training with logs with varying lengths, as was observed in the log measurements (nearly each well has different length).

The resultant tensor at the output is the dot-product of tensor out of the local model and that of the global model. This is an element-wise operation and can be shown as below:

$$A(x_i) = G(x_i)L(x_i) \tag{1}$$

where A denotes the output of the soft-attention block, G denotes the output of the global block and L, that of the local. x_i is the input log from the ith well. Based on the empirical studies that we have done, we noticed a few deviations from the original proposal by Wang et al, (2015) helped improve the predictions for our data. We noted that a *'tanh'* activation, instead of the sigmoid as originally proposed, gave us superior results. In addition to that, the use of dilated filters for each of the inception layers produced better feature maps. Eventually, the scaled tensor $A(x_i)$ is mapped to the output layer through a sigmoid activation. Additionally, in our work, an ablation study was performed that verified that the use of global and local model together provided a much better predictions than either model being used alone.

The model is trained to minimize binary cross-entropy loss with a condition to stop if there is no significant change in the validation loss. As can be seen in Figure 3, for each value of depth in the log there are only a handful markers in the ground-truth mask. This poses a label-imbalance problem as we can frame absence of a marker as class 0 and presence, class 1, thus, majority of the points in the log are class 0 and very few class 1. To remedy this to some extent, we propose to smoothen the labels by convolving a Gaussian kernel on the imbalanced ground truth. An ablation study of using ground truth labels as is, and with smoothening was done to show a clear advantage of the latter. As mentioned in the 'Data' section, the 144 wells from the entire Bakken field are split into train, test and validation sets. The train and validation sets are used at the time of training, whereas test wells can be treated as blind wells which are used to test the quality of the model since the model never saw them.

Evaluation Metric and Uncertainty

To provide a quantity to the quality predictions produced by our Methodology, we calculate an error (fts) metric between the prediction and the ground truth for each marker as follows:

$$e_i = \left| D_i^{\text{Expert}} - D_i^{\text{ML}} \right| \tag{2}$$

where e_i indicates the error, D_i^{Expert} indicates marker pick by the expert, D_i^{ML} indicates the prediction out of the machine learning algorithm for the ith target well. The depth of the well where a marker should be present according to the ML, is calculated by the formulation below:

$$D_i^{\text{ML}} = \underset{j}{\text{argmax}} P_{ij} \tag{3}$$

where P_{ij} indicates the probability prediction over the entire length of the well j of the ith well and $max\ P_{ij}$ is the probability of the marker at that location. To allow for a much more meaningful representation of the predictions, we calculate two additional metrics for each marker spanning all the test wells - Precision and Recall. Precision evaluates, of all of the markers that we predicted in a given well how many of those markers actually had a matching ground truth annotation.

Precision is calculated as below:

$$Pr_{\blacksquare\blacksquare\blacksquare\blacksquare dT} = \frac{\Sigma_i \, \mathbf{I}[e_i <= dl_{\blacksquare\blacksquare}]}{\blacksquare} \tag{4}$$

For each marker we calculate the precision of the predictions for multiple error tolerances dT in {1ft, 2ft, 5ft, 10ft}. The function I is essentially a gated operation, which outputs 1 if the error is within a given tolerance, or 0 if it is not. M is the total number of markers predicted in the well.

Recall evaluates, of all of the markers annotated in the ground truth how many did we capture as positive predictions? Recall is calculated as follows, where N is the total number of test wells and M is the total number of markers predicted in the well.

$$\text{Recall} = \frac{\blacksquare}{\blacksquare} \tag{5}$$

To calculate, a single metric that amalgamates both these quantities, we generate the F1 score for each marker. In this paper, we have maintained dT to 2ft.

$$\text{F1} = \frac{2\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare(\blacksquare\blacksquare\blacksquare\blacksquare_{dT})}{\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare(\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare)} \Big/ \frac{2\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare(\blacksquare\blacksquare\blacksquare\blacksquare_{dT})+}{\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare(\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare)} \tag{6}$$

In order to provide a sense of uncertainty and reliability of the predictions, our work employs Monte Carlo (MC) Dropout (Gal and Ghahramani, 2016) to generate a confidence interval of prediction at each depth in the well. To estimate this uncertainty band for a well, we generate N predictions of the same machine learning model, each time dropping the same number of neural network nodes with a different spatial distribution – we call each model a perturbed version of the original. We then use these N predictions to finally calculate the standard deviation of the distribution of predictions – this serves as the model uncertainty of the predictions.

Results

In this section we demonstrate the quality of results as produced by the proposed methodology on the dataset mentioned previously. Since each well consists of multiple log measurements (GR, RES, DEN) as described in the introduction of this paper, we will show the results for two different schemes:

Experiment 1: Using GR as input logs
Experiment 2: Using multiple logs as multiple input channels

Using GR only

Table 1 and Figure 4 summarize the results of predictions using GR as input log only. In this experiment setup, 25% of the total 144 logs with GR measurements were used as training logs, and 75% for doing blind well testing. Out of the training wells, 20% we kept aside for validation purposes, for tracking the training progress for each epoch, and 80% were used as examples to supervise the machine learning process. The columns 1ft, 2ft, 5ft, 10ft and 20ft show the precision of the predicted markers with those as error tolerances #■■. Each row is an average of the respective marker precisions over all test wells.

wells.

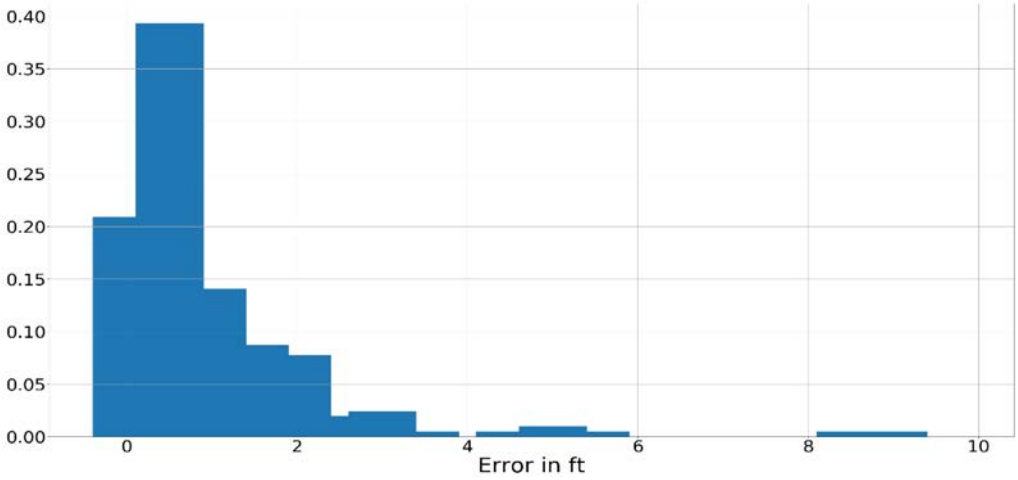

Figure 4—Error Histogram of GR-only results.

Table 1—Results with GR as input log only

Picks	1ft	2ft	5ft	10ft	20ft	Recall
UB000	0.85	0.98	0.99	0.99	0.99	0.95
MB000	0.93	0.95	0.95	0.95	0.95	0.92
LB000	0.88	0.93	0.96	0.97	0.97	0.93
LB009	0.68	0.89	1	1	1	0.92
BB000	0.75	0.91	1	1	1	0.86
TF100	0.25	0.46	0.71	0.94	1	0.68
TF180	0.5	0.64	0.89	0.94	0.94	0.78
TF200	0.82	0.9	0.96	0.96	0.96	0.77
TF300	0.42	0.64	0.89	0.97	0.98	0.82
TF330	0.54	0.8	0.98	1	1	0.91

Additionally, we further observed the performance of our machine learning model in wells that had only very subtle responses in GR to formation changes. The predictions were highly accurate. Examples of predictions on TF330 and TF200 in Bakken are shown in Figure 5 and additional predicted markers tracked through a few wells can be seen in Figure 6.

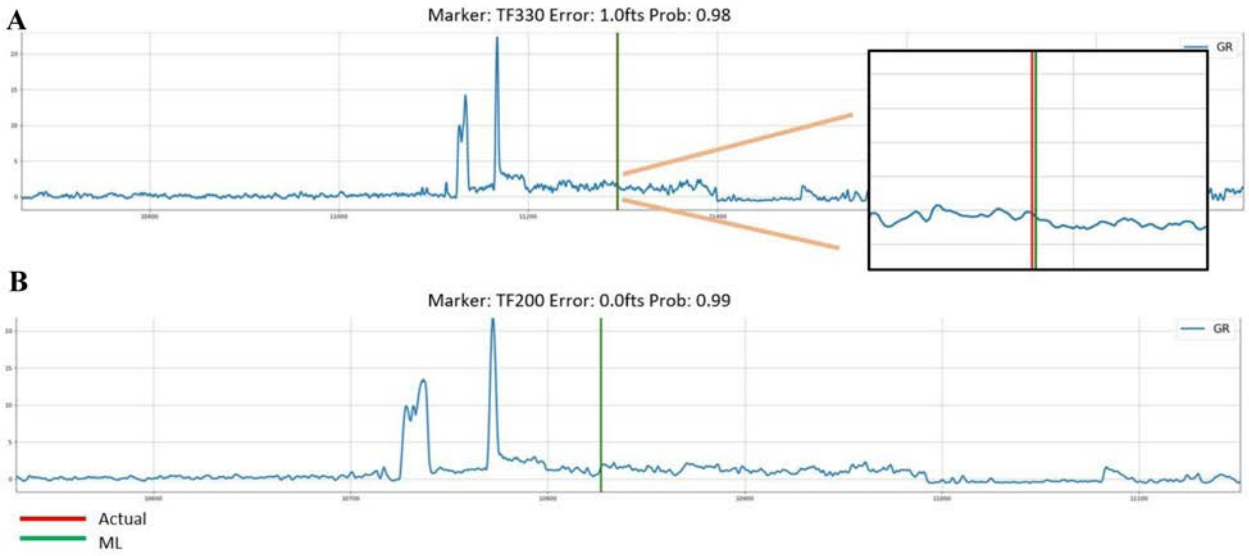

Figure 5—Detection of subtle markers for TF330 and TF200 The prediction in B overlaps that of the marker picked by the interpreter The error is calculated as $e_i = |D_i^{Expert} - D_i^{ML}|$

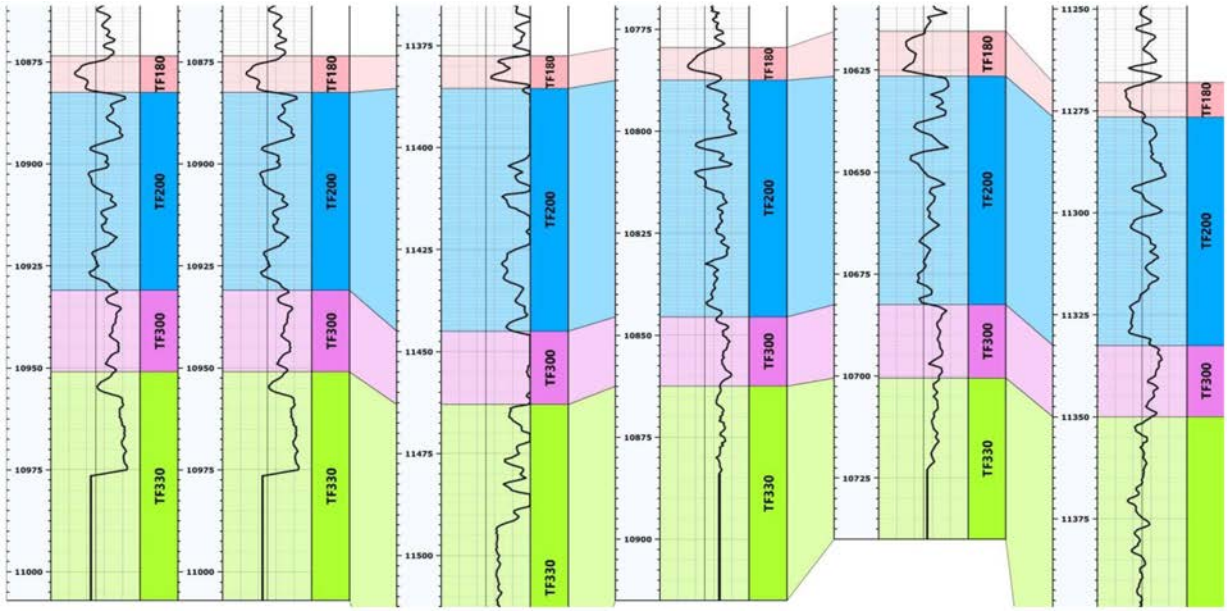

Figure 6—Tracking subtle markers

Using Multiple logs

Similar studies were performed as in the previous section, to understand the uplift of adding additional independent log measurements. The assumption is that adding an uncorrelated measurement in additional to GR would provide information otherwise not captured in the GR log. Therefore, we added RES in one experiment and RES and DEN in a separate experiment. We noted that the performance of majority of the markers improved. A summary of these two experiments is depicted in Table 2 and Table 3. The results improve especially for subtle marker cases.

Table 2—Results with GR-RES logs as input logs. Training Wells - 25%, Test Wells – 75%

Picks	1ft	2ft	5ft	10ft	20ft	Recall
UB000	0.92	0.98	0.99	0.99	0.99	0.95
MB000	0.91	1	1	1	1	0.92
LB000	0.99	1	1	1	1	0.92
LB009	0.8	0.84	0.96	1	1	0.93
BB000	0.85	0.97	0.97	1	1	0.83
TF100	0.13	0.23	0.6	0.85	0.94	0.82
TF180	0.4	0.64	0.94	1	1	0.85
TF200	0.73	0.88	0.97	0.97	0.97	0.75
TF300	0.4	0.67	0.91	1	1	0.81
TF330	0.38	0.75	1	1	1	0.85

Table 3—Results with GR-RES-DEN logs as input logs Training Wells - 25%, Test Wells – 75%

Picks	1ft	2ft	5ft	10ft	20ft	Recall
UB000	0.98	1	1	1	1	0.98
MB000	0.97	1	1	1	1	0.94
LB000	0.99	1	1	1	1	0.95
LB009	0.57	0.75	0.98	0.98	0.98	0.98
BB000	0.87	0.95	0.95	1	1	0.95
TF100	0.27	0.35	0.77	0.88	0.96	0.77
TF180	0.41	0.72	0.98	1	1	0.79
TF200	0.61	0.91	1	1	1	0.87
TF300	0.51	0.76	0.91	0.98	1	0.82
TF330	0.48	0.67	0.97	1	1	0.56

Ablation Study 1 : Global Vs Local View Models

As mentioned in the workflow section, an ablation study was performed that verified that the use of global and local model together provided a much better predictions than either model being used alone. In order to achieve this, we trained the machine learning model independently with the global and local setting and jointly. To quantify the ablation study, F1 scores of the predictions of test wells was compared keeping all other assumptions and parameters the same. Table 4 summarizes this ablation study and proves that the usage of Global and Local models to maximize the context for delineating formation tops outperforms the scenario of using then individually. The experiment was performed on a different dataset, Permian, to ensure that this finding can be generalized.

Table 4—Comparison of F1 scores averaged for all test wells

Dataset	Global (G)	Local (L)	Global + Local
Bakken	0.915	0.929	0.937
Permian	0.913	0.821	0.939

Ablation Study 2 : Label smoothening to remedy label imbalance

Another ablation study was performed to understand the effect of smoothening the marker labels(See Figure 7) on the performance of the proposed algorithm. We see from Table 5 that there is a significant improvement, moreover, in general predicting a marker within a certain feet is acceptable to the interpreters anyway. This indicated that we need to treat the label imbalance issue in order to obtain accurate results.

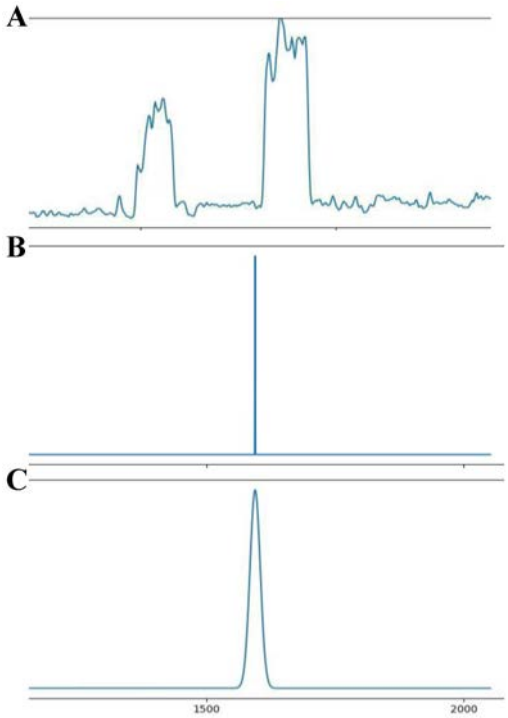

Figure 7—(A) Gamma Ray log (B) Label marker picked by the interpreter (C) Label after gaussian kernel smoothening

Table 5—Comparison of F1 scores for label smoothening

Dataset	With	Without
Bakken	0.549	0.937
Permian	0.652	0.939

Conclusion

In this paper we propose a soft attention based convolutional neural network for log correlation workflow. The machine learning architecture allows for local and global view of the well to make predictions. Additionally, we proved that label smoothening at training time, remedies label imbalance and provides an uplift in the machine learning predictions. The proposed architecture can be trained with single log such as gamma ray or multiple basic logs such as gamma ray, density, resistivity, etc. Metrics such as precision and recall were used to quantify architecture and paradigm choices. In order to have robust architecture, each marker is trained independently. In this way the complexity of the machine learning architecture can remain constant when we are dealing with different number of markers in a dataset and different fields. Numerical results on various unconventional land dataset showed the excellent potential of the proposed approach.

References

1. Chen, L.-C., Yang, J. Wang, W. Xu, and A.L. Yuille, 2016, Attention to Scale: Scale-aware Semantic Image Segmentation: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3640–3649
2. Gal, Y. and Z. Ghahramani, 2016, Dropout as a Bayesian approximation: representing model un certainty in deep learning: international conference on machine learning, 1050–1059
3. Gibson, R.I., 1995, Basement tectonics and hydrocarbon production in the Williston Basin: An interpretive overview: 7th Int'l. Williston Basin Symposium, 1995 *Guidebook*, p. 3–11

4. H. Maniar, S. Ryali, M. S. Kulkarni, A. ABUBAKAR, Machine-learning methods in geoscience, *SEG Technical Program Expanded Abstracts*, 4638–4642, 2018.
5. M. Kulkarni, A. Abubakar, Soft Attention Convolutional Neural Networks for Rare Event Detection in Sequences, *arXiv e-prints, 2011.02338*, Nov 2020.
6. Ronnenberger, 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*, Springer, 234–241
7. Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, 2015, Going deeper with convolutions: Presented at The IEEE Conference on Computer Vision and Pattern Recognition(CVPR)
8. Wang, F., M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, 2017, Residual attention network for image classification: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3156 – 3164