geoconvention
Calgary • Canada • May 12-14 2025

Automated Formation Top Picking with 1D CNN: Application in the Mannville Formation

Lawrence Kanyan, Juan Melendez
McDaniel and Associates

Summary

Over the past several decades, thousands to tens of thousands of wells have penetrated geological formations, particularly in Canadian and US sedimentary basins. Mapping geological surfaces on a basin scale involves identifying formation tops from these wells. Tops in public datasets, when available, are often inconsistent due to variations in interpretation by different geologists over time. As a result, geologists frequently need to re-pick these tops to ensure consistency based on well log signatures, a process that is both time-consuming and resource intensive.

In this study, we present a method that utilizes advanced machine learning techniques to predict formation tops. By training on a subset of labelled wells, our method can predict tops in unlabeled wells using 1D Convolutional Neural Networks (CNNs), a one-dimensional adaptation of the 2D CNNs commonly used in image recognition tasks. Our approach builds on previous methodologies (Brazell et al., 2019, Mainar et al., 2019, and Godwin et al., 2018) but introduces novel elements in the use of model ensembles, data processing techniques before and after model training, and assessing model performance. Furthermore, our method provides a confidence score for each prediction, facilitating quick quality control of the results. This approach has been replicated and internally tested across various geological plays, incorporating a range of data augmentation techniques both before and after predictions.

We present the results from the application of the above workflow to predict the Top of Mannville across a large region within the Western Canadian Sedimentary Basin, Alberta, Canada. We were able to consistently pick the Mannville top in ~20,000 wells from a training dataset of ~3,000 wells.

Workflow

Several authors have previously attempted algorithmic methods, such as cross correlation and dynamic depth warping, to project formation tops from nearby interpreted wells to uninterpreted wells. However, these methods face limitations, including intensive feature engineering, poor generalization, sensitivity to variations in log data, and a reliance on the accuracy of the seed wells, which may propagate errors across predictions. In contrast, supervised deep learning methodologies, like those utilizing convolutional neural networks (CNNs), offer several advantages for predicting formation tops. These include the ability to learn complex patterns directly from well log data without extensive feature engineering, adaptability to varying geological conditions, and improved generalization to unseen wells. Additionally, deep learning models can be enhanced with ensemble techniques and data augmentation to further increase prediction accuracy and provide confidence estimates, enabling more robust and consistent formation top identification across large datasets. One disadvantage of CNNs is that they do not directly output a confidence metric. We solved this problem by using an ensemble of models, which in addition to providing a confidence metric, also improves the overall model accuracy.

Our automated formation tops picking workflow utilizes a supervised learning methodology, with the key steps outlined below:

Preprocessing:

1. We create a surface grid based on manual formation top picks from seed wells (Figure 1). This grid serves two purposes: (a) it provides an initial depth estimate for the expected pick, and (b) it acts as a baseline to evaluate model performance in the test set.
2. To expand the training dataset, we generate random slices of well logs around the estimated picks from step 1 and apply random zooming. This slicing helps speed up model training and ensures the model does not always predict tops at the midpoint of a slice. Random zooming prevents the model from failing due to stretch/squeeze caused by thickness variations or well deviations if the logs are in Measured Depth (MD) domain.
3. We split the labeled data into training and test sets for model evaluation.

Model Training:

1. The processed logs are fed into a 1D CNN deep learning model, with multiple logs treated as different input channels, similar to Red, Green and Blue (RGB) channels in image classification tasks.
2. We train an ensemble of CNN models by varying hyperparameters such as learning rate and batch size.
3. For the test set, we produce multiple predictions for a top by combining the ensemble models with data augmentations, sufficient to calculate robust statistics such as the mean and standard deviation.

Final Predictions on Unlabeled Data:

1. We augment the prediction data set by using the same methods as the training and test sets, generating multiple predictions for each data point.
2. These predictions are used to compute the mean and standard deviation. The mean serves as the final predicted top, while the standard deviation acts as a confidence metric, with higher values indicating lower confidence.
3. Continual manual cross-checking of auto picks with well data to confirm correct placement. Low-confidence picks and those significantly deviating from the grid pick are specifically flagged for manual review or exclusion.

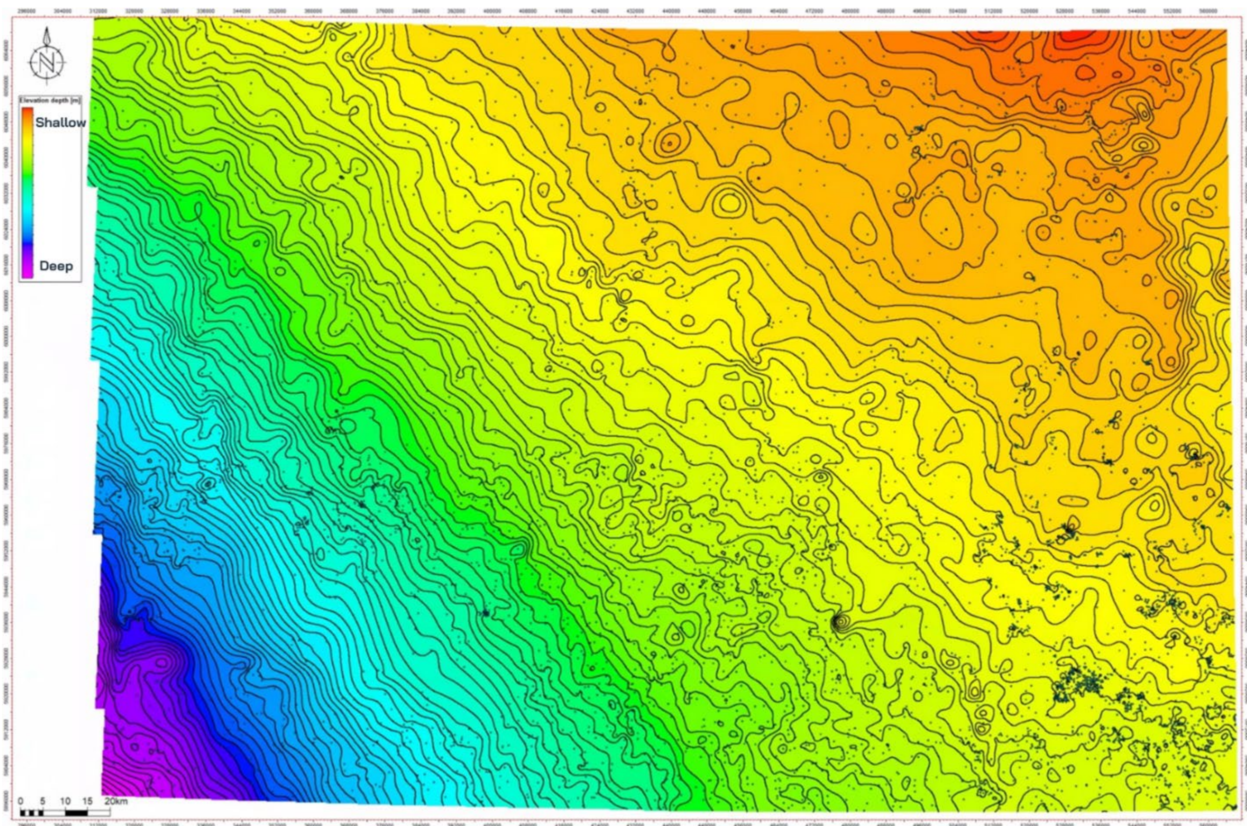

Figure 1: Section of Top Mannville structure constructed using ~3,000 manual picks.

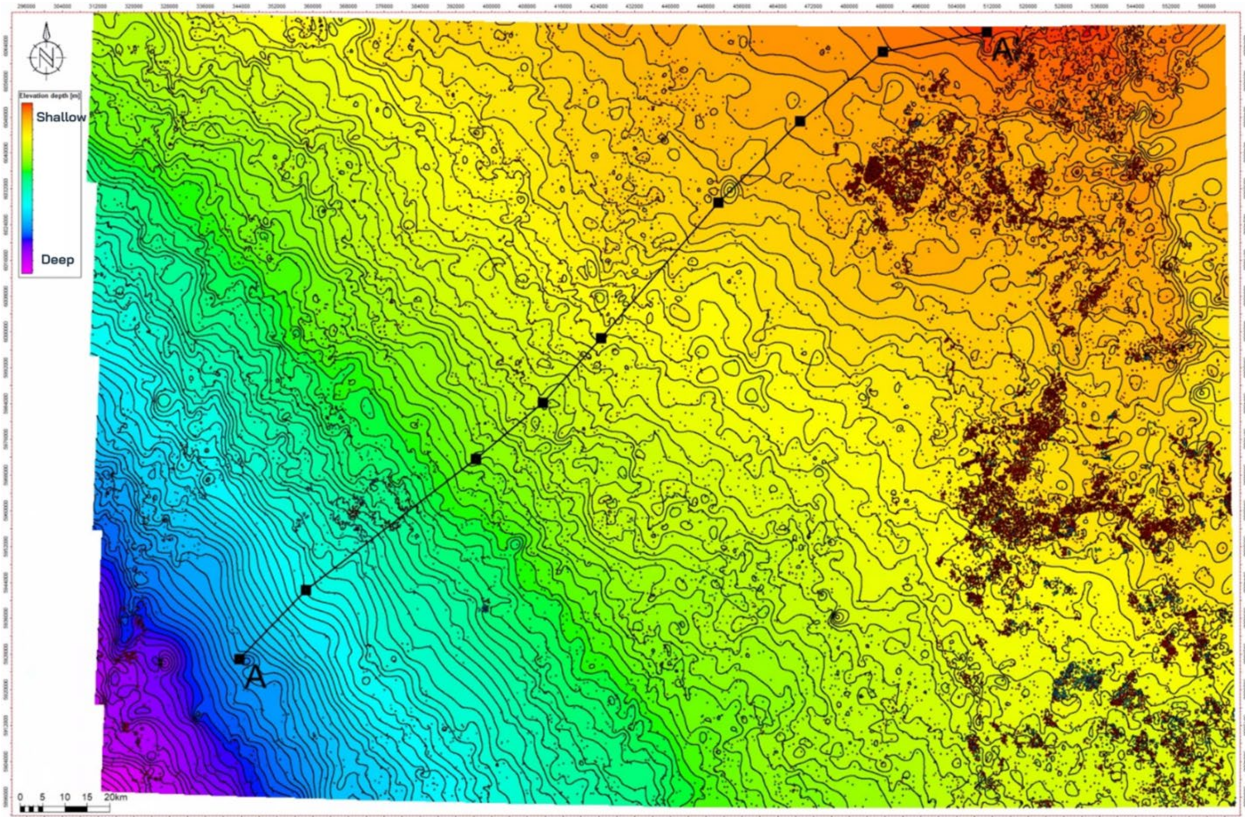

Figure 2: Same section of Top Mannville structure constructed using ~3,000 manual picks and ~20,000 auto picks. Reference cross-section also shown. The greater data density compared to Figure-1 allows structure mapping in greater detail.

Results, Observations, Conclusions

The gridded map (Figures 2 and 3) could predict 90% of the formation tops within ±15 meters in the test set. Using individual models, we could predict 90% of formation tops with an accuracy of ±6 meters. Using the model ensemble and data augmentation, we could predict 90% of formation tops with an accuracy of ±2 meters. We conducted a manual quality check of the pick on a random selection of wells to verify the accurate placement of the model pick.

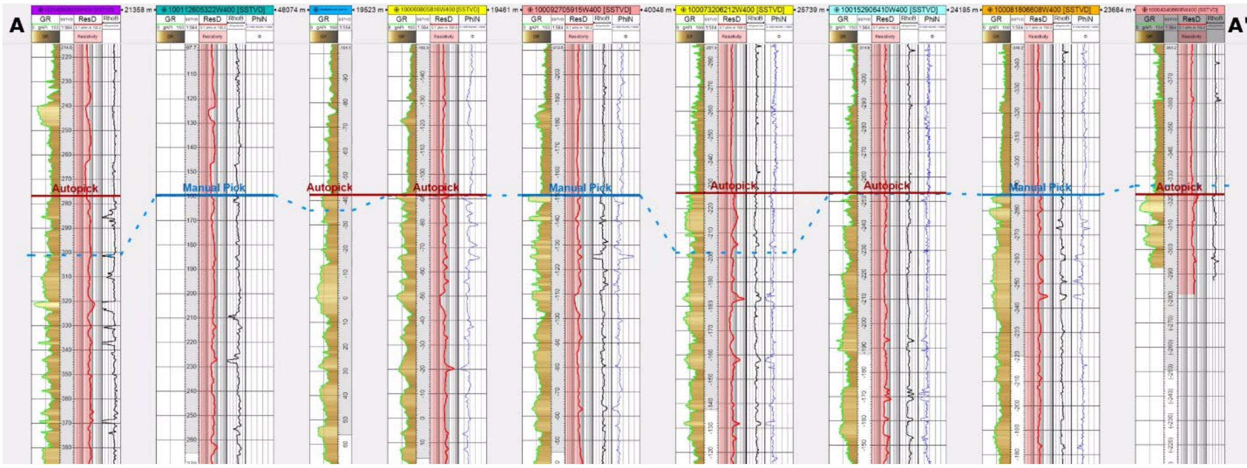

Figure 3: Reference cross section (see figure 2) with manual picks (blue) and auto picks (red). Auto picks demonstrate much better results than using only a grid made from a limited number of picks (dashed blue line, same grid as figure 1).

A significant advantage is that we could flag predictions with confidence levels. We found that the most inaccurate predictions are also low confidence. Upon deeper inspection, these were found to have issues with well logs, such as missing data over the intervals where formation tops were expected. Notably, we did not extensively clean the input data to minimize manual work, and it would be ideal if the workflow could automatically handle this through confidence scores. The low confidence picks were not used to create the final grid.

Our automated workflow significantly reduced the time required to complete tasks, bringing it down from several weeks or months to just a couple of days.

References

Godwin, J., J. Roberts, and E. Panetti, 2019, Constrained automatic tops picking using convolutional neural networks: SEG International Exposition and Annual Meeting, San Antonio, Texas, USA, https://doi.org/10.1190/segam2019-3215174.1.

Maniar, H., S. Ryali, M. S. Kulkarni, and A. Abubakar, 2018, Machine-learning methods in geoscience: SEG International Exposition and Annual Meeting, Anaheim, California, USA, https://doi.org/10.1190/segam2018-2997218.1.

Brazell, S., A. Bayeh, I. Tobar, and D. Burton, 2019, A Machine Learning-based approach to assistive well log correlation: Houston Geological Society, Applied Geoscience Conference, https://doi.org/10.30632/PJV60N4-2019a1