GC Accelerate Well Correlation with Deep Learning*

Bo Zhang[1], Yuming Liu[2], Xinmao Zhou[3], and Zhaohui Xu[2]

Search and Discovery Article #42429 (2019)
Posted August 19, 2019

*Adapted from the Geophysical Corner column, prepared by the authors, in AAPG Explorer, August, 2019. Editor of Geophysical Corner is Satinder Chopra (schopra@arcis.com). Managing Editor of AAPG Explorer is Brian Ervin. AAPG © 2019. DOI:10.1306/42429Zhang2019

[1]University of Alabama, Tuscaloosa, Alabama, USA (bzhang33@ua.edu)
[2]China University of Petroleum-Beijing, Beijing, China
[3]Research Institute of Petroleum Exploration and Development (RIPED), CNPC, Beijing, China

General Statement

With the relatively low price of oil and gas, the identification of bypassed pay in mature oil fields has become increasingly important. Mature – often called “brown” – fields may be characterized by thousands of wells acquired at different times by different operators. The high density of wells makes this type of field a candidate for analysis using techniques originally developed for 3-D seismic interpretation, such as the development of well-log attributes. There are two kinds of attributes used in seismic interpretation: (1) explicitly generated attributes, such as coherence and impedance volumes used in interactive interpretation, and (2) implicitly generated attributes computed internally by machine learning algorithms.

Conventional well log analysis is based on interactive, interpreter-based pattern recognition. A skilled interpreter identifies similar patterns (such as upward fining and coarsening) in neighboring wells and links them using a conscious or subconscious stratigraphic sequence model. Tying dozens of wells is time consuming. Tying thousands of wells is both time consuming and error-prone. In this case study from northeast China, there are approximately 7,000 wells acquired over a period of 40 years. Careful, interactive interpretation has provided 100 intersecting cross sections tying a total of 1,786 wells. Investing precious human interpreter time to analyze the remaining five thousand wells is not cost-effective. We therefore propose the following workflow:

- First, interactively interpret key cross sections using a sequence stratigraphic framework as we have done in the past.

- Next, use these logs and picked well tops as training data for convolutional neural network-aided interpretation of the remaining well tops.

- Finally, generate thickness and porosity-thickness maps using all the wells and quality control the results.

Geological Background

The research oil field located within the Songliao Basin of northeast China has produced oil for more than 45 years. The target is a Middle Cretaceous lacustrine delta. The research oil field covers an area of 107 square kilometers encompassing 6,992 wells. The field engineers divided the reservoir formation into four orders of hierarchal units to better manage the field development (Figure 1). There are one, four, 10 and 31 members for the first-, second-, third-, and fourth-order units, respectively.

Well Correlation Using the CNN SegNet Algorithm

Human interpreters define sequence boundaries and lithologic units by their well log patterns. A machine learning algorithm like CNN is trained by a human interpreter to recognize these patterns and subsequently “segment” the well log data into units it has been trained to identify. In machine learning, the interpreted sequences are called “labels.” Some of the human-interpreter labels are used to “train” the algorithm, others to “test” them and still others to “validate” the prediction. Once validated as being sufficiently accurate, the trained CNN is applied to all the wells. CNN generation of the predicted facies is fast and quite accurate; however, as with interpretation by multiple geoscientists, the results need to be quality controlled, and if necessary, modified. The first, second, and third panels in Figure 2 show spontaneous potential (SP), micro gradient (RMG), and micro potential (RMN) logs, respectively. The fourth panel in Figure 2 shows a 1-D image generated using SP, RMG and RMN logs. The fifth panel in Figure 2 shows the interpreted second order sequence by human interpreters. We chose 463 wells out of human-correlated 1,786 wells to construct our CNN.

Training, Testing, Validation and Accuracy

We randomly select 65 percent, 40 percent, 20 percent and 10 percent subsets of the 463 wells to form the training data set to determine how the accuracy of our prediction depends on the size of labeled input data. Thus, the first (65 percent) prediction used 300 wells for training and 163 wells for testing, the second (40 percent) prediction used 185 wells for training and 163 wells for testing. The third (20 percent) prediction used 93 wells for training and 370 wells for testing. Finally, the fourth (10 percent) prediction used only 46 wells for training process and 417 wells for testing. The “prediction accuracy” is defined as:

$$\text{Accuracy}(\%) = \frac{\text{Samples with correct prediction}}{\text{Total samples of the well logs}} \quad (1)$$

Figure 3 shows the prediction accuracy for the second-, third-, and fourth-order geological units, showing that we obtain a high accuracy for all the second and third order units.

In contrast, we obtain a significantly lower accuracy for some of the fourth order units. We attribute some of this lower fourth-order accuracy to greater lateral lithologic heterogeneity. Figure 4 shows the predicted results for well W438 at different orders for the four training data experiments. The colored panel marked “V” shows the interactively interpreted units provided by the oil company data owner. The colored panels marked 1, 2, 3, and 4 show the results when using 65 percent, 40 percent, 20 percent, and 10 percent of the wells in training.

The end product is a suite of highly detailed maps (Figure 5). Such maps can then be integrated with 3-D seismic data for a more complete geostatistical inversion.

Conclusions

Convolutional neural networks have been used to identify objects for self-driving cars as well as faults and salt domes on seismic data. Here we show that CNN holds great promise in interpreting the large amount of well log data in mature oil fields. Current CNN predictions are not perfect and need to be quality controlled. However, in this data volume, only 25 percent of the available data justified reinterpretation using traditional interactive workflows, such that 75 percent of the data are currently ignored. CNN provides a means to incorporate these additional data into a more complete analysis in a timely manner. Reinterpretation of mature, data-rich oil fields may require a year of interpreter time. Rather than using a team of geoscientists to accelerate this process and introducing inconsistencies, CNN may provide a means to provide an interpretation consistent with the patterns identified by the lead interpreter.

Figure 1. Three representative well logs showing the first-, second-, third-, and fourth-order sequences.

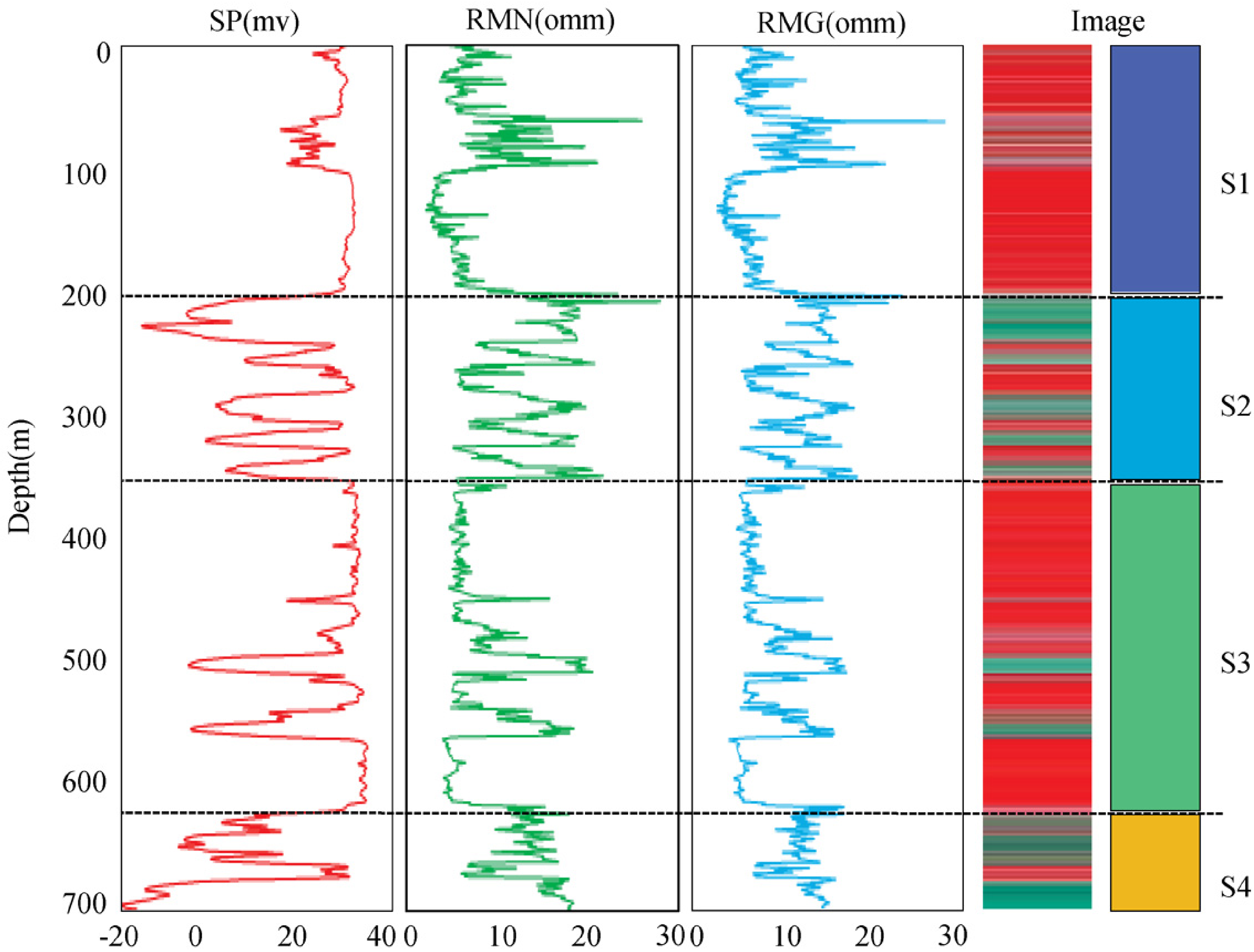

Figure 2. The 1-D image converted using SP, RMG, and RMN logs, and the coarser second-order sequence labels are used as input for the well correlation using deep learning.

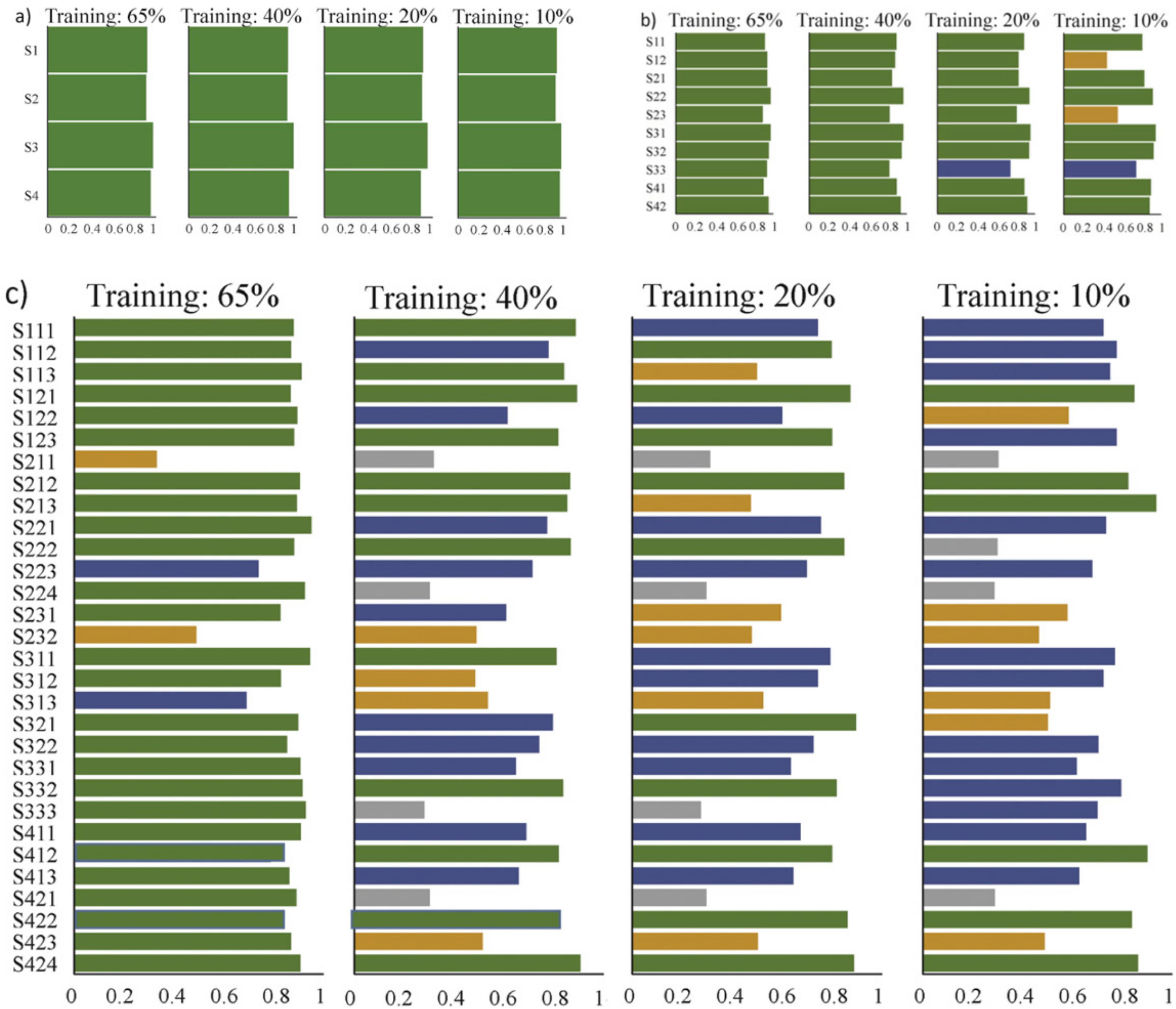

Figure 3. Chart showing the accuracy in predicting (a) second, (b) third, and (c) fourth order units using 65 percent, 40 percent, 20 percent and 10 percent of 417 interactive interpreter generated labels as training data.

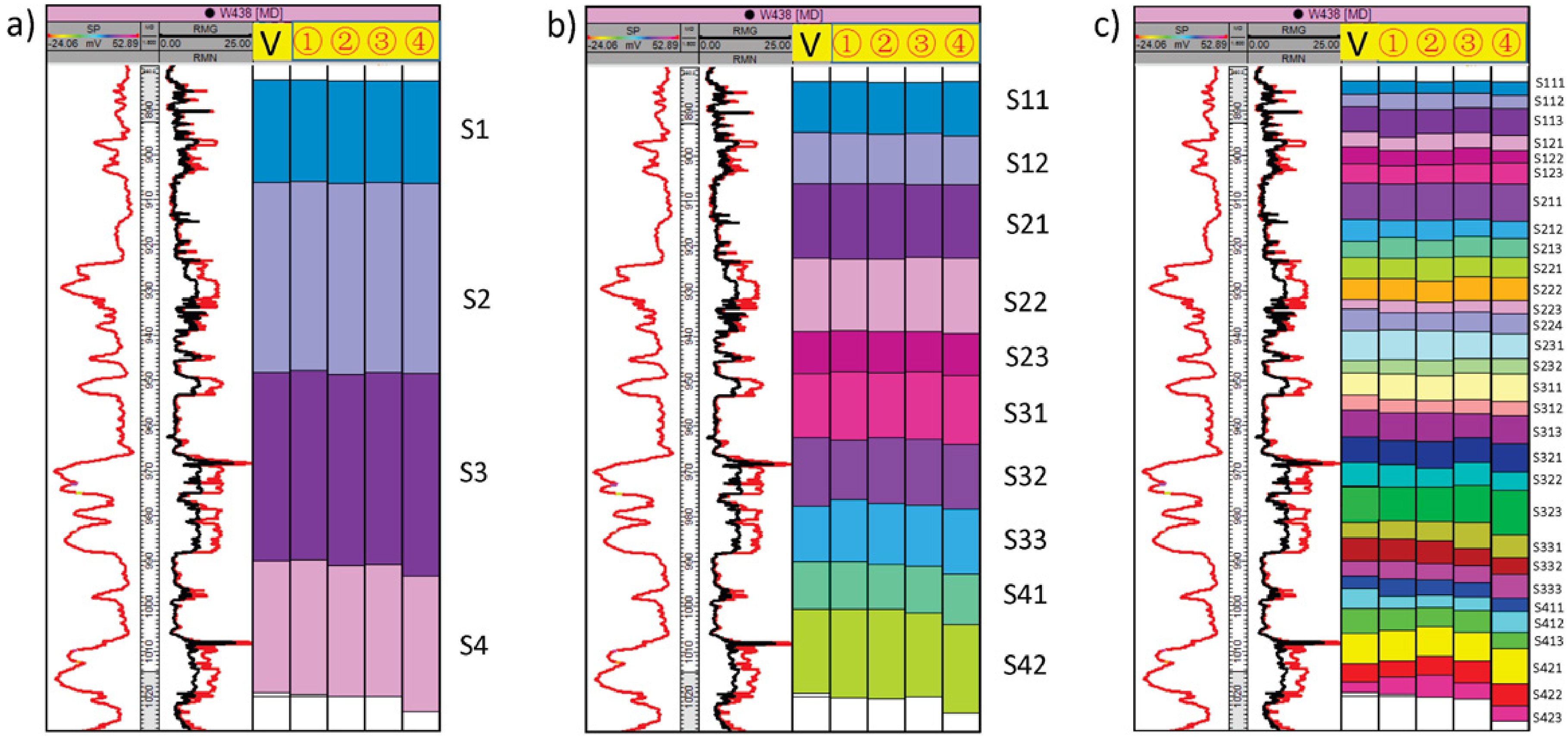

Figure 4. The predicted results for a representative well showing the SP, RMN and RMG data used as input for (a) second-order, (b) third-order, and (c) fourth-order sequences. Colored column "V" three shows the sequence boundaries constructed through conventional interactive interpretation, providing a "validation" of our prediction. Colored columns 1, 2, 3, and 4 show the predicted sequences for the four training tests described in Figure 3. Note the greater misalignment of the predicted sequences in (c) when using only 10 percent of the data for training.

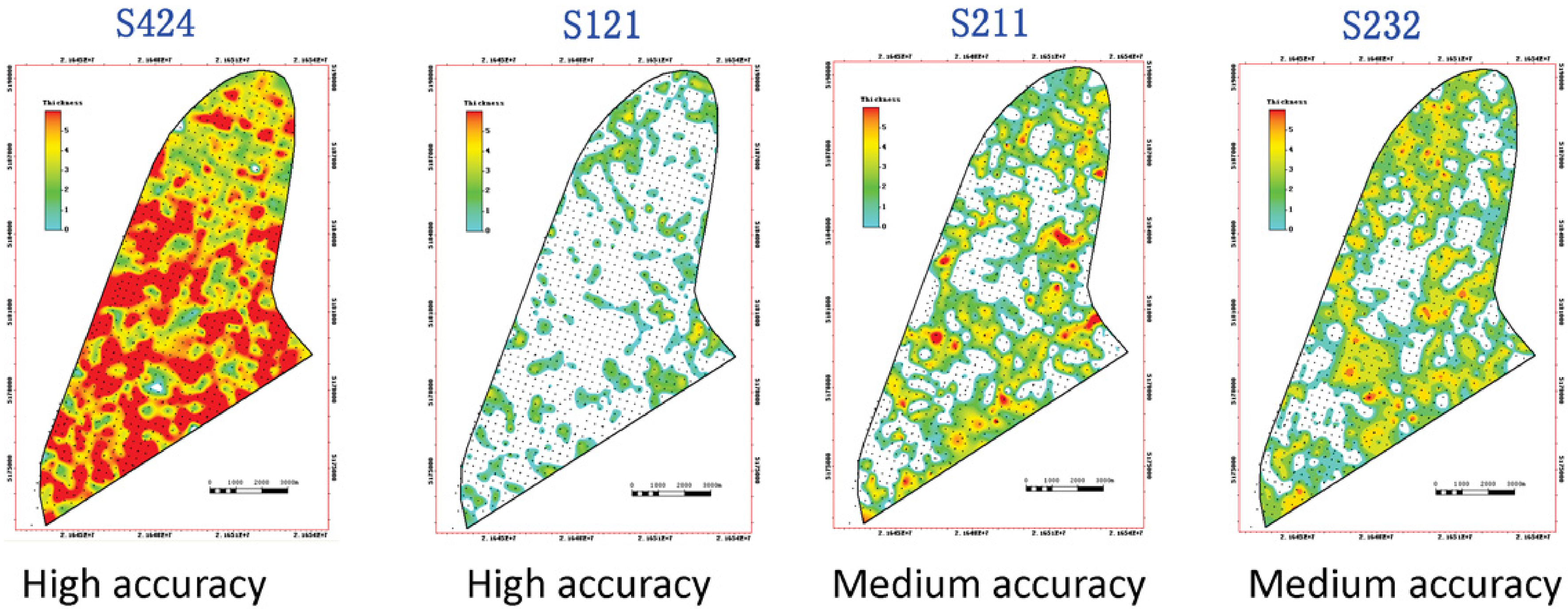

Figure 5. Representative fourth-order sequence thickness maps obtained by using CNN applied to predict the tops of 463 wells, using 20 percent of wells as training, and 80 percent as testing data. The higher accuracy prediction of sequences S424 and S121 are a due to relatively uniform lithofacies across the whole oil field. In contrast, the medium accuracy for sequences S211 and S232 are due to the high lateral lithologic heterogeneity, implying that patterns or in-context information identified by the skilled.