Deep-Learning Based Surrogate Model for Subsurface Flow Problem
Category: Others (Surrogate Modeling)

Haoyu Tang
Department of Energy Resources Engineering
Stanford University
hytang@stanford.edu

Su Jiang
Department of Energy Resources Engineering
Stanford University
sujiang@stanford.edu

Jingru Cheng
Department of Energy Resources Engineering
Stanford University
cjr63@stanford.edu

1 Introduction

There are plentiful existing reservoir simulators trying to predict the production rate, and have already achieved quite reliable effects. However, almost all of them suffer from being both time-consuming and expensive. This problem is even worse for our specific topic, as there are tremendous amount of possibilities of well locations to investigate. Both time and cost can be saved if we can successfully build a neural network based on only a small fraction of all possible simulations, and have robust and reliable performance on predicting long-term production rate of different well location combinations. The main challenges of this project are to determine the proper architecture of neural network for this problem and the proper format to input the well location as well as other geological data into the model.

In this project, we introduce a surrogate model aiming to replace the reservoir simulator, which can accurately predict production rate based on input well location and geophysical characteristics of this certain reservoir, and thus addressing the optimization problem of well location. At the early stage of our project, we started to try out using CNN directly to make a end-to-end prediction from well location directly to the cumulative oil production rate, which turned out to be a poor prediction model. Later, we borrowed the idea of U-net, especially the one discussed in [1]. The surrogate model employs Convolutional Neural Networks (CNNs) to capture the non-linear relationship between the spatial distribution of wells, geophysical parameter map and production rate. And then add LSTM layer to predict the production rate for each time step. We performed some trial and error on top of the original model for the preferred parameter selections.

2 Dataset specification

The data sets we used were generated from our in-house simulator AD-GPRS, and our well locations which are generated randomly are reformulated into a 2D matrix. To be specific, the input includes a permeability field matrix which would give the geological information of the reservoir and a matrix of well locations indicating where are the injectors and producers. The output data are simulated by AD-GPRS and they are the oil/water rate of both injectors and producers. There are in total 500 different cases in which 400 is training set and 100 is test set.

CS230: Deep Learning, Winter 2020, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

3 Approach

Our method to build the surrogate model for subsurface flow problem follows the basic structure of that in [1]. We will describe briefly about the model architecture, training process, and data pre-processing used here.

First of all, the model architecture we used here is a Recurrent Residual U-Net. As is shown in Fig. 1, the architecture includes a encoding net consisting of Convolutional and Residual Blocks, a decoding net (similar to encoding), and a ConvLSTM for the time sequence evolution. The encoding net would take the input of the permeability field matrix (which is a measure of the ability of a material (such as rocks) to transmit fluids) and the well location matrix to output some useful features for both ConvLSTM and decoding net; the decoding net would input the same features as well as the time sequence feature output by ConvLSTM to output the pressure/saturation maps for each time step. These maps are essential for the calculation of oil/water rate which would in turn give us the profitability of developing the oil reservoir.

Second, to train the model, we define the following loss function:

$$\ell = \arg\min_{\theta} \frac{1}{n_t} \sum_{t=1}^{n_t} ||\hat{y}^t - y^t||_p^p + \lambda \frac{1}{n_t} \frac{1}{n_w} \sum_{t=1}^{n_t} \sum_{w=1}^{n_w} ||\hat{y}_w^t - y_w^t||_p^p. \tag{1}$$

Where λ is the weight for the well state in which we want to in particular stress on since the near well has a much stronger effect on the well rate. Besides, n_t stands for number of time step and n_w is the number of wells. Here, p is the norm we use, e.g. L2 norm. And, the $\hat{y}$ and y are the state (pressure/saturation) of the test case from our model prediction and from simulation results respectively. During the training, we set 500 epochs using mini-batch with a batch size of 8. The training is performed on our home department's cluster (CEES-Mazama) using a Nvidia Tesla V100 GPU. The training takes about 3-4 hours to finish.

Finally, to be able to input useful information about the well location (which is generated randomly for different cases), we introduce a matrix where "1" denoting the injection well, "-1" as the production well, and "0" meaning no well. This matrix together with the permeability field matrix will be lumped together as two channels to input into our neural network. Since the well location matrix is sparse, we introduce some 3×3 mask mini-matrix around the well locations to make our model more aware of the well locations. After inputting large enough training set and training for appropriate iterations, this surrogate model is justified to accurately give consistent predictions to simulator results.

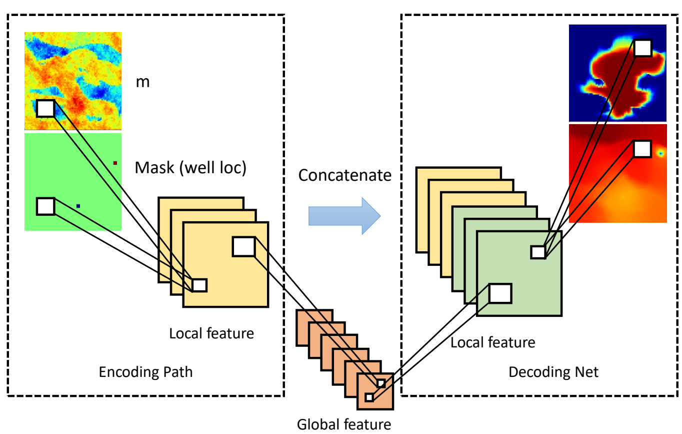

Figure 1: The architecture of U-Net

In addition, since the cumulative oil production or the oil rate is the most important data for drilling decision, we introduce the formula to compute the rate from the pressure and saturation which are the primary output of our model:

$$(q_o^w)_i = WI_i(\lambda_o\rho_o)_i(p_i - p^w). \tag{2}$$

In this equation, $(q_o^w)_i$ is the oil rate of producer i, λ_o and ρ_o are oil phase mobility and density respectively, p_i and p^w are the field pressure at grid block i and well bottom hole pressure (BHP) respectively, and the WI_i is the well index computed by,

$$WI_i = \frac{2_i\Delta z}{log(r_o/r_w)}. \tag{3}$$

Here, k_i is the permeability at grid block i, Δz is the grid block thickness, r_w is the wellbore radius, and $r_o = 0.14\sqrt{((\Delta x)^2 + (\Delta y)^2)}$, in which Δx and Δy are grid block length and width respectively in x and y directions.

4 Results

Figure 2 shows the permeability field of the channelized model and a training sample of well locations. As is shown in Fig. 6 and Fig. 7, the lower part which is our model prediction results for pressure and saturation respectively seem to be quite close to the upper part which is the simulation results.

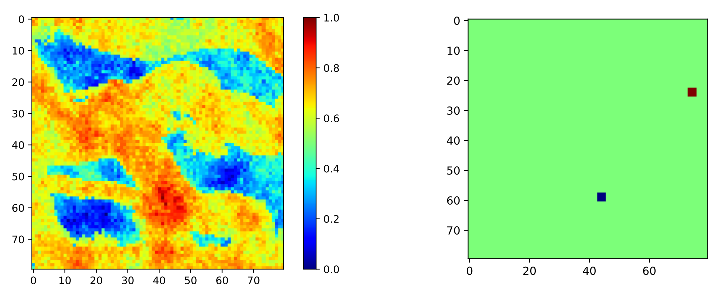

Figure 2: Permeability field (shown in the left plot) and well location (shown in the right plot)

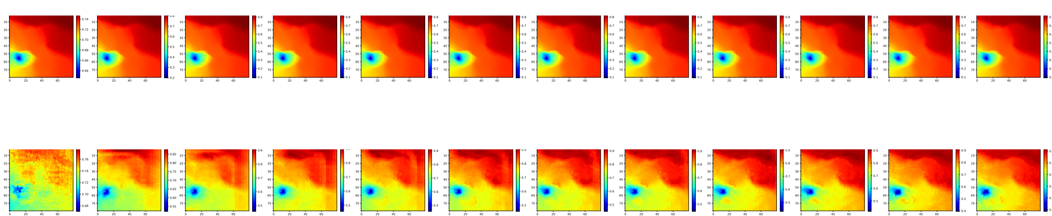

Figure 3: Pressure of test sample (Plots on the top show the simulation results, and plots on the bottom show the prediction)

To realize the actual situation in simulator, we further implemented our model to four-well case.Figure 5 shows a training sample of well locations, permeability is still the same as two-well case.

The learning process of the model is shown in Fig. 8, which represents the change of relative error, which is defined as the absolute error between real value and predicted value divided by real value. From the learning curve, we can see that the model is learning quickly at the beginning and decrease drastically later, which may results from the complexity of our model.

As is shown in Fig. 6 and Fig. 7, the lower part which are our model prediction results for pressure and saturation respectively are still quite close to the upper part which are the simulation results. And Fig. 9 shows the oil rate of the producer, which indicates the very close approximation between predicted result and analytical results.

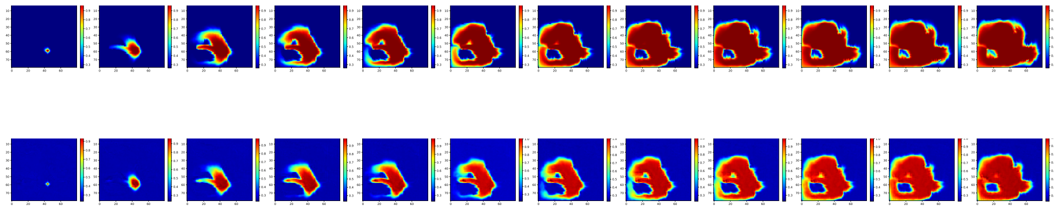

Figure 4: Saturation of test sample (Plots on the top show the simulation results, and plots on the bottom show the prediction)

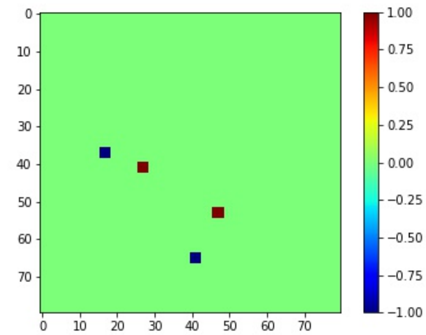

Figure 5: Well location for 4-well case

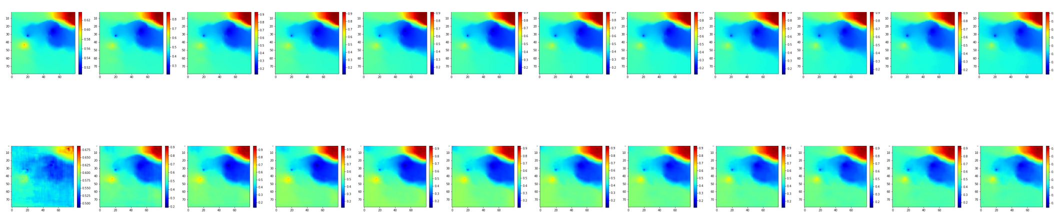

Figure 6: Pressure of test sample (Plots on the top show the simulation results, and plots on the bottom show the prediction)

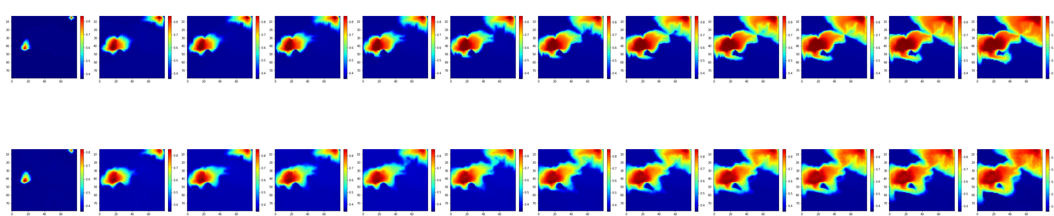

Figure 7: Saturation of test sample (Plots on the top show the simulation results, and plots on the bottom show the prediction)

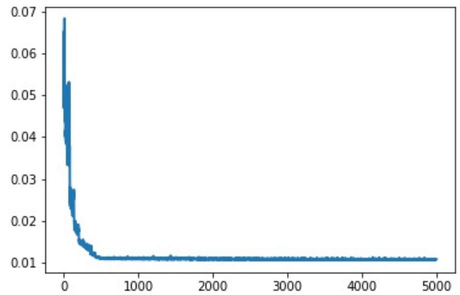

Figure 8: Learning Curve

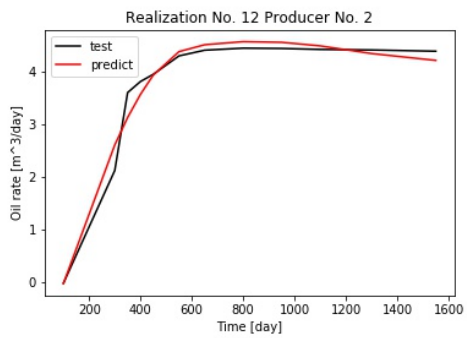

Figure 9: oil rate

5 Summary

In this project, we implement a Recurrent Residual U-Net based surrogate model to generate flow results for channelized reservoir model with different well locations. We consider the same geological model with two different well settings (two wells and four wells). The results show that the surrogate model could generate relatively close pressure and saturation maps to the simulation results. However, some discrepancies are observed in the flow rate result, since the flow results mainly relies on the pressure and saturation around the well blocks. So we will further test the model and add more weight to the blocks around the well in the loss term to improve the flow results. In the future, we will also consider to combine the surrogate model with the well location optimization to maximize the production and profit for the oil and gas problem.

[1] Tang, M., Liu, Y., & Durlofsky, L. J. (2019). A deep-learning-based surrogate model for data assimilation in dynamic subsurface flow problems. *arXiv preprint arXiv:1908.05823*.

[2] Zhou, Y. (2012) Parallel general-purpose reservoir simulation with coupled reservoir models and multisegment wells. Doctoral dissertation, Stanford University.