UCFER 2021 Annual Review Meeting

Upscaling Experimental Measurements to the Field Scale Using a Machine-Learning-Based, Scale-Bridging Data Assimilation Approach (04-VaT-O1-09)

Cheng Chen, Ph.D.
Adjunct Associate Professor
Virginia Tech

10/5/2021

Team Members

Virginia Tech

Faculty: Cheng Chen (PI), Heng Xiao, James McClure, Nino Ripepi, Ming Fan

Graduate Students: Hongsheng Wang, Xuhui Zhou

NETL Collaborators

Dustin Crandall

Laura Dalton

Illinois State Geological Survey Collaborators

Sallie Greenberg

Sherilyn Williams-Stroud

Tasks and Milestones

- Project period: 11/01/2019 – 12/31/2021
- Task 1: Develop *supervised* and *unsupervised* machine learning (ML) methods to automatically segment the enormous volume of CT images of reservoir rocks in NETL's Geoimaging Lab (Aim R1; completed)
- Task 2: Use physics-informed ML models to predict rock's transport properties (e.g., absolute permeability [k] and relative permeability [k_r]) from the segmented CT images (Aim R1; partially completed)
- Task 3: Develop a Graphical User Interface (GUI) to manage and visualize rock CT images and lab-measured k_r curves from Geoimaging Lab (Aim R2; completed)
- Task 4: Work with the UI team to collect well logging data from the IBDP project (CCS1) and ICCS project (CCS2) (Aim R3; completed)
- Task 5: Develop a scale-bridging data assimilation framework to integrate rock's transport properties obtained at various spatial scales and to calibrate the upscaled k_r curves using well-scale observation data of CO_2 plume migration (Aim R3; partially completed)

Publications and Products

Journal papers:

1. Fan, M., Y. Han, X. Tan, L. Fan, E. S. Gilliland, N. Ripepi, and C. Chen (2021), Experimental and Numerical Characterization of Lower Huron Shale as a Heterogeneous Material, ***Rock Mechanics and Rock Engineering***, 54(8), 4183-4200. https://doi.org/10.1007/s00603-021-02491-2.
2. Fan, M., J. McClure, R. Armstrong, M. Shabaninejad, L. E. Dalton, D. Crandall, and C. Chen (2020), Influence of Clay Wettability Alteration on Relative Permeability, ***Geophysical Research Letters***, 47, e2020GL088545, https://doi.org/10.1029/2020GL088545.
3. Guo, R., L. E. Dalton, M. Fan, J. McClure, L. Zeng, D. Crandall, and C. Chen (2020), The Role of the Spatial Heterogeneity and Correlation Length of Surface Wettability on Two-Phase Flow in a CO2-Water-Rock System, ***Advances in Water Resources***, 146, 103763, https://doi.org/10.1016/j.advwatres.2020.103763.
4. Zhou, X., J. McClure, C. Chen, and H. Xiao (2021), Neural Network Based Velocity Prediction in Porous Media from Geometry assistedwith Super-resolution, ***Physical Review Fluids***, under review.
5. Wang, H., L. Dalton, M. Fan, R. Guo, J. McClure, D. Crandall, and C. Chen (2021), Deep-learning-based workflow for partial volume segmentation in rock CT images: Application of entropy-assisted indicator kriging and UNet++, ***Advances in Water Resources***, under review.
6. Wang, H, D. Crandall, L. Dalton, and C. Chen (2021), Multi-phase flow simulation based on multi-class segmented digital rock images: Application of the unsupervised learning algorithm, in preparation.

Publications and Products

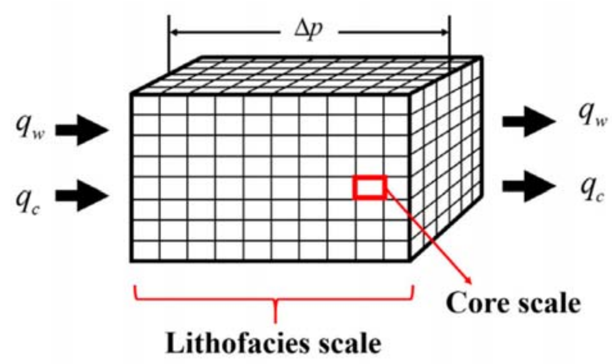
Conference papers:

1. Wang, H., D. Crandall, L. Dalton, and C. Chen (2021), Application of Convolutional Neural Networks in Digital Rock Segmentation: Supervised and unsupervised learning algorithms, 45th International Technical Conference on Clean Energy, July 26 - 29, Clearwater, Florida, USA, Paper ID: 154, paper length: 10 pages.
2. Wang, H., D. Crandall, L. Dalton, and C. Chen (2021), Applications of Convolutional Neural Networks in Digital Rock Segmentation: Supervised and Unsupervised Machine Learning Algorithms, Poster ID: 756, InterPore2021, May 31 – June 4, Berlin, German.
3. Wang, H, D. Crandall, L. Dalton, and C. Chen (2020), Multi-phase Segmentation of Digital Rock Images Using Convolution Neural Network: Training Dataset Generation, Model Training, and Result Visualization, Poster ID: IN011-13, American Geophysical Union (AGU) 2020 Fall Meeting.

Software:

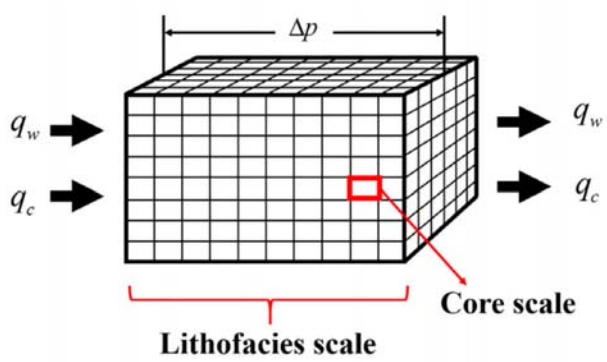
1. A Jupyter Notebook based GUI software package for the management and visualization of rock CT images and core flooding experiments.

Core to Large-Scale Upscaling



Gray-scale CT → **Manuel image processing** (**min to hour**) → Binary-scale CT → **Direct LB modeling** (**10 ~ 100 hours**) → k and kr

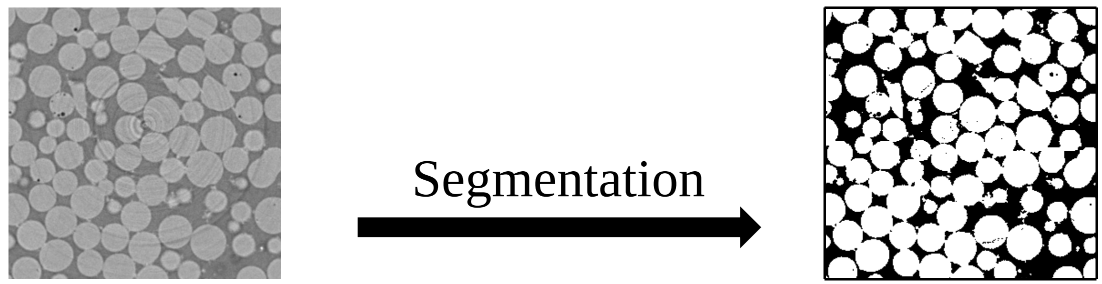
Core to Large-Scale Upscaling



Gray-scale CT → **Manuel image processing** (**min to hour**) → Binary-scale CT → **Direct LB modeling** (**10 ~ 100 hours**) → k and kr

Gray-scale CT → **Image processing with ML** (**millisecond**) → Binary-scale CT → **Flow field prediction with ML** (**second**) → k and kr

Digital Image Segmentation

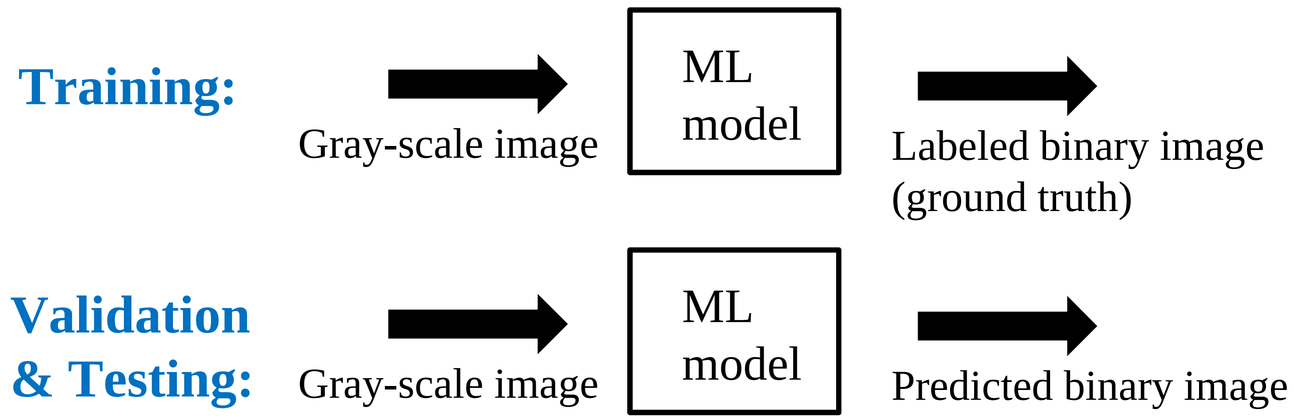


Raw CT image   Binary image

- Segmentation of gray-scale digital images into binary images before pore-scale modeling.
- Manual segmentation can be slow and has uncertainties.
- An industry CT scanner can generate 1500 gigabytes of raw CT image data (more than 50000 2D images) per week.
- Need a consistent, efficient, and automatic approach.
- Supervised machine learning can do the job.

1. Supervised Learning

How to Find Ground Truth in Supervised ML?



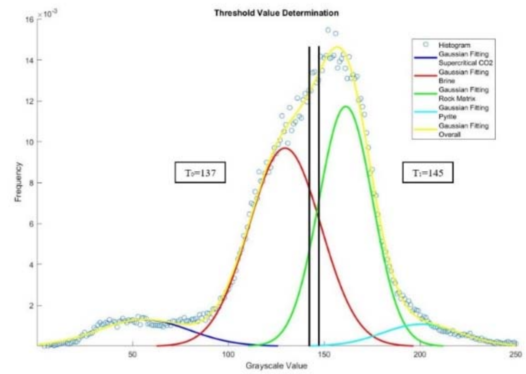
- Need labeled data for model training and validation.
- It is kind of a chicken and egg problem…
- Developed an indicator kriging workflow to generate high-accuracy labeled image data sets.

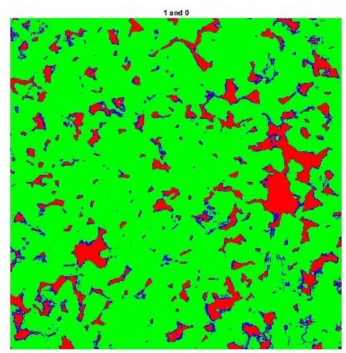
Generation of High-Accuracy Labeled Data

Information Entropy:

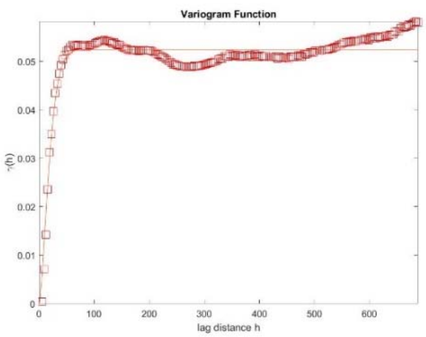
$$H = -\sum_{i=1}^{n} p(i) log_2 p(i)$$

Pixel with a higher entropy value has a higher chance being a boundary pixel





Red: void space
Green: solid
Blue: uncertain



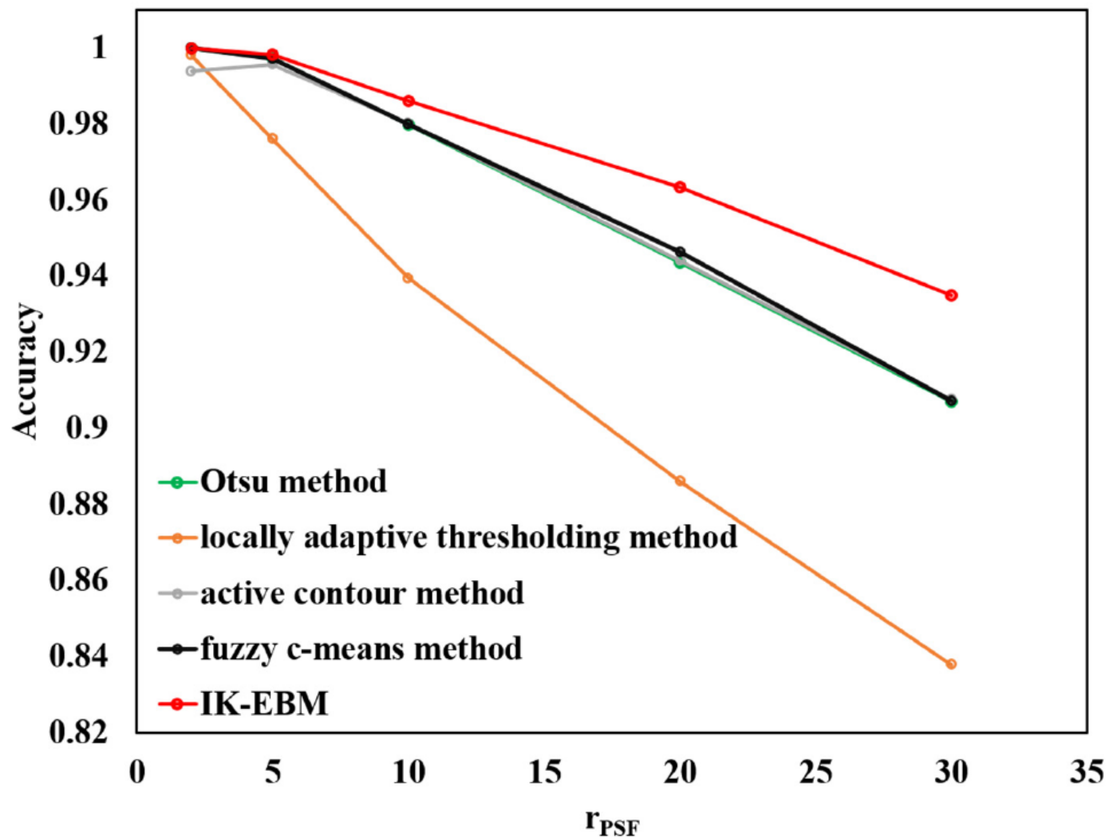
$$Label = \begin{cases} 0, if\, G^*(x_0) < 0.5 \\ 1, otherwise \end{cases}$$

$$where\ G^*(x_0) = \sum_{n=1}^{N} b_n G(x_n)$$

$$\begin{cases} \sum_{m=1}^{N} b_m \bar{\gamma}(x_n; x_m) + \lambda = \bar{\gamma}(x_n; x_0), n = 1, \dots, N \\ \sum_{m=1}^{N} b_m = 1 \end{cases}$$

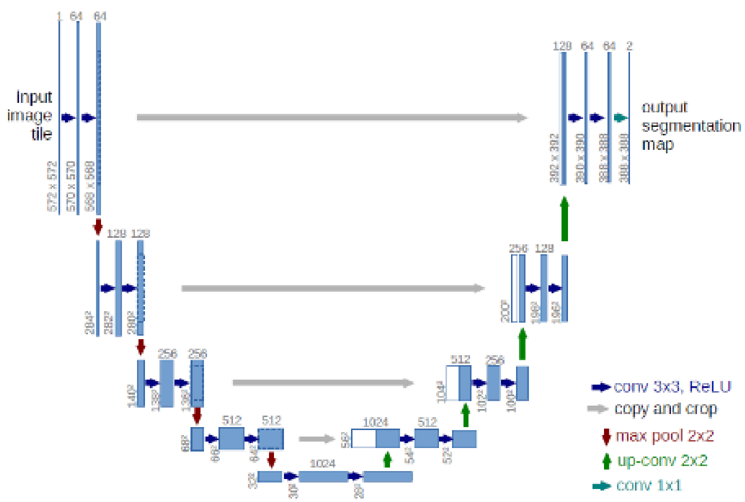
Use indicator kriging to determine the label of unknown pixels based on the variogram function

Generation of High-Accuracy Labeled Data

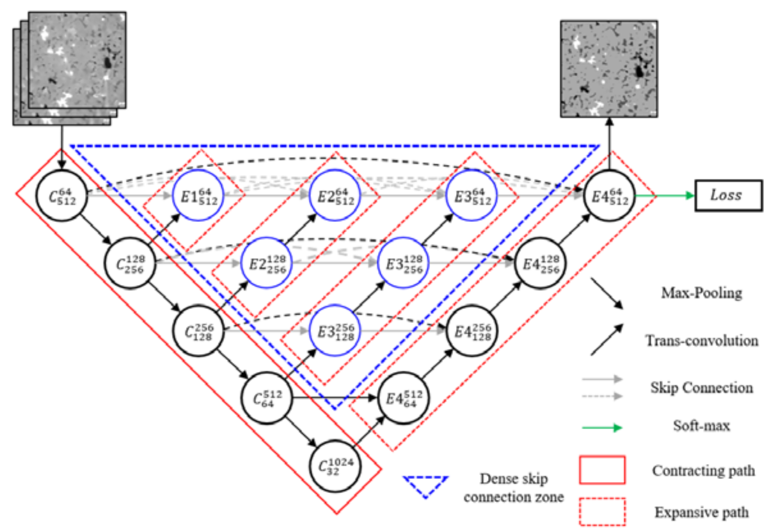


Entropy-based-masking indicator kriging (IK-EBM) mitigates the partial volume blurring (PVB) effect and leads to the highest segmentation accuracy.

U-Net ++



U-net (Ronneberger et al., 2015)

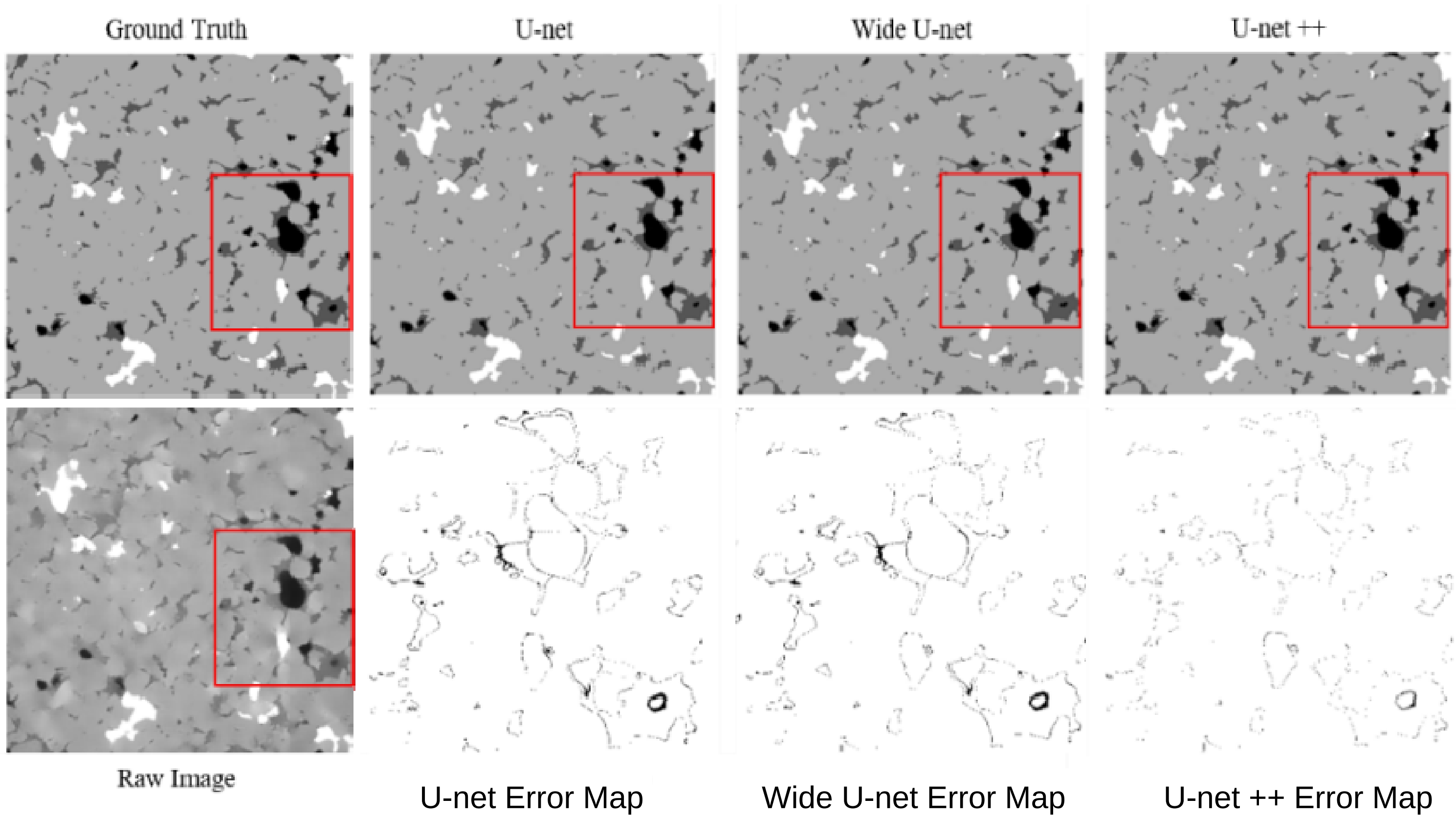


U-net++ (Wang et al., 2021; after Zhou et al., 2019)

- A state-of-the-art supervised learning model, U-net ++.
- Re-designed skip connections, which improve segmentation accuracy and make training faster.

1. Supervised Learning

U-Net ++

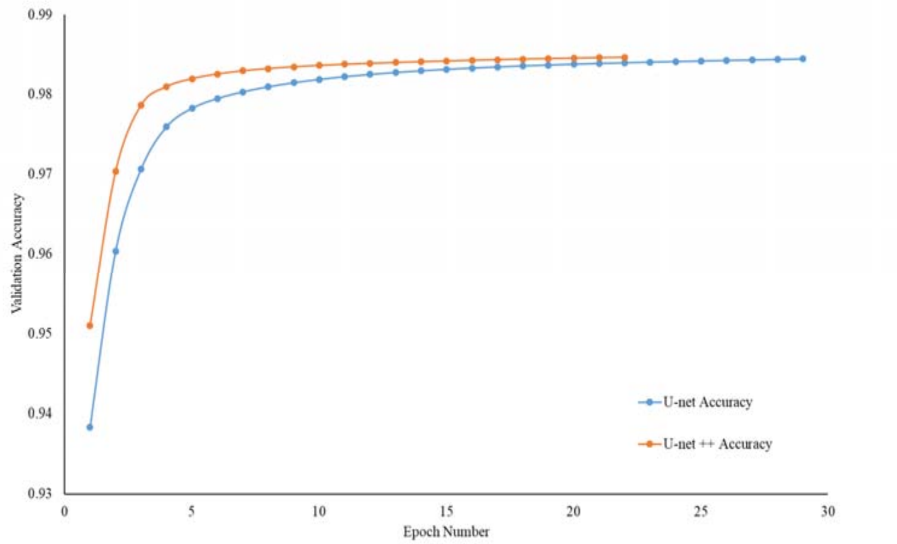


Model comparison

- Improved segmentation performance.
- Main improvements on boundary and small target segmentation accuracy.

1. Supervised Learning

U-Net ++



Validation accuracy vs. Epoch

Test accuracy: Mean and class accuracy

	Mean	$scCO_2$	Brine	Rock	Pyrite
U-net	0.9714	0.8480	0.7441	0.9983	0.9514
Wide U-net	0.9724	0.8737	0.7479	**0.9987**	0.9778
U-net ++	**0.9805**	**0.9202**	**0.8153**	0.9985	**0.9832**

➢ Multiscale feature aggregation in U-net++ better extracts fine-scale features (e.g., solid-void boundaries) and leads to faster convergence.

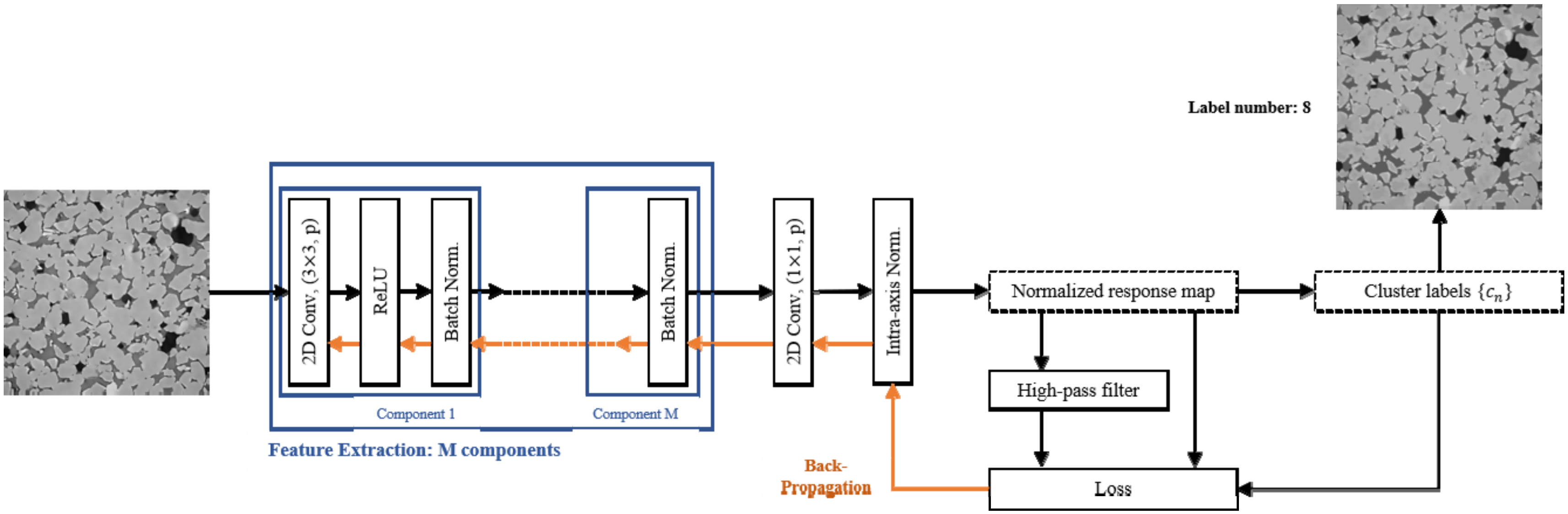
1. Unsupervised Learning

- How about the cases in which we do not have label data
- Or no time to generate large-volume label datasets?
- Can unsupervised machine learning help?

Clustering: pixels having similar features receive the same label

Three criteria [Kanezaki, A. (2018, April)]:
(a) Pixels with similar features are desired to be assigned the same label.
(b) Spatially continuous pixels are desired to be assigned the same label.
(c) The number of cluster labels is desired to be large.

1. Unsupervised Learning



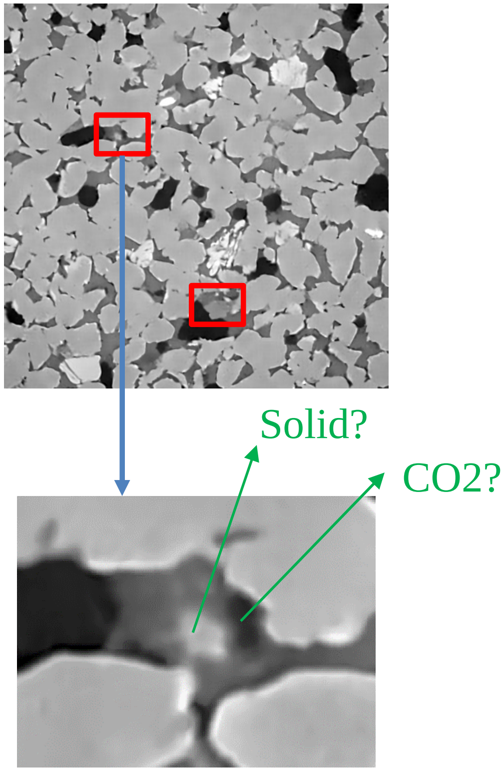
Unsupervised machine learning algorithm(Wang et al., 2021; after Kim, et al. 2020)

Step 1: Prediction of cluster labels with a fixed network.
Step 2: Training of network parameters with predicted cluster labels.
Step 3: Repeat Steps 1 and 2 until the loss function minimized.
Step 4: Re-clustering to three labels: **$scCO_2$, brine, and rock**.

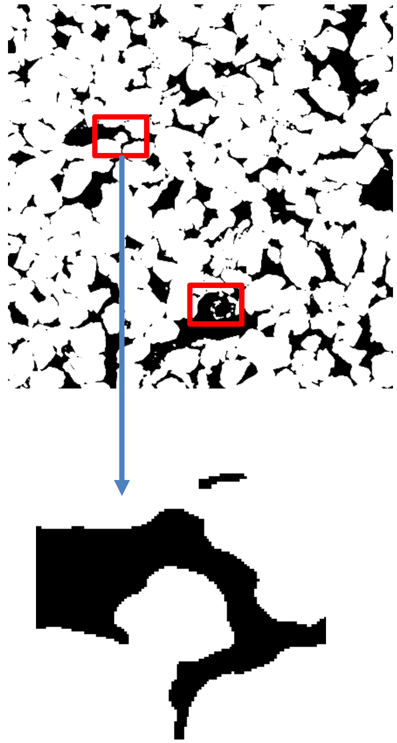
Three criteria [Kanezaki, A. (2018, April)]:
(a) Pixels of similar features are desired to be assigned the same label.
(b) Spatially continuous pixels are desired to be assigned the same label.
(c) The number of unique cluster labels is desired to be large.

1. Unsupervised Learning

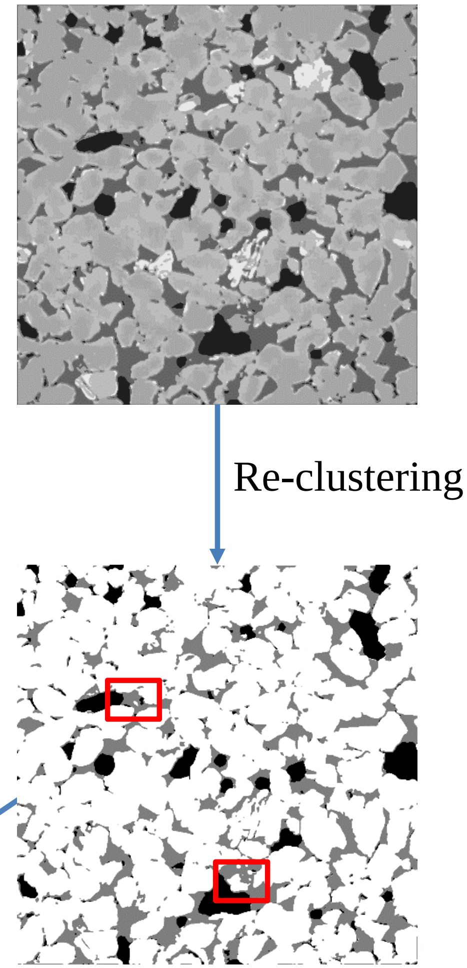
Raw Image:
three phases: CO_2, Brine, and Rock.



Grayscale-based image segmentation

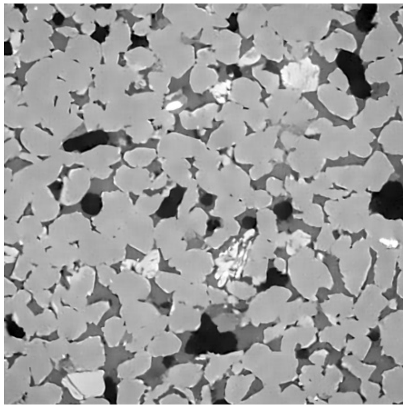


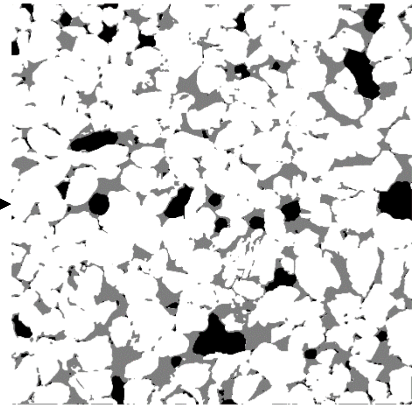
Unsupervised-learning segmentation: 8 labels.



Advantages:
(1) Automatic, fast and accurate;
(2) Feature continuity and local smoothness.

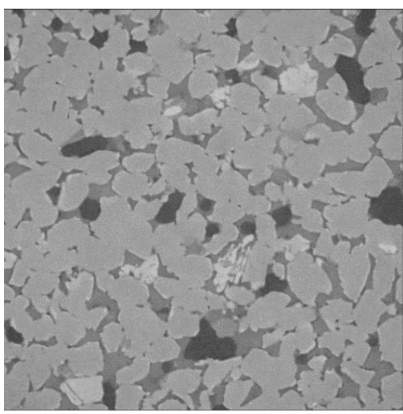
1. Unsupervised Learning

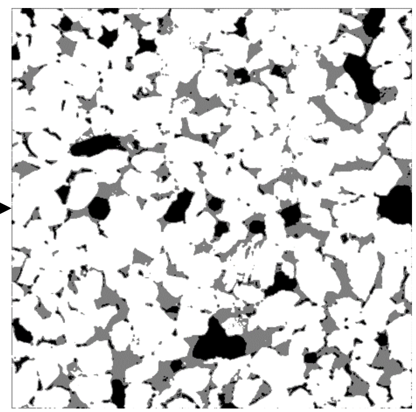




Raw Image - three phases: CO_2, Brine, and Rock.

Unsupervised segmentation





Raw Image with Gaussian noise: Mean: 0; S.d.: 20.

Unsupervised segmentation

- Great noise tolerance.

Result Discussions

- The IK-EBM mitigates the PVB effect and leads to improved segmentation accuracy and high-quality training datasets.
- Convolutional Neural Networks achieve good performance of image segmentation because of local connectivity, weight sharing, and scalability robustness.
- Features extracted by CNN not only include grayscale values, but also contain geometric info and pattern.
- Supervised learning algorithms maximize the value of existing data, including raw images and segmentation images.
- A well-trained supervised ML model can achieve segmentation quality as good as training data at the **millisecond scale**.
- If training data are unavailable, unsupervised learning algorithms provide automatic, fast, and accurate segmentation. Processing time can be as short as **seconds**.

Physics-Informed Machine Learning

Data-driven training:

High-resolution pore geometry → ML model → High-resolution flow field (LB simulation on a fine mesh)

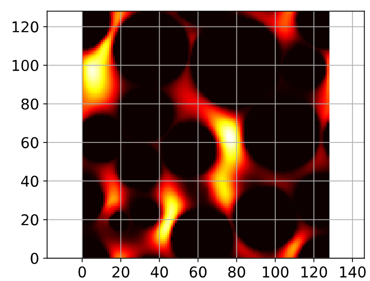
Physics-informed training:

High-resolution pore geometry
+ Low-resolution flow field (LB simulation on a coarse mesh) → ML model → High-resolution flow field (LB simulation on a fine mesh)

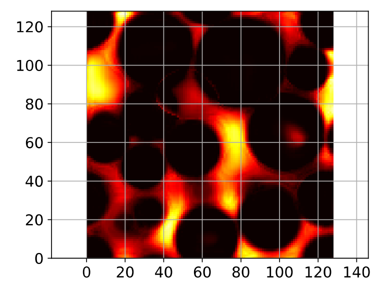
- Low-resolution flow field provides physics information and constraints.
- Enhances model accuracy at a relatively low cost.
- Effective in simulating flow fields in vuggy pore space, which is highly relevant to energy recovery in carbonate reservoirs.

2. Physics-Informed Machine Learning

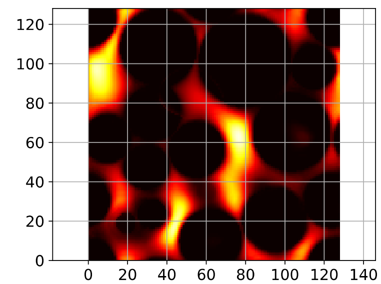
$$\epsilon = \frac{||v_{predict} - v_{truth}||}{||v_{truth}||}$$



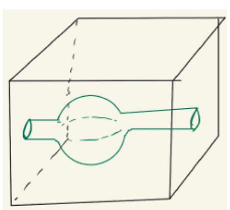
Ground truth

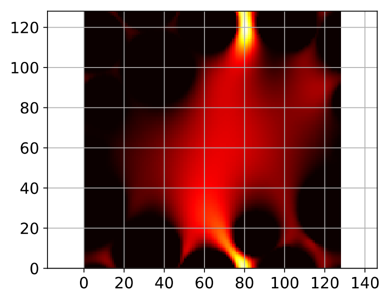


pred. w/o coarse velocity
Validation error: 11.52%

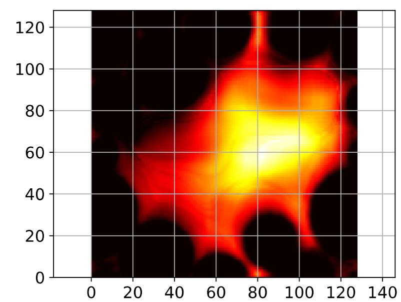


pred. w/ coarse velocity
Validation error: 7.44/%

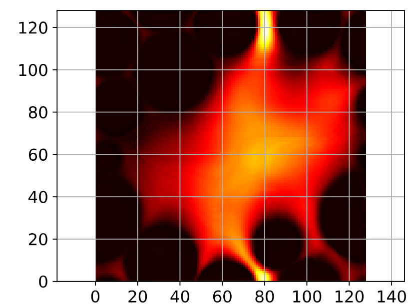




Ground truth



pred. w/o coarse velocity
Validation error: 60.16%

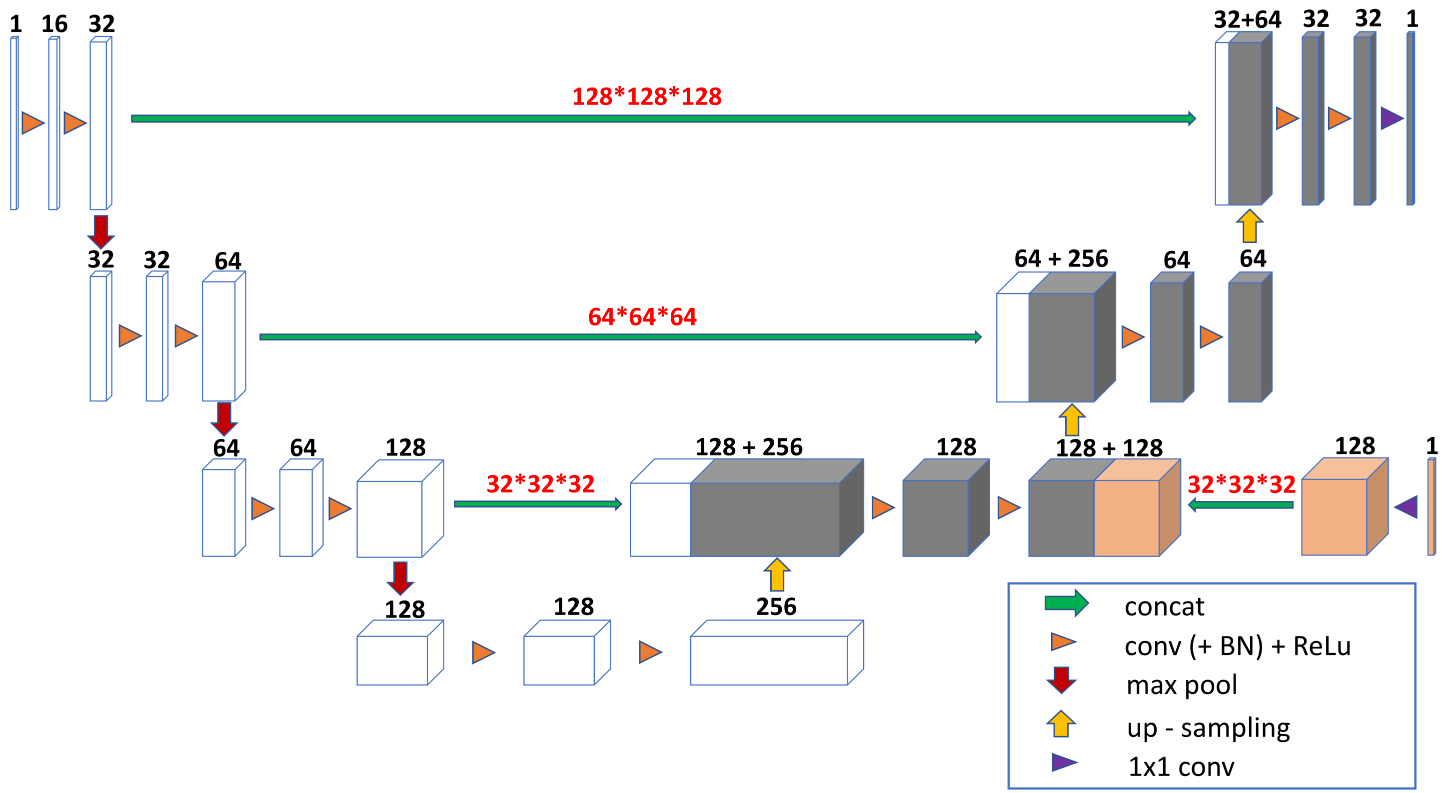


pred. w/ coarse velocity
Validation error: 17.74/%

- Pure geometry input leads to discontinuity of flow field in vuggy space.
- Low-resolution flow field input alleviates this issue, leading to lower generalization error.

2. Physics-Informed Machine Learning

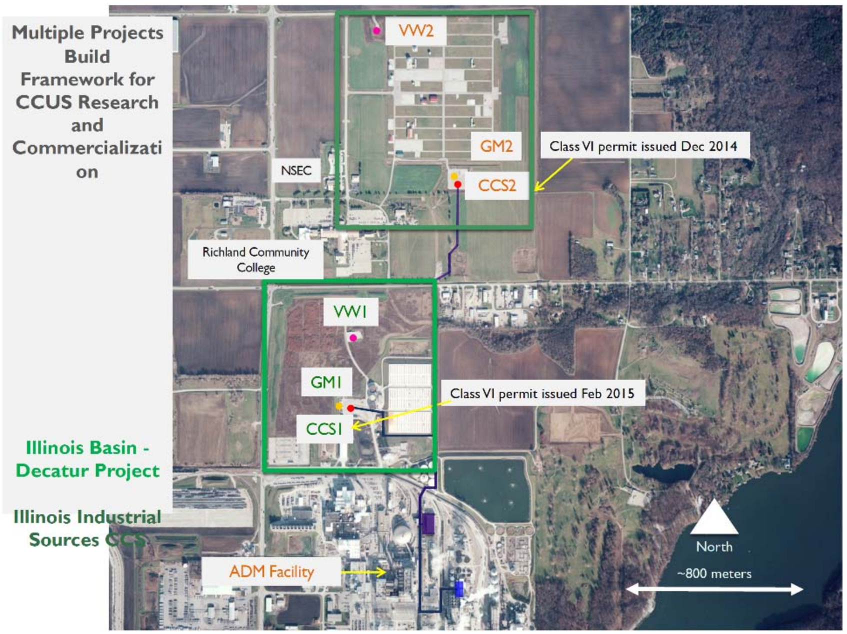
Neural Network Architecture



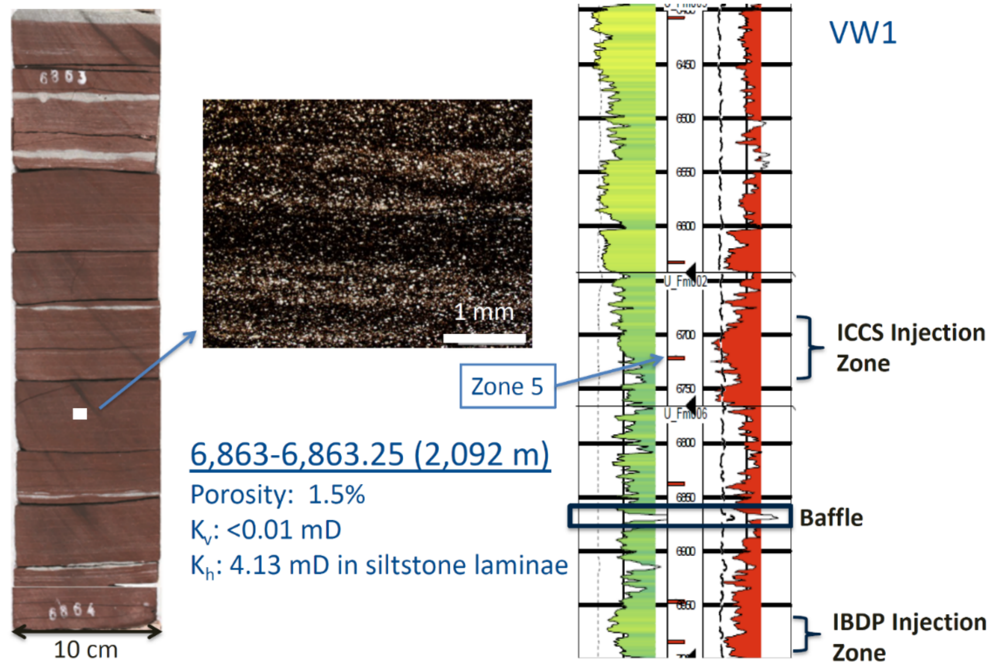
Schematic of the proposed convolutional neural network to predict the flow field in porous media. Gray boxes represent feature maps. White boxes represent copied feature maps from the encoder part; brown boxes represent the feature maps from the input of coarse scale velocity.

The number of channels is denoted above each feature map. The sizes of 3D images in each level are denoted above the concatenation arrows. The arrows with different colors denote the different operations.

3. Well Logging Data and ML



Well Locations



Perforation Depth and Target Reservoir: Baffle guarantees project independence

Project Name	Well Name	Lithology Interpretation Data	CO2 Saturation Monitoring Data
Illinois Basin – Decatur Project (IBDP)	CCS1	Available	Available
	VW1		
Illinois Industrial Carbon Capture and Storage Project (IL-ICCS)	CCS2		
	VW2		

CCS: Carbon capture and storage; VW: Verification well

3. Well Logging Data and ML

- Well logging interpretation requires rich experience, time, and expense.
- Well-trained model allows beginners to obtain good interpretation results.

Goals:

(1) Mineral composition prediction from raw log data (in progress)
(2) Rock type interpretation from raw log data (in progress)
(3) CO_2 saturation interpretation using supervised or unsupervised learning. (in progress)

Goal (1) and (2) use same dataset; Goal (1) and (3) use the same algorithm.

(1) Algorithms:
KNN; Decision tree; Random forest; XGBoost; LightGBM with or without wavelet transform data; Support vector machine.

(2) Useful Python packages:
Data analysis: Pandas, Numpy, Seaborn;
Well logging data pre-processing: Cegal;
Machine learning implementation: Keras, TensorFlow, sklearn.

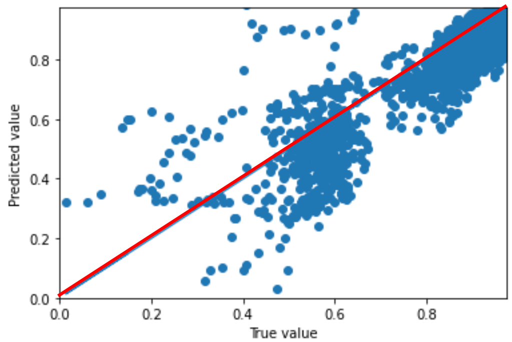
3. Well Logging Data and ML

x:True value and y: predicted value

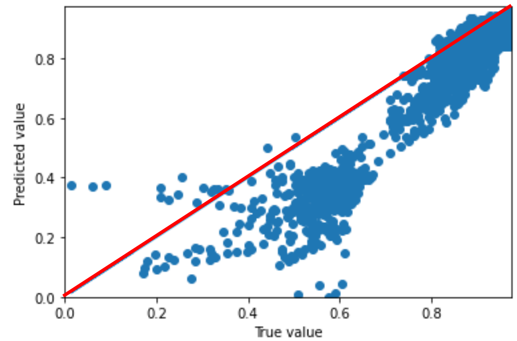
Reference line: y = x

Methodology: Quartz composition as an example.

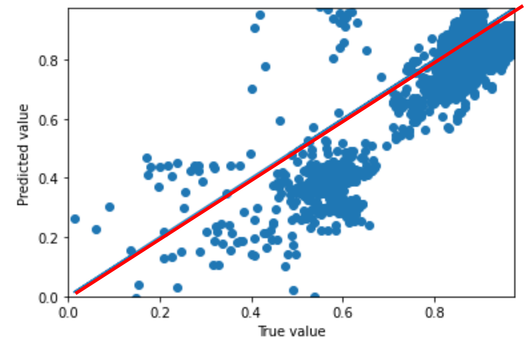
(1) Support Vector Regression (SVR) with different kernels.



Radial basis function kernel
MSE: 0.00874

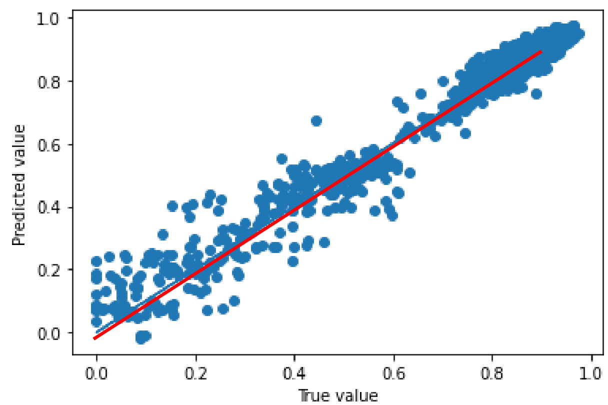


Linear kernel
MSE: 0.01212



Polynomial kernel
MSE: 0.02586

(2) Deep Neural Network (DNN): Works well for quartz. MSE still high for minerals having small fractions. Need to test more network architectures and hyper-parameters.



Fully connected networks:
Three layers with 100 nodes in each layer
MSE: 0.00159

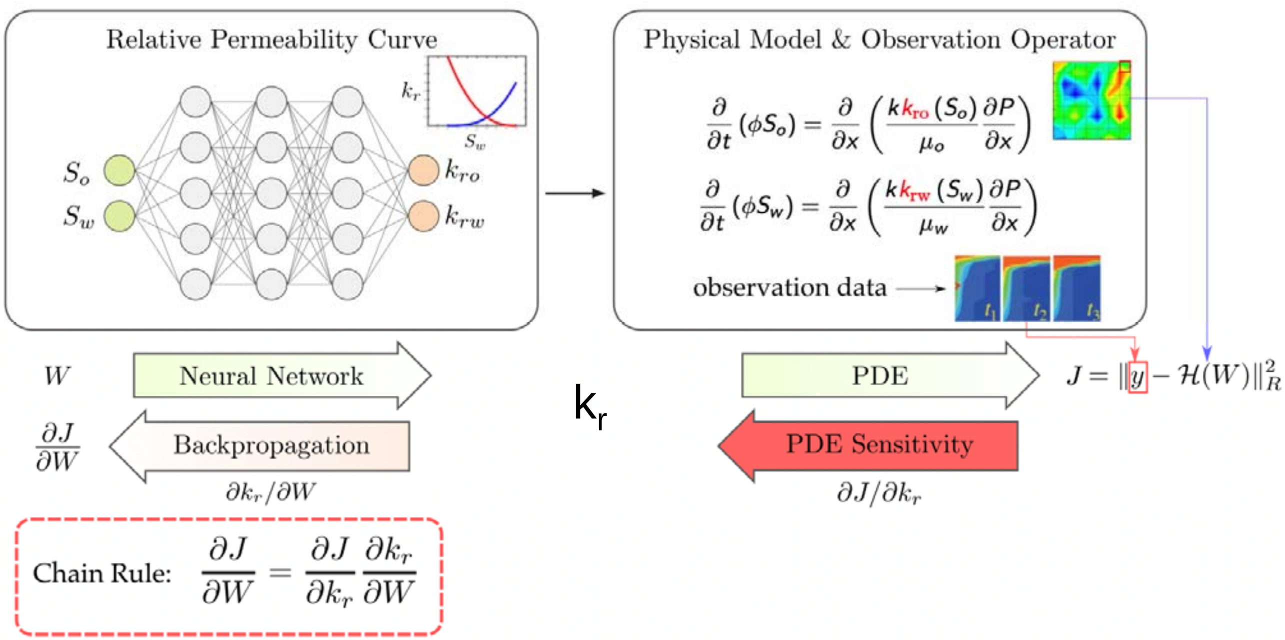
4. Data Assimilation Framework

Goals: Learning relative permeability curve by data assimilation

Major modules:

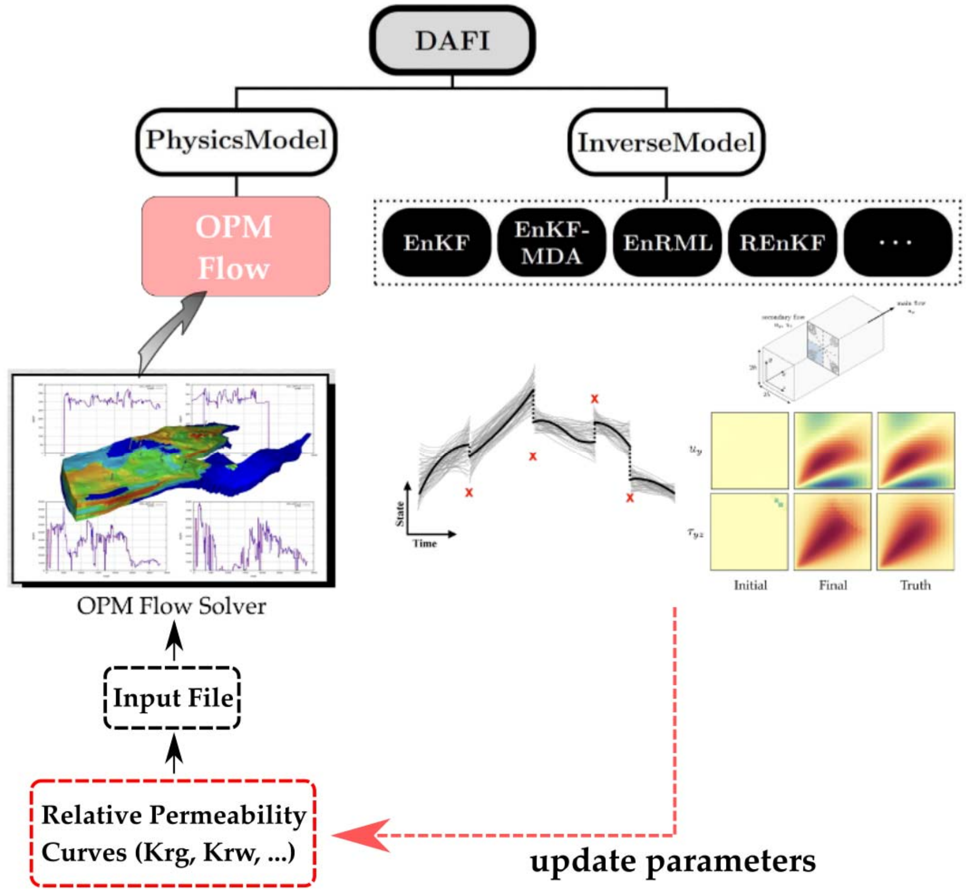
(1) Neural network for k_r modelling;

(2) DAFI + OPM Flow Solver for computing gradient $\frac{\partial J}{\partial k_r}$.



DA framework for k_r curve learning: k_r is an intermediate, hidden variable because it cannot be directly observed. Observation data (e.g. oil/gas saturations in some cells) are used to update the weights.

4. Data Assimilation Framework

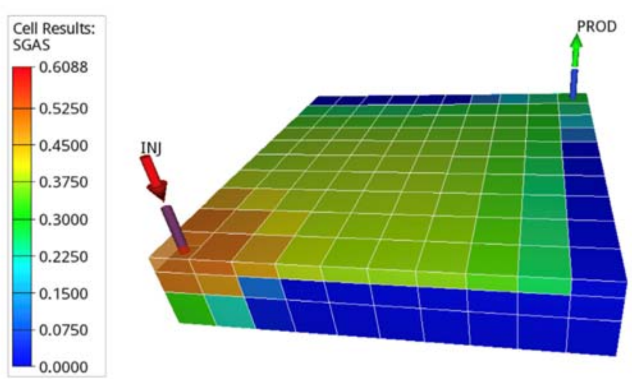


Overview of the framework: DAFI contains two main classes, the PhysicalModel and InverseMethod. The physical model is provided by the OPM Flow Solver and input file (including the Kr curves). The InverseModel updates the parameters in Kr curves based on the observation data.

4. Data Assimilation Framework

Study with A Benchmark Case – SPE1

- The Society of Petroleum Engineers (SPE) has run a series of comparative solutions projects with the aim of comparing and benchmarking different simulators.
- In this study, the first project (SPE1) is introduced as the three-dimensional black oil simulation benchmark case.
- The first well (red) injects gas from the top layer at a rate of 100 MMscf/day with a BHP limit of 9014 psia; the other well (green) is set to produce oil from the bottom layer with a production target of 20000 stb/day and BHP limit of 1000 psia.



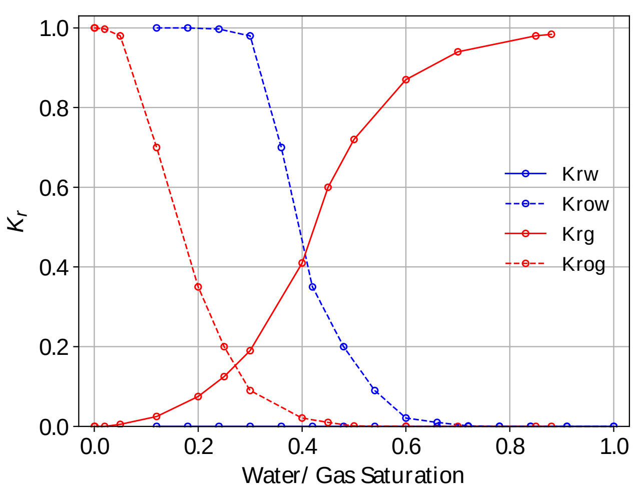
- Grid: 10 x 10 x 3 cells;
- Computation domain: 10000 ft x 10000 ft x 100 ft;
- Thickness of each layer: 20 ft, 30 ft, 50ft;
- The reservoir is initially undersaturated;
- Constant porosity of 0.3 throughout all 300 cells;
- The absolute permeability are 500 mD, 50 mD and 200 mD in the top, middle and bottom layers for all x, y and z directions.

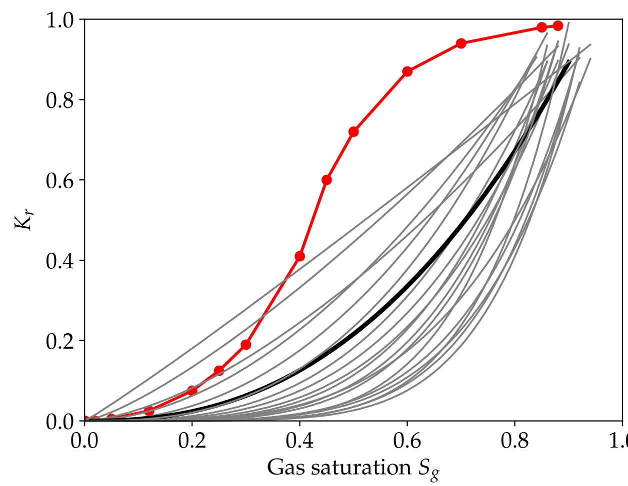
Parameterization of Relative Permeability Curve

- The ground truths of relative permeability curves (Krw, Krow, Krg, Krog) are shown below. Here, we assume the Krg curve is unknown, which is to be learned by data assimilation.
- Corey-model is used for the approximation of gas relative permeability (Krg) curve:

  $K_{rg} = K_{rg,\,max}\,(\frac{S_g - S_{gc}}{1 - S_{or} - S_{wc} - S_{gc}})^{N_g}$. For convenience, we set $S_{gc} = S_{or} = 0$ as they exactly are. Three variables S_{wc}, N_g and $K_{rg,max}$ are used for the parameterization.
- Latin Hypercube Sampling (LHS) method are used for sampling S_{wc}, N_g and $K_{rg,max}$.

  S_{wc} ~ [0.05, 0.15], N_g ~ [1, 6] and $K_{rg,max}$ ~ [0.9, 1].
- The commonly used Corey-model cannot accurately describe the true Krg curve (red). Here, we assume a baseline Krg curve (**black**) to be modelled.





Conclusions and Discussion

- Machine learning (ML) and data analytics have been successful in solving many technical challenges.
- ML significantly accelerates digital rock image segmentation and flow property estimation.
- ML and data assimilation can be combined to match both the observations (e.g., saturation) and intermediate, hidden variables (e.g., k_r).
- How we use them to identify and solve the *big questions* in subsurface engineering?

Thank You !