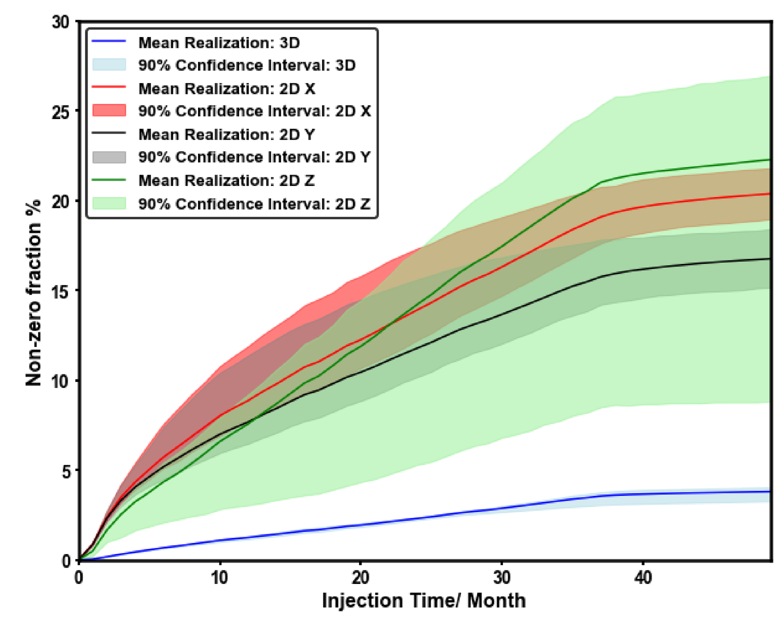
Efficient Dimension Reduction of Complex Three-dimensional CO_2 Saturation Using Deep Learning Models

Hongsheng Wang (hongsheng.wang@beg.utexas.edu) [1], Seyyed Hosseini (seyyed.hosseini@beg.utexas.edu) [1]
[1] Bureau of Economic Geology, Jackson School of Geosciences, The University of Texas at Austin, Austin, TX 78758, USA.

Science-informed Machine Learning to Accelerate Real Time (SMART) Decisions in Subsurface Applications

Motivation

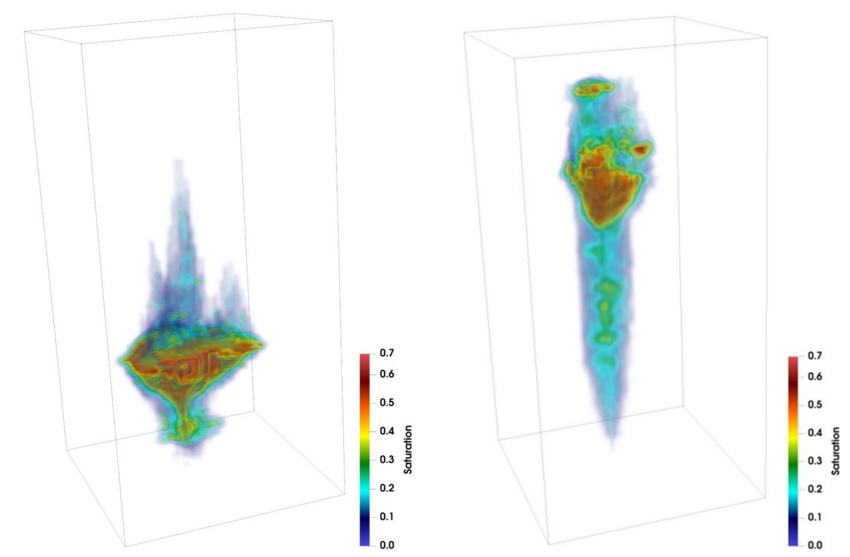
- Dimension reduction optimizes deep learning training by extracting latent variables with fewer numbers and retaining most data features, effectively reducing overfitting risks in complex models with limited training data.
- Sparse nature (i.e., non-zero values exist in few grid cells) and abrupt changes at plume boundaries (i.e., shock front) of 3D CO_2 saturation data pose significant challenges to dimension reduction models.
- Existing dimension reduction approaches inadequately address the complexities of 3D saturation data, underscoring the need for a more robust and efficient solution.



Grid cell fractions with non-zero values in 2D and 3D CO_2 saturation data

Research Objectives

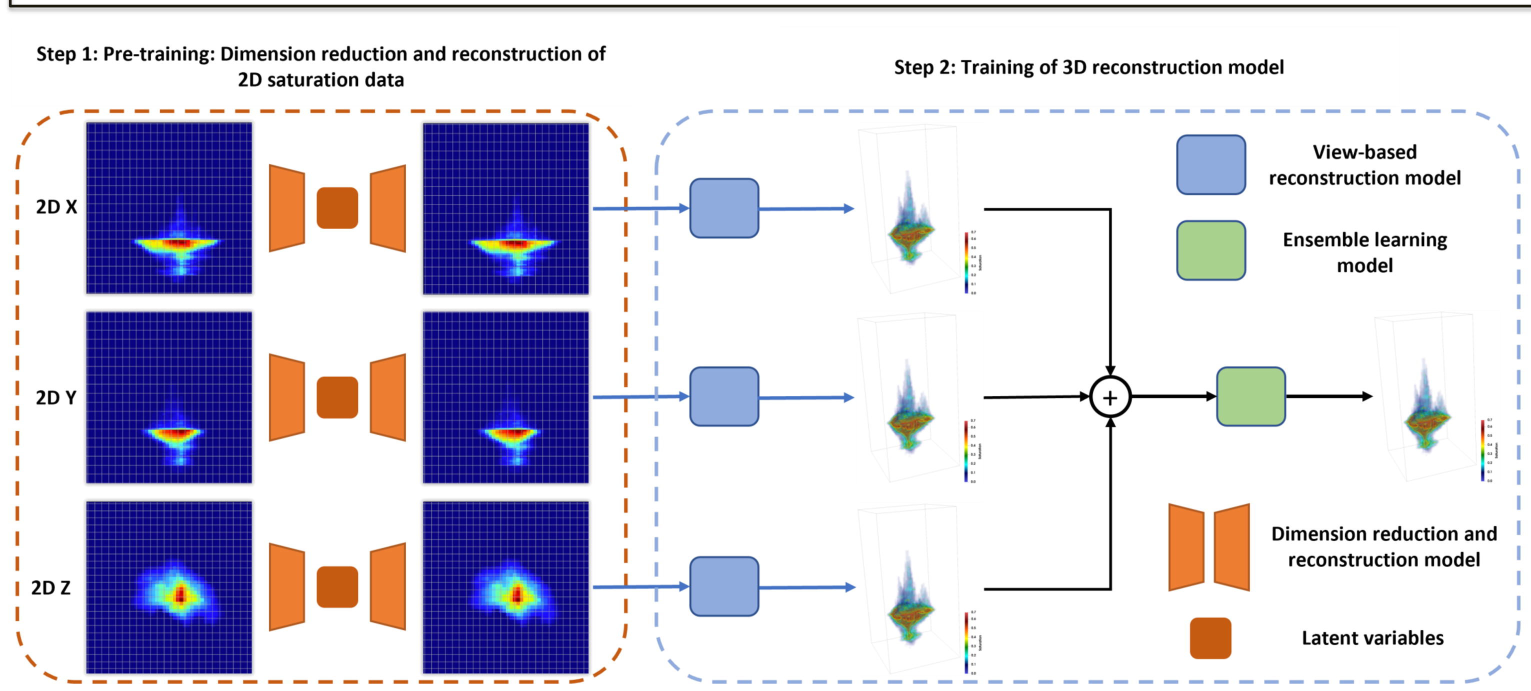
- Design and implement a model to effectively reduce the dimensionality of complex 3D CO_2 saturation data, facilitating more efficient analysis and processing.
- Improve predictive accuracy and model robustness by employing a hybrid loss function that combines data loss with partial differential equation (PDE) loss, aligning model outputs with physical laws governing fluid dynamics.



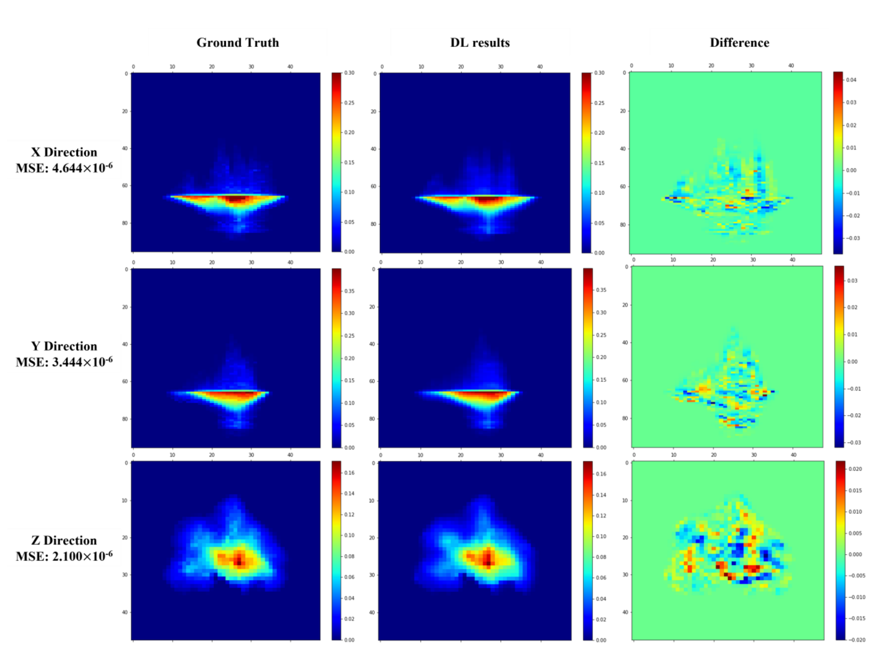
3D CO_2 saturation (Baffle at middle) | 3D CO_2 saturation (No Baffle)

Methodology

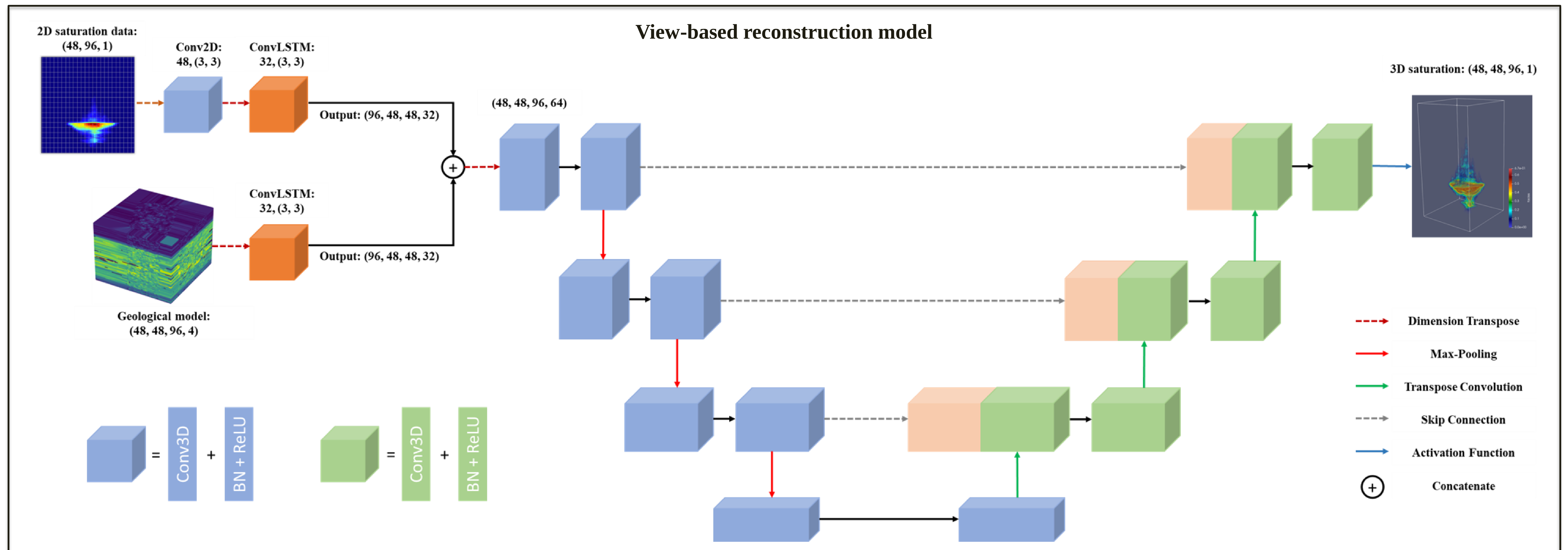
- Step 1: Pre-training of dimension reduction and reconstruction models for 2D saturation data.
  - Obtained by averaging in three directions: X, Y, and Z.
  - More effective non-zero values, making dimension reduction easier.
  - Dimension reduction and reconstruction model in this work: convolutional autoencoders.
  - Latent variable numbers: 128 for X, Y, and Z saturation separately. 384 latent variables in total.
- Step 2: Training of 3D reconstruction model.
  - Inputs: 2D average saturation data in X, Y, and Z directions, and geological data.
  - Output: 3D saturation data.
  - Include three view-based reconstruction model, followed with a ensemble learning model.

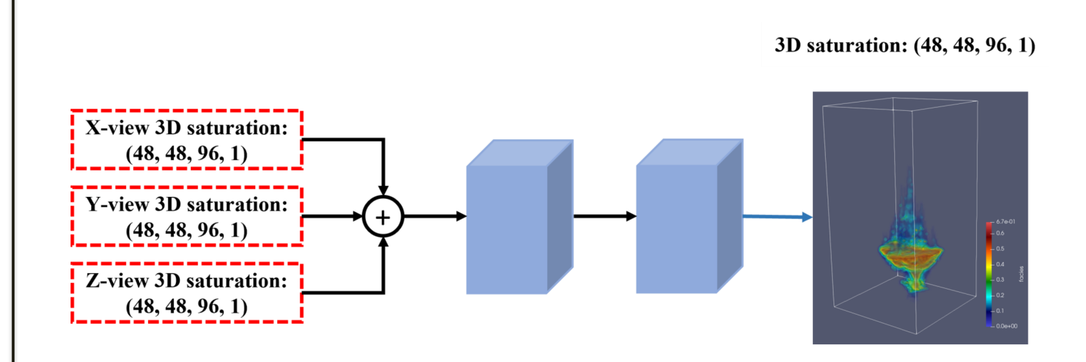


Overall methodology workflow



Dimension reduction and reconstruction for 2D data





Ensemble learning model

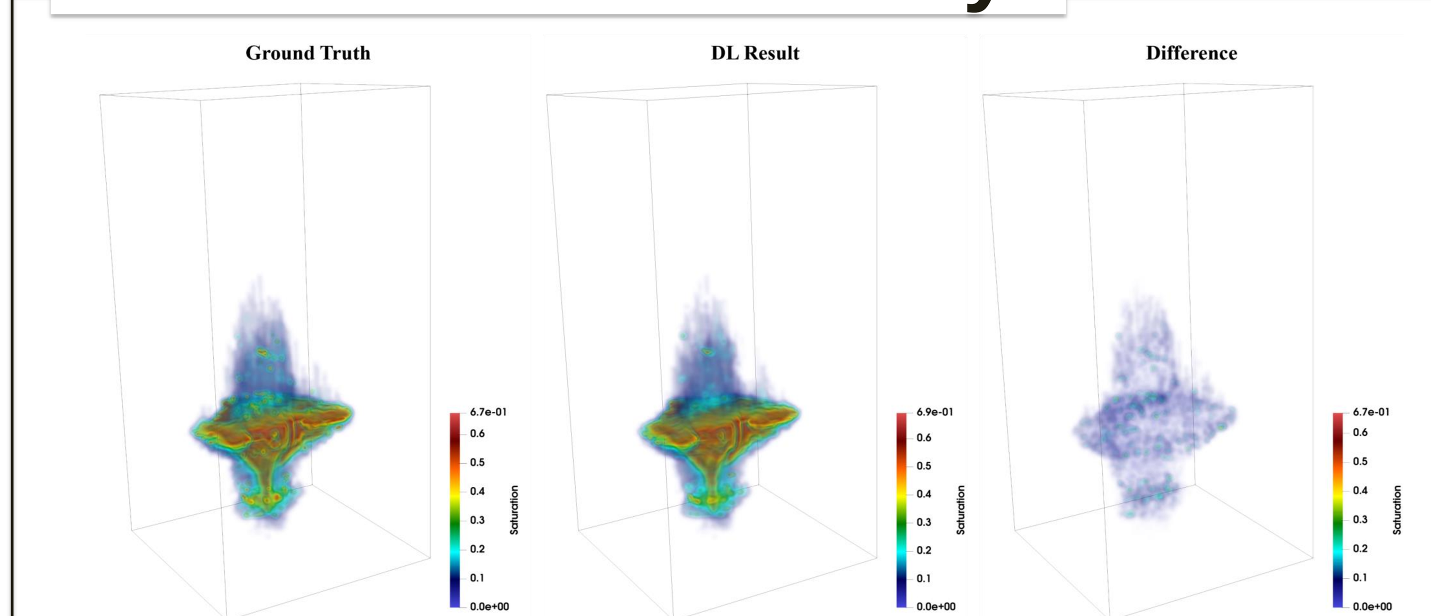
Total Loss = *Reconstruction Loss* + *Masked Loss* + *PDE Loss*

Reconstruction Loss: $-\frac{1}{N}\sum_{i=1}^{N} S_g \log(\widehat{S_g}) + (1 - S_g)\log(1 - \widehat{S_g})$

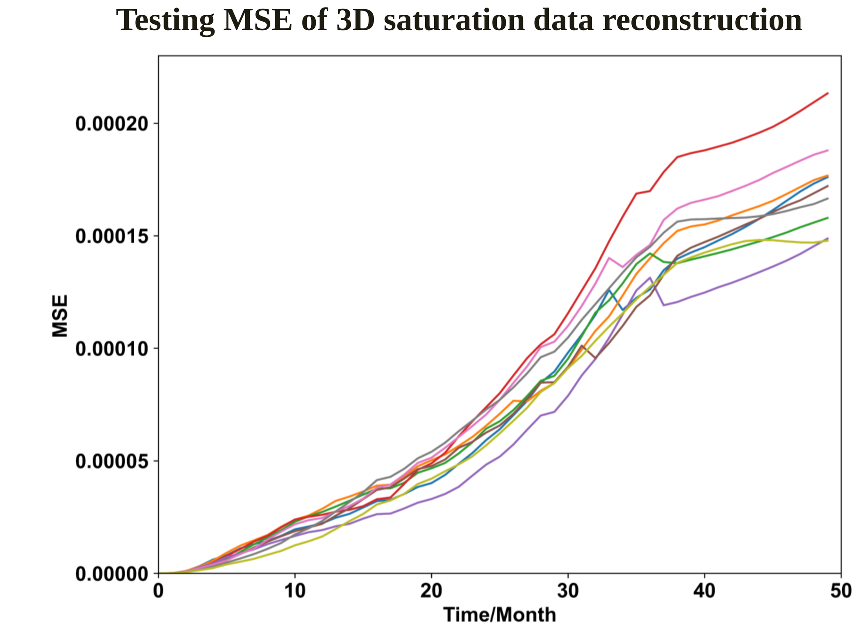
Masked Loss: $\text{Binary Mask} \times (S_g - \widehat{S_g})^2$, $\text{Binary Mask} = (S_g > 0)$

PDE Loss: $(\nabla S_g - \nabla \widehat{S_g})^2$

Results and Summary



Testing MSE of 3D saturation data reconstruction



- Dimension reduction and reconstruction of 2D saturation data are easy to implement and achieve high accuracy. In this work, we used 384 latent variables in total for 2D data in three directions.
- The 3D reconstruction model captures the relationship between 2D inputs and 3D CO_2 saturation, which has great improvement potential by using more strategies, such as residual connection and attention mechanism.
- The proposed workflow can be incorporated into the applications of DL in CCS.

Acknowledgment and Disclaimer

This work is part of a sub-task from the Science-informed Machine Learning to Accelerate Real-Time (SMART) Decisions in Subsurface Applications (edx.netl.doe.gov/SMART). Authors wish to thank the Gulf Coast Carbon Center at UT-Austin BEG for their support.

This project was funded by the United States Department of Energy, National Energy Technology Laboratory, in part, through a site support contract. Neither the United States Government nor any agency thereof, nor any of their employees, nor the support contractor, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.