# Sandia National Laboratories

# Multi-label classification of mineralogy distribution in shale images using machine learning methods

H41G-1754

Jack Ringer[1,2], Hongkyu Yoon[1] (hyoon@sandia.gov)
[1] Geomechanics Department, Sandia National Laboratories, Albuquerque, NM, USA; [2] University of New Mexico, Albuquerque, NM, USA

## Motivations

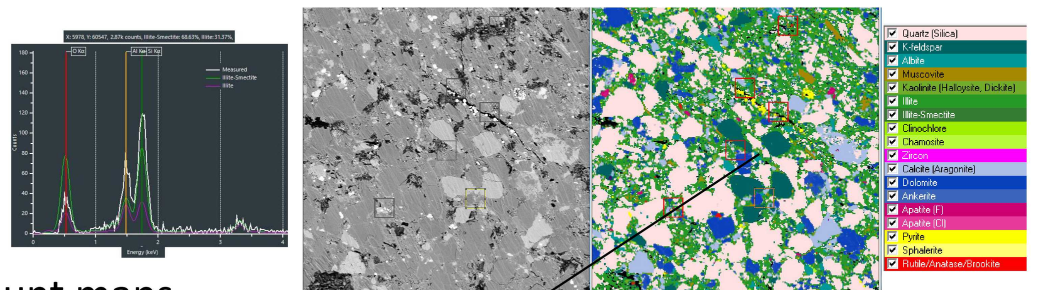
Recent advances in multiscale imaging and elemental analysis techniques have revolutionized our ability to quantitatively characterize geomaterials. Emerging machine learning (ML) methods will enable us to analyze these big data more accurately and computationally efficiently so that we can get more insightful understanding of interactions of complex geometries and minerals.

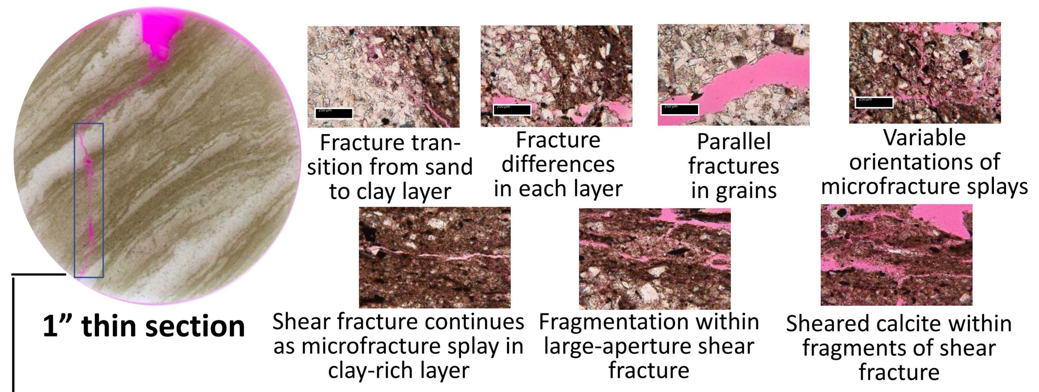
## Objectives

- Present advanced imaging and elemental analysis methods for physical and chemical characterization of shale samples
- Multicalss segmentation for major phases in shale
- Mullticlass and multilabel classification for mineralogy classification using data based on MAPS (Modular Automated Processing System) Mineralogy

## MAPS Mineralogy

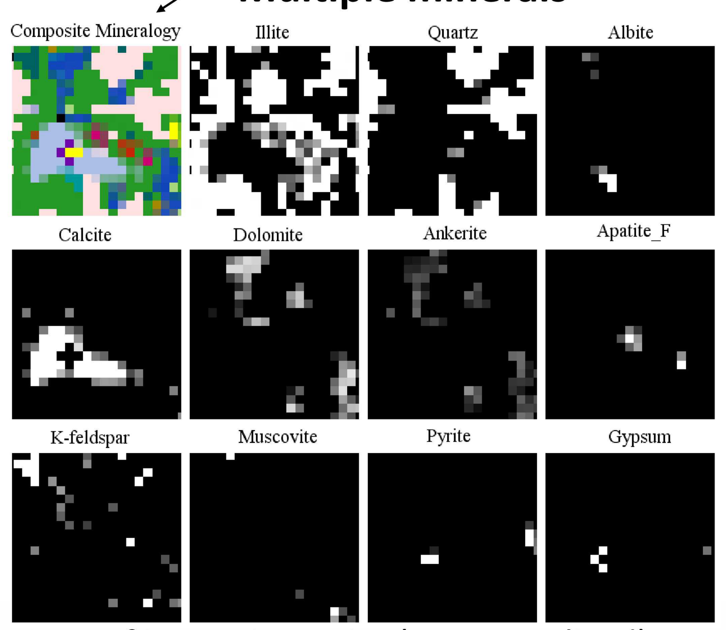
- **SEM-based automated mineralogical measurement, analysis, data integration**
  - Collection, overlay and re-registration of multiple images from different modalities
  - SEM, SEM-EDS, optical, CL, EBSD
  - QEMSCAN measurement algorithms
- **Mineral identification**
  - Spectral matching
  - Each pixel – single/multiple minerals
  - Simultaneous mineral element and count maps





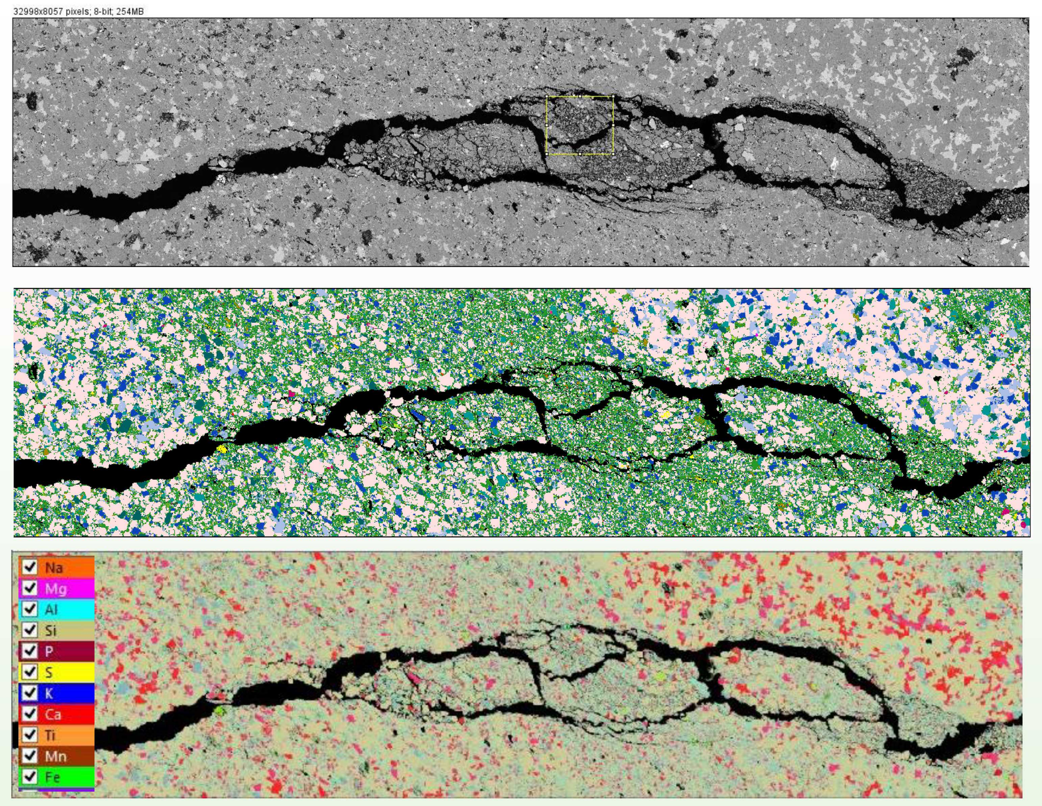

Crack in Mancos shale after BT (Na and Yoon et al. JGR2017)

**Multiple minerals**

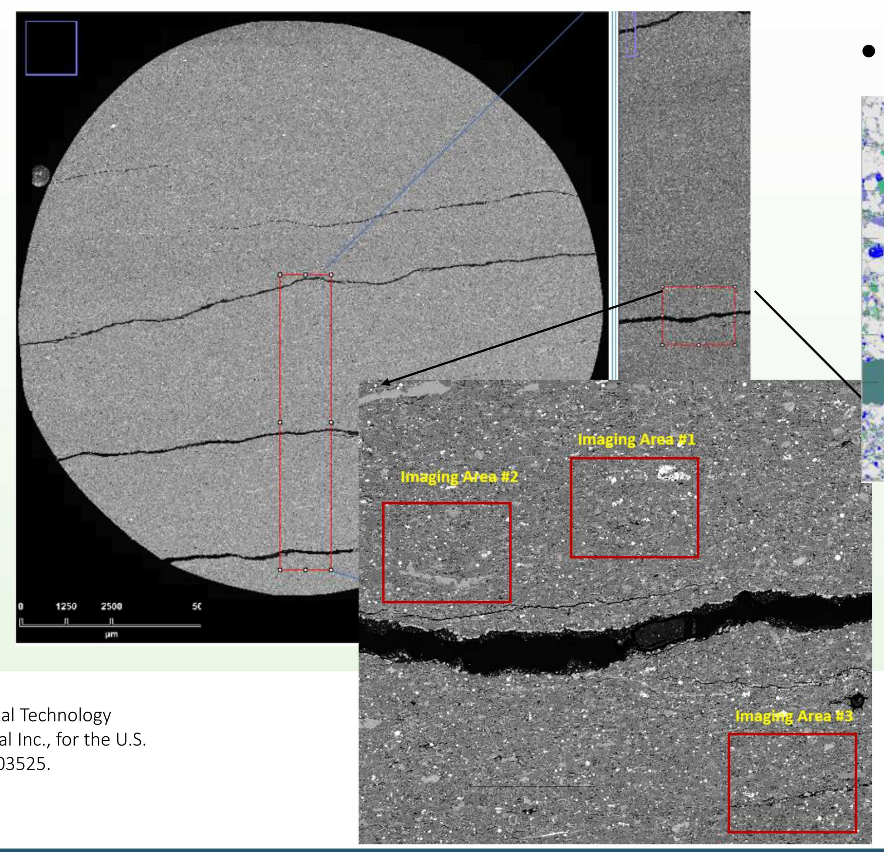



Sum of image intensity (wt normalized) =1

Focus Area: Back Scattered Electron image, MAPS mineralogy distribution (composite color), and elemental mapping
[32999 x 8057 pixels] [>20 minerals] [>20 elements]

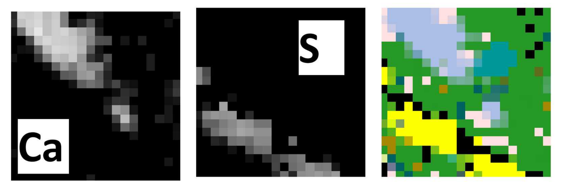


Marcellus shale sample [16645x17733]

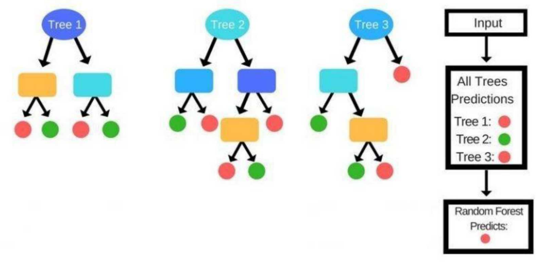


## Multi-class classification

- Setting when there are > 2 possible class labels
- Classifiers: Gradient Boosting, Decision Tree, Random Forest, Linear SVM, Logistic Regression, Naïve Bayes, Nearest Neighbors [Scikit-learn]
- Input: element fractions per pixel
- Output: single mineral class per pixel



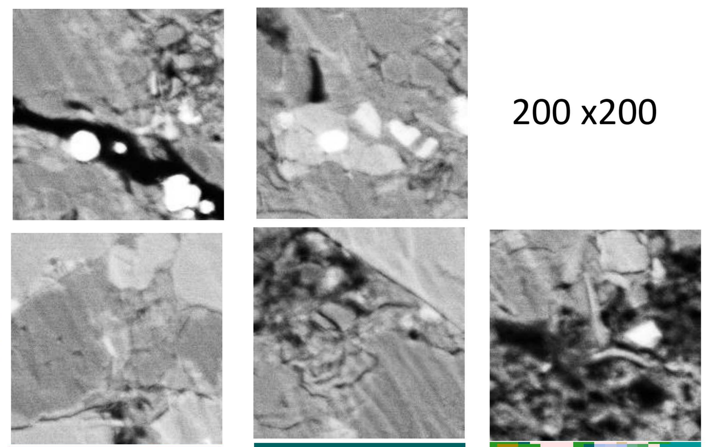

**Schematic of Random Forest**




- Inherently deals with multiclass datasets
- Robustness & taking care of overfitting
- High accuracy
- Slow in generating prediction
- Difficult to interpret

**Comparison of methods (1 training images and five testing images)**

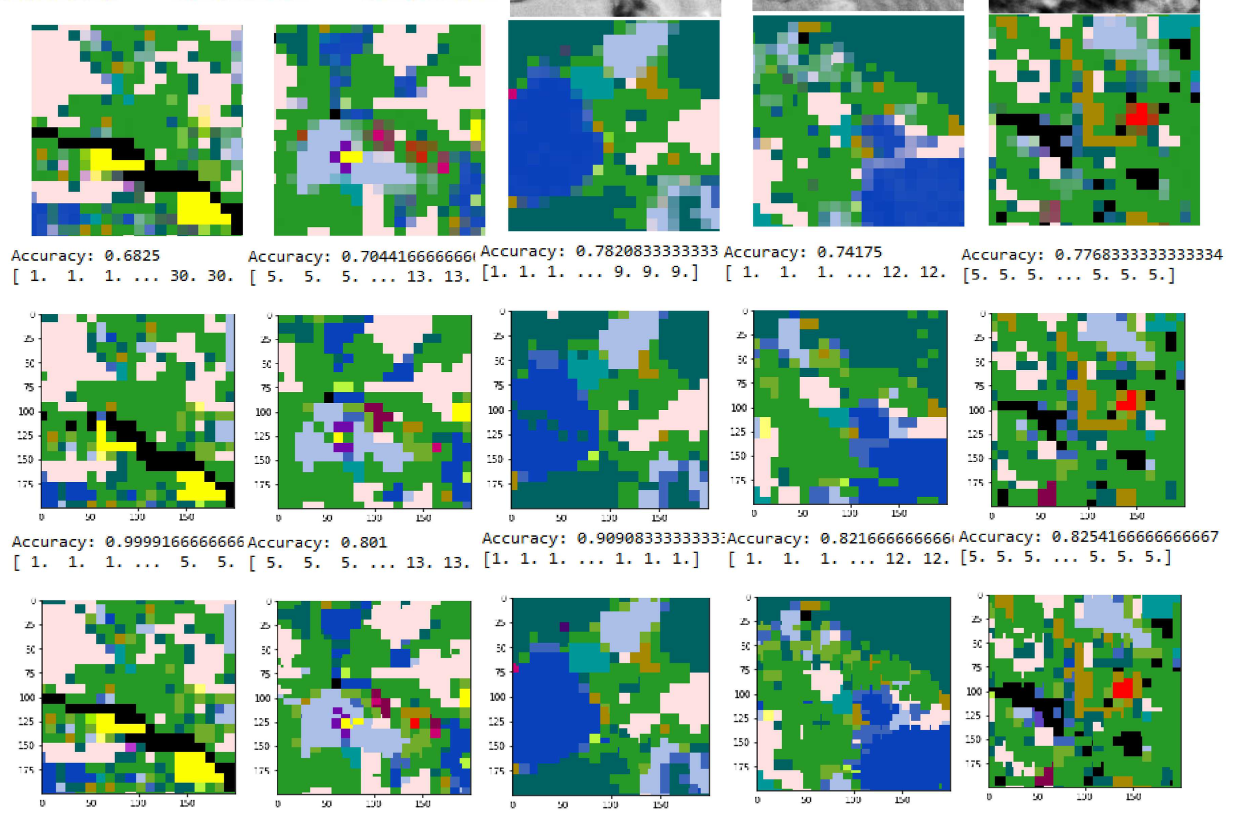
| classifier | train_score | test_score | train_time |
|---|---|---|---|
| Gradient Boosting Classifier | 1.000000 | 0.808772 | 148.355674 |
| Random Forest | 1.000000 | 0.739833 | 7.057065 |
| Naive Bayes | 1.000000 | 0.808562 | 34.234749 |
| Decision Tree | 1.000000 | 0.700906 | 0.024480 |
| Nearest Neighbors | 0.999501 | 0.700432 | 0.460318 |
| Linear SVM | 0.991098 | 0.681583 | 0.014426 |
| Neural Net | 0.980060 | 0.700432 | 8.122163 |
| Logistic Regression | 0.707520 | 0.680083 | 1.602031 |




**Mineral composite images with BSE images**

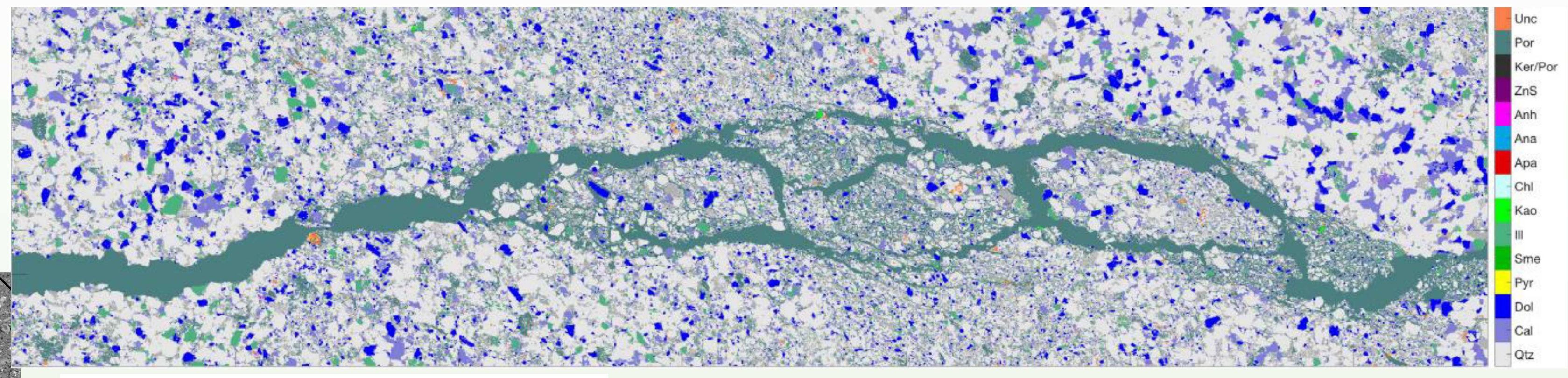
**Testing results with a trained model using 1 image (RF)**

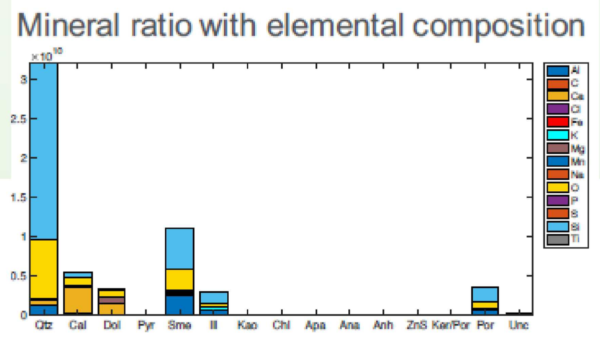
**Testing results with a trained model using 2 images (RF)**



## Multi-label classification [Li, 2018]

- Multiple attributes to an image [15 minerals are the list of labels per pixel]
- Convolutional neural network-based multi-label classification
- Input data: 15 elemental maps, elemental distribution ratio, BSE image
- Output: multiple minerals per pixel






Results from Weichang Li @ARAMCO

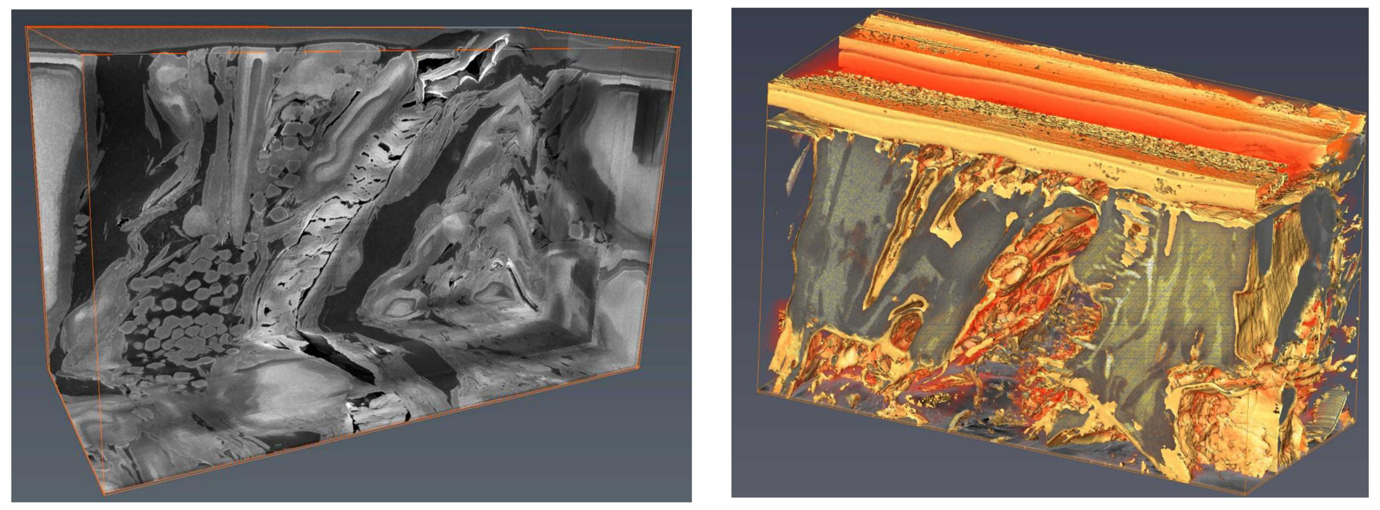
**Key references**


- Weichang Li, Robust Learning with Noisy Labels in Multispectral Source Rock Characterization, Machine learning in Solid Earth Geosciences, Sanfa Fe, Feb 20-22, 2018, Page 18.
- Long, J., Shelhamer, E. and Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).
- Badrinarayanan, V., Kendall, A. and Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12), pp.2481-2495.

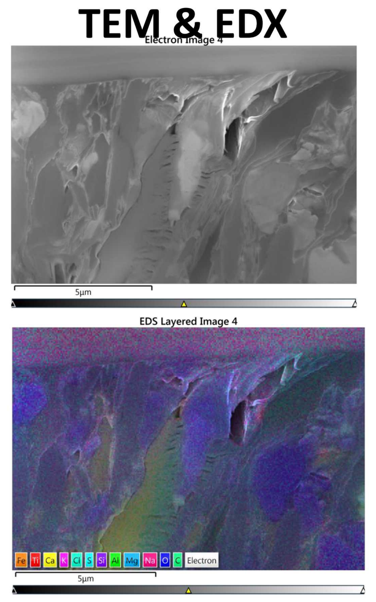

## Image Segmentation

- Pixel-based segmentation for three phase (organic, pore, and solid phases)
- Focused Ion Beam-SEM (FIB-SEM) and Transmission electron microscopy (TEM) images with BSE+MAPS mineralogy at larger scales and EDX mapping data
- Training (120 images) and validation (30 images)

**3D view of FIB-SEM images of Marcellus shale**



**TEM & EDX**

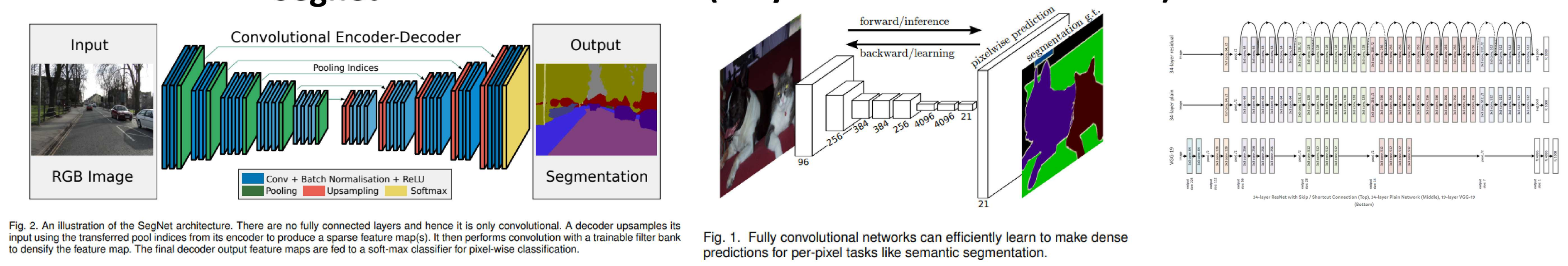


- Deep image segmentation models (https://github.com/divamgupta/image-segmentation-keras)

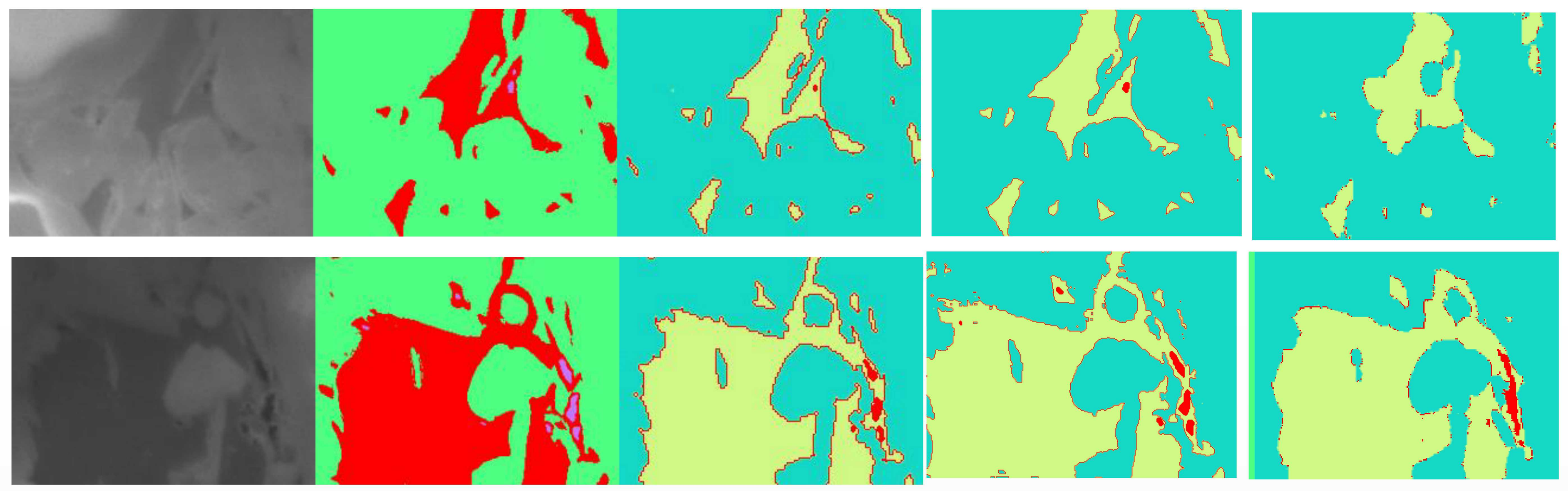
| | Model | | Training data | | | Validation data | | |
|---|---|---|---|---|---|---|---|---|
| | Base Model | Segmentation model | Solid | Organic | Pore | Solid | Organic | Pore |
| Case 1 | Resnet | Segnet | 0.9555 | 0.935 | 0.782 | 0.9276 | 0.89 | 0.634 |
| Case 2 | Vanilla CNN | Segnet | 0.9714 | 0.956 | 0.68 | 0.9397 | 0.906 | 0.5 |
| Case 3 | Vanilla CNN | FCN32 | 0.8957 | 0.8435 | 0.3326 | 0.8921 | 0.8298 | 0.2714 |

Loss

**Segnet** **FCNs (fully convolutional networks)** **ResNet**




**Original Image** **Truth** **Case 1** **Case 2** **Case 3**



## Summary

- Advanced imaging and analysis methods can generate "big data" of geochemical related information which can provide information-rich data for ML applications.
- Multiclass and multilable classification can be utilized for accurate mineralogy classification due to high volume of training data. We expect the classification accuracy will increase dramatically with better algorithms and various training data.
- Image segmentation of shale samples can be doable using existing training models available and transfer learning can enhance segmentation accuracy dramatically. Conventional segmentation requires manual and tedious local operations

**Related poster: H51I-1595 (Friday morning) for DCGAN vs SAGAN for image reconstruction**



This work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories.