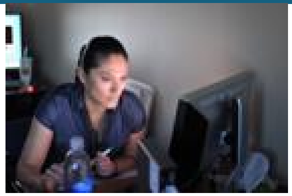
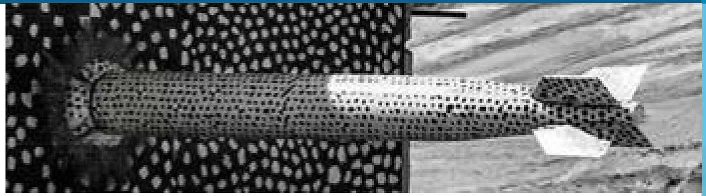
SAND2022-6838C

Semantic segmentation of rock images and ensemble approach for deep learning methods





Jack Ringer* and Hongkyu Yoon

Geomechanics Department

Sandia National Laboratories

Albuquerque, NM, USA

* University of New Mexico Undergraduate student

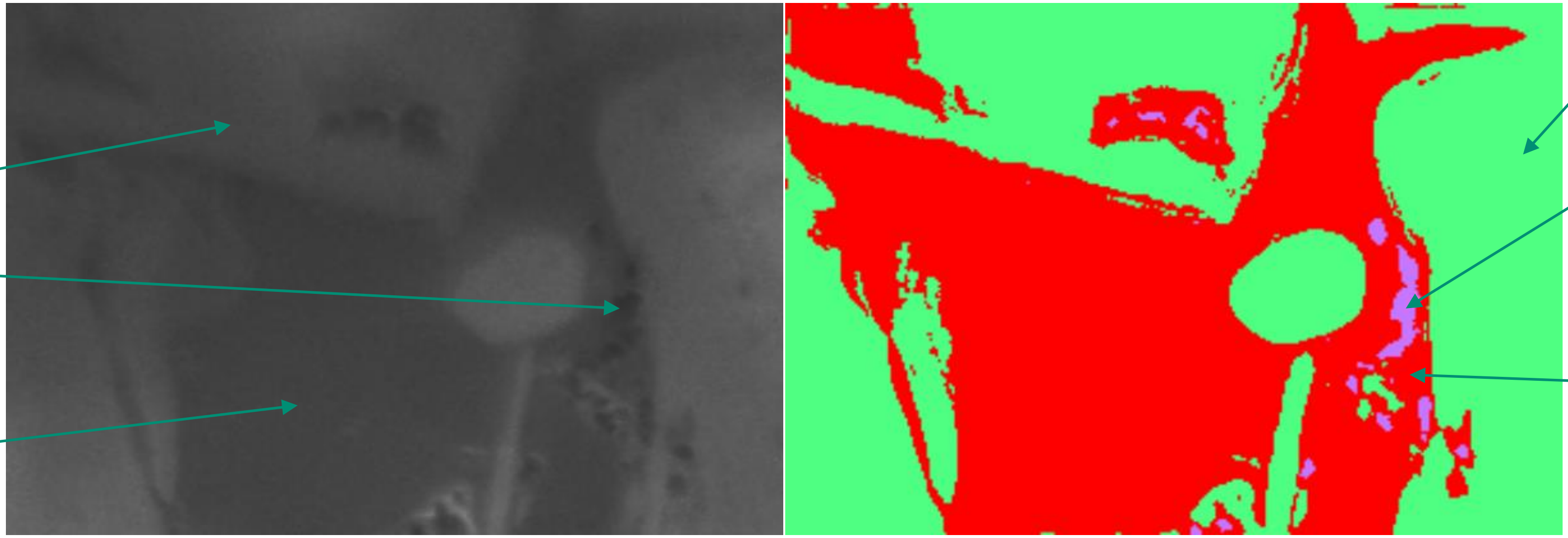
InterPore 2022

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

Background: What is Semantic Segmentation?

- Semantic segmentation aims to classify images at the pixel-level
- This is done by assigning a value to each pixel based upon its class

One pixel =
10 nm resolution



Focused Ion Beam-Scanning Electron Microscopy image of
Marcellus Sandstone (left) and manually segmented image (right)

- Generating accurate segmentations of rock images is critical to geomaterial characterizations
- Challenges: Complexity in geometry, size, scale, and compositions

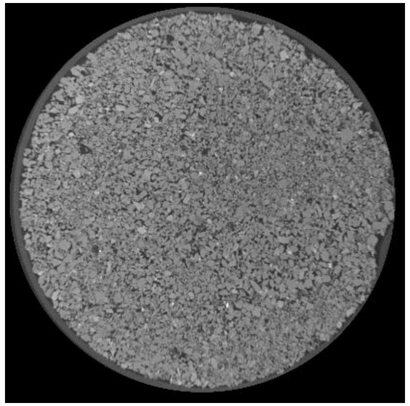
Image Data

- Original datasets include 3D microCT and FIB-SEM images
- Original images have been segmented with various traditional methods (e.g., Yoon and Dewers, 2013 GRL for S-Chalk)
- 128x128 and 128x128xD images are used for 2D and 3D models (D= # of images in depth)
- Dataset is split randomly into training (70%), validation (15%), and testing (15%)

Dataset	Num. of Training Images	Num. of Validation Images	Num. of Testing Images
Sandstone	27173	5869	5870
Carbonate Chalk	4812	1050	1050
Shale*	1593	231	231
Liege Chalk	11827	2534	2535

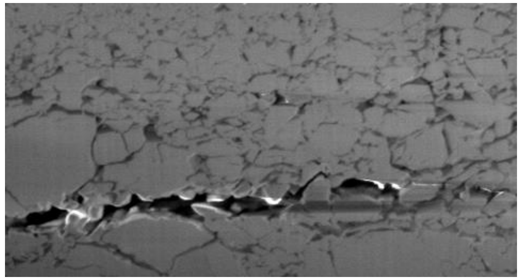
*Shale training data was augmented to increase # of pore samples

Boise Sandstone



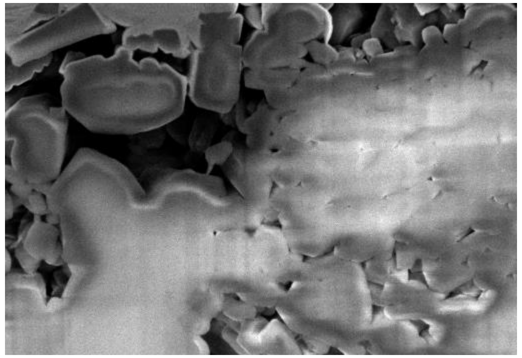
MicroCT images at 30 micron resolution (1500x1500x1800)

Carbonate Chalk (S-Chalk)



FIB-SEM images at 15 nm resolution (932x620x930)

Carbonate Chalk (L-Chalk)



FIB-SEM images at 10 nm resolution (900x700x900)

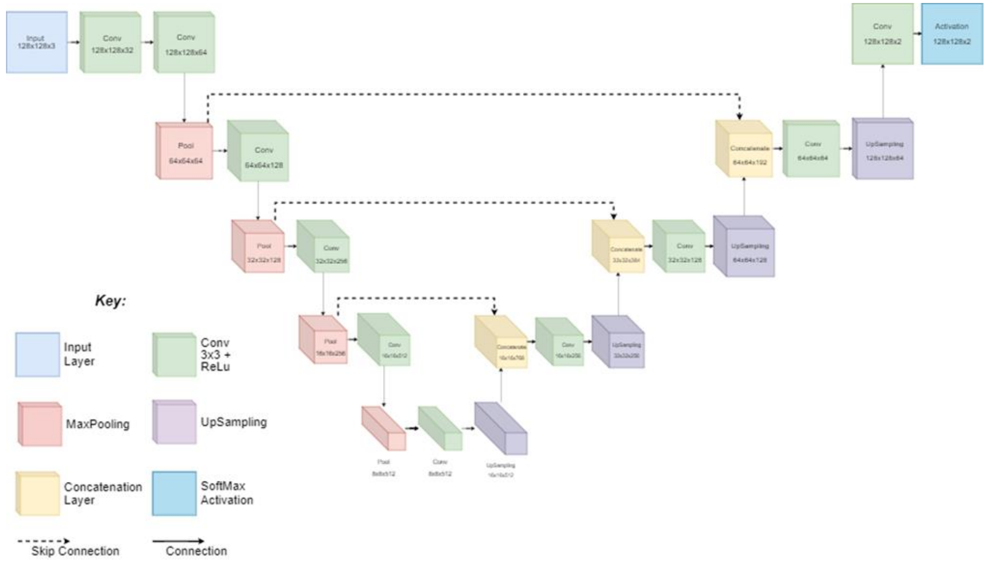
Marcellus Shale

FIB-SEM images at 10 nm resolution (900x700x900)

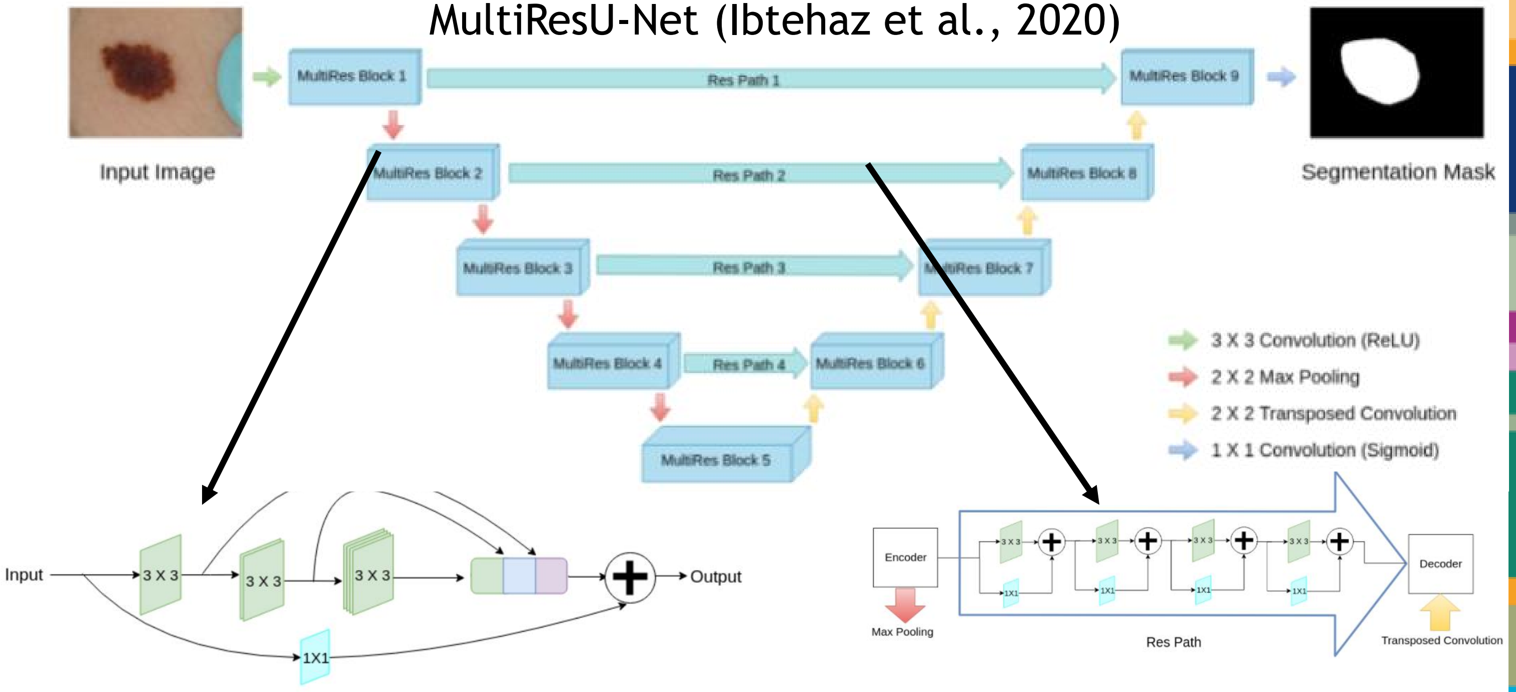
Methods: Model Architectures

- Models are U-Net based architectures (U-Net 2D/3D, U-VGG16, U-ResNet, MultiResUnet)
- U-Net follows "U" shape of convolutional neural network architecture with a feature of skip connection & MultiResU-Net accounts for multi-scale features.
- All models follow an encoder-decoder architecture
  - Encoder extracts feature maps from input image
  - Decoder transforms these feature maps into a prediction

U-Net (Ronneberger et al., 2015)



MultiResU-Net (Ibtehaz et al., 2020)

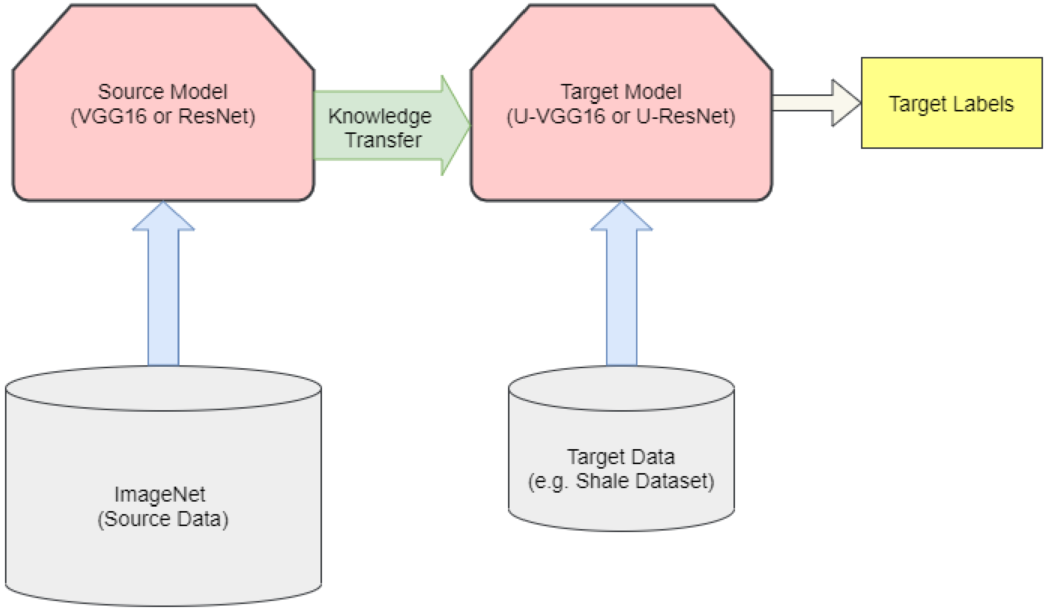


Methods: Hyperparameters & Transfer Learning

- Adam optimizer with a learning rate of 0.001
- Loss: categorical focal loss
- Batch size: 196 for U-Net and U-VGG16, 128 for U-Resnet, 64 for MultiResUnet
- Early stopping: 100 epochs

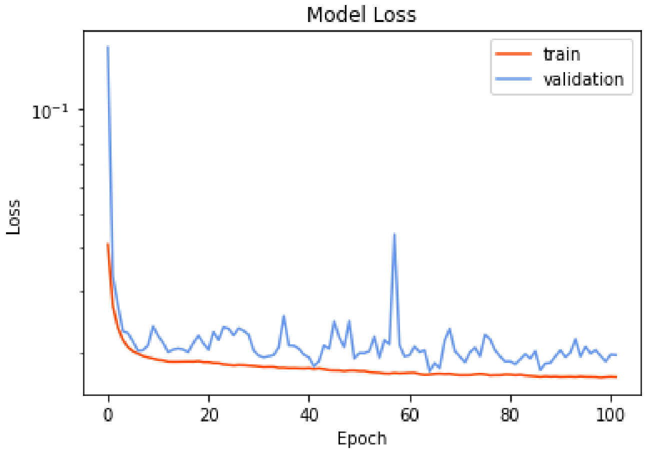
Transfer Learning Approach

- Method where model originally trained on one task leveraged for another
- Useful because it can allow for training to converge at a faster rate and lead to more robust models
- Investigated by initializing VGG16 and ResNet models with weights from ImageNet and fine tuning of decoder portion or all parameters



Base Results (test data)

- “F-IoU” refers to frequency IoU
- “Pixel Acc” refers to the pixelwise accur

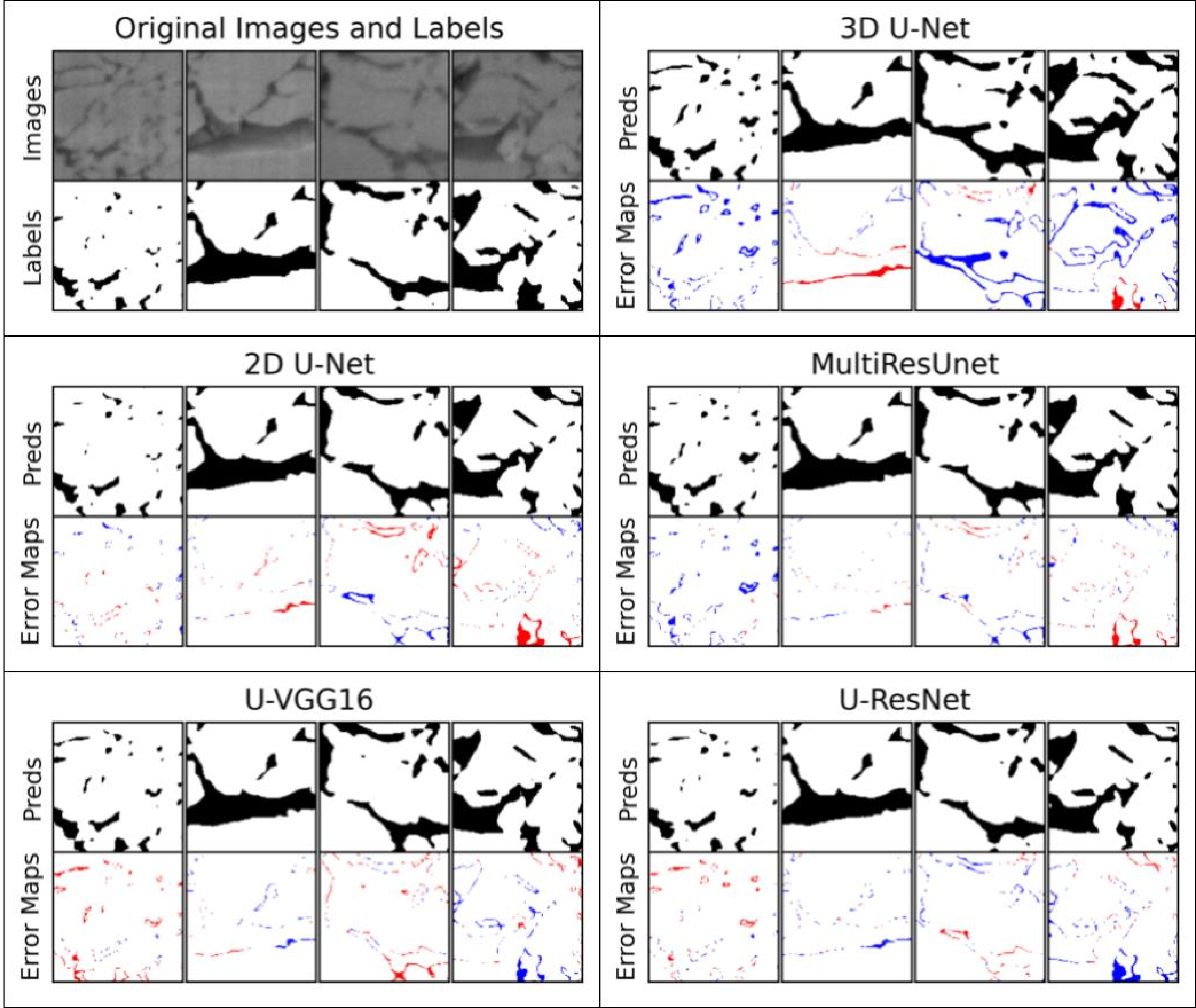


Example training plot showing validation and training loss as MultiResUnet is trained on Sandstone data

	Sandstone		S-chalk		L-chalk		Shale	
	F-IoU	*Pixel Acc*	*F-IoU*	*Pixel Acc*	*F-IoU*	*Pixel Acc*	*F-IoU*	*Pixel Acc*
U-Net	0.9522	0.9753	0.9514	0.9748	0.9219	0.9573	0.9332	0.9653
U-Net-3D	0.9270	0.9623	0.9006	0.9457	0.8752	0.9243	0.5989	0.7480
U-VGG16	0.9687	0.9840	0.9519	0.9749	0.9250	0.9572	0.7966	0.8863
U-ResNet	**0.9829**	**0.9913**	0.9397	0.9687	**0.9476**	**0.9719**	0.8948	0.9444
MultiResUnet	**0.9826**	**0.9912**	**0.9601**	**0.9794**	**0.9492**	**0.9730**	**0.9444**	**0.9713**

Predictions for Selma Chalk Data (Only Base Models Shown)

Sample images, predictions, and error maps from the testing split of the Selma chalk dataset. On the error maps, white indicates label and prediction agree, blue indicates pore was predicted but expected solid, red indicates solid was predicted but expected pore.



Transfer Learning Results (test data)

- "Fine-tune" refers to the revision of weights from ImageNet as training progresses.
- Numbers below given values indicate a comparison to the base case (e.g., (+0.1) indicates the metric improved by 0.1)

	Sandstone		S-chalk		L-chalk		Shale	
	F-IoU	*Pixel Acc*	*F-IoU*	*Pixel Acc*	*F-IoU*	*Pixel Acc*	*F-IoU*	*Pixel Acc*
U-VGG16-fine tune	0.9617 (-0.0070)	0.9805 (-0.0035)	0.9580 (+0.0061)	0.9784 (+0.0036)	0.9347 (+0.0097)	0.9655 (+0.0083)	0.6839 (-0.1127)	0.8121 (-0.0742)
U-Resnet-fine tune	0.9732 (-0.0097)	0.9863 (-0.0050)	0.9525 (+0.0128)	0.9751 (+0.0064)	0.9459 (-0.0017)	0.9709 (-0.0017)	0.9363 (+0.0415)	0.9671 (+0.0227)

MultiResUnet	**0.9826**	**0.9912**	**0.9601**	**0.9794**	**0.9492**	**0.9730**	**0.9444**	**0.9713**

Methods: Ensemble Approach

- Rather than generating predictions from a single model, multiple models are trained and then their predictions are combined
- In theory can reduce the variance of models and lead to better predictions
- In our implementation:
  - Model trained for 200 epochs, saving weights every 5 epochs
  - Best 3 models (based on validation sets) from 200 epochs are used to make ensemble predictions (i.e., average prediction)
  - Also tested special cases where models are drawn from a certain period of epochs in training (e.g. 3 models only taken from epochs 100-150)

Ensemble Results (Special Cases)

- Models were trained for 300 epochs
- Each respective case comes from a different epoch range:
  - C1: [100-300], C2: [100-200], C3: [250-300]
- In table case which achieved highest F-IoU is shown (indicated by C1, C2, or C3 below values)

	Sandstone		S-chalk		L-chalk		Shale	
	F-IoU	*Pixel Acc.*	*F-IoU*	*Pixel Acc.*	*F-IoU*	*Pixel Acc.*	*F-IoU*	*Pixel Acc.*
U-Net ensemble	0.9768 (C1)	0.9882	0.9569 (C1)	0.9775	0.9512 (C2)	0.9741	0.9497 (C3)	0.9741
U-Resnet ensemble	0.9878 (C2)	0.9938	0.9621 (C1)	0.9804	0.9568 (C1)	0.9773	0.9522 (C3)	0.9754
MultiResU-Net ensemble	0.9840 (C3)	0.9920	0.9646 (C1)	0.9818	0.9575 (C1)	0.9776	0.9503 (C1)	0.9745

Conclusions

- Deep learning architectures can successfully be applied to the task of semantic segmentation for rock images and can perform better than manual segmentation to recover natural morphology of original images
- Ensemble approach consistently improved performance
- Use of transfer learning improved accuracy and training speed in most cases
- 3D model underperformed
  - May be due to lack of training data and enough training (a small # of epochs in this work)
  - Complications can arise when using depth data, such as unpredictable variations in illumination between images
- Ensemble approach with hyperparameter tuned (results not shown) tend to improve performance in all cases
- Data labeling and curation will be explored to improve supervise learning process