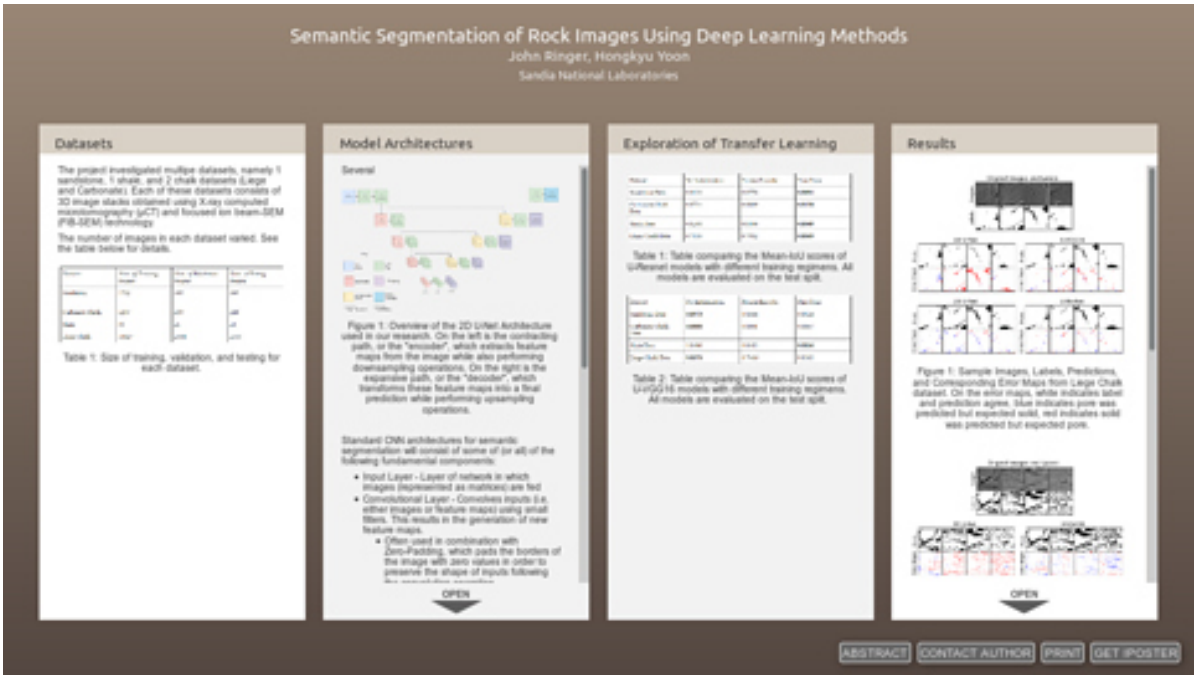
SAND2020-13144C

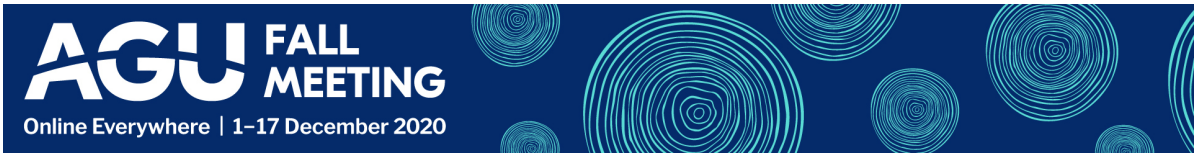
Semantic Segmentation of Rock Images Using Deep Learning Methods



John Ringer, Hongkyu Yoon

Geomechanics Department, Sandia National Laboratories, Albuquerque, NM

PRESENTED AT:



OVERVIEW

Motivations

- Semantic Segmentation is to label the specific regions of an image.
- Generating accurate segmentations of rock images is critical to accurately describing geomaterials.
- Conventional segmentation methods (e.g. watershed segmentation and thresholding) are labor intensive, often subject to user-bias, and can fail to adapt to certain datasets.

Objective

- Evaluate the capability of various convolutional neural network (CNN) architectures to segment rock images
- Establish an effective and reliable way for automatic segmentation via transfer learning

Datasets

- Two three dimensional (3D) rock images obtained using X-ray computed microtomography (μCT) and focused ion beam-SEM (FIB-SEM)
- 128x128 images are sampled from a full stack of 3D images to train and evaluate models (see Table 1)
- For sandstone (from CT images) a large number of images were used. For shale (from FIB-SEM image) a small number of samples were used. All data have been segmented (i.e., labelled) using traditional segmentation methods

Table 1. Number of training, validation, and testing images for the sandstone and shale datasets.

Dataset	*Num. of Training Images*	*Num. of Validation Images*	*Num. of Testing Images*
Sandstone	1702	365	365
Shale	95	21	21

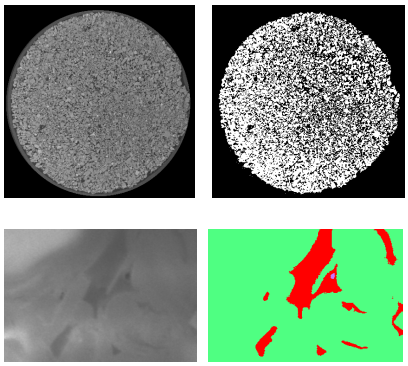


Figure 1. Examples of two rock original and segmented (i.e., labelled) image: (top) Boise sandstone and (bottom) Marcellus shale

MACHINE LEARNING METHODS

Model Architectures

Four different CNN architectures: three 2D and one 3D architectures.

- Each of these CNNs followes an encoder-decoder design.
- The encoder serves to extract feature maps from an input image whereas the decoder transforms these condensed feature maps into a single corresponding prediction map.

Each of three 2D architectures used the same decoder as the standard U-Net, but differed in the encoder as followings:

- The 2D U-Net follows a standard U-Net architecture (Figure 2).
- The U-VGG16 uses a VGG16 architecture.
- The U-ResNet uses a ResNet architecture.

The 3D model (referred to as 3D U-Net) follows an identical architecture to the 2D U-Net with 3D conv and pooling layers

- This allows the 3D model to use 3D image volumes as input.

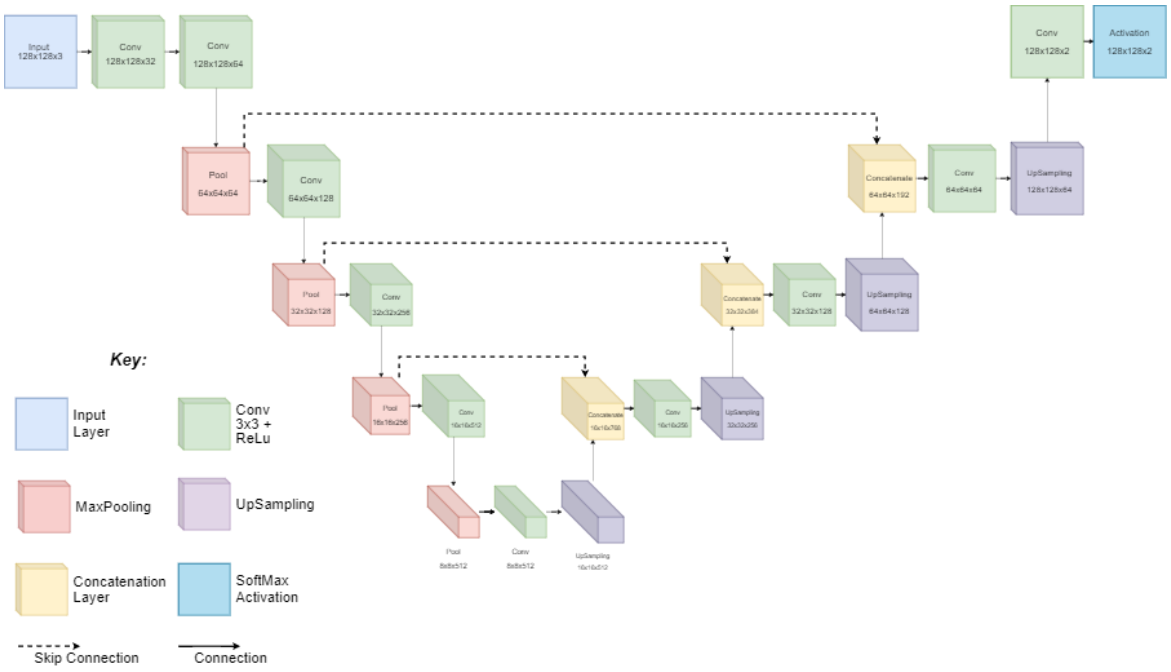


Figure 2. Overview of the 2D U-Net Architecture.

Key components:

- Input Layer (128x128x3 to utilize the pretrained model of the VGG16 and ResNet)
- Convolutional Layer
- Relu activation function
- Max pooling
- Upsampling
- BatchNormalization
- Skip-Connections
- SoftMax Activation

Use of Transfer Learning

Transfer learning is a process that utilize a model trained with different data for another data to take advantage of weigths and biases optimized:

- Improve training times
- Enhance appliability of pre-trained model for different datasets
- Improve model performance

To explore the use of transfer learning with our rock images, experiments were performed with the U-VGG16 and U-ResNet networks

- The encoder portion is initialized using weights from pretrained models (VGG16 & ResNet) with ImageNet; the decoder uses a random initialization.

Three different transfer learning approaches:

- ***From-scratch*** models used a random parameter initialization.
- ***Frozen encoder*** refers to models whose encoder is initialized and fixed with parameters from pretrained mdoels with ImageNet, but the parameters in the decoder are updated during training.
- ***Fine-tuned*** refers to models whose encoder is initialized with parameters from ImageNet and afterwhich is trained end-to-end (i.e. all parameters are updated).

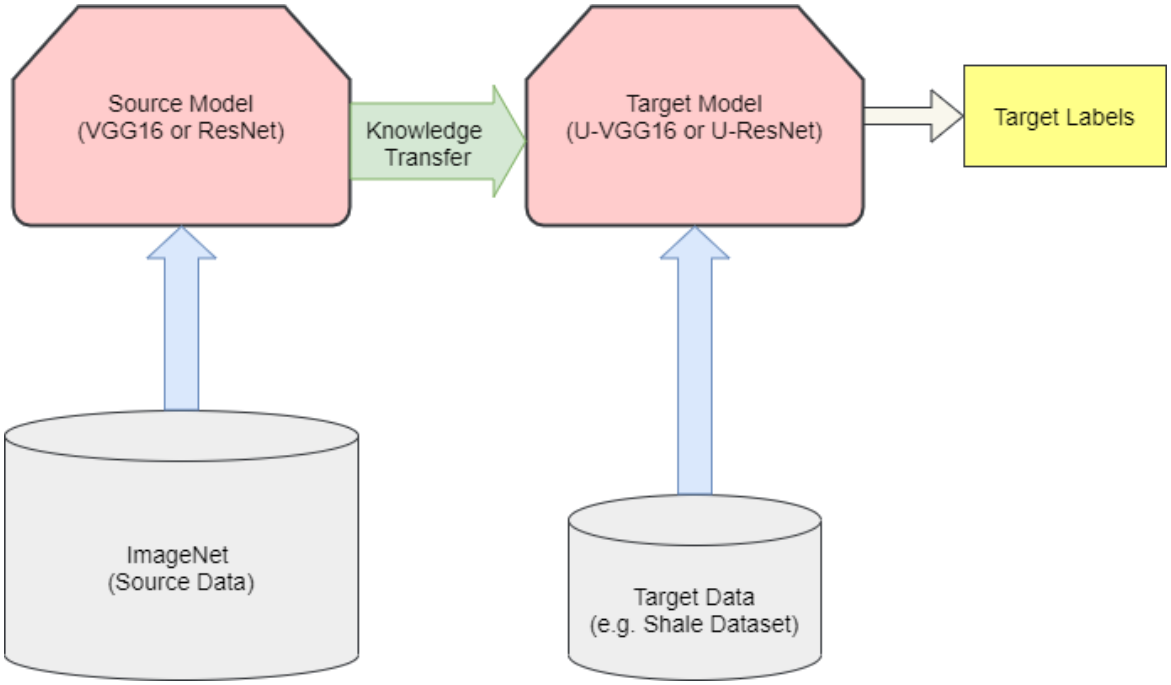


Figure 3. Overview of the transfer learning process. During the "Knowledge Transfer" stage parameters are transfered from a VGG16 (or ResNet) model trained on ImageNet to the encoder of a U-VGG16 (or U-ResNet).

RESULTS

Comparison of different CNN architectures

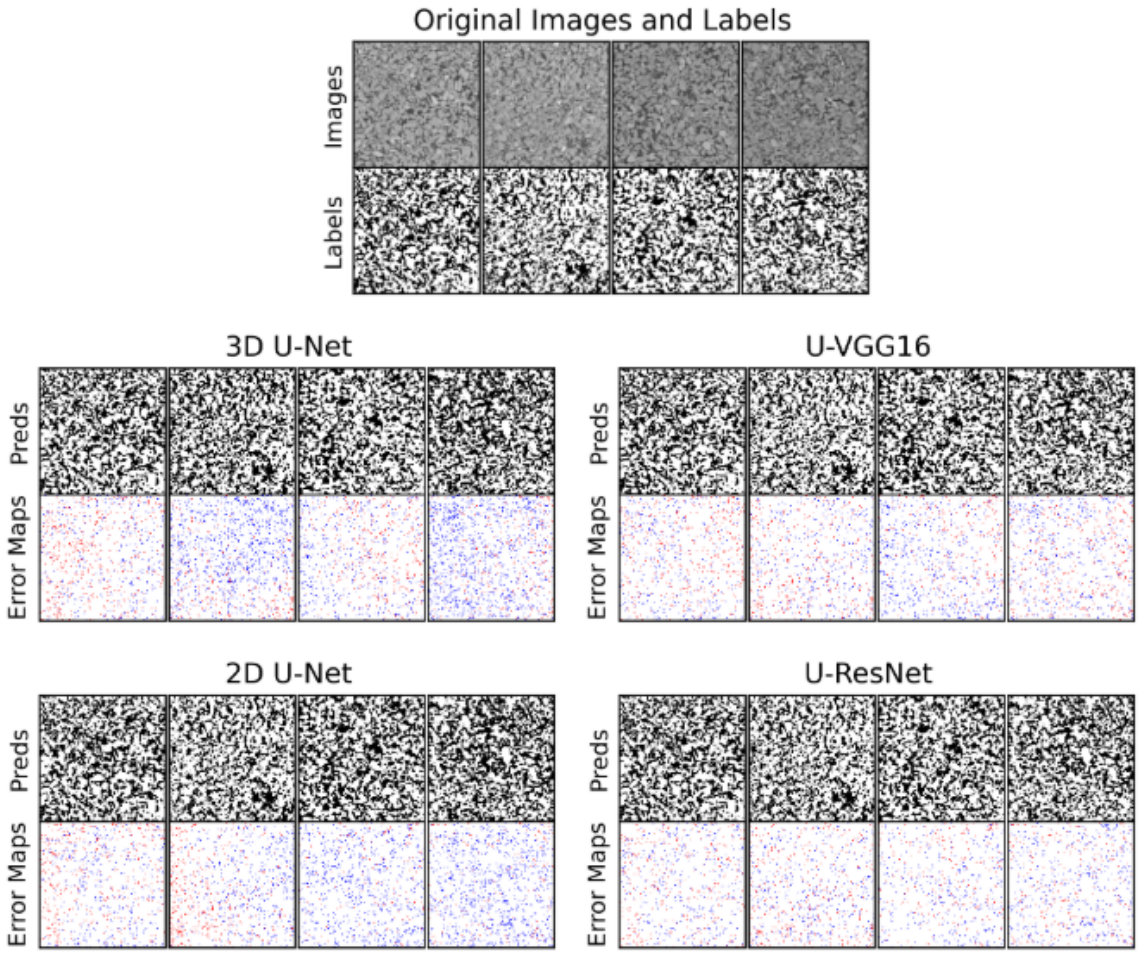


Figure 4. Sample images, labels, predictions, and error maps from the testing split of the sandstone dataset. On the error maps, white indicates label and prediction agree, blue indicates pore was predicted but expected solid, red indicates solid was predicted but expected pore.

Table 2. Comparison of trained 3D-Unet, the U-Net, U-VGG16, and U-ResNet models after evaluation on ths Sandstone test data.

Model	*Freq. IoU*	*Mean IoU*	*Pixel-Wise Acc.*	*Solid IoU*	*Pore IoU*
3D U-Net	0.8707	0.8652	0.9306	0.8913	0.8391
U-Net	0.8886	0.8833	0.9408	0.9074	0.8593
U-VGG16 (fine-tune)	0.8975	0.8924	0.9460	0.9159	0.8689
U-ResNet (fine-tune)	**0.9131**	**0.9091**	**0.9544**	**0.9275**	**0.8907**

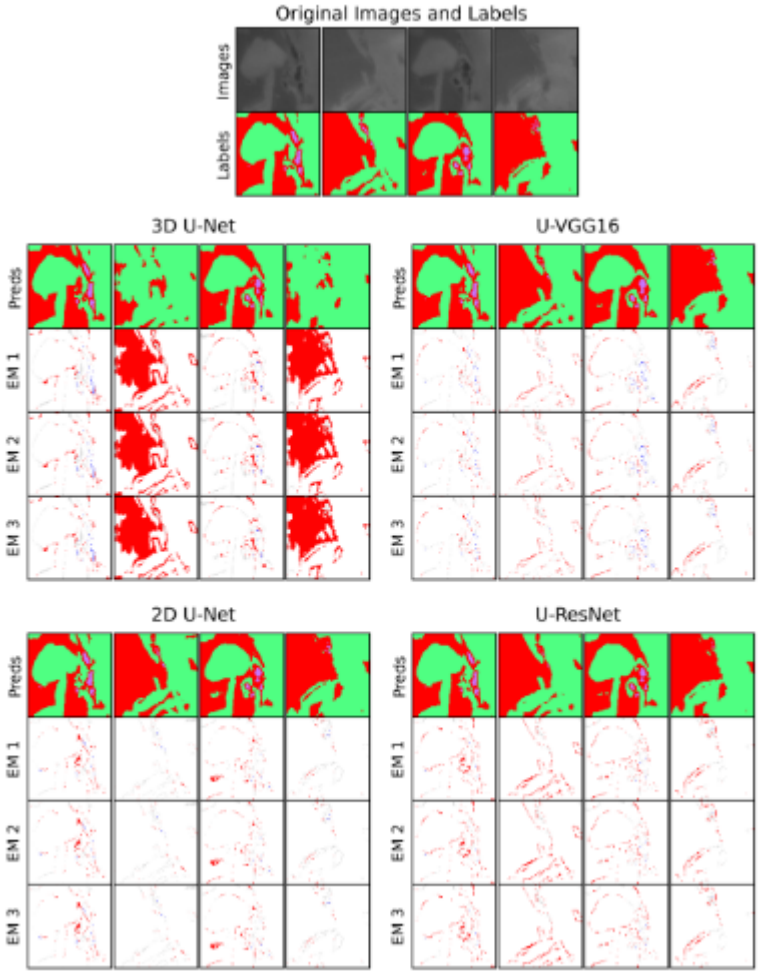


Figure 5. Sample images, labels, predictions, and corresponding error maps* from the testing split of the Shale dataset.

*Error Map Key:

- White: Label and prediction agree
- Light-grey: Other error
- Error Map 1 (EM 1):
  - Blue: Pore predicted, expected organic
  - Red: Solid predicted, expected organic
- Error Map 2 (EM 2):
  - Blue: Pore predicted, expected solid
  - Red: Organic predicted, expected solid
- Error Map 3 (EM 3):
  - Blue: Organic predicted, expected pore
  - Red: Solid predicted, expected pore

Table 3. Comparative performance of trained 3D-Unet, the U-Net, U-VGG16, and U-ResNet models after evaluation on the shale test data.

Model	*Freq. IoU*	*Mean IoU*	*Pixel-Wise Acc.*	*Solid IoU*	*Organic IoU*	*Pore IoU*
3D U-Net	0.9175	0.6441	0.8164	0.7525	0.5905	0.5891
U-Net	**0.9423**	0.8827	**0.9702**	**0.9543**	**0.9313**	0.7625
U-VGG16 (fine-tune)	0.9406	**0.8836**	0.9692	0.9526	0.9293	**0.7691**
U-ResNet (fine-tune)	0.9375	0.8549	0.9678	0.9520	0.9252	0.6876

Evaluation metrics:The Intersection over Union (IoU); the pixel-wise accuracy (PA), Mean IoU, and Frequency-Weighted IoU (Freq IoU). For more details see (Garcia-Garcia et. al).

Table 4. Table comparing the Mean-IoU scores of U-Resnet models with different training regimens. All models are evaluated on the test split.

Dataset	*No Initialization*	*Frozen Encoder*	*Fine-Tune*
Sandstone Data	0.8121	0.8776	**0.9091**
Shale Data	0.8241	0.8146	**0.8549**

Table 5. Table comparing the Mean-IoU scores of U-VGG16 models with different training regimens. All models are evaluated on the test split.

Dataset	*No Initialization*	*Frozen Encoder*	*Fine-Tune*
Sandstone Data	**0.8973**	0.8364	0.8924
Shale Data	0.8406	0.8483	**0.8836**

SUMMARY AND CONCLUSIONS

- Our results indicate deep learning architectures can successfully be applied to the task of semantic segmentation for μCT and FIB-SEM images of porous geomaterials.
- Although models are heavily reliant on the quality of input data, we found that these deep learning methods can recover information not captured by original segmentation labels.
- The use of pretrained weights from ImageNet was found to significantly improve the accuracy of U-ResNet models after training with rock images, indicating that transfer learning can improve performance for these networks (see Table 4).
  - This indicates that features learned from a large image database include common features for rock images.
  - U-VGG16 benefitted less from transfer learning, likely because the U-VGG16 has a simpler architecture compared to U-ResNet, making the model less robust with new training data as in this study(see Table 5).
- The 3D model was found to unexpectedly underperform in comparison to the 2D case.
  - One explanation for this underperformance of the 3D case is a lack of training data.
  - However, the poor performance of the 3D U-Net may not be completely explained by lack of data, as it has been found that complications arise when using depth data, such as unpredictable variations in illumination between images (Garcia-Garcia et. al).

Acknowledgement: Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

ABSTRACT

Recent advances in multiscale imaging techniques for the analysis of complex pore structures and compositions have revolutionized our ability to characterize various porous media systems. Segmentation of images obtained from different image techniques such as X-ray computed microtomography (μCT) and scanning electron microscopy (SEM) is the first step to quantitatively describe various features of geomaterials. However, conventional methods such as thresholding, watershed segmentation, and other predefined algorithms are subject to user bias, often require specific segmentation processes for each image set, and can fail to adapt to certain datasets. In this work we evaluate the capability of convolutional neural networks (CNNs)-based algorithms to segment both μCT and focused ion beam-SEM (FIB-SEM) images with varying degree of challenges for image segmentation. The performance of three different 2D CNN architectures (VGG16, ResNet, and U-Net) as well as a 3D CNN architecture (U-Net) is assessed on four independent datasets including sandstone, carbonate chalk and shale. Each of these datasets is composed of three-dimensional image stacks and corresponding ground truth labels that were constructed with various image processing algorithms. Our results indicate that deep learning architectures can successfully be applied to the task of semantic segmentation for μCT and FIB-SEM images and perform better than manual segmentation to recover natural morphology of original images. In addition, our preliminary results indicate that transfer learning can allow for models to converge more quickly during training and that generic image features (learned from a large dataset such as ImageNet) are applicable for our own image sets, and that the applicability of these generic features varies slightly from dataset to dataset. Performance comparison among different CNN architectures highlights the linkage of classification outcomes to underlying features of each CNN architecture and hyperparameters. SNL is managed and operated by NTESS under DOE NNSA contract DE-NA0003525.

REFERENCES

Garcia-Garcia, A., Orts, S., Oprea, S., Villena-Martinez, V., & Rodriguez, J. (2017). A Review on Deep Learning Techniques Applied to Semantic Segmentation. ArXiv, abs/1704.06857.

He, K., Zhang, X., Ren, S. and Sun, J. 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778.

Ronneberger, O., Fischer, P. and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention, 234–241.

Simonyan, K. and Zisserman, A. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. ICLR 2015, arXiv preprint arXiv:1409.1556.