EGU2020-4697
https://doi.org/10.5194/egusphere-egu2020-4697
EGU General Assembly 2020

Machine learning to identify sandstone properties from thin sections

Benjamin Emmel, Ole-Andre Roli, and Anouar Romdhane
SINTEF, SINTEF Industry, Norway (benjaminudo.emmel@sintef.no)

Sand volume and porosity measurements on sandstones are routine work in geoscientific applications, providing useful input to flow simulation in porous media-based analyses (e.g., in CO_2 storage and/or hydrocarbon migration studies). The classic way to gain knowledge about these parameters is point counting on thin sections. This time-consuming, repetitive, and subjective work is usually done by an experienced petrographer. Attempts to automate and digitize this process are therefore promising. An example using image analysis has been discussed in Roduit, 2007. However, one step further is combining image analysis with machine learning.

In this work, we evaluate the use of a neural network learning algorithm to classify selected sandstone properties from thin section images. Our database consists of ca. 3500 thin section images from different sandstone types with known properties. The images are grouped into 8 different sand volume and 8 different porosity classes. We split the dataset into a training (85 %) and validation dataset (15 %). In the processing stage, we normalize and scale all the images to a reference number of 128 pixels. For both classifications, we trained a convolutional neural network consisting of 5 convolutional layers and 4 max pool layers. The batches are normalized after each pooling layer and a dropout layer used to reduce overfitting before flattening. A final soft max layer is added so that the recovered output can be interpreted as probability distributions. We perform the training phase with a varying number of epochs ranging between 20 and 200. A training and validation accuracy > ca. 90 % is obtained after 25 epochs. For both cases, we observe that initially high model loss for the validation data reaches low values after 50 epochs.

To further test the approach, we analyse in a second stage a holdout dataset of sandstones from the Norwegian Continental Shelf. Preliminary results show that the derived sand volumes classification reproduce the point counting results well (80 % accuracy of predicting classes or neighbouring classes). More problematic is the reproducibility of porosities. Here, models using different epochs show variable results and the ≥100 epochs models systematically underestimates the measured rock porosities. We observe that only porosity classes well represented in the initial population of training images are reproduced with high accuracy. We finally discuss strategies to overcome such limitations.

Roduit, N., 2007. JMicroVision: un logiciel d'analyse d'images pétrographiques polyvalent. PhD Thesis University of Geneva,116 pp.