Quantitative Data Synthesis for Improving Machine Learning in Geophysical Applications: A case study on Fracture Detection in Formation Logs

J. Al-Mulhim¹, K. Katterbauer², A. Al Alshehri², R. Al-Zaidy ³, ¹

¹Information and Computer Science Department, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia
²Saudi Aramco, Dhahran, Saudi Arabia
³Center for Integrative Petroleum Research, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

Abstract

Deep learning models have become the de facto approach to address most image analysis problems. The quality of the parameters that make up these models rely on the amount and quality of labeled data used for training. Although the geophysical domain typically generates massive amounts of data, labeling the data remains a bottleneck in any machine learning application pipeline. This led to the increased interest in data synthesis in order to prevent overfitted models that are not transferable to unseen data due to limited training data. Data synthesis studies focus mostly on photographic images, which differ than software-generated images that are very common geophysical problems. In this study we aim to address this gap.

We propose a data synthesis approach that is capable of expanding the training dataset while maintaining underlying distribution assumptions of the image features. Our framework performs data synthesis using four augmentation techniques: Horizontal Flip, Gaussian Blur, Sharpness and a special data augmentation algorithm named TrivialAugmentWide. The techniques utilize a seed dataset to produce a new set of images with similar perceptual quality to that of the seed set. Each data set is used to fine-tune an EfficientNet neural network model on an image classification task for fracture detection in acoustic and resistivity formation log images.

The proposed system is evaluated on an image classification task using the public benchmark dataset for the well Pohokura-1 that is part of the Pohokura gas field located in the Taranaki Basin. Both acoustic and resistivity formation image logs were available for the well along with an indicator of an existence of a fracture in the image. Our framework synthesized an additional data from this seed set and our results demonstrate that for geophysical images that are generated via logs the TrivialAugmentWide technique generated the dataset that resulted in highest improvement in model performance.

The method proposes a novel qualitative data augmentation approach for geophysical images to increase deep learning model accuracies. The assessment of the various approaches is significant due to the perceptual distinction between log-generated images as opposed to photographic images, which have been the focus of most data synthesis studies. Additionally, qualitative data synthesis is a powerful technique for mitigating overfitting of deep learning models, which is a perpetual problem in the geophysical domain, and consequently enhancing model transferability to unseen data.