Towards Self-Labelled Carbonate Datasets with Stable Diffusion: Strategies and Implications

Ahmed Elkady[1], Ardiansyah Koeshidayatullah[1]

[1]King Fahd University of Petroleum and Minerals

Abstract

Machine learning and artificial intelligence have revolutionized geoscience data analytics and enhanced conventional methods for solving geological problems. In carbonate geology, petrographic analysis provides pivotal information in understanding the evolution of depositional and diagenetic processes affecting the carbonates. These aspects are important to better predict the quality and heterogeneity of carbonate reservoirs, which is crucial for sustainable energy exploration. However, the traditional and manual classification is labor- intensive and subjective, hindering reproducibility between geologists. Recent works have started to employ deep learning to optimize carbonate petrographic analysis. While deep learning applications in geosciences are showing promising results, their full potential is yet unrealized because of the scarcity of available data. Recently, Generative adversarial networks (GANs) have recently been the way to solve this problem by artificially generating a high-quality labeled dataset; however, GANs are often still limited as they require a substantial amount of training data. Our research addresses this challenge by exploring the effectiveness of latent diffusion models, specifically the state-of-the-art Stable Diffusion XL, and the varying strategies for fine-tuning them.

Utilizing a limited dataset of only 52 carbonate petrographic images, our diffusion model was able to generate high-quality self-labeled images that achieved a Fréchet Inception Distance (FID) score of around 51 and a Kernel Inception Distance (KID) mean score of 0.0098, which is a novel advancement given the dataset's size. Our study delves into the fine-tuning techniques employed and illustrates that latent diffusion models hold significant potential in overcoming data scarcity challenges in geosciences. This study highlights the potential of generative AI in creating high-quality, self-labelled datasets that can be used to further train or benchmark various downstream tasks in carbonate geology.