PTE572 Final Project Report (Fall 2021)

Parametrization and generation of geological facies models with Deep Convolutional GAN based on a training image

Mahammad Valiyev, PhD Student

This report was prepared for the Fall-2019 PTE572 class project (Due Date 12/06/2019).

Abstract

Generation of geologically realistic and reliable predictive models, honoring both prior conceptual geological model and field measurements has always been one of key components of sound reservoir management workflow. Effective parametrization of geological models is one way of making it possible to generate geologically realistic and reliable predictive models. In this work, GAN is proposed as a novel and effective method for the generation and parametrization of a binary facies channel model. Specifically, one representative training image for a channelized system is randomly sampled to generate training data for training the GAN. Once trained and tuned, GAN is used to generate unconditional geological models. Results show that, although, there is some connectivity loss and sometimes, artifacts being introduced, GAN can reproduce most of channel features that are present in a training image. Quantitative evaluation of results and comparison with PCA and VAE generated geological models are made to further investigate and compare the performance of proposed workflow.

Introduction

One of the central problems in the context of accurate reservoir performance prediction, is generation of reliable reservoir models. Reliability of geological model mostly depends on how effective it is in honoring prior conceptual geological model and field measurements. Two related subproblems to generation of sound reservoir models are 1) development of tools enabling generation of geologically complex and realistic prior models, effectively capturing uncertainty and 2) development of effective low-dimensional parametrization workflows for computationally efficient data conditioning.

A range of tools exist for the generation of both petrophysical and facies models, under realm of geostatistics (Deutsch et al., 1996). One of important pitfalls of some of geostatistical approaches is that they usually do not fully preserve the geological realism in the developed model. This is especially an issue during history matching process, where prior, more geologically looking, models are morphed significantly to match the dynamic data. Object-based modeling (OBM) approach was invented to address the lack of geological realism in the reservoir models, by generating realistic geobodies based on distributions of geometric shapes using marked point processes (Deutsch and Wang, 1996). However, OBM uses Markov Chain Monte Carlo (MCMC) algorithms to perform data conditioning (Holden et al., 1998), therefore, it becomes extremely slow and often fails to converge (Hauge et al., 2007; Skorstad et al., 1999).

Recent research on applying deep machine learning to reservoir modeling using methods called Generative Adversarial Networks (GANs), originally proposed by Goodfellow et al. (2014), becomes more active. Specifically, Chan & Elsheikh (2017) proposed parameterizing a geological model using GANs. The relevance of GANs for reservoir modeling is in the fact that they can define a low-dimensional representation of the original high-dimensional data (e.g., reservoir grid data in subsurface applications). The main specificity of GANs is that they are learned by creating a competition between the actual generative model or ‘‘generator’’ and a discriminator in a zero-sum game framework (Goodfellow et al., 2014), in which these components are learned jointly.

In this project, GAN is proposed as a parametrization method for a 2D channelized reservoir. Training image for a system is randomly cropped to produce training data for training the GAN. After being trained, generator within a GAN is used to produce an ensemble of geological models. Based on results, we can see that most of channel features from training image are reproduced, though there is a loss of some connectivity in generated models. Performance of proposed workflow is evaluated using some quantative methods and also compared with Variational Autoencoders (VAE) and Principal Component Analysis (PCA) based parametrization. GAN-generated models have been observed to have a higher quality.

Methodology

Workflow

The following diagrams show the general workflow and workflow elements for the project:

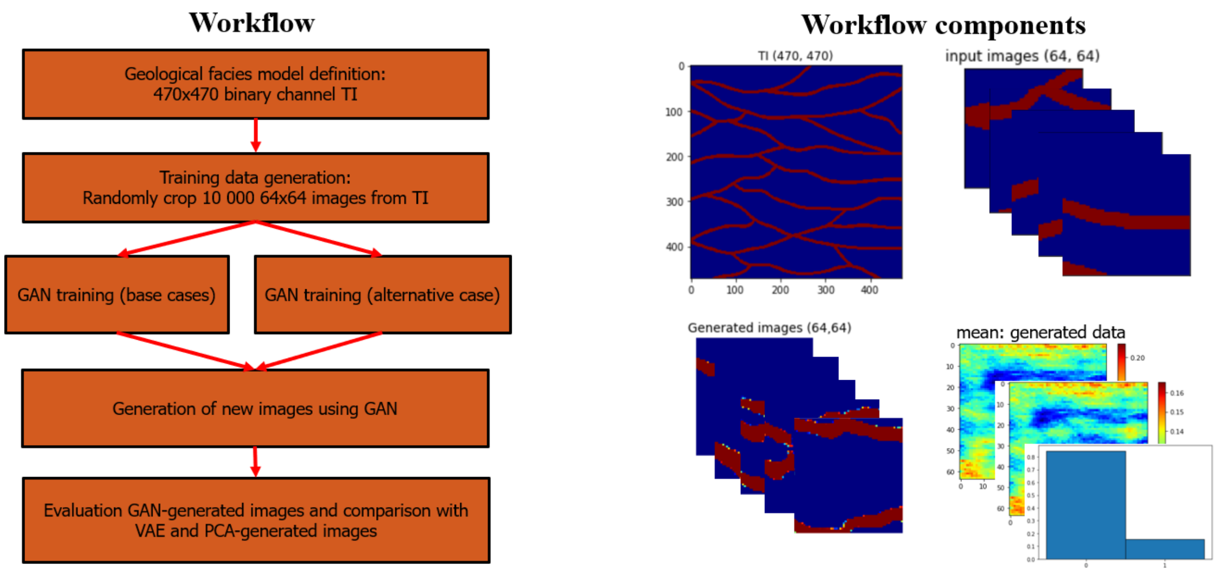

Figure 1. Workflow for the project

470x470 sized training image is used to represent the expected patterns in channelized system. 10 000 64x64 images have been randomly sampled from training image to generate training data for GAN. Base and some set of alternative data and GAN configurations are used in data generation and training process. Once the GAN is trained, generator within a GAN was used to generate new realizations. Mean, variance and histogram of values for generated and training images have been compared to quality check the results. The performance of GAN has also been compared to the ones generated using PCA and VAE.

Generative Adversarial Neural Networks

Generative Adversarial Networks (GANs) are way to make a generative deep learning model by having two neural networks (generator: G and discriminator: D) to compete with each other. The goal of D is to correctly differentiate between synthetic and real samples, whereas G aims to fool D. The interplay between G and D is formulated as a two-player minimax

$$\min_G \max_D \{\mathbb{E}_{\mathbf{y}\sim p_{\text{data}}} \log D(\mathbf{y}) + \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}} \log(1 - D(G(\mathbf{z})))\}$$

game:

Compared to other generative models, GANs do not explicitly model the data probability density function. Instead, generator function, represented by a neural network is trained in adversarial manner, that takes a random vector, sampled from gaussian or uniform distribution, and outputs an image similar to the ones provided as training data (Figure 2). There are some advantages and disadvantages of the GANs, presented in the literature:

Advantages:

- Highest resolution image generation among available generative models
- Fast inference (fast generation of new data)

Disadvantages

- Unstable training (open research problem): non-convergence and mode collapse

- Lack of intrinsic evaluation metrics (mainly by visual inspection): no explicit metric to optimize or compare generated and true images
- No density estimation: formulation of problem is different compared to other generative models

Results and Discussion

Base case

Figure 2 shows the training data and GAN-generated images for the base case (64x64 images, 10 000 images in training data).

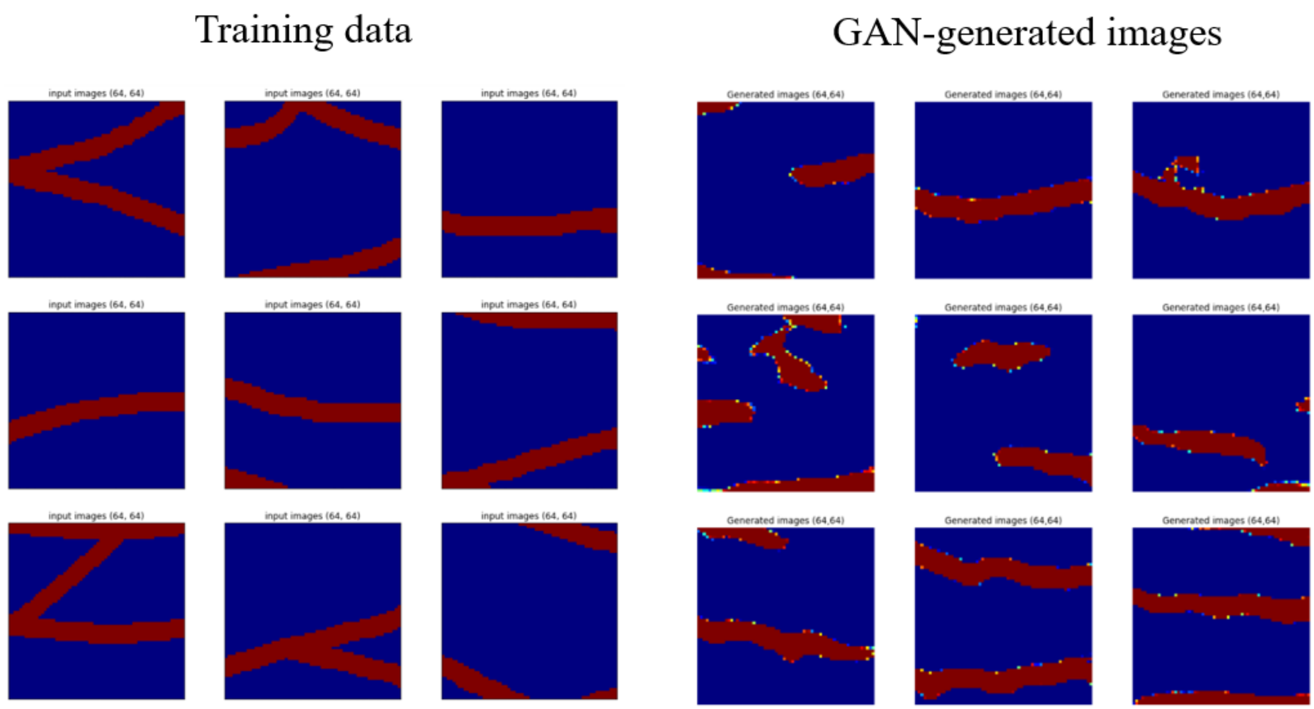

Figure 2. Training data and GAN-generaed images for the base case

Largely, the patterns in training data are reproduced and majority of generated channels preserve connectivity. However, there are some examples with connectivity being lost and some artifacts being introduced.

Sensitivities

Sensitivities to training image resolution (64x64, 100x100, 200x200), size of training data (10 000, 5000, 1000) and random gaussian vector dimension (500, 100, 1000) are done to evaluate the effect of those parameters to the quality of generated images.

100x100 and 200x200 generated images, naturally, contain more channel patterns, due to increased size of the image. However, they also contain channel patterns with lost connectivity. Therefore, increasing the size of training image data, although, leads to more natural and bigger channel pattern generation, the issue of lost connectivity is still persistent (Figure 3).

When training data volume is reduced from 10 000 to 5000 and 1000, there is clear degradation in the quality of the images. Concretely, the generated channel patterns have reduced resemblance to the ones present in training data (Figure 4).

Reducing the dimension of random gaussian vector from 1000 to 500 does not seem to have a significant effect on quality of generated images, whereas when reduced to 100, the thickness of channels become larger than the ones in training data, evidencing degradation in quality of results (Figure 5).

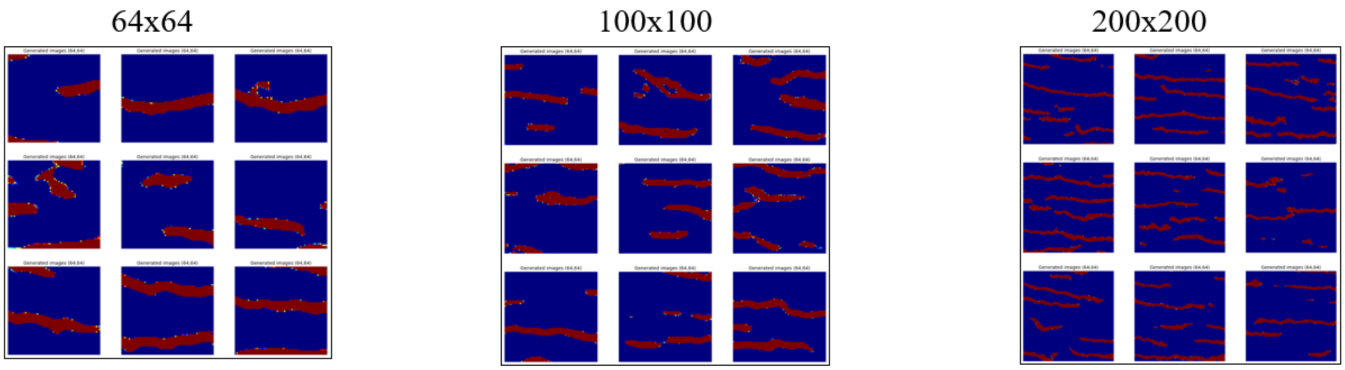

Figure 3. Sensitivity to resolution of training data

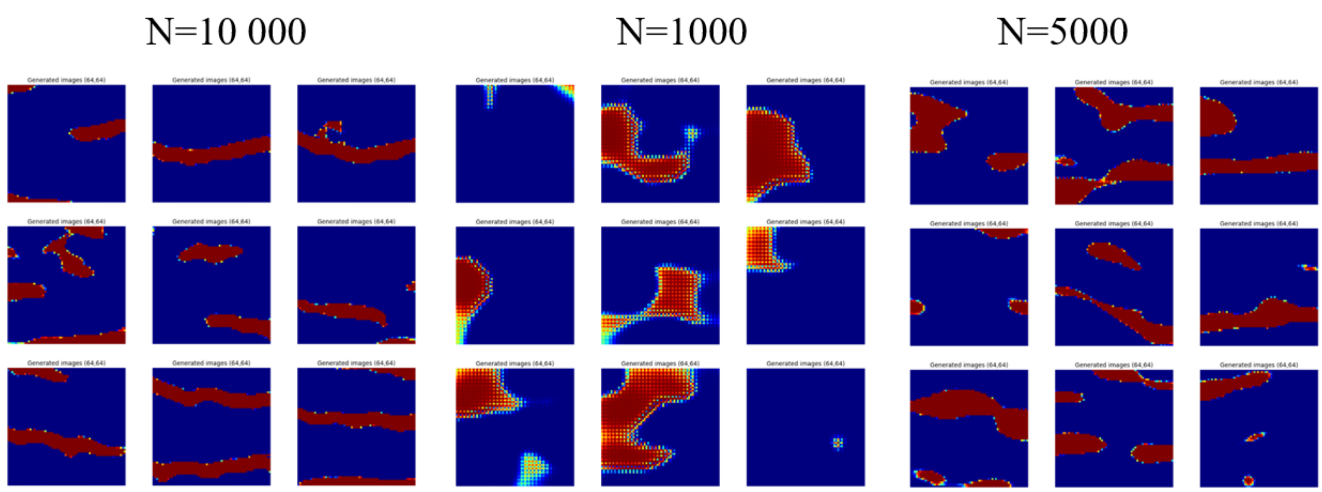

Figure 4. Sensitivity to size of training data

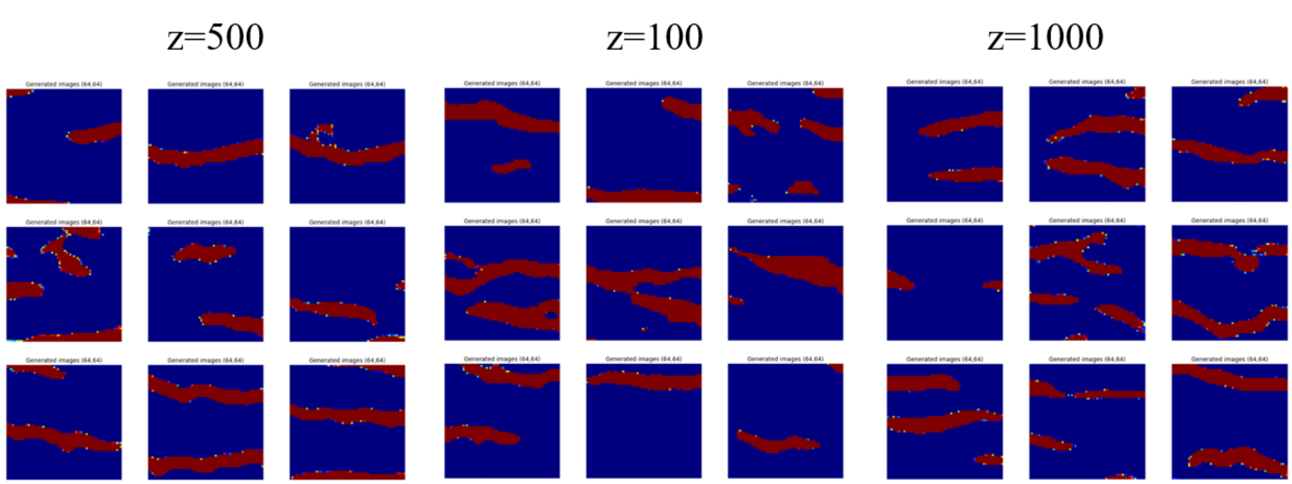

Figure 5. Sensitivity to gaussian noise vector dimension

Evaluation of results

To evaluate quantitatively the validity of results, mean, variance and histogram of generated and training data images have been compared. There is some discrepancy in position of channels between generated and training data images based on mean and variance of images, though the difference is not significant (approximately 0.05). Also, largely, the mean and variance values for each pixel are close to each other, meaning that channels are well distributed. There is also a very strong match between histograms for training and generated images, meaning that the ratio of channel to non channel facies is preserved in generated images (Figure 6).

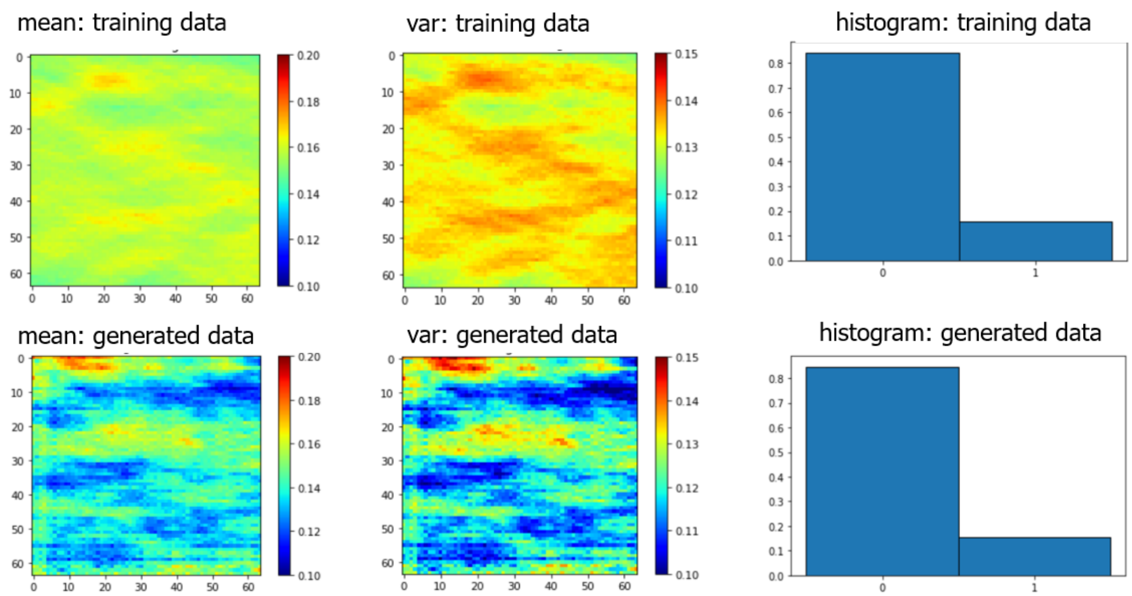

Figure 6. Mean, variance maps and histograms for training and generated data

Comparison with VAE and PCA generated images

In this section, the comparison between GAN-generated images and VAE, PCA based generated images are made. PCA-based images show the best connectivity patterns, though the images are quite blurry and generated images do not have discrete values. VAE based images, although mostly discrete, contain disconnected, erratically distributed channel patterns, that are not aligned with the ones present in training image. GAN-based images have reasonable connectivity patterns, and they are binary. Overall, this makes GAN-generated images have the best reproduction quality (Figure 7).

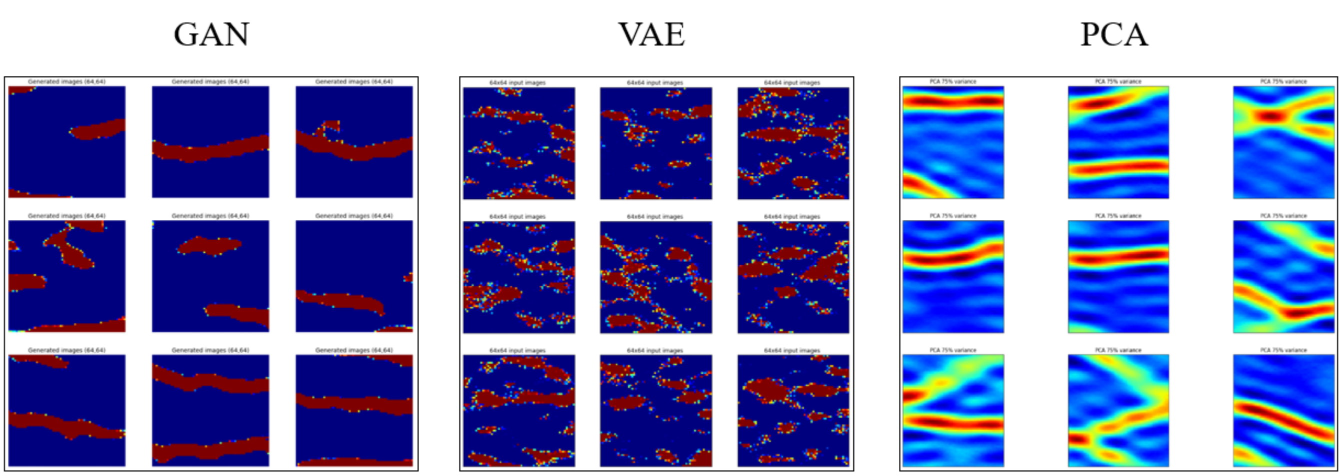

Figure 7. GAN vs VAE vs PCA based generated images

Conclusion and Further work

Study introduced GAN-based workflow for unconditional geological model generation and parametrization. Outcomes from sensitivity analysis on some input parameters and QC of obtained results have been presented. GAN has been able to reproduce the data histogram and channel patterns from TI, but there are some artifacts introduced, and spatial distribution of channels in generated samples does not fully match the ones in training data. Overall, GAN makes it possible to parametrize the geological model with low number of parameters and generate very fast, moderately realistic geological models. Results from workflow are superior to the ones, generated based using VAEs and PCA. Main challenges in the workflow are GAN-training complexity and effective data generation/sampling from TI.

As a future work, different data generation procedures 1) feeding OBM based geological models directly into GAN, 2) using different sampling strategy (sampling method, window size, multigrid sampling) can be tested. More systematic and comprehensive training of GAN (different loss function formulations, different architecture, and layer configurations) can also be introduced. Moreover, as a potential addition to the project scope, conditional GAN generated models and inversion workflows can be developed.

References

1. Arne Skorstad, Ragnar Hauge, and Lars Holden. Well conditioning in a fuvial reservoir model. Mathematical Geology 31, 7 (1999), 857–872.
2. Chan, S. and Elsheikh, A.H., 2017. Parametrization and generation of geological models with generative adversarial networks. arXiv preprint arXiv:1708.01810.
3. Clayton V Deutsch and Libing Wang. Hierarchical object-based stochastic modeling of fuvial reservoirs. Mathematical Geology 28, 7 (1996), 857–880.
4. Han, Jichao , Chen, Rongqiang , and Akhil Datta-Gupta. "Multiscale Method for History Matching Channelized Reservoirs Using Level Sets." Paper presented at the SPE Annual Technical Conference and Exhibition, Dubai, UAE, September 2016
5. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Advances in Neural Information Processing Systems. 2672–2680 (2014).
6. Rosca, M., Lakshminarayanan, B., Warde-Farley, D. and Mohamed, S., 2017. Variational approaches for auto-encoding generative adversarial networks. arXiv preprint arXiv:1706.04987.
7. Shing Chan and Ahmed H Elsheikh. Parametrization and Generation of Geological Models with Generative Adversarial Networks. arXiv preprint arXiv:1708.01810 (2017).
8. Zhang, L.F., Pan, M. and Li, Z.L., 2020. 3D modeling of deepwater turbidite lobes: a review of the research status and progress. Petroleum Science, pp.1-17.