Catalogação na publicação
Seção de Catalogação e Classificação

D192g Dantas, Lucas Gomes.
Geração de fácies geológicas a partir de um modelo de difusão latente: um estudo comparativo de embeddings / Lucas Gomes Dantas. - João Pessoa, 2025.
25 f. : il.

Orientação: Thaís Gaudencio do Rêgo.
TCC (Graduação) - UFPB/CI.

1. Fácies geológicas. 2. Modelo de difusão latente. 3. Embeddings. 4. VAE. 5. Geração de imagens sintéticas. I. Rêgo, Thaís Gaudencio do. II. Título.

UFPB/CI CDU 004.92

Elaborado por MICHELLE DE KASSIA FONSECA BARBOSA - CRB-738

Geração de Fácies Geológicas a partir de um Modelo de Difusão Latente: Um Estudo Comparativo de *Embeddings*

Lucas Gomes Dantas[1], **Thaís Gaudencio do Rêgo**[1]

[1]Centro de Informática – Universidade Federal da Paraíba (UFPB)
João Pessoa – PB – Brasil

lgd@academico.ufpb.br, gaudenciothais@gmail.com

Abstract. *Geological facies are rock units with similar depositional characteristics, essential for modeling subsurface heterogeneity in applications such as petroleum reservoir simulation and geological CO_2 storage. In scenarios with limited real data, synthetic generation becomes a useful alternative. This study investigates the generation of 2D synthetic geological facies using a Latent Diffusion Model (LDM) combined with different embedding strategies: Variational Autoencoder (VAE), VAE-GAN, and Vector Quantized VAE (VQ-VAE). Each embedder was trained under the same conditions and integrated into a common diffusion pipeline. The models were evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Fréchet Inception Distance (FID), and Learned Perceptual Image Patch Similarity (LPIPS). Among the strategies, the VAE achieved the best quantitative performance: $\overline{MAE}$ = 0.0214, $\overline{MSE}$ = 0.0092, FID = 65.0449 and $\overline{LPIPS}$ = 0.5643. The other approaches showed higher error values but exhibited characteristics suggesting potential for improved visual detail with extended training. The findings highlight a trade-off between model complexity and training efficiency in synthetic geological modeling.*

Resumo. *Fácies geológicas são unidades rochosas com características deposicionais semelhantes, essenciais para modelar a heterogeneidade da subsuperfície em aplicações como simulação de reservatórios de petróleo e armazenamento geológico de CO_2. Em cenários com escassez de dados reais, a geração sintética torna-se uma alternativa útil. Este estudo investiga a geração de fácies geológicas sintéticas 2D por meio de um Modelo de Difusão Latente (LDM), combinado com diferentes estratégias de embedding: Autoencoder Variacional (VAE), VAE-GAN e Vector Quantized VAE (VQ-VAE). Cada modelo foi treinado sob as mesmas condições e acoplado a um pipeline de difusão comum. As avaliações quantitativas foram conduzidas com base nas métricas MAE, MSE, FID e LPIPS. Dentre as abordagens, o VAE apresentou o melhor desempenho quantitativo: $\overline{MAE}$ = 0,0214, $\overline{MSE}$ = 0,0092, FID = 65,0449 e $\overline{LPIPS}$ = 0,5643. As demais estratégias apresentaram erros mais elevados, mas exibiram características que indicam um possível potencial de detalhamento visual com treinamentos mais longos. Os resultados evidenciam um equilíbrio entre a complexidade do modelo e a eficiência de treinamento na geração sintética de fácies geológicas.*

1. Introdução

Fácies geológicas podem ser definidas como conjuntos de características litológicas, estruturais e paleontológicas, que refletem os processos deposicionais e as condições ambientais sob as quais um determinado corpo de rocha foi formado. Sua identificação é fundamental para a construção de modelos geológicos coerentes, visto que permite inferir a distribuição de propriedades petrofísicas na subsuperfície, como porosidade e permeabilidade, que impactam diretamente na modelagem de fluxo em reservatórios [Ruppel e Harrington 2015].

Nesse contexto, a modelagem de fácies geológicas desempenha um papel crucial na caracterização de reservatórios, especialmente em ambientes sedimentares complexos, onde a variabilidade espacial e a heterogeneidade estrutural desafiam métodos tradicionais. Historicamente, essa modelagem se apoiou em técnicas de interpolação determinísticas e estocásticas para preencher lacunas entre dados escassos, como perfis de poço e interpretações sísmicas [Zhang et al. 2025]. Dentre esses, a Krigagem foi amplamente adotada por sua fundamentação geoestatística, embora frequentemente resulte em modelos excessivamente suavizados, devido à sua dependência do variograma. Métodos estocásticos, como a Simulação Indicadora Sequencial (SIS), oferecem maior variabilidade ao incorporar incertezas por meio de múltiplas realizações [Zhang et al. 2025]. Ainda assim, tanto abordagens determinísticas, quanto estocásticas baseadas em voxels, enfrentam limitações na representação de geometrias complexas e transições abruptas.

Com o avanço das técnicas computacionais, surgiram abordagens baseadas em objetos e, mais recentemente, métodos de aprendizado profundo voltados à geração sintética de fácies geológicas [Zhu e Zhang 2019]. Essas técnicas têm se mostrado promissoras na superação das limitações dos métodos tradicionais, especialmente na representação de padrões hierárquicos e heterogeneidades multiescalares [Chehrazi et al. 2011, Lee et al. 2023]. Entre essas, os Modelos de Difusão Latente (*Latent Diffusion Models* – LDM, do inglês) vêm se destacando como alternativas superiores às Redes Adversariais Generativas (*Generative Adversarial Network* - GAN, do inglês), por oferecerem maior estabilidade de treinamento e maior diversidade [Lee et al. 2023].

Entretanto, a eficácia dos LDMs depende criticamente da estratégia de *embedding* utilizada para projetar dados geológicos em espaços latentes semanticamente ricos. *Embeddings* são vetores de baixa dimensionalidade que preservam relações semânticas dos dados originais, facilitando o aprendizado e a manipulação de padrões complexos [Schroff et al. 2015, Chen et al. 2018]. Os *Autoencoders* variacionais (*Variational Autoencoder* - VAE, do inglês) são *autoencoders* probabilísticos que, além de reconstruir os dados de entrada, aprendem a modelar a incerteza no espaço latente por meio de um termo de regularização baseado na Divergência de Kullback-Leibler (*Kullback–Leibler Divergence* - KL, do inglês) [Kingma e Welling 2013]. Abordagens como VAE-GAN e VQ-VAE (*Vector Quantized VAE* - do inglês) aprimoram a preservação de detalhes geométricos por meio, respectivamente, de um discriminador adversarial ou de quantização vetorial [Lee et al. 2023, Hoover et al. 2023]. Estudos recentes mostram que a escolha do *embedding* impacta não apenas a fidelidade visual das amostras geradas, mas também sua coerência geológica, especialmente em reservatórios de alta heterogeneidade [Lee et al. 2023].

Este trabalho apresenta uma avaliação sistemática de três abordagens de *embed-*

ding — VAE, VAE-GAN e VQ-VAE — integradas a um LDM para a geração de fácies geológicas sintéticas, a partir do conjunto de dados GANRiver-I [Sun et al. 2023]. A análise é conduzida com base em métricas quantitativas como o Erro Médio Quadrático (*Mean Squared Error* – MSE, do inglês), Erro Absoluto Médio (*Mean Absolute Error* – MAE, do inglês), a Distância de Fréchet (*Fréchet Inception Distance* – FID, do inglês) e a Similaridade Perceptual Aprendida (*Learned Perceptual Image Patch Similarity* – LPIPS, do inglês), visando identificar os *trade-offs* entre fidelidade estrutural, qualidade perceptual e diversidade de padrões. Os resultados desta investigação contribuem para o avanço de modelos generativos aplicados à exploração de hidrocarbonetos e ao armazenamento geológico de CO_2, ampliando o potencial preditivo e a confiabilidade de modelos de reservatórios.

2. Trabalhos Relacionados

A literatura científica recente demonstra um interesse crescente na aplicação de modelos generativos profundos (*Deep Generative Models* - DGM, do inglês), como VAEs e LDMs, para a complexa tarefa de geração de modelos de fácies geológicas. Esta seção revisita contribuições chave que informam diretamente a metodologia e os objetivos desta investigação. O propósito é examinar concisamente as abordagens propostas, os conjuntos de dados utilizados e os principais resultados reportados, estabelecendo o contexto científico no qual se insere o presente trabalho.

O VAE, proposto no trabalho de [Kingma e Welling 2013], estabeleceu um paradigma para o aprendizado de modelos generativos, com espaços latentes contínuos de forma tratável. Utilizando o princípio da inferência variacional Bayesiana, o VAE aprende um mapeamento probabilístico dos dados para um espaço latente (via *encoder*) e deste de volta para o espaço de dados (via *decoder*), otimizando o limite inferior de evidência (*Evidence Lower Bound* - ELBO, do inglês). A introdução da técnica de reparametrização foi um avanço metodológico chave, permitindo a otimização por gradiente estocástico. A capacidade do VAE de aprender representações latentes ricas e estruturadas o qualifica como uma estratégia fundamental de *embedding* para modelos generativos mais complexos.

O trabalho de [Rombach et al. 2021] apresentou os LDMs como uma solução eficiente para síntese de imagens de alta qualidade. A abordagem central separa a compressão perceptual da geração propriamente dita: primeiro, um *autoencoder* mapeia a imagem para um espaço latente compacto; em seguida, um processo iterativo de difusão refina essa representação. Essa arquitetura em duas etapas reduz significativamente o custo computacional e atinge desempenho de ponta em diversas tarefas. Além disso, [Rombach et al. 2021] demonstraram a flexibilidade na escolha do *autoencoder*, incorporando tanto a regularização via KL (similar ao VAE) — que mede a dissimilaridade entre duas distribuições de probabilidade e age como penalização [Shlens 2014] — quanto a quantização vetorial (VQ), destacando o papel essencial do *embedding* latente para o sucesso do modelo.

A aplicabilidade dos LDMs à modelagem de fácies geológicas foi confirmada por investigações recentes. O estudo de [Lee et al. 2023] demonstrou a eficácia de um LDM adaptado para geração condicional de fácies de reservatório em dados sintéticos 2D (GEOPARD), preservando informações de poços e superando GANs em diversidade e fidelidade visual. Complementarmente, [Federico e Durlofsky 2024] empregaram LDMs na

parametrização de geomodelos de fácies e validaram os resultados por meio de métricas de consistência geológica e simulações de fluxo, utilizando dados gerados via Modelagem Baseada em Objetos (*Object-Based Modeling* – OBM, do inglês), no Petrel, software da Schlumberger que cria geocorpos paramétricos (canais, dunas etc.) condicionados a poços e sísmica, capturando naturalmente formas curvilíneas e heterogeneidades complexas [Hassanpour e Deutsch 2010].

A literatura recente, portanto, consolida os LDMs [Rombach et al. 2021] como uma abordagem promissora e eficiente para tarefas generativas complexas, com aplicações validadas no domínio da modelagem de fácies geológicas [Lee et al. 2023, Federico e Durlofsky 2024]. Observa-se também a relevância crucial do estágio inicial de *autoencoding*, que define o espaço latente onde opera a difusão, sendo o VAE [Kingma e Welling 2013] uma escolha frequente e eficaz [Federico e Durlofsky 2024]. No entanto, uma análise comparativa direta do impacto de diferentes estratégias de *embedding* — como VAE, VAE-GAN, VQ-VAE — na qualidade e características dos modelos de fácies gerados por LDMs ainda não foi extensivamente explorada. O presente estudo visa abordar esta questão, conduzindo uma **avaliação sistemática e quantitativa dessas estratégias de *embedding* integradas a um LDM**. Utilizando o conjunto de dados *benchmark* GANRiver-I [Sun et al. 2023], busca-se elucidar os *trade-offs* inerentes a cada abordagem – avaliados por MSE, MAE, FID e LPIPS –, contribuindo para o entendimento de como otimizar a arquitetura de LDMs para a geração realista de fácies geológicas. O Quadro 1 apresenta um panorama comparativo dos principais trabalhos utilizados neste estudo.

Quadro 1. Resumo sobre os trabalhos relacionados.

Autor	Objetivo	Algoritmo	Conjunto de Dados	Métrica
[Kingma e Welling 2013]	Introduzir o VAE para modelagem generativa e inferência aproximada via ELBO e reparametrização.	VAE	MNIST, Frey Face	*Log-Likelihood Lower Bound* (ELBO), Qualidade Visual
[Rombach et al. 2021]	Desenvolver LDMs que alcancem desempenho estado da arte em síntese de imagem (geração incondicional, texto-imagem, super-resolução, etc.).	LDM, ADM, ADM-G, BigGAN-deep, CogView, DALL-E, DC-VAE, DDPM, Modelos de Regressão, ImageBART, Lafite, LSGM, PGGAN, ProjectedGAN, SR3, StyleGAN, StyleGAN-2, U-Net GAN, UDM e VQGAN + T	CelebA-HQ, DIV2K, FFHQ, ImageNet, LSUN-Beds, LSUN-Churches e MS-COCO	Precisão, Revocação, Distância de Fréchet (*FID*), *Inception Score* (IS), PSNR, SSIM e LPIPS
[Lee et al. 2023]	Aplicar LDM para geração condicional de fácies geológicas, preservando dados e superando GANs.	LDM (adaptado p/ dados categóricos), U-Net, GAN (comparativo)	Conjunto de Dados Sintético 2D Fácies (GEOPARD)	Taxa de Erro de Preservação, Divergência JS, Avaliação Visual (Fidelidade, Diversidade)
[Federico e Durlofsky 2024]	Usar LDM com VAE para parametrização de fácies e assimilação de dados (amostragem DDIM).	LDM (VAE *encoder* + UNet Diffusion + DDIM *sampling*)	Conjunto de Dados Sintético 2D/3D Fácies (Petrel OBM)	Visual, Conectividade 2-pontos, Estatísticas de Fluxo, SSIM
[Sun et al. 2023]	Descrever o conjunto de dados GANRiver-I como *benchmark* para ML em modelagem de fácies.	Geração de dados via FLUMY (*Process-Based Simulator*)	GANRiver-I (Fluvial Meandrante)	Descrição do Conjunto de Dados
Este estudo	Comparar sistematicamente *embeddings* (VAE, VAE-GAN, VQ-VAE) em um LDM para gerar fácies geológicas.	LDM, VAE, VAE-GAN, VQ-VAE	GANRiver-I [Sun et al. 2023]	Erro Médio Quadrático (*MSE*), Erro Absoluto Médio (*MAE*), Distância de Fréchet (*FID*), Similaridade Perceptual Aprendida (*LPIPS*)

3. Metodologia

Esta seção detalha os procedimentos metodológicos adotados na presente pesquisa, englobando a caracterização da base de dados, as arquiteturas dos modelos generativos implementados, os parâmetros de treinamento, as métricas utilizadas para avaliação quantitativa e a configuração do ambiente experimental.

3.1. Base de Dados

As imagens de fácies geológicas utilizadas nesta pesquisa foram extraídas do conjunto de dados público GANRiver-I [Sun et al. 2023] (Figura 1). Este conjunto de dados foi desenvolvido especificamente para servir como um *benchmark* desafiador para algoritmos de aprendizado de máquina e geoestatística, voltados à reprodução de geometrias complexas e não-estacionárias de fácies em reservatórios fluviais meandrantes de baixa razão neta sobre bruta (*Net-to-Gross* - NTG, do inglês) [Sun et al. 2023].

O GANRiver-I consiste em 16.000 imagens 2D, com sistema de cor RGBA, cada uma com dimensões de 256x256 pixels. Estas imagens são *slices* horizontais extraídas de 25 simulações 3D distintas, geradas através do simulador baseado em processos FLUMY^{TM} [Sun et al. 2023]. Uma característica importante do GANRiver-I é a sua disponibilização em três níveis de complexidade, com 3, 7 e 9 classes de fácies, obtidas por amalgamação hierárquica das fácies originais do FLUMY^{TM}, conforme detalhado por [Sun et al. 2023]. Para os propósitos desta investigação comparativa de estratégias de *embedding*, optou-se pela utilização da versão com 9 fácies, a qual preserva a maior quantidade de detalhes e classes distintas geradas pelo simulador original. A Figura 2 ilustra amostras de imagens que compõem essa versão do conjunto.

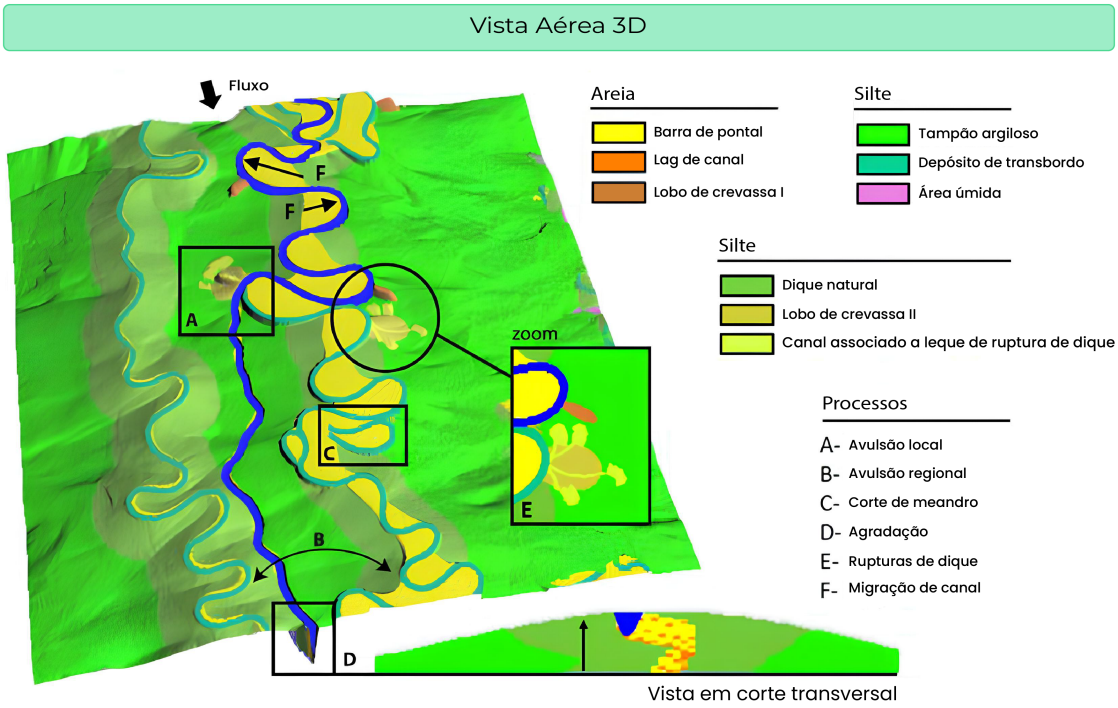
3.1.1. Pré-processamento de Dados

Apresenta-se, nesta etapa, a estratégia de filtragem adotada para equilibrar a distribuição de fácies no conjunto GANRiver-I, removendo imagens excessivamente dominadas por uma única classe. Em seguida, descreve-se o processo de preparação dos dados, que engloba a conversão das imagens para RGB, a normalização dos pixels e a divisão reprodutível em conjuntos de treinamento, validação e teste.

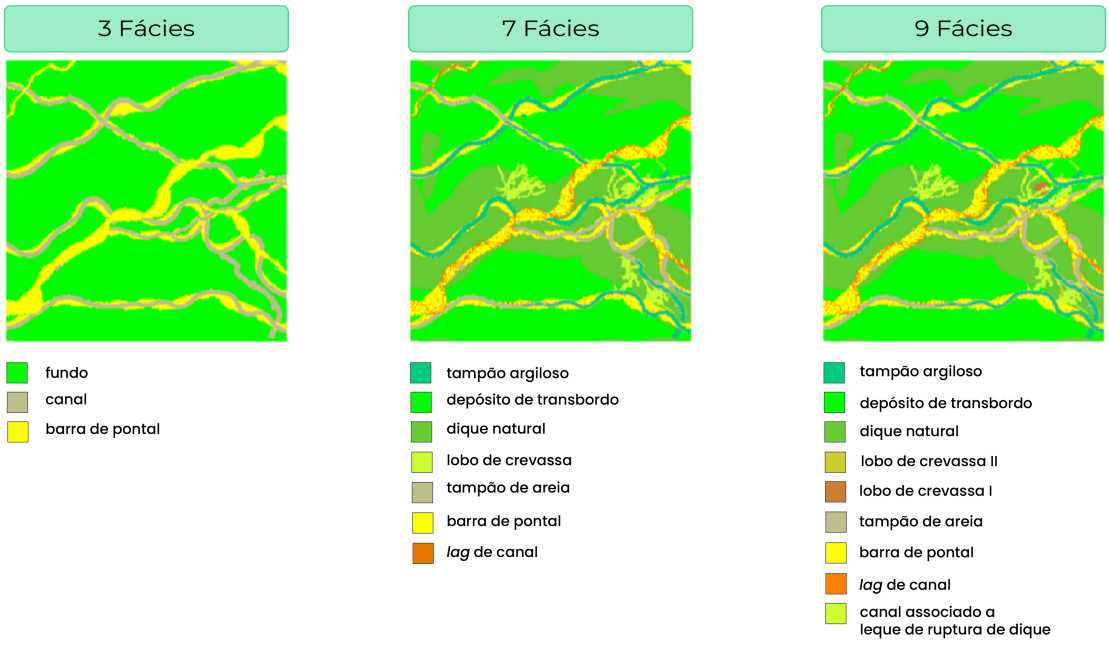
3.1.1.1. Filtragem do Conjunto de Dados

Uma análise preliminar da distribuição das fácies no conjunto de dados GANRiver-I indicou uma representação desbalanceada. Verificou-se que um subconjunto considerável das imagens era majoritariamente dominado por uma única fácies, caracterizada pela cor verde (RGB: 0, 255, 0). A super-representação de uma classe em um conjunto de dados de treinamento pode enviesar modelos generativos, levando-os a focar excessivamente nos padrões da classe dominante e a negligenciar a aprendizagem de características associadas às classes minoritárias. Conforme discutido por [Everaert et al. 2023], esse problema pode ser particularmente acentuado em modelos de difusão, devido ao viés de vazamento de sinal (*signal-leaky bias* - do inglês).

O viés de vazamento de sinal refere-se a um fenômeno em que, durante o treinamento, o ruído adicionado às imagens não elimina completamente o conteúdo semântico



(a)



(b)

Figura 1. (a) Representação tridimensional dos processos de um sistema meandrante simulados pelo FLUMY^{TM}. (b) Exemplos de imagens 2D horizontais geradas pelo FLUMY^{TM} e utilizadas no conjunto de dados GANRiver-I, com três níveis de complexidade: 3, 7 e 9 fácies. Fonte: Adaptado de [Sun et al. 2023].

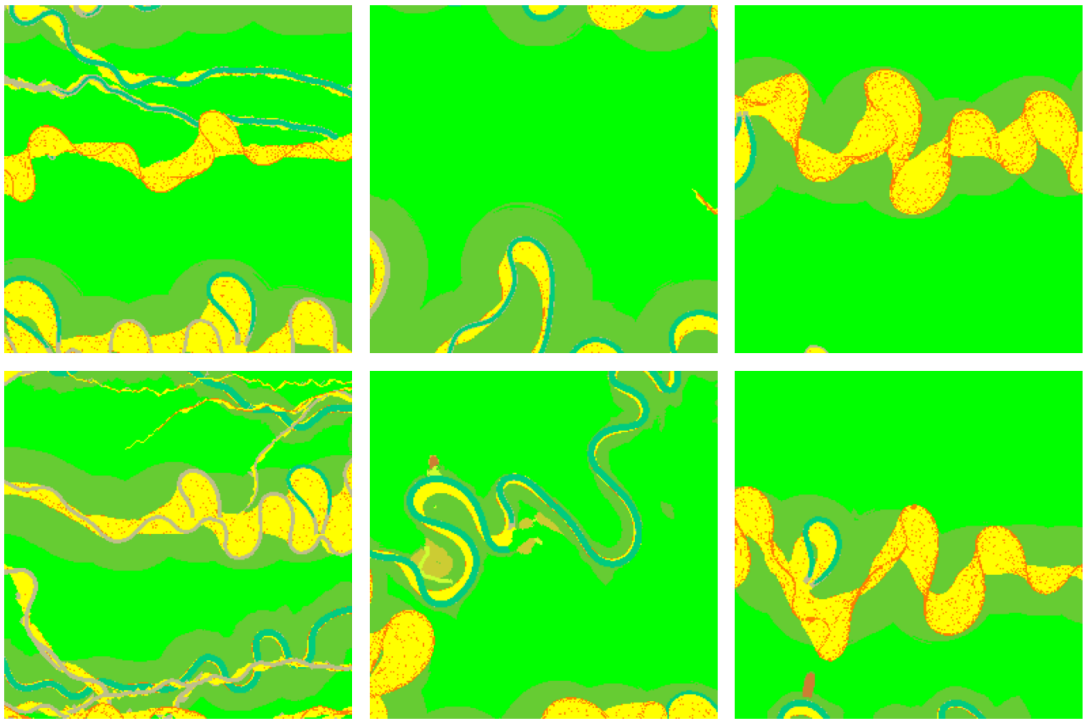


Figura 2. Amostras de imagens que compõem o conjunto de dados GANRiver-I. Fonte: [Sun et al. 2023]

original, permitindo que traços da estrutura dos dados (o sinal) ainda estejam presentes mesmo nos estágios finais do processo de difusão. Com isso, o modelo aprende a reconstruir imagens, com base em resíduos estatísticos da distribuição original dos dados, que podem estar enviesados — como no caso da predominância da fácies verde. Essa discrepância entre a distribuição dos dados ruidosos ao final do treinamento — que incorporam um viés de vazamento de sinal, refletindo as estatísticas do conjunto de dados — e a distribuição de ruído puro, usada para iniciar a inferência, pode impedir que o modelo capture adequadamente a diversidade do domínio-alvo e represente corretamente a heterogeneidade geológica.

Visando mitigar esse viés potencial e promover um aprendizado mais robusto e generalizável por parte do modelo de difusão, optou-se por aplicar um critério de filtragem baseado na predominância da fácies verde. Especificamente, todas as imagens do conjunto de dados original foram processadas para quantificar a porcentagem de pixels cujo valor era RGB $(0, 255, 0)$. Instâncias em que essa fácies cobria $\geq 89\%$ da área total foram consideradas excessivamente representativas da classe dominante e, portanto, removidas.

A Figura 3 evidencia o efeito desse filtro: o histograma à esquerda mostra o conjunto original, com uma cauda longa de amostras quase inteiramente verdes; o histograma à direita, após a filtragem, revela que essa cauda desaparece enquanto o corpo principal da distribuição permanece praticamente inalterado.

Este procedimento de filtragem resultou na exclusão de 2.370 imagens. Consequentemente, o conjunto de dados efetivamente empregado nas etapas subsequentes foi

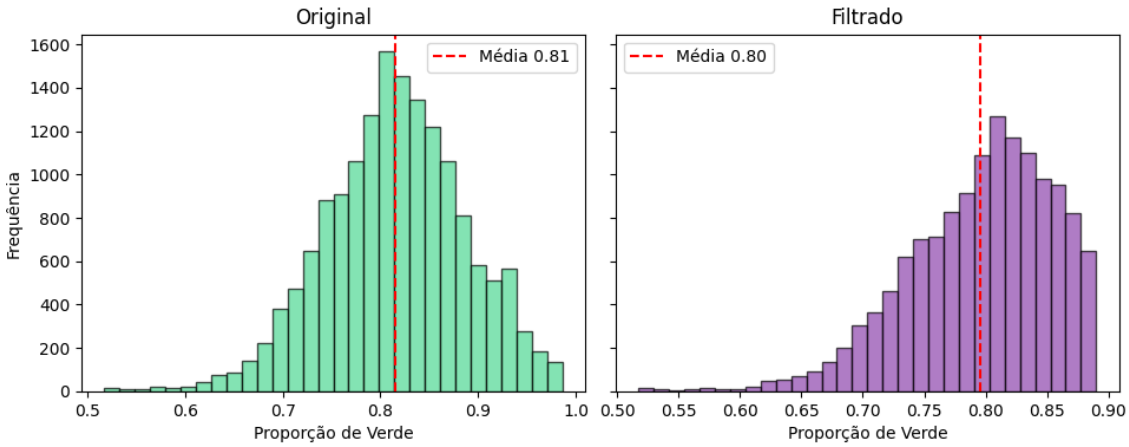


Figura 3. Distribuição da proporção de pixels verdes antes e depois da filtragem.

composto por 13.630 imagens, buscando uma base de dados com uma distribuição de fácies mais equilibrada.

3.1.1.2. Preparação dos dados

Após a filtragem inicial, as imagens remanescentes foram preparadas, utilizando as seguintes etapas implementadas em PyTorch, com o intuito de adequá-las ao treinamento dos modelos gerativos propostos:

1. As imagens foram convertidas explicitamente para o espaço de cores RGB, removendo-se o canal alfa originalmente presente no conjunto GANRiver-I.
2. Não foi realizado redimensionamento, uma vez que todas as imagens já possuíam as proporções compatíveis com o experimento, permanecendo em 256×256 pixels.
3. Em seguida, as imagens foram convertidas em tensores PyTorch com a transformação `ToTensor()`, a qual reescala os valores dos pixels do intervalo original [0, 255] para o intervalo [0, 1].
4. Após essa conversão, os valores dos pixels foram normalizados para o intervalo [-1, 1] por meio da transformação linear:

$$p' = \frac{p - 0{,}5}{0{,}5},$$

onde p representa o valor do pixel após a conversão inicial. Essa normalização é uma prática comum em modelos generativos baseados em redes neurais profundas, como GANs e modelos de difusão, por contribuir para uma melhor estabilidade numérica durante o treinamento [Ho et al. 2020a].

Após essas etapas, o conjunto de dados resultante foi dividido aleatoriamente e de forma reprodutível em três subconjuntos: treinamento (80%, 10.904 imagens), validação (10%, 1.363 imagens) e teste (10%, 1.363 imagens). A divisão utilizou uma semente fixa para garantir consistência entre experimentos.

3.2. Dependências e ambiente

Os experimentos foram realizados utilizando a linguagem de programação Python 3.8.18, com execução em ambiente local baseado no sistema operacional Windows 11. O código-fonte foi desenvolvido a partir de adaptações do repositório *Medfusion*[1], que fornece implementações dos *embedders* utilizados neste estudo, bem como o modelo de difusão latente aplicado originalmente à geração de imagens médicas. As principais bibliotecas utilizadas foram:

- `Torch` – versão 2.5.0+cu118: Biblioteca principal de aprendizado profundo, com suporte à execução em GPU;
- `Torchvision` – versão 0.20.0+cu118: Utilizada para transformações de imagens e integração com os conjuntos de dados;
- `Pytorch-lightning` – versão 1.9.5: Estrutura de alto nível para simplificação do treinamento de modelos;
- `Torchmetrics` – versão 1.6.0: Utilizada para cálculo de métricas durante o treinamento;
- `MONAI` – versão 1.4.0: Framework voltado ao processamento de imagens médicas, adaptado para o contexto geológico;
- `Numpy` – versão 1.26.4: Processamento de dados matriciais;
- `Pandas` – versão 2.2.2: Manipulação e análise de dados tabulares;
- `Matplotlib` – versão 3.8.4: Visualização gráfica de resultados.

A configuração de hardware utilizada para os experimentos foi composta por:

- **Processador:** Intel® Core™ Ultra 7 155H CPU @ 4,80 GHz;
- **Memória RAM:** 64 GB, 4800 MHz;
- **Placa de vídeo:** NVIDIA GeForce RTX 4060 Laptop GPU com 8 GB de VRAM;
- **CUDA:** versão 11.8; **cuDNN:** versão 90100.

3.3. Métricas de reconstrução e geração

A análise quantitativa dos *embedders* é fundamental para avaliar sua capacidade de reconstruir fielmente as imagens de fácies geológicas, enquanto a análise quantitativa do LDM permite mensurar a qualidade e a diversidade visual das imagens sintéticas geradas.

Para as reconstruções, foram utilizadas métricas consolidadas na literatura que quantificam a similaridade entre a imagem original e a imagem reconstruída, como o MAE [Willmott e Matsuura 2005] e o MSE [Tan et al. 2013].

Já para as amostras geradas pelo LDM, foi utilizada a FID [Heusel et al. 2018], métrica amplamente empregada na avaliação de modelos generativos, que compara distribuições estatísticas entre as imagens reais e as imagens geradas no espaço de ativação de uma rede convolucional treinada. Complementarmente, visando capturar aspectos perceptuais não refletidos pelo FID, adotou-se também a métrica LPIPS [Zhang et al. 2018], que avalia a similaridade visual entre imagens com base em distâncias de características profundas, fornecendo uma perspectiva mais alinhada à percepção humana de qualidade visual. No presente estudo, o LPIPS foi empregado para quantificar a semelhança perceptual média entre amostras sintéticas geradas pelo LDM e imagens reais do conjunto de teste, reforçando a análise da qualidade visual das gerações.

[1] `https://github.com/mueller-franzes/medfusion`

3.3.1. Erro Médio Absoluto

O MAE mede a média das diferenças absolutas entre os valores dos pixels da imagem original e da reconstruída, como mostra a Equação (1). É menos sensível a valores discrepantes que o MSE:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_{i_{\text{real}}} - y_{i_{\text{gen}}}|. \tag{1}$$

3.3.2. Erro Médio Quadrático

O MSE calcula a média das diferenças elevadas ao quadrado entre cada pixel da imagem original e da reconstruída. Por penalizar erros maiores de forma mais intensa, é útil para detectar distorções localizadas. A Equação (2) define o cálculo:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_{i_{\text{real}}} - y_{i_{\text{gen}}})^2. \tag{2}$$

3.3.3. Distância de Fréchet

Já a FID avalia o quão próximas estatisticamente são as distribuições das imagens reais e geradas, comparando seus vetores de características extraídos de uma rede neural treinada — neste estudo, a ResNet50 [Mascarenhas e Agarwal 2021]. A FID considera a média e a matriz de covariância das distribuições para medir essa semelhança, conforme expresso na Equação (3):

$$\text{FID}(x, g) = ||\mu_x - \mu_g||^2 + \text{Tr}\left(\Sigma_x + \Sigma_g - 2\sqrt{\Sigma_x \Sigma_g}\right), \tag{3}$$

onde:

- μ_x e μ_g são as médias dos vetores de características das imagens reais (x) e geradas (g);
- Σ_x e Σ_g representam as respectivas matrizes de covariância;
- Tr indica o traço da matriz (soma dos elementos da diagonal principal).

3.3.4. Similaridade Perceptual Aprendida

O LPIPS, no entanto, avalia a similaridade perceptual entre duas imagens, baseando-se em distâncias no espaço de características extraídas por redes convolucionais profundas. Diferentemente de métricas tradicionais como o MAE e o MSE, que operam no espaço de pixels, o LPIPS busca refletir a percepção visual humana ao comparar as imagens em níveis mais abstratos. A métrica calcula, para cada camada l da rede, a distância euclidiana entre os mapas de características normalizados das imagens avaliadas, ponderando

cada canal por pesos treinados que representam sua relevância perceptual. A Equação (4) apresenta a definição formal do LPIPS:

$$\text{LPIPS}(x, y) = \sum_{l} \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot (f_l(x)_{hw} - f_l(y)_{hw})\|_2^2, \tag{4}$$

em que:

- $f_l(\cdot)$ representa os mapas de características extraídos na camada l;
- w_l são os pesos treinados para cada canal da camada;
- H_l e W_l correspondem à altura e largura dos mapas de características;
- $\odot$ denota a multiplicação elemento a elemento.

No presente estudo, considerando que o modelo de difusão latente gera amostras de forma não condicional, o LPIPS foi aplicado para mensurar a similaridade perceptual entre cada imagem gerada e um conjunto amplo de imagens reais. A métrica final é expressa como a média das distâncias perceptuais, permitindo avaliar o quão próxima, em termos visuais, a distribuição das imagens sintéticas está da distribuição real.

3.4. Modelos

Neste estudo, foram utilizados dois componentes principais no processo de geração de imagens sintéticas de fácies geológicas: as arquiteturas de *Autoencoder*, utilizadas para gerar os *embeddings* — ou seja, mapear as imagens reais para um espaço latente comprimido —, e o LDM, responsável pela geração das imagens sintéticas a partir desse espaço latente.

3.4.1. Arquiteturas de *Embedding*

Para a composição do espaço latente utilizado no processo de difusão, foram implementadas três arquiteturas distintas de *autoencoder*, cada uma desempenhando o papel de *embedder* no LDM:

- **VAE** [Kingma e Welling 2013]: modelo probabilístico que introduz uma regularização no espaço latente, favorecendo a geração de representações contínuas e estruturadas;
- **VAE-GAN** [Larsen et al. 2016]: extensão adversarial do VAE, combinando a reconstrução probabilística com o refinamento visual promovido por um discriminador;
- **VQ-VAE** [van den Oord et al. 2017]: abordagem que discretiza o espaço latente via quantização vetorial, promovendo representações mais robustas e consistentes.

Cada *embedder* foi treinado com o mesmo esboço de rede — 4 blocos convolucionais, espaço latente de 8 canais, 10 épocas — e, em seguida, acoplado ao *pipeline* do LDM para gerar amostras sintéticas.

A Figura 4 apresenta a estrutura fundamental do VAE, destacando os componentes essenciais: o Codificador ($\mathcal{E}$), responsável por mapear a entrada x para o espaço latente z; e o Decodificador ($\mathcal{D}$), que reconstrói a amostra $\tilde{x}$ a partir dessa representação. Embora representem abordagens distintas, tanto o VAE-GAN quanto o VQ-VAE preservam essa base estrutural, diferenciando-se pela inclusão de mecanismos adversariais ou de quantização vetorial, respectivamente.

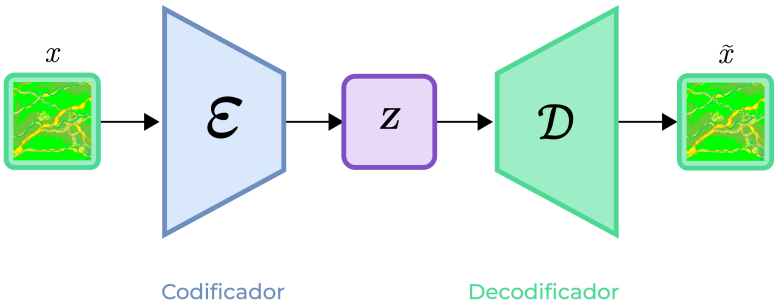


Figura 4. Esquema simplificado da arquitetura do VAE. Fonte: Adaptado de [Wang et al. 2023].

3.4.2. Modelo de Difusão Latente

Os LDMs são modelos generativos que operam em espaços latentes comprimidos e semanticamente ricos, em vez de atuarem diretamente no espaço original das imagens. Essa abordagem é baseada em processos estocásticos de *denoising*, em que vetores latentes são progressivamente corrompidos por ruído e, posteriormente, reconstruídos por meio de um processo reverso aprendido [Ho et al. 2020b].

A principal motivação por trás dessa reformulação é a redução do custo computacional e a manutenção de estruturas semânticas relevantes durante o processo de geração. A viabilidade na caracterização de dados complexos em contextos diversos é possível devido a estratégia de associar modelos de difusão a *autoencoders* [Rombach et al. 2021].

A arquitetura adotada neste trabalho segue esse princípio. Primeiramente, cada imagem geológica x é transformada em uma representação latente por meio de um *autoencoder*, que comprime as informações espaciais e semânticas em um vetor z. Sobre essa representação, aplica-se o processo de difusão direta, no qual ruído gaussiano é adicionado de forma incremental ao longo de T etapas, até que o vetor se torne praticamente indistinguível de uma amostra de ruído puro [Rombach et al. 2021]. A Figura 5 apresenta o processo de adição e remoção de ruído.

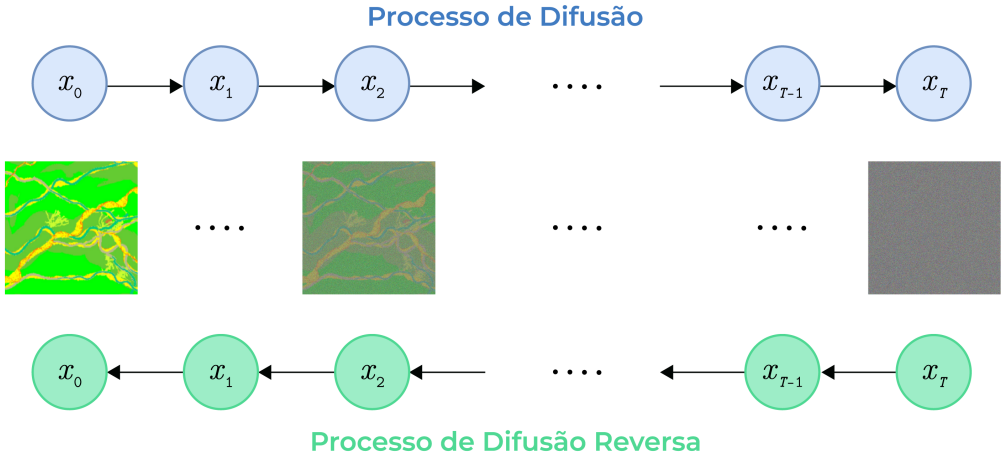


Figura 5. Processo de difusão e difusão reversa. Fonte: Adaptado de [Wang et al. 2023].

Em seguida, durante o processo de difusão reversa, uma rede neural convolucional do tipo U-Net é utilizada para estimar e remover o ruído adicionado em cada estágio, reconstruindo gradualmente uma representação latente coerente com os dados originais [Ronneberger et al. 2015]. Ao final desse processo, o vetor latente restaurado é decodificado pelo *autoencoder*, resultando em uma nova imagem sintética. A Figura 6 ilustra o processo de difusão no espaço latente.

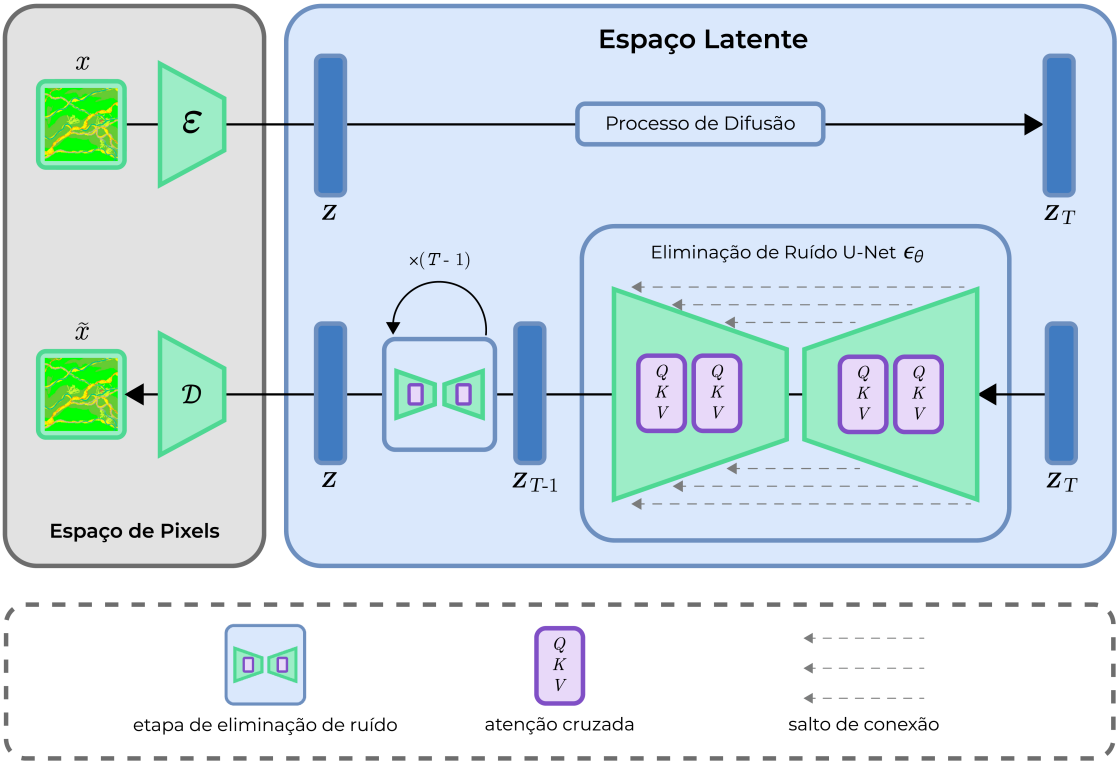


Figura 6. Ilustração do processo de difusão no espaço latente. A imagem real x é codificada em um vetor latente z pelo *autoencoder* ($\mathcal{E}$), que então é corrompido por ruído ao longo de T etapas até chegar em z_T. No processo reverso, a U-Net estima e remove o ruído passo a passo, até recuperar z, que é decodificado ($\mathcal{D}$) para gerar a imagem sintética $\tilde{x}$. A figura também destaca os blocos de atenção (QKV), conexões de salto e o módulo de *denoising* repetido $T-1$ vezes. Fonte: Adaptado de [Rombach et al. 2021].

3.4.3. Treinamento

Esta seção descreve a etapa de treinamento dos modelos envolvidos na geração de fácies geológicas sintéticas. O objetivo central foi comparar diferentes estratégias de *embedding* utilizadas na etapa de codificação do espaço latente — VAE, VAE-GAN e VQ-VAE — mantendo a arquitetura do LDM constante entre os experimentos. Dessa forma, buscou-se isolar os efeitos da escolha do *autoencoder* na qualidade estrutural e diversidade das amostras geradas.

3.4.4. Treinamento dos *Embedders*

O treinamento dos modelos de *embedding* foi conduzido de forma sistemática, utilizando o mesmo conjunto de dados em todos os experimentos. Abaixo, são detalhados os principais aspectos da configuração experimental:

- **Base de dados**: Foi utilizado um conjunto de 13.630 imagens RGB com resolução de 256×256 pixels. A divisão dos dados seguiu a proporção de 80% para treinamento, 10% para validação e 10% para teste. A semente 42 foi utilizada para replicabilidade.
- **Arquitetura base**: Todos os *embedders* foram construídos a partir de uma estrutura convolucional comum, composta por:
  - Quatro blocos convolucionais no *encoder* e no *decoder*.
  - Canais intermediários: `[64, 128, 256, 512]`.
  - *Kernel size*: 3×3.
  - *Strides*: `[1, 2, 2, 2]`.
  - Espaço latente de 8 canais: `emb_channels = 8`.
  - Convoluções bidimensionais: `spatial_dims = 2`.
  - Supervisionamento profundo ativado: `deep_supervision = 1`.
  - Mecanismos de atenção desabilitados: `use_attention = "none"`.
- **Configurações específicas por modelo**:
  - **VAE**: Treinado com função de perda MAE (`L1Loss`) e regularização do espaço latente com peso `embedding_loss_weight = 1e-6`. A estrutura latente é contínua e incentivada a seguir uma distribuição Gaussiana padrão.
  - **VAE-GAN**: Adota a mesma base do VAE, incorporando um discriminador adversarial ativado desde o início do treinamento (`start_gan_train_step = -1`), com o objetivo de aumentar o realismo visual das imagens reconstruídas.
  - **VQ-VAE**: Emprega uma representação latente discreta baseada em um dicionário com `8192` vetores. A perda utilizada foi a MAE (`L1Loss`), com penalização de quantização controlada pelo parâmetro `beta = 1`.
- **Configurações de treinamento**:
  - Otimizador: *Adam*.
  - Taxa de aprendizado: 0,0001.
  - Número de épocas: 10.
  - Tamanho do lote: 4 imagens.
  - Estabilização numérica: uso de `GradScaler`.
  - Monitoramento contínuo das perdas de reconstrução durante o treinamento.

3.4.5. Treinamento do Modelo de Difusão Latente

Após o treinamento de cada *autoencoder*, o respectivo modelo foi integrado a um *pipeline* de difusão latente, com o objetivo de gerar amostras sintéticas a partir do espaço latente aprendido. A implementação foi realizada com base na classe `DiffusionPipeline`, composta por três módulos principais: estimador de ruído, agendador de ruído e decodificador.

- **Estimador de ruído (`UNet`)**:
  - Estrutura com quatro blocos convolucionais.
  - Canais intermediários: `[256, 256, 512, 1024]`.
  - *Kernel sizes*: `[5, 3, 3, 3]`.
  - *Strides*: `[1, 2, 2, 2]`.
  - Dimensionalidade espacial: `spatial_dims = 2`.
  - Blocos residuais ativados: `use_res_block = True`.
  - Mecanismos de atenção aplicados a partir do segundo bloco: `use_attention = [False, True, True, True]`.
  - *Embeddings* temporais e condicionais foram desativados.
- **Agendador de ruído (`GaussianNoiseScheduler`)**:
  - Número de etapas: `timesteps = 1000`.
  - Intervalo de ruído: `beta_start = 0.002` a `beta_end = 0.02`.
  - Estratégia de agendamento: `scaled_linear`.
- **Configurações adicionais da pipeline**:
  - Objetivo da estimação: reconstrução direta do vetor latente final corrompido (`estimator_objective = "x_T"`).
  - Média exponencial de pesos ativada: `use_ema = True`.
  - Validação desativada durante o treinamento: `limit_val_batches = 0`.
- **Parâmetros de treinamento**:
  - Função de perda: MAE.
  - Otimizador: *AdamW*, com taxa de aprendizado de 0,0001.
  - Número de épocas: 10.
  - Tamanho do lote: 4 imagens.
  - Geração de amostras e salvamento de *checkpoints* a cada 100 etapas.
  - O modelo da última época foi armazenado e utilizado para geração de amostras sintéticas.

Essa estrutura modular permitiu avaliar de forma controlada o impacto das diferentes estratégias de *embedding* sobre o desempenho do processo generativo, mantendo fixos os demais componentes do *pipeline* de difusão.

4. Resultados e Discussões

Nesta seção, são apresentados e analisados os resultados obtidos ao longo dos experimentos conduzidos com os diferentes modelos de *embedding* — VAE, VAE-GAN e VQ-VAE —, bem como com o LDM associado a cada uma dessas estratégias. A avaliação quantitativa foi realizada com base nas métricas descritas na Seção 3, enquanto a análise qualitativa se baseou na observação visual das amostras geradas.

4.1. Resultados dos *Embedders*

O primeiro conjunto de resultados refere-se à capacidade de reconstrução dos modelos de *embedding* utilizados. As três arquiteturas — VAE, VAE-GAN e VQ-VAE — foram treinadas sob as mesmas condições e com um espaço latente de 8 canais, de forma a permitir uma comparação justa. A Tabela 1 apresenta os valores das métricas quantitativas calculadas no conjunto de teste, incluindo o MAE e o MSE.

Tabela 1. Resultados quantitativos médios dos *embedders* no conjunto de teste.

Modelo	$\overline{\text{MAE}}$ (↓)	$\overline{\text{MSE}}$ (↓)
VAE	0,0214	0,0092
VAE-GAN	0,0642	0,0243
VQ-VAE	0,0357	0,0218

A Figura 7 complementa esses dados, exibindo a evolução do MAE ao longo das dez épocas de treinamento. Observa-se que o VAE converge rapidamente e mantém o menor MAE em todas as épocas; o VQ-VAE apresenta desempenho intermediário, enquanto o VAE-GAN parte de erro mais alto, mas o reduz de forma consistente ao longo do treinamento.

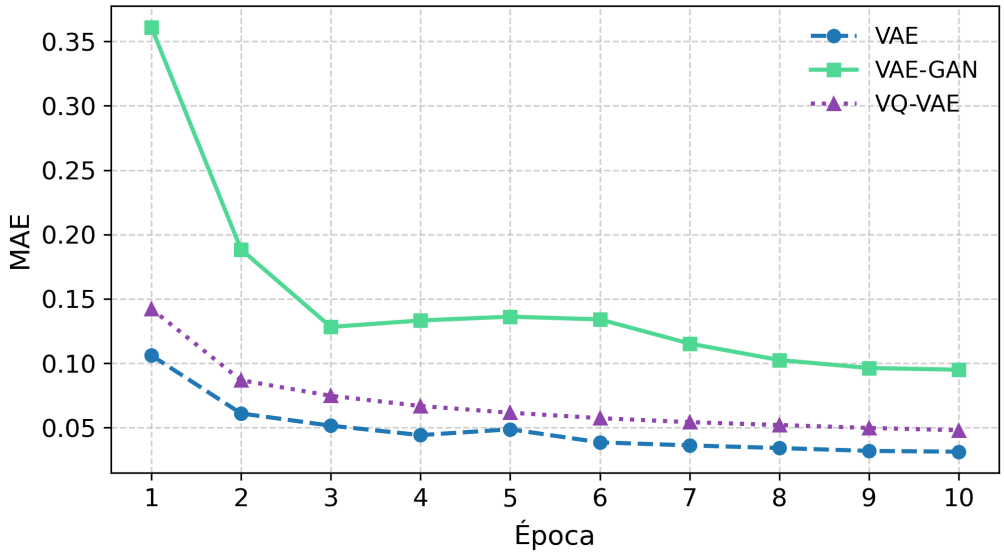


Figura 7. MAE durante o treinamento dos *embedders*.

Na análise quantitativa das reconstruções, os valores do MAE e do MSE — métricas em que, quanto mais próximas de zero, melhor — corroboram as observações visuais. O modelo **VAE** base apresentou os melhores valores de $\overline{\text{MAE}}$ (0,0214) e $\overline{\text{MSE}}$ (0,0092), indicando uma boa capacidade de reconstrução com alta fidelidade das imagens. Em contraste, o **VAE-GAN** obteve valores mais elevados de $\overline{\text{MAE}}$ (0,0642) e $\overline{\text{MSE}}$ (0,0243), sugerindo que, apesar de sua habilidade em gerar imagens visualmente mais detalhadas, o modelo pode não ter sido totalmente refinado, devido ao número limitado de épocas de treinamento (10 épocas). Isso sugere que um treinamento adicional poderia melhorar a qualidade da reconstrução, reduzindo os erros quantitativos. O **VQ-VAE**, com $\overline{\text{MAE}}$ de 0,0357 e $\overline{\text{MSE}}$ de 0,0218, ficou entre o VAE e o VAE-GAN, mostrando um desempenho equilibrado. Esse valor intermediário indica que o VQ-VAE, embora preservando uma boa parte das estruturas das fácies geológicas, ainda apresenta um erro maior quando comparado ao VAE base, mas é mais estável em regiões geomorfológicas complexas, como observado nas imagens reconstruídas na Figura 8.

A Figura 8 ilustra exemplos visuais de reconstruções realizadas por cada um dos

modelos. Observa-se que o VAE-GAN tende a preservar mais detalhes visuais, enquanto o VQ-VAE apresenta maior estabilidade em regiões geomorfológicas complexas.

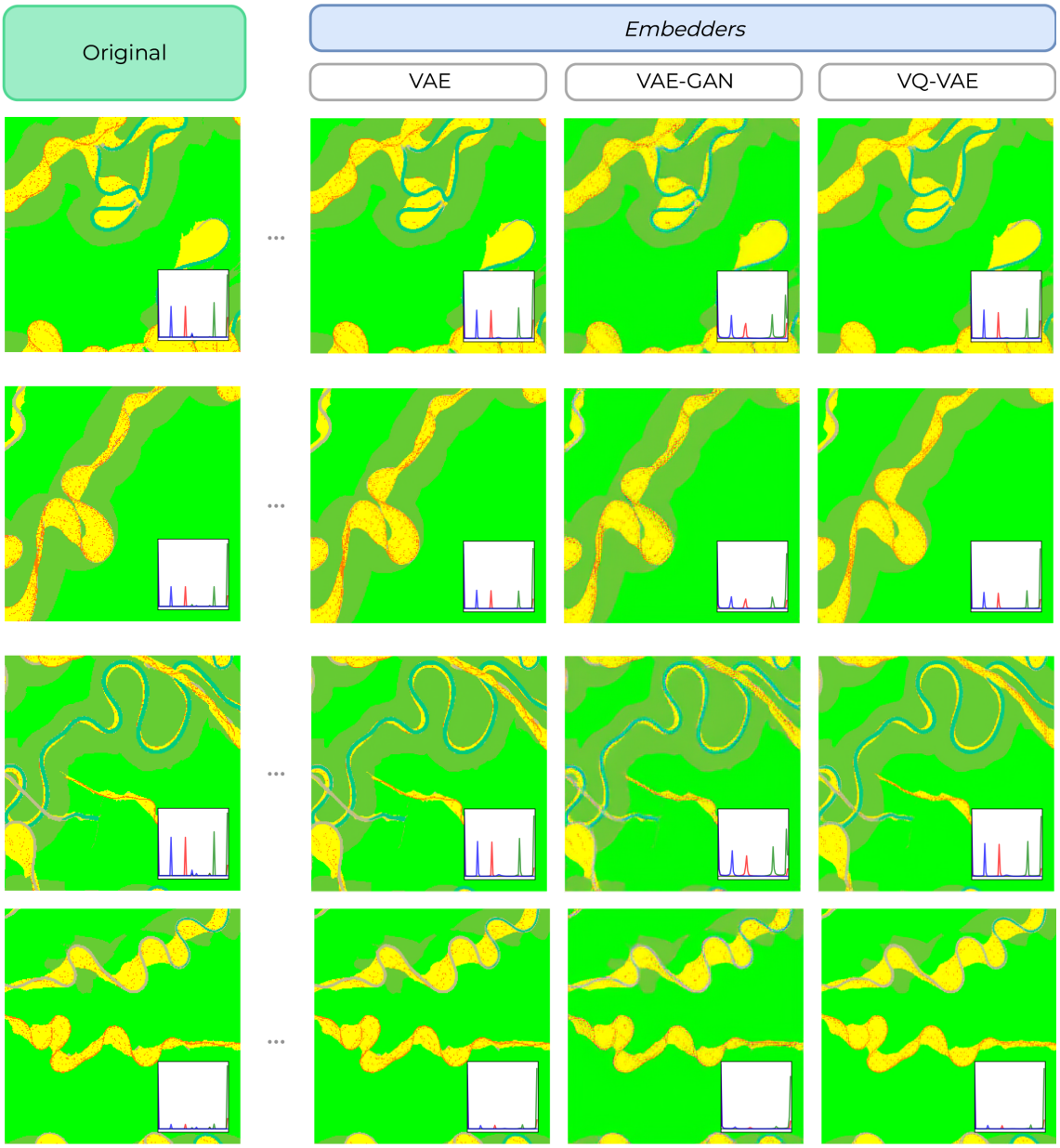


Figura 8. Exemplos de reconstruções de fácies geológicas realizadas por cada um dos *embedders*.

Visualmente, as reconstruções dos três *embedders* apresentam a mesma distribuição geral de fácies — os principais corpos (canais, barras de meandro, tampões) aparecem nas mesmas posições e com semelhança de cores. No entanto, ao examinar detalhes finos, surgem diferenças sutis:

- **VAE:** reconstrói imagens mais suaves, com transições graduais entre fácies. Os contornos das feições são menos definidos, o que reduz ruídos pontuais, mas tende a "borrar" bordas nítidas.

- **VAE-GAN:** introduz texturas adicionais e contornos mais marcados, conferindo aparência visualmente rica; porém, esse ganho de nitidez vem acompanhado de pequenos artefatos e ruídos dispersos, especialmente em áreas homogêneas (note os pontilhados irregulares no interior das fácies na Figura 8).
- **VQ-VAE:** graças à quantização vetorial, mantém bordas mais bem definidas que o VAE puro, mas pode gerar efeitos de "blocos" em regiões geomorfológicas complexas (observe pequenas falhas de continuidade nas transições), sem o excesso de ruído visto no VAE-GAN.

Em suma, embora a fidelidade estrutural global seja equivalente entre os modelos, a escolha do *embedder* afeta o equilíbrio entre suavidade, nitidez e ocorrência de artefatos — informações cruciais para aplicações que exigem detalhes finos em fácies geológicas.

4.2. Resultados do Modelo de Difusão Latente

Após o treinamento dos modelos de *embedding*, cada um deles foi acoplado a um LDM, cuja tarefa foi gerar amostras sintéticas a partir do espaço latente aprendido. A Tabela 2 apresenta os valores obtidos para as métricas quantitativas FID e LPIPS, permitindo comparar o desempenho dos LDMs resultantes de forma objetiva.

Tabela 2. Resultados quantitativos do LDM com diferentes *embedders*.

Embedder	**FID (↓)**	$\overline{\textbf{LPIPS}}$ **(↓)**
VAE	65,0449	0,5643
VAE-GAN	69,7406	0,5886
VQ-VAE	104,6080	0,5610

A Figura 9 detalha a evolução do MAE ao longo das dez épocas de treinamento do LDM. Percebe-se que a combinação do LDM com o VAE-GAN reduz o erro mais rapidamente, enquanto as versões com VAE e VQ-VAE apresentam curvas quase sobrepostas e convergem de forma mais lenta e semelhante entre si.

Ambas as métricas seguem o princípio de que valores mais baixos indicam melhor desempenho: o FID mede a distância entre as distribuições das imagens reais e geradas, no espaço de ativação de uma rede neural treinada; quanto mais próximo de zero, maior a similaridade estatística entre esses conjuntos. Já o LPIPS avalia a similaridade perceptual entre imagens geradas e imagens reais. No presente estudo, como o modelo de difusão latente gera amostras de forma não condicional, cada imagem gerada foi comparada individualmente a todas as imagens do conjunto de teste, e a métrica final foi expressa como a média dessas distâncias perceptuais. Dessa forma, valores mais próximos de zero, em ambas as métricas, indicam maior fidelidade visual e melhor qualidade de geração.

A Figura 10 apresenta exemplos de imagens sintéticas geradas pelos LDMs condicionados por cada arquitetura. As imagens revelam variações no nível de detalhamento estrutural, coerência espacial e diversidade morfológica entre as fácies geradas. Embora todas conservem a distribuição geral de fácies, observam-se diferenças sutis:

- **VAE:** texturas suaves e transições graduais entre fácies; mantém coerência espacial com pouca granulação, mas tende a reproduzir os contornos de forma mais

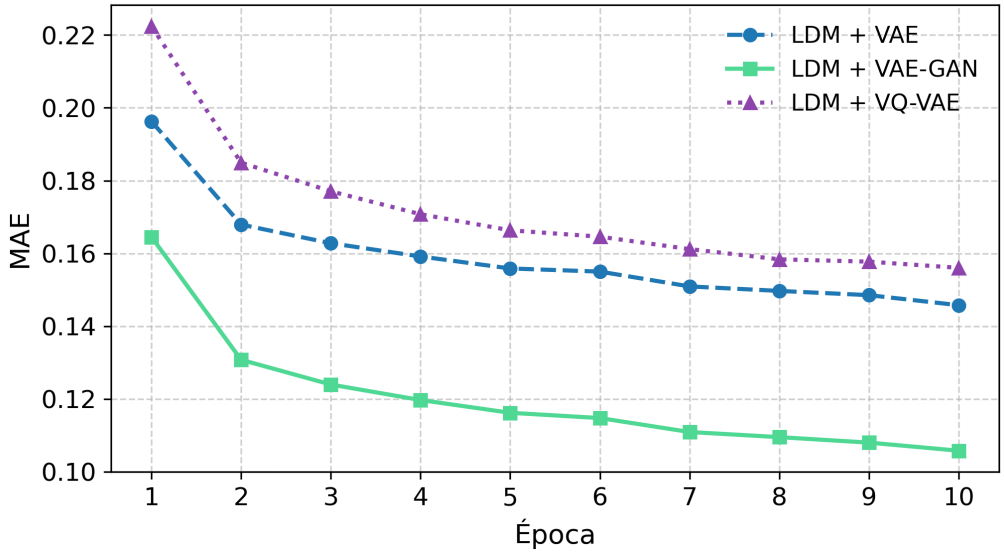


Figura 9. MAE durante o treinamento do LDM com diferentes *embeddings*.

difusa — menor ruído, porém menos definição de detalhes finos e menos complexidade estrutural.

- **VAE-GAN:** realce de bordas, feições geológicas mais nítidas e com mais complexidade estrutural; introduz maior variedade de texturas, mas também pequenos saltos de intensidade (ruídos finos) em áreas homogêneas, refletindo o caráter adversarial do *embedder*.
- **VQ-VAE:** preservação acentuada de contornos; entretanto, apresentou baixa geração de estrutura de fácies, especialmente em estruturas curvilíneas complexas.

4.3. Discussão

Os resultados obtidos evidenciam que a escolha da arquitetura de *embedding* exerce influência direta sobre o desempenho do modelo generativo, baseado em difusão latente para geração de fácies geológicas. O modelo VAE apresentou o melhor desempenho quantitativo nas métricas avaliadas, alcançando um FID de 65,0449, $\overline{\text{MAE}}$ de 0,0214, $\overline{\text{MSE}}$ de 0,0092 e $\overline{\text{LPIPS}}$ de 0,5643. Este resultado reflete a simplicidade relativa dessa arquitetura, permitindo uma convergência mais rápida e eficiente nas 10 épocas de treinamento realizadas.

Por outro lado, o modelo VAE-GAN demonstrou desempenho quantitativo inferior em relação ao VAE, com valores mais elevados de FID de 69,7406, $\overline{\text{MAE}}$ de 0,0642, $\overline{\text{MSE}}$ de 0,0243 e $\overline{\text{LPIPS}}$ de 0,5886. Apesar disso, observações qualitativas revelam que o VAE-GAN foi capaz de gerar imagens visualmente mais detalhadas e complexas, ainda que acompanhadas de ruídos perceptíveis. Acredita-se que esse desempenho intermediário deve-se à maior complexidade da arquitetura adversarial, que exige maior capacidade computacional e, principalmente, um número maior de épocas de treinamento para alcançar resultados satisfatórios.

A arquitetura VQ-VAE apresentou resultados quantitativos intermediários, com uma FID de 104,6080, $\overline{\text{MAE}}$ de 0,0357, $\overline{\text{MSE}}$ de 0,0218 e $\overline{\text{LPIPS}}$ de 0,5610. Visual-

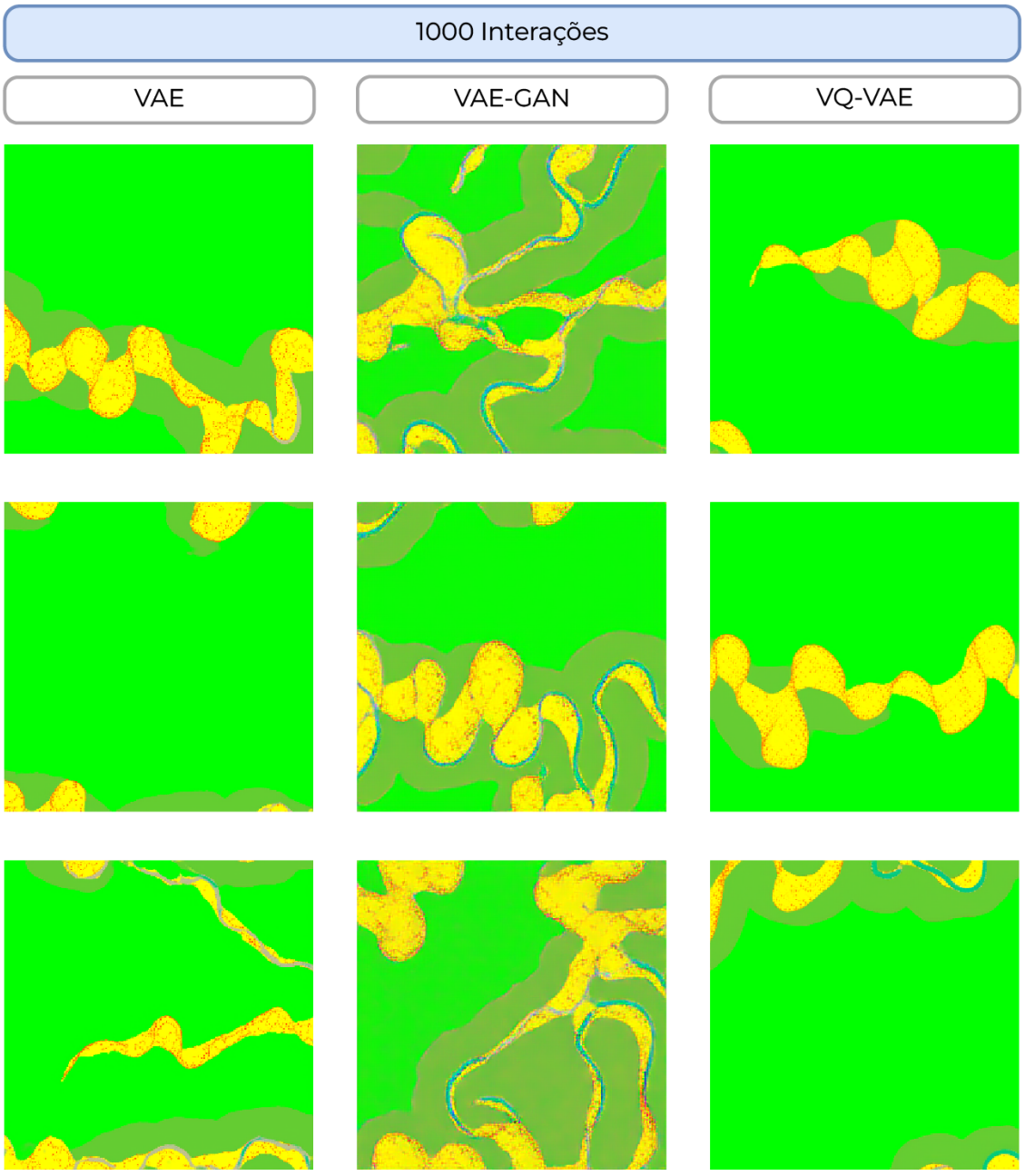


Figura 10. Exemplos de imagens sintéticas geradas com 1000 interações pelo LDM com diferentes *embedders*.

mente, o modelo conseguiu preservar estruturas geomorfológicas importantes, embora tenha apresentado menor fidelidade estrutural que o VAE puro e menor riqueza visual em comparação ao VAE-GAN.

Vale ressaltar que, apesar de o LPIPS do VQ-VAE ter se mostrado comparável ao do VAE, isso não implica necessariamente que sua qualidade perceptual seja superior ou equivalente em todos os aspectos. Devido à natureza do LPIPS, que avalia a similaridade perceptual com base em características profundas extraídas por redes convolucionais, modelos adversariais como o VAE-GAN podem gerar imagens com maior complexidade visual e texturas mais detalhadas, o que pode introduzir ruídos perceptíveis que não são fortemente penalizados pela métrica. Assim, valores mais baixos de $\overline{\text{LPIPS}}$ para o VQ-VAE podem refletir uma suavidade ou menor variação nas imagens geradas, enquanto o VAE-GAN, embora apresente valores ligeiramente mais altos, produz amostras mais ricas visualmente, como observado nas análises qualitativas.

Nesse contexto, fica evidente um compromisso entre complexidade do modelo, número de épocas de treinamento e qualidade das imagens geradas. Modelos mais complexos como o VAE-GAN e o VQ-VAE necessitam de treinamentos mais longos para alcançar seu potencial máximo, enquanto modelos mais simples, como o VAE tradicional, atingem resultados satisfatórios com maior rapidez, embora limitados na geração de detalhes mais sofisticados.

Portanto, em aplicações onde detalhes visuais complexos sejam essenciais, recomenda-se investir em modelos mais complexos, como o VAE-GAN, prolongando o treinamento para obter melhores resultados. Em contrapartida, para situações onde se deseja resultados rápidos, com boa fidelidade estrutural e menor custo computacional, o VAE tradicional mostra-se uma solução prática e eficaz.

5. Conclusão

Este trabalho investigou a geração de fácies geológicas sintéticas, utilizando um LDM e diferentes estratégias de *embedding*: VAE, VAE-GAN e VQ-VAE, treinadas sob condições experimentais semelhantes, com espaço latente de 8 canais.

Os resultados obtidos indicaram que o modelo VAE, devido à sua menor complexidade estrutural, apresentou a melhor performance quantitativa geral, atingindo valores mais baixos nas métricas avaliadas — FID = 65,0449, $\overline{\text{MAE}}$ = 0,0214, $\overline{\text{MSE}}$ = 0,0092 e $\overline{\text{LPIPS}}$ = 0,5643. Em contrapartida, o VAE-GAN, apesar de ter obtido métricas quantitativas menos satisfatórias — FID = 69,7406, $\overline{\text{MAE}}$ = 0,0642, $\overline{\text{MSE}}$ = 0,0243 e $\overline{\text{LPIPS}}$ = 0,5886 —, destacou-se qualitativamente na geração de imagens visualmente mais complexas, porém ainda com ruídos perceptíveis, sugerindo a necessidade de um treinamento mais extenso para aproveitar plenamente sua capacidade representacional. O modelo VQ-VAE apresentou resultados intermediários em termos quantitativos e qualitativos — FID = 104,6080, $\overline{\text{MAE}}$ = 0,0357, $\overline{\text{MSE}}$ = 0,0218 e $\overline{\text{LPIPS}}$ = 0,5610 — indicando boa preservação de estruturas geomorfológicas, ainda que com limitações em detalhes mais finos.

Cabe destacar que, embora o LPIPS tenha apresentado valores similares entre o VAE e o VQ-VAE, essa métrica deve ser interpretada com cautela, pois modelos adversariais, como o VAE-GAN, tendem a gerar imagens com maior complexidade visual e ruídos texturais, características que nem sempre são penalizadas fortemente pelo LPIPS.

Diante disso, conclui-se que o desempenho dos modelos generativos baseados em LDM depende diretamente da complexidade das estratégias de *embedding* adotadas, do número de épocas utilizadas no treinamento e da métrica escolhida para avaliação. Modelos mais complexos exigem maior tempo de treinamento para alcançar seu potencial máximo e, dependendo da métrica avaliada, podem ter suas qualidades visuais melhor capturadas por análises qualitativas complementares.

Apesar dos resultados promissores, algumas limitações devem ser ressaltadas. O ambiente utilizado (GPU NVIDIA RTX 4060) e o treinamento combinado de cada par *embedder* com LDM demandaram um tempo elevado de processamento, restringindo a possibilidade de variação de hiperparâmetros e do número de épocas testados. Esse alto custo computacional, somado ao esforço necessário para desenvolver e integrar os *pipelines* de pré-processamento, treinamento e geração, limitou o escopo experimental e impôs restrições à exploração de cenários adicionais.

5.1. Trabalhos futuros

Para aprofundar e aprimorar os resultados obtidos, sugere-se:

- Treinar os modelos mais complexos, como o VAE-GAN e VQ-VAE, por um número maior de épocas para verificar se os resultados quantitativos e qualitativos melhoram significativamente.
- Avaliar outras arquiteturas de *embedding*, como VQGAN [Esser et al. 2020], para explorar outras combinações potenciais de fidelidade estrutural e complexidade visual.
- Investigar estratégias alternativas de regularização e funções de perda, que possam contribuir para uma melhor convergência e qualidade final dos modelos generativos.

Esses avanços permitirão uma exploração mais aprofundada e completa do potencial de Modelos de Difusão Latente aplicados à geração de fácies geológicas sintéticas, contribuindo de forma significativa para a modelagem geológica de reservatórios.

Referências

[Chehrazi et al. 2011] Chehrazi, A., Rezaee, R., e Rahimpour, H. (2011). Pore-facies as a tool for incorporation of small-scale dynamic information in integrated reservoir studies. *Journal of Geophysics and Engineering*, 8:202–224.

[Chen et al. 2018] Chen, H., Perozzi, B., Al-Rfou, R., e Skiena, S. (2018). A tutorial on network embeddings.

[Esser et al. 2020] Esser, P., Rombach, R., e Ommer, B. (2020). Taming transformers for high-resolution image synthesis.

[Everaert et al. 2023] Everaert, M. N., Fitsios, A., Bocchio, M., Arpa, S., Süsstrunk, S., e Achanta, R. (2023). Exploiting the signal-leak bias in diffusion models.

[Federico e Durlofsky 2024] Federico, G. D. e Durlofsky, L. J. (2024). Latent diffusion models for parameterization and data assimilation of facies-based geomodels.

[Hassanpour e Deutsch 2010] Hassanpour, R. M. e Deutsch, C. V. (2010). An introduction to gridfree objectbased facies modeling. Technical report.

[Heusel et al. 2018] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., e Hochreiter, S. (2018). Gans trained by a two time-scale update rule converge to a local nash equilibrium.

[Ho et al. 2020a] Ho, J., Jain, A., e Abbeel, P. (2020a). Denoising diffusion probabilistic models.

[Ho et al. 2020b] Ho, J., Jain, A., e Abbeel, P. (2020b). Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33.

[Hoover et al. 2023] Hoover, B., Zaengle, D., Mark-Moser, M., Wingo, P., Suhag, A., e Rose, K. (2023). Enhancing knowledge discovery from unstructured data using a deep learning approach to support subsurface modeling predictions. *Frontiers in Big Data*, 6:1227189.

[Kingma e Welling 2013] Kingma, D. P. e Welling, M. (2013). Auto-encoding variational bayes.

[Larsen et al. 2016] Larsen, A. B. L., Sonderby, S. K., Larochelle, H., e Winther, O. (2016). Autoencoding beyond pixels using a learned similarity metric. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, pages 1558–1566.

[Lee et al. 2023] Lee, D., Ovanger, O., Eidsvik, J., Aune, E., Skauvold, J., e Hauge, R. (2023). Latent diffusion model for conditional reservoir facies generation.

[Mascarenhas e Agarwal 2021] Mascarenhas, S. e Agarwal, M. (2021). A comparison between vgg16, vgg19 and resnet50 architecture frameworks for image classification. In *Proceedings of IEEE International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications, CENTCON 2021*, pages 96–99. Institute of Electrical and Electronics Engineers Inc.

[Rombach et al. 2021] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., e Ommer, B. (2021). High-resolution image synthesis with latent diffusion models.

[Ronneberger et al. 2015] Ronneberger, O., Fischer, P., e Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241.

[Ruppel e Harrington 2015] Ruppel, S. C. e Harrington, R. R. (2015). *Facies and Sequence Stratigraphy*, pages 5–48. American Association of Petroleum Geologists and Bureau of Economic Geology.

[Schroff et al. 2015] Schroff, F., Kalenichenko, D., e Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering.

[Shlens 2014] Shlens, J. (2014). Notes on kullback-leibler divergence and likelihood.

[Sun et al. 2023] Sun, C., Demyanov, V., e Arnold, D. (2023). Gan river-i: A process-based low ntg meandering reservoir model dataset for machine learning studies. *Data in Brief*, 46:108785.

[Tan et al. 2013] Tan, H. L., Li, Z., Tan, Y. H., Rahardja, S., e Yeo, C. (2013). A perceptually relevant mse-based image quality metric. *IEEE Transactions on Image Processing*, 22:4447–4459.

[van den Oord et al. 2017] van den Oord, A., Vinyals, O., e Kavukcuoglu, K. (2017). Neural discrete representation learning. *arXiv preprint arXiv:1711.00937*.

[Wang et al. 2023] Wang, D., Ma, C., e Sun, S. (2023). Novel paintings from the latent diffusion model through transfer learning. *Applied Sciences (Switzerland)*, 13.

[Willmott e Matsuura 2005] Willmott, C. e Matsuura, K. (2005). Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate Research*, 30:79.

[Zhang et al. 2018] Zhang, R., Isola, P., Efros, A. A., Shechtman, E., e Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric.

[Zhang et al. 2025] Zhang, Y., Li, C., Li, J., Luo, X., Cheng, M., Zhang, X., e Lu, B. (2025). A new method of geological modeling for the hydrocarbon secondary migration research. *Applied Sciences*, 15:3377.

[Zhu e Zhang 2019] Zhu, L. e Zhang, T. (2019). Generating geological facies models with fidelity to diversity and statistics of training images using improved generative adversarial networks.