

Investigação de técnicas de Deep Learning para melhor eficiência em *NeRF*'s

1 Introdução

A histologia é uma área da biologia que estuda os tecidos biológicos em nível microscópico, sendo fundamental para a compreensão de diversas condições médicas e o desenvolvimento de possíveis tratamentos, diagnósticos e terapias. Tradicionalmente, a análise histológica envolve a coleta de amostras de tecidos, sua preparação e coloração, seguida pela observação microscópica, tradicionalmente computadas em cortes bidimensionais (2D). No entanto, com o avanço das tecnologias informacionais, novas abordagens estão sendo desenvolvidas para melhorar a eficiência das avaliações e inferências sobre os tecidos. Embora cruciais para diagnóstico e pesquisa, as imagens 2D possuem uma limitação para a visualização de arquiteturas e tecidos complexos. Desta forma, a renderização 3D se introduz nesta seara ao facilitar a visualização integrada de tais, como órgãos, vasos sanguíneos ou tecidos orgânicos com perspectivas dinâmicas, aprimorando a compreensão de relações espaciais críticas para diagnósticos e planejamentos cirúrgicos. Além disso, modelos 3D personalizados, gerados a partir de dados de pacientes, podem simular intervenções médicas com maior segurança, servir como ferramenta educativa para treinamento de profissionais e até apoiar a criação de próteses ou implantes sob medida. Existem diversas metodologias e técnicas para se alcançar bons resultados e modelos 3D, seja através da modelagem a partir de um artista ou, de maneira avançada, obter-se modelos a partir das já existentes imagens.

Em dissonância a métodos tradicionais de síntese gráfica, os campos de radiação neural (*NeRF*) utilizam-se de técnicas de renderização volumétricas comuns, apoiando-se nas redes neurais artificiais MLP's (multilayer perceptron) para processar a entrada de coordenadas 5D $((x, y, z, \theta, \phi))$, cada posição compreendida como um vetor 3D de posições (x, y, z) computado com o ângulo e direção de câmera correspondentes (θ, ϕ) em pontos de densidade volumétrica $(\sigma(x))$ e cor $(c = (r, g, b))$, capazes de, posteriormente, se transformarem em volume 3D convencional [1]. Porém, para conseguir modelos de alta confiabilidade, o *NeRF* depende de um conjunto consideravelmente grande de imagens e diferentes rotações dos ângulos da câmera do objeto de referência. Vários autores buscaram (e buscam) aprimorar seus datasets com inúmeras técnicas, sempre buscando qualificar mais e mais os dados e, conseqüentemente, obter imagens mais próximas do *ground-truth*. Por exemplo, pesquisadores utilizaram-se do SFM (Structure from Motion), uma técnica de imagem de alcance fotogramétrico, para se capturar imagens do ambiente em questão com maior precisão e qualidade, como no método de gaussianos 3D [2], ou mesmo o MVS (multi-view stereo), escolhido pelo uso subsequente de CNN's (*convolutional neural network*) [3], tudo isso chegando no estado da arte. Estas técnicas, conquanto fiéis às suas referências, são, ora computacionalmente caras, ora, ainda, demandantes de uma quantidade exacerbada de imagens do item a ser renderizado. *Datasets* histológicos, entretanto, costumam não possuir tamanha diversidade ou redundância de ângulos necessária, devido aos desafios como a complexidade estrutural dos tecidos e restrições éticas na coleta de amostras [4].

Consonantemente, neste trabalho, buscar-se-á possibilidades de se minimizar a distância entre um *dataset* ideal (repleto de redundância e ângulos de câmera distintos) e os datasets comuns na histologia. Para tal, seguindo trabalhos similares, se faz necessária uma investigação sobre a viabilidade de se utilizar técnicas de *Deep Learning*, seja com pré-processamento intensivo dos dados, ou mesmo com a geração de com GAN's (*generative adversarie network*), como feito previamente por Rozendo [5] em imagens bidimensionais. GAN's permitem a criação de imagens através de duas redes neurais treinadas para competição mútua (G - gerador e D - discriminador). G tenta criar imagens que emulam o pertencimento ao domínio de D, enquanto este o avalia, classificando-o como pertencente ou não. Eventualmente, G e D atingem um estado de equilíbrio (chamado de equilíbrio de Nash), no qual, mesmo após maiores treinamentos, a diferença não mais se altera significativamente [6]. Ao utilizar-se destes conceitos, espera-se, portanto, promover a redução da dependência de *datasets* extensos de *NeRF*'s, enquanto mantem-se sua qualidade intrínseca.

2 Objetivos

Neste projeto, tem-se como objetivo investigar e aplicar métodos de geração de imagem para aumentar um *dataset* limitado na histologia e obter-se modelos 3D próximos ao *ground-truth*. Para tal, pretende-se:

- Investigar técnicas de pré-processamento a fim de garantir a aceitabilidade da arquitetura de GAN's e NeRF's aqui utilizada.
- Identificar, dentre diferentes tipos de redes generativas adversárias, bibliograficamente, àquela com maior possível adaptabilidade ao caso de uso deste projeto.
- Avaliar o impacto do aumento de dados com imagens sintéticas (geradas por GANs) na qualidade da reconstrução 3D do NeRF, comparando-o com um modelo treinado apenas com dados reais.
- Implementar um modelo NeRF utilizando como base na implementação original.
- Definir as principais associações e limites observados no contexto aqui explorado.

Considerando o previsto em Edital PROPe Unesp Nº 08/2025 - PIBIC, esta proposta tem aderência plena com as áreas Prioritárias do Ministério da Ciência, Tecnologia, Inovações e Comunicações (estabelecidas na Portaria MCTIC nº 1.122/2020, com texto alterado pela Portaria MCTIC nº 1.329/2020), especificamente com as Áreas de Tecnologias para Qualidade de Vida (Saúde) e Tecnologias Habilitadoras (Inteligência Artificial). No que tange à lista dos Objetivos do Desenvolvimento Sustentável (ODS), esta pesquisa apresenta aderência direta com a ODS 9 (indústria, inovação e infraestrutura).

3 Metodologia

A pesquisa será conduzida em cinco etapas principais, dispostas a seguir.

3.1 Etapa 1 - Revisão Bibliográfica

Esta etapa permite o levantamento bibliográfico necessário para manter o projeto atualizado, proporcionar uma fundamentação teórica sólida e subsidiar a exploração proposta. A revisão focará em: (i) arquiteturas de Redes Adversárias Generativas (GANs) aplicadas à síntese de imagens médicas, com ênfase em GANs 3D-Aware (e.g., EG3D, StyleSDF) e Condicionais (cGANs), que permitem o controle sobre atributos da imagem como a pose da câmera; (ii) técnicas de aumento de dados para o treinamento de Campos de Radiância Neural (NeRF); e (iii) métricas de avaliação de estado-da-arte para modelos generativos 2D e de reconstrução 3D.

3.2 Etapa 2 - Pré-processamento *dataset* de imagens histológicas

Dado conjunto de dados de imagens histológicas [nome] a ser utilizado, observa-se 2 desafios principais: a necessidade de pré-processamento e padronização das imagens e a organização do conjunto de dados em segmentos distintos. Estes, a seguir, serão os segmentos tratados nesta etapa:

- Os NeRF's requerem os dados sobre a pose da câmera para cada imagem, ou seja, a posição e orientação da câmera no momento da captura da imagem. Para tal, é necessário que as imagens sejam organizadas em grupos, os quais contenham, individualmente, imagens de um único tecido e pose, em específico, sendo utilizado o *Nerfstudio* para permitir o *parsing* dos dados de entrada.
- O pré-processamento subsequente incluirá a normalização de cores e resolução, assim como a possível remoção de artefatos.

- Além disso, para efeito de treinamento do modelo, é necessário que sejam estabelecidos sub-conjuntos de imagens, cada qual usado em diferentes contextos. Serão criados 3 conjuntos principais: **Treinamento**, **Validação** e **Teste**.

Os conjuntos provenientes desta etapa serão utilizados na etapa 3 e na etapa 4.

3.3 Etapa 3 - Treinamento e Geração de imagens

Nesta etapa, um modelo GAN será treinado para gerar novas imagens histológicas a partir do conjunto de dados pré-processado. O modelo GAN deverá aprender as características dos tecidos histológicos presentes no *dataset*, permitindo, assim, a sintetização de imagens que preservem as propriedades visuais e estruturais dos tecidos.

Em especial, serão investigados os modelos DCGAN, RAGAN e WGAN-GP, evidenciados no trabalho do Botazzo [5]. Estas GAN's assimilam o problema da não competição gerada pelo discriminador (D), que usualmente tem vantagem acerca do gerador (G), cada um da sua maneira. Para se endereçar o problema das poses de câmera não existentes no conjunto de dados, será investigado o uso das cGAN's (*Conditional Generative Adversary Network*), cujo objetivo é conseguir produzir novos tipos de ângulos e direções de câmera para a imagem referente. Diferentemente de uma GAN tradicional, que gera amostras de forma não controlada, a cGAN introduz um mecanismo de condicionamento. Tanto o gerador (G) quanto o discriminador (D) recebem uma informação adicional, y — neste caso, o vetor de pose da câmera. Assim, o gerador não aprende apenas a criar imagens realistas, mas a criar imagens que correspondam à condição de pose fornecida. O discriminador, por sua vez, aprende a validar se a imagem gerada é autêntica e se corresponde à pose y , forçando o gerador a respeitar o controle de entrada. No entanto, possíveis novas arquiteturas podem ser utilizadas a depender do resultado da Etapa 1.

Para a avaliação das imagens sintetizadas e posterior inserção nos *datasets* de treinamento e validação, será utilizada a métrica conhecida como *Fréchet Inception Distance* (FID) [7], que mede a distância entre as distribuições de características de imagens reais e sintéticas modelando-as como distribuições Gaussianas.

Desta forma, um dentre os GAN's supracitados será escolhido de acordo com o melhor desempenho para o treinamento e a geração de novas imagens de alta qualidade para serem, posteriormente utilizadas na etapa 4, ao serem inseridas no conjunto de treinamento e validação.

3.4 Etapa 4 - Treinamento do modelo NeRF e Criação dos volumes 3D

O objetivo central desta etapa é avaliar o impacto do aumento de dados com GANs na qualidade da reconstrução 3D do NeRF. Para isso, serão conduzidos experimentos controlados:

- **Baseline:** Um modelo NeRF será treinado utilizando apenas o conjunto de treinamento original (real).
- **Aumento de Dados:** Um segundo modelo NeRF será treinado com o conjunto de treinamento original aumentado com as imagens sintéticas de alta qualidade geradas na Etapa 3. A proporção de dados sintéticos será um hiperparâmetro a ser explorado.

Para permitir a implementação do Nerf, será utilizado como base o código oficial, que se utiliza especialmente da biblioteca *python* conhecida como *tensorflow* e otimizações de GPU para viabilizar o intenso processamento. Ao se utilizar deste código, hiperparâmetros poderão ser retirados e otimizados, buscando o melhor resultado para os modelos 3D.

Desta forma, poder-se-á compreender a relação entre as imagens 2D e a reconstrução 3D dos tecidos histológicos. A partir do modelo NeRF treinado, serão obtidos volumes 3D dos tecidos histológicos, permitindo a visualização das estruturas internas dos tecidos com alta fidelidade. Estes volumes produzidos serão futuramente avaliados quanto à sua qualidade e fidelidade em relação às imagens originais do *dataset* na etapa 5.

3.5 Etapa 5 - Validação dos resultados

Para obter-se resultados precisos, factíveis e replicáveis, este projeto buscará utilizar-se de métricas de validação dos modelos 3D reconstruídos pela análise de poses específicas, ou seja, imagens 2D tiradas sobre o modelo.

Serão empregados tanto o PSNR (*Peak Signal-To-Noise Ratio*) quanto o SSIM (*Structural Similarity Index*). O PSNR faz uma comparação direta entre 2 imagens e consegue descrever o ruído discrepante entre estas, utilizando-se de uma razão sinal-ruído em escala logarítmica. Já o SSIM trabalha com a estrutura da imagem em si, ou seja, nos agrupamentos dos pixels da referência quando comparados à imagem gerada pela arquitetura GAN, capturando melhor a percepção humana de qualidade visual do que o PSNR. Devido à inacessibilidade a um *modeloground-truth* para o domínio das reconstruções 3D, utilizar-se-á a métrica LPIPS (*Learned Perceptual Image Patch Similarity*), que compara as imagens no espaço de características de uma rede neural profunda, fornecendo uma medida de distância que se alinha de forma mais robusta com a percepção visual humana de similaridade.

Essas métricas fornecerão uma base quantitativa robusta para avaliar e comparar a qualidade dos modelos testados, dimensionando uma alta fidelidade ao obter-se valores elevados de PSNR e SSIM e de LPIPS.

A metodologia seguirá o fluxo de trabalho a seguir, uma *pipeline* estruturada para melhor entendimento das etapas, entradas e saídas:

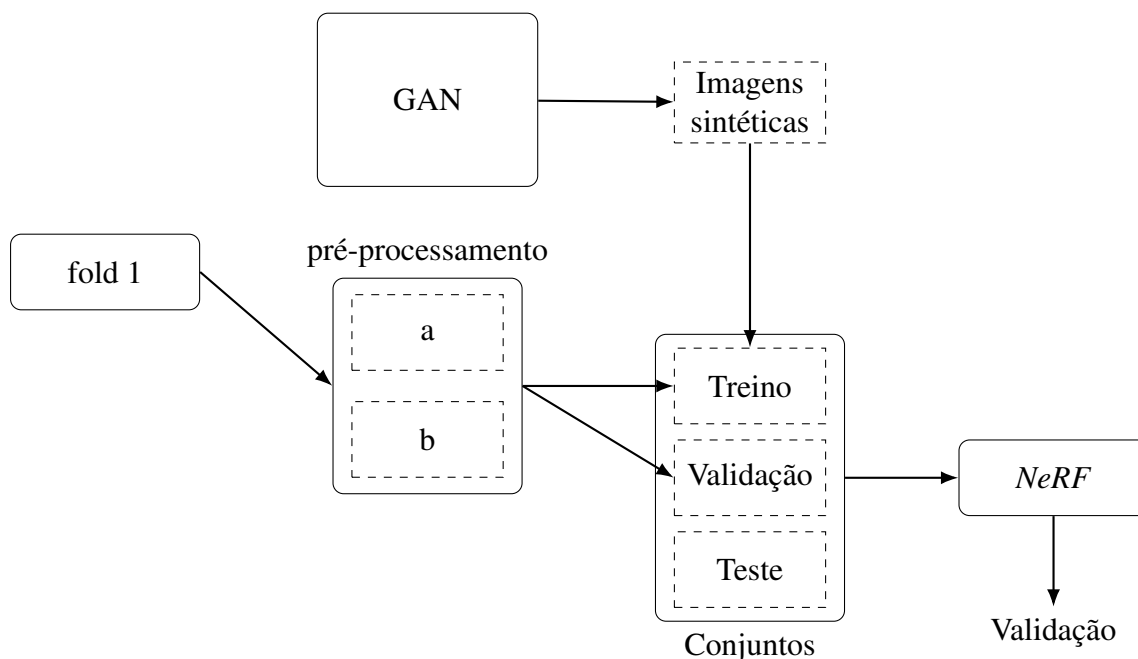


Figura 1: Pipeline da metodologia

4 Plano de Trabalho e Cronograma de Execução

Este projeto é parte de outros projetos que estão em pleno desenvolvimento pelo orientador. O plano de trabalho consiste em realizar as etapas descritas na seção: Etapa 1 - Revisão Bibliográfica; Etapa 2 - Pré-processamento *dataset* de imagens histológicas; Etapa 3 - Treinamento e Geração de imagens; Etapa 4 - Treinamento do modelo NeRF e Criação dos volumes 3D; Etapa 5 - Validação dos resultados. O cronograma de execução proposto para 12 meses está na Tabela 1.

Tabela 1: Proposta de Cronograma para o atendimento das etapas previstas no escopo do projeto, considerando o período de 12 meses.

Etapas	Ano 1: 2025-2026											
	1	2	3	4	5	6	7	8	9	10	11	12
1	x	x	x						x			
2	x	x	x									
3			x	x	x	x						
4					x	x	x					
5						x	x	x	x	x		
Escrita: Processos e Resultados		x		x		x		x		x	x	x

5 Resultados Esperados

Este projeto visa investigar a aplicabilidade de métodos convencionais de *Deep Learning* em dados histológicos, ao gerar imagens histológicas sintéticas e reconstruir volumes 3D a partir dessas imagens. Os resultados esperados incluem a avaliação da qualidade dessas imagens e dos volumes 3D reconstruídos e a contribuição para a área de pesquisa em imagens médicas, ao concluir a viabilidade de se utilizar GAN associados a NeRF.

Referências

- [1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” 2020.
- [2] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, “3d gaussian splatting for real-time radiance field rendering,” *ACM Transactions on Graphics*, vol. 42, July 2023.
- [3] A. Chen, Z. Xu, F. Zhao, X. Zhang, F. Xiang, J. Yu, and H. Su, “Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo,” 2021.
- [4] Y. Xue, J. Ye, Q. Zhou, L. R. Long, S. Antani, Z. Xue, C. Cornwell, R. Zaino, K. C. Cheng, and X. Huang, “Selective synthetic augmentation with histogan for improved histopathology image classification,” *Medical Image Analysis*, vol. 67, p. 101816, 2021.
- [5] G. B. Rozendo, B. L. d. O. Garcia, V. A. T. Borgue, A. Lumini, T. A. A. Tosta, M. Z. d. Nascimento, and L. A. Neves, “Data augmentation in histopathological classification: An analysis exploring gans with xai and vision transformers,” *Applied Sciences*, vol. 14, no. 18, 2024.
- [6] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” 2014.
- [7] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “Gans trained by a two time-scale update rule converge to a local nash equilibrium,” 2018.