

Segmentação de Núcleos em Imagens Citológicas do Colo do Útero com Pix2Pix e U-Net

João Pedro Lobato de Pinho¹, Alexei Manso Correa Machado^{1,2}

¹Departamento de Ciência da Computação

Pontifícia Universidade Católica de Minas Gerais (PUC Minas), Belo Horizonte, Brasil

²Departamento de Anatomia e Imagem - Faculdade de Medicina

Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brasil

emails joao.pinho.1275438@sga.pucminas.br, alexeimcmachado@gmail.com

Abstract. *The segmentation of cell nuclei is essential for the analysis of cytological images and the early detection of diseases such as cervical cancer. This work compares the U-Net and Pix2Pix architectures in the task of semantic segmentation of cervical cell nuclei. Pix2Pix is a conditional GAN that uses adversarial learning. The experiments were conducted using the CNSeg dataset, which contains cytological images in various contexts. U-Net achieved better performance in AJI (+2.90%) and PQ (+3.12%) metrics. However, Pix2Pix showed promising results by producing accurate segmentations. These findings suggest that adversarial networks can be a promising alternative for this task, which remains underexplored in the literature.*

Resumo. *A segmentação de núcleos celulares é essencial para a análise de imagens citológicas e a detecção precoce de doenças como o câncer cervical. Este trabalho compara as arquiteturas U-Net e Pix2Pix na tarefa de segmentação semântica de núcleos de células do colo do útero. O Pix2Pix é uma GAN condicional que utiliza aprendizado adversarial. Os experimentos foram realizados com a base de dados CNSeg, composta por imagens citológicas em diferentes contextos. A U-Net teve desempenho superior nas métricas AJI (+2,90%) e PQ (+3,12%). No entanto, o Pix2Pix mostrou-se promissor, produzindo segmentações precisas. Isso indica que redes adversariais podem ser uma alternativa promissora para essa tarefa, ainda pouco explorada na literatura.*

1. Introdução

O câncer cervical é uma neoplasia que frequentemente leva à morte de muitas mulheres. Embora seja um dos tipos de câncer mais preveníveis, ele permanece como uma das principais causas de mortes por câncer em mulheres de 20 a 39 anos [Siegel et al. 2024]. A detecção precoce, geralmente por meio do exame de Papanicolau, desempenha um papel crucial na prevenção e tratamento efetivo dessa doença. Contudo, a análise dos exames citológicos é realizada de forma manual, um processo complexo, subjetivo e demorado, que depende de profissionais qualificados.

Nas últimas décadas, o uso de algoritmos de aprendizado profundo demonstrou grande potencial para automatizar a análise de exames, especialmente na segmentação de núcleos de células. A segmentação precisa desses núcleos é fundamental para se extraírem características citológicas detalhadas, que auxiliam na identificação de células atípicas.

Contudo, as particularidades das imagens, como a sobreposição e a variação de formas dos núcleos, tornam essa tarefa desafiadora, exigindo o desenvolvimento de modelos de segmentação avançados.

Arquiteturas de redes neurais convolucionais, como a U-Net, são amplamente utilizadas para segmentação em imagens médicas devido à sua eficácia em problemas de segmentação semântica [Ronneberger et al. 2015]. Recentemente, redes adversariais generativas (GANs) têm sido aplicadas com sucesso em áreas como segmentação de imagens de retina e ressonância magnética, demonstrando seu potencial em gerar segmentações mais detalhadas [Ali et al. 2024]. No entanto, a aplicação de GANs na segmentação de núcleos de células do colo do útero é pouco explorada, sugerindo espaço para novas investigações.

O objetivo deste trabalho é apresentar uma discussão sobre as potencialidades da arquitetura de GAN condicional Pix2Pix na segmentação semântica de núcleos de células do colo do útero, utilizando a base de dados CNSeg [Zhao et al. 2023]. O trabalho busca comparar o desempenho do Pix2Pix com o da arquitetura U-Net tradicional, amplamente utilizada na área, tanto de forma qualitativa quanto quantitativa. A análise se concentra em avaliar como o Pix2Pix, por meio de sua abordagem adversarial, pode gerar segmentações mais precisas, especialmente em casos complexos.

2. Trabalhos Relacionados

As GANs foram introduzidas em 2014, trazendo uma nova abordagem para o aprendizado profundo através de um treinamento competitivo entre duas redes neurais: uma geradora, que cria amostras semelhantes aos dados reais, e uma discriminadora, que tenta distinguir as amostras geradas das reais [Goodfellow et al. 2014]. Esse processo é formalizado como um problema de minimax, onde o gerador G busca minimizar a habilidade do discriminador D em identificar as amostras falsas, enquanto o discriminador tenta maximizar sua capacidade de diferenciação. No mesmo ano, Mirza e Osindero introduziram as GANs condicionais (cGANs), que incorporaram informações adicionais, como rótulos ou imagens de referência, na entrada tanto do gerador quanto do discriminador, orientando o processo de geração de acordo com um objetivo específico [Mirza and Osindero 2014]. Essa modificação ampliou as possibilidades das GANs, tornando-as particularmente eficazes para tarefas que exigem um controle mais direto sobre o processo de geração, como a segmentação de imagens.

A partir de 2016, quando as GANs passaram a ser aplicadas no problema de segmentação semântica, [Luc et al. 2016] propuseram uma abordagem pioneira em que uma rede adversarial foi integrada a uma rede de segmentação para discriminar entre mapas de segmentação reais e preditos, corrigindo inconsistências e melhorando a precisão dos resultados nos conjuntos de dados Stanford Background e PASCAL VOC 2012. Já em 2018, a arquitetura SegGAN, proposta por [Zhang et al. 2018], combinou uma rede GAN com um modelo de segmentação semântica pré-treinado, em que a GAN refina as máscaras de segmentação para uma correspondência mais fiel com as imagens originais. A arquitetura emprega o modelo DeepLab para a segmentação inicial e integra uma rede de gerador e discriminador para o refinamento, mostrando melhoria na segmentação nos mesmos conjuntos de dados.

A aplicação de GANs em segmentação no domínio de imagens médicas também

se destacou nesse período. Em 2018, foi proposta a SegAN [Xue et al. 2018], uma arquitetura de segmentação para imagens médicas que usa uma rede completamente convolucional como segmentadora e uma rede crítica adversarial que, com uma função de perda multi-escala, reforça o aprendizado de relações espaciais locais e globais. A SegAN foi avaliada no conjunto de dados MICCAI BRATS, composto por imagens de ressonância magnética de tumores cerebrais de múltiplas modalidades, mostrando-se eficaz para segmentar regiões tumorais como tumor completo, núcleo tumoral e núcleo realçado com contraste. No trabalho de [Mahmood et al. 2020], uma cGAN foi utilizada para a segmentação de núcleos em imagens de múltiplos órgãos, integrando dados sintéticos e reais para o treinamento da rede. Essa abordagem ofereceu maior precisão na segmentação de núcleos individuais e sobrepostos em comparação com métodos convencionais como U-Net, Mask R-CNN e FCN.

A arquitetura RetinaGAN proposta por [Son et al. 2019] introduziu um modelo cGAN para a segmentação de vasos da retina e do disco óptico, apresentando ganhos em métricas de avaliação como AU-ROC, AU-PR e Dice em comparação com abordagens baseadas em CNN tradicionais como a U-Net e trabalhos anteriores. De forma complementar, [Park et al. 2020] propuseram uma nova arquitetura de GAN condicional, denominada M-GAN, voltada para a segmentação de vasos sanguíneos retinais. A M-GAN emprega uma estrutura de gerador e discriminador com blocos residuais profundos e conexões de *pooling multi-kernel* para suportar a invariância de escala dos vasos sanguíneos, segmentando vasos de diferentes espessuras e tamanhos. Avaliada em conjuntos de dados como DRIVE, STARE, HRF e CHASE-DB1, a M-GAN demonstrou-se superior em métricas como IoU, F1 e MCC, com precisão aumentada na segmentação de vasos finos e redução de falsos negativos. Além desses trabalhos, [Popescu et al. 2021] apresentaram uma abordagem baseada na arquitetura Pix2Pix também para segmentação de vasos em imagens de retina, utilizando imagens dos conjuntos de dados CHASE, DRIVE e STARE. Mesmo o modelo não sendo projetado e otimizado para esse tipo de segmentação, uma acurácia média de 0,93 foi alcançada e demonstrou alta precisão e especificidade, com baixos valores de falsos positivos.

Trabalhos recentes também exploraram a versatilidade das redes adversariais em outros domínios. Um exemplo é o modelo proposto por [Khaled et al. 2022] para segmentação de estruturas cerebrais em ressonâncias magnéticas, que adota uma abordagem de múltiplos estágios e utiliza uma arquitetura GAN para refinar contornos de tecidos como substância branca, substância cinzenta e fluido cerebrospinal. Em outra aplicação, [van der Schot et al. 2023] utilizaram a arquitetura Pix2Pix para segmentação de vasos placentários em imagens fetoscópicas, superando a U-Net em algumas métricas.

3. Materiais

O *dataset* utilizado neste trabalho, CNSeg, foi desenvolvido especificamente para a segmentação de núcleos de células cervicais e contém anotações de 124.353 núcleos extraídos de amostras de 1.530 pacientes [Zhao et al. 2023]. Esse *dataset* é dividido em três *subdatasets* principais — PatchSeg, ClusterSeg e DomainSeg — cada um projetado para representar diferentes níveis de complexidade e variação visual, permitindo uma avaliação abrangente dos modelos em diversos cenários.

O PatchSeg consiste em 3.487 imagens de 512×512 pixels, sendo 477 delas des-

tinadas ao conjunto de teste. Essas imagens capturam uma grande diversidade celular, incluindo células queratinizadas, infecções microbianas, manchas escuras e células com bordas pouco definidas. Ao todo, o PatchSeg contém 85.882 núcleos anotados.

O ClusterSeg é composto por 1.795 imagens de treino e 567 imagens de teste, totalizando 34.311 núcleos anotados. As imagens variam em tamanho, de 150 a 400 pixels, e são caracterizadas por aglomerações celulares e núcleos sobrepostos. O conjunto de teste é subdividido em duas categorias: Normal e Difícil. O conjunto Difícil contém imagens com núcleos mais intensamente sobrepostos.

Por fim, o DomainSeg permite a avaliação do desempenho dos modelos em domínios visuais desconhecidos, e é dividido em dois conjuntos: TargetA e TargetB. O TargetA contém 381 imagens, das quais 145 são reservadas para teste, apresentando núcleos agrupados com variações visuais distintas em relação ao ClusterSeg. O TargetB inclui 332 imagens, sendo 57 de teste, caracterizadas por alta complexidade, núcleos sobrepostos e borradinhos. Uma amostra representativa de cada um desses *subdataset* é exibida na Figura 1, que ilustra as variações visuais e os desafios específicos de cada conjunto.

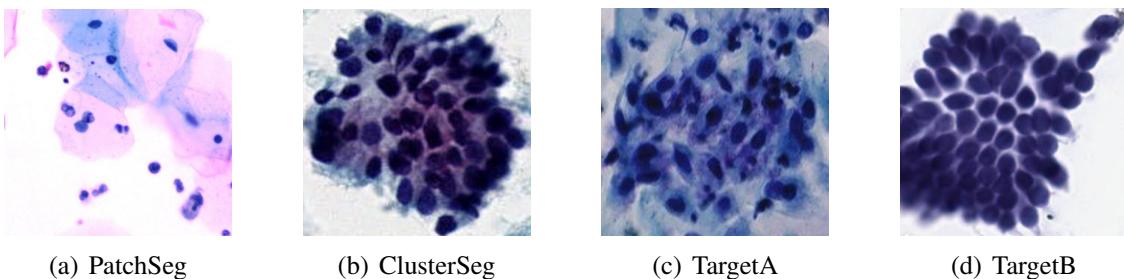


Figura 1. Amostra de imagens dos subdataset utilizados no estudo [Zhao et al. 2023].

4. Métodos

Inicialmente o modelo U-Net foi treinado utilizando-se a base de dados CNSeg, com o objetivo de garantir a comparabilidade com resultados publicados na literatura [Zhao et al. 2023]. A implementação da U-Net seguiu a abordagem clássica proposta por [Ronneberger et al. 2015], consistindo em uma arquitetura do tipo codificador-decodificador. O modelo é composto por camadas de convolução, *pooling* e *upsampling*. O processo de *downsampling* (codificação) é realizado através de camadas de convolução seguidas de operações de *max pooling*, que reduzem progressivamente a resolução espacial da imagem, ao mesmo tempo em que extraem características importantes. As operações de convolução aplicadas têm a função de ativação ReLU, e são configuradas com *stride* e *padding* igual a 1. No caminho do *upsampling* (decodificação), a resolução da imagem é restaurada utilizando operações de *upsample* e convoluções, com a utilização de *skip connections* para preservar informações espaciais dos estágios iniciais da rede. A função de ativação utilizada na camada de saída é a sigmoide. A arquitetura do modelo U-Net é ilustrada na Figura 2, que destaca suas principais operações de convolução, *pooling*, *upsampling* e as *skip connections* que conectam as camadas do codificador e decodificador.

Em seguida, foi implementado o modelo Pix2Pix descrito em [Isola et al. 2017], que possui uma arquitetura composta por dois submodelos principais: o gerador e o dis-

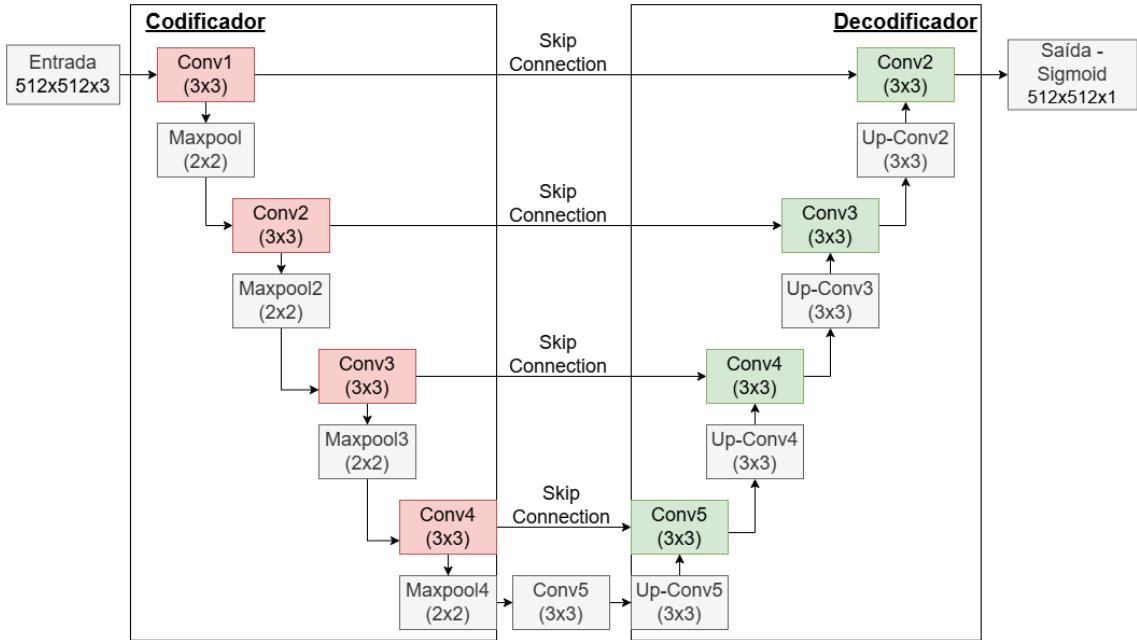


Figura 2. Arquitetura da U-Net apresentada por [Zhao et al. 2023].

crimador. O gerador adota uma arquitetura U-Net, que processa a imagem de entrada (redimensionada no pré-processamento para $512 \times 512 \times 3$) em uma série de operações de convolução, *downsampling* e *upsampling*. Durante a fase de *downsampling*, são aplicadas convoluções de tamanho 4×4 com a função de ativação Leaky ReLU, reduzindo progressivamente a resolução da imagem e extraíndo características importantes. Na camada *bottleneck*, a imagem atinge uma representação comprimida, e o processo é então invertido na fase de *upsampling*, onde convoluções transpostas com a função de ativação ReLU restauram gradualmente a resolução da imagem. O uso de (*skip connections*) transfere diretamente informações entre camadas correspondentes das fases de *downsampling* e *upsampling*. A camada de saída do gerador utiliza a função de ativação *tanh*.

O discriminador utilizado é o PatchGAN com tamanho 70×70 , que avalia pequenos *patches* da imagem em vez de processá-la por inteiro, permitindo uma análise mais detalhada e a captura de padrões locais. A arquitetura do PatchGAN é composta por uma série de camadas convolucionais, com aumento progressivo no número de filtros para extrair características mais complexas. Cada camada usa a função de ativação Leaky ReLU, e a última camada convolucional gera um mapa de previsão, indicando se cada região da imagem é real ou gerada. Com essa abordagem, o PatchGAN atua como uma função de perda adversarial que força o gerador a criar saídas mais realistas e detalhadas, aprimorando a qualidade da imagem gerada [Isola et al. 2017]. A figura 3 mostra a disposição do fluxo de treinamento com a arquitetura completa do modelo.

Para a avaliação dos modelos, foram utilizadas as mesmas métricas descritas pelos autores da base de dados CSeg[Zhao et al. 2023], que incluem o *Aggregated Jaccard Index* (AJI), *Aggregated Dice* (Adice), *Panoptic Quality* (PQ) e *Hausdorff Distance* (HD). O *Aggregated Jaccard Index* (AJI) é uma métrica específica para segmentação de núcleos, baseada na correspondência entre os contornos reais e previstos, computando o índice de Jaccard de forma agregada. Já o *Aggregated Dice* (Adice) é inspirado no AJI, mas cal-

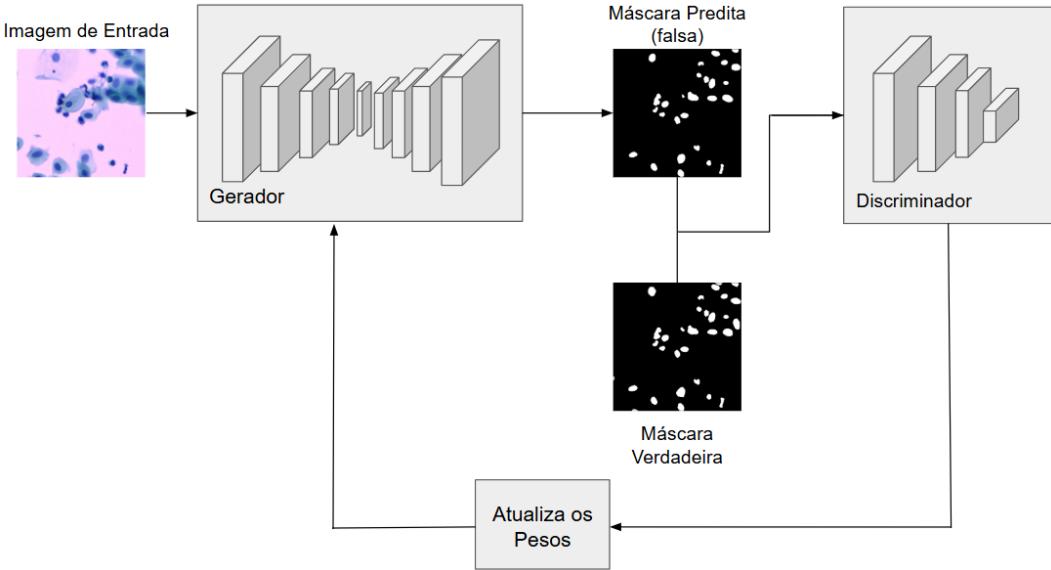


Figura 3. Arquitetura do gerador e discriminador Pix2Pix implementado.

cula o índice de Dice para cada par de contornos correspondentes e soma seus valores. A *Panoptic Quality* (PQ) avalia a consistência entre objetos segmentados individualmente, combinando aspectos de precisão e completude. Por fim, a *Hausdorff Distance* (HD) mede a maior distância entre os contornos reais e previstos, capturando discrepâncias geométricas significativas. Além dessas, também foram empregadas outras métricas, como precisão e *recall*, calculadas com a biblioteca scikit-learn, para fornecer uma análise mais detalhada do desempenho dos modelos.

5. Experimentos

Os experimentos foram realizados em dois ambientes distintos, devido ao grande volume de treinamento e testes necessários. O treinamento da U-Net foi conduzido em uma GPU Nvidia A100 e uma CPU com 40 GB de memória RAM. Os hiperparâmetros utilizados no trabalho original [Zhao et al. 2023] foram mantidos: tamanho de batch de 4, taxa de aprendizado de 2×10^{-4} , máximo de 200 épocas, função de perda *binary cross entropy* e otimizador Adam. Além disso, foram aplicadas as mesmas técnicas de aumento de dados, incluindo transformações geométricas, inversões axiais e cortes aleatórios.

Para o treinamento do modelo Pix2Pix, foi utilizado um processador AMD Ryzen 7 5700X 3.40GHz, GPU Nvidia RTX 4070 SUPER e 16 GB de memória RAM. A implementação foi realizada na linguagem Python (versão 3.10.12) com a biblioteca PyTorch (versão 2.4.0+cu121), e as métricas de avaliação foram calculadas utilizando a biblioteca scikit-learn (versão 1.5.1). O Pix2Pix foi treinado com os mesmos hiperparâmetros da U-Net, ou seja, 200 épocas, taxa de aprendizado de 2×10^{-4} , otimizador Adam e um valor de λ igual a 100, sendo λ um hiperparâmetro que ajusta o peso do termo L1 na função de perda do modelo:

$$L_{\text{pix2pix}} = \arg \min_G \max_D L_{\text{cGAN}}(G, D) + \lambda L_{L1}(G). \quad (1)$$

A função de perda do modelo Pix2Pix combina dois componentes principais: a

perda adversarial condicional (L_{cGAN}) e a perda L1, ponderadas pelo hiperparâmetro λ . O gerador (G) tenta minimizar a função de perda total, buscando gerar imagens indistinguíveis das reais, enquanto o discriminador (D) tenta maximizar a perda adversarial, distinguindo as imagens reais das geradas. Já o termo $L_{L1}(G)$ avalia a diferença absoluta entre os pixels das imagens geradas e das reais. O parâmetro λ ajusta a influência relativa entre essas duas perdas, equilibrando a qualidade adversarial do modelo e a similaridade estrutural das imagens geradas em relação às reais.

Os experimentos seguiram a mesma abordagem de testes utilizada pelos autores da base de dados [Zhao et al. 2023]. O primeiro experimento consistiu no treinamento do modelo utilizando o *subdataset* PatchSeg, seguido pelo teste no respectivo conjunto de teste. Em seguida, foi realizado o treinamento no *subdataset* ClusterSeg, com avaliações tanto no conjunto de teste completo quanto nos conjuntos separados, denominados Normal e Difícil. Por fim, o *subdataset* ClusterSeg foi utilizado como base de treino para os testes nos *subdatasets* TargetA e TargetB, a fim de testar a generalização do modelo.

6. Resultados

Nesta seção, são apresentados os resultados obtidos a partir dos testes realizados com os modelos U-Net e Pix2Pix. O objetivo desta análise é avaliar tanto de forma qualitativa quanto de forma quantitativa as segmentações geradas pelos modelos e comparar seu desempenho em diferentes *subdatasets* da base de dados. O treinamento dos modelos levou cerca de 5 horas e meia para a U-Net e 7 horas para o Pix2Pix. Devido ao alto custo computacional e ao tempo de execução necessário para treinar ambos os modelos, não foi possível realizar validação cruzada.

Os resultados quantitativos são apresentados em termos das métricas descritas na seção de Métodos, incluindo o *Aggregated Jaccard Index* (AJI), *Aggregated Dice* (Adice), *Panoptic Quality* (PQ), *Hausdorff Distance* (HD), e métricas de segmentação como precisão e *recall*.

Em seguida, os resultados são organizados de acordo com cada *subdataset* utilizado nos testes, PatchSeg, os conjuntos Normal e Difícil do *subdataset* ClusterSeg, e os *subdatasets* TargetA e TargetB, que representam diferentes domínios de validação.

Tabela 1. Métricas de desempenho no subdataset PatchSeg para U-Net e Pix2Pix

Modelo	AJI	PQ	Adice	HD	Precisão	Recall
U-Net	0,64	0,62	0,77	17,44	0,84	0,82
Pix2Pix	0,64	0,61	0,78	17,14	0,86	0,78

6.1. PatchSeg

Os modelos U-Net e Pix2Pix exibem desempenhos semelhantes no *subdataset* PatchSeg, conforme ilustrado nas métricas apresentadas na Tabela 1. Praticamente não houve diferença entre as métricas, ambos os modelos obtiveram valores bastante próximos nas métricas AJI, PQ e Adice, com a U-Net superando ligeiramente o Pix2Pix em termos de PQ (0,62 vs 0,61), enquanto o Pix2Pix obteve uma leve vantagem sobre a U-Net na métrica Adice (0,77 vs. 0,76). Embora o Pix2Pix tenha alcançado uma precisão mais alta (0,85 contra 0,84 da U-Net), o modelo U-Net apresentou melhores resultados em termos de *recall*. A Tabela 2 exibe uma comparação visual das segmentações geradas pelos dois modelos.

Tabela 2. Comparativo entre as previsões realizadas no subdataset PatchSeg.

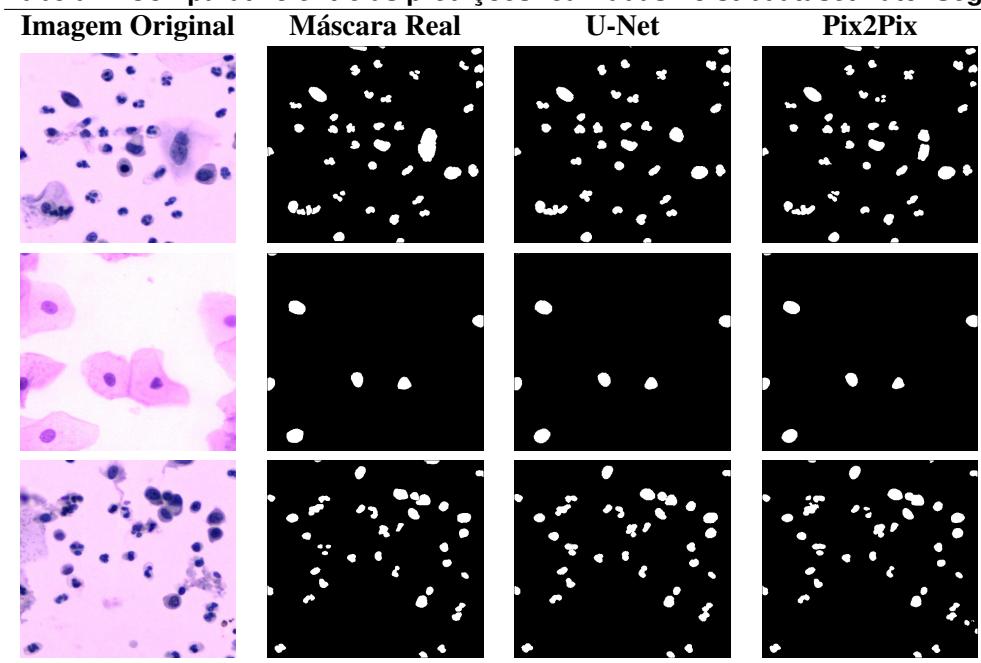


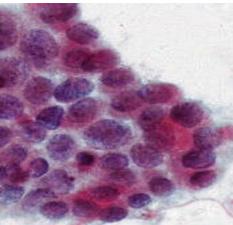
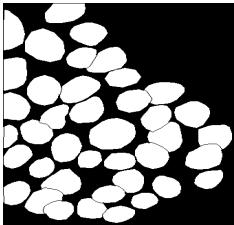
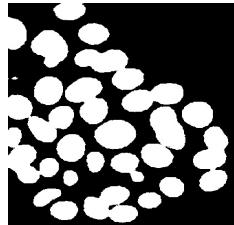
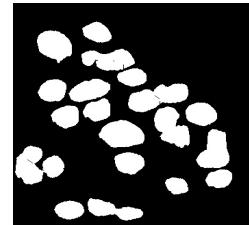
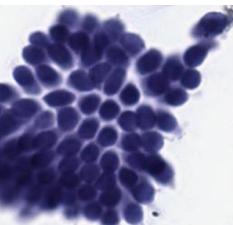
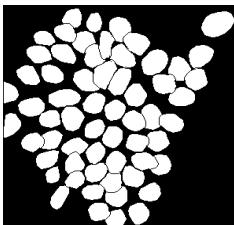
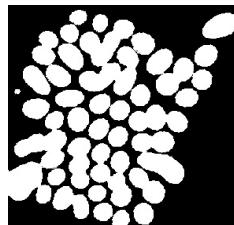
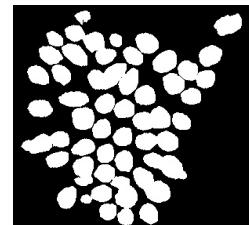
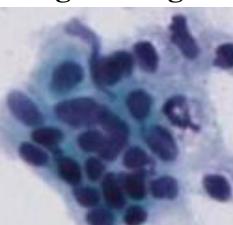
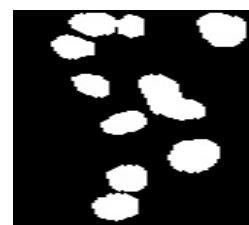
Tabela 3. Métricas de desempenho para o subdataset ClusterSeg e seus conjuntos Normal e Difícil para U-Net e Pix2Pix.

ClusterSeg						
Modelo	AJI	PQ	Adice	HD	Precisão	Recall
U-Net	0,60	0,59	0,73	25,03	0,86	0,87
Pix2Pix	0,58	0,57	0,73	26,34	0,85	0,84
ClusterSeg - Conjunto Difícil						
Modelo	AJI	PQ	Adice	HD	Precisão	Recall
U-Net	0,51	0,50	0,66	29,82	0,87	0,85
Pix2Pix	0,53	0,49	0,66	31,80	0,86	0,76
ClusterSeg - Conjunto Normal						
Modelo	AJI	PQ	Adice	HD	Precisão	Recall
U-Net	0,62	0,61	0,76	22,28	0,86	0,87
Pix2Pix	0,60	0,59	0,73	21,70	0,83	0,84

6.2. ClusterSeg

Os resultados apresentados na tabela 3 mostram que ambos os modelos tiveram um desempenho semelhante para o *subdataset* ClusterSeg. No ClusterSeg geral (conjunto Difícil e Normal), os modelos obtiveram desempenhos relativamente próximos em todas as métricas, com a U-Net apresentando ligeira vantagem. No Conjunto Difícil, ambos os modelos apresentaram uma queda no desempenho. Por outro lado, no Conjunto Normal, o modelo Pix2Pix apresentou um desempenho inferior tanto no aspecto quantitativo quanto no qualitativo. Isso é evidenciado na tabela 4 que mostra que as segmentações do Pix2Pix apresentaram bordas menos precisas, mais irregulares e com menos ajustes em comparação com a U-Net.

Tabela 4. Comparativo entre as predições realizadas no subdataset ClusterSeg e seus conjuntos Difícil e Normal.

ClusterSeg - Conjunto Difícil			
Imagen Original	Máscara Real	U-Net	Pix2Pix
			
			
ClusterSeg - Conjunto Normal			
Imagen Original	Máscara Real	U-Net	Pix2Pix
			
			

6.3. TargetA e TargetB

Ao se analisarem as segmentações geradas pelos modelos para o *subdataset* ClusterSeg, observa-se que ambos os modelos apresentaram dificuldades significativas na

segmentação de núcleos sobrepostos, o que é evidente já que a tarefa envolve um problema de segmentação de instâncias, o que requer a capacidade de distinguir núcleos que se encontram sobrepostos ou muito próximos. A mesma observação pode ser feita para os resultados dos testes nos *subdatasets* TargetA e TargetB. A Tabela 5 mostra que, de forma geral, o desempenho do modelo Pix2Pix foi inferior ao da U-Net nesses conjuntos, e a Tabela 6 ilustra as diferenças nas máscaras geradas.

Tabela 5. Métricas de desempenho para os subdatasets TargetA e TargetB para U-Net e Pix2Pix.

TargetA						
Modelo	AJI	PQ	Adice	HD	Precision	Recall
U-Net	0,54	0,53	0,69	12,62	0,78	0,81
Pix2Pix	0,50	0,52	0,70	17,54	0,65	0,80
TargetB						
Modelo	AJI	PQ	Adice	HD	Precision	Recall
U-Net	0,47	0,50	0,62	25,02	0,77	0,75
Pix2Pix	0,46	0,47	0,62	30,92	0,69	0,71

7. Conclusão

Este trabalho apresentou uma discussão sobre o uso da GAN condicional Pix2Pix na segmentação semântica de núcleos em imagens citológicas do colo do útero, comparando seu desempenho com uma arquitetura de segmentação convencional, a U-Net. Ambos os modelos apresentaram bons resultados no *subdataset* PatchSeg, onde o desafio de segmentação é mais leve e os núcleos estão menos sobrepostos. No entanto, nos *subdatasets* ClusterSeg e DomainSeg, ambos os modelos encontraram dificuldades, especialmente em regiões com núcleos sobrepostos ou onde há grande variação visual. Esse comportamento se deve às limitações de suas arquiteturas, que dificultam a segmentação de instâncias.

Apesar dessas limitações, o uso de GANs para a segmentação de núcleos provou ser uma alternativa promissora. Para trabalhos futuros, a exploração de técnicas baseadas em GANs adaptadas para segmentação de instâncias, ou para aprimoramento das bordas dos núcleos, pode trazer melhorias expressivas na precisão da segmentação. Além disso, o uso de GANs como técnica de pós-processamento para correção de erros de segmentação pode ajudar a refinar a saída dos modelos, promovendo uma segmentação mais detalhada e precisa. A expansão da base de dados para incluir uma maior variedade de tipos de núcleos e condições citológicas diversas também pode fortalecer a capacidade de generalização dos modelos e aprimorar seu desempenho em cenários mais desafiadores.

Agradecimentos — AMC Machado agradece o auxílio financeiro do Fundo de Incentivo à Pesquisa FIP-PUCMinas 2025/32467 e da FAPEMIG através dos projetos APQ-02753-24 e APQ-06556-24.

Tabela 6. Comparativo entre as previsões realizadas nos subdatasets TargetA e TargetB.

TargetA			
Imagen Original	Máscara Real	U-Net	Pix2Pix
TargetB			
Imagen Original	Máscara Real	U-Net	Pix2Pix

Referências

- Ali, M., Ali, M., Hussain, M., and Koundal, D. (2024). Generative adversarial networks (gans) for medical image processing: Recent advancements. *ARCH COMPUT METHOD E*.
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. C., and Bengio, Y. (2014). Generative adversarial networks. *COMMUN ACM*, 63:139 – 144.
- Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *PROC CVPR IEEE*, pages 5967–5976.
- Khaled, A., Han, J.-J., and Ghaleb, T. A. (2022). Multi-model medical image segmentation using multi-stage generative adversarial networks. *IEEE Access*, 10:28590–28599.
- Luc, P., Couprie, C., Chintala, S., and Verbeek, J. (2016). Semantic segmentation using adversarial networks. *ArXiv*, abs/1611.08408.
- Mahmood, F., Borders, D., Chen, R. J., Mckay, G. N., Salimian, K. J., Baras, A., and Durr, N. J. (2020). Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE T MED IMAGING*, 39(11):3257–3267.

- Mirza, M. and Osindero, S. (2014). Conditional generative adversarial nets. *ArXiv*, abs/1411.1784.
- Park, K.-B., Choi, S. H., and Lee, J. Y. (2020). M-gan: Retinal blood vessel segmentation by balancing losses through stacked deep fully convolutional networks. *IEEE Access*, 8:146308–146322.
- Popescu, D., Deaconu, M., Ichim, L., and Stamatescu, G. (2021). Retinal blood vessel segmentation using pix2pix gan. *MED C CONTR AUTOMAT*, pages 1173–1178.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *ArXiv*, abs/1505.04597.
- Siegel, R. L., Giaquinto, A. N., and Jemal, A. (2024). Cancer statistics, 2024. *CA: A Cancer Journal for Clinicians*, 74(1):12–49.
- Son, J., Park, S. J., and Jung, K.-H. (2019). Towards accurate segmentation of retinal vessels and the optic disc in fundoscopic images with generative adversarial networks. *J DIGIT IMAGING*, 32(3):499–512.
- van der Schot, A., Sikkel, E., Niekolaas, M., Spaanderman, M., and de Jong, G. (2023). Placental vessel segmentation using pix2pix compared to u-net. *Journal of Imaging*, 9(10).
- Xue, Y., Xu, T., Zhang, H., Long, L. R., and Huang, X. (2018). Segan: Adversarial network with multi-scale l1 loss for medical image segmentation. *Neuroinformatics*, 16(3):383–392.
- Zhang, X., Zhu, X., Zhang, X., Zhang, N., Li, P., and Wang, L. (2018). Seggan: Semantic segmentation with generative adversarial network. *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)*, pages 1–5.
- Zhao, J., jun He, Y., Zhou, S.-H., Qin, J., and ning Xie, Y. (2023). Cnseg: A dataset for cervical nuclear segmentation. *COMPUT METH PROG BIO*, 241:107732.