Seleção de Backbone Para Extração de Características com a U-Net na Segmentação de Patologias Renais

Ana A. F. Rocha[1], Rodrigo N. Borges[1], Rodrigo E. C. Batista[2], Rhaylson S. Nascimento[1], Émery F. Moriconi[1], Justino D. Santos[1], Rodrigo M. S. Veras[1]

[1]Departamento de Computação,
Universidade Federal do Piauí - Teresina, Brasil

[2]Instituto Federal do Piauí - Teresina, Brasil

{allyciaroxha,rodrigoelyelcb,emerymoriconi09,justinoduarte}@gmail.com,
{r_borges,rveras}@ufpi.edu.br, rhaylson.silva@ifpi.edu.br

Abstract. *This paper presents a comparative analysis of different backbones in combination with the U-Net architecture for the segmentation of renal pathologies, with an emphasis on glomerular sclerosis lesions. The study's main objective is to demonstrate the feasibility and effectiveness of using pre-trained backbones in this task. Five convolutional neural networks were evaluated on a set of 271 images. At the end of the experiments, VGG19 stood out, presenting the best performance, with a Dice coefficient of 35.88% in the test set and an accuracy of 89.84%.*

Resumo. *Este artigo apresenta uma análise comparativa de diferentes backbones em combinação com a arquitetura U-Net para a segmentação de patologias renais, com ênfase em lesões de esclerose glomerular. O objetivo principal do estudo é demonstrar a viabilidade e a eficácia do uso de backbones pré-treinados nessa tarefa. Foram avaliadas cinco redes neurais convolucionais em um conjunto de 271 imagens. Ao final dos experimentos, a VGG19 destacou-se, apresentando o melhor desempenho, com um coeficiente Dice de 35,88% no conjunto de teste e uma acurácia de 89,84%.*

1. Introdução

Os rins, dois órgãos localizados na região do abdomen, têm a principal função de filtrar o sangue, eliminando substâncias nocivas ao organismo através da urina. Esse processo ocorre nos néfrons, um conjunto de capilares sanguíneos envoltos pela cápsula de Bowman, que contém os glomérulos. É no glomérulo que chega o sangue para ser filtrado [Ferron and Rancano 2007], sendo assim exercem um papel crucial para a manutenção das funções renais.

As glomerulopatias são doenças que afetam os glomérulos renais, cujos diagnósticos necessitam ser precisos, para que o tratamento a ser fornecido seja o mais adequado. A glomeruloesclerose, um exemplo de glomerulopatia, caracteriza-se pela cicatrização e endurecimento dos glomérulos, podendo levar à perda gradual da função renal e, sem tratamento adequado, à insuficiência renal crônica. A identificação e quantificação dessas lesões são essenciais para evitar a progressão da doença. Na maioria dos casos, o nefrologista precisa lançar mão da biópsia renal para determiner o tipo de glomerulopatia com precisão [SBN 2023].

No contexto da saúde renal no Brasil, as glomerulopatias surgem como a terceira principal causa de doença renal crônica terminal, impactando 11% dos pacientes em diálise [Costa et al. 2017]. Tal constatação destaca a relevância desse tema, gerando a necessidade do desenvolvimento de estudos afim de determinar métodos eficazes para enfrentamento das glomerulopatias no âmbito da gestão da saúde renal. Diagnósticos sem o auxílio tecnológico incluem tarefas repetitivas e podem ser comprometidos por condições humanas adversas como estresse e cansaço [Doi 2007]. Desse modo, o uso de sistemas de diagnóstico auxiliado por computador (CAD – *Computer Aided Diagnosis*) mostra-se essencial, proporcionando diagnósticos mais precisos e menos subjetivos. Esta abordagem facilita a identificação de lesões nos glomérulos e impulsiona avanços significativos no diagnóstico e monitoramento das doenças renais, beneficiando os profissionais de saúde.

Diante da importância na identificação e diagnóstico de doenças renais, o objetivo primordial desta pesquisa consiste em desenvolver uma ferramenta para segmentação de lesões renais, denominadas esclerose glomerular, em imagens histológicas de rim. A execução deste estudo visa otimizar a precisão e eficiência do processo diagnóstico, reforçando a importância da integração de métodos computacionais para aprimorar a prática médica.

A Seção 2 aborda a literatura existente sobre o tema. A Seção 3 descreve as bases de imagens, método proposto e métodos de avaliação. As seções 4 e 5 apresentam os resultados parciais alcançados e, por fim, a Seção 6 discute o progresso do projeto e perspectivas futuras.

2. Trabalhos Relacionados

Barros et al. [Barros et al. 2017] introduzem o PathoSpotter-K, um sistema derivado do projeto PathoSpotter para detectar lesões glomerulares proliferativas. Utilizando técnicas convencionais de processamento de imagens e reconhecimento de padrões, o PathoSpotter-K foi testado em um conjunto de 811 imagens, incluindo 300 imagens de glomérulos normais e 511 imagens de glomérulos de rins afetados por glomerulopatias proliferativas. O sistema utiliza um classificador binário baseado no k-nearest neighbor (KNN), alcançando uma acurácia de 85% no conjunto de validação.

No estudo de Bel et al. [Bel et al. 2018], o objetivo central foi a segmentação do tecido renal utilizando Redes Neurais Convolucionais. Foram investigadas três arquiteturas de redes neurais: uma rede totalmente convolucional, uma multi-escala e uma U-net. Os resultados revelaram que as redes convolucionais alcançaram uma acurácia média de 90% para a maioria das classes, destacando-se pela capacidade precisa de segmentar estruturas como glomérulos, túbulos distais, capilares, entre outras.

A abordagem de Gadermayr et al. [Gadermayr et al. 2019] para segmentar glomérulos em imagens de lâminas de rim utilizou Redes Neurais Convolucionais (CNNs) em cascata. Comparada às redes totalmente convolucionais convencionais, o melhor modelo alcançou um coeficiente DICE de 90%, precisão de 89% e *recall* de 92% por meio de validação cruzada de tamanho oito. Os resultados foram considerados excelentes em análises qualitativas e de objeto, superando abordagens anteriores.

No estudo de Kaur et al. [Kaur et al. 2023], um modelo U-Net modificado foi desenvolvido para detectar glomérulos com precisão em imagens de tecido renal coradas com o corante *Periodic Acid-schif* (PAS). Foram feitos ajustes nos filtros, mapas de

características e profundidade do modelo, além da adição de camadas de convolução, normalização em lote, *max pooling* e *upsampling*. O modelo alcançou 95,7% de acurácia, 97,2% de precisão, 96,4% de *recall* e 96,7% de F1-score, superando o desempenho de outras abordagens, como o UNet original com EfficientNetB3.

Ao explorar a literatura, é evidente que a implementação de CNNs é uma alternativa benéfica para superar os desafios associados à segmentação de patologias no rim. Estudos anteriores confirmam sua eficácia para diagnósticos precisos e rápidos de doenças renais. Além disso, é possível identificar áreas em que modelos com objetivos semelhantes ao proposto ainda têm margem para aprimoramentos. Diante desse cenário, reforça-se a relevância do desenvolvimento do nosso projeto, visando preencher essas lacunas e contribuir para o avanço no campo da segmentação da esclerose glomerular.

3. Materiais e Métodos

Nesta seção, apresentamos os recursos empregados no desenvolvimento do projeto, com ênfase na metodologia criada para a segmentação de lesões em imagens de lâminas de exames patológicos renais.

Com esse objetivo, foram selecionadas, ajustadas e testadas as Redes Neurais Convolucionais pré-treinadas EfficientNetB7, DenseNet201, VGG16, VGG19 e MobileNetV2. A implementação das CNNs foi realizada utilizando a linguagem de programação *Python*, junto com as bibliotecas de Aprendizado Profundo *Keras* e *TensorFlow*.

Nos próximos tópicos, detalhamos o conjunto de imagens utilizado, as CNNs selecionadas, a implementação da rede U-Net como estrutura principal, e as métricas de avaliação aplicadas para mensurar o desempenho do modelo. Até o momento, conduzimos uma análise aprofundada da literatura, fundamental para a elaboração de um esquema robusto e abrangente no desenvolvimento da pesquisa.

3.1. Base de Imagens

Para conduzir testes e avaliações, foram unificados dois conjuntos de imagens de glomérulos, tratadas com corante PAS. O primeiro conjunto, contendo 115 imagens, foi obtido em parceria com um Departamento de Medicina Especializada (DME) e faz parte de uma base maior com cerca de 515 imagens classificadas com esclerose glomerular. O segundo conjunto, contendo 183 imagens com o mesmo tipo de lesão, foi fornecido pelo grupo PathoSpotter.

Após a união das bases e a remoção de imagens repetidas, o conjunto final totalizou 271 imagens A Figura 1 apresenta exemplos de imagens e das marcações realizadas pelos patologistas. As áreas de lesão foram cuidadosamente marcadas por especialistas em patologia renal.

3.2. U-Net

Para realizar as segmentações necessárias, utilizaremos uma arquitetura baseada na U-Net [Ronneberger et al. 2015]. A U-Net é uma rede neural convolucional nomeada por sua aparência em forma de "U", composta por um codificador e um decodificador, conforme ilustrado na Figura 2. Essas duas partes principais são conhecidas como bloco de contração e bloco de expansão [Long et al. 2015].

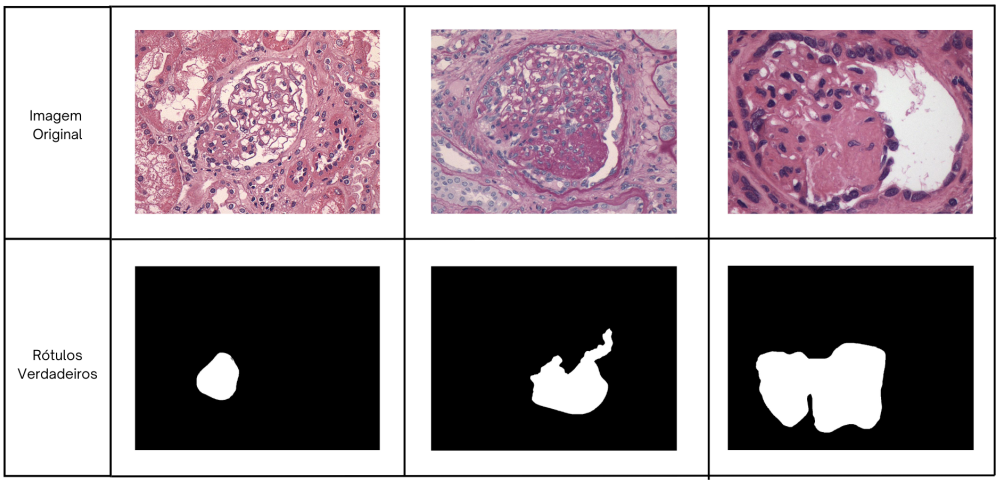

Figura 1. Imagens de glomérulos com lesões e suas respectivas máscaras segmentadas.

Na primeira parte da rede, denominada camada de descida, uma sequência de camadas de convolução seguida por camadas de *pooling* é empregada, permitindo a extração de características em múltiplas resoluções. Na segunda parte, chamada camada de subida, ocorre a expansão em níveis, restaurando a resolução original da imagem.

Uma característica distintiva da U-Net é a presença de conexões diretas entre a saída de um bloco de contração e a entrada do bloco de expansão correspondente [Ronneberger et al. 2015]. No decorrer da rede, essas conexões permitem a combinação de informações de diferentes níveis de resolução, resultando em maior precisão na segmentação.

Em nosso trabalho, investigamos diferentes arquiteturas pré-treinadas no conjunto de dados *ImageNet* como extratoras de características, substituindo o codificador padrão da U-Net, como será detalhado na seção a seguir.

3.3. Backbones Avaliados

No contexto de redes neurais convolucionais, o *backbone* é a parte inicial da estrutura, responsável pela extração de características das entradas, influenciando diretamente o desempenho e a capacidade de generalização do modelo. A utilização desse componente é uma estratégia viável para potencializar o desempenho das CNNs, dado seu potencial de extração otimizada de características, inclusive permitindo a utilização de redes pré-treinadas [Ciaparrone et al. 2020].

Na abordagem utilizada pelo método proposto, a U-Net atua como decodificador da arquitetura, enquanto o codificador da rede é substituído por diversas arquiteturas pré-treinadas no conjunto de dados ImageNet, tais como EfficientNetB7, DenseNet201, VGG-16, VGG-19 e MobileNetV2. Nesse sentido, trata-se de cinco redes diferentes concatenadas com o mesmo decodificador, ajustado para se adequar a rede codificadora, para que seja possível realizar o treinamento e análise.

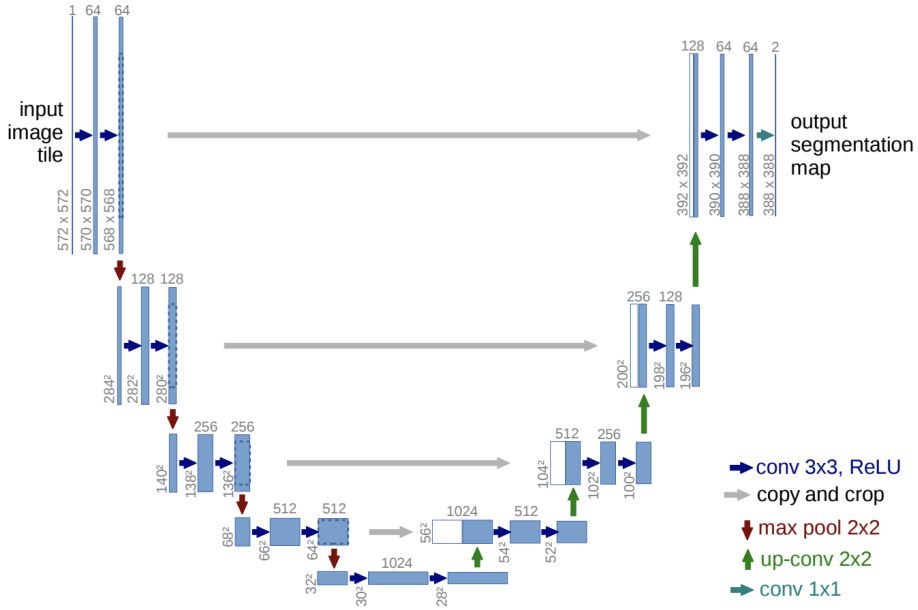

Figura 2. Arquitetura da U-Net (Parâmetros ilustrativos).

Na Figura 3, é ilustrado o modelo estruturado com apenas uma das redes testadas, incluindo suas entradas e saídas esperadas.

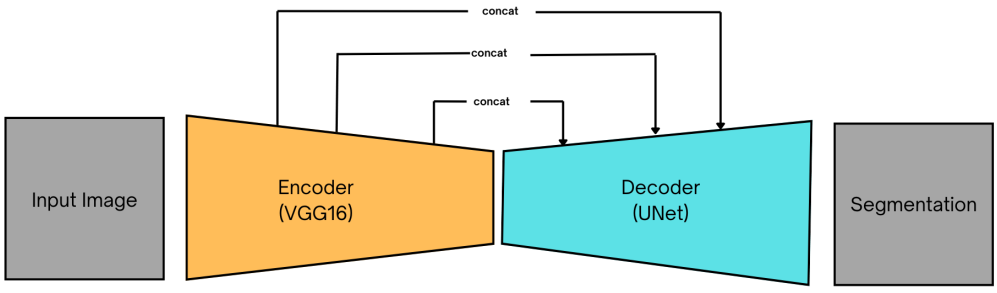

Figura 3. Fluxograma da arquitetura utilizada (Encoder ilustrativo).

3.4. Transferência de Aprendizado

Em nossa abordagem, a transferência de aprendizado foi um das técnicas cruciais para o desenvolvimento do modelo. A ideia é aproveitar o conhecimento aprendido em um campo e aplicá-lo em outro campo relacionado [Tajbakhsh et al. 2016]. Dessa forma, acelera o processo de treinamento e aprimorando o desempenho na nova tarefa.

Utilizando redes pré-treinadas na ImageNet como *backbone* da U-Net, é possível obter benefícios da capacidade desses modelos de extrair características complexas baseado no que já tiveram acesso, adaptando-as especificamente para a segmentação precisa de lesões em glomérulos.

O método desenvolvido utiliza o *Deep Fine-Tuning* (DFT), que, de acordo com o trabalho de Claro [Claro 2022], a técnica DFT permite o treinamento completo da CNN,

mas requer um custo computacional elevado quanto maior a quantidade de dados. Logo, essa abordagem torna-se viável em contextos com conjuntos de dados limitados. Ao ajustar os pesos da rede neural pré-treinada, inicialmente treinada na base da ImageNet, para se adaptarem aos padrões específicos dos dados de segmentação, o DFT possibilita melhorar a precisão e a generalização do modelo, facilitando sua aplicação em cenários clínicos e de pesquisa.

3.5. Métricas de Avaliação

A avaliação do desempenho da rede envolve a quantificação de sua eficácia para obter conclusões sobre a viabilidade de seu uso. No método proposto, as métricas utilizadas para avaliar os resultados foram Acurácia, Sensibilidade, Especificidade e Coeficiente de Sørensen (DICE), como em [Govind et al. 2018]. Com a definição desses indicadores de desempenho, é possível analisar a evolução do modelo e sua performance.

Especificamente, a Acurácia mede a proporção de predições corretas, considerando tanto as regiões de interesse quanto as que não são de interesse, sendo uma métrica global de desempenho. A Sensibilidade avalia a capacidade do modelo de identificar corretamente as regiões de interesse, enquanto a Especificidade mede a capacidade de identificar corretamente as regiões que não são de interesse. O DICE é uma métrica que quantifica a similaridade entre as regiões segmentadas pelo modelo e as regiões de referência, sendo particularmente útil para avaliar a sobreposição entre conjuntos preditos e verdadeiros.

4. Resultados e Discussão

Fazendo o uso de técnicas de transferência de aprendizado utilizando redes pré-treinadas e o *Deep Fine-Tuning*, das 271 imagens disponíveis em nossa base de imagens, 80% foram selecionadas para treino, 10% para validação e 10% para teste. A cada época, o conjunto de treino era embaralhado, já os de teste e validação sempre se tratavam apenas dos últimos 20% do vetor de imagens, portanto, as últimas 55 imagens.

A partir disso, treinamos cinco diferentes redes individualmente, utilizando cada uma delas como *backbone* para a arquitetura U-Net. Cada arquitetura foi treinada cinco vezes, totalizando vinte e cinco execuções no total. Cada execução consistiu em cinquenta épocas de treinamento.

Após o treinamento, calculamos a média e o desvio padrão dos valores das métricas de desempenho, incluindo DICE, Sensibilidade, Especificidade e Acurácia, para cada rede, como apresentados na Tabela 1. Em seguida, selecionamos a rede com a melhor média de DICE no conjunto de teste, identificando os valores máximo e mínimo entre as cinco repetições. Esses valores foram utilizados para a análise comparativa da segmentação.

A rede de melhor desempenho foi a VGG19, que apresentou o maior valor DICE de 35,88% e acurácia de 89,85%. Tais resultados indicam que a VGG19 foi a mais eficaz em termos de sobreposição entre previsões e rótulos verdadeiros, bem como na classificação correta das áreas relevantes e não relevantes. Os exemplos de segmentação correspondentes a melhor e pior execução de da VGG19 são ilustrados na Figura 4 e 5, respectivamente.

Tabela 1. Resultados das arquiteturas avaliadas.

Arquitetura	Dice(%)	Sen(%)	Spe(%)	Acc(%)
EfficientNet B7	28,64 ± 0.029	57,61 ± 0.092	84,84 ± 0.026	82,93 ± 0.016
DenseNet 201	23,87 ± 0.038	21,14 ± 0.018	96,37 ± 0.012	91,56 ± 0.014
VGG-16	34,81 ± 0.051	49,23 ± 0.032	90,79 ± 0.019	88,13 ± 0.018
MobileNet V2	23,88 ± 0.054	54,29 ± 0.163	83,15 ± 0.109	80,96 ± 0.091
VGG-19	**35,88 ± 0.032**	**44,44 ± 0.027**	**92,97 ± 0.013**	**89,84 ± 0.012**

A Figura 4 apresenta uma combinação de resultados tanto bons quanto ruins obtidos durante a melhor execução do modelo, considerando que cada arquitetura foi treinada e testada cinco vezes, com os pesos sendo reajustados a cada repetição. De maneira similar, a Figura 5 exibe resultados positivos e negativos, mas desta vez durante a pior execução do modelo. Em ambas as figuras, é possível visualizar a máscara original da imagem além da máscara predita pelo modelo, permitindo uma comparação direta entre a segmentação real e a segmentação prevista.

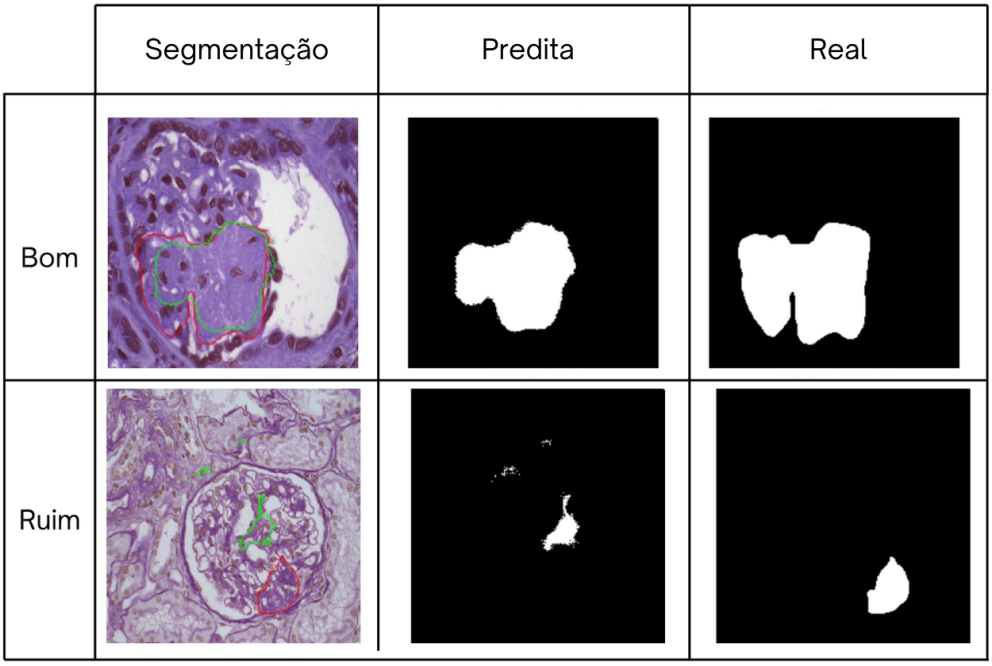

Figura 4. Resultados dos testes que teve melhor execução na VGG19, com as segmentações obtidas pelo método (em verde) e a máscara original da imagem (em vermelho).

5. Conclusão e Trabalhos Futuros

O presente estudo teve como objetivo avaliar a utilização de diferentes *backbones* como codificador para a arquitetura U-Net na segmentação de lesões em uma estrutura renal, utilizando modelos pré-treinados no conjunto de dados ImageNet. A melhor combinação encontrada foi a da U-Net com a VGG-19, que apresentou um DICE de 35,88% e Acurácia de 89,84%. A utilização de *backbones* como codificador oferece uma vantagem significativa ao facilitar a obtenção de resultados promissores na tarefa de segmentação, especialmente durante as fases iniciais do desenvolvimento de modelos mais avançados.

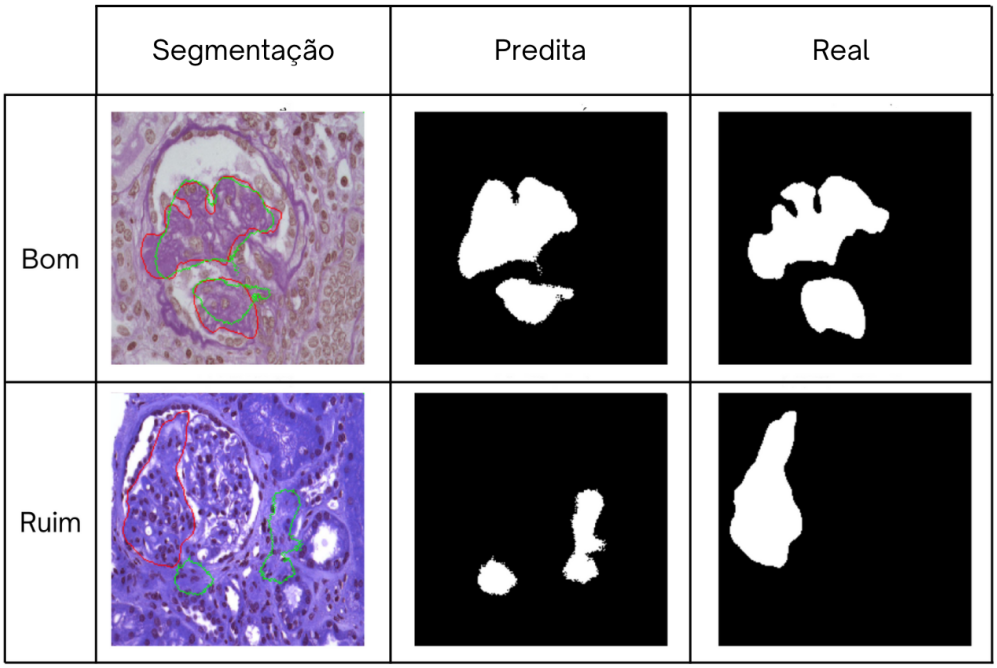

Figura 5. Resultados dos testes que teve pior execução na VGG19, com as segmentações obtidas pelo método (em verde) e a máscara original da imagem (em vermelho).

Embora a rede tenha demonstrado bom desempenho em métricas como acurácia e especificidade, os resultados em termos de DICE não foram tao satisfatórios, ou seja, a sobreposição entre previsões e rótulos verdadeiros. Além disso, ao analisar as máscaras preditas e as reais, nota-se que em ambas as execuções (melhor e pior) há casos em que o resultado obtido pelo método não se aproxima do original. Há possíveis causas que justificam essas situações, sendo uma delas a quantidade limitada de dados para treinamento, composto por 271 imagens. Conjuntos de dados pequenos podem resultar em modelos com menor capacidade de generalização, devido à reduzida diversidade de exemplos durante o treinamento. Este aspecto pode influenciar negativamente o desempenho das redes em termos de precisão na segmentação.

Além disso, este estudo pode servir como um ponto de referência para futuras pesquisas em segmentação de patologias renais, oferecendo uma baseline para comparar modificações na arquitetura da U-Net. Recomenda-se explorar o treinamento com outros pré-treinados, como ResNet201 e InceptionV3. Aumentar o tamanho do conjunto de dados é essencial, assim como considerar técnicas de aumento de dados se a base de imagens permanecer pequena. Por fim, investigar outras arquiteturas base além da U-Net, como a LinkNet [Chaurasia and Culurciello 2017], também é sugerido.

Referências

Barros, G. O., Navarro, B., Duarte, A., and Dos-Santos, W. L. (2017). Pathospotter-k: A computational tool for the automatic identification of glomerular lesions in histological images of kidneys. *Scientific reports*, 7:46769.

Bel, T., Hermsen, M., van der Laak, J., Litjens, G., Smeets, B., and Hilbrands, L. (2018). Automatic segmentation of histopathological slides of renal tissue using deep learning. page 37.

Chaurasia, A. and Culurciello, E. (2017). Linknet: Exploiting encoder representations for efficient semantic segmentation. In *2017 IEEE Visual Communications and Image Processing (VCIP)*, pages 1–4.

Ciaparrone, G., Bardozzo, F., Priscoli, M. D., Kallewaard, J. L., Zuluaga, M. R., and Tagliaferri, R. (2020). A comparative analysis of multi-backbone mask r-cnn for surgical tools detection. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

Claro, M. d. L. (2022). *Classificação de leucemias utilizando aumento de dados, transferência de aprendizado e combinação de cnns*. PhD thesis, Universidade Federal do Maranhão.

Costa, D. M. N., Valente, L. M., Gouveia, P. A. C., Sarinho, F. W., Fernandes, G. V., Cavalcante, M. A. G. M., Oliveira, C. B. L., Vasconcelos, C. A. J., and Sarinho, E. S. C. (2017). Comparative analysis of primary and secondary glomerulopathies in the northeast of brazil: data from the pernambuco registry of glomerulopathies – repeg. *Brazillian Journal of Nephrology*, 39(1):29–35.

Doi, K. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, 31(4-5):198–211.

Ferron, M. and Rancano, J. (2007). *Grande Atlas do Corpo Humano*. MANOLE, [S.l.].

Gadermayr, M., Dombrowski, A. K., Klinkhammer, B. M., Boor, P., and Merhof, D. (2019). Cnn cascades for segmenting sparse objects in gigapixel whole slide images. *Comput. Med. Imaging Graph*, 71:40–48.

Govind, D., Ginley, B., Lutnick, B., Tomaszewski, J., and Sarder, P. (2018). Glomerular detection and segmentation from multimodal microscopy images using a butterworth band-pass filter. page 39.

Kaur, G., Garg, M., Gupta, S., Juneja, S., Rashid, J., Gupta, D., Shah, A., and Shaikh, A. (2023). Automatic identification of glomerular in whole-slide images using a modified unet model. *Diagnostics*, 13:3152.

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer.

SBN (2023). Glomerulopatias. `https://www.sbn.org.br/orientacoes-e-tratamentos/doencas-comuns/glomerulopatias`. Acesso em: 20 fev. 2024.

Tajbakhsh, N., Shin, J., Gurudu, S., Hurst, R. T., Kendall, C. B., Gotway, M. B., and Liang, J. (2016). Convolutional neural networks for medical image analysis: Fine tuning or full training? *IEEE Transactions on Medical Imaging*, 35:1299–1312.