

Dual-Scale SE UNet: Uma Rede Neural Convolucional Aprimorada com SE Blocks para Segmentação Renal em Imagens de TC

Marcus Vinicius Oliveira¹, Caio Eduardo Falcão Matos¹, Geraldo Braz Júnior¹,
João Dallyson Sousa de Almeida¹

¹Núcleo de computação aplicada – Universidade Federal do Maranhão (UFMA)
65.080-805 – São Luís – MA – Br

{marcus.oliveira, caioefalcao, geraldo, jdallyson}@nca.ufma.br

Abstract. *Renal cancer is one of the most common neoplasms of the genitourinary tract, making automatic segmentation an essential tool for assisting in early diagnosis. This work aims to develop and evaluate a neural network for segmenting computed tomography images and identifying kidneys, tumors and cyst. We propose replacing the standard U-Net convolutional layers with Dual-Scale SE blocks to enhance feature extraction. The proposed methodology achieved a Dice coefficient of 0.93 and 0.86, respectively.*

Resumo. *O câncer renal é uma das neoplasias mais comuns do trato genitourinário, tornando a segmentação automática uma ferramenta essencial para auxiliar no diagnóstico precoce. Este trabalho visa desenvolver e avaliar uma rede neural para segmentação de imagens de tomografia computadorizada, identificando rins, tumores e cisto. Propomos a substituição das camadas convolucionais da U-Net por blocos Dual-Scale SE para aprimorar a extração de características. A metodologia proposta obteve um coeficiente Dice de 0,93 e 0,86, respectivamente.*

1. Introdução.

Entre 2008 e 2012, a incidência de câncer aumentou de 12,7 milhões para 14,1 milhões de casos, com projeções indicando que esse número pode chegar a 25 milhões nas próximas duas décadas. Além disso, estima-se que as mortes decorrentes da doença atinjam 14,6 milhões até 2035 [Stewart et al. 2014]. Esses dados ressaltam a urgência de investir em prevenção, diagnóstico precoce e controle da doença, especialmente em países menos desenvolvidos, onde o câncer é frequentemente detectado em estágios avançados e o acesso ao tratamento é limitado. O impacto será particularmente severo em nações de baixa e média renda, que enfrentam dificuldades crescentes para lidar com o aumento dos casos.

O carcinoma renal, também chamado de adenocarcinoma de rim, é a terceira neoplasia mais prevalente do trato genitourinário, correspondendo a aproximadamente 3% dos tumores malignos em adultos e figurando entre os dez tipos de câncer mais comuns [Siegel et al. 2024]. Segundo [Gloeckler Ries et al. 2003], a incidência do câncer de rim é de 15,4 por 100.000 homens e 7,7 por 100.000 mulheres. Diante desse cenário, a detecção precoce desempenha um papel crucial na melhoria do prognóstico e na eficácia do tratamento, aumentando consideravelmente as chances de cura. Quando identificado

nos estágios iniciais, antes da disseminação para outras partes do corpo, a taxa de sobrevivência relativa em cinco anos é de aproximadamente 93%. No entanto, em casos mais avançados, com metástase, essa taxa cai drasticamente para 12%. O câncer de rim está entre as neoplasias com maior impacto prognóstico quando diagnosticadas tardiamente, evidenciando a importância da detecção precoce para ampliar as chances de sucesso terapêutico [American Cancer Society 2024].

A Tomografia Computadorizada (TC) é amplamente utilizada no sistema de saúde devido ao seu baixo custo, qualidade de imagem satisfatória, rapidez na execução e aplicabilidade em diferentes diagnósticos [Brenner and Hall 2007]. No entanto, o processo de análise dessas imagens para a identificação e extração de informações relevantes ainda é altamente complexo, realizado de forma manual e visual pelos médicos. Esse procedimento exige um alto nível de atenção e repetição, podendo resultar em fadiga física e visual, além de desvio de atenção, o que pode comprometer a precisão da avaliação diagnóstica.

Diante desse cenário, a aplicação de aprendizado profundo em problemas médicos tem se tornado cada vez mais comum, auxiliando os profissionais de saúde no diagnóstico. As abordagens baseadas em *deep learning* têm demonstrado resultados promissores e conquistado espaço na prática clínica. A segmentação automática de tumores renais contribui para avaliações mais objetivas e reduz o esforço manual. Nesse sentido, o desafio de segmentação de tumores renais de 2019 (KiTS19 [Heller et al. 2019]), o KiTS2021 [Heller et al. 2023] e o KiTS2023 [Heller et al. 2023] buscam acelerar o desenvolvimento de ferramentas confiáveis para essa tarefa. Algumas pesquisas têm utilizado essa base de dados pública para avaliar algoritmos que podem auxiliar na construção de uma solução.

Entre os desafios de segmentação de regiões de rins e tumores em imagens de TC, está a extração de características de maneira eficiente em diferentes escalas espaciais. Este trabalho visa desenvolver uma versão modificada da rede U-Net por meio de blocos Dual-Scale SE Block para segmentação de rim e tumor em imagens de tomografia computadorizada visando não somente aprimorar a segmentação das estruturas anatômicas e patológicas, como também diminuir a carga de trabalho dos profissionais de saúde, automatizando parte do processo e minimizando a necessidade de intervenção manual, tornando a análise mais eficiente e confiável. Os blocos foram adicionados em substituição às convoluções da rede, visando ampliar a extração de características das imagens, combinando informações locais e globais e filtrando somente o que é mais relevante.

Pode-se destacar como contribuição do método proposto os seguintes pontos (a) um novo modelo baseado nos conceitos multinível e U-Net, visando aumentar a qualidade das características extraídas para a segmentação semântica de imagens de TC; (b) Adaptação e integração dos conceitos de multinível, atenção e U-Net para o desenvolvimento de um novo bloco. (c) Um método robusto para segmentação precisa de rins, cistos e tumores renais. Dessa forma, o método proposto pode ser integrado a sistemas Computer Aided Diagnosis (CAD), auxiliando no diagnóstico do câncer renal.

2. Trabalhos Relacionados

Os estudos [Zhao et al. 2022] e [Golts et al. 2022] utilizam o conjunto de dados KiTS21, porém adotam abordagens distintas para a segmentação. Os autores de [Zhao et al. 2022] propõem um método coarse-to-fine baseado no nnU-Net, no qual a segmentação ocorre

em duas etapas: primeiro, uma segmentação grosseira identifica a região de interesse (ROI) do rim; em seguida, uma segmentação refinada é aplicada dentro dessa ROI, separando tumores e cistos individualmente.

Além dessas abordagens, outros estudos exploram técnicas alternativas para segmentação em conjuntos de dados relacionados. Por exemplo, [Oliveira et al. 2022] utilizaram uma U-Net modificada com um bloco Pyramid Pooling Module na fase de contração para segmentar imagens do KiTS19, enquanto [Matos et al. 2023] empregaram a PPM-DeepLab para segmentar dados dos conjuntos KiTS21 e KiTS23.

Já os autores de [Golts et al. 2022] adotam uma estratégia distinta, baseada em um *Ensemble* de modelos 3D U-Net para segmentação de rins e tumores em tomografias computadorizadas no desafio KiTS21. Eles incorporam técnicas como aprendizado por transferência, perda regularizada, amostragem de múltiplas anotações e pós-processamento, visando aprimorar a precisão das predições.

Os pesquisadores de [Hou et al. 2019] apresentaram uma abordagem em cascata composta por duas redes neurais totalmente convolucionais: a primeira para a localização dos rins e a segunda para a segmentação dos rins e tumores. De maneira semelhante, [Zhang et al. 2019] adotaram uma metodologia em três etapas, utilizando uma nnU-Net para a localização renal. Por sua vez, [Isensee and Maier-Hein 2019] propuseram um *Ensemble* composto por três variações da U-Net 3D, Residual 3D U-Net e Pre-activation Residual 3D U-Net, enquanto [Türk et al. 2020] apresentaram um modelo híbrido denominado V-Net, ampliando as possibilidades para a segmentação de rins e tumores. Complementando essas técnicas, [George 2022] propôs um esquema de treinamento em dois estágios com U-Net 3D: primeiro em volumes reduzidos, depois na ROI em resolução total. Além disso, [George 2022] implementou um esquema de treinamento em dois estágios com U-Net 3D: inicialmente em volumes reduzidos, seguido pela região de interesse em resolução total. A Tabela 1 resume os trabalhos que apresentam a métrica dice. As abordagens, em geral, baseiam-se em variações da U-Net 3D, ensembles ou arquiteturas em cascata, alcançando ótimos resultados na segmentação renal. No entanto, a segmentação de tumores ainda apresenta maior dificuldade, refletida nos valores de Dice inferiores.

Tabela 1. Resumo dos trabalhos relacionados que apresentam dice.

Trabalho	Método	Dataset	Dice Rim	Dice Tumor
[Zhao et al. 2022]	nnU-Net (coarse-to-fine)	KiTS21	0,97	0,86
[Golts et al. 2022]	Ensemble 3D U-Net	KiTS21	-	0,83
[Matos et al. 2023]	PPM-DeepLab	KiTS21	0,94	0,84
[Hou et al. 2019]	3D UNet em Cascata	KiTS19	0,90	0,80
[Zhang et al. 2019]	nnU-Net em Cascata	KiTS19	0,97	0,83
[Isensee and Maier-Hein 2019]	Ensemble 3D U-Net	KiTS19	0,97	0,85
[Türk et al. 2020]	V-Net Híbrida	KiTS19	0,97	0,86
[George 2022]	U-Net 3D (2 estágios)	KiTS21	0,97	0,87

3. Metodologia

Nesta seção, a metodologia proposta neste trabalho é demonstrada. A Figura 1 ilustra uma visão geral das três etapas desenvolvidas. A primeira consiste na seleção das imagens

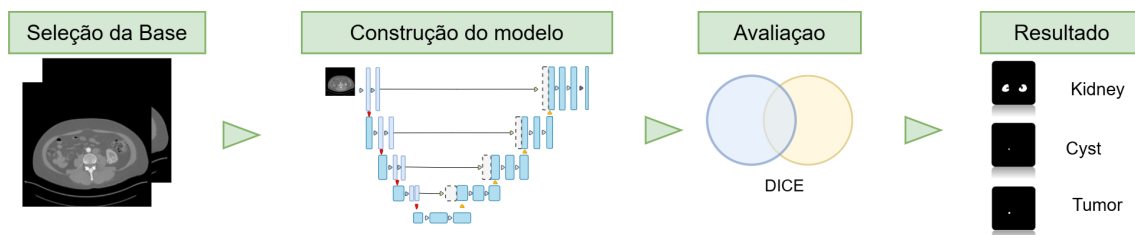


Figura 1. Etapas da metodologia proposta.

usadas no trabalho, proveniente do conjunto de dados do desafio de segmentação denominado KiTS21 [Heller et al. 2023]. Na segunda etapa, é definida a construção do modelo de segmentação, sendo este baseado em uma modificação da arquitetura U-Net, onde os blocos de convolução do codificador da rede são substituídos pelo bloco Dual-Scale SE proposto neste trabalho. Na última etapa, compreende a experimentação e avaliação do método proposto. Cada uma das etapas é apresentada nas seções seguintes.

3.1. Seleção da base.

Para a avaliação do método proposto, foi utilizado o *Kits21 Database* [Heller et al. 2023], um banco de dados público voltado para segmentação automática de estruturas renais em exames de tomografia computadorizada (CT). Essa base de dados reúne o *Kidney and Kidney Tumor Segmentation Challenge*, cujo objetivo principal é impulsionar pesquisas relacionadas à segmentação de tumores e massas renais, contribuindo para avanços no diagnóstico e tratamento de câncer renal. O conjunto de dados inclui 300 exames de CT, cada um referente a um paciente, armazenados no formato original NIFTI (*Neuroimaging Informatics Technology Initiative*), amplamente utilizado em neuroimagem, com resolução de 512×512 *pixels* em escala de cinza. Conforme ilustrado na Figura 2, as imagens apresentam estruturas renais segmentadas manualmente por especialistas, incluindo rins (roxo), tumores (verde) e cistos (azul).

O *dataset* disponibiliza segmentações manuais realizadas por especialistas, classificando as imagens em três categorias principais: rins, tumores renais e cistos renais. No total, são fornecidas 65.164 fatias, das quais 22.997 correspondem a rins, 8.341 a tumores e 2.869 a cistos.

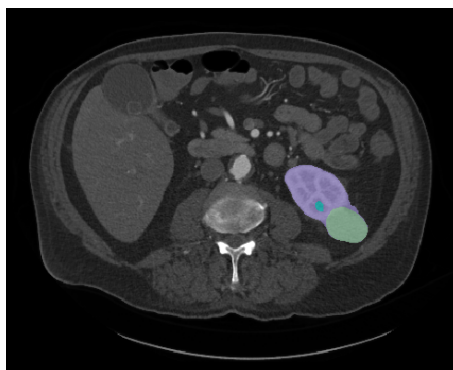


Figura 2. Exemplo de uma fatia segmentada.

3.2. Construção do modelo.

A arquitetura de rede analisada neste estudo foi inspirada na U-Net [Ronneberger et al. 2015], uma rede totalmente convolucional que se destaca por sua maior eficiência, devido à eliminação de camadas densas e à redução do número de parâmetros, possibilitando sua aplicação em imagens de diferentes dimensões. Essa rede realiza uma etapa inicial de redução de amostragem por meio de operações de convolução e *pooling*, seguida por uma fase de aumento da resolução. No caso da U-Net, essa etapa de reconstrução ocorre por meio de conexões de salto (*skip connections*), que conectam diretamente as camadas de redução de amostragem às camadas de aumento de resolução, permitindo o aproveitamento de informações anteriores para aprimorar a qualidade final das imagens [Long et al. 2014].

Devido à sua estrutura característica, essa rede recebeu o nome de U-Net, refletindo seu formato em “U”. A arquitetura é dividida em duas etapas principais: a primeira fase, chamada de *encoder*, consiste na contração da imagem para capturar suas principais características; já a segunda fase, o *decoder*, é responsável por restaurar a resolução original da imagem utilizando deconvoluções e conexões de salto (*skip connections*) para, ao final, gerar a predição de cada *pixel*.

Apesar da eficácia da U-Net, algumas variações arquiteturais foram propostas para aprimorar a extração de características, incorporando estratégias utilizadas em redes como a Inception [Szegedy et al. 2016] e a HRNet [Wang et al. 2019]. A aplicação do Dual-Scale SE visa capturar informações em escalas espaciais distintas, aprimorando a extração de características relevantes para a segmentação e combinando-as para obter uma visão mais precisa do que foi recebido. Isso é alcançado por meio da divisão em dois níveis, utilizando escalas de 3×3 e 7×7 , com uma distância razoável entre essas escalas, medida pelo tamanho do núcleo da convolução. O uso desses dois níveis permite capturar tanto detalhes locais quanto contextos mais amplos, otimizando a segmentação.

Para otimizar ainda mais a retenção de informação para a segmentação, além das variações estruturais mencionadas, técnicas de atenção como o Squeeze-and-Excitation (SE) [Hu et al. 2017] foram exploradas. Esse mecanismo melhora redes neurais ao recalibrar dinamicamente a importância de cada canal de um mapa de características. Ele funciona em duas etapas principais: *Squeeze* (Compressão), onde a rede aplica o *Global Average Pooling* (GAP) para obter uma representação global de cada canal, resumindo a informação espacial em um vetor; e *Excitation* (Recalibração), onde o vetor gerado passa por duas camadas totalmente conectadas (com funções *ReLU* e *Sigmoid*) para calcular pesos para cada canal, indicando sua importância relativa. Ao fim, os pesos recalibram os canais da *feature map* original, ajustando sua contribuição para a rede. O bloco SE melhora a capacidade de discriminação da rede, ajudando-a a focar nas características mais relevantes.

A integração do módulo SE [Hu et al. 2017] na arquitetura estudada, combinada com a concatenação de convoluções (Figura 3), visa aprimorar a extração de características, permitindo que a rede capture informações em diferentes escalas e filtre as partes mais relevantes dos mapas de características. Essa abordagem segmenta o mapa de características em múltiplos níveis, possibilitando um refinamento progressivo da informação extraída. O bloco modificado aplica convoluções específicas em cada nível, inserindo um módulo SE ao final de cada etapa para destacar as características mais im-

portantes, garantindo que somente as informações mais relevantes sejam transmitidas adiante.

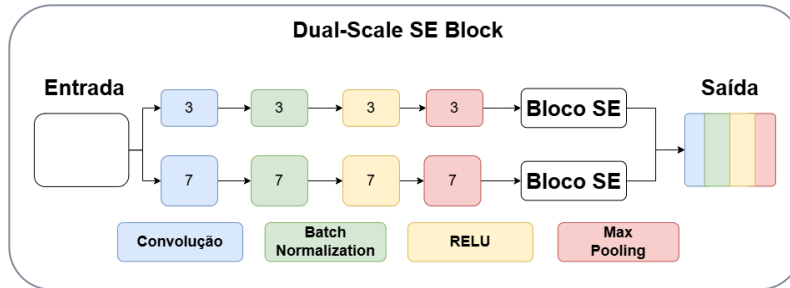


Figura 3. Bloco Dual-Scale SE.

A escolha dessa estrutura foi fundamentada na análise do conjunto de dados, onde muitos pacientes apresentavam perda total ou parcial do órgão renal. Além disso, os tumores e cistos frequentemente não estão completamente separados do órgão, apresentando tamanhos variados e localizando-se em regiões de alta similaridade estrutural. A modificação proposta, ilustrada na Figura 4, contribui para lidar com essa complexidade ao aprimorar a segmentação das regiões de interesse. A rede recebe como entrada imagens axiais de tomografia computadorizada (TC) com uma única fatia, e produz como saída mapas de segmentação multiclasse com três canais correspondentes às regiões de rim, cisto e tumor.

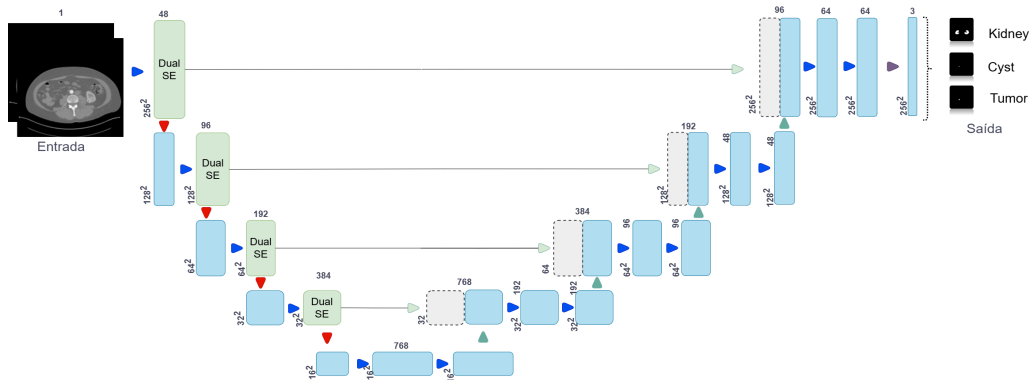


Figura 4. Dual-Scale SE UNet.

Por fim, a Figura 5 evidencia o impacto do mecanismo de atenção SE na redução do ruído presente na ativação original das convoluções convencionais (Conv 3×3 e Conv 7×7). As ativações resultantes (SE 3×3 e SE 7×7) demonstram um foco mais seletivo nas regiões mais relevantes da imagem, melhorando a discriminação dos padrões úteis e minimizando a dispersão para áreas menos significativas. Dessa forma, a proposta fortalece a capacidade da rede em identificar estruturas complexas, tornando a segmentação mais precisa e robusta.

3.3. Avaliação.

Para medir a precisão desta metodologia, aplicaram-se métricas amplamente validadas em sistemas Computer Aided Diagnosis (CADx) e diversos trabalhos presentes na literatura voltados para segmentação de imagens médicas.

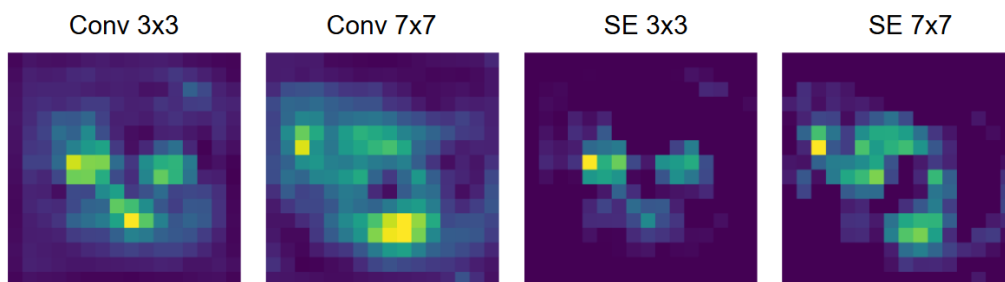


Figura 5. A comparação entre a atenção antes e depois.

A métrica utilizada para avaliação foi o coeficiente de similaridade de Dice [Fleiss et al. 1981], amplamente empregado em tarefas de segmentação. Essa métrica considera simultaneamente a precisão e a sensibilidade, refletindo o equilíbrio entre os acertos e os erros do modelo. Para isso, são levados em conta os casos corretamente identificados (verdadeiros positivos) e os erros por omissão ou inclusão (falsos negativos e falsos positivos).

4. Resultados e Discussão.

Nessa seção, são apresentados os resultados dos experimentos realizados. O método foi aplicado usando a linguagem Python. Utiliza-se a biblioteca Keras, com o TensorFlow GPU de back-end. A placa de vídeo utilizada foi a NVIDIA GeForce GTX 3060. As imagens foram pré-processadas e redimensionadas de 512×512 para 256×256 pixels, adotando-se uma separação dos dados por paciente para evitar vazamento entre conjuntos.

Realizou-se uma validação cruzada com 5 *folds*, utilizando-se quatro partes para treinamento e uma para teste em cada iteração. Em cada execução, 80% do conjunto de dados foi usado para treinamento e 20% para teste, sendo que 10% dos dados de treinamento foram reservados para validação durante o ajuste do modelo, garantindo uma avaliação robusta e generalizável do desempenho da rede neural. O treinamento foi conduzido por até 50 épocas, com interrupção antecipada (*early stopping*) caso a função de perda combinada (Dice Loss + Focal Loss [Lin et al. 2018]) no conjunto de validação não apresentasse redução por 8 épocas consecutivas.

Inicialmente, os hiperparâmetros do modelo foram ajustados utilizando o algoritmo *Tree-Structured Parzen Estimator* [Bergstra et al. 2011]. O Espaço de busca de hiperparâmetros é mostrado na Tabela 2. O multinível refere-se ao uso de convoluções com diferentes tamanhos de kernel. Foram testadas duas configurações: uma com dois níveis (3x3 e 7x7) e outra com quatro níveis (3x3, 5x5, 7x7 e 11x11). O *batch size* foi avaliado com dois valores distintos: 4 e 8. Por fim, a taxa de aprendizado foi otimizada em um espaço de busca contínuo entre 1×10^{-5} e 1×10^{-2} .

A Tabela 3 apresenta os valores médios do coeficiente Dice obtidos pela Dual-Scale SE UNet durante o processo de validação cruzada. Os resultados médios foram de 0,93 para o rim, 0,86 para o tumor e 0,93 para o cisto, com variações discretas entre as iterações, conforme evidenciado pelo baixo desvio padrão. Esses resultados foram alcançados utilizando o melhor conjunto de hiperparâmetros encontrados durante a otimização, uma estrutura multinível de dois níveis, *batch size* igual a quatro e uma taxa de aprendizado reduzida.

Tabela 2. Espaço de busca de hiperparâmetros.

Parâmetros	Valores
Multinível	2 ou 4
<i>Batch size</i>	4 ou 8
Taxa de aprendizado	$[1 \times 10^{-5}, 1 \times 10^{-2}]$

Tabela 3. Resultados dos valores de Dice obtidos na validação cruzada.

Interação	Dice Tumor	Dice Rim	Dice Cisto
1	0,87168	0,92703	0,93000
2	0,85379	0,93170	0,92878
3	0,83939	0,92799	0,91821
4	0,88846	0,95062	0,94945
5	0,84973	0,93012	0,95026
Média	0,86021	0,93389	0,93534
Desvio padrão	0,01870	0,00851	0,01236

Ao compararmos com a U-Net padrão (Tabela 4), observa-se uma melhora consistente nos resultados: a Dual-Scale SE UNet superou a U-Net nos três alvos de segmentação, especialmente nos casos de tumor (0,86 contra 0,83) e cisto (0,93 contra 0,86). Esses dados indicam que modificações no caminho de contração da U-Net, como a inserção de mecanismos de atenção do tipo SE, podem contribuir significativamente para o aprimoramento do desempenho da segmentação.

4.1. Discussão.

Nessa seção apresentamos alguns resultados qualitativos para analisar as segmentações realizadas. A rede demonstra aprendizado ao analisar o caso 00170 e 00051 da base Kits21 apresentado na Figura 6, onde a primeira imagem corresponde à tomografia computadorizada original, no meio a segmentação realizada pelo médico, destacando os rins (verde), o tumor (vermelho) e o cisto (azul) e a direita a segmentação gerada pela Dual-Scale SE UNet, que também identifica essas estruturas.

Nesta seção, apresentamos resultados qualitativos das segmentações obtidas. A rede demonstra aprendizado consistente ao analisar os casos 00170 e 00051 da base KiTS21, mostrados respectivamente nas Figuras 6 e 7. Na primeira coluna observa-se a tomografia original, seguida pela segmentação manual do médico com rins (verde), tumor (vermelho) e cisto (azul) e à direita, a segmentação gerada pela Dual-Scale SE UNet, que mostra boa correspondência com estas estruturas anatômicas.

A análise do caso 00170 e 00051 revela que a rede segmenta as estruturas de maneira semelhante à do especialista, embora apresente pequenas diferenças, especialmente nos contornos e nas regiões mais complexas. Essas variações podem ser resultados de desafios na diferenciação de estruturas semelhantes ou de diferenças anatômicas individuais, o que indica que, apesar do desempenho da rede, ainda é necessário realizar ajustes ou aperfeiçoamentos para alcançar maior precisão.

Tabela 4. Resultados dos valores de Dice pela metodologia proposta, comparados com a U-Net padrão.

Modelos	Dice Rim	Dice Tumor	Dice Cisto
U-Net	0,93	0,83	0,86
Dual-Scale SE UNet	0,93	0,86	0,93

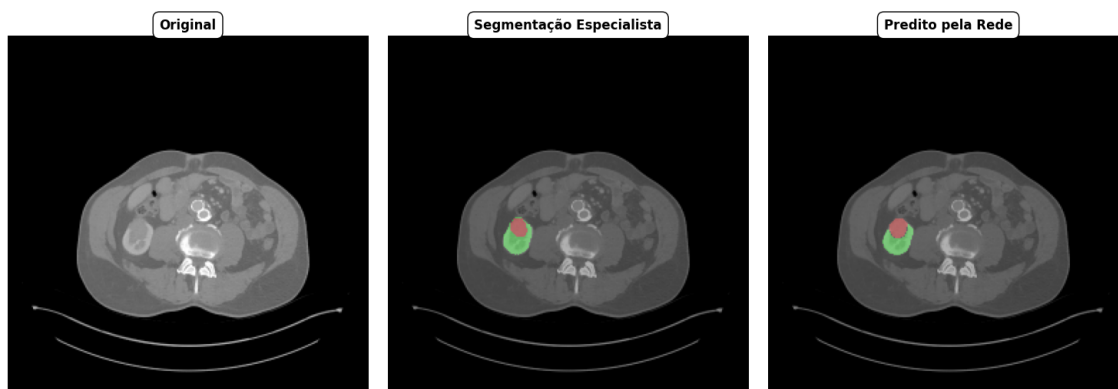


Figura 6. Estudo de caso 00170 da base Kits21.

Comparando com os trabalhos publicados na literatura, apresentados na Tabela 5, observa-se que algumas abordagens, como as de [Zhao et al. 2022], [Türk et al. 2020], [Zhang et al. 2019] e [George 2022], alcançaram um Dice de 0,97 na segmentação do rim, enquanto a Dual-Scale SE UNet obteve um valor inferior, de 0,93. Ao compararmos com outro estudo que também utilizou imagens 2D [Matos et al. 2023], nota-se um padrão de desempenho inferior, sugerindo uma limitação intrínseca a esse tipo de abordagem. Especificamente, a utilização de imagens 2D pode levar à perda de continuidade espacial entre as fatias, comprometendo a representação tridimensional da anatomia. Diferentemente das imagens 3D, que preservam a coerência entre os cortes axiais, coronais e sagitais, as imagens 2D tratam cada fatia isoladamente, o que pode resultar na perda de informações contextuais importantes para a segmentação precisa de estruturas anatômicas contínuas como o rim.

Na segmentação do Tumor, a Dual-Scale SE UNet alcançou um Dice de 0,86, posicionando-se entre os melhores resultados da literatura, cujos valores variam de 0,80 a 0,87. Esse desempenho competitivo reforça a eficácia da abordagem proposta para essa estrutura específica.

Já na segmentação de Massas (Tumor + Cisto), somente três estudos reportaram valores, e a Dual-Scale SE UNet obteve o maior Dice (0,89), seguido por [Zhao et al. 2022] com 0,88 e [Matos et al. 2023] com 0,84.

De maneira geral, os resultados demonstram que a Dual-Scale SE UNet apresenta um desempenho competitivo, especialmente na segmentação do Tumor e das Massas. No entanto, comparando com outros estudos, a segmentação do Rim ainda pode ser refinada para alcançar valores próximos aos do estado da arte.

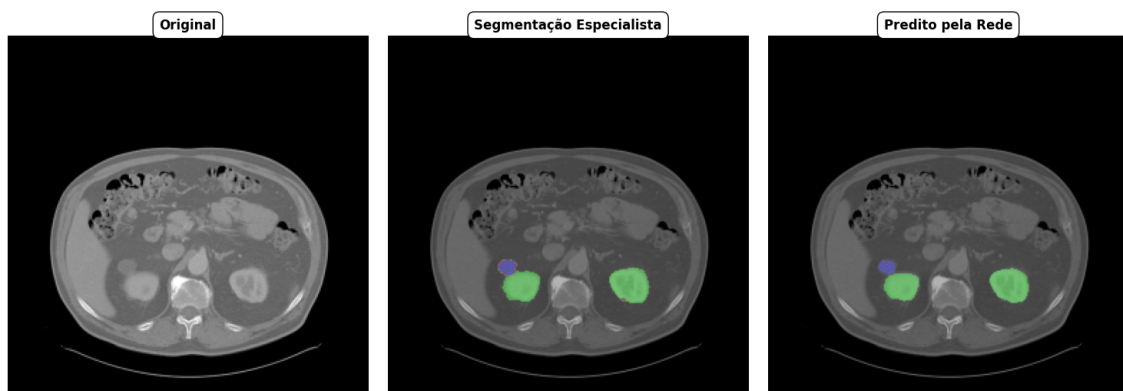


Figura 7. Estudo de caso 00051 da base Kits21.

Tabela 5. Comparando com a literatura.

Redes	Dice Rim	Dice Tumor	Dice Massas
[Golts et al. 2022]	–	0,83	–
[Zhao et al. 2022]	0,97	0,86	0,88
[Hou et al. 2019]	0,90	0,80	–
[Zhang et al. 2019]	0,97	0,83	–
[Isensee and Maier-Hein 2019]	0,97	0,85	–
[Türk et al. 2020]	0,97	0,86	–
[George 2022]	0,97	0,87	–
[Matos et al. 2023]	0,94	0,84	0,84
Dual-Scale SE UNet	0,93	0,86	0,89

5. Conclusão.

Este estudo propôs a Dual-Scale SE UNet, uma arquitetura neural inovadora que combina blocos de atenção Squeeze-and-Excitation (SE) com convoluções em escalas duais (3×3 e 7×7) para aprimorar a segmentação de rins, tumores e cistos em imagens de tomografia computadorizada. Os resultados demonstram a eficácia da abordagem, com coeficiente Dice de 0,93 para rins, 0,86 para tumores e 0,93 para cistos, superando a U-Net tradicional e competindo com o estado da arte.

Apesar do desempenho promissor, a segmentação de rins ainda pode ser melhorada, especialmente quando comparada a métodos baseados em imagens 3D, que preservam melhor a continuidade espacial. No entanto, a Dual-Scale SE UNet destacou-se na segmentação de tumores e massas, alcançando resultados competitivos e reforçando seu potencial para aplicações clínicas. A metodologia proposta não somente reduz a carga de trabalho dos profissionais de saúde, mas também contribui para diagnósticos mais precisos e eficientes.

Como perspectivas futuras, os autores sugerem explorar novos módulos de atenção, a automação do processo de configuração dos blocos e a aplicação do modelo em outras modalidades de imagem e tipos de tumores. Esses avanços podem ampliar a aplicabilidade clínica da rede, consolidando-a como uma ferramenta valiosa no auxílio ao diagnóstico precoce e no tratamento do câncer renal.

Agradecimento(s)

Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brasil – Código de Financiamento 001, ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brasil, e à Fundação de Amparo à Pesquisa e Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA) (Brasil) pelo apoio financeiro.

Referências

- American Cancer Society (2024). What is kidney cancer? Disponível em: <https://www.cancer.org/cancer/types/kidney-cancer/about/what-is-kidney-cancer.html>. Acessado em: 1 de maio de 2024.
- Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B. (2011). Algorithms for hyperparameter optimization. *Advances in neural information processing systems*, 24.
- Brenner, D. J. and Hall, E. J. (2007). Computed tomography—an increasing source of radiation exposure. *New England journal of medicine*, 357(22):2277–2284.
- Fleiss, J. L., Levin, B., Paik, M. C., et al. (1981). The measurement of interrater agreement. *Statistical methods for rates and proportions*, 2(212-236):22–23.
- George, Y. (2022). A coarse-to-fine 3d u-net network for semantic segmentation of kidney ct scans. In Heller, N., Isensee, F., Trofimova, D., Tejpaul, R., Papanikolopoulos, N., and Weight, C., editors, *Kidney and Kidney Tumor Segmentation*, pages 137–142, Cham. Springer International Publishing.
- Gloeckler Ries, L. A., Reichman, M. E., Lewis, D. R., Hankey, B. F., and Edwards, B. K. (2003). Cancer survival and incidence from the surveillance, epidemiology, and end results (seer) program. *The Oncologist*, 8(6):541–552.
- Golts, A., Khapun, D., Shats, D., Shoshan, Y., and Gilboa-Solomon, F. (2022). An ensemble of 3d u-net based models for segmentation of kidney and masses in ct scans. In Heller, N., Isensee, F., Trofimova, D., Tejpaul, R., Papanikolopoulos, N., and Weight, C., editors, *Kidney and Kidney Tumor Segmentation*, pages 103–115, Cham. Springer International Publishing.
- Heller, N., Isensee, F., Trofimova, D., Tejpaul, R., Zhao, Z., Chen, H., Wang, L., Golts, A., Khapun, D., Shats, D., Shoshan, Y., Gilboa-Solomon, F., George, Y., Yang, X., Zhang, J., Zhang, J., Xia, Y., Wu, M., Liu, Z., Walczak, E., McSweeney, S., Vasdev, R., Hornung, C., Solaiman, R., Schoepfoerster, J., Abernathy, B., Wu, D., Abdulkadir, S., Byun, B., Spriggs, J., Struyk, G., Austin, A., Simpson, B., Hagstrom, M., Virnig, S., French, J., Venkatesh, N., Chan, S., Moore, K., Jacobsen, A., Austin, S., Austin, M., Regmi, S., Papanikolopoulos, N., and Weight, C. (2023). The kits21 challenge: Automatic segmentation of kidneys, renal tumors, and renal cysts in corticomedullary-phase ct.
- Heller, N., Sathianathan, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., Rosenberg, J., Blake, P., Rengel, Z., Oestreich, M., Dean, J., Tradewell, M., Shah, A., Tejpaul, R., Edgerton, Z., Peterson, M., Raza, S., Regmi, S., Papanikolopoulos, N., and Weight, C. (2019). The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. Disponível em: <https://github.com/neheller/kits19>.

- Hou, X., Xie, C., Li, F., and Nan, Y. (2019). Cascaded semantic segmentation for kidney and tumor. *Submissions to the*.
- Hu, J., Shen, L., Albanie, S., Sun, G., and Wu, E. (2017). Squeeze-and-excitation networks. *arXiv preprint arXiv:1709.01507*, updated in 2019.
- Isensee, F. and Maier-Hein, K. H. (2019). An attempt at beating the 3d u-net. *arXiv preprint arXiv:1908.02182*.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2018). Focal loss for dense object detection.
- Long, J., Shelhamer, E., and Darrell, T. (2014). Fully convolutional networks for semantic segmentation.
- Matos, C., Oliveira, M., Diniz, J., Fernandes, A., Junior, G. B., and Paiva, A. (2023). Ppm-deeplab: Módulo de pirâmide de pooling como codificador da rede deeplabv3+ para segmentação de rins, cistos e tumores renais. In *Anais do XXIII Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 210–221, Porto Alegre, RS, Brasil. SBC.
- Oliveira, M., Matos, C., Júnior, G. B., Paiva, A., Almeida, J., Costa, G., Levy, M., and Freitas, M. (2022). Ppm-unet: Uma rede neural convolucional para a segmentação de rins em imagens de tc. In *Anais do XXII Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 268–276, Porto Alegre, RS, Brasil. SBC.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation.
- Siegel, R. L., Giaquinto, A. N., and Jemal, A. (2024). Cancer statistics, 2024. *CA: A Cancer Journal for Clinicians*, 74(1):12–49.
- Stewart, B. W., Wild, C. P., et al. (2014). *World cancer report 2014*. IARC Press, International Agency for Research on Cancer, Lyon, France.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Türk, F., Lüy, M., and Barışç1, N. (2020). Kidney and renal tumor segmentation using a hybrid v-net-based model. *Mathematics*, 8(10):1772.
- Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., and Xiao, B. (2019). Deep high-resolution representation learning for visual recognition. *TPAMI*.
- Zhang, Y., Wang, Y., Hou, F., Yang, J., Xiong, G., Tian, J., and Zhong, C. (2019). Cascaded volumetric convolutional network for kidney tumor segmentation from ct volumes. *arXiv preprint arXiv:1910.02235*.
- Zhao, Z., Chen, H., and Wang, L. (2022). A coarse-to-fine framework for the 2021 kidney and kidney tumor segmentation challenge. In Heller, N., Isensee, F., Trofimova, D., Tejpaul, R., Papanikolopoulos, N., and Weight, C., editors, *Kidney and Kidney Tumor Segmentation*, pages 53–58, Cham. Springer International Publishing.