UNIVERSIDADE FEDERAL FLUMINENSE

LUCAS MENDONÇA DE MORAIS CAVALCANTE

Análise do Compartilhamento de Mapas de Características entre Fatias Consecutivas em Arquiteturas U-Net Bidimensionais para Segmentação Mandibular

NITERÓI

2026

LUCAS MENDONÇA DE MORAIS CAVALCANTE

Análise do Compartilhamento de Mapas de Características entre Fatias Consecutivas em Arquiteturas U-Net Bidimensionais para Segmentação Mandibular

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do Grau de Mestre em Computação. Área de concentração: Ciência da Computação

Orientador:
Aura Conci

Coorientador:
Leandro Augusto Frata Fernandes

NITERÓI

2026

Ficha catalográfica automática - SDC/BEE
Gerada com informações fornecidas pelo autor

C376a Cavalcante, Lucas Mendonça de Morais
Análise do Compartilhamento de Mapas de Características entre Fatias Consecutivas em Arquiteturas U-Net Bidimensionais para Segmentação Mandibular / Lucas Mendonça de Morais Cavalcante. - 2026.
64 f.: il.

Orientador: Aura Conci.
Coorientador: Leandro Augusto Frata Fernandes.
Dissertação (mestrado)-Universidade Federal Fluminense, Instituto de Computação, Niterói, 2026.

1. Aprendizado Profundo. 2. Segmentação. 3. Mandíbula. 4. Produção intelectual. I. Conci, Aura, orientadora. II. Fernandes, Leandro Augusto Frata, coorientador. III. Universidade Federal Fluminense. Instituto de Computação. IV. Título.

CDD - XXX

Bibliotecário responsável: Debora do Nascimento - CRB7/6368

Lucas Mendonça de Morais Cavalcante

Análise do Compartilhamento de Mapas de Características entre Fatias Consecutivas em Arquiteturas U-Net Bidimensionais para Segmentação Mandibular

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do Grau de Mestre em Computação. Área de concentração: Ciência da Computação.

Aprovada em Agosto de 2026.

BANCA EXAMINADORA

Prof. Aura Conci - Orientadora, UFF

Prof. Leandro Augusto Frata Fernandes - Co-orientador, UFF

Prof. Flávio Luiz Seixas, UFF

Prof. João Paulo Papa, UNESP

Niterói

2026

Dedico este trabalho a Janaina Mendonça de Morais, minha mãe, e Rosembergue da Silva Cavalcante Junior, meu pai. Pessoas que sacrificaram uma parte não trivial de suas vidas em minha criação, da qual sou eternamente grato. Também dedico aos meus familiares que me apoiarem durante toda a minha trajetória acadêmica, em especial minha dinda Gildete Mendonça de Morais, que me apoiou durante todos os momentos.

Agradecimentos

A minha orientadora Aura Conci e meu co-orientador Leandro Augusto Frata Fernandes, que me mostraram os caminhos a serem seguidos e pela confiança depositada.

Gostaria de expressar meu sincero agradecimento à CAPES pelo apoio concedido ao longo desta pesquisa, fundamental para a realização deste trabalho.

Resumo

No contexto de segmentação de mandíbulas em tomografias computadorizadas axiais, esta dissertação investigou os efeitos do compartilhamento progressivo de mapas de características produzidos pelas camadas internas da U-Net entre imagens consecutivas. O foco não foi o desenvolvimento de uma nova arquitetura de estado da arte, mas a investigação sistemática do efeito desse compartilhamento em diferentes níveis da arquitetura U-Net. O objetivo central da pesquisa foi analisar se o compartilhamento progressivo de mapas de características seria capaz de incorporar contexto entre fatias consecutivas em uma arquitetura U-Net bidimensional, assim como compreender como diferentes níveis e direções de propagação influenciam o comportamento da rede. Para isso, foram realizados experimentos nos quais a reutilização de informações é progressivamente aumentada. Inicialmente, apenas a fatia adjacente é utilizada como informação adicional. Em seguida, os mapas de características gerados pelas camadas do *encoder* e *decoder* passaram a ser compartilhados progressivamente, de forma que todas as representações internas fossem compartilhadas. Foram avaliadas duas estratégias de propagação: *forward*, da entrada para a saída da rede, e *reverse*, da saída em direção à entrada, aumentando progressivamente o nível de transmissão de informações entre as camadas da rede. Utilizando o conjunto de dados *Public Domain Database for Computational Anatomy* (PDDCA), contendo 40 tomografias computadorizadas anotadas por especialistas, os resultados obtidos mostram que as configurações mais superficiais mantiveram desempenho semelhante ao modelo U-Net de referência, sem diferenças estatisticamente significativas, enquanto o compartilhamento em camadas mais profundas resultou em degradação da Interseção sobre União (IoU), Índice de Similaridade Dice e 95% Distância de Hausdorff (95HD). Esses resultados permitiram analisar o comportamento do compartilhamento de representações internas em diferentes níveis da arquitetura, fornecendo evidências experimentais de que, no contexto investigado, a propagação direta de mapas de características entre fatias consecutivas, sem mecanismos mais sofisticados de seleção de características, não produz ganhos consistentes para a segmentação mandibular em arquiteturas U-Net bidimensionais.

Palavras-chave: Aprendizado Profundo; Mapas de Características; Contexto Inter-Fatias; U-Net; Tomografia Computadorizada; Segmentação Mandibular;

Abstract

In the context of mandibular segmentation in axial computed tomography images, this dissertation investigated the effects of progressively sharing feature maps generated by the internal layers of U-Net across consecutive slices. The focus was not on developing a new state-of-the-art architecture, but rather on systematically investigating the impact of feature map sharing at different levels of the U-Net architecture. The primary objective was to determine whether the progressive sharing of feature maps could incorporate inter-slice contextual information into a two-dimensional U-Net architecture, as well as to understand how different sharing levels and propagation directions influence network behavior. To this end, a series of experiments was conducted in which information reuse was progressively increased. Initially, only the adjacent slice was used as additional contextual information. Subsequently, the feature maps generated by the encoder and decoder layers were progressively shared until all internal representations were propagated between consecutive slices. Two propagation strategies were evaluated: forward, from the network input toward the output, and reverse, from the output toward the input, progressively increasing the level of information transmission across network layers. Experiments were performed using the Public Domain Database for Computational Anatomy (PDDCA), which contains 40 computed tomography scans manually annotated by experts. The results show that feature sharing at shallower layers achieved performance comparable to the baseline U-Net, with no statistically significant differences, whereas sharing at deeper layers led to degradation in Intersection over Union (IoU), Dice Similarity Coefficient (DSC), and the 95% Hausdorff Distance (95HD). These findings provide insights into the behavior of internal representation sharing at different levels of the U-Net architecture and offer experimental evidence that, within the investigated setting, the direct propagation of feature maps across consecutive slices, without more sophisticated feature selection mechanisms, does not provide consistent improvements for mandibular segmentation using two-dimensional U-Net architectures.

Keywords: Deep Learning; Feature Maps; Inter-Slice Context; U-Net; Computed Tomography; Mandibular Segmentation.

Lista de Figuras

Lista de Tabelas

Sumário

1 Introdução

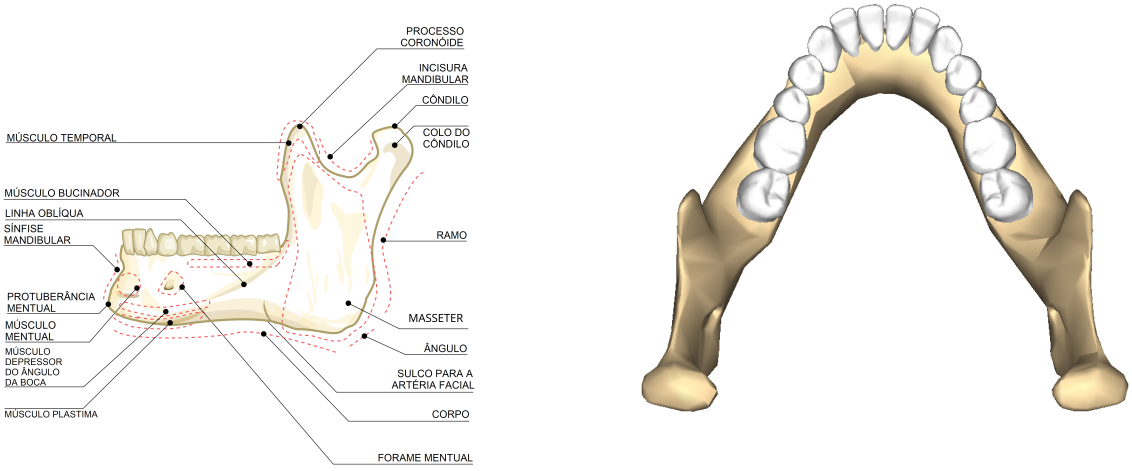
A segmentação de imagens médicas consiste no processo de delimitação de estruturas anatômicas de interesse em exames de imagem. É uma etapa fundamental em diversas aplicações, incluindo diagnóstico, planejamento cirúrgico, reconstrução tridimensional e acompanhamento de tratamentos. Tradicionalmente, esse processo é realizado de forma manual por especialistas, demandando tempo e um alto grau de atenção. É uma tarefa de natureza repetitiva e cansativa para profissionais qualificados, a qual poderia ser direcionada para tarefas de maior complexidade. Essas limitações têm motivado o desenvolvimento de métodos automáticos de segmentação.

Diversas técnicas foram propostas para a segmentação automática de estruturas anatômicas em imagens médicas [(Wang, R. *et al.*, 2022), (Xu *et al.*, 2024), (Xia *et al.*, 2025)], incluindo métodos baseados em limiarização (Otsu, 1979), crescimento de regiões (Adams; Bischof, 1994), modelos deformáveis (Kass; Witkin; Terzopoulos, 1988) e, mais recentemente, abordagens de aprendizado de máquina (Criminisi; Shotton, 2013). Nos últimos anos, métodos baseados em aprendizado profundo, uma subárea do aprendizado de máquina, passaram a apresentar desempenho superior em diversas tarefas de segmentação médica, devido à sua capacidade de aprender representações diretamente a partir dos dados.

O avanço do poder computacional, especialmente com o uso de unidades de processamento gráfico (GPUs), aliado à disponibilidade de grandes bases de dados, impulsionou a popularização das redes neurais profundas, em particular das redes neurais convolucionais (CNNs) (Long; Shelhamer; Darrell, 2015). Essas redes passaram a alcançar desempenho competitivo ou superior ao estado da arte em diversas tarefas de segmentação, eliminando a necessidade de iterações manuais de extração de características e permitindo o aprendizado automático de representações relevantes. Nesse contexto, a U-Net (Ronneberger; Fischer; Brox, 2015) destacou-se como uma das arquiteturas mais utilizadas em segmentação biomédica, devido à sua estrutura em forma de *encoder-decoder* com *skip connections*, que permite a preservação de informações de baixa e alta resolução. Devido

à sua eficácia, a U-Net passou a ser amplamente empregada na segmentação de diferentes estruturas anatômicas, incluindo a mandíbula.

A mandíbula é uma estrutura anatômica de grande interesse na área odontológica, sendo essencial em aplicações como planejamento cirúrgico e fabricação de próteses, que dependem de uma delimitação precisa dessa região. Sua morfologia apresenta variações significativas ao longo de sua extensão, podendo ser dividida em duas regiões principais, sendo elas o corpo e o ramo mandibular, como pode ser visto na Figura 1. O corpo da mandíbula geralmente apresenta uma estrutura mais homogênea e com formato característico em "U", sendo relativamente mais fácil de ser distinguido em imagens tomográficas devido ao contraste de intensidades em unidades de Hounsfield (Hounsfield, 1973) em relação aos tecidos adjacentes. Em contraste, o ramo mandibular apresenta maior complexidade anatômica, com menor contraste e maior presença de estruturas vizinhas com intensidades semelhantes, o que dificulta sua segmentação automática. Nessa região, mesmo especialistas frequentemente recorrem à análise de fatias adjacentes e ao conhecimento anatômico para auxiliar na delimitação correta da estrutura (Breeland; Aktar; Patel, 2023).



(a) Adaptado de: Hariadhi (2020). (b) Fonte: Anatomography (2012).

Figura 1: Anatomia da mandíbula humana.

1.1 Motivação

A segmentação automática de mandíbula em imagens de tomografia computadorizada ainda apresenta desafios significativos. Embora modelos baseados em redes neurais profundas tenham alcançado bons resultados em diferentes avaliadores de desempenho, sua

performance em cenários clínicos reais ainda pode ser afetada por variações anatômicas, artefatos e ambiguidades estruturais. Essas limitações podem impactar a consistência e a confiabilidade dos modelos em aplicações práticas.

Abordagens de segmentação baseadas em aprendizado profundo para imagens médicas podem ser divididas em métodos bidimensionais e tridimensionais (Sabanci; Aslan; Aslan, 2026), cada um com vantagens e limitações distintas. Métodos 2D tendem a ser mais leves computacionalmente e amplamente utilizados devido à maior disponibilidade de dados anotados. Por outro lado, abordagens 3D são capazes de explorar o contexto volumétrico entre fatias, o que pode melhorar a consistência espacial das segmentações, porém com maior custo computacional e demanda de memória. No caso da mandíbula, o contexto é particularmente relevante. Enquanto o corpo mandibular apresenta delimitação relativamente mais simples, regiões como o ramo e o côndilo possuem maior complexidade anatômica e pouco contraste com estruturas adjacentes. Em abordagens 2D, cada fatia é processada de forma independente, o que pode levar à perda de informação contextual entre cortes consecutivos, prejudicando a performance. Já as abordagens 3D, apesar de mais completas, podem ser inviáveis em muitos cenários devido ao alto custo computacional.

Uma alternativa é reutilizar representações internas extraídas de uma fatia nas fatias adjacentes, buscando incorporar informações contextuais sem o elevado custo computacional das convoluções tridimensionais presentes em outros métodos. Diante desse cenário, esta dissertação investiga sistematicamente o compartilhamento de mapas de características entre fatias consecutivas como uma estratégia para introduzir contexto inter-fatia em arquiteturas U-Net bidimensionais. A segmentação de mandíbulas foi utilizada como estudo de caso por representar um problema em que a informação contextual entre cortes consecutivos possui grande relevância. O objetivo da pesquisa não é propor uma nova arquitetura de estado da arte, mas compreender como diferentes estratégias de compartilhamento influenciam as representações internas da rede e seu desempenho na tarefa de segmentação.

1.2 Metodologia

Os experimentos foram planejados para investigar sistematicamente o efeito do compartilhamento progressivo de mapas de características entre fatias consecutivas. Para isso, diferentes configurações de arquitetura, correspondentes a diferentes níveis de comparti-

lhamento na U-Net, foram avaliadas por meio de treinamentos independentes, permitindo isolar o impacto de cada estratégia. O desempenho de cada configuração foi analisado principalmente por meio das métricas Interseção sobre União (IoU), coeficiente de similaridade de Dice e Distância de Hausdorff a 95% (HD95), sendo as diferenças estatísticas avaliadas pelo teste não paramétrico de Wilcoxon.

Inicialmente, foi considerada a U-Net sem alterações como *baseline*. Em seguida, foram conduzidos experimentos nos quais a informação da fatia antecedente é incorporada à fatia atual. Progressivamente, esse compartilhamento é estendido para incluir os mapas de características gerados pelas camadas do *encoder* e do *decoder*, até abranger todas as representações internas e a saída da rede; denominamos esses experimentos de *forward*.

Adicionalmente, foi avaliada uma variação inversa da estratégia de compartilhamento, na qual a propagação de informações ocorre a partir da saída da rede em direção às camadas intermediárias. Essa estratégia tem o propósito de investigar se a direção da propagação das informações influencia o processo de reconstrução das representações no *decoder*, onde acontece a formação da imagem; denominamos esses experimentos de *reverse*.

1.3 Objetivos

O objetivo geral desta dissertação é investigar sistematicamente os efeitos do compartilhamento progressivo de mapas de características entre fatias consecutivas em arquiteturas U-Net bidimensionais, utilizando a segmentação mandibular como estudo de caso para analisar sua capacidade de incorporar contexto inter-fatia.

Nesse sentido, foram definidos os seguintes objetivos específicos:

- Implementar uma U-Net 2D *baseline* para segmentação de mandíbulas em tomografias computadorizadas;
- Implementar diferentes estratégias de compartilhamento de mapas de características entre fatias consecutivas.
- Investigar experimentalmente o impacto do compartilhamento em diferentes níveis hierárquicos da arquitetura.
- Investigar a influência da direção de propagação das informações.
- Avaliar quantitativamente e qualitativamente o efeito dessas estratégias.

- Analisar como o compartilhamento influencia o comportamento da rede e a incorporação de contexto entre fatias.

1.4 Principais Contribuições

As principais contribuições desse trabalho são.

- Investigação sistemática do compartilhamento progressivo de mapas de características entre fatias consecutivas em arquiteturas U-Net 2D.
- Avaliação sistemática das diferentes estratégias de compartilhamento em diferentes níveis hierárquicos da arquitetura, utilizando os esquemas de propagação *forward* e *reverse*.
- Comparação quantitativa e qualitativa com uma U-Net de referência.
- Apresentação de evidências experimentais de que, no contexto investigado, estratégias diretas de compartilhamento de mapas de características não produzem ganhos consistentes sem mecanismos adaptativos de seleção.

1.5 Organização dos Capítulos

Esta dissertação está organizada em seis capítulos. O Capítulo 2 apresenta os fundamentos teóricos que serviram de base para o desenvolvimento deste trabalho. No Capítulo 3 é realizada uma revisão da literatura com os trabalhos mais relevantes em relação à dissertação. No Capítulo 4, é descrita a metodologia adotada. O Capítulo 5 apresenta os resultados dos experimentos realizados e os procedimentos empregados em sua execução. No Capítulo 6, são discutidos e analisados os resultados obtidos. Por fim, o Capítulo 7 apresenta as conclusões do estudo, bem como as considerações finais e possíveis perspectivas para trabalhos futuros.

2 Fundamentação Teórica

Este capítulo apresenta a fundamentação teórica necessária para a compreensão dos conceitos discutidos nesta dissertação. Como o objetivo final é a investigação sistemática do compartilhamento de mapas de características entre fatias consecutivas em arquiteturas U-Net bidimensionais, este capítulo enfatiza os conceitos relacionados à representação interna das CNNs, às diferenças entre abordagens 2D e 3D e às estratégias existentes para a incorporação de contexto inter-fatia.

2.1 Segmentação de Imagens Médicas

A segmentação de imagens médicas consiste no processo de atribuição de rótulos semânticos a pixels em representações 2D ou a voxels em volumes 3D, em exames de diagnóstico por imagem, com o objetivo de delimitar estruturas anatômicas de interesse. Formalmente, dado um domínio de imagem R, correspondente ao conjunto de todos os pixels ou voxels, o objetivo é particionar R em sub-regiões disjuntas que representam classes anatômicas distintas, como órgãos, ossos, tumores ou tecidos moles, extraindo uma máscara binária ou multiclasse representando a localização dos mesmos (Gonzalez; Woods, 2010).

Essa extração é fundamental para auxílio no planejamento cirúrgico, confecção de próteses e radioterapia (Ma *et al.*, 2024). Contudo, a automatização desse processo enfrenta desafios intrínsecos aos dados de aquisição, como o baixo contraste entre estruturas adjacentes, presença de artefatos (ruídos térmicos, movimentos de pacientes e materiais metálicos) e alta variabilidade anatômica entre pacientes, devido a diferenças de idade, presença de patologias ou deformidades congênitas.

2.2 Redes Neurais Convolucionais

Redes Neurais Convolucionais (*Convolutional Neural Networks*) (CNNs) se estabeleceram como um método de referência para o processamento de imagens em tarefas como

a de segmentação. Diferentes das redes totalmente conectadas (*Fully Connected Neural Networks*), as CNNs preservam a topologia espacial dos dados por meio de três propriedades fundamentais: campos receptivos locais, compartilhamento de pesos e subamostragem (LeCun *et al.*, 1989).

A operação central é a convolução. Formalmente, dada uma imagem de entrada I e um kernel K de dimensões $M \times N$, em que i e j representam, respectivamente, os índices da linha e da coluna da imagem de saída, e u e v representam, respectivamente, os índices da linha e coluna dentro do kernel, o mapa de características S é obtido por:

$$S(i,j) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} I(i+u,j+v)\, K(u,v) \tag{2.1}$$

Na prática, devido à eficiência computacional, a operação implementada é a correlação cruzada, combinação do resultado de diferentes kernels definidos pelo par de canal de entrada, canal de saída e viés. Formalmente, dada uma entrada de dimensões (N,C_{in},H,W) e uma saída $(N,C_{out},H_{out},W_{out})$, em que N representa o tamanho do lote (*batch size*) C_{in} o número de canais de entrada, C_{out} o número de canais de saída, e H e W representam, respectivamente, a altura e a largura da entrada, o valor de cada mapa de características da saída é obtido pela soma da correlação cruzada entre cada canal de entrada e seu respectivo kernel, somado ao termo de viés (*bias*). Assim, $out(N_i, C_{out_j})$ representa o j-ésimo canal de saída para o i-ésimo elemento do lote; $bias(C_{out_j})$ corresponde ao termo de viés associado ao j-ésimo canal de saída; $weight(C_{out_j}, k)$ representa o kernel aprendido que conecta o k-ésimo canal da entrada ao i-ésimo elemento do lote; e $\star$ denota a operação de correlação cruzada bidimensional válida (Paszke *et al.*, 2019). Dessa forma, a saída da camada é dada por:

$$\text{out}(N_i, C_{\text{out}_j}) = \text{bias}(C_{\text{out}_j}) + \sum_{k=0}^{C_{\text{in}}-1} \text{weight}(C_{\text{out}_j},k) \star \text{input}(N_i,k) \tag{2.2}$$

As operações de convolução, parametrizadas pelos kernels aprendidos durante o treinamento, atuam como extratoras de características locais. À medida que o sinal avança pelas camadas profundas da rede, ocorre uma extração hierárquica de atributos, com as camadas iniciais detectando padrões de baixa abstração como bordas, gradientes e texturas, enquanto as camadas profundas representam características de maior nível semântico. No contexto da segmentação de imagens médicas, essas representações codificam as estruturas anatômicas de interesse.

Nesta dissertação, esses mapas assumem papel central, pois são justamente as representações compartilhadas entre fatias consecutivas.

Após a operação de convolução, é comum aplicar uma função de ativação não linear, sendo a Unidade Linear Retificada (*Rectified Linear Unit*) (ReLU) a mais empregada em CNNs. Com isso, os valores negativos são mapeados para zero, enquanto valores positivos permanecem inalterados. Essa operação introduz não linearidade ao modelo, permitindo que a rede aprenda representações mais complexas sem aumentar significativamente o custo computacional. A função é definida como:

$$\mathrm{ReLU}(x) = \max(0,x), \tag{2.3}$$

Outra operação amplamente empregada é o *Max Pooling*; ela reduz a dimensionalidade espacial dos mapas de características, diminui o custo computacional e torna as representações menos sensíveis a pequenas variações de posição. Formalmente, considerando um mapa de características de entrada X e uma janela de *pooling* de dimensões $p \times p$, em que i e j representam, respectivamente, os índices da linha e da coluna do mapa de saída, e m e n representam, respectivamente, os índices da linha e da coluna dentro da janela de *pooling* (Reza *et al.*, 2026), a saída é definida por:

$$P_{ij} = \max_{0 \le m < p} \max_{0 \le n < p} X_{(i+m)(j+n)}. \tag{2.4}$$

2.3 Funções de Perda

O treinamento dessas arquiteturas baseia-se na minimização de uma função de perda (*loss function*), responsável por quantificar a diferença entre a máscara gerada pelo modelo (p) e a máscara de referência (y). Uma das funções mais utilizadas em tarefas de segmentação binária é a Entropia Cruzada Binária (*Binary Cross-Entropy*) (BCE) (Ekman, 2021). Essa função avalia, pixel a pixel, a diferença entre as probabilidades do pixel pertencer à máscara ou ao fundo previstas pelo modelo e os dados de referência, penalizando predições incorretas de forma mais intensa do que predições corretas. Durante o treinamento, a minimização dessa perda ajuda o modelo a produzir máscaras cada vez mais próximas da referência.

Seja $y_i \in [0,1]$ o rótulo de referência associado ao i-ésimo pixel, em que $y_i = 1$ indica a presença da região de interesse e $y_i = 0$ representa o fundo da imagem. Seja $p_i \in [0,1]$ a

probabilidade estimada pelo modelo para a classe positiva no mesmo pixel. Considerando ainda um conjunto de N pixels pertencentes ao conjunto de treinamento. A função de perda BCE é então definida como a média da divergência entre os valores de referência e as probabilidades preditas pelo modelo, sendo expressa por:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)] \tag{2.5}$$

2.4 Métricas de Avaliação

A avaliação de modelos de segmentação médica necessita de métricas que sejam capazes de quantificar a sobreposição entre as segmentações preditas pelo modelo e as de referência, assim como a correspondência entre seus contornos. Métricas tradicionais, como acurácia e precisão, podem ser pouco representativas nesse contexto, especialmente em imagens com alto desbalanceamento entre as classes. Isso ocorre porque uma grande proporção dos pixels pode pertencer ao fundo ou a estruturas que não pertencem ao objeto de interesse, fazendo com que apresentem valores elevados mesmo quando a segmentação do objeto de interesse não é satisfatória. Dessa forma, neste trabalho foram utilizadas as métricas IoU, Dice 95HD.

A métrica IoU, quantifica a razão entre a interseção e a união de duas regiões. Considerando A como o conjunto de pixels pertencentes à segmentação predita e B como o conjunto de pixels pertencentes à segmentação de referência, a IoU é definida por:

$$(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{2.6}$$

O coeficiente de similaridade de Dice também quantifica a sobreposição entre a segmentação predita e a segmentação de referência. Diferentemente da IoU, o Dice normaliza a interseção pela soma das áreas das duas regiões, sendo definido por:

$$Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|} \tag{2.7}$$

Além da sobreposição das regiões, também é importante avaliar a correspondência espacial entre seus contornos. Para esse propósito, foi utilizada a distância de Hausdorff, que mede a maior distância mínima entre pontos pertencentes a dois conjuntos. Considerando A e B como os conjuntos de pontos correspondentes às superfícies ou contornos

das segmentações, a distância de Hausdorff é definida como:

$$d_H(A, B) = \max \left\{ \sup_{a \in A} d(a, B), \sup_{b \in B} d(A, b) \right\} \tag{2.8}$$

em que $d(a,B)$ representa a distância mínima entre o ponto a e qualquer ponto pertencente ao conjunto B. Dessa forma, a distância de Hausdorff permite quantificar as diferenças espaciais entre os contornos da segmentação predita e da segmentação de referência. Neste trabalho, foi utilizada a distância de Hausdorff no 95º percentil (95HD), em vez da distância máxima. Essa escolha reduz a influência de *outliers*, que poderiam produzir valores desproporcionalmente elevados. O cálculo da distância foi realizado diretamente no espaço discreto dos pixels da imagem.

2.5 A Arquitetura U-Net e suas Variantes Dimensionais

Proposta originalmente por Ronneberger, Fischer e Brox (2015), a U-Net tornou-se a arquitetura de referência em segmentação de imagens biomédicas. Sua estrutura simétrica assemelha-se à forma da letra "U", dividida estritamente em um caminho de encolhimento (*encoder*) e um caminho de expansão (*decoder*). Na Figura 2 é possível observar a arquitetura original da U-Net proposta por Ronneberger, Fischer e Brox (2015). O fluxo direto é representado pelas setas azuis claras e roxas, representando, respectivamente, as operações de subamostragem pelo máximo (*max pooling*) e convolução de ampliação (*up convolution*); as operações de convolução e ativação ReLu são representadas pelas setas azuis escuras e as conexões de salto (*skip connections*) representadas por setas vermelhas curvas. "D" representa os blocos de convolução de subamostragem (*encoder*); "B" representa o bloco de gargalo (*bottleneck*); "U" representa os blocos de convolução de ampliação das amostras (*decoder*).

Na Figura 2, a arquitetura não é apresentada em seu formato tradicional em "U", mas sim em uma disposição mais direta e vertical. Essa escolha foi adotada para facilitar a introdução posterior do conceito de compartilhamento de características, permitindo que sua representação seja apresentada de forma mais intuitiva e favorecendo a compreensão da arquitetura.

O *encoder* funciona como uma rede convolucional convencional, extraindo representações semânticas a partir da perda de resolução espacial. O *decoder* reconstrói a imagem por meio de interpolações e convoluções transpostas. O grande diferencial da U-Net re-

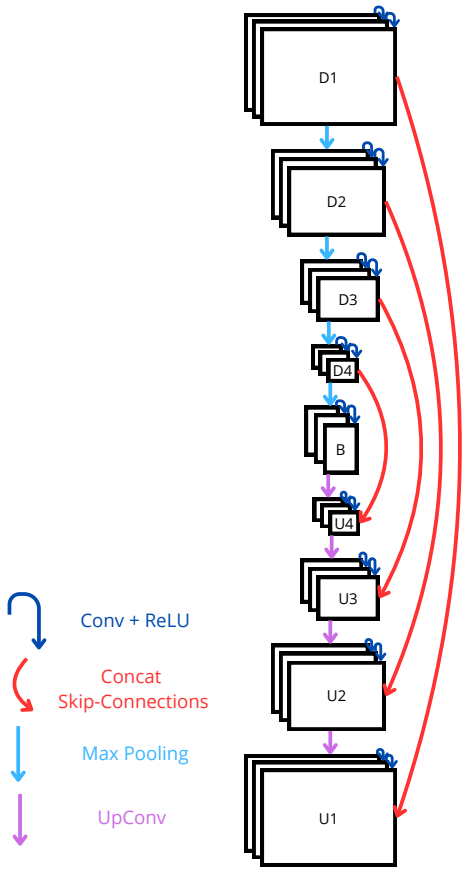


Figura 2: A arquitetura U-Net padrão proposta por Ronneberger, Fischer e Brox (2015).

side nas conexões de salto (*skip connections*), as quais realizam a concatenação direta dos mapas de características de alta resolução do *encoder* com as camadas equivalentes no *decoder*. Esse mecanismo mitiga a perda de detalhes geométricos finos, permitindo uma localização espacial precisa das bordas de estruturas anatômicas.

Nesta dissertação, os mapas de características produzidos ao longo dessas etapas são tratados como representações intermediárias armazenadas para a reutilização entre fatias consecutivas. Assim, em vez de modificar a estrutura básica da U-Net, investiga-se como o compartilhamento dessas representações influencia o comportamento da rede.

2.5.1 Abordagens 2D vs. 3D

Ao lidar com exames naturalmente tridimensionais, como a Tomografia Computadorizada, enfrentamos um dilema entre processamento bidimensional e tridimensional, ambos válidos com vantagens e desvantagens.

Modelos 2D processam cada fatia axial de forma independente, desconsiderando a correlação anatômica entre cortes adjacentes, como pode ser visto na Figura 3. Já as ex-

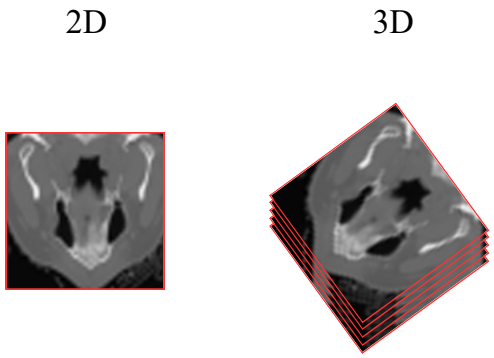


Figura 3: Exemplos de entradas das redes 2D vs 3D.

Tabela 1: Comparação entre abordagens de segmentação 2D e 3D.

Abordagem	2D	3D
Contexto Espacial	Limitado ao plano da fatia. Ignora a continuidade entre fatias.	Explora o contexto volumétrico completo nos três eixos.
Custo Computacional	Baixo consumo de memória (VRAM).	Elevado consumo de memória (VRAM).
Resolução dos Dados	Permite o uso de fatias na resolução original.	Frequentemente exige subamostragem do volume devido ao alto consumo de memória.

tensões tridimensionais, como a 3D U-Net, substituem os kernels $k \times k$ por kernels cúbicos $k \times k \times k$ (Çiçek *et al.*, 2016). Embora consigam capturar com eficiência a continuidade volumétrica, essas redes sofrem com o aumento da dimensionalidade, demandando volumes massivos de memória de vídeo (VRAM), o que restringe o tamanho dos lotes de dados (*batch size*) utilizados no processo iterativo de treinamento da rede ou força a fragmentação do volume em sub-blocos (*patches*), introduzindo inconsistências em suas bordas. A comparação entre as abordagens em pontos relevantes pode ser vista na Tabela 1.

A capacidade de introdução de contexto volumétrico e de custo computacional motiva a investigação de estratégias intermediárias, capazes de introduzir contexto entre fatias mantendo a arquitetura bidimensional. É nesse contexto que se insere a investigação desta dissertação.

2.6 Compartilhamento e Propagação de Características

Em abordagens baseadas em imagens bidimensionais, cada fatia de uma tomografia é processada de maneira independente. Embora essa estratégia simplifique o modelo e re-

duza o custo computacional, ela também limita a integração do contexto espacial entre fatias consecutivas. Como consequência, informações relevantes sobre continuidade anatômica podem ser ignoradas, o que pode afetar a consistência das predições em estruturas volumétricas.

Uma alternativa para incorporar informação contextual consiste em compartilhar características entre diferentes fatias de um mesmo paciente. Nesse cenário, representações aprendidas para uma fatia podem ser reutilizadas ou propagadas para fatias adjacentes, permitindo que a rede tenha acesso a informações complementares durante o processo de segmentação. Esse tipo de estratégia busca aproximar o comportamento de um modelo bidimensional do contexto volumétrico explorado por abordagens tridimensionais, porém sem empregar convoluções 3D completas, que geralmente apresentam maior custo computacional e demanda de memória.

Esse mecanismo é viabilizado, predominantemente, pela continuidade anatômica e pela baixa variação contextual entre as fatias tomográficas, o que permite reutilizar grande parte da informação presente nas imagens anteriores. Contudo, essa abordagem pode propagar erros e ruídos sequencialmente. Além disso, há o risco de incompatibilidade entre as representações compartilhadas e as características das imagens adjacentes, assim como o acúmulo de informações irrelevantes, as quais podem se tornar majoritárias dependendo do conjunto de dados de entrada.

O compartilhamento de características pode ocorrer em diferentes níveis da arquitetura. Ele pode ser aplicado diretamente sobre a entrada da rede, sobre mapas de características intermediários ou mesmo sobre representações mais profundas. Dependendo da profundidade em que ocorre o compartilhamento, a rede passa a integrar informações com diversos níveis de abstração. Camadas mais iniciais tendem a codificar padrões locais, como bordas e texturas, enquanto camadas mais profundas representam estruturas com maior conteúdo semântico.

Além disso, existem diferentes mecanismos para implementar essa propagação de informação. Uma estratégia simples consiste em empilhar os resultados obtidos em uma fatia nas adjacentes, permitindo que a rede aprenda correlações locais entre cortes consecutivos. Outra possibilidade envolve o uso de conexões explícitas entre as ativações de diferentes fatias, como mecanismos de concatenação (Qiu *et al.*, 2021) ou atenção espacial (Oktay *et al.*, 2018).

Abordagens mais sofisticadas incluem o uso de redes recorrentes, como ConvLSTM (Azad *et al.*, 2019), que modelam explicitamente a dependência sequencial entre fatias

adjacentes por meio da propagação de informações ao longo da sequência. Outros trabalhos também exploram mecanismos de atenção para incorporar informações contextuais de forma adaptativa. Embora essas abordagens sejam capazes de capturar dependências de longo alcance, elas tendem a apresentar maior custo computacional e, em alguns casos, maior demanda por dados para uma boa generalização.

A ideia de propagação de informação entre fatias consecutivas também pode ser interpretada como uma forma de introduzir dependência estrutural entre cortes adjacentes, o que é especialmente relevante em estruturas anatômicas com continuidade volumétrica. Em muitos órgãos e lesões, a aparência em uma única fatia pode ser ambígua, enquanto a observação de múltiplos cortes consecutivos reduz incertezas e melhora a confiabilidade das predições.

2.7 Anatomia e Segmentação da Mandíbula

A mandíbula é o maior e mais forte osso da face, desempenhando papel crucial na mastigação e no suporte estético facial. Na odontologia, sua segmentação precisa em exames médicos de imagem é requisito fundamental para planejamento de implantes e tratamentos de deformidades. Um exemplo de ilustração anatômica da mandíbula pode ser visto na Figura 4.

Do ponto de vista morfológico, a mandíbula é dividida em duas regiões principais, com densidades e texturas bem distintas (Carmo; Chaves, 2023):

- **Corpo Mandibular:** Região horizontal que se caracteriza por uma espessa camada de osso cortical homogêneo, o que geralmente se traduz em alto contraste e contornos bem definidos nas imagens de tomografia.
- **Ramo Mandibular e Côndilo:** O ramo mandibular corresponde à porção vertical da mandíbula e apresenta uma geometria mais complexa em comparação ao corpo. Nessa região, o osso tende a ser mais fino e com variações estruturais mais evidentes, apresentando transições de intensidade mais suaves. O côndilo, em particular, é uma sub-região do ramo mandibular de difícil delineamento, devido ao seu pequeno tamanho nas imagens tomográficas em comparação com as estruturas adjacentes.

Somado à complexidade geométrica, o ambiente bucal introduz diversos ruídos de imagem. A presença de restaurações metálicas, aparelhos ortodônticos fixos e implantes gera intensidades muito diferentes das encontradas normalmente na região.

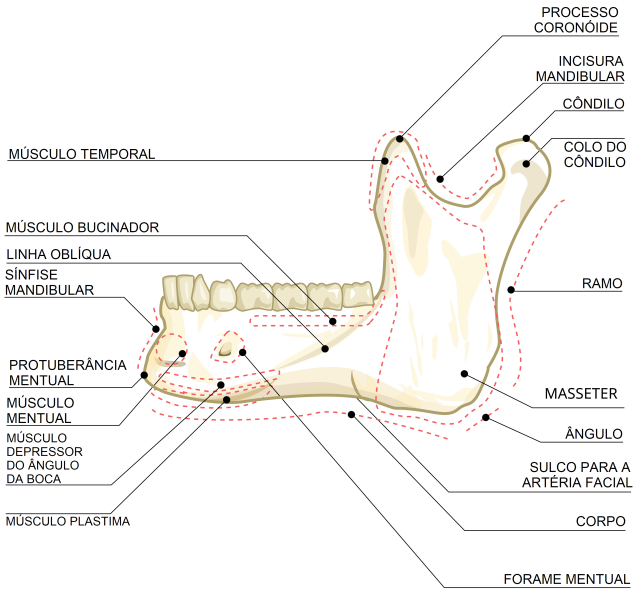


Figura 4: Anatomia da mandíbula humana. Adaptado de: Hariadhi (2020).

2.8 Síntese do Capítulo

Este capítulo estabeleceu as premissas teóricas para a sustentação deste trabalho. Foram discutidos os fundamentos da segmentação médica, a evolução das Redes Neurais Convolucionais, o funcionamento da arquitetura U-Net juntamente com uma análise das abordagens bidimensionais e tridimensionais, e por fim, foram delimitadas as propriedades anatômicas e os desafios de imagem na região da mandíbula, fornecendo o embasamento necessário para a introdução das técnicas de compartilhamento de características inter-fatias. Esta base conceitual fundamenta diretamente a metodologia investigada a ser apresentada no Capítulo 4.

3 Revisão da Literatura

Como o objetivo desta dissertação é investigar o compartilhamento de mapas de características entre fatias consecutivas em arquiteturas U-Net bidimensionais, esta revisão enfatiza trabalhos que exploram mecanismos de propagação de informações, recorrência e reutilização de representações internas, além de estudos específicos sobre segmentação mandibular.

Embora existam inúmeras variantes da U-Net voltadas ao aumento da acurácia da segmentação, esta revisão concentra-se principalmente em métodos que incorporam contexto entre imagens. Como discutido anteriormente, tais métodos podem ser implementados em arquiteturas 2D ou 3D. Neste trabalho, o foco é em abordagens bidimensionais devido ao menor custo computacional e à possibilidade de investigar estratégias de incorporação de contexto sem recorrer a convoluções tridimensionais.

3.1 Aprimoramento de Representações Internas

Após a introdução da U-Net, pesquisadores buscaram melhorar a capacidade de representação da U-Net adicionando mecanismos de atenção; Oktay *et al.* (2018) propuseram o mecanismo de *Attention Gate*, capaz de aprender a destacar regiões relevantes e suprimir áreas irrelevantes da imagem durante a segmentação, melhorando a qualidade dos resultados sem aumento significativo do custo computacional. Os autores alcançaram um Índice de Similaridade Dice de 84% no conjunto de dados de segmentação multiclasse abdominal CT-150 (Roth *et al.*, 2016), superando o desempenho da U-Net convencional. Diferentemente do presente trabalho, que investiga a propagação de características entre fatias adjacentes de tomografias computadorizadas, este estudo concentra-se no aprimoramento da própria fatia por meio de mecanismos de atenção espacial.

Inspirados por redes convolucionais recorrentes e redes residuais profundas, Alom *et al.* (2018) introduziram recorrência nos blocos internos da U-Net, reutilizando mapas de características gerados em cada camada da rede. A arquitetura proposta, denominada *Re-*

current Residual Convolutional Neural Network based on U-Net (R2U-Net), permite uma melhor compreensão do contexto espacial em estruturas complexas, mantendo eficiência de memória por reutilizar os mesmos pesos ao longo da propagação direta. No conjunto de dados de lesão de pele *International Skin Imaging Collaboration* (ISIC) (Codella *et al.*, 2018), o método atingiu IoU médio de 94%. Vale ressaltar que, embora reutilize características, essa recorrência ocorre apenas dentro da própria camada, diferentemente deste trabalho, que busca reutilizar características provenientes de fatias vizinhas.

Com a popularização dos Transformers na área de visão computacional, Xie *et al.* (2021) propuseram o *SegFormer*, uma arquitetura para segmentação semântica que combina um *encoder* Transformer hierárquico com um *decoder* baseado em *MultiLayer Perceptron* (MLP). Diferentemente de abordagens convolucionais tradicionais, o *encoder* produz representações em múltiplas escalas sem utilizar codificação posicional, enquanto o *decoder* integra essas representações de forma leve e eficiente. Os autores demonstraram desempenho competitivo em diversos conjuntos de dados de segmentação, superando arquiteturas anteriores com menor número de parâmetros e menor custo computacional. Diferentemente deste trabalho, que investiga o compartilhamento de mapas de características entre fatias consecutivas mantendo a arquitetura U-Net, o *SegFormer* propõe uma nova arquitetura baseada em Transformers para melhorar a representação das imagens durante a segmentação.

3.2 Incorporação de Contexto

A forte dependência espacial entre imagens consecutivas motivou, logo em seguida, o desenvolvimento de arquiteturas com recorrência temporal. Wei Wang *et al.* (2019) combinaram redes neurais recorrentes com a U-Net, compartilhando segmentações e características convolucionais entre quadros consecutivos de vídeo. Por meio de unidades *Single-Gated-Unit* (SGU), o modelo passou a codificar continuidade e contexto espacial em dados 2D. O método foi validado no conjunto de dados *Digital Retinal Images for Vessel Extraction* (DRIVE) (Staal *et al.*, 2004), que faz segmentação de vasos sanguíneos na retina, obtendo desempenho superior ao da U-Net convencional.

Seguindo essa direção, Azad *et al.* (2019) propuseram a *Bi-Directional Convolutional Long Short Term Memory U-Net* (BCDU-Net), uma extensão da U-Net que substitui as conexões de salto tradicionais por camadas *Bidirectional ConvLSTM* e incorpora convoluções densamente conectadas. A arquitetura favorece a reutilização de características e

alcança 93% de similaridade de Jaccard no conjunto ISIC. Diferentemente da R2U-Net, cuja recorrência ocorre dentro dos blocos convolucionais, a BCDU-Net utiliza ConvLSTM para modelar dependências entre diferentes níveis da arquitetura, aumentando a capacidade de propagação de contexto, porém, ao custo de maior complexidade computacional. Enquanto a BCDU-Net emprega mecanismos recorrentes complexos para modelar dependências espaciais, esta dissertação investiga uma hipótese mais simples: se o compartilhamento direto de mapas de características, sem módulos recorrentes ou mecanismos de atenção, seria suficiente para incorporar contexto entre fatias consecutivas.

Recentemente, os autores de Li *et al.* (2026) propuseram a arquitetura *Plug-and-Play Neighborhood Explorer* (PnP-NE), cujo objetivo é incorporar informações de fatias vizinhas preservando as vantagens computacionais das redes bidimensionais. O método utiliza um módulo de fusão de informações entre 3 fatias consecutivas, permitindo que características provenientes da vizinhança contribuam para a segmentação da fatia central de forma adaptativa. Os experimentos demonstraram melhorias em diferentes conjuntos de dados de segmentação médica quando comparados a arquiteturas 2D convencionais. Diferentemente do presente trabalho, que investiga o compartilhamento direto de mapas de características entre fatias consecutivas, essa abordagem emprega um mecanismo específico de fusão para integrar as informações da vizinhança antes da predição.

3.3 Otimização de Pipeline

Paralelamente às propostas que modificam a arquitetura da U-Net, outra linha de pesquisa concentrou-se na automatização do processo de treinamento. Um dos trabalhos de maior impacto na área é o *no-new-Net U-Net* (nnU-Net), desenvolvido por Isensee *et al.* (2021). Trata-se de um *framework* autoconfigurável baseado na U-Net, capaz de adaptar automaticamente estratégias de pré-processamento, treinamento e inferência às características do conjunto de dados utilizado. Essa abordagem reduz significativamente a necessidade de ajuste manual de hiperparâmetros e permite aplicação em múltiplas modalidades de imagem médica, incluindo tomografia computadorizada, ressonância magnética e raios X. O nnU-Net apresentou desempenho elevado em diversas tarefas de segmentação, como tumores cerebrais pediátricos e espaços perivasculares, ambos com Índice de Similaridade Dice de 90% (Vossough *et al.*, 2024; Pham *et al.*, 2026), além de segmentação craniofacial com Índice de Similaridade Dice de 97% para a mandíbula (Dot *et al.*, 2022).

3.4 Estudos Específicos para Segmentação da Mandíbula

Especificamente para segmentação mandibular, Qiu *et al.* (2021) utilizaram uma combinação de redes convolucionais e recorrentes, permitindo que diferentes fatias de uma mesma TC compartilhassem informações relevantes para a segmentação. Como os dados de entrada eram exames tridimensionais, a avaliação foi realizada sobre o volume 3D da mandíbula. Em um conjunto proprietário de TC, o método alcançou um índice de similaridade Dice de 0,97.

Pankert *et al.* (2023) propuseram um processo em duas etapas utilizando múltiplas 3D U-Nets em cascata. A primeira rede gera uma segmentação de baixa resolução para definir uma caixa delimitadora (*bounding box*), enquanto a segunda refina a segmentação na região de interesse detectada. Em um conjunto proprietário de TC, o método atingiu um índice de similaridade Dice de aproximadamente 0,95.

Em um estudo comparativo recente, Irshad *et al.* (2024) avaliaram diferentes softwares de segmentação mandibular, incluindo abordagens totalmente automáticas e métodos assistidos manualmente. Entre as ferramentas analisadas destaca-se o Relu (Ileşan *et al.*, 2023), que utiliza uma arquitetura 3D U-Net para segmentar a mandíbula a partir de arquivos de *Digital Imaging and Communications in Medicine* (DICOM). O software alcançou um índice de similaridade dice médio de aproximadamente 0,92, demonstrando desempenho competitivo entre as ferramentas avaliadas. Os autores discutem ainda aspectos relacionados ao desempenho e à aplicabilidade clínica das diferentes soluções, evidenciando que a escolha da ferramenta depende não apenas da acurácia da segmentação, mas também de fatores como fluxo de trabalho e necessidade de intervenção do usuário.

3.5 Síntese do Capítulo

A comparação dos principais pontos e limitações dos trabalhos citados neste capítulo pode ser vista na Tabela 2.

De forma geral, os trabalhos anteriores exploraram mecanismos de atenção, recorrência, integração de ConvLSTM e arquiteturas tridimensionais para melhorar o desempenho da segmentação. Contudo, a maioria dessas abordagens prioriza o aumento da complexidade arquitetural, enquanto relativamente pouca atenção tem sido dedicada à investigação sistemática da propagação de características latentes entre fatias adjacentes em cenários de segmentação 2D com baixo custo computacional. A existência de poucos estudos

nessa direção motivou o presente trabalho, que busca analisar como o compartilhamento de mapas de características produzidos entre fatias consecutivas pode contribuir para a segmentação mandibular mantendo baixo custo computacional.

Tabela 2: Comparação entre trabalhos relacionados sob a perspectiva da incorporação de contexto entre imagens.

Trabalho	**Pontos fortes**	**Limitações**
Attention U-Net (Oktay *et al.*, 2018)	Melhora a representação da própria imagem por meio de mecanismos de atenção espacial.	O compartilhamento é limitado a informações contidas na mesma imagem.
R2U-Net (Alom *et al.*, 2018)	Reutiliza características internas da rede sem aumento expressivo do número de parâmetros.	A recorrência ocorre apenas dentro da própria imagem, não explorando continuidade entre imagens.
SegFormer (Xie *et al.*, 2021)	Melhora a representação das imagens por meio de uma arquitetura baseada em Transformers, capturando dependências de longo alcance com alta eficiência.	Propõe uma nova arquitetura para segmentação, sem investigar o compartilhamento de mapas de características ou a incorporação de contexto entre fatias consecutivas.
Recurrent U-Net (Wang, W. *et al.*, 2019)	Modela dependências entre imagens consecutivas utilizando unidades recorrentes.	Depende de mecanismos recorrentes específicos, dificultando a análise isolada do efeito do compartilhamento de mapas de características.
BCDU-Net (Azad *et al.*, 2019)	Captura contexto entre representações por meio de ConvLSTM bidirecional.	Introduz maior complexidade arquitetural e custo computacional, tornando mais difícil avaliar o impacto isolado do compartilhamento de características.

Trabalho	**Pontos fortes**	**Limitações**
PnP-NE (Li *et al.*, 2026)	Incorpora contexto entre fatias vizinhas mantendo uma arquitetura 2D por meio de um mecanismo adaptativo de fusão de informações.	Não investiga de forma sistemática o impacto do compartilhamento de mapas de características em diferentes níveis da arquitetura.
nnU-Net (Isensee *et al.*, 2021)	Obtém elevado desempenho por adaptar automaticamente o pipeline ao conjunto de dados.	Seu foco está na otimização da configuração do treinamento da arquitetura original da U-Net, e não na investigação de mecanismos de compartilhamento de características.
Qiu *et al.* (2021)	Utiliza módulos de recorrência entre fatias para segmentação mandibular.	Emprega uma arquitetura recorrente tridimensional; além do custo computacional, não avalia o efeito do compartilhamento direto de mapas de características entre fatias.
Pankert *et al.* (2023)	Alcança elevada precisão utilizando U-Nets 3D em cascata.	Alto custo computacional e não há uma investigação de estratégias de compartilhamento de representações entre fatias em modelos bidimensionais.

4 Metodologia

Neste capítulo são apresentados os fundamentos metodológicos deste trabalho, incluindo a ideia central da investigação, a descrição da base de dados utilizada, as modificações realizadas na arquitetura da U-Net e os procedimentos experimentais empregados para a avaliação do método.

4.1 Ideia Central

A mandíbula apresenta variações anatômicas significativas ao longo de sua extensão, o que torna sua segmentação um problema desafiador. Na região correspondente ao corpo mandibular, a estrutura possui formato característico em "U" e ocupa uma parcela considerável do corte axial. Nessa região, poucos tecidos ósseos apresentam intensidades semelhantes na tomografia computadorizada, permitindo que arquiteturas bidimensionais convencionais, como a U-Net original, realizem a segmentação com elevada precisão, como pode ser observado na Figura 5, onde temos na extrema esquerda, a imagem de referência de um modelo 3D com porção significativa do corpo circulado em vermelho, e as três últimas imagens representando respectivamente, a tomografia original, a segmentação dos especialistas, ambos pertencentes a base de dados PDDCA (Raudaschl *et al.*, 2017) e o por fim resultado gerado pela U-Net.

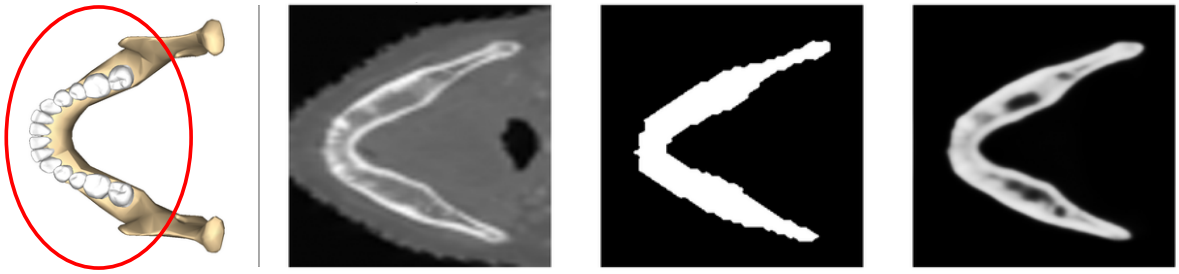


Figura 5: Exemplo de tomografia contendo o corpo da mandíbula. Adaptado de: Hariadhi (2020).

Entretanto, nas regiões do ramo mandibular e principalmente do côndilo, a tarefa

torna-se significativamente mais complexa. Nessas regiões, a mandíbula apresenta intensidades muito semelhantes às de estruturas vizinhas, como o crânio e o maxilar, dificultando sua distinção apenas com base nas informações presentes em um único corte axial. Além disso, o côndilo está presente em um número menor de fatias quando comparado às outras regiões da mandíbula, o que reduz a quantidade de exemplos disponíveis durante o treinamento e dificulta o aprendizado da rede. Exemplos podem ser vistos nas Figuras 6 e 7. Na extrema esquerda da Figura 6 temos a imagem de referência de um modelo 3D com porção significativa do ramo circulado em vermelho, já na Figura 7 igualmente na extrema esquerda, temos uma imagem de referência de um modelo 3D com o côndilo circulado em vermelho e o processo coronoide em rosa. Em ambas as figuras temos as três últimas imagens representando, respectivamente, a tomografia original, a segmentação dos especialistas, ambas pertencentes à base de dados PDDCA e, por fim, o resultado gerado pela U-Net;

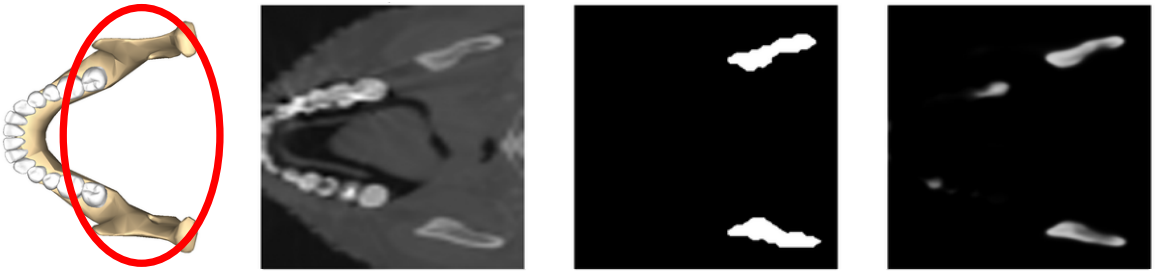


Figura 6: Exemplo de tomografia contendo o ramo da mandíbula. Adaptado de: Hariadhi (2020).

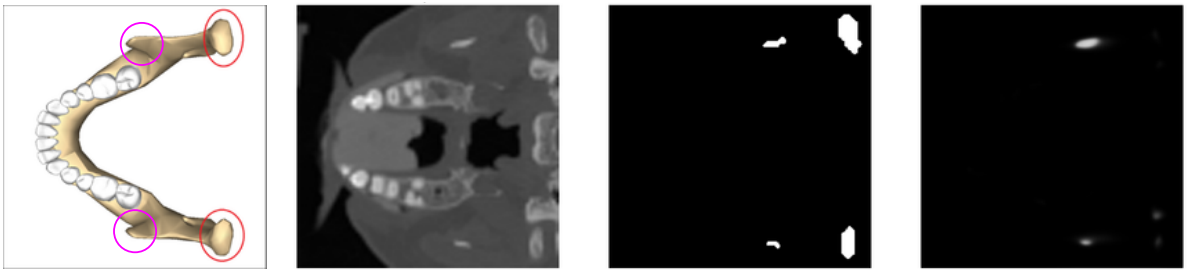


Figura 7: Exemplo de tomografia contendo o côndilo e o processo coronoide da mandíbula. Adaptado de: Hariadhi (2020).

Em situações como essa, informações provenientes de cortes adjacentes são importantes para a correta identificação das estruturas anatômicas. No entanto, arquiteturas bidimensionais processam cada imagem de forma independente, desconsiderando as relações espaciais entre fatias consecutivas. Essa limitação reduz sua capacidade de incorporar o contexto anatômico, o que tende a comprometer o desempenho justamente nas regiões em que esse contexto é mais relevante.

Uma alternativa consiste na utilização de arquiteturas tridimensionais, capazes de processar simultaneamente todo o volume da tomografia e explorar diretamente as relações espaciais entre voxels. Embora essas abordagens apresentem resultados de estado da arte, elas demandam maior capacidade computacional, maior quantidade de memória e tempos de treinamento significativamente superiores. Esses fatores limitam sua utilização em ambientes com recursos computacionais restritos.

Dessa forma, este trabalho investiga uma abordagem intermediária que busca incorporar informações contextuais entre fatias consecutivas, preservando a simplicidade computacional das arquiteturas bidimensionais.

4.2 Arquiteturas Investigadas

A principal estratégia investigada nesta dissertação em relação à U-Net original proposta por Ronneberger, Fischer e Brox (2015) consiste no compartilhamento de representações intermediárias entre fatias axiais consecutivas. Em vez de processar cada imagem de forma completamente independente, as representações geradas durante o processamento da fatia anterior são armazenadas temporariamente e reutilizadas durante o processamento da fatia corrente.

Para isso, mapas de características provenientes de camadas correspondentes são concatenados ao longo da dimensão dos canais antes da aplicação da convolução. Após o processamento de cada fatia, as representações armazenadas são substituídas pelas representações da fatia recém-processada, previamente desacopladas do grafo computacional, evitando a propagação do gradiente entre diferentes fatias. Todo esse processo é realizado sequencialmente ao longo da direção axial da tomografia, sendo os *buffers* reinicializados sempre que um novo paciente é processado.

Matematicamente, seja $F_l^{(t)}$ o mapa de características produzido pela camada l para a fatia t. O cálculo da representação pode ser expresso por

$$F_l^{(t)} = \sigma\left(W_l * [F_{l-1}^{(t)}, F_{l-1}^{(t-1)}] + b_l\right) \tag{4.1}$$

em que $[\cdot,\cdot]$ representa a concatenação ao longo da dimensão dos canais, W_l corresponde aos filtros convolucionais da camada, b_l representa o vetor de bias e σ denota a função de ativação. Para a primeira fatia de cada exame, como não existe uma fatia anterior disponível, a própria é concatenada consigo mesma.

Foram investigadas duas estratégias distintas de compartilhamento de informações. Ambas contêm diversos conjuntos de experimentos em relação às estratégias adotadas. Para cada modificação da rede foram conduzidos novos treinamentos completos e testes da rede.

Na primeira, a qual denominamos *forward*, o compartilhamento foi realizado progressivamente da entrada em direção à saída da rede. Inicialmente, o primeiro experimento consistiu apenas na concatenação da imagem da fatia anterior com a fatia corrente. Em seguida, no próximo experimento, passaram a ser compartilhados também os mapas de características produzidos pelo primeiro bloco convolucional. Esse procedimento foi repetido sucessivamente, produzindo novos experimentos cumulativos até a última camada da rede, aumentando progressivamente a quantidade de contexto espacial fornecida à rede.

Na segunda estratégia, a qual denominamos *reverse*, o compartilhamento foi realizado da mesma forma que no *forward*, mas de forma reversa, da saída da rede em direção à entrada. Inicialmente, apenas o último mapa de características produzido da fatia anterior é reutilizado. Em seguida, passam a ser compartilhados os mapas produzidos anteriormente pelos blocos do *decoder*, avançando progressivamente em direção ao gargalo. Esse segundo conjunto de experimentos foi conduzido principalmente com o objetivo de observar o impacto do compartilhamento de características no processo de formação da imagem da rede. Neste trabalho não foram implementadas as demais configurações até a camada de entrada, devido aos retornos decrescentes observados nos experimentos realizados.

4.2.1 Configurações das Arquiteturas

Como o objetivo deste trabalho é analisar o efeito do compartilhamento progressivo de mapas de características, foram desenvolvidas 14 configurações distintas da arquitetura investigada, cada uma representando um nível diferente de compartilhamento de informações entre fatias consecutivas. As configurações variaram desde o compartilhamento exclusivo da imagem da fatia anterior até o compartilhamento de todos os mapas intermediários de características.

As configurações dos experimentos *forward* são descritas na Tabela 3.

Tabela 3: Descrição dos experimentos de compartilhamento contidos nas Figuras 8, 9 e 10.

Subfigura	**Mudanças na Arquitetura**
4.8(a)	Compartilhamento executado somente entre as fatias tomográficas consecutivas.
4.8(b)	Compartilhamento executado desde as fatias tomográficas consecutivas até o mapa de características do *Encoder 1*.
4.8(c)	Compartilhamento executado desde as fatias tomográficas consecutivas até o mapa de características do *Encoder 2*.
4.9(a)	Compartilhamento executado desde as fatias tomográficas consecutivas até o mapa de características do *Encoder 3*.
4.9(b)	Compartilhamento executado desde as fatias tomográficas consecutivas até o mapa de características do *Bottleneck*.
4.9(c)	Compartilhamento executado desde as fatias tomográficas consecutivas até o mapa de características do *Decoder 1*.
4.10(a)	Compartilhamento executado desde as fatias tomográficas consecutivas até o mapa de características do *Decoder 2*.
4.10(b)	Compartilhamento executado desde as fatias tomográficas consecutivas até o mapa de características do *Decoder 3*.
4.10(c)	Compartilhamento executado desde as fatias tomográficas consecutivas até o mapa de características do *Decoder 4*.

As configurações dos experimentos *reverse* são descritas na Tabela 4.

Tabela 4: Descrição das etapas de compartilhamento apresentadas nas subfiguras das Figuras 8, 9 e 10.

Subfigura	**Mudanças na Arquitetura**
4.11(a)	Compartilhamento é executado somente no último mapa de características produzido pelo *Decoder 4*.
4.11(b)	Compartilhamento é executado desde o mapa de características do *Decoder 4* até o produzido pelo *Decoder 3*.
4.11(c)	Compartilhamento é executado desde o mapa de características do *Decoder 4* até o produzido pelo *Decoder 2*.

Subfigura	Mudanças na Arquitetura
4.12(a)	Compartilhamento é executado desde o mapa de características do *Decoder 4* até o produzido pelo *Decoder 1*.
4.12(b)	Compartilhamento é executado desde o mapa de características do *Decoder 4* até o produzido pelo *Bottleneck*.

4.3 Procedimentos Experimentais

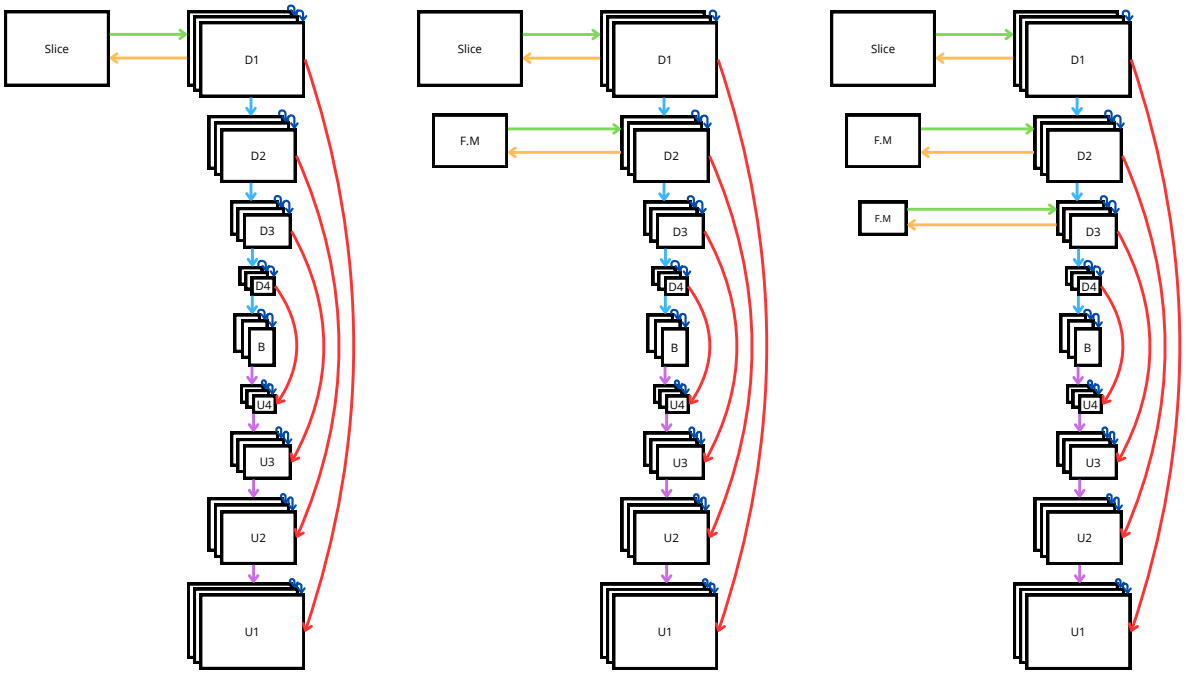
Todos os modelos foram implementados utilizando PyTorch 2.11, treinados em uma GPU NVIDIA RTX PRO 6000 Blackwell Server Edition com 95 gigas de memória, equipada com o driver NVIDIA versão 580.82.07, CUDA 13.0 e a biblioteca de aceleração de redes neurais CUDNN 91900.

Todas as arquiteturas foram treinadas utilizando a mesma configuração experimental. O treinamento foi realizado durante 30 épocas, empregando (*batch size*) igual a um, em que cada lote corresponde ao exame completo de um paciente. Como diferentes exames possuem números distintos de fatias, cada iteração processa sequencialmente todas as imagens pertencentes ao mesmo volume tomográfico.

Foi utilizada a implementação do Pytorch (Paszke *et al.*, 2019) da função de perda, `torch.nn.BCEWithLogitsLoss`, que incorpora internamente a função sigmoide antes do cálculo da BCE, proporcionando maior estabilidade numérica. A taxa inicial de aprendizado é igual a 0,01 e há um mecanismo de redução automática da taxa de aprendizado (*Learning Rate Scheduler*) configurado com fator de redução de 0,5 e paciência de cinco épocas. As operações de convolução, ReLU e *Max Pooling* também foram implementadas utilizando Pytorch, respectivamente integrando os módulos `torch.nn.Conv2d`, `torch.nn.ReLU` e `torch.nn.MaxPool2d`.

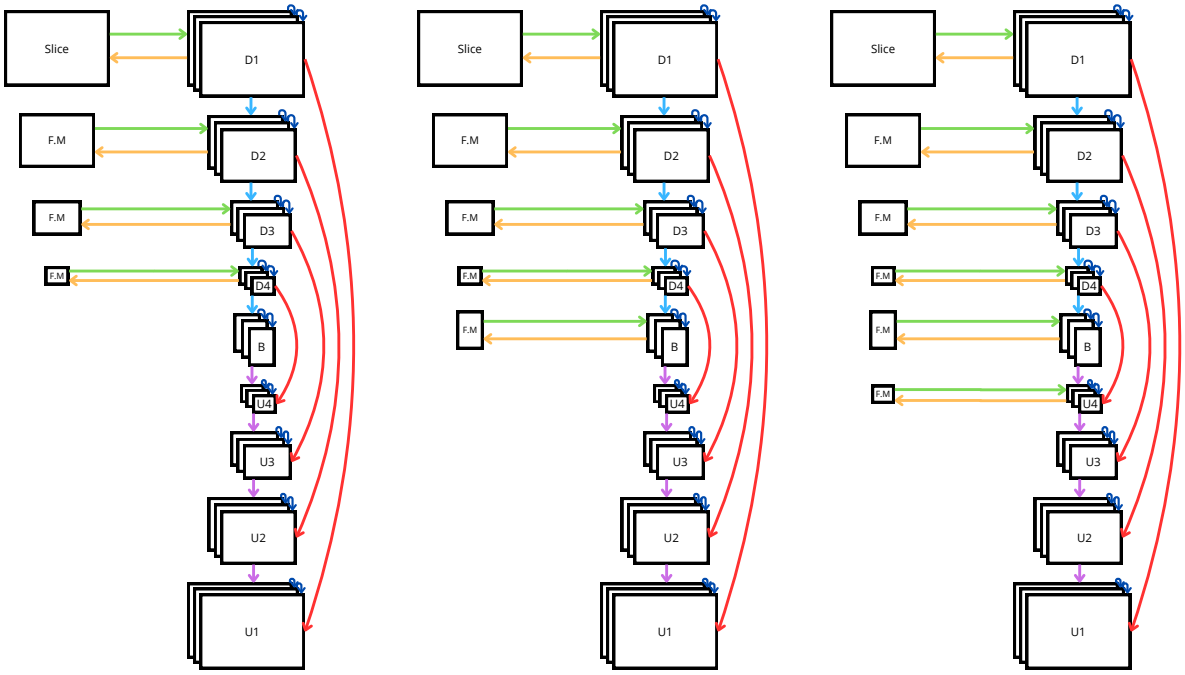
A avaliação quantitativa foi realizada utilizando as métricas IoU, Dice e 95HD, todas amplamente empregadas na literatura para avaliação de métodos de segmentação médica.

Para verificar se as diferenças observadas entre as configurações investigadas e a U-Net convencional eram estatisticamente significativas, foi aplicado o teste não paramétrico de Wilcoxon para amostras pareadas. Cada configuração foi treinada seis vezes utilizando diferentes sementes aleatórias. A hipótese nula (H_0) estabelece que a mediana das diferenças entre as estratégias investigadas e a arquitetura de referência é igual a zero, indicando ausência de diferença estatisticamente significativa. Adotou-se nível de significância de



(a) *Slice.* (b) *Encoder 1.* (c) *Encoder 2.*

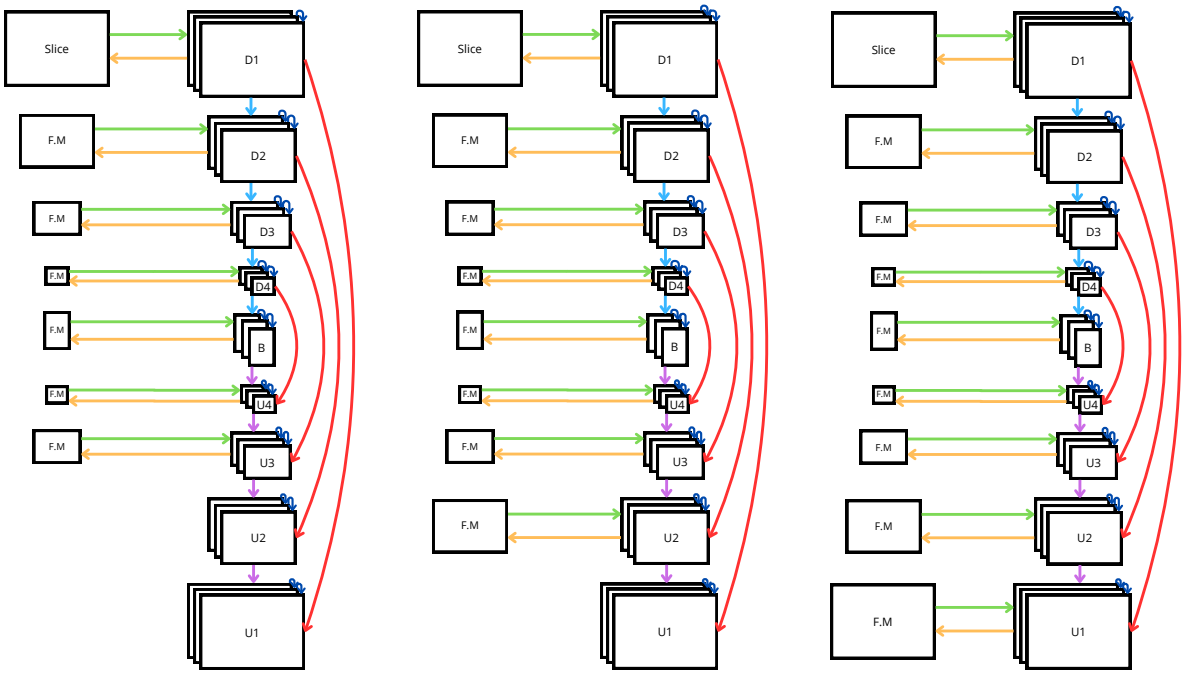
Figura 8: Arquiteturas *forward*: *Slice*, *Encoder 1* e *Encoder 2*.



(a) *Encoder 3.* (b) *Bottleneck.* (c) *Decoder 1.*

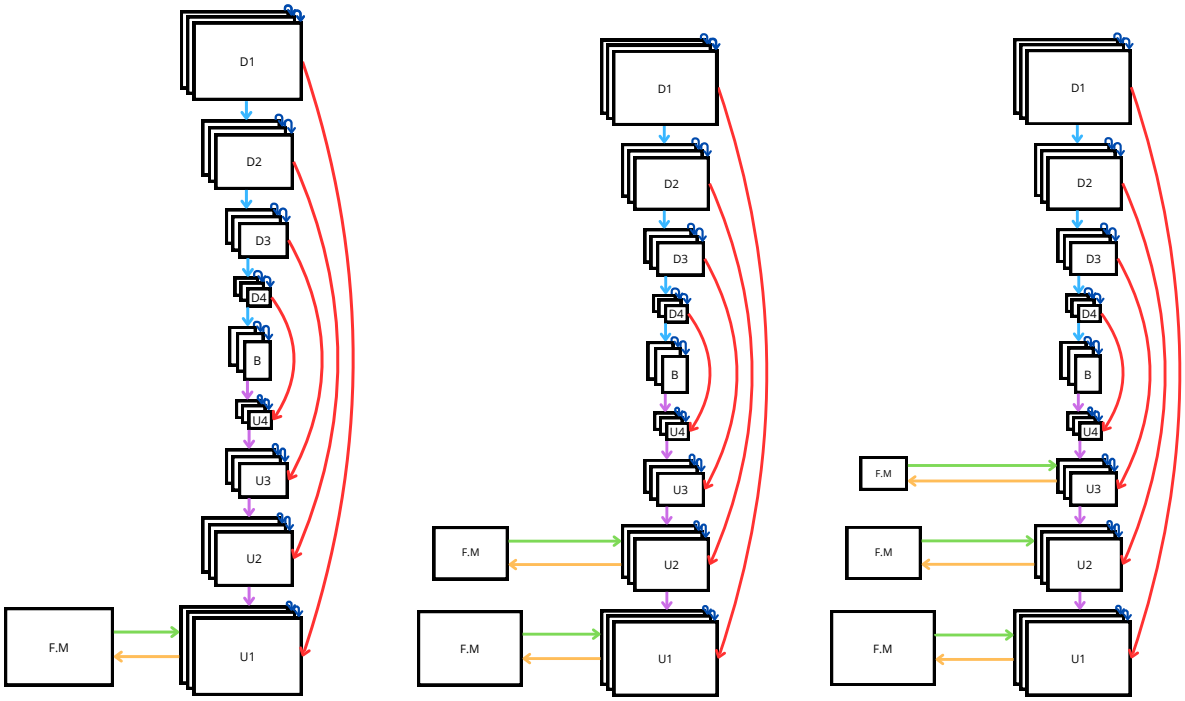
Figura 9: Arquiteturas *forward*: *Encoder 3*, *Bottleneck* e *Decoder 1*.

$\alpha = 0{,}05$. Nos casos em que a hipótese nula foi rejeitada, a direção da diferença foi determinada pela comparação das medianas das métricas obtidas, permitindo identificar se a modificação avaliada produziu melhora ou degradação no desempenho em relação à



(a) *Decoder 2.* (b) *Decoder 3.* (c) *Decoder 4.*

Figura 10: Arquiteturas *forward*: *Decoder 2*, *Decoder 3* e *Decoder 4*.

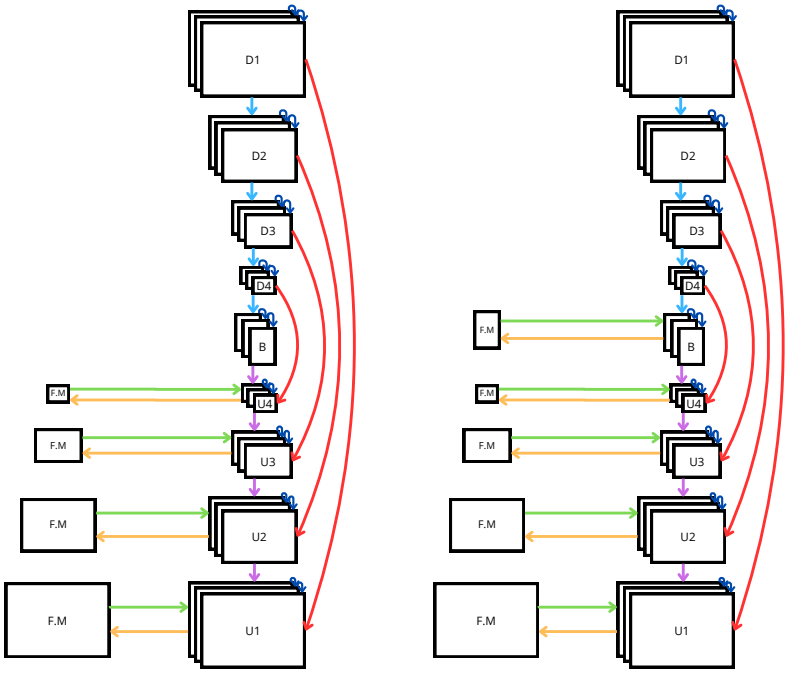


(a) *Decoder 4.* (b) *Decoder 3.* (c) *Decoder 2.*

Figura 11: Arquiteturas *reverse*: *Decoder 4*, *Decoder 3* e *Decoder 2*.

U-Net convencional.

Posteriormente, foram aplicados mapas de atenção às redes por meio da técnica *Gradient-weighted Class Activation Mapping* (Grad-CAM) (Selvaraju *et al.*, 2017), com



(a) *Decoder 1.* (b) *Bottleneck.*

Figura 12: Arquiteturas *reverse*: *Decoder 1* e *Bottleneck*.

o objetivo de identificar as regiões da imagem que mais influenciaram as decisões do modelo. Essa análise permite interpretar onde a rede concentra sua atenção e compreender os fatores que levaram à segmentação de determinadas regiões.

4.4 Base de Dados

Os experimentos foram realizados utilizando a base de dados PDDCA, disponibilizada originalmente para o desafio *Medical Image Computing and Computer Assisted Intervention* (MICCAI) *Head and Neck Challenge* de 2015. O conjunto contém 48 exames de tomografia computadorizada da região de cabeça e pescoço com estruturas anatômicas diferentes segmentadas manualmente. Desses exames, apenas 40 possuem segmentações manuais da mandíbula, sendo estes os utilizados neste trabalho.

Durante o pré-processamento, inicialmente foram selecionados apenas os exames contendo segmentações da mandíbula. Em seguida, foram mantidas somente as fatias axiais nas quais existia alguma porção dessa estrutura anatômica.

Posteriormente, foi realizada uma etapa de recorte manual de uma Região de Interesse (ROI), definida por meio de caixas delimitadoras centralizadas na mandíbula, com o objetivo de eliminar regiões irrelevantes da imagem e concentrar o processamento na

estrutura de interesse. Após essa etapa, todas as imagens foram redimensionadas para resolução de 256×256 pixels, correspondente ao tamanho de entrada da arquitetura U-Net utilizada. Todas as métricas foram calculadas sobre as predições do modelo na resolução de saída, que coincide com a resolução dos dados de entrada redimensionados.

A divisão entre os conjuntos de treinamento, validação e teste foi realizada no nível dos pacientes, evitando que fatias diferentes de um mesmo exame estivessem presentes em conjuntos distintos. Foram destinados 60% dos exames para treinamento, 20% para validação e 20% para teste. Não foram aplicadas técnicas adicionais de normalização ou limitação dos valores de intensidade.

4.5 Síntese do Capítulo

Este capítulo apresentou a metodologia utilizada para investigar o compartilhamento progressivo de mapas de características entre fatias consecutivas em arquiteturas U-Net bidimensionais, utilizando a segmentação mandibular como estudo de caso. Foram descritas as duas estratégias de compartilhamento investigadas, os procedimentos de treinamento e avaliação estatística, bem como a base de dados PDDCA e as etapas de pré-processamento adotadas. No capítulo seguinte são apresentados os resultados experimentais obtidos pelas configurações diferentes adotadas e sua comparação com a U-Net convencional.

5 Resultados

Neste capítulo, apresentamos os resultados quantitativos e qualitativos das estratégias de compartilhamento de características investigadas em comparação com a arquitetura U-Net de referência. Avaliamos os esquemas de propagação nas já mencionadas abordagens *forward* e *reverse* utilizando as métricas IoU, Dice e 95HD.

5.1 Resultados Quantitativos

Nesta seção são apresentados os resultados quantitativos obtidos pelas configurações investigadas, comparando-as com a arquitetura U-Net de referência. As Tabelas 5 e 7 apresentam os valores médios e desvios padrão das métricas 95HD, Dice e IoU obtidos ao longo das seis execuções independentes de cada configuração. Complementarmente, as Tabelas 6 e 8 apresentam os valores de p obtidos pelo teste não paramétrico de Wilcoxon, permitindo avaliar se as diferenças observadas em relação ao modelo de referência foram estatisticamente significativas.

5.1.1 Experimentos *Forward*

A Tabela 5 resume os experimentos do *forward*, nos quais mapas de características progressivamente mais profundos foram compartilhados entre fatias, da entrada em direção à saída do modelo. Em negrito, temos os melhores resultados obtidos dentre todos os experimentos em determinada métrica. Nenhuma das configurações investigadas superou o desempenho da U-Net de referência. Observou-se um declínio geral nos índices IoU e Dice à medida que o compartilhamento de características ocorria em camadas mais profundas. As configurações de *Slice* a *Encoder 3* mantiveram níveis de desempenho comparáveis aos do modelo base; no entanto, como mostra a Tabela 6, onde em negrito temos os experimentos que alcançaram um *p-value* menor que 0.05, o teste de Wilcoxon falhou em rejeitar a hipótese nula, indicando que essas diferenças não foram estatisticamente signifi-

Tabela 5: Resultados quantitativos dos experimentos *forward*.

Métricas	**95HD**	**Dice**	**IoU**
U-Net	24,13 ± 2,88	**0,85** ± 0,01	**0,70** ± 0,04
Slice	23,80 ± 3,03	0,83 ± 0,04	0,68 ± 0,04
Encoder 1	22,83 ± 3,35	0,84 ± 0,01	0,68 ± 0,04
Encoder 2	**22,12** ± 2,33	0,84 ± 0,01	0,69 ± 0,03
Encoder 3	25,73 ± 10,17	0,80 ± 0,07	0,62 ± 0,13
Bottleneck	30,04 ± 8,32	0,78 ± 0,05	0,56 ± 0,10
Decoder 1	30,51 ± 7,89	0,81 ± 0,02	0,59 ± 0,05
Decoder 2	40,39 ± 13,08	0,73 ± 0,08	0,44 ± 0,16
Decoder 3	47,66 ± 11,25	0,64 ± 0,08	0,28 ± 0,18
Decoder 4	53,58 ± 23,12	0,60 ± 0,07	0,18 ± 0,14

Tabela 6: Teste de Wilcoxon dos experimentos *forward*.

Comparação	**95HD**	**Dice**	**IoU**
U-Net vs *Slice*	0,84	0,37	0,31
U-Net vs *Encoder 1*	0,43	1,00	0,78
U-Net vs *Encoder 2*	0,31	0,62	0,59
U-Net vs *Encoder 3*	0,56	0,21	0,21
U-Net vs *Bottleneck*	0,15	**0,03**	**0,03**
U-Net vs *Decoder 1*	**0,03**	**0,03**	**0,03**
U-Net vs *Decoder 2*	0,06	0,06	0,06
U-Net vs *Decoder 3*	**0,03**	**0,03**	**0,03**
U-Net vs *Decoder 4*	0,06	**0,03**	**0,03**

cativas. As configurações que rejeitaram a hipótese nula (*Bottleneck*, *Decoder 1*, *Decoder 3* e *Decoder 4*) apresentaram resultados substancialmente inferiores aos do modelo de referência, refletindo uma degradação na precisão da segmentação.

Embora *Slice*, *Encoder 1*, *Encoder 2* tenham alcançado valores ligeiramente melhores (menores) de 95HD do que a U-Net padrão, sugerindo maior consistência de borda e menos erros extremos de contorno, isso ocorreu em detrimento da área geral segmentada. Além disso, como mencionado anteriormente, essa pequena melhoria no 95HD não apresentou significância estatística.

Tabela 7: Resultados quantitativos dos experimentos *reverse*.

Métricas	**95HD**	**Dice**	**IoU**
U-Net	**24,13** ± 2,88	**0,85** ± 0,01	**0,70** ± 0,04
Decoder 4	41,37 ± 23,45	0,73 ± 0,13	0,51 ± 0,23
Decoder 3	31,08 ± 13,22	0,80 ± 0,08	0,59 ± 0,16
Decoder 2	36,17 ± 15,92	0,76 ± 0,11	0,51 ± 0,23
Decoder 1	48,80 ± 32,60	0,73 ± 0,15	0,45 ± 0,29
Bottleneck	39,08 ± 13,90	0,74 ± 0,11	0,46 ± 0,23

Tabela 8: Teste de Wilcoxon dos experimentos *reverse*.

Comparação	**95HD**	**Dice**	**IoU**
U-Net vs *Decoder 4*	0,15	0,21	0,15
U-Net vs *Decoder 3*	0,06	0,15	0,06
U-Net vs *Decoder 2*	0,03	**0,03**	**0,03**
U-Net vs *Decoder 1*	0,06	0,09	0,06
U-Net vs *Bottleneck*	**0,03**	**0,03**	**0,03**

5.1.2 Experimentos *Reverse*

Nos experimentos *reverse*, apresentados nas Tabelas 7 e 8, onde, respectivamente, temos os valores em negrito indicando: os melhores resultados em determinadas métricas e os experimentos capazes de rejeitar a hipótese nula com *p-value* inferior a 0,05. As mesmas estratégias de compartilhamento de características foram aplicadas, mas agora a partir da saída em direção à entrada, os modelos exibiram uma tendência comportamental semelhante. Apenas as configurações *Decoder 1* e *Decoder 3* demonstraram diferenças estatisticamente significativas em relação à linha de base, com ambas apresentando métricas de segmentação inferiores.

De modo geral, as estratégias investigadas de compartilhamento de características não proporcionaram melhorias quantitativas significativas em relação à U-Net de referência e, em cenários de propagação mais profunda, resultaram em uma degradação do desempenho.

5.2 Resultados Qualitativos

Para complementar os dados quantitativos, apresentamos os resultados qualitativos representativos das regiões do corpo e do ramo mandibular. Todas as visualizações foram

geradas a partir do conjunto de dados de teste utilizando a biblioteca `matplotlib`.

5.2.1 Experimentos do *Forward*

Nos experimentos do *forward*, as segmentações mais precisas foram majoritariamente observadas na região do corpo mandibular. A Figura 13 demonstra uma fatia típica dessa área. Mesmo entre as arquiteturas de pior desempenho, o tecido alvo permaneceu bem definido e precisamente localizado.

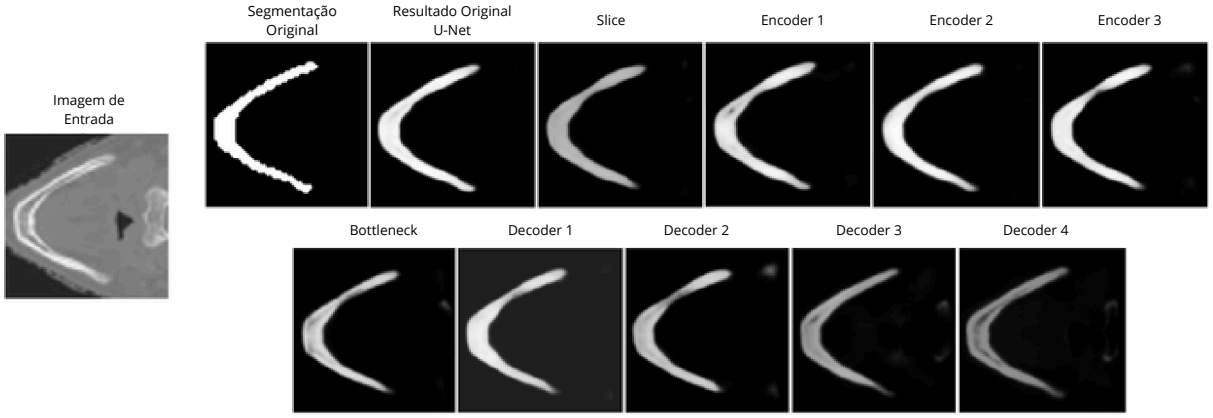


Figura 13: Resultados bons do corpo mandibular nos experimentos *forward*.

Um dos poucos cenários em que as redes apresentaram dificuldades na região do corpo mandibular ocorreu nas primeiras fatias tomográficas. Nessa configuração, uma pequena parte da mandíbula está presente, com apenas uma fração do seu formato característico e muito próxima do limite dos tecidos pertencentes ao paciente. Nesses casos houve uma maior presença de falsos positivos, afetando principalmente as métricas de precisão. Apesar disso, foram observados apenas em poucos exemplos qualitativos. Um exemplo pode ser visto na Figura 14.

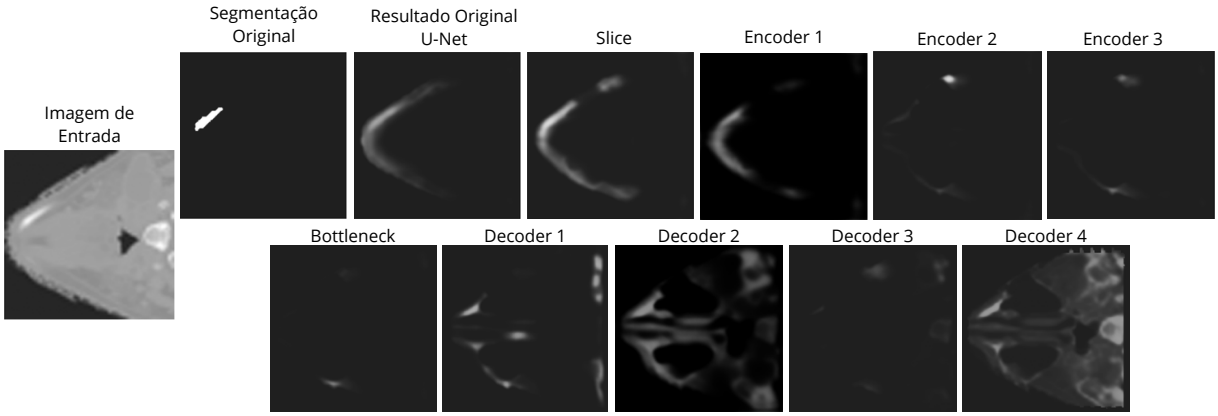


Figura 14: Resultados ruins do corpo mandibular nos experimentos *forward*.

Segmentações satisfatórias também foram alcançadas na região do ramo. Algumas configurações conseguiram discernir com sucesso as duas estruturas ósseas distintas, como é ilustrado na Figura 15.

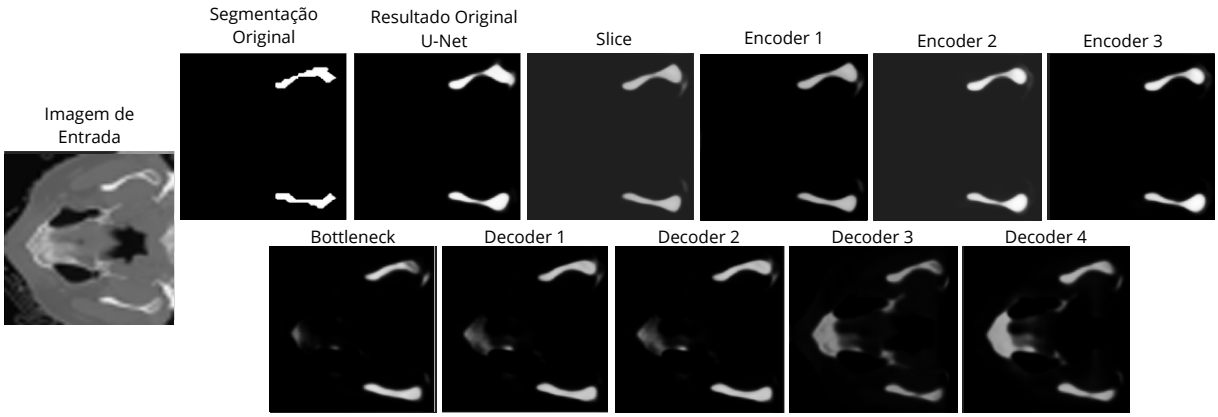


Figura 15: Resultados bons do ramo mandibular nos experimentos *forward*.

Por outro lado, a Figura 16 destaca os casos típicos de falha, particularmente a respeito da segmentação do côndilo. Nesses casos, o tecido-alvo não é o alvo principal da fatia. Mesmo as arquiteturas de melhor desempenho identificaram e delimitaram corretamente apenas o processo coronoide, falhando completamente em segmentar a região do côndilo. Nos casos mais graves de classificação incorreta, os modelos segmentaram erroneamente a maxila e os ossos pertencentes ao crânio em vez da mandíbula. Nas arquiteturas que se mantiveram competitivas em relação à U-Net padrão, observou-se um maior número de falsos negativos, ignorando a estrutura de interesse. Como é uma região presente consistentemente em todas as tomografias, afetou negativamente as métricas de forma significativa.

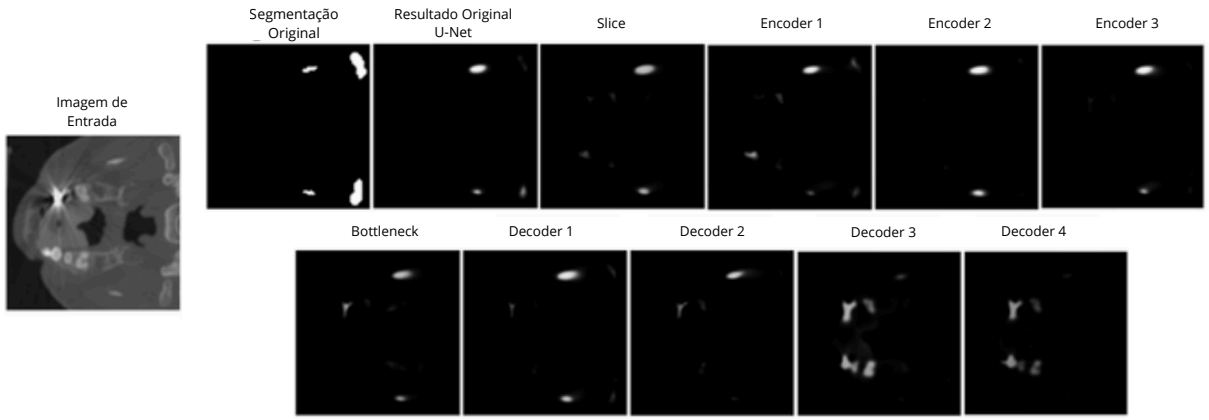


Figura 16: Resultados contendo o côndilo nos experimentos *forward*.

As observações qualitativas mostraram-se consistentes com os resultados quantitativos.

5.2.2 Experimentos *Reverse*

Seguindo o que observamos nos experimentos anteriores, o *reverse* apresentou comportamento similar em todas as regiões. No corpo da mandíbula, todas as configurações conseguem delimitar bem a região mandibular devido às características já mencionadas do entorno do órgão de interesse. Um exemplo característico pode ser visto na Figura 17

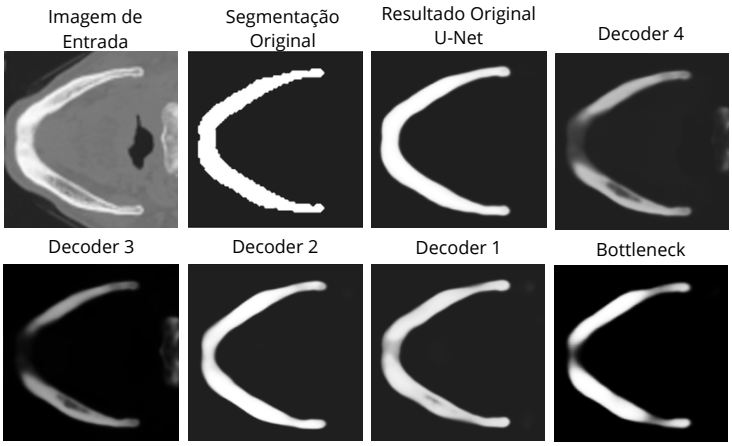


Figura 17: Resultados do corpo mandibular nos experimentos *reverse*.

Também foram observadas as mesmas dificuldades para segmentar corretamente o côndilo, com uma quantidade considerável de falsos negativos e falsos positivos na delimitação da mandíbula, como pode ser visto na Figura 18.

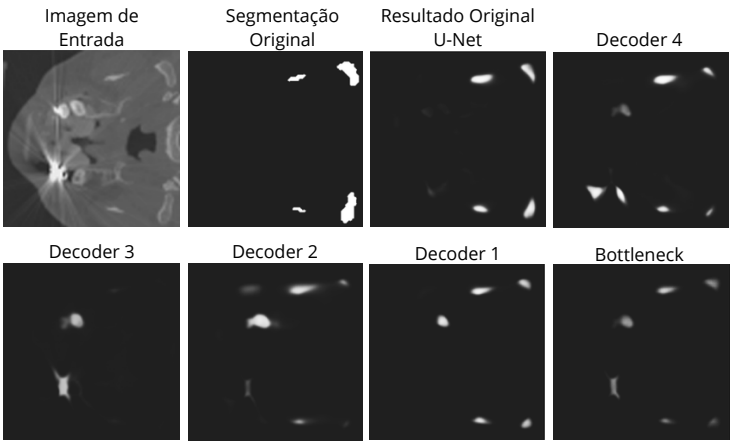


Figura 18: Resultados do côndilo nos experimentos *reverse*.

Devido a essa similaridade entre resultados, não continuamos os experimentos até a entrada da rede, observando os retornos decrescentes obtidos após os treinamentos.

5.3 Síntese do Capítulo

Os resultados quantitativos e qualitativos mostraram que, nas condições investigadas, o compartilhamento progressivo de mapas de características entre fatias consecutivas não produziu ganhos consistentes em relação à U-Net de referência. Entretanto, os experimentos permitiram caracterizar como diferentes níveis de compartilhamento influenciam o comportamento da rede, fornecendo evidências experimentais que servirão de base para a discussão apresentada no capítulo seguinte.

6 Discussão e Análise

A hipótese deste trabalho era que as modificações arquiteturais investigadas seriam capazes de melhorar o desempenho da segmentação, particularmente na região do côndilo mandibular, em comparação com a U-Net convencional. Contudo, os experimentos não demonstraram melhorias consistentes tanto nas métricas quantitativas quanto na qualidade das segmentações. Em alguns casos, as segmentações na região do corpo da mandíbula, anteriormente observadas como consistentes, também se tornaram menos estáveis.

É possível que a propagação progressiva de representações intermediárias tenha introduzido erros acumulados entre fatias adjacentes, dificultando a capacidade da rede de distinguir as estruturas mandibulares dos tecidos adjacentes. Isso é corroborado quando comparamos os experimentos do *forward* e *reverse*; ambos apresentaram métricas comparáveis nos primeiros experimentos, mas à medida que novos mapas de características foram sendo progressivamente empilhados, houve uma deterioração notável. Essa é uma explicação possível para o fato do *Encoder 2* apresentar melhores resultados, pois as características compartilhadas não atrapalharam o suficiente a capacidade de discernimento da rede e, como a diferença observada não foi estatisticamente significativa, a pequena vantagem numérica no 95HD obtida pela configuração pode ser atribuída à variabilidade inerente ao processo de treinamento.

A arquitetura U-Net de referência permaneceu como a mais consistente ao longo dos experimentos, o que sugere que a informação contextual adicional, introduzida pelo empilhamento de fatias e pela propagação de características, não foi suficientemente discriminativa para a segmentação da mandíbula. Diferentemente dos trabalhos de Wei Wang *et al.* (2019) e Qiu *et al.* (2021), que utilizam módulos de recorrência capazes de selecionar dinamicamente quais informações devem ser propagadas, a estratégia proposta realiza a concatenação direta de representações, sugerindo que mecanismos mais sofisticados de seleção de características podem ser necessários, ainda que impliquem maior custo computacional.

Uma vez que a maioria dos pixels nas imagens pertence ao fundo e a tecidos não mandibulares, e que essas regiões permanecem relativamente consistentes ao longo das fatias adjacentes, o empilhamento de fatias vizinhas pode ter reforçado informações redundantes em vez de características anatomicamente relevantes. Como resultado, o contexto adicional pode ter atuado como ruído, dificultando a distinção entre as estruturas-alvo e os tecidos adjacentes por parte da rede. Essa hipótese também é compatível com o tratamento adotado para a primeira fatia de cada tomografia. Como não existe uma fatia anterior da qual possam ser compartilhadas as representações, a própria representação da primeira fatia é reutilizada. Ainda assim, essa estratégia não produziu melhorias estatisticamente significativas em relação à U-Net, sugerindo que a simples reutilização de informações altamente correlacionadas não é suficiente para beneficiar a segmentação.

De forma semelhante, a propagação de mapas de características intermediários entre as fatias pode ter introduzido representações latentes redundantes ao longo das camadas. Como as camadas mais profundas já agregam uma ampla informação contextual por meio da arquitetura *encoder-decoder*, a adição de mapas de características externos provenientes de fatias adjacentes pode ter gerado redundância, comprometendo potencialmente a preservação de detalhes anatômicos locais discriminativos. Embora fatias consecutivas apresentem alta similaridade anatômica, os mapas de características gerados pela U-Net são produzidos especificamente para a imagem de entrada correspondente. Dessa forma, a reutilização direta dessas representações em uma fatia distinta pode introduzir ativações que deixam de refletir corretamente o conteúdo da imagem atual. Esse efeito tende a tornar-se mais pronunciado em camadas profundas, cujas representações são mais abstratas e dependentes do contexto global da imagem.

A estratégia investigada foi motivada pela hipótese de que a continuidade espacial existente entre fatias consecutivas poderia ser explorada por um mecanismo simples de compartilhamento de representações intermediárias, com o intuito de melhorar o desempenho das redes na região do côndilo. Essa hipótese era plausível, uma vez que estruturas anatômicas apresentam elevada correlação visual entre fatias adjacentes e abordagens recorrentes já haviam demonstrado benefícios ao explorar essa continuidade. Entretanto, os resultados indicam que a simples concatenação de mapas de características não foi suficiente para capturar essa informação de forma eficaz, sugerindo que mecanismos adaptativos de seleção ou atualização das representações são necessários. O côndilo permaneceu como uma região desafiadora para segmentação; mesmo utilizando informação entre fatias consecutivas, sua delimitação permaneceu inferior ao modelo de referência. Isso sugere que o principal desafio dessa região não está apenas na falta de contexto espacial, mas também

no baixo contraste encontrado entre o côndilo e estruturas ósseas vizinhas.

6.1 Análise por Grad-CAM

Como complemento à análise dos resultados, foram realizados experimentos utilizando o Grad-CAM, uma técnica de interpretação que produz mapas de importância visual a partir dos gradientes e dos mapas de ativação gerados durante a inferência. Esses mapas destacam as regiões da imagem que mais contribuíram para a predição da classe de interesse, permitindo uma análise qualitativa do comportamento da rede.

Em aplicações de classificação, é comum aplicar o Grad-CAM às últimas camadas convolucionais da rede, uma vez que elas representam características de maior nível semântico. Entretanto, em arquiteturas voltadas para segmentação, como a U-Net, a aplicação da técnica nas últimas camadas tende a produzir mapas fortemente correlacionados com a própria máscara prevista, reduzindo sua capacidade de fornecer informações adicionais sobre o processo de decisão do modelo. Por esse motivo, neste trabalho os gradientes e mapas de ativação foram extraídos da última camada do *encoder* (*Encoder 4*), imediatamente anterior ao *bottleneck*. Nessa posição, as representações ainda preservam características discriminativas de alto nível antes do processo de reconstrução espacial realizado pelo *decoder*, tornando os mapas mais informativos para interpretação.

Os mapas gerados para as diferentes configurações investigadas, contendo exemplos representativos das regiões do corpo mandibular, do ramo mandibular e do côndilo, são apresentados nas Figuras 19, 20, 21, 22 e 23. As cores nas figuras representam um mapa de calor (*heatmap*) que indica quais regiões da imagem tiveram maior influência na decisão do modelo. As cores mais quentes, como vermelho e amarelo, indicam as regiões que exerceram maior influência na decisão do modelo; as cores intermediárias, como verde, representam regiões com influência moderada, enquanto as cores mais frias, como azul e roxo, correspondem a áreas com pouca ou nenhuma contribuição para a decisão final.

Os mapas obtidos pelas configurações entre *Slice* e *Bottleneck* apresentaram comportamento semelhante. Na região do corpo mandibular, observou-se maior contribuição das regiões correspondentes aos contornos da mandíbula e da garganta para a predição da estrutura de interesse. Na região do ramo mandibular foi observado comportamento semelhante, com maior contribuição das bordas da mandíbula do que da região interna da estrutura. Também foram observadas ativações sobre os dentes da maxila em algumas imagens, sugerindo que essas estruturas exerceram influência sobre as decisões da

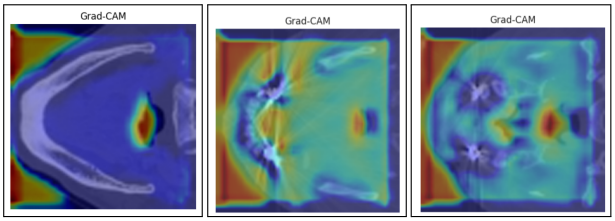
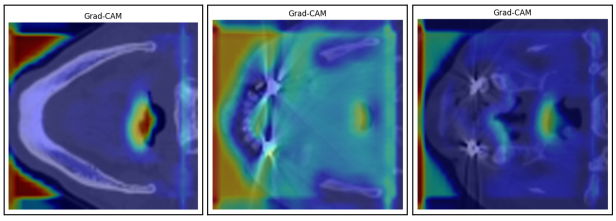
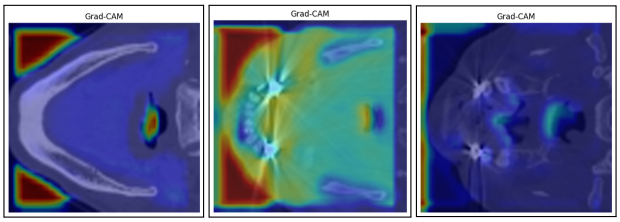
Configuração	Grad-Cam
U-Net	
Slice	
Encoder 1	

Figura 19: Grad-Cam: U-Net, *Slice* e *Encoder 1*

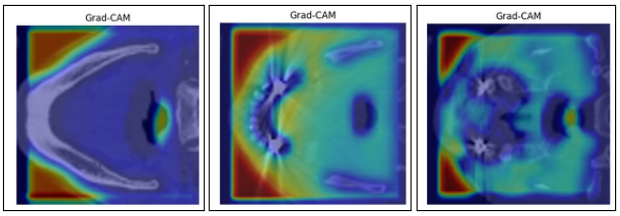
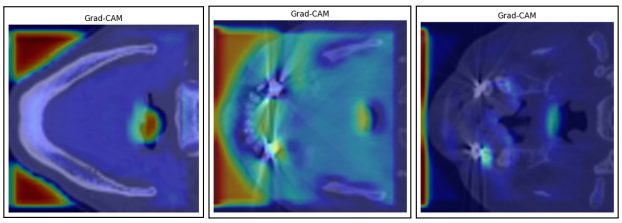
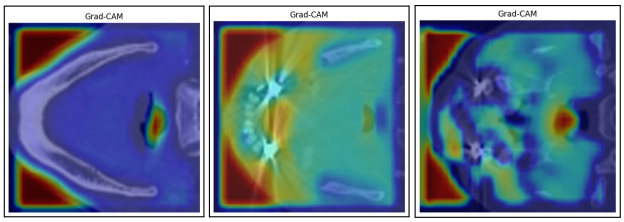
Configuração	Grad-Cam
Encoder 2	
Encoder 3	
Bottleneck	

Figura 20: Grad-Cam: *Encoder 2*, *Encoder 3*, *Bottleneck*

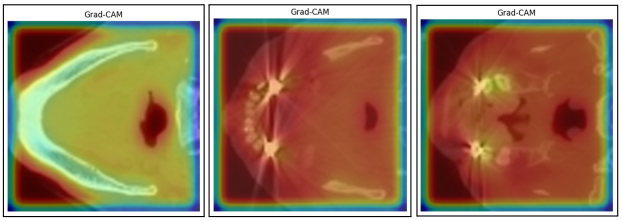
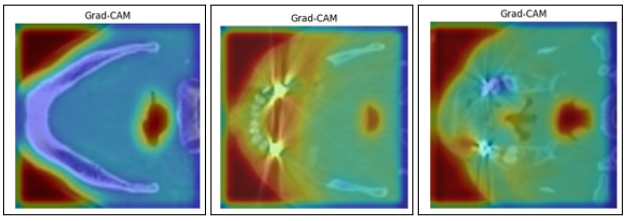
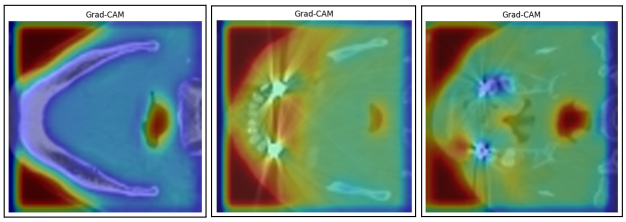
Configuração	Grad-Cam
Decoder 1	
Decoder 2	
Decoder 3	

Figura 21: Grad-Cam: *Decoder 1*, *Decoder 2* e *Decoder 3*

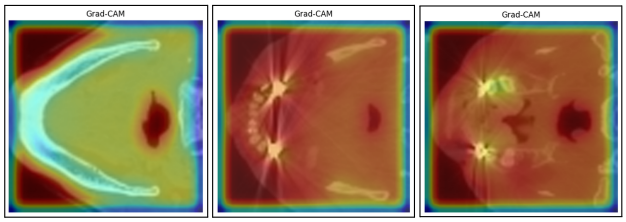
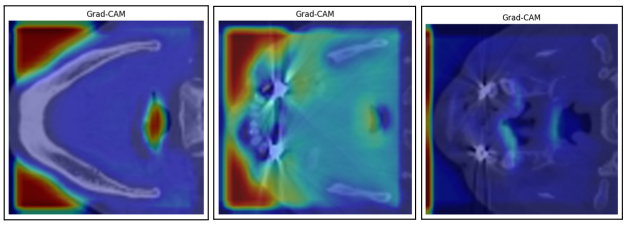
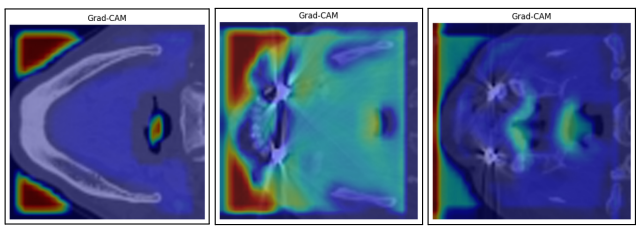
Configuração	Grad-Cam
Decoder 4	
Reverse Decoder 1	
Reverse Decoder 2	

Figura 22: Grad-Cam: *Decoder 4*, *Reverse Decoder 1* e *Reverse Decoder 2*

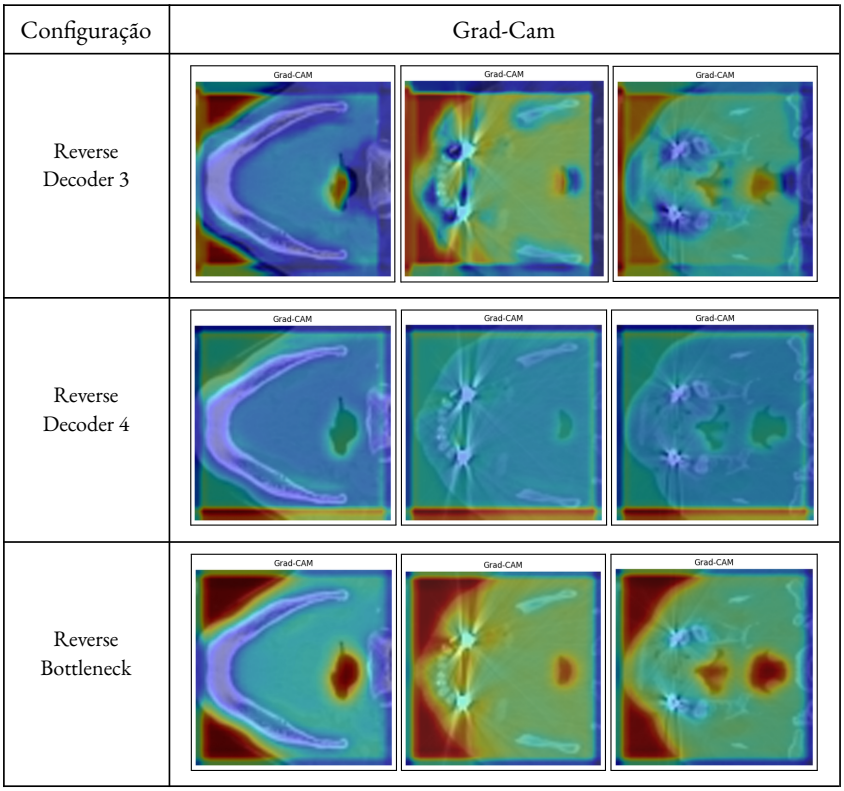


Figura 23: Grad-Cam: *Reverse Decoder 3*, *Reverse Decoder 2* e *Reverse Bottleneck*

rede, mesmo não pertencendo à classe alvo da segmentação. Na região do côndilo, os mapas apresentaram maior variabilidade entre as diferentes configurações. Ainda assim, mesmo nos casos em que as segmentações foram qualitativamente superiores, observou-se uma concentração menos localizada das regiões de maior contribuição, enquanto as áreas pertencentes ao côndilo apresentaram menor intensidade de ativação, com demasiada contribuição provinda dos contornos da região maxilar.

A partir da configuração *Bottleneck*, observou-se um aumento gradual da dispersão das regiões de maior contribuição ao longo da mandíbula, nas quais os mapas passaram a apresentar ativações mais distribuídas pelas áreas da imagem. Esse padrão é consistente com a degradação quantitativa observada nos experimentos, sugerindo que o compartilhamento progressivo de mapas de características pode ter reduzido a capacidade discriminativa das representações internas do modelo.

Comportamento semelhante foi observado nos experimentos *reverse*, nos quais os mapas também se tornaram progressivamente mais difusos à medida que níveis mais profundos da arquitetura passaram a compartilhar representações entre fatias consecutivas. Em conjunto, essas observações qualitativas reforçam a interpretação de que o compartilhamento direto de mapas de características, principalmente nas camadas mais profundas da

rede, tende a introduzir representações menos discriminativas, o que é compatível com a degradação de desempenho observada nas métricas quantitativas.

6.2 Limitações

O presente trabalho apresenta algumas limitações que devem ser consideradas na interpretação dos resultados. Primeiramente, os experimentos foram conduzidos exclusivamente utilizando a arquitetura U-Net, não sendo possível afirmar se o mesmo comportamento ocorreria em arquiteturas mais recentes baseadas em mecanismos de atenção ou Transformers. Além disso, a avaliação foi realizada apenas sobre a base de dados PDDCA, composta por exames de pacientes de cabeça e pescoço, o que pode limitar a generalização dos resultados para outros conjuntos de dados e protocolos de aquisição.

Outra limitação refere-se à própria estratégia de compartilhamento investigada. Os mapas de características foram propagados diretamente entre fatias consecutivas, sem qualquer mecanismo capaz de selecionar ou ponderar automaticamente quais informações deveriam ser reutilizadas. Dessa forma, dados redundantes ou pouco relevantes podem ter sido transmitidos entre as imagens, influenciando negativamente o desempenho da segmentação.

Por fim, este estudo concentrou-se exclusivamente em abordagens bidimensionais, não investigando arquiteturas tridimensionais nem estratégias híbridas que possam explorar simultaneamente informações intra-fatia e inter-fatias.

6.3 Síntese do Capítulo

Neste capítulo apresentamos uma discussão sobre os resultados obtidos, indicando que a simples propagação direta de mapas de características entre fatias consecutivas não foi suficiente para melhorar a segmentação mandibular. Embora a hipótese inicial não tenha sido confirmada, os experimentos permitiram identificar limitações importantes dessa estratégia e apontar direções promissoras para futuras investigações envolvendo mecanismos adaptativos de compartilhamento de informações.

7 Conclusão

Esta dissertação investigou a estratégia de compartilhamento de mapas de características entre fatias consecutivas em uma arquitetura U-Net 2D, avaliada por diferentes configurações de propagação (*forward* e *reverse*), analisando o impacto do empilhamento de cortes tomográficos e mapas de características entre fatias adjacentes na segmentação de mandíbulas.

Os experimentos mostraram que a U-Net permaneceu como a arquitetura mais consistente. O compartilhamento em camadas superficiais apresentou desempenho semelhante ao modelo de referência, sem diferenças estatisticamente significativas. Em contrapartida, o compartilhamento em camadas profundas promoveu degradação progressiva das métricas Dice, IoU e 95HD.

Esses resultados indicam que a propagação de mapas de características entre cortes adjacentes pode introduzir redundância nas representações latentes, sem contribuir efetivamente para a delimitação e discriminação dos tecidos anatômicos. Consequentemente, a precisão das representações espaciais e locais é reduzida, comprometendo o desempenho da segmentação. Portanto, nas condições avaliadas, a propagação direta de mapas de características entre fatias consecutivas, sem mecanismos adaptativos de seleção, não produziu ganhos consistentes para a segmentação mandibular em arquiteturas U-Net 2D.

Em síntese, este trabalho evidencia que a incorporação de informações contextuais adjacentes em modelos 2D permanece um desafio na visão computacional aplicada à área médica. Nesse contexto, estratégias mais sofisticadas de seleção e integração das informações propagadas representam uma direção promissora para ampliar o uso de contexto entre fatias.

7.1 Trabalhos Futuros

Os resultados obtidos sugerem diferentes possibilidades para trabalhos futuros. Uma direção promissora consistiria em substituir a concatenação direta dos mapas de características por mecanismos capazes de selecionar adaptativamente quais representações devem ser propagadas entre fatias, como módulos de atenção ou unidades ConvLSTM.

Outra linha de investigação consiste em investigar o compartilhamento apenas de subconjuntos específicos dos mapas de características ou aplicar mecanismos de ponderação capazes de reduzir a propagação de informações redundantes. Da mesma forma, a integração da estratégia investigada com arquiteturas mais recentes, como *Attention* U-Net, nnU-Net ou modelos baseados em Transformers, pode permitir avaliar se representações mais discriminativas tornam o compartilhamento entre fatias mais eficaz.

Também seria interessante investigar diferentes formas de exploração do contexto inter-fatias. Em vez de utilizar apenas a fatia imediatamente anterior, podem ser consideradas janelas compostas por múltiplas fatias consecutivas, estratégias bidirecionais que utilizem simultaneamente informações das fatias anterior e posterior, ou ainda mecanismos que realizem o compartilhamento apenas quando houver elevada similaridade estrutural entre as imagens adjacentes.

Neste trabalho, cada experimento avaliou o compartilhamento de mapas de características de forma progressiva ao longo da U-Net, de modo que o compartilhamento em uma determinada camada necessita do compartilhamento das camadas anteriores. Assim, não foram investigadas combinações em que níveis específicos fossem compartilhados de forma independente, como, por exemplo, compartilhar apenas o *Encoder 2* sem compartilhar previamente *Slice* e *Encoder 1*. Uma extensão natural consiste em investigar o compartilhamento simultâneo entre diferentes níveis da arquitetura, sem a restrição sequencial, explorando combinações entre camadas do *encoder*, *bottleneck* e *decoder*. Essa abordagem permitiria avaliar se determinadas combinações de representações intermediárias são mais eficazes do que o compartilhamento progressivo adotado neste trabalho.

Outra direção promissora consiste em investigar estratégias que permitam compreender de forma mais detalhada o impacto do compartilhamento de mapas de características ao longo da arquitetura. Análises da distribuição das ativações, da similaridade entre representações produzidas por fatias consecutivas e da evolução das características compartilhadas em diferentes níveis do *encoder* e do *decoder* podem contribuir para identificar em quais etapas ocorre a degradação do desempenho e quais tipos de informação são efe-

tivamente reutilizados pela rede.

Outra possibilidade consiste em realizar uma análise mais detalhada do impacto computacional das diferentes estratégias de compartilhamento, considerando consumo de memória, tempo de treinamento, tempo de inferência e escalabilidade para volumes tomográficos maiores. Essa avaliação permitiria comparar de forma mais abrangente o custo-benefício das abordagens bidimensionais com compartilhamento de características em relação às arquiteturas tridimensionais.

Por fim, a abordagem poderia ser validada em outros conjuntos de dados e modalidades de imagem, bem como investigar técnicas híbridas que combinem o baixo custo computacional das arquiteturas bidimensionais com mecanismos mais sofisticados de modelagem da continuidade espacial entre fatias consecutivas.

REFERÊNCIAS

ADAMS, R.; BISCHOF, L. Seeded Region Growing. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 16, n. 6, p. 641–647, 1994. DOI: 10.1109/34.295913.

ALOM, M. D. Zahangir *et al.* Recurrent Residual Convolutional Neural Network Based on U-Net (R2U-Net) for Medical Image Segmentation. **arXiv preprint arXiv:1802.06955**, 2018.

ANATOMOGRAPHY. **Mandible Close-Up Superior**. [*S. l.: s. n.*], 2012. Wikimedia Commons. Acessado em 28 jun. 2026. Licença: Creative Commons Attribution-ShareAlike 2.1 Japan (CC BY-SA 2.1 JP). Disponível em: https://commons.wikimedia.org/wiki/File:Mandible_close-up_superior.png.

AZAD, Reza *et al.* Bi-Directional ConvLSTM U-Net with Densley Connected Convolutions. **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops**, 2019.

BREELAND, Grant; AKTAR, Aylin; PATEL, Bhupendra C. Anatomy, Head and Neck, Mandible. **StatPearls Publishing**, 2023.

CARMO, Rafael Lourenço do; CHAVES, Catarina. **Mandíbula**. [*S. l.: s. n.*], 2023. Disponível em https://www.kenhub.com/pt/library/anatomia/a-mandibula, acessado em 25/06/2026.

ÇIÇEK, Özgün *et al.* 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. **International Conference on Medical Image Computing and Computer-Assisted Intervention**, p. 424–432, 2016.

CODELLA, Noel C. F. *et al.* Skin Lesion Analysis Toward Melanoma Detection. **International Symposium on Biomedical Imaging (ISBI)**, p. 168–172, 2018.

CRIMINISI, Antonio; SHOTTON, Jamie. **Decision Forests for Computer Vision and Medical Image Analysis**. [*S. l.*]: Springer Science & Business Media, 2013.

DOT, Gauthier *et al.* Fully Automatic Segmentation of Craniomaxillofacial CT Scans for Computer-Assisted Orthognathic Surgery Planning Using the nnU-Net Framework. **European Radiology**, Springer, v. 32, n. 6, p. 3639–3648, 2022.

EKMAN, Magnus. **Learning Deep Learning: Theory and Practice of Neural Networks, Computer Vision, Natural Language Processing, and Transformers Using TensorFlow**. [*S. l.*]: Addison-Wesley Professional, 2021.

GONZALEZ, Rafael C.; WOODS, Richard C. **Processamento Digital de Imagens**. 3. ed. São Paulo: Pearson, 2010.

HARIADHI. **Mandible**. [*S. l.: s. n.*], 2020. Wikimedia Commons. Acesso em: 27 jun. 2026. Licença CC BY-SA 4.0. Disponível em: https://commons.wikimedia.org/wiki/File:Mandible_svg_hariadhi.svg.

HOUNSFIELD, Godfrey N. Computerized transverse axial scanning (tomography): Part 1. Description of system. **The British journal of radiology**, The British Institute of Radiology, v. 46, n. 552, p. 1016–1022, 1973.

ILEŞAN, Robert R. *et al.* Comparison of Artificial Intelligence-Based Applications for Mandible Segmentation: From Established Platforms to In-House-Developed Software. **Bioengineering**, MDPI, v. 10, n. 5, 2023. Disponível em: https://www.mdpi.com/2306-5354/10/5/604.

IRSHAD, Talal Bin *et al.* Mandibular Bone Segmentation from CT Scans: Quantitative and Qualitative Comparison Among Software. **Dental Materials**, Elsevier, v. 40, n. 8, e11–e22, 2024.

ISENSEE, Fabian *et al.* nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. **Nature Methods**, Nature Publishing Group US New York, v. 18, n. 2, p. 203–211, 2021.

KASS, Michael; WITKIN, Andrew; TERZOPOULOS, Demetri. Snakes: Active Contour Models. **International Journal of Computer Vision**, Springer, v. 1, n. 4, p. 321–331, 1988.

LECUN, Yann *et al.* Backpropagation Applied to Handwritten Zip Code Recognition. **Neural Computation**, MIT Press, v. 1, n. 4, p. 541–551, 1989.

LI, Qiankun *et al.* Empowering 2D neural network for 3D medical image segmentation via neighborhood information fusion. **Pattern Recognition**, p. 113035, 2026.

LONG, Jonathan; SHELHAMER, Evan; DARRELL, Trevor. Fully Convolutional Networks for Semantic Segmentation. **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**, 2015.

MA, Jun *et al.* Segment Anything in Medical Images. **Nature Communications**, Nature Publishing Group UK London, v. 15, n. 1, p. 654, 2024.

OKTAY, Ozan *et al.* Attention U-Net: Learning Where to Look for the Pancreas. **arXiv preprint arXiv:1804.03999**, 2018.

OTSU, Nobuyuki. A Threshold Selection Method from Gray-Level Histograms. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 9, n. 1, p. 62–66, 1979. DOI: 10.1109/TSMC.1979.4310076.

PANKERT, Tobias *et al.* Mandible Segmentation from CT Data for Virtual Surgical Planning Using an Augmented Two-Stepped Convolutional Neural Network. **International Journal of Computer Assisted Radiology and Surgery**, Springer, v. 18, n. 8, p. 1479–1488, 2023.

PASZKE, Adam *et al.* Pytorch: An imperative style, high-performance deep learning library. **Advances in neural information processing systems**, v. 32, 2019.

PHAM, William *et al.* A Comprehensive Framework for Automated Segmentation of Perivascular Spaces in Brain MRI with the nnU-Net. **Neuroradiology**, Springer, p. 1–19, 2026.

QIU, Bingjiang *et al.* Recurrent Convolutional Neural Networks for 3D Mandible Segmentation in Computed Tomography. **Journal of Personalized Medicine**, MDPI, v. 11, n. 6, p. 492, 2021.

RAUDASCHL, Patrik F. *et al.* Evaluation of Segmentation Methods on Head and Neck CT: Auto-Segmentation Challenge 2015. **Medical Physics**, Wiley Online Library, v. 44, n. 5, p. 2020–2036, 2017.

REZA, Md Himel *et al.* A comprehensive review of convolutional neural networks: foundations, enhancements and applications. **Neural Computing and Applications**, Springer, v. 38, n. 4, p. 56, 2026.

RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. U-Net: Convolutional Networks for Biomedical Image Segmentation. **International Conference on Medical Image Computing and Computer-Assisted Intervention**, p. 234–241, 2015.

ROTH, Holger *et al.* **Data From Pancreas-CT (Version 2)**. [*S. l.*]: The Cancer Imaging Archive, 2016. DOI: 10.7937/K9/TCIA.2016.tNB1kqBU. Disponível em: https://doi.org/10.7937/K9/TCIA.2016.tNB1kqBU.

SABANCI, Kadir; ASLAN, Busra; ASLAN, Muhammet Fatih. Medical Image Segmentation Methods: A Decision-Guided Survey Covering 2D/3D CNNs, Transformers, VLMs, SAM-Based Models and Diffusion Approaches. **Bioengineering**, MDPI, v. 13, n. 5, p. 555, 2026.

SELVARAJU, Ramprasaath R *et al.* Grad-cam: Visual explanations from deep networks via gradient-based localization. **Proceedings of the IEEE/CVF International Conference on Computer Vision**, p. 618–626, 2017.

STAAL, Joes *et al.* Ridge-Based Vessel Segmentation in Color Images of the Retina. **IEEE Transactions on Medical Imaging**, IEEE, v. 23, n. 4, p. 501–509, 2004.

VOSSOUGH, Arastoo *et al.* Training and Comparison of nnU-Net and DeepMedic Methods for Autosegmentation of Pediatric Brain Tumors. **American Journal of Neuroradiology**, American Society of Neuroradiology, v. 45, n. 8, p. 1081–1089, 2024.

WANG, Risheng *et al.* Medical image segmentation using deep learning: A survey. **IET image processing**, Wiley Online Library, v. 16, n. 5, p. 1243–1267, 2022.

WANG, Wei *et al.* Recurrent U-Net for Resource-Constrained Segmentation. **Proceedings of the IEEE/CVF International Conference on Computer Vision**, p. 2142–2151, 2019.

XIA, Qingling *et al.* A comprehensive review of deep learning for medical image segmentation. **Neurocomputing**, Elsevier, v. 613, p. 128740, 2025.

XIE, Enze *et al.* SegFormer: Simple and efficient design for semantic segmentation with transformers. **Advances in neural information processing systems**, v. 34, p. 12077–12090, 2021.

XU, Yan *et al.* Advances in medical image segmentation: A comprehensive review of traditional, deep learning and hybrid approaches. **Bioengineering**, MDPI, v. 11, p. 1034, 2024.