

Avaliação da qualidade e explicabilidade das imagens gerada pela rede U-Mamba

Carlos Emanuel Silva e Melo Oliveira¹, Gustavo Andrade Alves¹

¹Engenharia de Software – Pontifícia Universidade Católica de Minas Gerais (PUC MG)
Belo Horizonte – MG – Brazil

{carlos.oliveira.1405439, gaalves}@sga.pucminas.br

Abstract. *Medical image segmentation (MIS) is an extremely important activity for computer-aided diagnosis (CAD), therefore advances in this area of research have been intense in recent years and have culminated in U-Mamba, a deep learning application that combines Convolutional Neural Networks (CNN) and State Space Models (SSM). This research studies the impact of changing the CNN-SSM block of the U-Mamba network on the MIS process and relates this impact to classic metrics which evaluate the performance of image segmentation models: the Dice Similarity Coefficient (DSC) and the Normalized Surface Distance (NSD). A memory module was introduced in the original model and this new version of U-Mamba was tested on a dataset of abdominal organs. The main results found by the research indicate that the introduction of a memory block after SSM produces a model with better results compared to current literature.*

Resumo. *A segmentação de imagens biomédicas (MIS, do inglês Medical Image Segmentation) é uma atividade de suma importância para o diagnóstico auxiliado por computador (DAC), portanto os avanços nessa área de pesquisa foram intensos nos últimos anos e culminaram no U-Mamba, uma aplicação de deep learning que combina Rede Neurais Convolucionais (CNN, do inglês Convolutional Neural Networks) e Modelos de Estado de Espaço (SSM, do inglês State Space Model). Esta pesquisa estuda o impacto da alteração do bloco CNN-SSM da rede U-Mamba no processo de MIS e relaciona esse impacto com métricas clássicas as quais avaliam o desempenho de modelos de segmentação de imagem: o Coeficiente de Similaridade de Dice (DSC, do inglês Dice Similarity Coefficient) e a Distância de Superfície Normalizada (NSD, do inglês Normalized Surface Distance). Para realizar este trabalho foi implementado um módulo de memória no modelo original e essa nova versão do U-Mamba foi testada em um dataset de órgãos do abdômen. Os principais resultados encontrados pela pesquisa apontam que a introdução de um bloco de memória após o SSM produz um modelo com resultados comparáveis com a literatura atual.*

Bacharelado em Engenharia de Software - PUC Minas
Trabalho de Conclusão de Curso (TCC)

Orientador de conteúdo (TCC I): Joana Souza - joanasouza@pucminas.br
Orientador de conteúdo (TCC I): João Pedro Batisteli - jp.batisteli@hotmail.com
Orientador de conteúdo (TCC I): Leonardo Vilela - leonardocardoso@pucminas.br
Orientador acadêmico (TCC I): Cleiton Tavares - cleitontavares@pucminas.br
Orientador do TCC II: Leonardo Vilela - leonardocardoso@pucminas.br

Belo Horizonte, 23 de Novembro de 2025.

1. Introdução

O DAC é uma importante ferramenta para profissionais da área de saúde que interpretarem imagens médicas geradas por ressonância magnética (RMI), tomografia computadorizada (TC), raio-X e tomografia por emissão de pósitron (TEP) [Patel 2023]. Para a realização do DAC o processo de MIS é de suma importância, pois permite separar estruturas ou regiões de interesse em imagens de modo a conceder uma avaliação precisa que auxilia na tomada de decisões terapêuticas [Ronneberger et al. 2015]. Na última década, arquiteturas baseadas em CNNs tornaram-se relevantes no âmbito de MIS [Patel 2023] e, mais recentemente, transformadores (em inglês *transformers*) também tem ganhado popularidade nesse campo [Armand et al. 2024].

Neste contexto, SSMs emergem como uma nova abordagem para a construção de redes profundas e o *U-Mamba* é uma das encarnações a qual adota essa estratégia [Ma et al. 2024]. Existem diversos trabalhos que abordam métricas e *frameworks* de avaliação de modelos de MIS como os trabalhos de [Gopigari and Singh 2022] e [Salvatierra-Guillermo and Zavaleta Jave 2024], contudo poucos estudos abordam a questão da visualização dos resultados do processo de segmentação. Desse modo, **o problema que este trabalho aborda é a carência de melhorias na visualização de imagens médicas geradas pelos modelos de MIS.**

Avaliar as redes que realizam o processo de MIS é uma tarefa crítica, porque garante a efetividade e precisão dos resultados [Rayed et al. 2024]. Para realizar essa atividade existem diversas métricas como: DSC, Interseção sobre união (IoU, do inglês *Intersection-Over-Union*), Característica de Operação do Receptor (ROC, do inglês *Receiver Operating Characterist*), Área sob a curva de ROC (AUC, do inglês *Area under the ROC curve*), e a Distância média de Hausdorff (AHD, do inglês *Average Hausdorff Distance*). [Müller et al. 2022]. Ainda que essas métricas sejam uma forma sólida de medir a qualidade do modelo é fortemente aconselhado fornecer visualizações das segmentações obtidas de modo a permitir comparações com o *ground truth*, avaliar visualmente os resultados e evitar vieses estatísticos [Müller et al. 2022].

O objetivo geral deste trabalho é **investigar o impacto de uma inclusão de memória na arquitetura do modelo *U-Mamba* e analisar os efeitos dessa alteração na qualidade do modelo explorando seu potencial evolutivo.** Os objetivos específicos incluem: (I) investigar o impacto de modificações estruturais no bloco CNN-SSM do *U-Mamba*; (II) avaliar o desempenho das modificações utilizando as métricas supracitadas; (III) avaliar a explicabilidade e qualidade das imagens geradas pelo modelo de MIS.

Como resultado deste estudo, encontrou-se uma relação entre as modificações fei-

tas na rede *U-Mamba* e a qualidade das imagens resultantes do processo de segmentação. Além disso, também foi obtida uma relação entre as métricas de avaliação de modelos (DSC, IoU, AUC, etc) com a explicabilidade da imagem gerada ao final do processo.

Este documento está organizado da seguinte forma: na seção 2 são apresentados os termos e conceitos relevantes para o presente estudo, em seguida, na Seção 3, são apresentados os trabalhos relacionados, discutindo os avanços recentes em segmentação médica com *UNet* e *Transformers*. Na Seção 4, detalham-se os materiais e métodos utilizados, incluindo as configurações experimentais e os *datasets* empregados. Em seguida, a seção 5 apresenta os resultados obtidos, a seção 6 as ameaças à validade do trabalho e, por fim, a seção 7 sintetiza as principais conclusões encontradas durante a pesquisa.

2. Fundamentação Teórica

As subseções a seguir apresentam os conceitos necessários para o entendimento da experiência e estão organizadas da seguinte forma: a seção 2.1 introduz o conceito de Segmentação de Imagem, em seguida a seção 2.2 foca em explicar o que *Machine Learning* e *Deep Learning* e finalmente a seção 2.3 apresenta o modelo *U-Mamba*.

2.1. Segmentação de Imagem

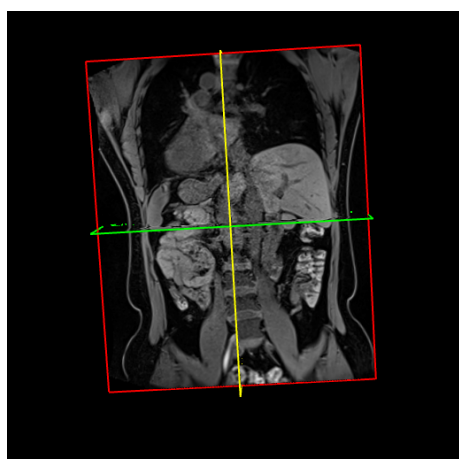
O processo de segmentar uma imagem consiste em encontrar grupos de *pixels* os quais combinam de modo a identificar objetos em uma imagem. [Szeliski 2022]. Portanto, um algoritmo capaz de segmentar uma imagem deve ser capaz de dividir, isolar e reconhecer objetos nessa imagem [Brice and Fennema 1970]. No contexto de imagens médicas (RMI, CT, raio x, etc) a segmentação serve como meio de dividir imagens em estruturas anatômicas significativas, esse processo é chamado de MIS [Ma et al. 2024].

A Figura 1 apresenta um exemplo de MIS: na Figura 1a é visto uma ressonância magnética do abdômen em escala de cinza e na Figura 1b é possível ver que alguns órgãos foram separados e o resto da imagem, o qual é uma região de pouco ou nenhum interesse, foi suprimido. Dito de outra forma, os *pixels* que formam os órgãos de interesse foram agrupados e isolados de modo a serem facilmente reconhecíveis no resultado da segmentação.

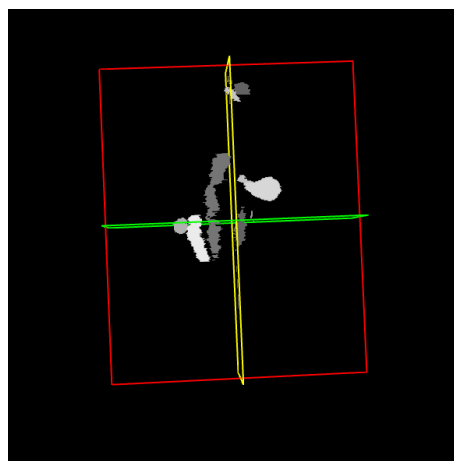
2.2. Machine Learning e Deep Learning

O *Machine Learning* é um campo de estudo que visa programar computadores para que eles sejam capazes de aprender através da experiência [Samuel 1959]. Em um de seus trabalhos pioneiros, Arthur Samuel demonstrou que um computador poderia aprender a jogar damas melhor do que seu próprio programador, ajustando parâmetros internos com base nos resultados dos jogos. Esse princípio fundamental de ajustar um modelo para melhorar seu desempenho em uma tarefa, com base em dados, é o cerne do *Machine Learning*.

O *Deep Learning* é uma classe de *Machine Learning* que revolucionou a área ao introduzir o conceito de *representation learning* [LeCun et al. 2015]. Em vez de exigir uma extração manual das características para transformar dados brutos (como os valores de pixels de uma imagem) em uma representação adequada, os próprios métodos de *Deep Learning* descobrem automaticamente as representações necessárias para a tarefa, como detecção ou classificação.



(a) Imagem de ressonância magnética de um abdômen



(b) Resultado da segmentação

Figura 1: Exemplo de segmentação de imagens do MICCAI 2022 AMOS

Esse paradigma deu origem a diversas arquiteturas especializadas, cada uma otimizada para diferentes tipos de dados e tarefas. Três abordagens se destacam no cenário atual: em primeiro lugar, as CNNs as quais recebem esse nome por causa da convolução, uma operação matemática que recebe como argumento duas funções e resulta em uma combinação dessas duas funções, desse modo, no contexto de *Deep Learning*, uma CNN é uma rede neural artificial responsável por processar dados com um formato de grade e, em ao menos uma de suas camadas, faz uso da convolução ao invés de uma multiplicação de matrizes [Goodfellow et al. 2016].

Em segundo lugar, os **Transformers** que foram originalmente propostos para tarefas de processamento de linguagem natural e introduziram o mecanismo de atenção. Essa inovação permite que o modelo atribua pesos de forma dinâmica para diferentes partes de uma sequência de entrada ao gerar uma saída. Isso viabiliza a captura de dependências de longo alcance, ou seja, a capacidade de correlacionar características (*features*) espacialmente distantes em uma imagem [Vaswani et al. 2023], em um contexto de segmentação de imagens médicas, isso ajuda, por exemplo, a identificar que um órgão é o fígado pela sua posição em relação a coluna vertebral.

Finalmente, os SSMs que são uma classe de modelos de sequência que se destacam pela eficiência em capturar dependências em sequências muito longas. Eles mapeiam uma sequência de entrada para um estado latente e o evoluem ao longo do tempo. Arquiteturas recentes, como o *Mamba* [Gu and Dao 2023], aprimoraram os SSMs com um mecanismo de seleção de dados, fazendo com que tenham uma melhor eficiência computacional, e tornando-os promissores para uma vasta gama de aplicações.

2.3. U-Mamba

O *U-Mamba* é uma arquitetura de Redes Neurais Artificiais (ANN, do inglês *Artificial Neural Network*) híbrida a qual combina CNN e *Structures State Space* (S4) para realizar segmentação de imagens médicas em 2D ou 3D [Ma et al. 2024]. No modelo são introduzidos os blocos *U-Mamba* os quais são compostos por dois blocos residuais consecutivos e um bloco *Mamba* em seguida.

A arquitetura geral, como apresentado na Figura 2, segue o *framework encoder-decoder* formado por dois blocos principais: o *encoder* (blocos à esquerda) que recebe uma sequência de tamanho variado e o *decoder* (blocos à direita) o qual gera uma sequência desejada. O *encoder* e o *decoder* no presente modelo são formados, respectivamente, por blocos *U-Mamba* e por blocos residuais e convoluções transpostas [Ma et al. 2024]. A escolha desse *framework* é justificada pelo fato de ser a base para diferentes modelos que lidam com longas sequências [Zhang et al. 2023].

Por fim, existem dois tipos de bloco *U-Mamba*: o *U-Mamba_Bot* e *U-Mamba_Enc*. O primeiro designa quando o bloco *U-Mamba* é usado apenas no *bottleneck* e o segundo quando o bloco é usado em todo o *encoder* [Ma et al. 2024].

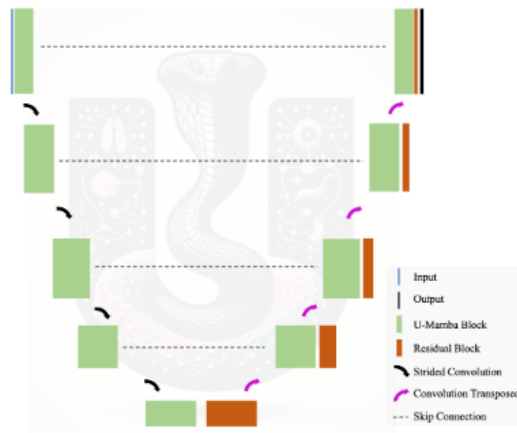


Figura 2: Representação conceitual do *framework encoder-decoder* [Ma et al. 2024]

3. Trabalhos Relacionados

Nesta seção são discutidos os estudos de outros autores os quais auxiliam na contextualização do trabalho, apresentam métodos reutilizados neste artigo e executam processos e análises semelhantes às desta pesquisa. Mais especificamente, os textos detalhados neste segmento envolvem a proposição de novos métodos para a segmentação de imagens médicas, a apresentação de métodos para avaliar a qualidade dessas segmentações e estudos os quais comparam diferentes métodos.

Ma et al. (2024) propõe o *U-Mamba*, uma arquitetura híbrida que combina CNNs e SSMs para MIS. Essa abordagem visa capturar tanto características locais quanto dependências de longo alcance em imagens. Foi realizado experimentos em quatro regiões distintas para segmentação (órgãos abdominais em CT e MRI, instrumentos em imagens endoscópicas e células em imagens de microscopia) e o resultado demonstrou que o *U-Mamba* supera redes baseadas em CNNs e *Transformers*, como *nnU-Net* e *SwinUNETR*, em métricas como DSC e NSD. Este trabalho é importante para o presente artigo que visa aprimorar o *U-Mamba* utilizando seu repositório público como base, explorando otimizações na integração CNN-SSM.

Armand et al. (2024) explora a eficácia dos *transformers* em arquiteturas baseadas na *UNet* para segmentação de imagens médicas, comparando quatro modelos: *UNet*, *UNetR*, *Trans-UNet* e *Swin-UNet*. A pesquisa busca entender como os *transformers* podem melhorar a precisão da segmentação médica, tarefa essencial para diagnósticos e

tratamentos clínicos. Os autores utilizaram quatro conjuntos de dados distintos (retinais, endoscópicos, histopatológicos e multi-órgãos) para avaliar as arquiteturas com métricas padrão como DSC e IoU. Os resultados obtidos pelos autores mostram que os modelos baseados em *transformers* (especialmente *Trans-UNet* e *Swin-UNet*) superaram o *UNet* tradicional em muitos casos, mas o desempenho varia dependendo do conjunto de dados. Esse trabalho avalia modelos de segmentação de imagens assim como é proposto neste artigo, contudo esta pesquisa avança nas comparações incluindo modelos baseados em SSM.

Salvaierra-Guillermo e Zavaleta Jave (2024) apresenta um estudo o qual compara dois algoritmos *open-source* de segmentação de imagens: MONAI (*Medical Open Network for AI*) e SAMM (*Segment Any Medical Model*) e tem como principal objetivo identificar seus pontos fortes, fracos e a utilidade em aplicações reais. Para realizar essa comparação foi utilizada a biblioteca *Python MISEval* (*Medical Image Segmentation Evaluation*) e, mais especificamente, as seguintes métricas: DSC, AHD e Similaridade Volumétrica (VS, do inglês *Volumetric Similarity*). Durante os testes, os pesquisadores utilizaram uma base de dados de imagens do baço e concluíram que o MONAI apresenta melhor performance nessa tarefa. Os autores realizam um estudo comparativo no âmbito da segmentação de imagens biomédicas e as métricas de avaliação utilizadas pelos autores são empregadas neste trabalho.

Gopigari e Singh (2022) realizam uma comparação entre as técnicas de segmentação de imagem e técnicas de detecção de borda no contexto da detecção de células cancerígenas originárias da Leucemia. Para realizar a comparação das técnicas os pesquisadores utilizaram 260 imagens (130 de células saudáveis e 130 de células não saudáveis) com tamanho de 257 x 257 as quais foram obtidas na base de dados ALL-IDB2 (*Acute Lymphoblastic Leukemia Image Database*) e seguiram a seguinte metodologia: (1) redimensionar a imagem para o tamanho 256 x 256, (2) transformar a imagem para a escala de cinza, (3) aplicar filtros, (4) performar a detecção de bordas, (5) performar a segmentação de imagem e (6) obter as métricas. Cada um desses passos forma executados para cada técnica avaliada no estudo. As métricas utilizadas foram *Mean Absolute Error* e *Peak Signal to Noise Ratio*. Este trabalho também avalia comparativamente técnicas de segmentação de imagem, contudo foca no uso de métricas e não na imagem gerada ao final da segmentação, desse modo serve como um contraponto para este trabalho.

Luo et al. (2025) apresenta *Arena Mechanism (ARM)*, uma abordagem baseada em teoria dos jogos para melhorar a segmentação de imagens médicas utilizando o modelo *nnU-Net*. Enquanto o *nnU-Net* já se destaca por sua capacidade de automação e adaptação, o ARM busca aprimorar as camadas convolucionais com três etapas principais: *Recruit* (seleção de módulos como *SCConv* e *ConvNeXt* como "desafiantes"), *Train* (ajuste de pesos via decomposição SVD) e *Fight* (fusão colaborativa com funções de perda e recompensa). Experimentos demonstraram que o ARM alcança resultados melhores em relação a arquiteturas como *MedNeXtL*, *STU-Net*, *SwinUNETRV2* e *CoTr*. Este trabalho busca otimização de convoluções, e sua integração estruturada de módulos do ARM podem inspirar melhorias no *U-Mamba*.

4. Materiais e Métodos

A pesquisa realizada neste trabalho é classificada como quantitativa e explicativa. Em primeiro lugar, o caráter explicativo é justificado pelo fato deste estudo buscar as causas e explicações dos dados observados, isto é, os fatores determinantes desses dados [Wazlawick 2009]. Em segundo lugar, o caráter quantitativo tem origem nas métricas avaliadas, DSC e NSD, as quais são representadas por valores numéricos que serão usadas para medir a qualidade da imagem resultante do processo.

As subseções a seguir são organizadas da seguinte forma: a Seção 4.1 apresenta e justifica os métodos utilizados durante a pesquisa, a Seção 4.2 descreve os instrumentos utilizados, a Seção 4.3 descreve os métodos de coleta e a Seção 4.4 apresenta as métricas avaliadas no estudo.

4.1. Métodos

A principal contribuição deste trabalho foi a introdução de um módulo de memória no bloco de construção do *U-Mamba*, como mostrado na Figura 3b. Este módulo é inspirado no mecanismo de atualização de memória detalhado por Cardoso et al (2021). Essa implementação visa aumentar a capacidade do modelo de reter e utilizar informações contextuais relevantes de entradas anteriores durante o processamento sequencial realizado pelo módulo *Mamba*.

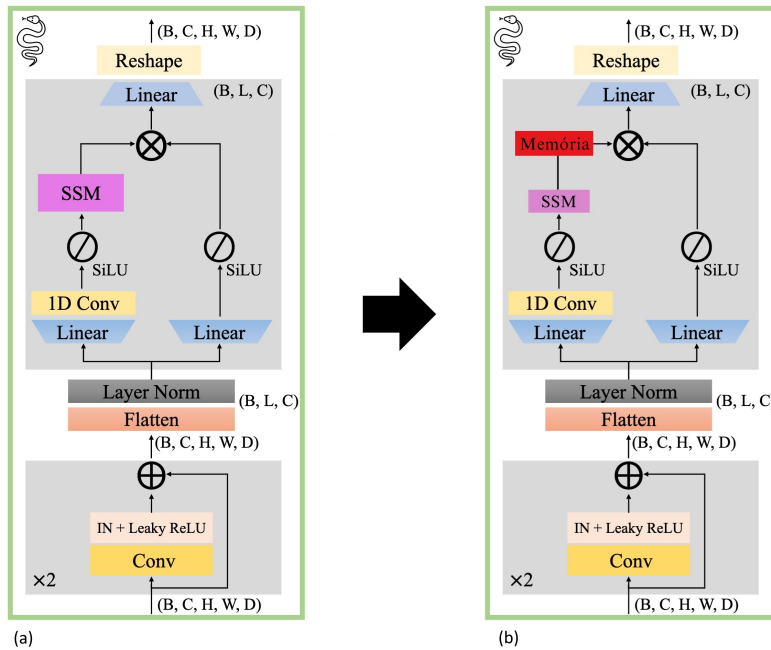


Figura 3: Mudança do modelo em nível arquitetural
[Ma et al. 2024]

Como mostrado na Figura 4, esse módulo possui duas entradas, a *Ht*, proveniente do processamento atual do bloco *U-Mamba*, e a entrada *Mt-1*, proveniente do processamento do bloco *U-Mamba* anterior. Ambas passam pelo bloco *Multi - Head Attention* responsável por avaliar as informações de *Ht* em relação a *Mt-1*, e gerando um contexto *St*, esta que gera um estado de memória *Ct* e também passa por um bloco de *Squeeze-and-excitation (SE)* responsável por reponderar a importância das informações e gerar o estado

de memória Z_t . Ao final é gerado uma combinação ponderada resultando na informação M_t , permitindo que a arquitetura não apenas capture dependências de longo alcance, mas que também diminua a redundância e mantenha um histórico contextual melhor.

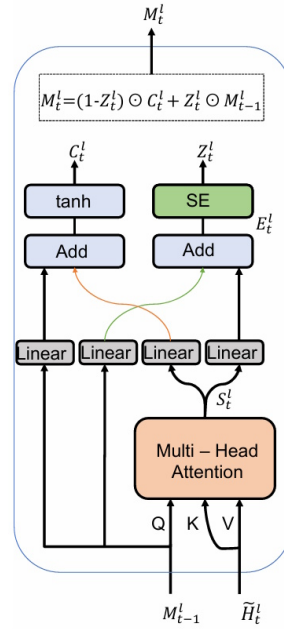


Figura 4: Módulo de memória inserido no bloco *U-Mamba* [Cardoso et al. 2021]

Os seguintes passos ilustrados na Figura 5 foram executados para experimentar o novo modelo proposto integrando o módulo de memória. Em primeiro lugar, implementou-se o módulo de memória utilizando o *PyTorch*¹. Em seguida, variou-se em qual posição do bloco *U-Mamba* o módulo de memória produziria os melhores resultados. O terceiro passo foi executar o modelo modificado repetidas vezes para extrair a média e desvio padrão, pois este modelo não é determinístico e, desse modo, os resultados apresentam variações. Por fim, foram feitas análises dos resultados e a comparação dos resultados com modelos da literatura atual os quais também utilizam SSMs e o mesmo *dataset*.

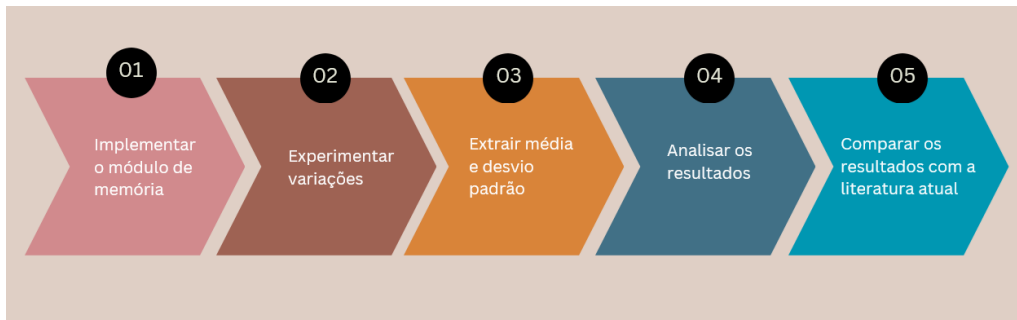


Figura 5: Metodologia do trabalho

¹O código de implementação do módulo de memória está disponível no repositório do trabalho cujo link de acesso encontra-se disponível no final deste texto, na seção Pacote de Replicação

4.2. Instrumentos

Para a condução desta pesquisa, os principais instrumentos utilizados foram: o *Visual Studio Code* (VSCode) para escrever e gerenciar todo o código do projeto. No que tange às bibliotecas de software, foi usado o *PyTorch* para a construção e o treinamento das redes neurais.

O novo modelo *U-Mamba* com inclusão de memória foi executado localmente em uma máquina com o sistema operacional Linux, distribuição Ubuntu na versão 20.04 e com as seguintes especificações: processador Intel(R) Core(TM) i5-14600KF 3.50 GHz, 32GB de memória RAM, 2TB de SSD com uma partição de 300GB para o Linux e placa de vídeo NVIDIA Geforce RTX 4060. Nesse ambiente o *U-Mamba* original também foi executado para realizar as comparações.

4.3. Métodos de coleta

O *dataset* utilizado nesta pesquisa é o mesmo utilizado pelos autores do artigo original que apresenta o *U-Mamba*, isto é, uma modificação da base de dados do *MICCAI 2022 AMOS* [Ji et al. 2022]. O conjunto de dados é formado por imagens de ressonância magnética dos órgãos do abdômen e contém sessenta imagens de treino e vinte imagens de teste (tanto as imagens de treino e de teste possuem suas respectivas anotações -*ground truth*-disponíveis).

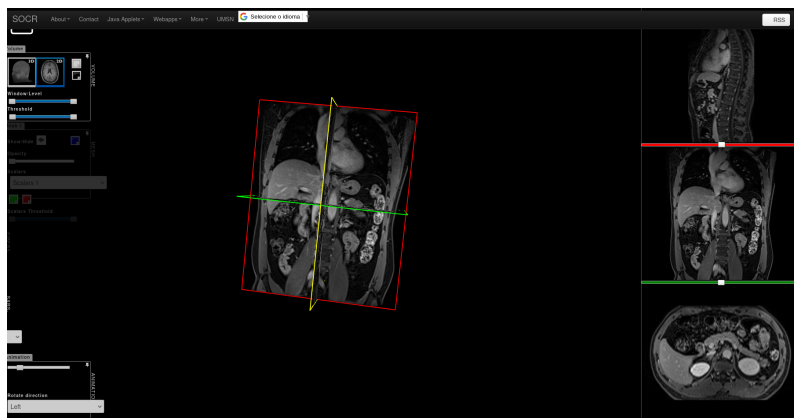


Figura 6: Exemplo de imagem do *dataset*

Como é possível observar na Figura 6, as imagens fornecidas no *dataset* estão na escala de cinza e são salvas no formato *.nii.gz*.²

Os órgãos presentes nas imagens do *dataset* são: fígado, rim direito, baço, pâncreas, aorta, veia cava inferior, glândula adrenal direita, glândula adrenal esquerda, vesícula biliar, esôfago, estômago, duodeno e rim esquerdo. Cada um desses órgãos são etiquetado com números começando com 1 para o fígado e seguindo até 13 para o rim esquerdo.

4.4. Métricas

Para medir a qualidade das imagens de segmentação geradas pelos modelos, comparando a predição (P) com a referência (G), utilizamos o DSC e NSD.

²Para visualizar imagens com esta extensão foi utilizado o site disponível em: <https://socr.umich.edu/HTML5/BrainViewer/>

A primeira métrica mede o quanto a imagem de segmentação prevista se sobrepõe à imagem de referência. Um valor mais próximo de 1 indica melhor qualidade e maior sobreposição. Para calculá-la utiliza-se a fórmula a seguir:

$$DSC(P, G) = \frac{2 \times |P \cap G|}{|P| + |G|}$$

A segunda métrica mede o DSC nos *pixels* da borda da imagem com certa margem de incerteza. Assim como o Coeficiente de Dice, quanto mais próximo de 1 for o resultado do NSD melhor o resultado. O cálculo desta métrica é realizado da seguinte maneira:

$$NSD(P, G) = \frac{|S_P \cap B_G^t| + |S_G \cap B_P^t|}{|S_P| + |S_G|}$$

5. Resultados

Esta seção apresenta os resultados obtidos após a aplicação dos métodos especificados na seção anterior.

5.1. Análise dos resultados

Em primeiro lugar, para encontrar o melhor resultado para o modelo, foram testadas diversas variações na arquitetura do modelo: com memória e *Squeeze and Excitation* antes do SSM, com memória e SE após SSM, somente memória antes do SSM, etc. Todas as variações e respectivos resultados podem ser encontrados na tabela 1. O melhor resultado encontrado foi módulo de memória e o SE depois do bloco SSM, portanto todos os resultados apresentados a seguir se referem a essa variação do modelo.

Tabela 1: Resultados dos testes com variação nos módulos para Dataset de MRI

Varição	DSC	NSD
U-Mamba_Bot 2d	0.7588 ± 0.1051	0.8285 ± 0.1074
Sem Memória	0.7400	0.7907
SE + Memória (antes)	0.7806	0.8421
SE + Memória (depois)	0.7916	0.8543
Memória (depois)	0.7886	0.8493
Memória (antes)	0.7395	0.7872
SE (depois)	0.7745	0.8348
SE (antes)	0.7879	0.8504

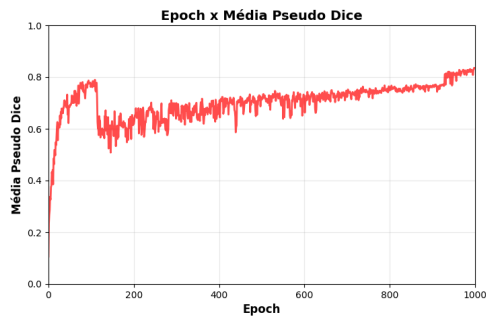
Em segundo lugar, dadas as restrições de *hardware*, foi conduzida uma análise de estabilidade para verificar a reprodutibilidade dos resultados. O modelo de SE + Memória pós SSM no U-Mamba.bot 2d foi treinado mais cinco vezes com *batch size* de 8. *Batch size* é um hiperparâmetro o qual define quantas amostras são processadas simultaneamente em uma época do modelo. A Tabela 2 apresenta os valores brutos de DSC e NSD obtidos. Observa-se que, embora o Melhor Resultado (DSC 0.7916) — utilizado nas comparações com o estado da arte — seja superior à média das cinco execuções

(DSC 0.7791), o baixo desvio padrão (± 0.0042) confirma que a arquitetura apresenta consistência nos resultados.

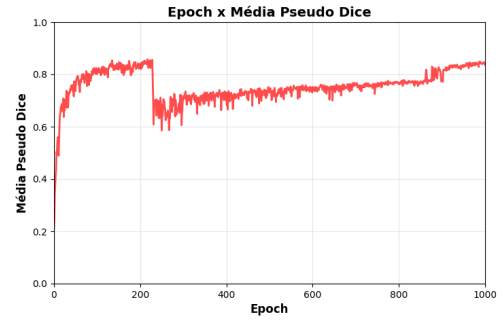
Tabela 2: Análise de estabilidade do modelo (Batch Size = 8) comparada ao melhor resultado reportado.

Execução	DSC	NSD
Run 1	0.7811	0.8457
Run 2	0.7770	0.8422
Run 3	0.7749	0.8412
Run 4	0.7769	0.8407
Run 5	0.7855	0.8461
Média (5 runs)	0.7791	0.8432
Desvio Padrão (5 runs)	0.0042	0.0025
Melhor Resultado	0.7916	0.8543

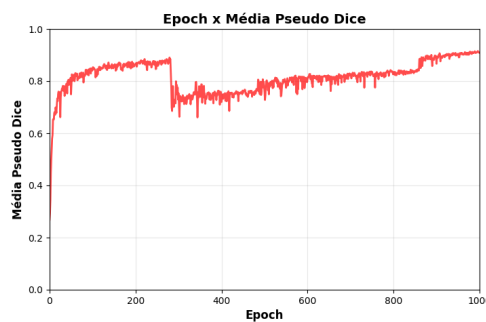
Em terceiro lugar, analisou-se o impacto do *batch size* no resultado do DSC do modelo. Para tal o modelo foi executado com este hiperparâmetro recebendo os valores de 2, 4, 6 e 8 e os gráficos da Figura 7 foram gerados utilizando a biblioteca *matplotlib* do *python*. Ao avaliar os gráficos é possível perceber que quanto maior o valor do *batch size* melhor é o resultado do DSC. Também é possível notar que em todos os tamanhos existe um momento em que o resultado cai drasticamente e durante todo o resto do treinamento o modelo recupera-se desta queda. Contudo, quanto maior o tamanho do lote, mais épocas são necessárias para que o DSC caia e mais rapidamente o modelo se recupera.



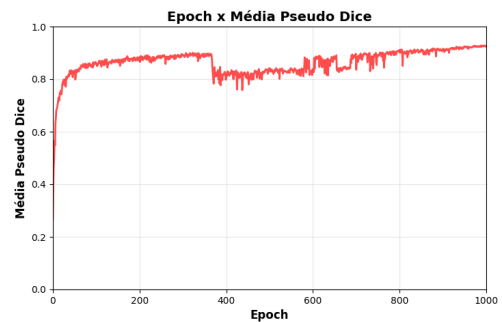
(a) Batch Size 2



(b) Batch Size 4



(c) Batch Size 6



(d) Batch Size 8

Figura 7: Evolução do Dice por época de treino dividido por Batch Size

Para entender a causa da queda abrupta do DSC durante o treino do modelo gerou-se gráficos para a métrica separado pelos órgãos segmentados pelo modelo. Desse modo foi possível encontrar os resultados apresentados na Figura 8 os quais apresentam a evolução do DSC da glândula adrenal direita. A partir desses novos gráficos é possível observar que o que limita o aumento da métrica durante os treino é um órgão chamado glândula adrenal (tanto a esquerda quanto a direita) que demoram mais que os outros órgãos para ser aprendido pelo modelo. Ainda é possível averiguar que, assim como nos resultados da Figura 7 a queda dos valor do DSC das glândulas adrenais é "empurrado para frente" conforme o *batch size* aumenta.

Uma hipótese do porquê esse órgão é tão complexo de segmentar e causa o impacto observado anteriormente é: as glândulas adrenais estão localizadas muito próximas aos rins e tem um formato relativamente semelhante a esses órgãos. Normalmente, o que facilita a identificação é a diferença de cor, contudo as imagens utilizadas para treinar o modelo não apresentam cores, desse modo a classificação da glândula adrenal torna-se muito mais complexa.

Por fim, a Tabela 3 detalha os resultados obtidos pelo modelo desenvolvido para esse artigo de forma comparativa a outros modelo que também utilizam o *SSM*. Os resultados apresentados mostram que o modelo proposto³ possui resultados comparáveis e até melhores que o da literatura atual: 79.16 para *DSC* e 85.54 para o *NSD*.

³Nome sujeito a mudanças

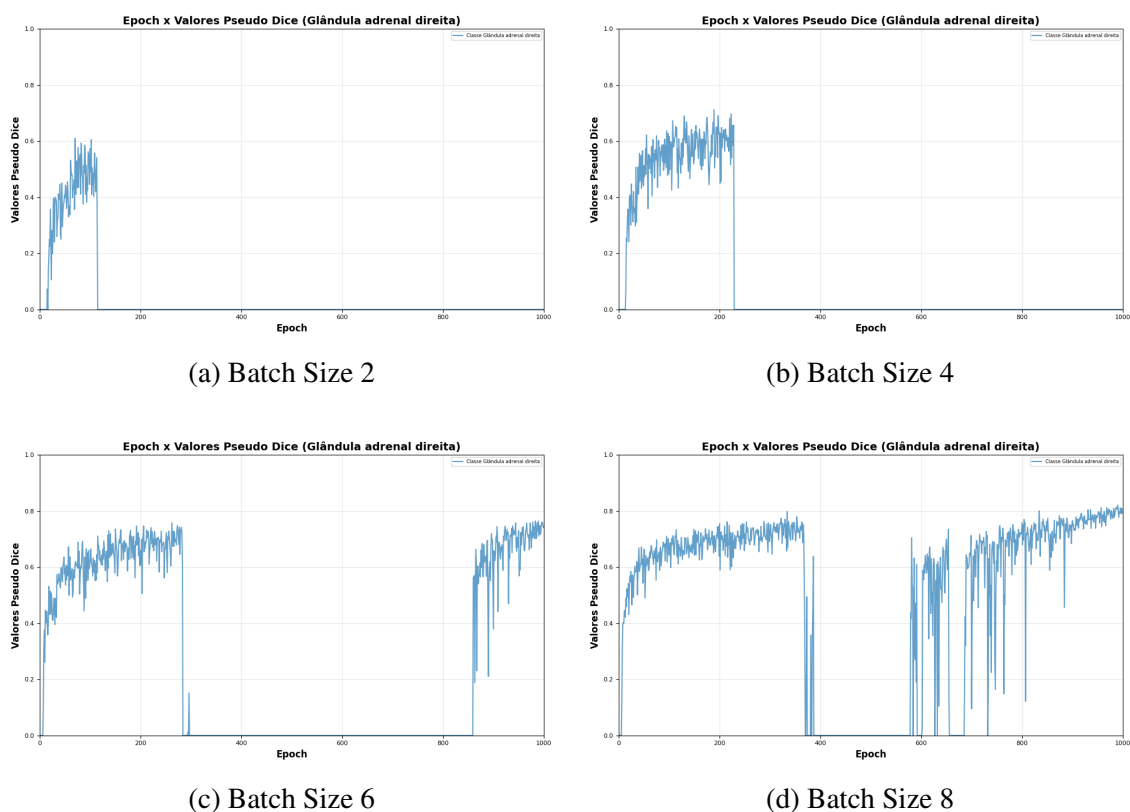


Figura 8: Evolução do DSC da glândula adrenal direita

Tabela 3: Resumo dos resultados da segmentação 2D: segmentação de órgãos do abdômen em imagens de ressonância magnética

Modelo	Dataset	DSC	NSD	F1
U-Mamba_Bot	2D Abdomen MRI	75.88±10.51	82.85±10.74	-
U-Mamba_Enc	2D Abdomen MRI	76.25±10.82	83.27±10.87	-
LKM-UNet	2D Abdomen MRI	77.35	83.80	-
Swin-UMamba	2D Abdomen MRI	77.60	84.21	-
Swin-UMamba (Decoder)	2D Abdomen MRI	77.05	83.76	-
Modelo Proposto	2D Abdomen MRI	79.16	85.54	-

6. Ameaças à validade

Validade de Construção: Como evidenciado na análise por órgãos, a métrica DSC penaliza erros em estruturas de pequena escala e morfologia complexa, como as glândulas adrenais, onde um desvio de poucos *voxels* resulta em quedas que podem não refletir a utilidade clínica real. Embora a incorporação da NSD mitigue parcialmente a sensibilidade a erros de borda, a ausência de um estudo de usuário formal com radiologistas para validar a explicabilidade visual e a relevância clínica das segmentações introduz um viés de subjetividade.

Validade Interna: a relação entre a introdução do módulo de memória e o ganho de desempenho é suscetível a variáveis de confusão introduzidas pelas restrições de *hardware*. O treinamento foi limitado a um *batch size* de 8 devido à capacidade da GPU

(RTX 4060), enquanto os experimentos de sensibilidade conduzidos neste estudo (Figura 7) indicam que arquiteturas baseadas em Mamba se beneficiam de lotes maiores para a estabilização dos gradientes. Essa restrição técnica atua como um fator limitante que pode ter impedido o modelo de alcançar seu potencial máximo.

Validade Externa: a validade é restringida pelo escopo do *dataset* AMOS 2022 e pela natureza da modalidade de imagem. O estudo concentrou-se exclusivamente em MRI, cujas intensidades de sinal não são padronizadas como no CT. Diferentemente do CT, onde os valores de *pixel* (Unidades Hounsfield) têm significado físico absoluto, as intensidades na MRI são relativas e variam significativamente dependendo dos parâmetros de aquisição e do fabricante do *scanner* [Nyul and Udupa 2000], o que pode restringir a extrapolação direta dos resultados obtidos para cenários que utilizem protocolos de imagem diferentes dos presentes neste conjunto de dados.

Validade de Conclusão: Os valores finais reportados (DSC de 0.7916) referem-se à melhor execução (*best run*) para demonstrar o potencial máximo da arquitetura, porém, a análise de estabilidade com cinco repetições (Tabela 2) revelou uma média inferior (0.7791 ± 0.0042), indicando que o pico de desempenho depende de uma inicialização favorável dos pesos. Somado ao tamanho reduzido do conjunto de teste (20 volumes), existe o risco de que a superioridade observada seja superestimada, sem testes de hipótese formais em uma amostra maior, não se pode descartar categoricamente que as melhorias marginais sejam fruto da variabilidade aleatória inerente ao treinamento de redes neurais profundas.

7. Conclusão

Este estudo investigou o impacto da implementação de um módulo de memória no *U-Mamba*, um modelo de segmentação de imagens que introduz o uso de SSMs nesse contexto, e como esse novo componente arquitetural pode influenciar os resultados obtidos pelo modelo em um *dataset* composto por imagens bidimensionais de ressonâncias magnéticas da região do abdômen. Para medir os resultados obtidos pelo modelo, foram utilizadas as métricas DSC e NSD.

Os resultados revelaram que as alterações propostas neste estudo apresentam resultados equiparáveis, quando não melhores, ao modelo original e também a outras modificações existentes na literatura. Em geral, o modelo desenvolvido durante esta pesquisa apresentou um DSC de 79,16 e um NSD de 85,54, superando tanto a variação *U-Mamba-Bot* (com DSC médio de 75,88 e NSD médio de 82,85) quanto a variação *U-Mamba-Enc* (com DSC médio de 76,25 e NSD médio de 83,27).

Além disso, a análise de estabilidade com cinco execuções demonstrou que o modelo proposto mantém consistência em seus resultados, com baixo desvio padrão ($\pm 0,0042$ para DSC), reforçando a confiabilidade da arquitetura mesmo sob restrições de *hardware*. A investigação do impacto do *batch size* e da segmentação por órgãos evidenciou que estruturas de menor escala e maior complexidade morfológica, como as glândulas adrenais, representam um desafio significativo para o modelo, influenciando diretamente a evolução do DSC durante o treinamento.

Em síntese, a introdução de um módulo de memória pós-SSM mostrou-se uma estratégia eficaz para melhorar a capacidade do modelo de capturar dependências de

longo alcance e manter contexto sequencial, resultando em ganhos mensuráveis de desempenho. Trabalhos futuros podem explorar a aplicação desse módulo em outras modalidades de imagem, como tomografia computadorizada, bem como a realização de validações clínicas com especialistas para avaliar a explicabilidade e utilidade prática das segmentações geradas.

Pacote de Replicação

O pacote de replicação deste trabalho encontra-se disponível em:

<https://github.com/ICEI-PUC-Minas-PPLES-TI/plf-es-2025-1-tcci-0393100-pes-carlos-emanuel-oliveira-e-gustavo-alves>

Referências

- Armand, T. P. T., Bhattacharjee, S., Choi, H.-K., and Kim, H.-C. (2024). Transformers effectiveness in medical image segmentation: A comparative analysis of unet-based architectures. In *2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, pages 238–242.
- Brice, C. R. and Fennema, C. L. (1970). Scene analysis using regions. *Artificial intelligence*, 1(3-4):205–226.
- Cardoso, L. V., Guimarães, S. J. F., and Patrocínio, Z. K. G. (2021). Enhanced-memory transformer for coherent paragraph video captioning. In *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 836–840.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>.
- Gopigari, N. and Singh, T. (2022). Comparision of segmentation techniques for acute lymphoblastic leukemia in leukemia cancer. In *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–5.
- Gu, A. and Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*.
- Ji, Y., Bai, H., Ge, C., Yang, J., Zhu, Y., Zhang, R., Li, Z., Zhanng, L., Ma, W., Wan, X., et al. (2022). Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. *Advances in neural information processing systems*, 35:36722–36732.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521:436–44.
- Luo, H., Guan, C., Shao, T., Li, S., Kishi, T., and Yoshie, O. (2025). Arm : nnu-net with arena mechanism for medical image segmentation. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5.
- Ma, J., Li, F., and Wang, B. (2024). U-mamba: Enhancing long-range dependency for biomedical image segmentation.
- Müller, D., Soto-Rey, I., and Kramer, F. (2022). Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, 15(1):210.

- Nyul, L. G. and Udupa, J. K. (2000). Standardizing the MR image intensity scales: making MR intensities have tissue-specific meaning. In Mun, S. K., editor, *Medical Imaging 2000: Image Display and Visualization*, volume 3976 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 496–504.
- Patel, S. (2023). An overview and application of deep convolutional neural networks for medical image segmentation. In *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, pages 722–728.
- Rayed, M. E., Islam, S. S., Niha, S. I., Jim, J. R., Kabir, M. M., and Mridha, M. (2024). Deep learning for medical image segmentation: State-of-the-art advancements and challenges. *Informatics in Medicine Unlocked*, page 101504.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation.
- Salvatierra-Guillermo, D. P. and Zavaleta Jave, V. O. (2024). Comparative assessment of samm and monai algorithms in spleen image segmentation: Approaches and metrics. In *2024 IEEE ANDESCON*, pages 1–6.
- Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):210–229.
- Szeliski, R. (2022). *Computer vision: algorithms and applications*. Springer Nature.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2023). Attention is all you need.
- Wazlawick, R. S. (2009). *Metodologia de pesquisa para ciência da computação*, volume 2. Elsevier Rio de Janeiro.
- Zhang, A., Lipton, Z. C., Li, M., and Smola, A. J. (2023). *Dive into Deep Learning*. Cambridge University Press. <https://D2L.ai>.