

EL-nnU-Net: Early Learning para Segmentação 3D do Hipocampo em Ressonância Magnética

Patrick Ryan Sales dos Santos¹, José Denes Lima Araújo²,
Antonio Oseas de Carvalho Filho¹

¹Universidade Federal do Maranhão (UFMA) – Programa de Pós-Graduação
Doutorado em Ciência da Computação - Associação (UFMA/UFPI)

²Universidade Federal do Piauí (UFPI) - Campus CSHNB
Laboratório de Pesquisas Aplicadas a Visão e Inteligência Computacional

{sales, jose.denes, antoniooseas}@ufpi.edu.br

Abstract. *Automatic hippocampus segmentation in magnetic resonance imaging is an important task for supporting the diagnosis of several neurological diseases. This work proposes EL-nnU-Net, which employs an Early Learning strategy based on self-distillation. The method explores intermediate feature maps through the Early Decoder, enabling the generation of auxiliary masks that help refine the network training process. The experiments were conducted using the Medical Segmentation Decathlon dataset, composed of 260 volumes. The proposed approach achieves competitive performance, obtaining the best result for the Anterior class and performance equivalent to the state of the art for the Posterior class.*

Resumo. *A segmentação automática do hipocampo em exames de ressonância magnética é uma tarefa importante para o apoio ao diagnóstico de diversas doenças neurológicas. Este trabalho propõe a EL-nnU-Net, que utiliza estratégia de Early Learning baseada em autodestilação. O método explora mapas intermediários por meio do Early Decoder, permitindo a geração de máscaras auxiliares que auxiliam no refinamento do treinamento da rede. Os experimentos foram conduzidos com o conjunto de dados do Medical Segmentation Decathlon, composto por 260 volumes. A abordagem alcança desempenho competitivo, obtendo o melhor resultado para a classe Anterior e desempenho equivalente ao estado da arte para a classe Posterior.*

1. Introdução

O hipocampo (HC) é uma estrutura cerebral que apresenta um formato arqueado e alongado, localizada nos dois lados do lobo temporal. Essa região exerce papel importante nos processos de aprendizagem, na construção e consolidação da memória, além de contribuir para a orientação espacial. O HC também tem um forte envolvimento na formação de novas informações, consolidação da memória de longo prazo e no suporte a memórias declarativas, episódicas e espaciais [Albadawi 2025].

A forma e o tamanho do HC estão inteiramente relacionados com um funcionamento adequado do cérebro. Alterações estruturais nessa região, como volume, espessura, simetria ou até mesmo integridade celular, têm sido amplamente associadas ao desenvolvimento e à progressão de diferentes doenças neurológicas e psiquiátricas, como Epilepsia

[Khan et al. 2025], Esquizofrenia [Swathi and Sankaran 2025] e Doença de Alzheimer [Bonarota et al. 2025].

A literatura apresenta duas maneiras manuais de determinar se o HC está sofrendo algum tipo de deformidade, sendo a primeira delas um método que envolve a subjetividade do especialista experiente, num processo demorado que pode levar até três horas para um único exame de Imagem de Ressonância Magnética (IRM). O segundo, auxiliado por alguns dispositivos eletrônicos, exige que o operador defina limiares, pontos iniciais (*seeds*) e contornos iniciais com base na experiência [Arabi and Zaidi 2024].

Diante disso, diversos esforços têm sido direcionados ao desenvolvimento de uma ferramenta automática para a segmentação do HC em exames de IRM, com o objetivo de auxiliar médicos especialistas em uma tarefa tão crucial e delicada. A proposta não é substituir o profissional, mas oferecer um suporte ao diagnóstico, contribuindo para maior agilidade, precisão e confiabilidade nos resultados [Isensee et al. 2021].

O presente trabalho tem como objetivo propor uma nova abordagem baseada em aprendizado profundo para a segmentação de HC em exames de IRM, no contexto de segmentação multiclasse. Especificamente, busca-se classificar cada voxel do exame em uma das seguintes categorias: Fundo, HC Anterior e HC Posterior. Sendo gerada assim uma máscara de segmentação que poderá ser utilizada pelo especialista para avaliar possíveis anormalidades no volume do HC.

Nesta abordagem, foi projetada uma nova rede neural baseada na nnU-Net, amplamente empregada em tarefas semelhantes de segmentação de órgãos humanos, incluindo, de forma específica, o HC. Contudo, foram realizadas modificações estruturais na arquitetura original, além do desenvolvimento de uma nova estratégia de treinamento para redes neurais profundas, denominada *Early Learning* (EL). As contribuições deste artigo são as seguintes:

- Uma rede 3D de segmentação que utiliza o módulo *Early Decoder* para explorar mapas intermediários da rede na construção das máscaras.
- Uma estratégia de treinamento baseada em *knowledge distillation* para otimização de redes neurais.

2. Trabalhos Relacionados

O trabalho de [Çiçek et al. 2016] estendeu a arquitetura U-Net para dados volumétricos, substituindo as convoluções bidimensionais por tridimensionais, o que permite explorar diretamente as informações locais ao longo do volume. Na segmentação 3D do HC, essa característica apresenta-se relevante, pois a estrutura do HC é constituída de forma contínua entre os *slices*. Entretanto, sua capacidade de modelar dependências espaciais globais continua restrita, o que pode limitar a consistência global ao longo do volume.

A nnU-Net consolidou-se como referência em segmentação de imagens médicas 2D e 3D. Os autores [Isensee et al. 2021] propuseram um *framework* autoconfigurável que adapta pré-processamento, arquitetura e hiperparâmetros às características do conjunto de dados. Contudo, um *framework* que se autoconfigura para imagens peculiarmente pequenas e com baixo contraste entre o HC e o fundo pode apresentar limitações.

[Hatamizadeh et al. 2022] Introduziu a UNETR, uma arquitetura onde o codificador é baseado em *Transformer*, substituindo o *backbone* convolucional tradicional por

um modelo do tipo *Vision Transformer*. Essa estratégia permite que a UNETR capture dependências globais entre diferentes regiões do volume. Entretanto, por apresentar um número mais elevado de parâmetros quando comparada a arquiteturas puramente convolucionais, a UNETR tende a exigir uma quantidade maior de dados para treinamento.

Ademais, [Hatamizadeh et al. 2021] introduziram o Swin-UNETR, que combina o conceito do UNETR com o *Swin Transformer*, que utiliza mecanismos de atenção baseados em janelas deslizantes. Essa estratégia reduz o custo computacional e melhora a escalabilidade, ao mesmo tempo em que preserva a capacidade de aprendizado global entre os *slices*. Todavia, a complexidade estrutural pode exigir estratégias adicionais de regularização para evitar sobreajuste em conjuntos de dados limitados.

Proposto por [Xie et al. 2021], a rede CoTr apresenta a inserção seletiva de módulos *Transformer* em arquiteturas convolucionais, substituindo determinadas operações de convolução por mecanismos de atenção contextual. Essa estratégia busca combinar os benefícios das convoluções com a capacidade dos *Transformers* de capturar dependências globais, promovendo um equilíbrio entre informações contextuais locais e globais. Entretanto, essa combinação pode se tornar um desafio, especialmente em cenários com conjuntos de dados limitados.

O TransBTS, introduzido por [Wang et al. 2021], adota uma abordagem híbrida ao combinar camadas convolucionais com módulos *Transformer*. Entretanto, assim como outras arquiteturas híbridas e mais complexas, o TransBTS pode enfrentar dificuldades quando treinado com conjuntos de dados reduzidos, devido à sua maior demanda por amostras para um aprendizado estável e generalizável.

[Liu et al. 2022] apresentou a PHTrans, arquitetura híbrida que integra convoluções e *Transformers*, possibilitando a extração de características locais e globais. Essa característica pode ser vantajosa para a segmentação do HC, cuja morfologia demanda uma análise multiescala para uma delimitação mais precisa. Ainda assim, embora essa modelagem contribua para uma representação mais rica, ela não elimina o risco de sobreajuste quando aplicada a conjuntos de dados com número reduzido de amostras.

Considerando as limitações discutidas nos trabalhos relacionados, principalmente no que se refere à segmentação de estruturas pequenas em relação ao tamanho da IRM em bases de dados com número de amostras reduzido, o presente trabalho propõe uma investigação de estratégias voltadas a estabilizar o processo de treinamento e a melhorar a generalização na segmentação tridimensional do HC. Por intermédio da EL-nnU-Net, os problemas descritos nos trabalhos relacionados podem ser mitigados, usando mecanismos baseados em autodestilação para aprimorar tanto a qualidade da segmentação quanto a estabilização do treinamento em conjuntos de dados com número reduzido de amostras, além de criar sub-redes que necessitam de menor poder computacional para inferência.

3. Materiais e Método

Esta seção descreve os materiais e métodos utilizados neste trabalho. Inicialmente, é apresentado o conjunto de dados empregado nos experimentos. Em seguida, são discutidos os fundamentos da EL-U-Net e da nnU-Net, culminando na descrição da arquitetura proposta, EL-nnU-Net, de seus principais componentes e do procedimento de treinamento. Por fim, são apresentadas as métricas adotadas para a avaliação do desempenho do modelo.

3.1. Conjunto de Dados

Os experimentos foram conduzidos utilizando um conjunto de dados do *Medical Segmentation Decathlon* [Antonelli et al. 2022], disponíveis publicamente, composto por 260 volumes, incluindo indivíduos saudáveis e doentes. As anotações disponíveis estão divididas em duas classes anatômicas, Anterior e Posterior. Considerando o número relativamente limitado de amostras, foi adotada validação cruzada com cinco *folds*, visando obter estimativas mais confiáveis do desempenho do modelo em questão. Todos os volumes foram padronizados para a resolução espacial de $[40 \times 56 \times 40]$.

3.2. Early Learning U-Net

Early Learning U-Net (EL U-Net) é uma extensão da U-Net proposta por [dos Santos et al. 2025], fundamentada no conceito de destilação de conhecimento, mais especificamente na estratégia de autodestilação. A literatura mostra diversas abordagens clássicas de destilação, nas quais uma rede professora de maior complexidade computacional transfere o conhecimento para uma rede estudante com menor complexidade computacional. De forma diferente, a autodestilação acontece internamente à própria arquitetura. Nesse caso, diferentes partes do mesmo modelo passam a atuar simultaneamente como professora e estudante, promovendo um processo de aprendizado colaborativo [Mansourian et al. 2025].

Na EL U-Net, os resíduos intermediários extraídos em cada nível do codificador são utilizados para a geração de máscaras de segmentação auxiliares por meio dos blocos de *Early Decoder* (ED). Cada um desses blocos tem a capacidade de produzir uma máscara de segmentação completa e funcional de forma independente da parte do decodificador da rede principal, o que torna possível a existência de múltiplas sub-redes internas dentro da rede U-Net. Em vez de depender apenas da saída tradicional do decodificador da U-Net, o modelo passa a ter diversas previsões ao longo da hierarquia de características [dos Santos et al. 2025].

A literatura aponta que a utilização da autodestilação traz benefícios nesses cenários, sendo a criação de sub-redes menores e mais eficientes uma delas. Além disso, estudos destacam que a utilização dessa técnica melhora a capacidade de generalização, resultado da regularização introduzida pela supervisão das camadas de ED. Elas atua como uma forma de suavização do aprendizado, reduzindo, assim, o risco de sobreajuste do conjunto de treinamento, especialmente em bases de dados com número reduzido de amostras de treinamento [Mansourian et al. 2025].

3.3. nnU-Net

A nnU-Net foi proposta por [Isensee et al. 2021] com a finalidade de automatizar a configuração de modelos de segmentação de exames médicos. Diferentemente de abordagens tradicionais, que necessitam de ajustes manuais de hiperparâmetros e arquitetura para cada conjunto de dados, a nnU-Net analisa as informações do conjunto de dados, como resolução e número de classes, e define estratégias customizadas para cada conjunto de dados, como etapas de pré-processamento, arquitetura e hiperparâmetros de treinamento [Isensee et al. 2021].

A metodologia proposta foi desenhada para imagens médicas tridimensionais. Entretanto, também oferece suporte para imagens bidimensionais. Em ambos os casos,

a rede se adapta dinamicamente a profundidade da rede, ao tamanho dos *patches* e ao número de filtros convolucionais, conforme as características das amostras, facilitando, assim, o equilíbrio entre a complexidade da rede e a disponibilidade de recursos computacionais, como memória [Isensee et al. 2021].

3.4. Early Learning nnU-Net

A EL U-Net apresenta um escopo restrito no que tange ao conjunto de dados, pois foi construída no domínio de segmentação de imagens bidimensionais de histopatologia e, por decorrência, existe uma forte dependência de alterações para extrapolar para outros domínios e conjuntos de dados. Para superar essa limitação, é proposta, como forma de evolução, a EL-nnU-Net, incorporando ao *framework* da nnU-Net o mecanismo de EL autoconfigurável, permitindo a flexibilidade existente e aumentando as possibilidades de segmentação com as máscaras auxiliares.

A Figura 1 mostra o esquema da arquitetura proposta, organizada, de maneira simplificada, em três componentes principais. O primeiro corresponde ao codificador, responsável por extrair representações mais abstratas do volume de entrada, capturando características cada vez mais complexas ao longo da hierarquia de níveis. O segundo componente é composto pelo decodificador, onde tem como função reconstruir o espaço latente produzido, gerando representações mais próximas da máscara de segmentação. Por fim, a terceira parte, denominada ED, recebe os resíduos intermediários gerados em diferentes níveis do codificador e os utiliza para produzir uma representação auxiliar da máscara final funcional.

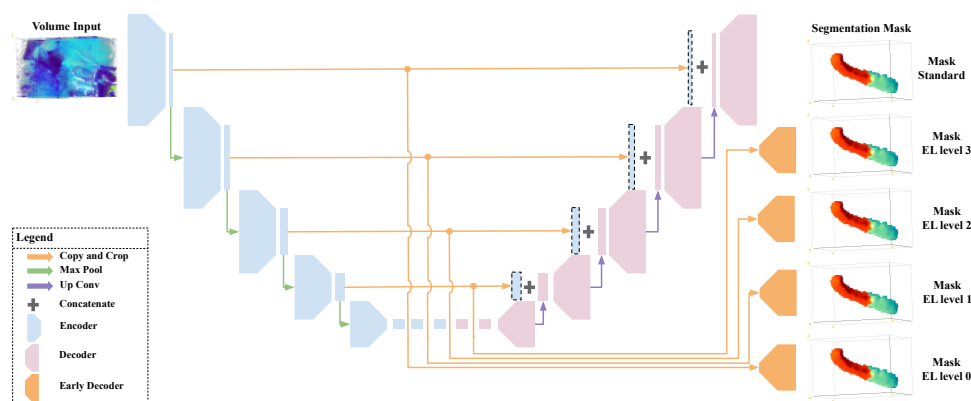


Figura 1. Arquitetura EL-nnU-Net. Esquema ilustrativo da geração das máscaras EL a partir dos mapas residuais extraídos das camadas do encoder.

Durante o treinamento, além da máscara tradicional gerada pelo nnU-Net, mais quatro máscaras são geradas também, denominadas máscaras de nível um a três. Cada máscara passa pelo cálculo de erro, fornecendo, assim, um valor que pode ser usado para o ajuste dos parâmetros da rede. Porém, é preciso definir o quanto cada determinada máscara pode influenciar no ajuste geral da rede. Para que isso seja controlado, é introduzido uma variável de peso no cálculo do erro, denominada *Early Learning Decoder Weight (ELD Weight)*, que pode variar entre 0,0 e 1,0, ou seja, nenhuma influência e influência total. O critério de erro utilizado consiste na combinação de três funções de perda, Cross-Entropy (CE) definida pela Equação 1, Dice Loss dada pela Equação 2 e Focal Loss pela Equação 3.

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

$$L_{Dice} = 1 - \frac{2 \sum_{i=1}^N p_i y_i + \epsilon}{\sum_{i=1}^N p_i + \sum_{i=1}^N y_i + \epsilon} \quad (2)$$

$$L_{Focal} = -\frac{1}{N} \sum_{i=1}^N (1 - p_i)^\gamma [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (3)$$

A perda associada à saída principal da rede é formulada pela Equação 4. Para cada saída auxiliar k proveniente do ED, define-se pela Equação 5. E o critério global do EL é então dado pela Equação 6.

$$L_{main} = 2 L_{CE}(O_{main}, T) + L_{Dice}(O_{main}, T) + L_{Focal}(O_{main}, T) \quad (4)$$

$$L_{aux}^{(k)} = (L_{CE}(O_k, T) + L_{Dice}(O_k, T) + L_{Focal}(O_k, T)) \cdot ELD_w^{(k)} \quad (5)$$

$$L_{total} = 2 L_{main} + \sum_k L_{aux}^{(k)} \quad (6)$$

3.4.1. Early Decoder

O ED é responsável por transformar os resíduos intermediários das camadas do codificador em máscaras auxiliares. No trabalho de [dos Santos et al. 2025], eram usados sete níveis de ED. Entretanto, na arquitetura proposta pelo presente trabalho, foram definidos quatro níveis, mudança essa motivada pelo custo computacional que camadas 3D necessitam para o processamento. A camada de EL foi desenhada para ser simples intencionalmente. O funcionamento interno pode ser visto na Figura 2, onde se encontram quatro estágios, sendo: i) composto por uma convolução 3D; ii) por duas convoluções 3D; iii) por uma função de interpolação; e iv) por duas convoluções 3D. Dessa forma, o bloco fica leve do ponto de vista computacional, não introduzindo acréscimos excessivos de parâmetros para a arquitetura principal.

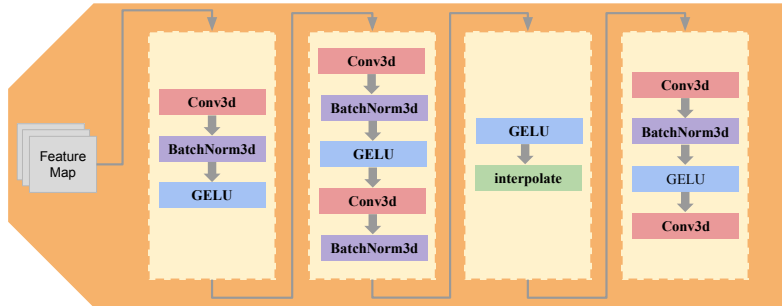


Figura 2. Detalhes do fluxo interno das camadas de ED.

A opção por um ED mais simples foi uma escolha de projeto. Diferentemente do decodificador principal da nnU-Net, ele não tem a função de apenas reconstruir o espaço latente para produzir a máscara final. A ideia aqui é ao limitar a profundidade e

o número de camadas convolucionais o modelo passa a exigir que o codificador entregue representações mais importantes já nos estágios intermediários da rede. Dessa forma um decodificador enxuto consegue entregar máscaras auxiliares funcionais e essa estratégia também favorece o enriquecimento do espaço latente, que é explorado pelo decodificador principal da rede.

3.5. Treinamento da rede

Em todos os experimentos, foram mantidas as configurações originais da nnU-Net, bem como suas estratégias padrão de aumento de dados, com exceção do otimizador. Foi adotado o Adam com *weight decay* de 0.0001 para atualização dos parâmetros. A taxa de aprendizado inicial foi definida em 0.001, utilizando a estratégia *poly learning rate*, na qual a taxa de aprendizado decai progressivamente ao longo do treinamento. O modelo foi treinado por 200 épocas, empregando validação cruzada com cinco *folds*.

3.6. Métricas de Validação

Para avaliar o desempenho da segmentação, foram utilizadas as métricas *Dice Similarity Coefficient* (DSC) e *Intersection over Union* (IoU), amplamente adotadas em tarefas de segmentação médica por quantificarem o grau de sobreposição entre a máscara predita e a máscara de referência.

3.7. Ambiente de Implementação

Os experimentos foram conduzidos a partir da base de código da nnU-Net [Isensee et al. 2021]. A implementação foi realizada em *PyTorch*, e todos os experimentos foram executados em uma única GPU GeForce RTX 5060 Ti.

4. Resultados

Nesta seção, apresentamos os resultados obtidos com a arquitetura proposta, estabelecendo uma comparação com os trabalhos relacionados. A análise considera tanto métricas quantitativas de segmentação quanto o comportamento do modelo sob diferentes configurações. O objetivo é avaliar o impacto da estratégia de EL na precisão final do modelo de segmentação do HC.

4.1. Influência das Camadas Early Decoder

O ELD weight é uma variável que tem como objetivo controlar a influência das máscaras auxiliares geradas em cada nível da rede no cálculo da função de erro. Como foi discutido na Seção 3.4, máscaras geradas em níveis mais superficiais ou mais profundos podem exercer maior ou menor influência na otimização do modelo, dependendo dos valores atribuídos a esses pesos.

Todavia, ajustar manualmente esses hiperparâmetros não é uma tarefa simples, uma vez que variações podem impactar no treinamento e principalmente no desempenho final do modelo. Com o objetivo de contornar esse problema, foi utilizado um mecanismo automático de busca de hiperparâmetros baseado em *Bayesian Optimization*, que permite explorar possíveis combinações de valores. A Figura 3 mostra o gráfico de linhas paralelas com as diferentes combinações testadas e a relação entre esses valores e o desempenho obtido no DSC.

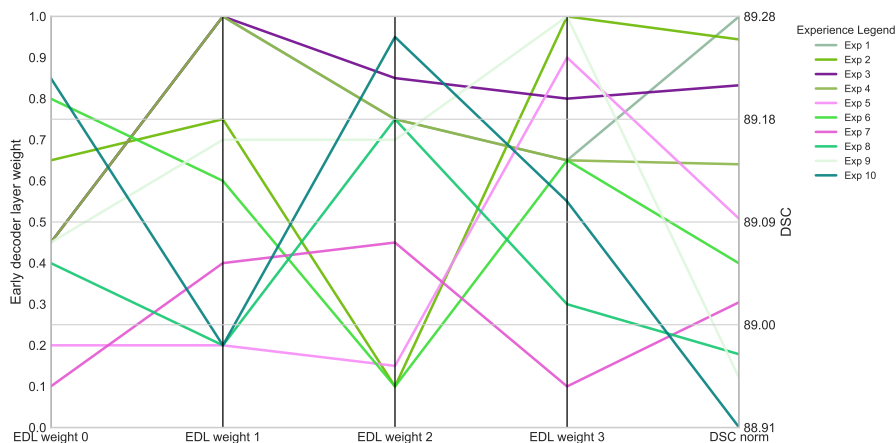


Figura 3. Gráfico de coordenadas paralelas mostrando a influência dos pesos das camadas EDs em relação ao DSC.

A busca dos hiperparâmetros foi realizada com dez iterações com o intuito de encontrar os valores ideais para o presente problema, para assegurar uma comparação justa entre os experimentos executados, todas as dez execuções utilizaram o mesmo conjunto de treino, validação e teste, dessa forma, foi possível analisar de maneira consistente a influência desses hiperparâmetros no desempenho do modelo. Onde foi encontrado os valores de $ELD\ weight\ 0 = 0.45$, $ELD\ weight\ 1 = 1$, $ELD\ weight\ 2 = 0.75$ e $ELD\ weight\ 3 = 0.65$.

4.2. Comparação com os trabalhos relacionados

De posse dos melhores valores de hiperparâmetros obtidos por meio da *Bayesian Optimization*, foi conduzido o processo completo de treinamento, validação e teste para as cinco *folds*, conforme a Figura 4 apresenta. Observa-se que as curvas de treinamento apresentam comportamento estável no decaimento do erro tanto para o conjunto de treinamento quanto para o de validação, não evidenciando assim sinais significativos de sobreajuste. Esse comportamento também se reflete na evolução da métrica DSC, que apresenta tendência crescente e estável ao longo das épocas, resultando nos valores de desempenho de teste apresentados na Tabela 1.

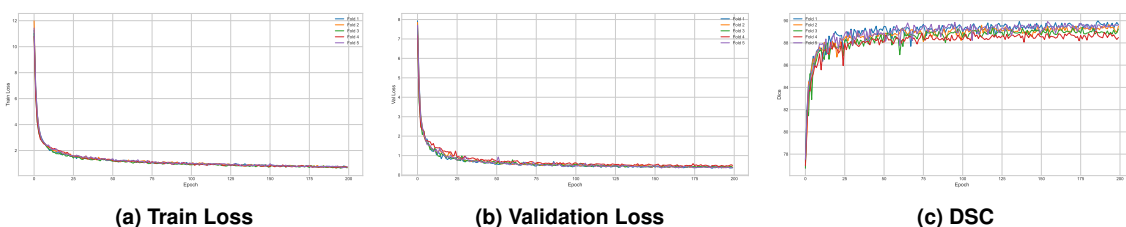


Figura 4. Gráficos de treinamento nas cinco folds.

Com o resultado médio das execuções, foi realizada a comparação com os trabalhos relacionados. Como forma de comparação justa, utilizou-se a mesma divisão de treino, validação e teste das cinco folds para todos os resultados encontrados na Tabela 2. Diante disso, nota-se que o presente trabalho obteve resultados competitivos, alcançando

Fold	DSC		IoU	
	Anterior	Posterior	Anterior	Posterior
1	90,26	88,28	82,40	79,22
2	90,49	89,00	82,82	80,34
3	90,60	88,64	83,00	79,82
4	90,15	88,13	82,28	79,03
5	89,94	88,58	81,82	79,61
Média	90,29 ± 0.23	88,53 ± 0.30	82,46 ± 0.41	79,60 ± 0.46

Tabela 1. Resultados quantitativos obtidos na validação cruzada em cinco folds para a segmentação do HC.

o melhor resultado de DSC para a classe Anterior e alcançando o mesmo resultado da classe Posterior que a nnU-Net, sendo também os melhores resultados. Acredita-se que tal desempenho se dê mediante a técnica de EL proposta, em que as máscaras auxiliares contribuíram para o refinamento dos parâmetros do codificador da rede e, como resultado, para a produção de uma máscara mais refinada no fluxo principal da EL-nnU-Net.

Modelos	Anterior DSC	Posterior DSC
3DU-Net	88,5	86,4
nnU-Net	90,0	88,5
UNETR	83,6	81,1
Swin-UNETR	87,8	85,1
CoTr	88,9	87,1
TransBTS	88,6	87,0
PHTrans	89,4	87,9
EL-nnU-Net	90,3	88,5

Tabela 2. Comparação dos modelos utilizando DSC anterior e posterior.

4.3. Impacto das máscaras de Early Decoder

Na Seção 4.2, foi apresentado o resultado da predição principal da rede, aquela que é produzida pelo fluxo natural da rede nnU-Net. Contudo, ao introduzir a técnica EL, a rede passa a produzir máscaras auxiliares, que empregam a funcionalidade principal de refinar os parâmetros do codificador. Como ganho colateral, é possível também utilizar essas máscaras para a segmentação final, uma vez que apresentam as mesmas propriedades de dimensão e classes que a máscara principal. A Tabela 3 apresenta os resultados de DSC e IoU para as classes Anterior e Posterior nas cinco folds, nos quatro níveis das máscaras.

Nota-se que os valores de média entre os níveis podem apresentar uma variação esperada, uma vez que representam máscaras construídas a partir de ramificações distintas da rede. Para o presente problema, o nível um obteve os melhores resultados na classe Anterior, e o nível dois obteve os melhores resultados na classe Posterior, estando tecnicamente empatados quando levado em consideração o desvio padrão. Acredita-se que tal fato se dê por decorrência dos valores de *ELD weight* escolhidos pela busca *bayesiana*, em que os níveis um e dois tiveram os valores de 1,0 e 0,75, respectivamente.

As predições auxiliares apresentam, de fato, uma qualidade inferior quando comparadas com a predição principal; entretanto, ainda apresentam um resultado promissor, pois obtiveram resultados melhores que alguns trabalhos relacionados, como os casos de [Çiçek et al. 2016], [Hatamizadeh et al. 2022], [Swathi and Sankaran 2025], [Xie et al. 2021] e [Wang et al. 2021]. Sendo que as predições de nível um necessitam

Nível	Fold	DSC Ant.	DSC Post.	IoU Ant.	IoU Post.
0	1	81,63	82,10	69,20	69,80
	2	79,30	81,49	66,62	70,17
	3	79,56	78,82	66,79	66,20
	4	78,99	80,58	65,98	68,17
	5	81,21	83,56	68,72	71,96
	Média	80,14 ± 1,07	81,31 ± 1,58	67,46 ± 1,26	69,26 ± 1,95
1	1	89,56	87,12	81,25	77,37
	2	89,64	87,61	81,40	78,13
	3	89,44	87,01	81,04	77,21
	4	89,52	87,20	81,22	77,49
	5	89,63	88,16	81,30	78,92
	Média	89,56 ± 0,07	87,42 ± 0,42	81,24 ± 0,12	77,82 ± 0,63
2	1	89,45	87,15	81,04	77,40
	2	89,68	88,04	81,47	78,81
	3	89,48	87,29	81,13	77,64
	4	89,30	86,87	80,84	77,03
	5	89,54	87,96	81,15	78,61
	Média	89,49 ± 0,12	87,46 ± 0,46	81,13 ± 0,20	77,90 ± 0,69
3	1	88,97	86,77	80,26	76,78
	2	89,04	87,54	80,42	77,99
	3	89,02	86,93	80,36	77,09
	4	88,80	86,96	80,03	77,14
	5	88,64	87,21	79,67	77,42
	Média	88,89 ± 0,15	87,08 ± 0,27	80,15 ± 0,27	77,28 ± 0,41

Tabela 3. Resultados por nível e fold considerando DSC e IoU para as regiões anterior e posterior.

de menos poder computacional para serem treinadas e geradas; em ambientes com pouca disponibilidade de recursos, podem ser especialmente úteis.

Como estudo de caso, as Figuras 5 e 6 apresentam uma comparação qualitativa entre a predição principal da rede, a predição gerada no nível um e a máscara real segmentada manualmente pelo especialista, correspondendo, respectivamente, ao pior e ao melhor caso observado.

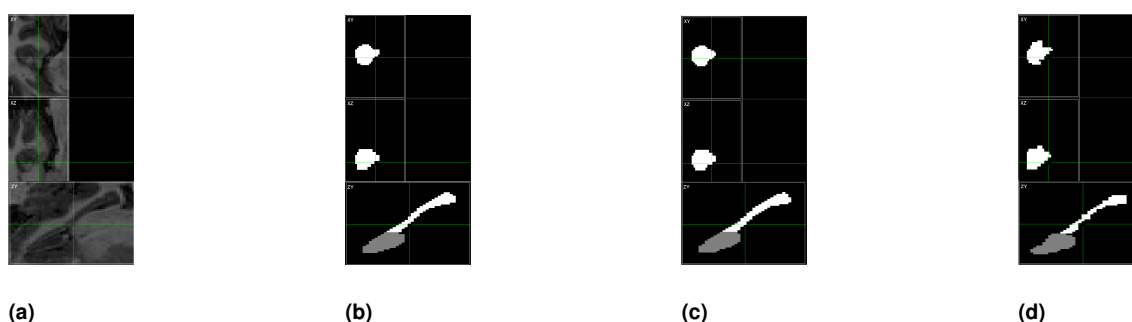


Figura 5. Comparação do caso com pior desempenho no conjunto, volume de entrada (a), predição padrão EL-nnU-Net (b), predição nível 1 e máscara real

Observa-se que, em ambos os cenários, as duas predições geradas pela rede apresentam segmentações coerentes com a região do HC, entretanto, nota-se uma tendência do modelo em produzir máscaras mais externa, correspondendo a áreas maiores do que aquelas delimitadas pelo especialista. Por outro lado, a segmentação manual apresenta contornos mais precisos e restritos a região do HC. Ainda assim, mesmo com essa diferença

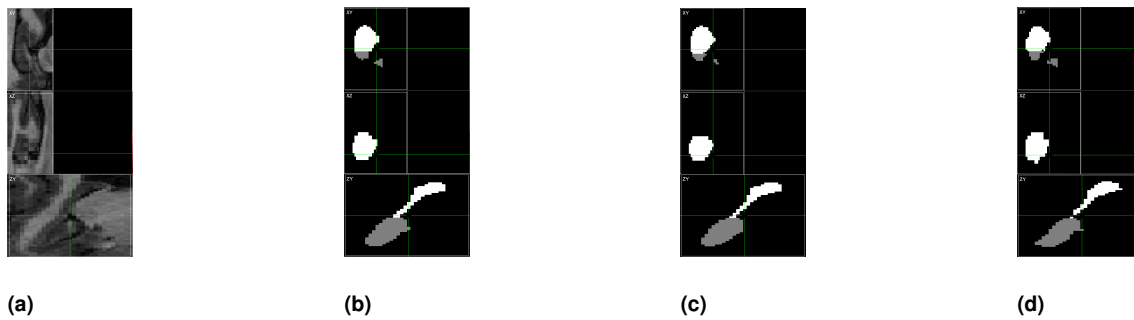


Figura 6. Comparação do caso com melhor desempenho no conjunto, volume de entrada (a), predição padrão EL-nnU-Net (b), predição nível 1 e máscara real

de delimitação, a máscara produzida no nível um mantém uma qualidade competitiva em relação a máscara do especialista.

5. Conclusão

O presente trabalho apresentou a rede EL-nnU-Net, sendo uma extensão da nnU-Net que utiliza a estratégia de EL para explorar mapas intermediários do ED. Como consequência, o método descrito permite a geração de máscaras auxiliares, que utilizam autodestilação para refinar as representações aprendidas pelo codificador. Além disso, a abordagem proposta foi integrada ao *framework* autoconfigurável da nnU-Net, permitindo sua aplicação em diferentes conjuntos de dados e em cenários tridimensionais, ampliando o escopo da EL U-Net original, que possui aplicação restrita a imagens 2D de histopatologia.

Os resultados experimentais demonstraram que a EL-nnU-Net alcançou desempenho competitivo na segmentação do HC, obtendo o melhor resultado para a classe Anterior e desempenho equivalente ao estado da arte para a classe Posterior. além de tudo, a análise das máscaras auxiliares construídas por intermédio dos blocos ED mostrou que essas predições também produzem segmentações competitivas, podendo ser úteis como alternativas de menor custo computacional em cenários com restrições de hardware.

Como trabalhos futuros, pretende-se avaliar a arquitetura proposta em outros conjuntos de dados e em diferentes tarefas de segmentação médica tridimensional. Durante os experimentos realizados, também foi observada uma limitação que ainda merece investigação mais aprofundada. Essa questão está relacionada à definição dos valores de *ELD weight*, que se mostrou bastante sensível ao longo do treinamento e pode influenciar de maneira significativa o desempenho final do modelo.

Referências

- Albadawi, E. A. (2025). Structural and functional changes in the hippocampus induced by environmental exposures. *Neurosciences Journal*, 30(1):5–19.
- Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B. A., Litjens, G., Menze, B., Ronneberger, O., Summers, R. M., et al. (2022). The medical segmentation decathlon. *Nature communications*, 13(1):4128.
- Arabi, H. and Zaidi, H. (2024). Single annotator versus multi-annotator: Challenge of segmenting two neighboring hippocampus head and body with high precision. *Biomedical Signal Processing and Control*, 97:106667.

- Bonarota, S., Caruso, G., Di Domenico, C., Sperati, S., Tamigi, F. M., Giulietti, G., Giove, F., Caltagirone, C., and Serra, L. (2025). Integration of automatic mri segmentation techniques with neuropsychological assessments for early diagnosis and prognosis of alzheimer’s disease. a systematic review. *Neuroimage*, 314:121264.
- Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016). 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer.
- dos Santos, P. V., dos Santos, P. R., Diniz, J. O., and de Carvalho Filho, A. O. (2025). U-net-el: A u-net and early learning-based method for histopathological breast image segmentation. In *Brazilian Conference on Intelligent Systems*, pages 593–606. Springer.
- Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H. R., and Xu, D. (2021). Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In *International MICCAI brainlesion workshop*, pages 272–284. Springer.
- Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H. R., and Xu, D. (2022). Unetr: Transformers for 3d medical image segmentation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 574–584.
- Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., and Maier-Hein, K. H. (2021). nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211.
- Khan, H., Alutaibi, A. I., Tejani, G. G., Sharma, S. K., Khan, A. R., Ahmad, F., and Mousavirad, S. J. (2025). Mri based early temporal lobe epilepsy detection using dgwo based optimized haetn and fuzzy-aal segmentation framework (fasf). *PLoS One*, 20(7):e0325126.
- Liu, W., Tian, T., Xu, W., Yang, H., Pan, X., Yan, S., and Wang, L. (2022). Phtrans: Parallely aggregating global and local representations for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 235–244. Springer.
- Mansourian, A. M., Ahmadi, R., Ghafouri, M., Babaei, A. M., Golezani, E. B., yasamani ghamchi, Z., Ramezani, V., Taherian, A., Dinashi, K., Miri, A., and Kasaei, S. (2025). A comprehensive survey on knowledge distillation. *Transactions on Machine Learning Research*.
- Swathi, N. and Sankaran, K. S. (2025). Advances and clinical implications of neuroanatomical biomarkers in schizophrenia by optimizing brain tissue segmentation. *Biomedical Signal Processing and Control*, 102:107215.
- Wang, W., Chen, C., Ding, M., Yu, H., Zha, S., and Li, J. (2021). Transbts: Multimodal brain tumor segmentation using transformer. In *International conference on medical image computing and computer-assisted intervention*, pages 109–119. Springer.
- Xie, Y., Zhang, J., Shen, C., and Xia, Y. (2021). Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 171–180. Springer.