

Liver Tumor Segmentation in CT Scans Using Deep Learning: A U-Net Approach with Transfer Learning

Marcos L. Alves¹, Murilo C. Salem¹, Daniel H. S. P. Barretos¹, Anderson P. Ferrugem¹

¹Centro de Desenvolvimento Tecnológico - Universidade Federal de Pelotas(UFPel)
Caixa Postal 354 - 96.010-900 - Pelotas - RS - Brasil

Abstract. *This work presents a pipeline for 2D segmentation of the liver and tumors from computed tomography volumes (NIfTI). We used the public Liver Tumor Segmentation dataset, which contains 131 volumes, from which 2D slices were extracted and converted into JPEG images to train a U-Net model with a ResNet50 backbone (transfer learning) implemented using the FastAI library. The preprocessing steps include reading the volumes with NiBabel, applying liver-specific DICOM windows, normalization, and resizing to 128×128. CrossEntropy (multiclass) and Dice Loss were adopted, achieving up to 0.98 foreground accuracy and consistent segmentation of hepatic regions, though small lesions remain challenging.*

Resumo. *Este trabalho apresenta um pipeline para segmentação 2D de fígado e tumores a partir de tomografia computadorizada (NIfTI). Utilizamos o dataset público Liver Tumor Segmentation com 131 volumes, do qual foram extraídas fatias 2D e convertidas em imagens JPEG (com compressão mínima, sem perda perceptível) para treinamento da U-Net com backbone ResNet50 (transfer learning) implementando a biblioteca FastAI. O pré-processamento inclui leitura dos volumes com NiBabel, aplicação de janelas DICOM específicas para fígado, normalização e redimensionamento para 128×128. Como função de perda foi adotada CrossEntropy (multiclasse) e métricas customizadas de acurácia sobre os pixels de interesse (foreground).*

1. Introdução

Tumores hepáticos são um problema de saúde pública, responsáveis por alta mortalidade. O diagnóstico precoce é crucial, e a tomografia computadorizada (TC) é amplamente utilizada [Ronneberger et al. 2015][Milletari et al. 2016]. A segmentação automática do fígado e tumores é promissora [Ronneberger et al. 2015] [Isensee et al. 2021], mas desafiadora devido à baixa diferenciação de contraste, variabilidade anatômica e múltiplos tipos de tumores. Este trabalho propõe um pipeline reproduzível baseado em U-Net com transferência de aprendizado [Howard and Gugger 2020][He et al. 2016] visando avaliar o impacto do pré-treinamento na segmentação hepática em TC, com contribuições em: (i) adaptação de dataset NIfTI para treino 2D, (ii) uso de U-Net com ResNet50 pré-treinado, (iii) técnicas específicas de pré-processamento e aumento de dados, e (iv) análise qualitativa e quantitativa.

2. Dataset e Metodologia

2.1. Dataset e Pré-processamento

Foi utilizado o dataset LiTS [Mvd 2017] (131 volumes NIfTI [Milletari et al. 2016] com anotações de fígado e tumores). As fatias foram convertidas com NiBabel em tenso-

res NumPy, padronizadas por rotação e submetidas a janelas de intensidade: hepática (150,30) e customizada (200,60). Após a normalização e redimensionamento (128×128), geraram-se imagens JPEG (dados) e PNG (máscaras), assegurando a compatibilidade com frameworks de Deep Learning. Os volumes foram divididos em 70% para treino, 15% para validação e 15% para teste, respeitando pacientes distintos entre os conjuntos.

2.2. Arquitetura do Modelo e Treinamento

A arquitetura central foi a U-Net [Ronneberger et al. 2015] com backbone ResNet50 pré-treinado em ImageNet [He et al. 2016]. O *transfer learning* acelerou a convergência e melhorou a generalização em datasets limitados [Isensee et al. 2021]. Foram aplicadas técnicas de aumento de dados (rotações, flips, ajustes de brilho/contraste e deformações elásticas [Milletari et al. 2016]). A função de perda combinou CrossEntropyLoss e Dice Loss [Dice 1945], balanceando *pixel-wise accuracy* e medidas de região [Ronneberger et al. 2015, Milletari et al. 2016]. O treinamento utilizou *learning rate finder*, *fine-tuning* em múltiplas fases [Howard and Gugger 2020], *batch* = 16, *weight decay* = 0.1 e early stopping. O melhor modelo foi salvo via `SaveModelCallback`.

2.3. Métricas de Avaliação

A avaliação usou Dice Coefficient (sobreposição), Sensibilidade (pixels de tumor detectados) e Especificidade (falsos positivos reduzidos). Métricas customizadas foram implementadas no notebook, calculando a acurácia apenas em pixels foreground [Dice 1945, Ronneberger et al. 2015, Milletari et al. 2016].

- **Dice Coefficient (F1 Score de segmentação):** mede a sobreposição entre máscara predita e referência [Dice 1945].
- **Sensibilidade (Recall):** avalia a proporção de pixels de tumor corretamente detectados [Ronneberger et al. 2015].
- **Especificidade:** quantifica a capacidade do modelo em identificar corretamente o *background*, evitando falsos positivos [Milletari et al. 2016].

3. Resultados

3.1. Métricas Quantitativas

A Tabela 1 apresenta os resultados quantitativos obtidos pelo modelo U-Net com backbone ResNet50 treinado sobre o dataset de tumores hepáticos. Foram consideradas as métricas: `train_loss`, `valid_loss`, `foreground_acc` e `cust_foreground_acc`. Os valores refletem o desempenho médio nos conjuntos de validação e teste.

3.2. Análise Qualitativa

Além das métricas quantitativas, foram avaliados exemplos visuais das segmentações. A Figura 1 mostra casos representativos: (a) segmentações corretas, onde o contorno do fígado e das lesões foi identificado com alta precisão; (b) falhas comuns, como subsegmentação de tumores pequenos e imprecisão nas bordas hepáticas em regiões de baixo contraste.

Tabela 1. Resultados quantitativos da segmentação automática de fígado e tumores.

Época	Train Loss	Valid Loss	Foreground Acc	Cust.Foreground acc
1	0.070	0.036	0.836	0.950
2	0.017	0.024	0.832	0.975
3	0.010	0.022	0.841	0.966
4	0.007	0.019	0.857	0.986
5	0.006	0.033	0.893	0.990

3.3. Discussão

Os resultados obtidos indicam que o uso de *transfer learning* com o backbone ResNet50 é eficaz em cenários médicos com dados limitados, proporcionando ganhos substanciais em relação a abordagens treinadas do zero. A aplicação de *data augmentation* foi particularmente importante para lidar com a variabilidade anatômica e as diferenças de aquisição entre exames, reduzindo casos de sobreajuste. O pré-processamento, especialmente a aplicação de janelas de intensidade específicas, também contribuiu para uma melhor distinção entre fígado, tumores e tecidos vizinhos, simplificando a tarefa de segmentação.

Entretanto, ainda foram observadas limitações importantes. O modelo apresentou dificuldades em identificar tumores muito pequenos ou localizados em regiões de baixo contraste. Além disso, a generalização para outros datasets, coletados em equipamentos distintos ou populações diferentes, permanece uma questão em aberto, ressaltando a necessidade de estudos futuros com dados mais diversos. Embora o foco não tenha sido a comparação direta com outros métodos, os resultados obtidos são compatíveis com os relatados na literatura para o dataset LiTS.

4. Conclusão e Trabalhos Futuros

Os resultados quantitativos demonstraram que o método alcança bom equilíbrio entre sensibilidade e especificidade, enquanto a análise qualitativa evidenciou que o modelo é capaz de segmentar corretamente a maioria dos tumores, embora apresente dificuldades em casos de lesões muito pequenas ou em regiões de baixo contraste.

As contribuições principais incluem: (i) adaptação de um dataset público em formato NIfTI para treinamento em pipelines modernos, (ii) validação da eficácia do uso de modelos pré-treinados em segmentação médica e (iii) análise combinada quantitativa e qualitativa dos resultados. Como limitações, destacam-se a dependência do dataset utilizado e a restrita capacidade de generalização para outros contextos clínicos sem ajustes adicionais. Trabalhos futuros devem explorar variantes recentes (*Attention U-Net*, *Swin-UNet*) e abordagens semi/auto-supervisionadas, hoje predominantes em segmentação médica, além da utilização de aprendizado semi-supervisionado ou auto-supervisionado para aproveitar melhor os dados não anotados.

Em síntese, os resultados confirmam o potencial do *transfer learning* para segmentação médica, destacando a importância do pré-processamento adequado e da análise crítica dos resultados no desenvolvimento de sistemas de apoio ao diagnóstico.

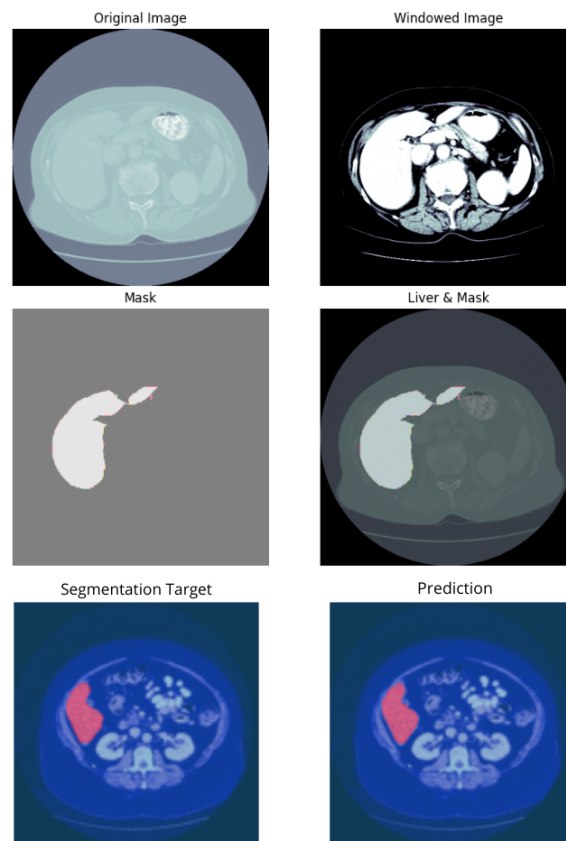


Figura 1. Exemplos qualitativos de segmentação: à esquerda, resultados corretos; à direita, erros típicos como subsegmentação de tumores pequenos.

Referências

- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.
- Howard, J. and Gugger, S. (2020). *Deep Learning for Coders with FastAI and PyTorch: AI Applications Without a PhD*. O'Reilly Media.
- Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., and Maier-Hein, K. H. (2021). nnu-net: Self-adapting framework for u-net-based medical image segmentation. *Nature Methods*, 18:203–211.
- Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation. *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571.
- Mvd, A. (2017). Liver tumor segmentation challenge (lits). <https://www.kaggle.com/datasets/andrewmvd/liver-tumor-segmentation>.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 9351:234–241.