

Aplicação de Redes Neurais Totalmente Convolucionais na Segmentação Semântica de Imagens Médicas

Thiago Faccio¹, Wallace Casaca²
IBILCE/UNESP, São José do Rio Preto, SP

A tarefa de segmentação de imagens médicas pode ser realizada de maneira inteiramente automática, ou alternativamente, por meio de mecanismos interativos [3]. Na abordagem automática, o modelo é projetado para interpretar os dados da imagem de forma independente, agrupando pixels com base em critérios de similaridade, como proximidade espacial ou intensidade de cores. Esses métodos fazem uso de algoritmos capazes de detectar padrões ou características-alvo específicas, aplicando funções matemáticas que analisam gradientes, bordas, texturas, ou outras características que possibilitam distinguir as regiões da imagem.

As chamadas Redes Neurais Totalmente Convolucionais (do inglês, *Fully Neural Networks* – FCNs) têm se mostrado uma abordagem efetiva em aplicações envolvendo a segmentação semântica de imagens médicas. Ao eliminar a necessidade de camadas densamente conectadas, as redes FCNs permitem o mapeamento direto dos pixels de entrada para a partição de saída, oferecendo maior eficiência em comparação com outras arquiteturas de Redes Neurais Convolucionais [1]. As FCNs tem sido amplamente empregadas para diferentes fins, incluindo sensoriamento remoto e análise biomédica, e a evolução dessas redes tem levado ao desenvolvimento de modelos mais refinados, como a rede U-Net [2], projetada para segmentação biomédica, e a DeepLabv3+ [2], desenvolvida para tarefas gerais de segmentação semântica, incorporando convoluções para melhorar a captação de diferentes níveis de contexto.

Considerando a importância das FCNs no contexto de segmentação, este estudo investigou o desempenho de diferentes arquiteturas FCNs aplicadas à segmentação semântica de imagens médicas. Para esse propósito, foi empregada a base de dados *Lung Segmentation* [4], e analisadas três arquiteturas: a U-Net, a rede DeepLabv3+, e uma FCN genérica, constituída por uma arquitetura em U, filtros de convoluções e operações de *max pooling* no processo de *encoder*, e convoluções transpostas no processo de *decoder*. Além disso, a referida rede contou com 512 filtros convolucionais em sua camada mais profunda, com intuito de comparar sua performance em termos de acurácia e eficiência computacional em um contexto de segmentação binária. A topologia implementada nas três redes é similar: todas utilizam as funções de ativação *ReLU* em suas camadas convolucionais, e por se tratar de um problema binário, a função *Sigmoid* em suas camadas de saída, com o diferencial que a DeepLabv3+ conta também com o módulo *Atrous Spatial Pyramid Pooling* (ASPP) [2] para capturar informações em múltiplas escalas via convoluções dilatadas.

A metodologia adotada consistiu no treinamento integral das redes descritas acima, sem a utilização de pesos pré-treinados, utilizando o mesmo conjunto de 704 imagens de raio-X disponíveis na base *Lung Segmentation* com suas respectivas máscaras binárias de referência. As imagens e máscaras foram redimensionadas para a resolução de 256×256 pixels e, para a divisão do conjunto de dados, adotou-se 80% das amostras para o treinamento e 20% para a validação.

¹thiago.faccio@unesp.br

²wallace.casaca@unesp.br

A Tabela 1 apresenta as métricas de avaliação, acurácia, perda (*loss*), e tempo de processamento por época de treinamento para as três arquiteturas implementadas após 40 épocas. Sob a perspectiva quantitativa, os resultados demonstraram que a arquitetura U-Net atingiu a maior acurácia média. Por outro lado, as redes DeepLabv3+ e FCN genérica, embora apresentando mensurações menores que a U-Net, mantiveram escores competitivos, destacando-se especialmente pelo desempenho computacional superior, particularmente da FCN genérica, que demonstrou melhor performance de processamento em comparação com as demais arquiteturas treinadas.

Tabela 1: Métricas de avaliação. Valores em negritos indicam os melhores escores.

Arquitetura	Acurácia	Perda	Tempo p/Época
FCN Genérica	0.9721	0.0735	73s
DeepLabv3+	0.9791	0.0896	258s
U-Net	0.9809	0.0520	320s

Este estudo realizou uma análise comparativa entre diferentes arquiteturas de redes neurais totalmente convolucionais para segmentação de imagens médicas de raio-x. Enquanto a U-Net se mostra ideal para cenários que exigem alta precisão, a FCN de implementação própria se destacou pela melhor eficiência computacional, sendo útil em contextos onde o tempo de inferência é um fator crítico. Além disso, por não depender de *skip connections*, a FCN genérica não só foi mais rápida como conseguiu sobrepor as variáveis de convoluções anteriores, demandando menos alocação de memória durante o treinamento. A DeepLabv3+ por sua vez, ofereceu uma solução intermediária, equilibrando qualidade e custo computacional. Os resultados obtidos reforçam a relevância das redes totalmente convolucionais na segmentação de imagens médicas, destacando a necessidade de um equilíbrio entre precisão, eficiência computacional e viabilidade prática conforme as exigências de cada aplicação.

Agradecimentos

Os autores agradecem à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP – Processos 2013/07375-0 e 2023/14427-8) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq – Processos 316228/2021-4 e 141073/2024-1) pelo apoio à pesquisa.

Referências

- [1] M. Agarwal, S. K. Gupta e K. K. Biswas. “Development of a compressed FCN architecture for semantic segmentation using Particle Swarm Optimization”. Em: **Neural Computing and Applications** 35.16 (2023), pp. 11833–11846. DOI: 10.1007/s00521-023-08324-3.
- [2] A. Akagic, M. Kapo, E. Kandić, M. Bećirović e N. Kadrić. “Brain tumor segmentation of MRI images with U-Net and DeepLabV3+”. Em: **IEEE 3rd International Conference on Computing and Machine Intelligence (ICMI)**. 2024, pp. 1–6.
- [3] A. Bruzadin, M. Boaventura, M. Colnago, R. G. Negri e W. Casaca. “Learning label diffusion maps for semi-automatic segmentation of lung CT images with COVID-19”. Em: **Neurocomputing** 522 (2023), pp. 24–38. DOI: 10.1016/j.neucom.2022.12.003.
- [4] S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma e C. J. McDonald. “Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration”. Em: **IEEE Transactions on Medical Imaging** 33.2 (2013), pp. 577–590.