

Relatório sobre aplicações de Deep Learning no tratamento de MRI

Relatório para a UC de Redes Neurais e Aprendizado Profundo

Professora Regina Coelho - ICT/UNIFESP

Aluno Gabriel Sardinha - RA:140853

1) Introdução

Métodos utilizando Deep Learning são amplamente usados para o tratamento/segmentação de imagens clínicas de ressonância magnética, visto que esse tipo de imagem costuma apresentar baixa qualidade e até mesmo ruídos. Nesse relatório irei abordar duas aplicações que buscam solucionar, de forma eficaz, certos problemas[3] que sondam esse tipo de trabalho, como explosão de gradiente, overfitting, tempo de treinamento, acurácia entre outros. Ambos os artigos propõem modelos que combinam CNN's com interação com usuário para apurar a segmentação. Irei discutir pontos em comum, principais diferenças e melhores resultados.

2) Métodos

2.1) Bounding box and Image-specific Fine-tuning-based Segmentation (BIFSeg)

2.1.1) Introdução

Esse primeiro artigo propõe um framework que combina uma CNN, que fornece uma segmentação inicial, com um *fine-tuning* de imagem específica, supervisionada (a partir de marcações do usuário) ou não-supervisionada (sem marcações do usuário). Isso faz com que essa CNN se torne mais especializada em determinados tipos de imagem, sendo que ela pode ser utilizada para imagens 2D ou 3D.

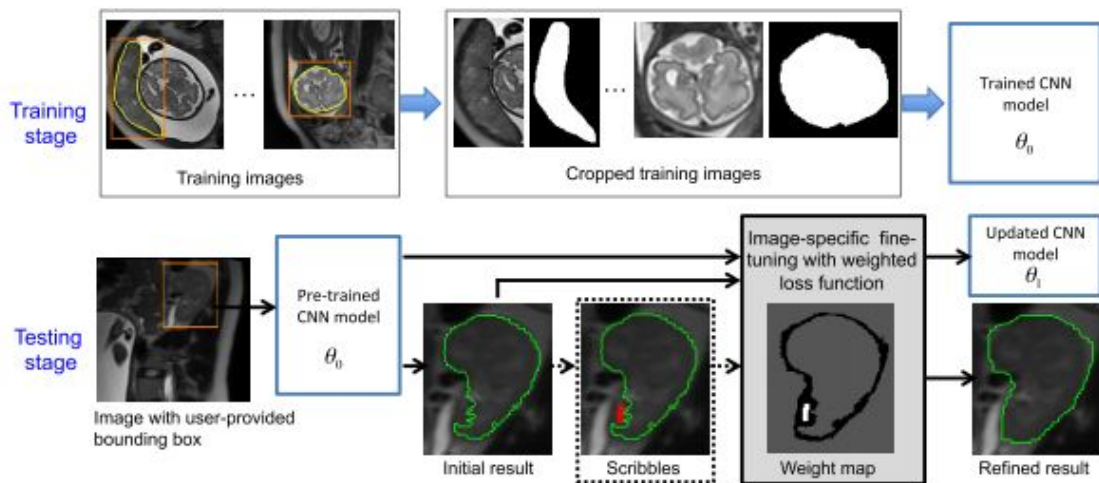
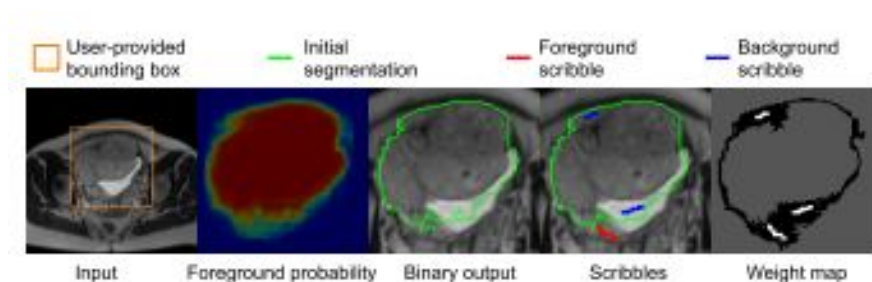


Fig 01: Esquema do método proposto.



2.1.2) Explicação

Para imagens 2D, os autores optam pelo método P-Net [4] para a segmentação binária inicial, a partir da *bounding box* fornecida pelo usuário. Esse método consiste em cinco blocos de layers convolucionais com unidades de ativação retificada (ReLU) que possuem parâmetros de dilatação diferentes, utilizados para coletar características em diferentes escalas. Essas características são concatenadas para um sexto bloco que será o classificador, que contém layer softmax para obter o output e executar o *fine-tuning*. Para imagens 3D, eles utilizam uma extensão de P-Net, chamada PC-Net para a segmentação inicial, que consiste em 2 blocos de kernels 3x3x3 e 3 blocos de kernels 3x3x1 para a coleta de características. Com isso, são comprimidas as características coletadas em layers 1x1x1 para concatená-las para o bloco 6 (também classificador), onde ocorrerá o fine-tuning.

Para treinamento, os autores utilizam uma *cross entropy loss function*. Para o *fine-tuning* da rede e do output os autores propõem um sistema de pesos (*weighted loss function*), que os separa a partir de sua confiabilidade. Os pixels marcados pelo usuário irão possuir uma confiabilidade muito maior perante aos outros pixels, com isso a rede tende a ignorar pixels com baixa confiabilidade. Essa confiabilidade girará em torno do output do layer softmax localizado no sexto bloco, para situações não-supervisionadas.

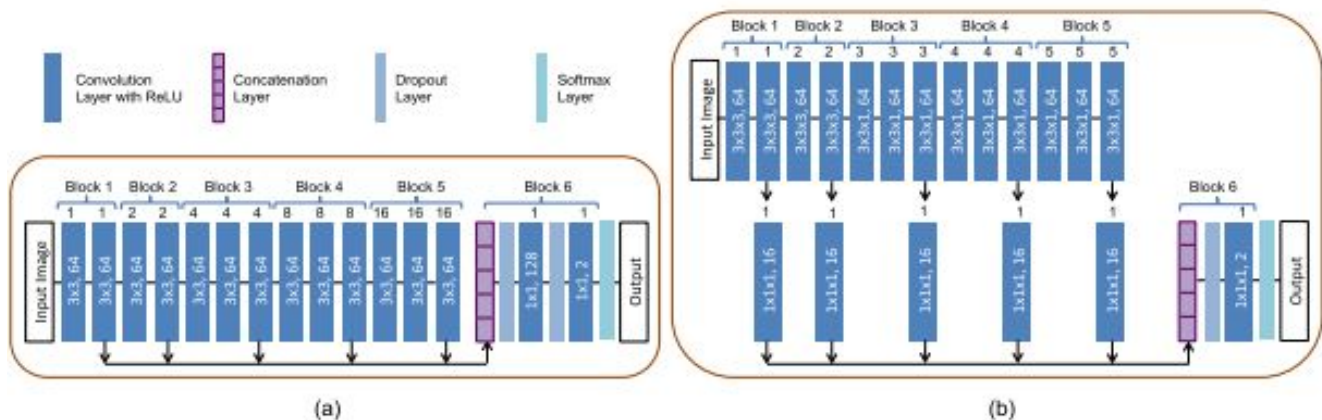


Fig 02: Camadas e pathing da rede para imagens 2D (a) e para imagens 3D (b)

2.1.3) Resultados

Primeiramente, conclui-se que o método estudado é aplicável, teoricamente, em outros modelos de CNN. Como a pesquisa é focada em segmentação com interação do usuário, o método perde em quesitos tempo e eficiência de memória, quando comparado com métodos totalmente convolucionais [4][5], porém ganha em acurácia em praticamente todos os testes. Além disso, se tratando de redes com interações de usuários, o BIFSeg alcança altas acurácias com um menor número de interações quando comparado com métodos tradicionais.

TABLE III

QUANTITATIVE COMPARISON OF P-NET AND THREE SUPERVISED REFINEMENT METHODS WITH SCRIBBLES FOR FETAL MRI SEGMENTATION. T_m IS THE MACHINE TIME FOR REFINEMENT. \wedge DENOTES PREVIOUSLY UNSEEN OBJECTS. IN EACH ROW, BOLD FONT DENOTES THE BEST VALUE. * DENOTES p -VALUE < 0.05 COMPARED WITH THE OTHERS

		P-Net	P-Net+CRF	BIFSeg(-w)	BIFSeg
Dice (%)	P	84.57 \pm 8.37	88.64 \pm 5.84	89.79 \pm 4.60	91.93\pm2.79*
	FB	89.44 \pm 6.45	94.04 \pm 4.72	95.31 \pm 3.39	95.58\pm1.94
	FL [^]	83.59 \pm 6.42	88.92 \pm 3.87	89.21 \pm 2.95	91.71\pm3.18*
	MK [^]	85.29 \pm 5.08	87.51 \pm 4.53	87.78 \pm 4.46	89.37\pm2.31*
T_m (s)		-	0.02\pm0.01*	0.72 \pm 0.11	0.74 \pm 0.12

P: Placenta, FB: Fetal brain, FL: Fetal lungs, MK: Maternal kidneys.

TABLE VI

QUANTITATIVE COMPARISON OF PC-NET AND THREE SUPERVISED REFINEMENT METHODS WITH ADDITIONAL SCRIBBLES FOR 3D BRAIN TUMOR SEGMENTATION. T_m IS THE MACHINE TIME FOR REFINEMENT. Δ DENOTES PREVIOUSLY UNSEEN OBJECTS. IN EACH ROW, BOLD FONT DENOTES THE BEST VALUE. * DENOTES p -VALUE < 0.05 COMPARED WITH THE OTHERS

		PC-Net	PC-Net+CRF	BIFSeg(-w)	BIFSeg
Dice (%)	TC	82.66 \pm 7.78	85.93 \pm 6.64	85.88 \pm 7.53	87.49\pm6.36*
	WT $^{\wedge}$	83.52 \pm 8.76	85.18 \pm 6.78	86.54 \pm 7.49	88.11\pm6.09*
$T_m(s)$	TC	-	0.14\pm0.06*	3.33 \pm 0.86	4.42 \pm 1.88
	WT $^{\wedge}$	-	0.12\pm0.05*	3.17 \pm 0.87	4.01 \pm 1.59

TC: Tumor core in T1c, WT: Whole tumor in FLAIR.

2.2) Deep Interactive Geodesic Framework (DeepGeoS)

2.2.1) Introdução

Neste segundo método, os autores propõem um framework que consiste em dois estágios: No primeiro estágio, eles usam uma CNN totalmente automática (P-Net + CRF-Net) para fornecê-los uma segmentação inicial. Posteriormente o usuário fornece alguns cliques ou rabiscos, indicando alguma possível falha da demarcação inicial, com isso, no segundo estágio, é utilizada outra CNN (R-Net + CRF-Net) para refinar a primeira segmentação a partir do input do usuário.

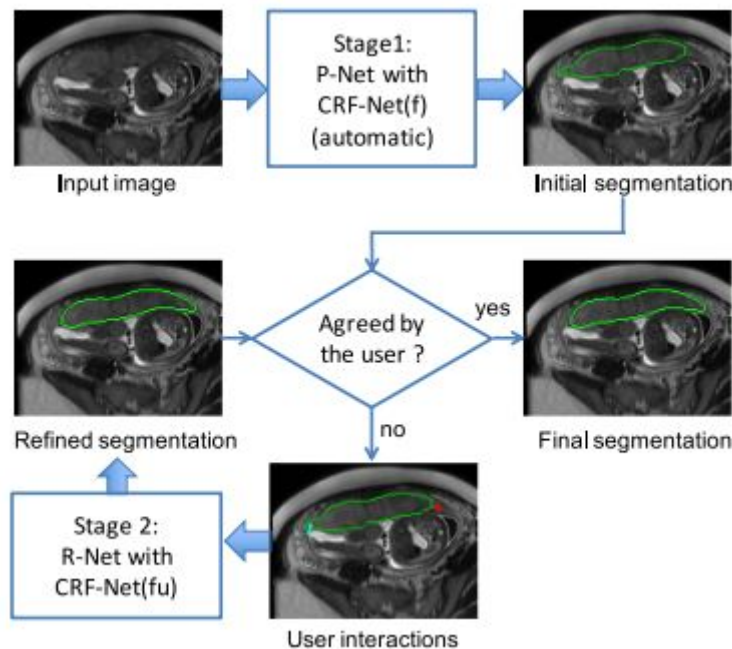


Fig 01: Esquema do método proposto.

2.2.2) Explicação

O primeiro estágio dessa método é similar ao da rede exemplificada acima, visto que ela também usa um método P-Net [4] para a segmentação inicial, com o adicional de uma rede *Conditional Random Field (CRF)* [6]. O que diferencia esse framework do já explicado é que na fase de *fine-tuning* é utilização de outro método (R-Net) juntamente com a mesma CRF para o refinamento do output, e é ela que irei detalhar a seguir. R-Net utiliza a mesma estrutura do P-Net, porém o que muda é o número de canais de entrada, que passa a ser 16, e emprega as interações do usuário no CRF-Net. Uma *Conditional Random Field* consiste em uma rede que modela pixel labels para variáveis randômicas que formam um *Markov Random Field (MRF)*, a partir de uma sequência de algoritmos de *Message Passing > Re-weighting > Compatibility Transform > Unary Addition > Normalization*. O método CRF foi modelado como uma rede neural recorrente (RNN) [6] com algumas modificações (Mudança de uma única função Gaussiana para *freeform functions* (Pairwise-Net); Integração com interações do usuário; Os *Unary Potentials* são obtidos através de P-Net/R-Net), para que ela seja treinada juntamente com a P-Net/R-Net por retropropagação.

Os autores propõem codificar as interações do usuário, através de transformações de Geodesic Distance, para servir de input juntamente com a segmentação inicial a fim de refinar a imagem recebida. A Pairwise-Net é uma CNN com kernels 1x1, para imagens 2D, ou 1x1x1 para imagens 3D, para treiná-la foi utilizado um algoritmo *Stochastic Gradient Descent (SGD)* com uma *quadratic loss function*.

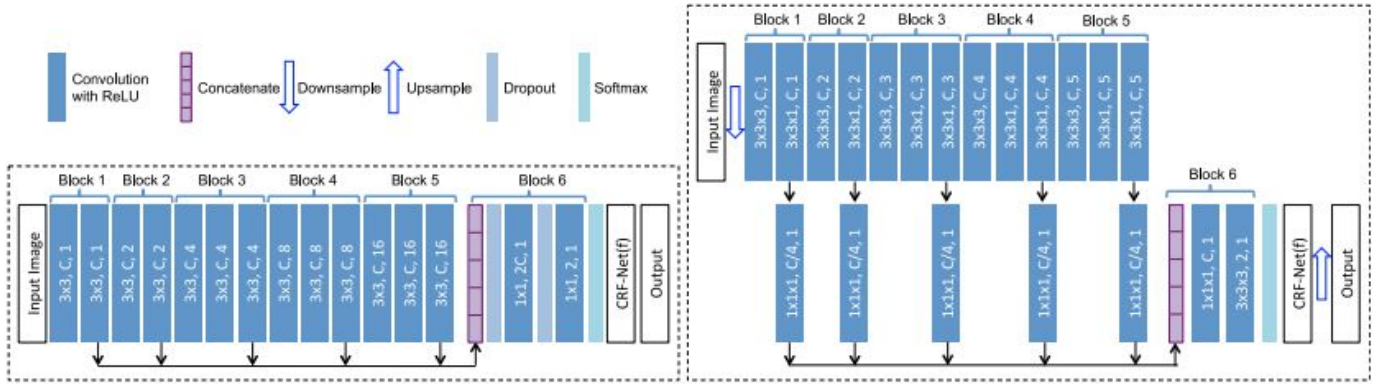


Fig 02: Camadas e pathing da rede para imagens 2D (a) e para imagens 3D (b)



Fig. 4. The Pairwise-Net for pairwise potential function $f(\tilde{f}_{ij}, d_{ij})$. \tilde{f}_{ij} is the difference of features between a pixel pair i and j . d_{ij} is the euclidean distance between them.

2.2.3) Resultados

Similarmente ao primeiro artigo, os autores concluem que, futuramente, esse framework pode ser aplicado para tratamento de imagens de outros órgão no futuro. Tanto para imagens 2D do *Fetal MRI*, quanto para imagens 3D de tumores cerebrais do FLAIR, esse método se mostrou mais eficaz, em questões de acurácia, quando comparado a métodos tradicionais. Porém ele demonstra os mesmos problemas que o anterior, visto que ele é um método que combina interações com o usuário.

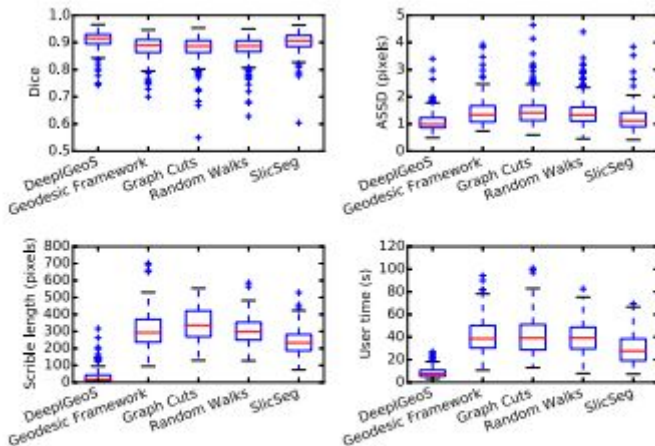


Fig. 10. Quantitative comparison of 2D placenta segmentation by different interactive methods in terms of Dice, ASSD, total interactions (scribble length) and user time.

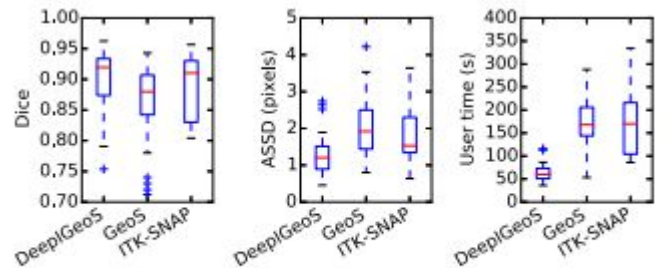


Fig. 15. Quantitative evaluation of 3D brain tumor segmentation by DeepiGeoS, GeoS and ITK-SNAP.

3) Discussão Final

3.1) Principais diferenças e semelhanças

Para principais semelhanças em ambos os projetos identifiquei os seguintes pontos:

Utilizaram um método já existente (P-Net) para a segmentação inicial;

- Ambos substituíram os layers de *max-polling* e *downsampling* do P-Net por *dilated convolution layers* em cada bloco, a fim de preservar a resolução da imagem;
- Ambos verificaram que interações com o usuário são benéficas para a acurácia;
- Ambos os métodos podem vir a ser utilizados em outras aplicações no futuro.

Para principais diferenças, identifiquei os seguintes pontos:

- Na hora do refinamento, um optou por uma função *weighted loss function* e o outro implementou uma CRF-RNN;
- Na hora do treinamento, um preza para que seu modelo treine de uma maneira que se torne *image-specific*, já o outro não foca nessa situação.

3.2) Comparação de Resultados

Podemos verificar que, posteriormente a todos os testes, o BIFSeg se mostrou mais eficiente no geral, quando comparado DeepGeoS no que tange imagens 2D. Quando o assunto são imagens 3D do cérebro, o DeepGeoS ganha tranquilamente, ambos os testes a métrica “Dice” para avaliação. Como se trata de modelos que possuem interações com o usuário, outros modelos *fully CNN* ganham quando o quesito é tempo de máquina. No geral, o DeepGeoS apresenta um resultado mais satisfatório.

3.3) Conclusões Finais

Com esse trabalho, podemos concluir que na área de segmentação de imagens clínicas a “variável humana” se mostra indispensável quando procuramos melhores acurácias, por enquanto. Além disso, para futuros trabalhos, podemos cogitar esse tipo de modelo sendo aplicado em outras áreas, até mesmo em questões de previsibilidade, que vão além da compreensão humana atual. Gostaria de agradecer a professora Regina por ter aplicado esse trabalho, já que possibilitou uma pesquisa enorme nessa área e aumentou ainda mais a minha vontade de seguir por essa área tão instigante e ainda misteriosa da computação.

Referências

- [1] G. Wang *et al.*, "Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning," in *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1562-1573, July 2018, doi: 10.1109/TMI.2018.2791721.
- [2] G. Wang *et al.*, "DeepGeoS: A Deep Interactive Geodesic Framework for Medical Image Segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1559-1572, 1 July 2019, doi: 10.1109/TPAMI.2018.2840695.
- [3] Hesamian, M.H., Jia, W., He, X. *et al.* Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *J Digit Imaging* 32, 582–596 (2019). <https://doi.org/10.1007/s10278-019-00227-x>
- [4] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834-848, 1 April 2018, doi: 10.1109/TPAMI.2017.2699184.
- [5] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. CVPR, Jun. 2015, pp. 3431–3440.
- [6] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr, "Conditional random fields as recurrent neural networks," in Proc. IEEE Int. Conf. Comput. Vis., 2015, pp. 1529–1537.