

Attention Efficient Residual U-Net: uma Rede Neural para a Segmentação de Lesões de Pele

José Guilherme Pereira Lima¹, Geraldo Braz Júnior¹

¹Departamento de Informática - Universidade Federal do Maranhão (UFMA)
guilherme.jose@discente.ufma.br, geraldo.braz@ufma.br

Abstract. *Melanoma is one of the most serious types of skin cancer due to its high mortality rate, which can reach 70%. An early diagnosis of the disease is crucial as it increases the ten-year survival rate by up to 97%. The segmentation of skin lesions is one of the essential steps in the diagnostic process for the accurate detection of melanoma. However, even for specialist doctors, segmenting these lesions can be costly and challenging due to the wide variety of skin lesions, which can have irregular edges, different dimensions and colors, and the large number of exams to be analyzed. Automatic detection of the lesion area for segmentation is an important area of study so the doctors can focus on the correct diagnosis of the disease itself. This work aims to compare popular convolutional neural network architectures and propose a new model based on encoder-decoder architectures that achieve efficiency and performance parameters for the segmentation of dermoscopic images. Thus, a method based on Attention U-net, using pre-trained encoder and residual blocks for better optimization called AER-Net (simplified from Attention Efficient Residual U-net) was proposed, which presented 85.25% and 77.29% in Dice Coefficient and Jaccard index respectively for the ISIC Archive image database and proved that the model can assist in the automatic diagnosis process of skin cancer.*

Resumo. *O melanoma é um dos tipos de câncer de pele mais graves devido à sua alta taxa de mortalidade, que pode chegar a 70%. Um diagnóstico precoce da doença é crucial, pois aumenta a taxa de sobrevivência de dez anos em até 97%. A segmentação das lesões cutâneas é uma das etapas essenciais do processo de diagnóstico para a detecção precisa do melanoma. Porém, mesmo para médicos especialistas, segmentar essas lesões é custoso e desafiador pela grande variedade de manchas, que podem ter bordas irregulares, dimensões e cores diferentes, e pela grande quantidade de exames a serem analisados. A detecção automática da área da lesão para segmentação mostra-se uma importante área de estudo para que o médico especialista possa se concentrar no diagnóstico correto da própria doença. Este trabalho teve como objetivo comparar arquiteturas de redes neurais convolucionais populares e propor um novo modelo baseado em arquiteturas encoder-decoder que atinja parâmetros de eficácia e desempenho para a segmentação de imagens dermatoscópicas. Assim, foi proposto um método baseado em Attention U-net, utilizando encoder pré-treinado e blocos residuais para melhor otimização. O modelo proposto, chamado de AER-Net (simplificado de Attention Efficient Residual U-net), apresentou 85,25% e 77,29% em Coeficiente Dice e índice Jaccard respectivamente para base de imagens do ISIC Archive e comprovou que o modelo pode auxiliar no processo de diagnóstico automático de câncer de pele.*

1. Introdução

O câncer de pele é o tipo mais comum de câncer, sendo responsável por um em cada três casos em todo o mundo [Ge et al. 2017]. O câncer de pele pode ser dividido em dois grupos principais: melanoma e não melanoma. Embora o melanoma seja responsável por apenas 22% dos casos [for Cancer Research 2018], é de longe o mais perigoso porque tem mais probabilidade de crescer e avançar para outros órgãos. As últimas estatísticas disponíveis no mundo mostram que os casos foram aumentando a cada ano a uma taxa alarmante. Nos Estados Unidos, estima-se que o número de novos casos de melanoma diagnosticados em 2021 aumentará 5,8% com 106.110 novos casos de melanoma sendo diagnosticados, resultando em cerca de 7.000 mortes [Society 2021]. No Brasil, estima-se que 8.450 novos casos de melanoma foram diagnosticados no país em 2020 [INCA 2021]. Embora o melanoma consista em apenas 4% dos diagnósticos de câncer de pele no país, é responsável por 70% de todas as mortes por este tipo de câncer [INCA 2021].

Embora as estatísticas indiquem a gravidade da doença, os números também indicam que o diagnóstico precoce do melanoma aumenta muito a prevalência de recuperação. Quando detectado precocemente, a taxa de sobrevivência de 5 anos para o melanoma é de 99%. A taxa de sobrevivência cai para 66% quando a doença atinge os nódulos linfáticos e 27% quando a doença metastatiza para órgãos distantes [Society 2021].

Uma das formas não invasivas desse diagnóstico é por meio da dermatoscopia, que consiste em um médico especialista examinando imagens dermatoscópicas de lesões cutâneas. Contudo, mesmo para médicos especialistas, segmentar essas lesões é custoso e difícil devido à grande variedade de manchas que às vezes têm bordas irregulares, diferentes dimensões e cores. Portanto, estudos vêm sendo realizados para detectar automaticamente esses ferimentos para ajudar os profissionais. Uma das etapas essenciais na análise computadorizada de imagens dermatoscópicas é a segmentação automática da lesão cutânea. Nos últimos anos, Redes Neurais Convolucionais (CNNs) surgiram como uma das ferramentas mais poderosas de processamento de imagens mostrando resultados promissores em vários domínios, incluindo análises de imagens médicas [Ronneberger et al. 2015]. A maioria das abordagens de segmentação de imagens médicas por CNNs alcançaram resultados de ponta através do desenvolvimento de arquiteturas encoder-decoder para redes neurais convolucionais profundas, por exemplo, a Segnet [Badrinarayanan et al. 2017], U-net [Ronneberger et al. 2015], e redes totalmente convolucionais [Long et al. 2015]. O encoder é responsável pela extração de características automática através de várias camadas de convolução e downsampling. O decoder incorpora e faz upsampling das características extraídas da parte do encoder para gerar a máscara de segmentação prevista.

Devido a clara importância que o diagnóstico precoce de câncer de pele exerce nas taxas de sobrevivência, métodos que ajudam na detecção automática da lesão em questão se mostram cada vez mais relevantes para auxiliar os profissionais médicos e facilitar a diagnose. A relevância desse estudo fundamenta-se na proposta de um modelo novo de CNN baseado em arquiteturas do tipo encoder e decoder que possa ser eficaz na segmentação de imagens médicas. O novo método proposto foi chamado de AER-Net é composto por três conceitos fundamentais: backbone pré-treinado da família Efficientnet, uso de conexões resíduas ao longo da arquitetura e skip connections em Attention Gates. Detalhes da metodologia proposta pode ser visto na seção 3.

2. Referencial Teórico

Nazi et al. [AI Nazi and Abir Tasnim 2020] descrevem uma abordagem automática para segmentar lesões de pele usando arquitetura profunda U-Net e transferência de aprendizado. A arquitetura U-Net, juntamente com a técnica de pós-processamento, limiar de Otsu, foram aplicadas na abordagem proposta por Souza et al. [Souza et al. 2020].

A metodologia aplicada por Ji et al. [Ji et al. 2018] propõe uma arquitetura de rede de agregação de recursos chamada FA-CNN. Para isso, a rede ResNet34 [He et al. 2016] é implementada como um backbone para construir o módulo do encoder. Como contribuição, este trabalho visa empregar supervisão auxiliar aos blocos de convolução entre o encoder e o decoder. Para a segmentação da lesão cutânea, Amin et al. [Amin et al. 2020] propõe o uso da arquitetura de rede Alexnet e VGG-16 para extração de recursos para em seguida fazer a fusão e seleção de recursos usando PCA.

3. Metodologia

3.1. Aquisição de dados e Pré-processamento

A base de dados utilizada para este estudo foi a ISIC 2018 dataset [Codella et al. 2019] [Tschandl et al. 2018], utilizada no desafio ISIC para a Análise de Lesão de Pele para a Detecção do Melanoma, edição 2018. O desafio proposto pela International Skin Imaging Collaboration (ISIC), uma parceria global que organizou o maior repositório mundial de imagens dermoscópicas disponíveis publicamente, disponibilizou um dataset de 2594 imagens dermatoscópicas.

Para cada imagem no banco de dados, estão disponíveis também a segmentação manual feita por um especialista, como é demonstrado na Figura 1.

Figura 1. Exemplo de lesão maligna encontrada na ISIC 2018 com sua máscara de segmentação.

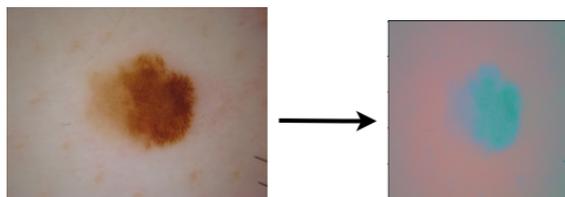


As imagens inicialmente RGB com dimensões variadas que iam de 576×768 a 6748×4499 pixels foram redimensionadas para 256×256 pixels. Foi utilizado o espaço de cores CIELAB, após testes em diferentes canais de cores como HSV, RGB, HSV+RGB e tons de cinza onde demonstrou a maior eficácia na segmentação de imagens da base. Cada canal da imagem foi então normalizado para o intervalo $[0, 1]$ para otimização do treinamento das redes.

3.2. Método Proposto

Neste trabalho propõe-se um novo modelo de CNN baseado na U-Net e uso de transferência de aprendizado, chamado de Attention Efficient Residual U-Net (AER-Net). A Figura 3 apresenta a visão esquemática da arquitetura. É dividida em dois caminhos, o encoder e decoder.

Figura 2. Exemplo de imagem de entrada após o pré-processamento.



Para o encoder, primeiramente, faz-se uso do backbone EfficientNetb1 pré-treinado na base de dados Imagenet como o encoder da arquitetura, o qual irá reduzir o tamanho da imagem de entrada e extrair características através das sucessivas operações de convolução.

Baseados na teoria de escalonamento de modelos, Tan et al. [Tan and Le 2020] propuseram um novo método de escala que dimensiona uniformemente todas as dimensões de profundidade, largura e resolução usando um coeficiente de composto simples, mas altamente eficaz de acordo com um número fixo de fatores de escala. Além disso, projetaram uma nova rede base EfficientnetB0, a qual foi possível escaloná-la para obter uma família de modelos chamados EfficientNets, nomeados a partir da EfficientNetb0 à EfficientNetb7.

O bloco convolucional que compõe a família EfficientNet é o MBConv [Sandler et al. 2018], o qual incluem otimizações de aperto e excitação para os mapas de features. A medida que o escalonamento varia da EfficientNetB0 à EfficientNetB7, o número de operações, profundidade, largura, resolução e o tamanho do modelo aumentam, assim como a acurácia. Nesta pesquisa, foram testados os modelos EfficientNetB0 ao EfficientNetB4 devido a limitações de hardware, com o backbone EfficientNetB1 obtendo a melhor performance para a tarefa.

Para construir o decoder, usa-se a cada nível uma camada de upsampling com stride = 2 que dobra o tamanho de um mapa de recursos enquanto reduz o número de canais pela metade, uma camada de concatenação que mescla o mapa de recursos gerado das etapas anteriores com o mapa do encoder correspondente, seguida por blocos residuais contendo duas camadas de convolução com filtros 3x3, seguidas por Batch Normalization (BN) e função de ativação ReLU. Por fim, é aplicada uma camada de convolução 1x1 e função de ativação sigmoid para gerar a máscara de segmentação final. A representação visual do bloco residual decoder é apresentada na Figura 4.

3.2.1. Attention Gates

Assim como na U-Net, as conexões dos mapas de recursos do encoder diretamente com o decoder correspondente são feitas através de skip connections. Técnicas foram propostas na literatura para modificações das skip connections, as reprojando para um treinamento mais eficiente e aumento da acurácia na saída final [Zhou et al. 2020]. No método proposto, utiliza-se Attention Gates (AGs) [Oktay et al. 2018] que implementam uma sequência de operações de convolução e multiplicação elemento a elemento para implicitamente aprender a suprimir regiões irrelevantes em uma imagem de entrada, enquanto destaca características salientes úteis para a tarefa de segmentação. Os AGs estão imple-

Figura 3. Arquitetura AER-Net.

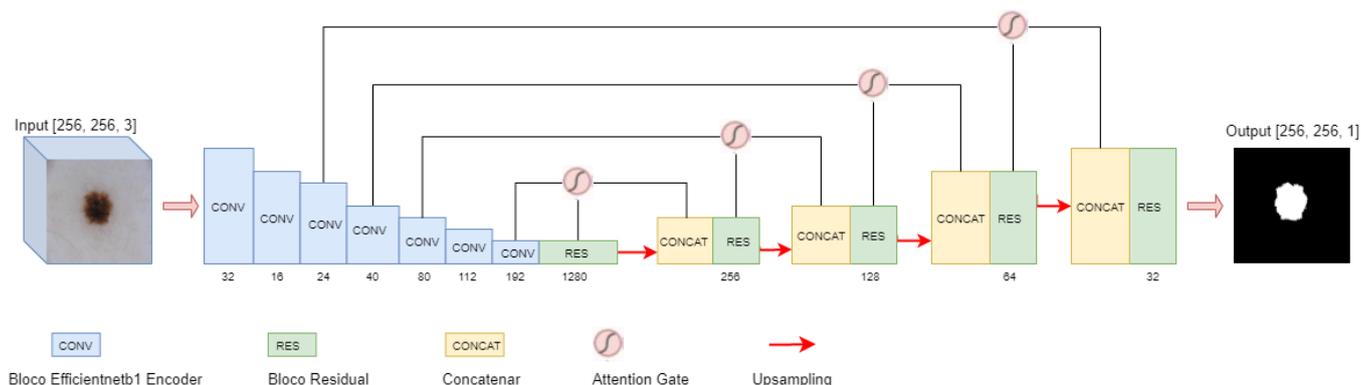
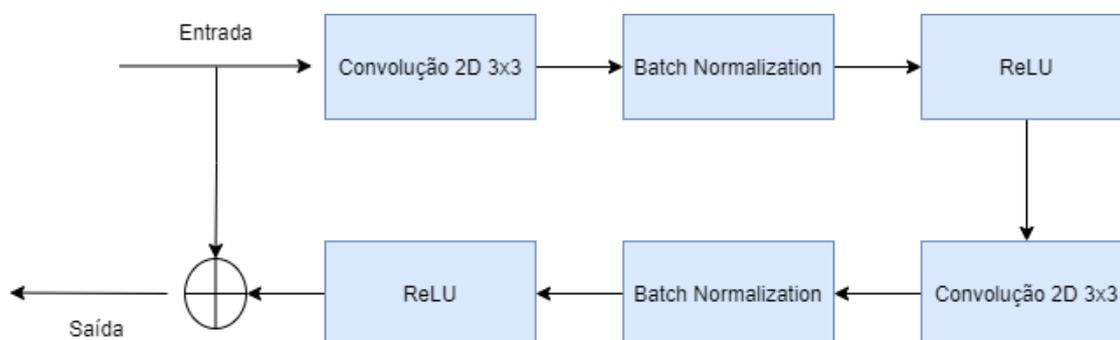


Figura 4. Bloco residual AER-Net.



mentados antes das operações de concatenação para mesclarem apenas as informações mais relevantes da imagem específica. Na Figura 5 está apresentado com detalhes as operações dos AGs como proposto por Oktay et al. [Oktay et al. 2018].

3.3. Treinamento

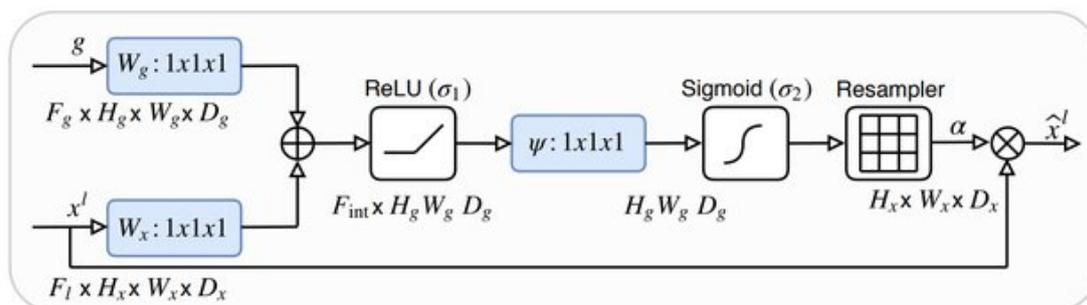
Para a etapa do treinamento, 80,72% das imagens foram destinadas para o treinamento da rede, totalizando 2094 imagens dermatoscópicas com suas máscaras de segmentação. Ainda para a realização dos testes, o método de otimização para treinamento da rede foi Adam [Kingma and Ba 2017] com taxa de aprendizado em 0.001 e definidos 100 épocas com batch de tamanho 16.

3.4. Função Loss

Neste estudo, a Dice Loss foi utilizada como a função loss para otimização do modelo. A Dice Loss irá buscar maximizar a sobreposição entre a máscara de saída e a máscara ground-truth e vem sendo utilizada com sucesso para trabalhos com classes desbalanceadas, bastante comuns na área médica onde a região de pixels de interesse ocupa apenas uma região muito pequena na imagem analisada [Milletari et al. 2016]. A função Dice Loss (DCL), portanto pode ser definida como:

$$DCL = 1 - \frac{2 \sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \quad (1)$$

Figura 5. Attention Gate.



onde as somas correm sobre os N pixels da máscara de segmentação binária de saída $p_i \in P$ e da máscara de segmentação ground truth $g_i \in G$.

4. Resultados e Discussão

4.1. Implementação

Os testes foram feitos no ambiente do Google Colaboratory [GOOGLE 2021], que é uma ferramenta para ensino e pesquisa na área de aprendizado de máquina, disponibilizando um ambiente de notebooks Jupyter executados na nuvem. A ferramenta conta com processador Intel(R) Xeon(R) CPU @ 2.30GHz e GPU Tesla T4. A linguagem de programação utilizada no desenvolvimento deste trabalho foi Python (versão 3.7). Além disso, foi utilizado a biblioteca Keras para a modelagem de todas as estruturas das CNNs utilizadas no trabalho.

4.2. Métricas de Avaliação

Índice Jaccard médio (JSI) e Coeficiente Dice (DC) foram as métricas utilizadas para avaliar os resultados de segmentação. Elas são definidas como:

$$JSI = \frac{TP}{TP + FP + FN} \quad (2)$$

$$DC = \frac{2 * TP}{2TP + FP + FN} \quad (3)$$

Onde TP, FP e FN representam os verdadeiros positivos, falsos positivos e falsos negativos, respectivamente.

Em segmentação semântica, para calcular o acerto da máscara de segmentação de saída de cada modelo (O) comparado às máscaras de segmentação ground-truth (G), as métricas foram definidas como:

$$JSI = \frac{|G \cap O|}{|G \cup O|} \quad (4)$$

$$DC = \frac{2|G \cap O|}{|G| + |O|} \quad (5)$$

Tabela 1. Backbones com melhor desempenho

Backbone	Índice Jaccard
EfficientNetb1	76.1%
Se-ResNeXt101	75.7%
EfficientNetb3	75.65%
SeResNet152	75,6%
EfficientNetb0	75,4%

Tabela 2. Experimentação para AER-Net

	Índice Jaccard	Coefficiente Dice
U-Net [Ronneberger et al. 2015]	55,5%	67,4%
Attention + U-Net	70%	79,6%
Attention + Residual connections + U-net	70,6%	79,9%
Attention + Efficientnetb1 + Residual + U-Net (AER-Net)	77,2%	85,25%

4.3. Estimação de Backbones

A transferência de aprendizado foi usada através de modelos de arquitetura pré-treinados no ImageNet [Deng et al. 2009]. Aproveitando os dados do primeiro dataset para extrair informações que possam ser úteis para o aprendizado da CNN [Bengio et al. 2017], o objetivo de usar a aprendizagem por transferência foi diminuir o tempo de treinamento e também resultar em menores erros de generalização. Para isso, foram realizados extensivos experimentos em várias redes pré-treinadas: ResNet, VGG, EfficientNet, DenseNet, Inception, MobileNet, SeNet, SE-ResNeXt, ResNeXt. Os resultados em método jaccard para similaridade das máscaras produzidas pelas 5 melhores redes estão dispostos na Tabela 1.

Efficientnetb1 obteve os melhores resultados para a tarefa de segmentação, portanto foi escolhido como encoder da arquitetura proposta.

4.4. Resultados

O modelo apresentado AER-Net é baseado na teoria proposta pela Attention U-net, que por sua vez é uma alteração com o intuito de melhorar da rede U-Net. As principais características distintivas desta proposta para a U-Net tradicional estão na substituição das skip connections com Attention Gates, uso de blocos de convoluções contendo conexões resíduas e um encoder pré-treinado. Na tabela 2, é apresentado cada componente mencionado e o impacto na performance em índice Jaccard (JSI) e Coeficiente Dice (DC) em relação a hipótese original, U-Net. A AER-Net obteve um JSI médio de 77.2% e DC médio de 85.25% a respeito das máscaras de segmentação de saída da rede comparado com as máscaras ground-truth na base ISIC 2018, portanto superando os resultados da U-Net clássica.

Como inferido pela Tabela 2, a adição de Attention Gates (AGs), seguida pela aplicação de conexões resíduas nos blocos de convolução e sobretudo o uso da rede pré-treinada Efficientnetb1 como a arquitetura encoder foram cruciais para aprimorar o desempenho final da AER-Net nas métricas de JSI e DC. Já na tabela 3, é apresentado o modelo AER-Net comparado com outros métodos na literatura. A AER-Net apresen-

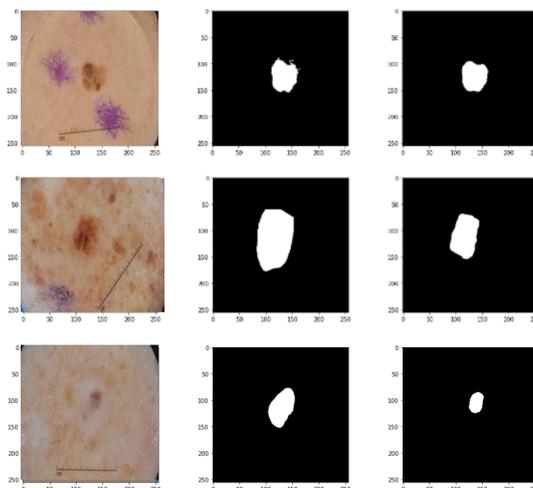
Tabela 3. Comparação de diferentes métodos no dataset ISIC 2018

Autor	Método	Jaccard	Dice
Souza et al. [Souza et al. 2020]	U-Net	68.0%	77.2%
Ji et al. [Ji et al. 2018]	baseado em Resnet34	79.51%	-
Amin et al. [Amin et al. 2020]	VGG-16	85.0%	82.0%
Oktay et al. [Oktay et al. 2018]	Attention U-Net	70%	79.6%
Método Proposto	AER-Net	77.29%	85.25%

tada mostrou resultados promissores em Coeficiente de Dice ultrapassando metodologias propostas como alguns dos métodos melhores colocados do desafio ISIC 2018 que não implantaram técnicas de ensemble, e superando o desempenho da Attention U-Net.

A Figura 6 apresenta exemplos de saídas de máscaras de segmentação da AER-Net proposta em comparação com a máscara ground-truth, exibindo os exemplos de saída de qualidade boa, média e ruim, respectivamente em cada linha. Na primeira coluna está representada a imagem RGB original. Na segunda coluna está representado a máscara ground-truth. A terceira coluna contém a saída da máscara de segmentação da rede AER-Net.

Figura 6. Exemplo de saída de segmentação na AER-Net.



Para a definição uma saída de qualidade boa na primeira linha da Figura 6, foi extraída uma máscara de segmentação de saída com Coeficiente Dice 96%, a qual mostra-se que o modelo conseguiu contornar bem a área da lesão respeitando as suas bordas. Para uma saída de qualidade média na segunda linha, é mostrada uma máscara que obteve 68% de Dice. Para esta imagem, a rede conseguiu segmentar bem a região central da lesão que equivale a região da mancha com cores mais fortes na imagem original. Porém, de acordo com a segmentação feita pelo especialista, a lesão inteira se estende pela pele até por regiões em que o contraste entre a lesão e a pele sejam muito baixos. Como resultado, a máscara de segmentação gerada obteve muitos falsos negativos nesta região em volta a mancha mais escura. Já para a terceira linha representando uma saída de qualidade ruim, uma máscara com Dice 45%, a rede apresentou resultado não satisfatório. Esta lesão de melanoma contém uma mancha de cor diferente, e com bordas com contrastes bastante

suavizados entre a pele e lesão. Por sua vez, a máscara final gerada respeitou o contorno da área da mancha mais escura e suas bordas, porém segmentando errado a região da lesão com tonalidade muito clara e sem contraste evidente entre pele e lesão.

5. Conclusão

Com o aumento de casos de câncer de pele em todo o mundo nos últimos anos, faz-se cada vez mais necessário o uso de ferramentas como CNNs aliadas a dermatoscopia para ajudar a medicina no diagnóstico mais rápido e eficiente, pois quanto mais cedo a diagnose, maior a chance de cura.

Neste trabalho, foi proposto um novo modelo de rede neural para segmentação de imagens, chamado de AER-Net. Baseado numa modificação da U-Net com Attention Gates e conexões residuais, apresentou métricas que superaram os outros modelos avaliados e também outros estudos no campo, se mostrando confiável para o uso na segmentação de lesões de pele.

Além disso, uma estimacão de backbones foi feita para avaliar o melhor método para a extração de características que serviria como o encoder das arquiteturas de CNNs avaliadas. Tal estimacão apontou que os backbones da família Efficientnet podem ser ótimos aliados na tarefa de segmentação automática de lesões cutâneas.

Para trabalhos futuros, propõe-se a continuacão do estudo da AER-Net, empregando diferentes tipos de aumento de dados e pré-processamento para melhorar o conteúdo dos dados como, ajuste de contraste, novos canais de cores, suavizacão e outras técnicas.

Referências

- Al Nazi, Z. and Abir Tasnim, A. (2020). Automatic skin lesion segmentation and melanoma detection: Transfer learning approach with u-net and dcnn-svm. In *Proceedings of International Joint Conference on Computational Intelligence*, pages 371–381. Springer.
- Amin, J., Sharif, A., Gul, N., Anjum, M. A., Nisar, M. W., Azam, F., and Bukhari, S. A. C. (2020). Integrated design of deep features fusion for localization and classification of skin cancer. *Pattern Recognition Letters*, 131:63–70.
- Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495.
- Bengio, Y., Goodfellow, I., and Courville, A. (2017). *Deep learning*, volume 1. MIT press Massachusetts, USA:.
- Codella, N., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., et al. (2019). Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*.
- Deng, J., Dong, W., Socher, R., Li, L., Kai Li, and Li Fei-Fei (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255.

- for Cancer Research, W. C. R. F. I. (2018). Diet, nutrition, physical activity and cancer: a global perspective. Acesso em Janeiro 17, 2021.
- Ge, Z., Demyanov, S., Chakravorty, R., Bowling, A., and Garnavi, R. (2017). Skin disease recognition using deep saliency features and multimodal learning of dermoscopy and clinical images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 250–258. Springer.
- GOOGLE (2021). Google colab. Acesso em Abril 19, 2021.
- 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*, pages 770–778.
- INCA (2021). Inca homepage. Acesso em Janeiro 17, 2021.
- Ji, Y., Li, X., Zhang, G., Lin, D., and Chen, H. (2018). Automatic skin lesion segmentation by feature aggregation convolutional neural network. *Technical report*.
- Kingma, D. P. and Ba, J. (2017). Adam: A method for stochastic optimization.
- Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.
- Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation.
- Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., and Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520.
- Society, A. C. (2021). Cancer Facts and Figures 2021. Acesso em Janeiro 17, 2021.
- Souza, L., Lelis, S., and Silva, R. (2020). Segmentação de lesões de pele utilizando algoritmos de aprendizagem profunda. In *Anais do XX Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 344–355. SBC.
- Tan, M. and Le, Q. V. (2020). Efficientnet: Rethinking model scaling for convolutional neural networks.
- Tschandl, P., Rosendahl, C., and Kittler, H. (2018). The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *scientific data* 5, 180161.
- Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. (2020). Unet++: Redesigning skip connections to exploit multiscale features in image segmentation.