UNIVERSIDADE FEDERAL DO MARANHÃO

Curso de Graduação em Ciência da Computação

José Guilherme Pereira Lima

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

São Luís - MA

2021

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Lima, José Guilherme Pereira.
Attention Efficient Residual U-Net: uma Rede Neural para a Segmentação de Lesões de Pele / José Guilherme Pereira Lima. - 2021.
39 f.

Orientador(a): Geraldo Braz Júnior.
Monografia (Graduação) - Curso de Ciência da Computação, Universidade Federal do Maranhão, São Luis, 2021.

1. Fully convolutional network. 2. Melanoma. 3. Segmentação semântica. I. Júnior, Geraldo Braz. II. Título.

José Guilherme Pereira Lima

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

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Geraldo Braz Júnior

São Luís - MA

2021

José Guilherme Pereira Lima

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

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Trabalho _____. São Luís - MA, 29 de Abril de 2021:

Prof. Dr. Geraldo Braz Júnior
Orientador
Universidade Federal do Maranhão

Prof. Msc. Italo Francyles Santos da Silva
Universidade Federal do Maranhão

Prof. Dr. João Dallyson Sousa de Almeida
Universidade Federal do Maranhão

São Luís - MA
2021

À minha família e meus amigos.

Agradecimentos

Primeiramente aos meus pais, Ana e Fábio, por me proporcionarem desde cedo todos os meios para que tenha oportunidade a uma boa educação.

Ao restante da minha família pelo incentivo.

Ao meu orientador, Geraldo Braz, que sem os ensinamentos e puxões de orelha este estudo não seria possível.

Aos professores do curso de Ciência da Computação pela contribuição na minha formação.

Ao meu amigo do curso e da vida Alfredo, que me ajudou incontáveis vezes ao longo dessa jornada acadêmica.

A Yago pelo apoio, parceria e incentivo.

Aos meus amigos Altivo, Rick, Rafael, Jake e Flávio.

Aos parceiros do VipLab.

A UFMA.

A Fapema pelo apoio financeiro a este trabalho.

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 tem como objetivo comparar arquiteturas de redes neurais convolucionais populares e propor uma 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 optimização. O modelo proposto, chamado de AER-Net (simplificado de Attention Efficient Residual Unet), apresentou 87,7% e 79.5% 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.

Palavras-chave: Melanoma, fully convolutional network, segmentação semântica.

Abstract

Melanoma is one of the most severe skin cancer types due to its high mortality rate, which can achieve 70%. An early diagnosis of the disease is crucial as it increases the ten-year survival rate up to 97%. The segmentation of skin lesions is one of the essential steps of the diagnosis process for accurate melanoma detection. However, even for specialist doctors, segmenting these lesions is costly and challenging due to the wide variety of stains, which can have irregular edges, different dimensions, and colors, and due to the high amounts of exams to analyze. Automatic detection of the lesion area for segmentation proves to be an important area of study so that the specialist doctor can focus on the correct diagnosis of the disease itself. This work aims to compare popular convolutional neural network architectures and to propose a new model based on encoder-decoder architectures that reach parameters of efficiency and performance for the segmentation of dermoscopic images. Thus, a method based on Attention U-net was proposed, using pre-trained encoder and residual blocks for better optimization. The proposed model, called AER-Net (short for Attention Efficient Residual Unet), presented 87.7% and 79.5% in Dice Coefficient and Jaccard index for ISIC Archive dataset and proved that the model can assist in the process for automatic skin cancer diagnosis.

Keywords: Melanoma, fully convolutional network, semantic segmentation.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

ABCD	Assimetry Border Color Diameter
AG	Attention Gate
BFL	Binary Focal Loss
BN	Batch Normalization
CNN	Convolutional Neural Network
CPU	Central Process Unit
DC	Dice Coefficient
DCL	Dice Coefficient Loss
FPN	Feature Pyramid Network
GPU	Graphics Process Unit
ISIC	International Skin Imaging Collaboration
JSI	Jaccard Similarity Index
PCA	Principle Component Analysis
ReLU	Rectified Linear Units
RGB	Red Green Blue
SWA	Stochastic Weight Averaging
VGG	Visual Geometry Group

Sumário

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 (RESEARCH, 2018), é de longe o mais perigoso porque tem mais probabilidade de crescer e espalhar. 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 automaticamente detectar esses ferimentos para ajudar os profissionais médicos. 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; FISCHER; BROX, 2015). A maioria das abordagens de segmentação de imagens médicas por CNNs alcançaram resultados de ponta através do desenvolvimento de arquiteturas encoder-decoder para sua rede neural convolucional profunda, por exemplo, Segnet (BADRINARAYANAN; KENDALL; CIPOLLA, 2017), U-net (RONNEBERGER; FISCHER; BROX, 2015), e rede totalmente convolucional (LONG; SHELHAMER; DARRELL, 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 comparação e análise de diferentes métodos de CNNs para apontar, justificar quais modelos têm melhor desempenho na tarefa de segmentação de lesões de pele e por fim, sugerir um modelo novo de CNN que possa ser eficaz na segmentação de imagens médicas.

1.1 Objetivo

Este trabalho tem como objetivo avaliar múltiplas arquiteturas encoder-decoder de redes neurais convolucionais profundas e propor uma nova CNN com eficácia comprovada para a segmentação automática de lesões cutâneas baseado na análise de resultados.

1.1.1 Objetivos específicos

- Empregar alguns dos mais populares tipos de arquiteturas de CNNs profundas que são amplamente utilizadas na comunidade de visão computacional para segmentação semântica de imagens dermatoscópicas de lesões cutâneas.
- Analisar o desempenho dos modelos de CNNs implementados e identificar um modelo de CNNs de alto desempenho por estudos quantitativos através de métricas estado da arte, e estudos qualitativos em relação a segmentação de melanoma.
- Propor um novo modelo de CNN baseado nos melhores métodos avaliados, ajustar hiperparâmetros da nova rede e constatar que tal modelo pode ser usado com eficácia na segmentação automática de imagens para auxiliar no diagnóstico de câncer de pele.

2 Trabalhos Relacionados

Diversos trabalhos na literatura realizaram a detecção automática da área da lesão de pele para segmentação por meio de técnicas computacionais. A seguir, é apresentado um conjunto de trabalhos que realizam a segmentação automática das lesões cutâneas.

Nazi e 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, Lelis e Silva (2020).

Por outro lado, Tang et al. (2019) propõe a separação de blocos convolucionais em conjunto com a arquitetura U-Net para a extração de características correlacionadas e com semântica alta. Um esquema baseado em média de peso estocástico (SWA) é aplicado para evitar sobreajuste.

A metodologia aplicada por 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.

Xie et al. (2020) propôs uma nova técnica de arquitetura de Rede Neural Convolucional (CNN) para segmentação de lesões cutâneas. Consequentemente, mapas de características de alta resolução são gerados para recuperar limites de lesões de pele mais precisos usando um bloco de recursos de alta resolução (HRFB). Para a segmentação da lesão cutânea, Amin et al. (2020) propõe o uso da arquitetura de rede Alexnet e VGG-16 para extração de recursos. Em sequência, fusão e seleção de recursos usando PCA. A Tabela 1 apresenta um sumário dos trabalhos citados, base de dados utilizadas e resultados dos métodos.

Neste trabalho, procura-se propor uma nova arquitetura de rede neural convolucional para a segmentação de imagens médicas de lesões de pele baseado em estruturas encoder-decoder, como a U-Net. O modelo sugerido busca tirar vantagem do aprendizado de transferência utilizando uma rede pré-treinada da família Efficientnet como encoder, assim como tentativa de mitigar o problema da degradação no treinamento de redes profundas utilizando blocos residuais (HE et al., 2016) e a adição de Attention Gates (AGs) para segmentação de imagens (OKTAY et al., 2018) para destacar características salientes que são transmitidas através das skip connections. Os detalhes da nova arquitetura são apresentados na Seção 4.3.5.

Tabela 1 – Sumário de trabalhos relacionados

Work	Method	Dataset	Dice Coefficient	Jaccard Index
Tang et al. (2019)	Separable-UNet	ISIC 2016	86.9%	79.2%
Xie et al. (2020)	CNN-HRFB	ISIC 2016	91.8%	85.8%
Xie et al. (2020)	CNN-HRFB	ISIC 2017	86.2%	78.3%
Tang et al. (2019)	Separable-UNet	ISIC 2017	93.3%	89.2%
Souza, Lelis e Silva (2020)	U-Net	ISIC 2018	77.2%	68.0%
Ji et al. (2018)	FA-CNN	ISIC 2018	-	79.5%
Nazi e Tasnim (2020)	U-Net	ISIC 2018	87.0%	80.0%
Amin et al. (2020)	VGG16/Alexnet	ISIC 2018	82.0%	85.0%

3 Fundamentação Teórica

Este capítulo tem por finalidade estabelecer os conceitos teóricos usados no trabalho e que são necessários para compreensão das técnicas utilizadas na metodologia proposta.

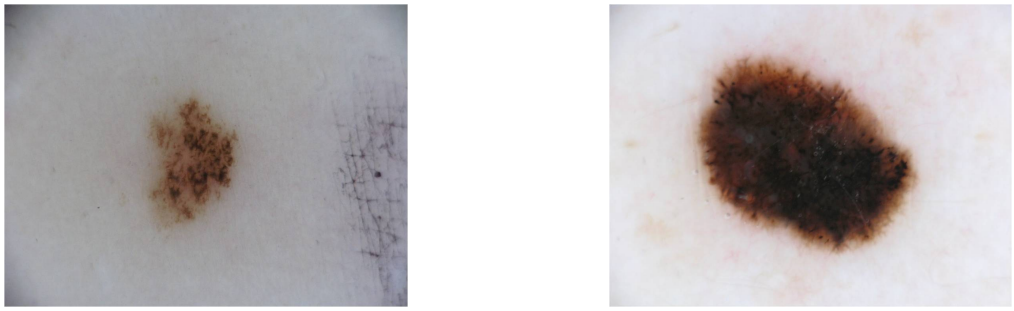
3.1 Dermatoscopia

A dermatoscopia é um tipo de exame dermatológico não invasivo que tem como objetivo fazer a análise da pele de forma mais detalhada, sendo útil na avaliação e diagnóstico de alterações cutâneas, como o câncer de pele. A técnica é feita através de um aparelho chamado dermatoscópio, semelhante a um microscópio binocular com uma fonte de luz embutida para o exame da pele (RP; JH; LE, 2004). A dermatoscopia é mais precisa do que o exame a olho nu para o diagnóstico de melanoma cutâneo em lesões suspeitas quando realizada em ambiente clínico (VESTERGAARD et al., 2008). A técnica aliada ao diagnóstico clínico pode aumentar em até 9% a precisão do diagnóstico correto de melanoma, em comparação com a avaliação clínica sozinha (M et al., 1994).

A aplicação da dermatoscopia é feita através do uso de vários líquidos de imersão (óleos, álcool, água ou gel) para tornar a superfície da pele translúcida e reduzir o reflexo, assim as estruturas morfológicas da pele são facilmente visíveis (RP; JH; LE, 2004), em seguida a pele é observada através do dermatoscópio.

Na Figura 1 tem-se um exemplo de duas imagens geradas por um dermatoscópio, sendo uma benigna e a outra cancerosa. Com a lesão ampliada e seus contornos e detalhes preservados, é possível na prática, reduzir o número de biópsias desnecessárias quando lesões benignas são diagnosticadas como malígnas e também melanomas que podem permanecer sem diagnóstico ou serem diagnosticados tarde demais.

Figura 1 – Exemplo de visualização de uma lesão de pele em dermatoscopia



(a) Exemplo de uma lesão benígna. (b) Exemplo de uma lesão malígna.

Fonte – ISIC Archive

O melanoma inicial pode ser reconhecido clinicamente utilizando a regra ABCD. As iniciais significam: Assimetria - muitas lesões recentes crescem de forma desigual, resultando em um padrão assimétrico; Bordas - o crescimento desigual também resulta em bordas irregulares; Cor - lesão sem uniformidade de tons, podendo possuir duas ou mais tonalidades como variâncias de escuro, claro, castanho, vermelho e azuis; Diâmetro - lesões com área da lesão de diâmetro maior que 5 mm podem ser consideradas suspeitas para o melanoma (SOARES, 2008).

3.2 Redes Neurais Convolucionais

As Redes Neurais Convolucionais (CNNs) são redes neurais profundas especializadas no processamento de dados que possuem uma topologia semelhante a uma grade, como por exemplo imagens digitais que são representadas por grades 2D de pixels. (GOODFELLOW; BENGIO; COURVILLE, 2016). O uso de CNNs para tarefas de reconhecimento visual foi primeiramente proposto por (Lecun et al., 1998) para resolver o problema de automatização do reconhecimento de números manuscritos. O uso de CNNs foi expandido e é comumente utilizado com sucesso para tarefas como classificação de imagens, segmentação de imagens, análise de imagens médicas e processamento de linguagem natural (COLLOBERT; WESTON, 2008).

A arquitetura básica de uma CNN é formada por três tipos de camadas, sendo: a camada de convolução, a de subamostragem (pooling) e a camada totalmente conectada (fully connected layer) (GU et al., 2017). Estas três camadas serão brevemente descritas a seguir nas subseções 3.2.1, 3.2.2 e 3.2.3 respectivamente.

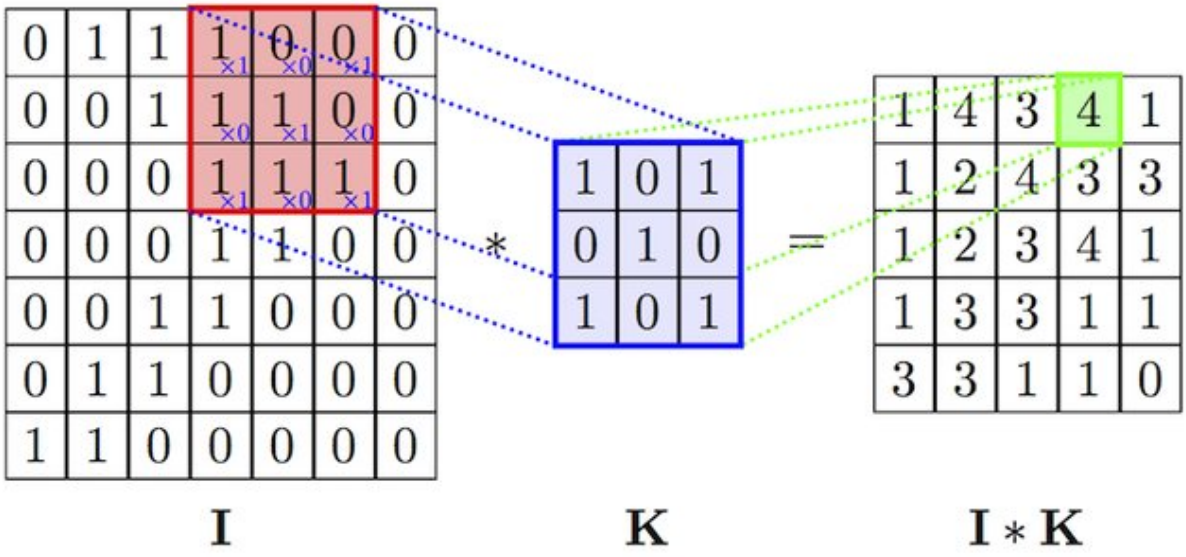
3.2.1 Camada de convolução

A camada de convolução é a camada central e mais importante de uma CNN e a qual faz maior parte do processamento computacional. Os dois parâmetros desta camada consistem da imagem de entrada (input) e um conjunto de filtros ou kernels. E a saída gerada é referida como um mapa de ativação (GOODFELLOW; BENGIO; COURVILLE, 2016).

Os filtros de uma camada convolucional é uma matriz de números são aplicados através de uma janela deslizante que percorre toda a imagem de entrada calculando o produto escalar entre os pesos do filtro e os pixels correspondentes da entrada. Esta operação é conhecida como convolução.

Em uma convolução, os filtros convolucionais devem se deslocar ao longo dos eixos x e y da matriz da imagem de entrada de acordo com um número pré-definido de pixels conhecido como passos (stride). O stride especificará a quantidade de passos ser deslocada a cada operação de convolução. A operação de convolução a convolução multiplica cada

Figura 2 – Exemplo de operação de convolução.



Fonte – (MOHAMED, 2017)

elemento do filtro por um elemento da imagem de entrada, e depois faz uma soma desse valores, para gerar um elemento do mapa de ativação como saída. Após cada filtro for convolvido com a imagem de entrada, será gerado um mapa de ativação de saída. A Figura 2 mostra a representação visual de um exemplo de convolução. Nela, é aplicado um filtro K 3x3, que percorre a imagem de entrada I, produzindo um mapa de ativação I * K.

Normalmente logo após a convolução aplica-se uma função de ativação não-linear. Uma função de ativação desempenha um papel importante na rede, decidindo quais pixels devem ser ativados em uma determinada camada. Um exemplo de função de ativação é a ReLU (do inglês Rectified Linear Units) que tem por objetivo retornar 0 se o input for negativo, porém se o valor do input for qualquer número positivo x, a função retorna o valor x sem modificação.

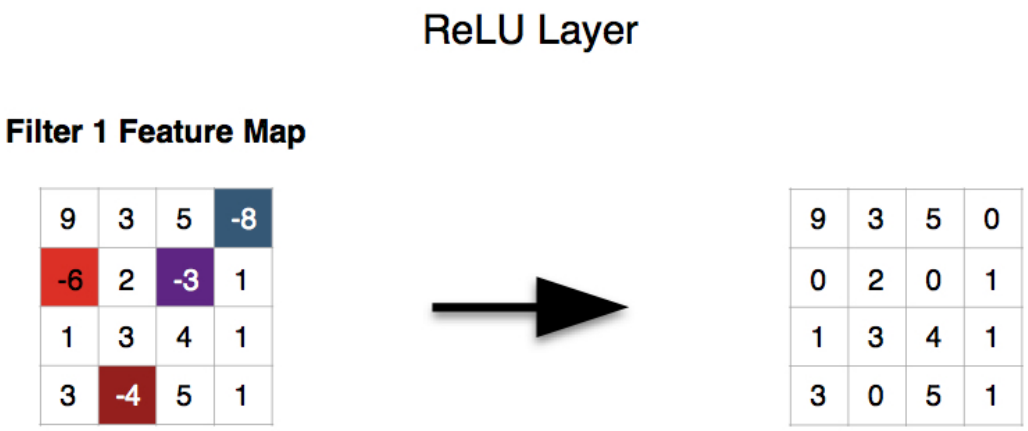
$$f(x) = max(0, x) \tag{3.1}$$

Matematicamente dada pela Equação 3.1, esta função é aplicada para cada pixel do mapa de ativação de forma a retificar os valores negativos. A figura demonstra visualmente como a função ReLU opera gerando um mapa de ativação sem pixels negativos. Este mapa final gerado será enviado às próximas camadas da rede neural convolucional.

3.2.2 Camada de pooling

A próxima camada chamada de pooling (em tradução livre, agrupamento) tem o objetivo de reduzir a resolução espacial da imagem gerada pela camada de convolução. Com o pooling há uma redução no custo computacional da rede e também evita-se o

Figura 3 – Aplicação da função de ativação ReLU em um mapa de ativação 4x4.

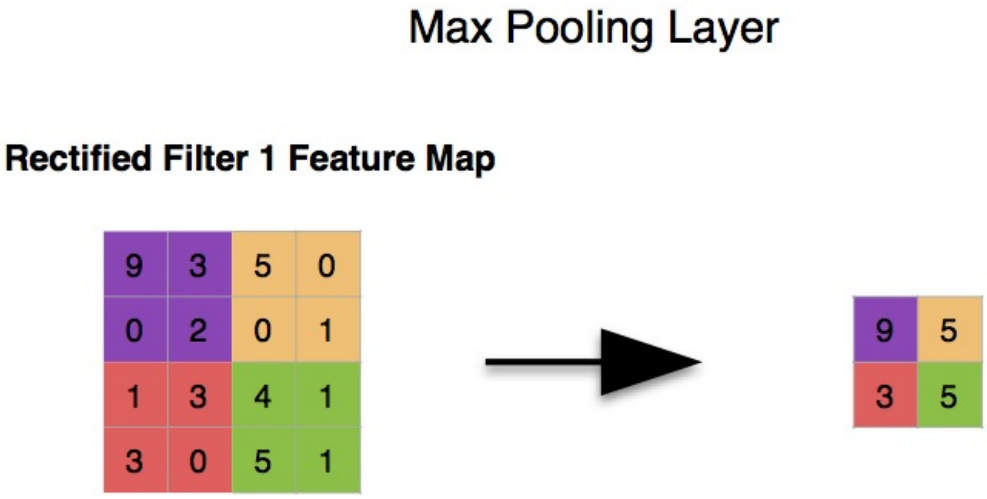


Fonte – (ANIEMEKA, 2017)

overfitting. Uma função de pooling substitui a saída da camada convolucional em um determinado local por uma estatística resumida das saídas próximas (GOODFELLOW; BENGIO; COURVILLE, 2016).

Um exemplo de função de pooling popular é a max pooling. Esta abordagem consiste em substituir os valores de uma região determinada pelo valor máximo, como demonstrado na Figura 4.

Figura 4 – Aplicação de max pooling utilizando kernel 2x2 e stride de 2.

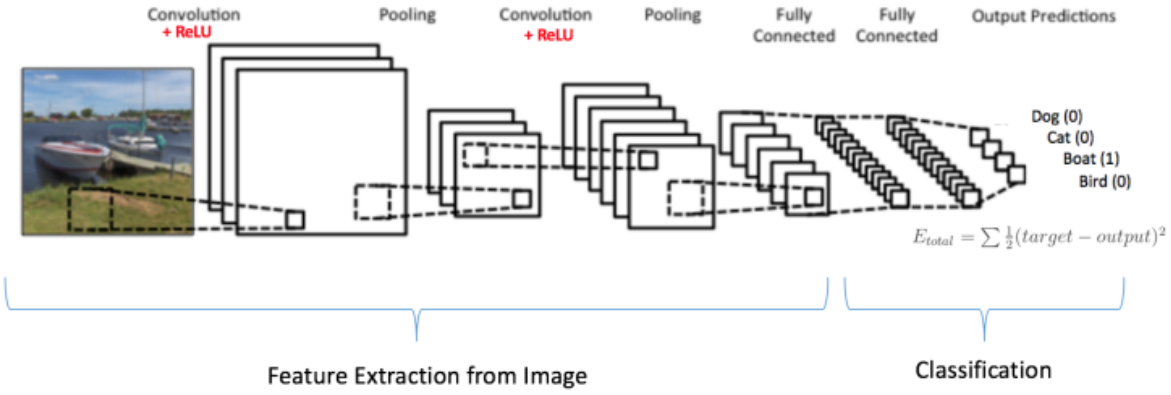


Fonte – Adaptado de (ANIEMEKA, 2017)

3.2.3 Camada totalmente conectada

As camadas totalmente conectadas estão localizadas após sucessivas camadas de convolução e pooling e tem o objetivo de classificar as entradas em uma das classes determinadas. Nesta camada, a CNN retorna a probabilidade de que um objeto em uma foto seja de um determinado tipo (ANIEMEKA, 2017).

Figura 5 – Representação visual de uma arquitetura Rede Neural Convolucional e suas camadas



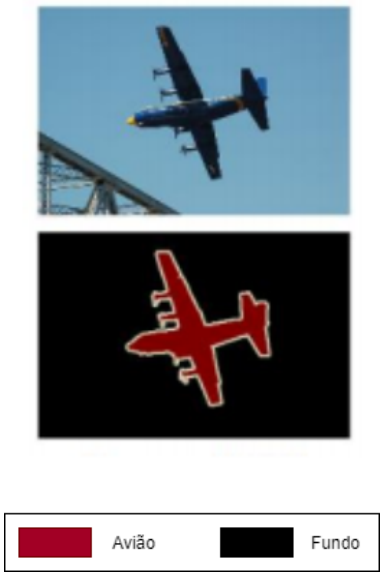
Fonte – (KARN, 2016)

O termo totalmente conectado se dá ao fato de que cada neurônio na camada anterior está ligado a cada neurônio na próxima camada, adicionando uma camada de saída com o número de neurônios equivalente ao número de classes determinadas no experimento para realizar a classificação aplicando algum tipo de função de ativação (KARN, 2016). A Figura 5 apresenta uma representação de uma arquitetura de CNN com a camada totalmente conectada no fim retratando a probabilidade da imagem ser rotulada para cada classe determinada.

3.3 Segmentação Semântica

A segmentação semântica é uma etapa do processamento de imagens que consiste em separar e distinguir diferentes partes de uma imagem de acordo com características similares. Sendo uma imagem representada digitalmente por uma matriz de pixel, o objetivo de uma segmentação semântica é atribuir a cada pixel da imagem uma classe correspondente. A Figura 6 mostra uma segmentação semântica contendo a classe avião e a classe do fundo da imagem. Os pixels analisados e classificados como representando o avião estão na cor vinho, enquanto os pixels classificados como fundo da imagem que não pertencem ao avião estão na cor preto.

Figura 6 – Exemplo de Segmentação Semântica.



Fonte – Traduzido e adaptado de (CHEN et al., 2016)

3.3.1 U-Net

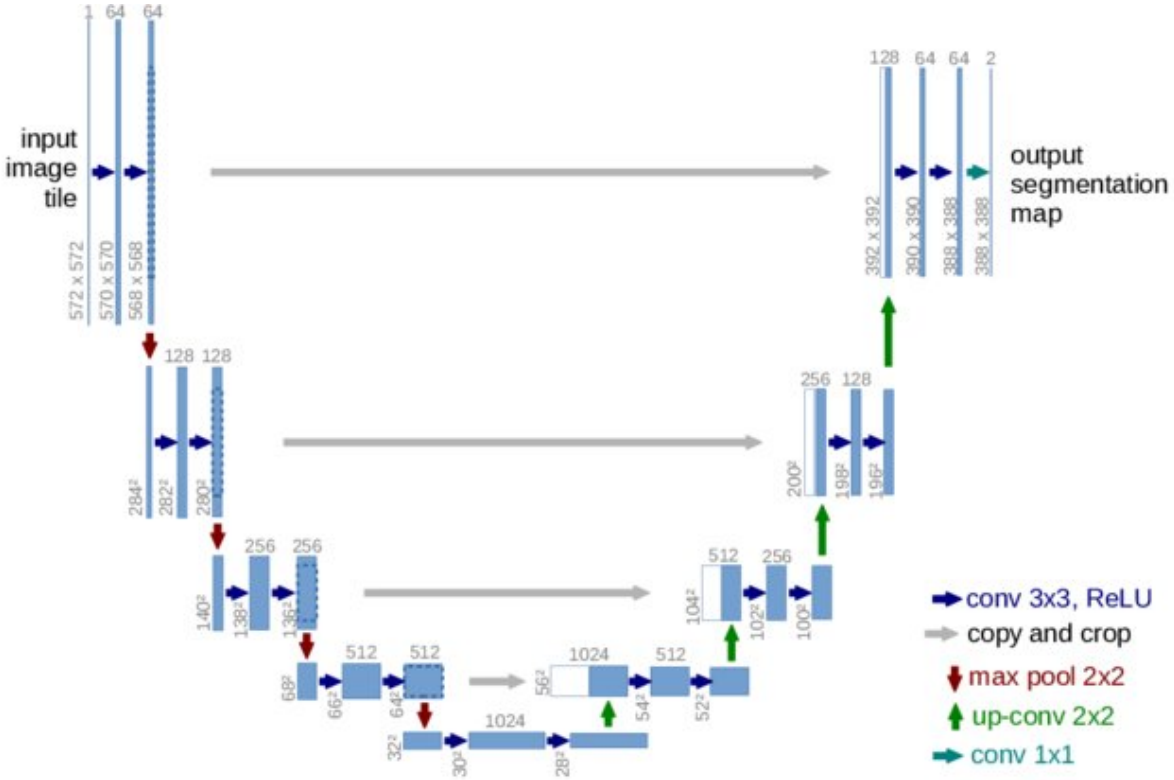
A U-Net é uma rede totalmente convolucional modificada e estendida para que ela possa trabalhar com muitas poucas imagens de treinamento e produzir segmentações mais precisas (RONNEBERGER; FISCHER; BROX, 2015). A arquitetura da U-net é mostrada na Figura 7. É composta de duas partes, dividas em quatro etapas cada gerando um formato de U em sua arquitetura. A primeira parte é a extração de recursos consistindo na aplicação de operações repetidas de convolução com kernel de 3x3, cada etapa seguida por uma função de ativação Rectified Linear Unit (ReLU) e uma operação de Max Pooling de 2x2 com stride = 2, e a segunda parte é o upsampling, onde os operadores de upsampling substituem os operadores de pooling. Essas camadas são responsáveis por aumentar a resolução da saída. Para uma melhor localização, os mapas de recursos do caminho de contratação são combinados com a saída após o upsampling.

3.3.2 Linknet

O Linknet é uma rede totalmente convolucional, de tamanho médio e projetada para aprender informações relevantes sem qualquer aumento significativo no número de parâmetros. Linknet visa contornar informações espaciais diretamente do encoder para o decoder correspondente, a fim de melhorar a precisão junto com uma diminuição significativa no tempo de processamento (CHAURASIA; CULURCIELLO, 2017).

A Figura 8 mostra a arquitetura da rede Linknet. Sendo uma evolução da U-Net, o primeiro bloco do modelo realiza uma convolução na imagem de entrada usando um kernel

Figura 7 – Arquitetura da U-net.



Fonte – (RONNEBERGER; FISCHER; BROX, 2015)

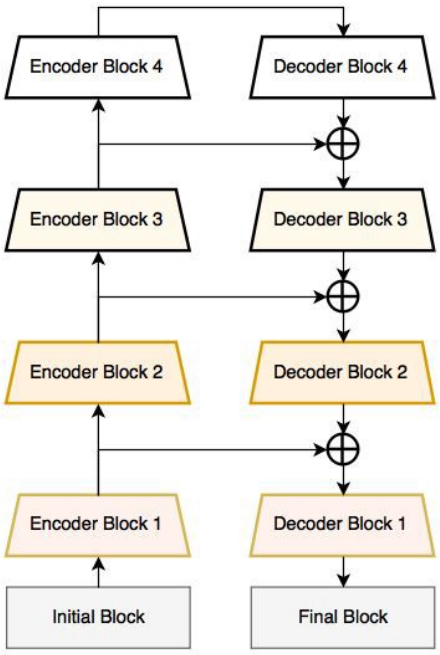
de tamanho 7x7 com stride = 2. Ele é seguido por uma camada de Max Pooling stride = 2. O próximo conjunto de camadas é formado por uma série de blocos com conexões residuais. Em cada bloco residual, a camada convolucional inicial terá um stride = 2 fazendo a redução da imagem de entrada e o restante das camadas convolucionais no respectivo bloco terá stride = 1. Além disso, o modelo consiste em uma série de blocos decoder. O bloco decoder faz uma operação de convolução 1x1 que é seguida por Batch Normalization (BN) e convoluções transpostas para o upsampling de mapas de recursos obtidos nos blocos anteriores.

3.3.3 FPN

Feature Pyramid Network (FPN) é um extrator de recursos que visa gerar várias camadas de mapas de recursos através de três partes principais: uma via de baixo para cima, uma via de cima para baixo e conexões laterais (Lin et al., 2017).

A via de baixo para cima funciona como o encoder da rede e tem o objetivo de gerar diferentes mapas de recurso a cada nível convolucional para serem usados nas conexões laterais. Cada conexão lateral mescla mapas de recursos de mesmo tamanho espacial da via de baixo para cima e do caminho de cima para baixo, após aplicar uma operação de convolução 1x1. Na via de cima para baixo que atua como o decoder da rede, nos mapas

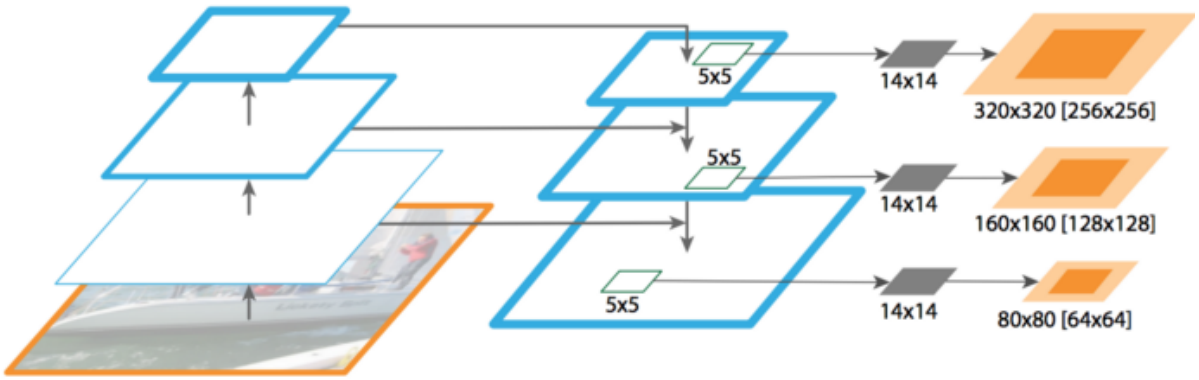
Figura 8 – Arquitetura da Linknet.



Fonte – (CHAURASIA; CULURCIELLO, 2017)

de recurso semanticamente mais fortes são aplicados operações de upsampling com stride = 2 e em seguida concatenados com os mapas de recursos do encoder via as conexões laterais. Cada mapa gerado, no fim, é acrescenteado uma convolução de kernel 3×3 e mesclados para gerar o mapa de segmentação final. A Figura 9 apresenta a arquitetura da rede FPN.

Figura 9 – Arquitetura da FPN.



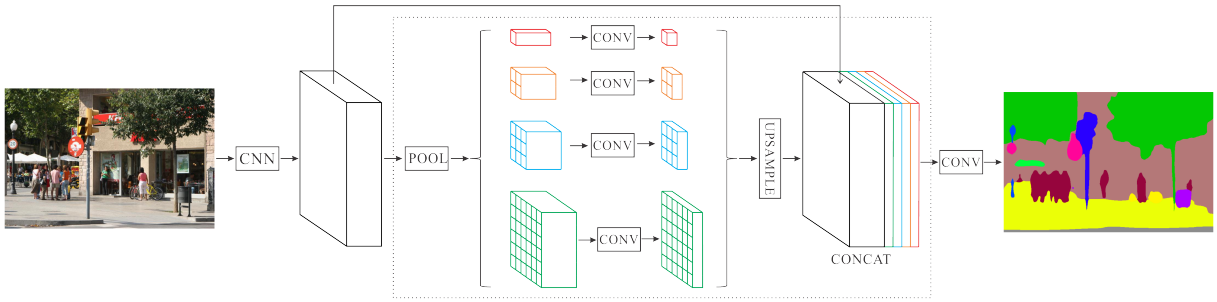
Fonte – (Lin et al., 2017)

3.3.4 PSP-Net

PSP-Net é um modelo de segmentação semântica bem conhecido que utiliza um módulo de análise de pirâmide que explora informações de contexto global por agregação de contexto baseada em regiões diferentes (ZHAO et al., 2017).

A Figura 10 mostra uma representação da arquitetura da rede PSP-Net. Dada uma imagem como entrada, o PSPNet utiliza uma CNN com convoluções dilatadas nos últimos blocos para extrair o mapa de recursos. O tamanho final do mapa de recursos é 1/8 do tamanho imagem de entrada. Após, utiliza-se o módulo de pooling da pirâmide, a parte principal da arquitetura da rede, para reunir informações de contexto. Nele, é aplicado diferentes operações de convolução para gerar mapas de recursos de diferentes níveis. Usando a pirâmide de 4 níveis, os núcleos agrupados cobrem toda, a metade e também pequenas porções da imagem. Estes mapas de recursos são seguidos por upsampling e concatenados com o mapa anterior original. Este mapa mesclado é seguido por uma camada de convolução para gerar o mapa de previsão final.

Figura 10 – Arquitetura da PSP-Net.



Fonte – (ZHAO et al., 2017)

4 Metodologia

Neste capítulo são apresentadas as etapas da metodologia de estudo implantadas para o desenvolvimento do trabalho. Logo após, cada etapa é explanada. Em primeiro lugar, são apresentadas as bases de imagens dermatoscópicas utilizadas. Em seguida, a etapa de testes e escolha do backbone para as CNNs propostas. Em seguida, as CNNs implementadas para avaliação. Após, são apresentadas as métricas utilizadas para a avaliação dos métodos. E finalmente, é apresentado a proposta de um novo modelo de arquitetura de CNN para segmentação semântica. Esta sequência de etapas da metodologia do trabalho é apresentada na Figura 11.

Figura 11 – Fases metodológicas do estudo.

Fonte – Elaborada pelo autor.

4.1 Aquisição de Dados e Pré-processamento

Duas bases de dados foram utilizadas para este estudo. A primeira, o ISIC 2018 dataset (CODELLA et al., 2019) (TSCHANDL; ROSENDAHL; KITTLER, 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 com suas máscaras de segmentação ground-truth. A segunda base de dados tirada diretamente do ISIC Archive, que será chamada de ISIC autoral, foi extraída através de um script automático proposto por (AVINERI; TALMOR, 2016), pelo qual obteve 4000 imagens dermoscópicas de lesão de pele com suas máscaras de segmentação ground-thruth. Destas imagens 3000 representam lesões de pele benígnas e 1000 lesões que representam câncer de pele malígno.

Para cada imagem no banco de dados, estão disponíveis também a segmentação manual feita por um especialista, como é demonstrado nas Figuras 12 e 13.

Figura 12 – Exemplo de lesão benígna encontrada na ISIC 2018 com sua máscara de segmentação.

Fonte – Autor.

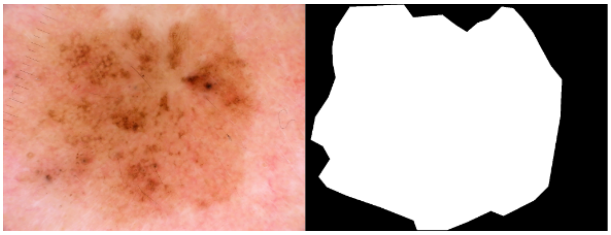


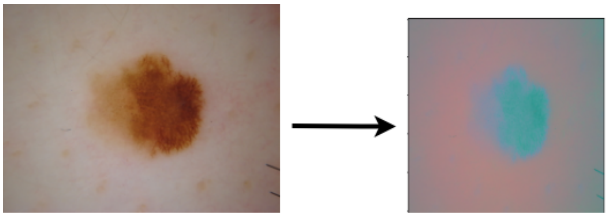
Figura 13 – Exemplo de lesão malígna encontrada na ISIC Autoral com sua máscara de segmentação.

Fonte – Autor.

4.1.1 Pré-processamento

As imagens RGB inicialmente com dimensões variadas que iam de 576 × 768 a 6748 × 4499 pixels foram redimensionadas para 256 x 256 pixels. Após um redimensionamento inicial de 128 x 128 pixels para diminuição do custo computacional, observou-se que se podia recortar até 256 x 256 pixels, este foi o tamanho máximo que contemplava as regiões para segmentar e que não sobrecarregasse a memória da máquina em que os testes foram realizados. 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 das bases. Cada canal da imagem foi então normalizado para o intervalo [0, 1] para otimização do treinamento das redes.

Figura 14 – Exemplo de imagem de entrada após o pré-processamento.



Fonte – Autor.

4.2 Estimação de backbones

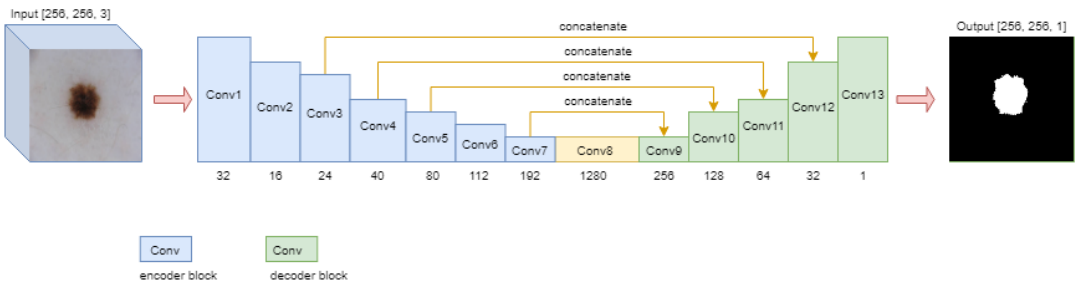
O aprendizado de transferência foi usado utilizando modelos de arquitetura pré-treinados no ImageNet (Deng et al., 2009). Aproveitando os dados do primeiro dataset para extrair informações que possam ser úteis para o aprendizado da CNN (BENGIO; GOODFELLOW; COURVILLE, 2017), o objetivo de usar a aprendizagem 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. Efficientnetb1 obteve os melhores resultados para a tarefa de segmentação, por isso foi escolhido como backbone das arquiteturas propostas.

4.3 Implementação de modelos de CNNs

Quatro diferente modelos de CNNs encoder-decoder foram implementados e avaliados: U-Net (RONNEBERGER; FISCHER; BROX, 2015), FPN (Lin et al., 2017), PSP-Net (ZHAO et al., 2017) e Linknet (CHAURASIA; CULURCIELLO, 2017). As próximas subseções explanam os como estas foram implementadas, utilizando Efficientnetb1 como backbone.

4.3.1 U-net

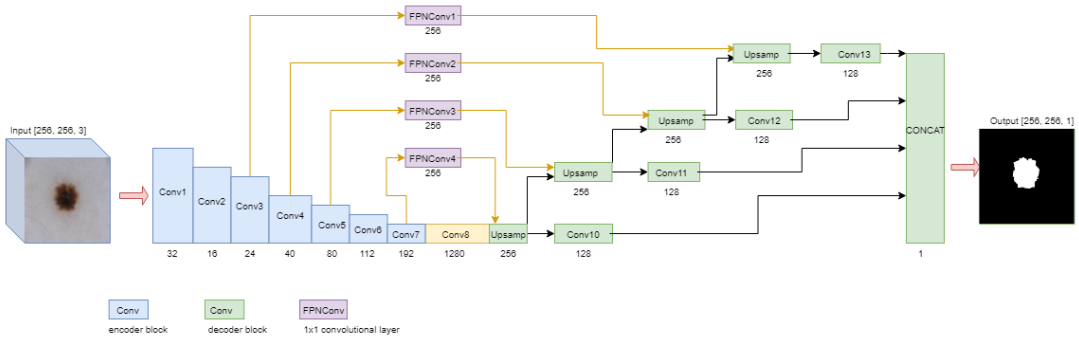
Figura 15 – Arquitetura da rede U-net avaliada.



Fonte – Autor.

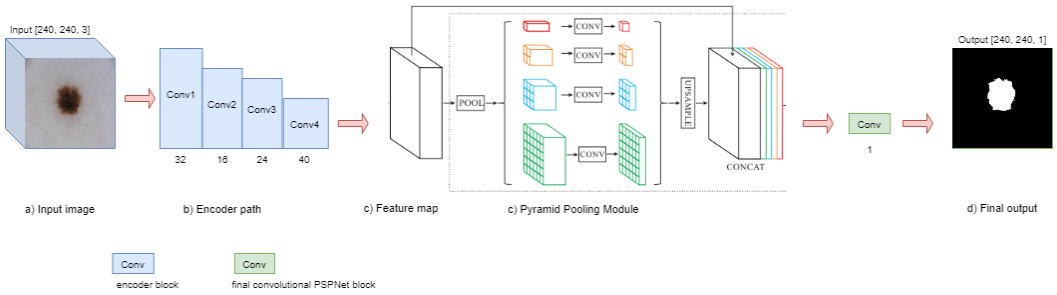
A Figura 15 mostra a arquitetura proposta do modelo U-Net. Contém a camada inicial de convolução totalmente com 32 filtros, seguida de blocos de convolução MBConv (SANDLER et al., 2018) inerente à Efficientnetb1. Cada bloco do decoder consiste em upsampling, concatenate, conv, BN, ReLU. Conv é a camada convolucional, BN é a camada de normalização de batch e ReLU a função de ativação. Na camada final, uma convolução 1x1 é aplicada para mapear cada mapa de recursos de 16 componentes para o número desejado de classes e ativação sigmoid. O bloco amarelo Conv8 na Figura 15 consiste em MBConv, Conv, BN, Relu6 para conectar os caminhos do encoder e decoder.

Figura 16 – Arquitetura da rede FPN avaliada. Cada mapa de característica no decoder é concatenado e, em seguida, aplicado ativação sigmoid para gerar a máscara de segmentação final.



Fonte – Autor.

Figura 17 – Arquitetura da rede PSP-Net avaliada.



Fonte – Autor.

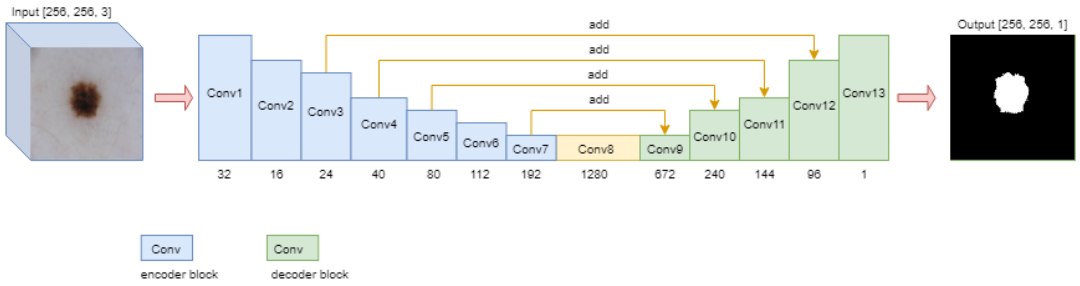
4.3.2 FPN

No modelo avaliado, o decoder da FPN faz o upsampling de mapas espacialmente mais grosseiros, mas semanticamente mais robustos e melhora-os com os mapas de recursos do encoder Efficientnetb1 através de conexões laterais. A Figura 16 apresenta as conexões laterais em amarelo, que aplicará camadas convolucionais 1x1 para os mapas de recursos a serem mesclados com o mapa de recurso do decoder correspondente depois do upsampling. Esse processo é iterado até que o mapa final de resolução seja gerado.

4.3.3 PSP-Net

No modelo PSP-Net avaliado, Efficientnetb1 extrai o mapa de recursos que é aplicado ao Módulo de Análise de Pirâmide (Pyramid Parsing Module) conforme exibido em (c) na Figura 18. Após uma camada convolucional final é aplicada consistindo em Conv, upsampling e ativação sigmoid para obter a máscara de saída final.

Figura 18 – Arquitetura da rede Linknet avaliada.



Fonte – Autor.

4.3.4 Linknet

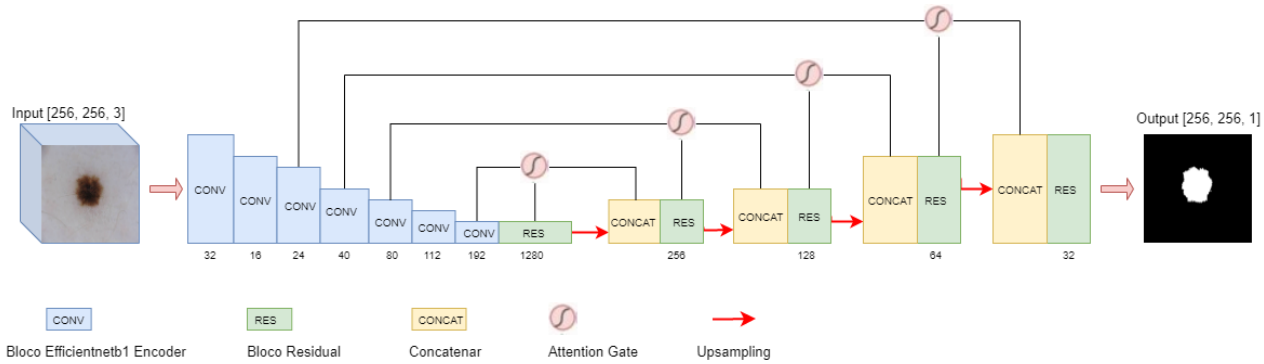
Cada bloco do decoder consiste em nas operações de Conv, BN, ReLU, upsampling e Add. A camada Add encaminha a saída do bloco de convolução do encoder correspondente para o mapa de recursos do decoder atual. Ela adiciona mapas de recursos de mesmo número de canais. Uma convolução final 1x1 é usada para mapear cada vetor de característica de 16 canais para o número desejado de classes junto com a ativação sigmoid.

4.3.5 AER-Net: Uma nova proposta de rede neural convolucional

Neste trabalho propõe-se um novo modelo de CNN baseado na U-Net e uso de aprendizado de transferência, chamado de Attention Efficient Residual U-Net (AER-Net). A Figura 19 apresenta a visão esquemática da arquitetura. É dividida em dois caminhos, o encoder e decoder. 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. 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. A representação visual do bloco residual decoder é apresentada na Figura 20.

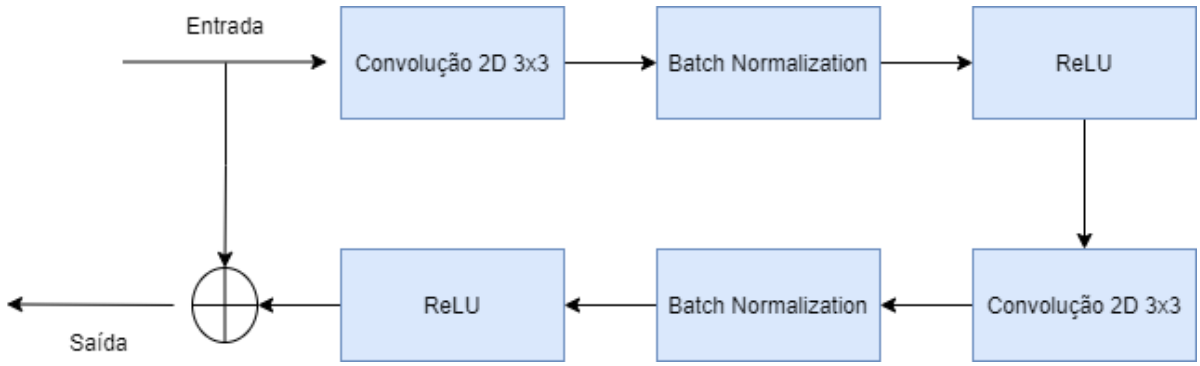
Assim como na U-Net, as conexões dos mapas de recursos do encoder diretamente com o decoder correspondente são feitas através de skip connections. Técnicas foram propostas na literatura para modificações das skip connections, as reprojetando para um treinamento mais eficiente e aumento da acurácia na 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,

Figura 19 – Arquitetura AER-Net.



Fonte – Autor

Figura 20 – Bloco residual AER-Net.



Fonte – Autor

enquanto destaca características salientes úteis para a tarefa de segmentação. Os AGs estão implementados antes das operações de concatenação para mesclarem apenas as informações mais relevantes da imagem específica. Na Figura 21 está apresentado com detalhes as operações dos AGs.

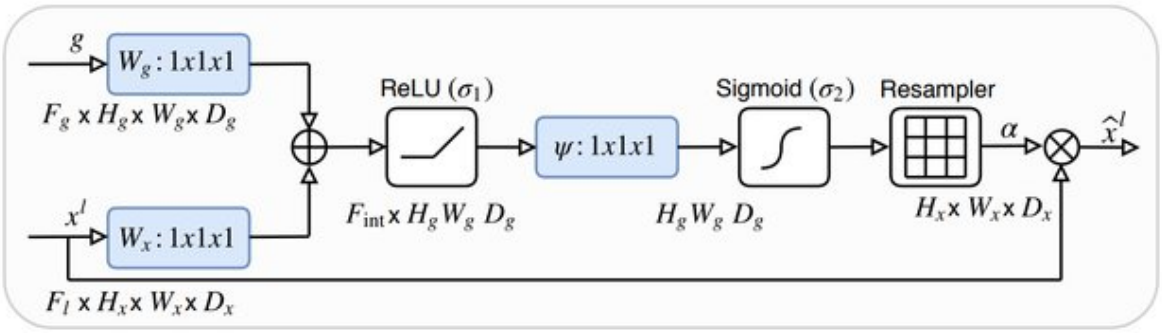
Por fim, é aplicado uma camada de convolução 1x1 e função de ativação sigmoid para gerar a máscara de segmentação final.

4.4 Métricas de Avaliação e Função Loss

Índice Jaccard médio (JSI) e Coeficiente Dice (DC) foram as métricas utilizadas para avaliar as máscaras de segmentação de saída de cada modelo (O) comparado às máscaras de segmentação ground-truth (G).

$$JSI = |G \cap O|/|G \cup O| \tag{4.1}$$

Figura 21 – Attention Gate.



Fonte – (OKTAY et al., 2018)

$$DC = 2|G \cap O|/|G| + |O| \tag{4.2}$$

Uma combinação de função Dice Loss (DCL) e Binary Focal Loss (BFL) foi usada para a função loss de treinamento (L) é definida como:

$$L = DCL + BFL \tag{4.3}$$

5 Resultados e Discussão

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.

Para a base ISIC 2018, o método de separação foi uma divisão aleatória na proporção de 80,7% para treinamento (2094 imagens) e 19,27% para o teste de validação (500 imagens). Para a base ISIC Autoral, foram separadas 50% das imagens para treinamento e os outros 50% restantes para teste. Para a realização dos testes, o método de otimização para treinamento da rede foi Adam (KINGMA; BA, 2017) com taxa de aprendizado em 0.001.

Para os experimentos no dataset ISIC 2018, a Tabela 2 apresenta os valores de Jaccard médio e Coeficiente de Dice obtidos pelas diferentes CNNs propostas. O modelo baseado na U-Net alcançou o valor de Jaccard médio mais alto entre os avaliados, enquanto o modelo baseado na arquitetura FPN alcançou o maior valor Dice. PSP-Net resultou na pior performance entre todos os métodos testados, e tal resultado pode ser explicado na falta de skip connections na arquitetura PSP-Net que iriam fazer a concatenação direta entre mapas de pixels do encoder para o decoder, com isso perdendo informações espaciais após sucessíveis camadas de convolução. O modelo proposto AER-Net conseguiu superar a performance dos modelos previamente avaliados tanto em Jaccard como em Dice.

Experimentos realizados na mais extensa base ISIC Autoral estão retratados na Tabela 3 revelando que U-Net, FPN e Linknet ainda mostraram superioridade em relação a PSP-Net na segmentação das imagens, validando a relevância das skip connections para esta tarefa. O método proposto AER-Net também provou-se superior nas métricas avaliadas, evidenciando a confiabilidade do modelo para tarefas de segmentação de lesões de pele.

Tabela 2 – Resultados de segmentação no dataset ISIC 2018

Método	Jaccard	Dice
Unet	76.08%	84.24%
Linknet	75.94%	84.25%
PSPNet	71.86%	80.98%
FPN	75.99%	84.46%
AER-Net	**77.29%**	**85.25%**

Tabela 3 – Resultados de segmentação na base ISIC Autoral

Método	Jaccard	Dice
Unet	78.48%	86.64%
Linknet	78.3%	86.2%
PSPNet	74.26%	82.92%
FPN	78.3%	86.43%
AER-Net	**79.5%**	**87.7%**

Tabela 4 – Comparação de diferentes métodos no dataset ISIC 2018

Autor	Método	Jaccard	Dice
Souza, Lelis e Silva (2020)	U-Net	68.0%	77.2%
Ji et al. (2018)	baseado em Resnet34	79.51%	-
Amin et al. (2020)	VGG-16	85.0%	82.0%
Oktay et al. (2018)	Attention U-Net	70%%	79.6%
Métodos Avaliados	PSPNet - Efficientnet	71.86%	80.98%
	Unet - Efficientnet	76.08%	84.24%
	Linknet - Efficientnet	75.94%	84.25%
	FPN - Efficientnet	75.99%	84.46%
Método Proposto	AER-Net	77.29%	85.25%

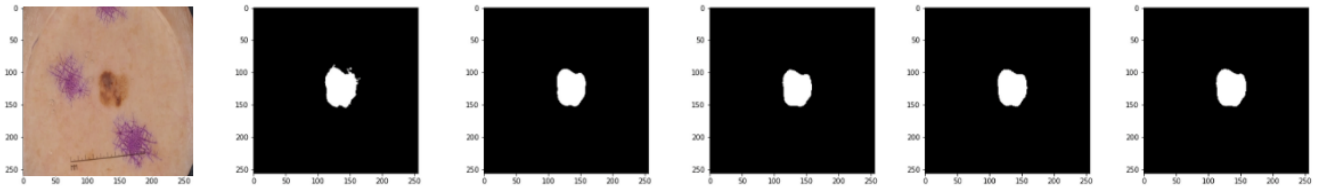
A Tabela 4 apresenta os métodos de segmentação analisados comparado com outros métodos na literatura. O modelo AER-Net apresentado mostrou resultados promissores em Coeficiente de Dice ultrapassando metodologias propostas como alguns dos métodos melhores colocados do desafio ISIC 2018 que não implantaram técnicas de ensemble, enquanto demonstrou equivalência em índice Jaccard médio, e superando o desempenho da Attention U-Net. A U-Net avaliada também mostrou resultados promissores via valores Jaccard comparado aos propostos. U-Net, LinkNet, and FPN também aumentaram a performance em Dice comparado a outras abordagens mostrando que essas arquiteturas também podem ser confiáveis na segmentação de lesões de pele médicas.

5.0.1 Estudos de Casos

As Figuras 22, 23 e 24 apresentam exemplos de saídas de máscaras de segmentação para cada modelo em comparação com a máscara ground-truth, exibindo os exemplos de saída de qualidade boa, média e ruim, respectivamente. A primeira coluna é a imagem RGB original. A segunda coluna mostra a máscara ground-truth. A terceira coluna contém a saída da rede FPN, a quarta coluna a saída da rede Linknet, a quinta coluna a saída da rede PSP-Net e a sexta coluna a saída da rede U-Net.

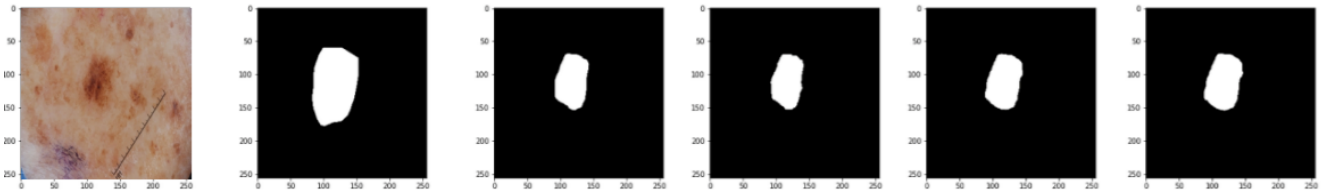
A Figura 22 mostra exemplos da execução da segmentação das quatro redes de uma lesão de fácil segmentação. Todas as redes testadas conseguiram contornar bem a área da lesão respeitando as suas bordas.

Figura 22 – Exemplo de saída de segmentação de boa qualidade.



Fonte – Elaborada pelo autor.

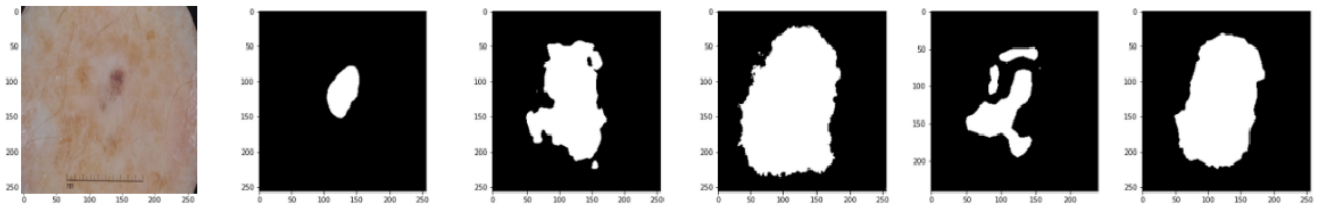
Figura 23 – Exemplo de saída de segmentação de média qualidade.



Fonte – Elaborada pelo autor.

A Figura 23 apresenta exemplos de saída em uma imagem dermatoscópica de complexidade média de segmentação. Nesta imagem, as redes conseguiram 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 das redes obtiveram muitos falsos negativos nesta região em volta a mancha mais escura.

Figura 24 – Exemplo de saída de segmentação de má qualidade.

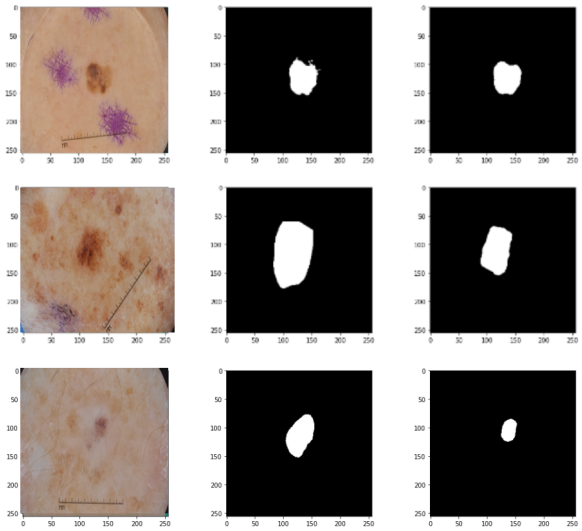


Fonte – Elaborada pelo autor.

Já a Figura 24 apresenta resultados não satisfatórios pelas metodologias aplicadas. Esta lesão de melanoma contém uma mancha de cor diferente, e com bordas com contrastes bastante suavizados entre a pele e lesão. Desta forma, todas as redes de CNN tiveram dificuldade em segmentar tal imagem. A FPN englobou parte da lesão e segmentou vários falsos positivos em volta gerando uma máscara de bordas irregulares e com falhas. A Linknet apesar de englobar toda a região da lesão, também errou em atribuir vários Falso positivos na região da pele que não a contém, tornando a área segmentada bastante desproporcional a segmentação pelo especialista. A PSP-Net resultou uma segmentação

com bordas irregulares que não englobavam a lesão e com falhas representado Falsos negativos e positivos na área em questão. Por fim, a U-Net apresentou condição análoga a Linknet gerando falsos positivos em volta da lesão não conseguindo fazer a distinção entre pele e lesão tão bem quanto o especialista.

Figura 25 – Exemplo de saída de segmentação na AER-Net.



Fonte – Elaborada pelo autor.

Para a AER-Net, a Figura 25 apresenta a saída de segmentação gerado pelo modelo das mesmas imagens anteriores, separados em cada linha. Para a imagem de boa segmentação da primeira linha e para a segmentação de resultado médio na segunda linha, o modelo apresentou resultados semelhantes aos métodos avaliados anteriormente, apresentando Dice de 96% para a primeira e 65% para a segunda. Já para o melanoma da terceira linha de difícil segmentação, o modelo apresentou um Dice de 47% o que não é satisfatório, porém nota-se que máscara de segmentação gerada respeitou o contorno da área da mancha mais escura e suas bordas, segmentando errado a região da lesão com tonalidade muito clara e sem contraste evidente entre pele e lesão, ao contrário dos métodos avaliados a priori que apresentaram uma máscara com contornos irregulares e buracos nas regiões de bordas.

6 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, primeiro foram avaliados quatro modelos de CNNs para a segmentação de lesões cutâneas em imagens dermatoscópicas: FPN, Linknet, PSP-Net e U-Net. U-Net, FPN e Linknet obtiveram os melhores resultados com base em Jaccard e Dice, apresentando tais métricas melhores que a PSPNet em todos os testes, demostrando que as arquiteturas com skip connections diretas entre o encoder e decoder apresentam melhores resultados na tarefa de segmentação automática de imagens dermatoscópicas do que aquelas que tem ausência de tais conexões.

Em cima disso, 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 estimação 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 estimação apontou que os backbones da família Efficientnet podem ser ótimos aliados na tarefa de segmentação automática de lesões cutâneas.

Para trabalhos futuros, propõe-se implementar e avaliar arquiteturas CNNs mais extensas e complexas. Além de continuar o 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, suavização, e etc.

Referências

AMIN, J.; SHARIF, A.; GUL, N.; ANJUM, M. A.; NISAR, M. W.; AZAM, F.; BUKHARI, S. A. C. Integrated design of deep features fusion for localization and classification of skin cancer. *Pattern Recognition Letters*, Elsevier, v. 131, p. 63–70, 2020. Citado 3 vezes nas páginas 14, 15 e 33.

ANIEMEKA, I. *A Friendly introduction to Convolutional Neural Networks.* 2017. Last accessed March 28, 2021. Disponível em: <https://hashrocket.com/blog/posts/afriendly-introduction-to-convolutional-neural-networks>. Citado 2 vezes nas páginas 19 e 20.

AVINERI, G.; TALMOR, O. *ISIC Archive Downloader*. 2016. Last accessed January 29, 2021. Disponível em: <https://github.com/GalAvineri/ISIC-Archive-Downloader>. Citado na página 25.

BADRINARAYANAN, V.; KENDALL, A.; CIPOLLA, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 39, n. 12, p. 2481–2495, 2017. Citado na página 12.

BENGIO, Y.; GOODFELLOW, I.; COURVILLE, A. *Deep learning*. [S.l.]: MIT press Massachusetts, USA:, 2017. v. 1. Citado na página 27.

CHAURASIA, A.; CULURCIELLO, E. Linknet: Exploiting encoder representations for efficient semantic segmentation. In: IEEE. *2017 IEEE Visual Communications and Image Processing (VCIP)*. [S.l.], 2017. p. 1–4. Citado 3 vezes nas páginas 21, 23 e 27.

CHEN, L.-C.; PAPANDREOU, G.; KOKKINOS, I.; MURPHY, K.; YUILLE, A. L. *Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs*. 2016. Citado na página 21.

CODELLA, N.; ROTEMBERG, V.; TSCHANDL, P.; CELEBI, M. E.; DUSZA, S.; GUTMAN, D.; HELBA, B.; KALLOO, A.; LIOPYRIS, K.; MARCHETTI, M. et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*, 2019. Citado na página 25.

COLLOBERT, R.; WESTON, J. A unified architecture for natural language processing: Deep neural networks with multitask learning. In: *Proceedings of the 25th International Conference on Machine Learning*. New York, NY, USA: Association for Computing Machinery, 2008. (ICML '08), p. 160–167. ISBN 9781605582054. Disponível em: <https://doi.org/10.1145/1390156.1390177>. Citado na página 17.

Deng, J.; Dong, W.; Socher, R.; Li, L.; Kai Li; Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2009. p. 248–255. Citado na página 27.

GE, Z.; DEMYANOV, S.; CHAKRAVORTY, R.; BOWLING, A.; GARNAVI, R. Skin disease recognition using deep saliency features and multimodal learning of dermoscopy

and clinical images. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. [S.l.], 2017. p. 250–258. Citado na página 12.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>. Citado 2 vezes nas páginas 17 e 19.

GOOGLE. *Google Colaboratory*. 2021. Acesso em April 19, 2021. Disponível em: <https://colab.research.google.com/>. Citado na página 32.

GU, J.; WANG, Z.; KUEN, J.; MA, L.; SHAHROUDY, A.; SHUAI, B.; LIU, T.; WANG, X.; WANG, L.; WANG, G.; CAI, J.; CHEN, T. *Recent Advances in Convolutional Neural Networks*. 2017. Citado na página 17.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778. Citado na página 14.

INCA. *INCA Homepage*. 2021. Last accessed January 17, 2021. Disponível em: <https://www.inca.gov.br>. Citado na página 12.

JI, Y.; LI, X.; ZHANG, G.; LIN, D.; CHEN, H. Automatic skin lesion segmentation by feature aggregation convolutional neural network. *Technical report*, 2018. Citado 3 vezes nas páginas 14, 15 e 33.

KARN, U. *An Intuitive Explanation of Convolutional Neural Networks*. 2016. Acesso em March 17, 2021. Disponível em: <https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnetstps://hashrocket.com/blog/posts/afriendly-introduction-to-convolutional-neural-networks>. Citado na página 20.

KINGMA, D. P.; BA, J. *Adam: A Method for Stochastic Optimization*. 2017. Citado na página 32.

Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998. Citado na página 17.

Lin, T.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. p. 936–944. Citado 3 vezes nas páginas 22, 23 e 27.

LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 3431–3440. Citado na página 12.

M, C.; G, Z.; P, B.; P, C.; S, B.; R., M. Dermatoscopy: usefulness in the differential diagnosis of cutaneous pigmentary lesions. 1994. Citado na página 16.

MOHAMED, I. S. *Detection and Tracking of Pallets using a Laser Rangefinder and Machine Learning Techniques*. Tese (Doutorado), 09 2017. Citado na página 18.

NAZI, Z. A.; TASNIM, A. A. Automatic skin lesion segmentation and melanoma detection: Transfer learning approach with u-net and dcnn-svm. In: SPRINGER. *Proceedings of International Joint Conference on Computational Intelligence*. [S.l.], 2020. p. 371–381. Citado 2 vezes nas páginas 14 e 15.

OKTAY, O.; SCHLEMPER, J.; FOLGOC, L. L.; LEE, M.; HEINRICH, M.; MISAWA, K.; MORI, K.; MCDONAGH, S.; HAMMERLA, N. Y.; KAINZ, B.; GLOCKER, B.; RUECKERT, D. *Attention U-Net: Learning Where to Look for the Pancreas*. 2018. Citado 4 vezes nas páginas 14, 29, 31 e 33.

RESEARCH, W. C. R. F. I. for C. *Diet, Nutrition, Physical Activity and Cancer: a Global Perspective*. 2018. Last accessed January 17, 2021. Disponível em: <dietandcancerreport.org>. Citado na página 12.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *International Conference on Medical image computing and computer-assisted intervention*. [S.l.], 2015. p. 234–241. Citado 4 vezes nas páginas 12, 21, 22 e 27.

RP, B.; JH, S.; LE, F. Dermoscopy of pigmented lesions: a valuable tool in the diagnosis of melanoma. *Swiss Medical Weekly*, v. 134, n. 7/8, p. 83–90, 2004. Citado na página 16.

SANDLER, M.; HOWARD, A.; ZHU, M.; ZHMOGINOV, A.; CHEN, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2018. p. 4510–4520. Citado na página 27.

SOARES, H. B. *Análise e classificação de imagens de lesões da pele por atributos de cor, forma e textura utilizando máquina de vetor de suporte*. Tese (Doutorado) — Universidade Federal do Rio Grande do Norte, 02 2008. Citado na página 17.

SOCIETY, A. C. *Cancer Facts and Figures 2021*. 2021. Last accessed January 17, 2021. Disponível em: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2021/cancer-facts-and-figures-2021.pdf>. Citado na página 12.

SOUZA, L.; LELIS, S.; SILVA, R. Segmentação de lesões de pele utilizando algoritmos de aprendizagem profunda. In: SBC. *Anais do XX Simpósio Brasileiro de Computação Aplicada à Saúde*. [S.l.], 2020. p. 344–355. Citado 3 vezes nas páginas 14, 15 e 33.

TANG, P.; LIANG, Q.; YAN, X.; XIANG, S.; SUN, W.; ZHANG, D.; COPPOLA, G. Efficient skin lesion segmentation using separable-unet with stochastic weight averaging. *Computer methods and programs in biomedicine*, Elsevier, v. 178, p. 289–301, 2019. Citado 2 vezes nas páginas 14 e 15.

TSCHANDL, P.; ROSENDAHL, C.; KITTLER, H. *The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data 5, 180161*. 2018. Citado na página 25.

VESTERGAARD, M.; MACASKILL, P.; HOLT, P.; MENZIES, S. Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. *British Journal of Dermatology*, v. 159, n. 3, p. 669–676, 2008. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2133.2008.08713.x>. Citado na página 16.

XIE, F.; YANG, J.; LIU, J.; JIANG, Z.; ZHENG, Y.; WANG, Y. Skin lesion segmentation using high-resolution convolutional neural network. *Computer methods and programs in biomedicine*, Elsevier, v. 186, p. 105241, 2020. Citado 2 vezes nas páginas 14 e 15.

ZHAO, H.; SHI, J.; QI, X.; WANG, X.; JIA, J. Pyramid scene parsing network. In: *CVPR*. [S.l.: s.n.], 2017. Citado 2 vezes nas páginas 24 e 27.

ZHOU, Z.; SIDDIQUEE, M. M. R.; TAJBAKHSH, N.; LIANG, J. *UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation*. 2020. Citado na página 29.