

HENRIQUE DUTRA ALVARES

Orientadora: Valéria de Carvalho Santos
Coorientadora: Roberta Barbosa Oliveira

**ANÁLISE DO IMPACTO DA APLICAÇÃO DE MÉTODOS
DE INTERPOLAÇÃO DE IMAGENS NA SEGMENTAÇÃO
DE LESÕES DE PELE USANDO REDE NEURAL
CONVOLUCIONAL SEGNET**

Ouro Preto
Dezembro de 2019

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**ANÁLISE DO IMPACTO DA APLICAÇÃO DE MÉTODOS
DE INTERPOLAÇÃO DE IMAGENS NA SEGMENTAÇÃO
DE LESÕES DE PELE USANDO REDE NEURAL
CONVOLUCIONAL SEGNET**

Monografia apresentada ao Curso de Bacharelado em Ciência da Computação da Universidade Federal de Ouro Preto como requisito parcial para a obtenção do grau de Bacharel em Ciência da Computação.

HENRIQUE DUTRA ALVARES

Ouro Preto
Dezembro de 2019

A473a

Alvares, Henrique Dutra.

Análise do impacto da aplicação de métodos de interpolação de imagens na segmentação de lesões de pele usando rede neural convolucional SegNet [manuscrito] / Henrique Dutra Alvares. - 2019.

31f.: il.: color; tabs.

Orientadora: Prof^a. Dr^a. Valéria de Carvalho Santos.
Coorientadora: Prof^a. Dr^a. Roberta Barbosa Oliveira.

Monografia (Graduação). Universidade Federal de Ouro Preto. Instituto de Ciências Exatas e Biológicas. Departamento de Computação.

1. Interpolação. 2. Dermoscopia. 3. Lesões. I. Santos, Valéria de Carvalho. II. Oliveira, Roberta Barbosa. III. Universidade Federal de Ouro Preto. IV. Título.

CDU: 004.032.26

Catálogo: ficha.sisbin@ufop.edu.br



UNIVERSIDADE FEDERAL DE OURO PRETO

FOLHA DE APROVAÇÃO

Análise do impacto da aplicação de métodos de interpolação de imagens na segmentação de lesões de pele usando rede neural convolucional SegNet

HENRIQUE DUTRA ALVARES

Monografia defendida e aprovada pela banca examinadora constituída por:

Valéria de Carvalho Santos

Dra. VALÉRIA DE CARVALHO SANTOS – Orientadora
Universidade Federal de Ouro Preto

Roberta Barbosa Oliveira

Dra. ROBERTA BARBOSA OLIVEIRA – Co-orientadora
Universidade de Brasília

Andrea G. Campos Bianchi

Dra. ANDREA GOMES CAMPOS BIANCHI
Universidade Federal de Ouro Preto

Guillermo Cámara Chávez

Dr. GUILLERMO CÁMARA CHÁVEZ
Universidade Federal de Ouro Preto

Ouro Preto, Dezembro de 2019

Resumo

Este trabalho tem como objetivo analisar o impacto de diversos métodos de interpolação de imagens nos resultados da segmentação de lesões de pele por meio de redes neurais convolucionais, e avaliar se existe diferença significativa entre os métodos. Para analisar o impacto dos métodos, foi utilizada a base de imagens do ISIC (*The International Skin Imaging Collaboration*) para treinar a rede neural convolucional SegNet com imagens redimensionadas por cinco diferentes métodos, sendo eles: Vizinho mais próximo, bilinear, bicúbico, Lanczos e *Area Sampling*. Foram realizados testes preliminares, que apresentaram resultados significativos, considerando a similaridade entre as imagens redimensionadas. Após o treinamento dos modelos da rede SegNet usando as bases de dados redimensionadas, foram encontrados resultados menos expressivos que os testes preliminares levaram a supor, porém congruentes em seu comportamento. Com os resultados obtidos pelos modelos e comparados entre si, foi concluído que existe uma pequena diferença a ser encontrada utilizando diferentes métodos de interpolação e que o método *Area Sampling* obteve os melhores resultados.

Palavras-chave: métodos de interpolação, redimensionamento de imagens, SegNet, segmentação de imagens, imagens dermatocópicas.

Abstract

This paper aims to analyze the impact of several image interpolation methods on the results of skin lesion segmentation through convolutional neural networks, and to evaluate if there is significant difference between the methods. To analyze the impact of the methods, the ISIC (The International Skin Imaging Collaboration) image base was used to train the SegNet convolutional neural network with images scaled by five different methods, namely: Nearest Neighbor, Bilinear, Bicubic, Lanczos and Area Sampling. Preliminary tests were performed, which presented significant results, considering the similarity between the resized images. After training the SegNet network models using the resized databases, results were less expressive than the preliminary tests led to suppose, but congruent in their behavior. With the results obtained by the models and compared with each other, it was concluded that there is a small difference to be found using different interpolation methods and that the Area Sampling method obtained the best results.

Keywords: interpolation methods, image resizing, SegNet, image segmentation, dermatocopic images.

Dedico esse trabalho a todos que partiram e que carrego a saudade no coração. Obrigado por acreditarem em mim, e me apoiarem nesta jornada.

Agradecimentos

Agradeço aos meus Pais Elizeu (*in memoriam*) e Miriam pelas oportunidades de crescer que me foram dadas, aos meus tios Eider e Maria Francisca (*in memoriam*) pelo investimento feito em minha educação, e à minha companheira Ellen, por estar ao meu lado nos momentos mais difíceis. Agradeço também aos professores Guillermo e Andrea, pelas contribuições dadas neste trabalho.

Sumário

1	Introdução	1
1.1	Justificativa	2
1.2	Objetivos geral e específicos	3
1.3	Método do trabalho	3
1.4	Organização do trabalho	3
2	Revisão de Literatura	5
3	Fundamentação Teórica	7
3.1	Métodos de interpolação	7
3.1.1	Vizinho mais próximo	7
3.1.2	Bilinear	9
3.1.3	Bicúbico	10
3.1.4	Lanczos	11
3.1.5	<i>Area sampling</i>	11
3.2	SegNet	12
3.2.1	Encoder	13
3.2.2	Decoder	14
3.3	Métodos de avaliação	14
3.3.1	Jaccard Index	14
3.3.2	SSIM	15
4	Metodologia	16
4.1	Redimensionamento de imagens	16
4.2	Segmentação de Imagens	17
4.3	Avalização dos resultados	18
5	Experimentos	19
5.1	Bases de imagens	19
5.2	Ambiente de desenvolvimento	19

5.3	Hardware	20
5.4	Estruturação dos arquivos	21
5.5	Testes Preliminares	21
5.5.1	Comparação com imagem original	21
5.5.2	Comparação com imagens redimensionadas	22
5.6	Treinamento e Teste	22
5.7	Resultados	24
6	Conclusão	26
6.1	Trabalhos futuros	26
	Referências Bibliográficas	28

Lista de Figuras

1.1	Exemplo de pares do ISIC: (a) uma lesão de pele e (b) sua respectiva máscara <i>Ground Truth</i>	2
3.1	Exemplo interpolação: (a) imagem original e (b) imagem redimensionada sem interpolação.	8
3.2	Exemplos de métodos de interpolação: (a) interpolação bicúbica e (b) interpolação vizinho mais próximo.	8
3.3	Exemplo da interpolação vizinho mais próximo: (a) imagem original e (b) imagem redimensionada.	9
3.4	Exemplo da interpolação bilinear: (a) imagem original e (b) imagem redimensionada.	10
3.5	Exemplo da interpolação bicúbico: (a) imagem original e (b) imagem redimensionada.	11
3.6	Exemplo da interpolação lanczos: (a) imagem original e (b) imagem redimensionada.	12
3.7	Exemplo da interpolação <i>area sampling</i> : (a) imagem original e (b) imagem redimensionada.	12
3.8	Estrutura da Segnet (Badrinarayanan et al., 2017).	13
4.1	Visão geral das etapas que serão realizada neste trabalho.	17
5.1	Exemplos de imagens dermatoscópicas (Codella et al. (2017); Tschandl et al. (2018a)).	20
5.2	Um exemplo da qualidade da segmentação: (a) máscara <i>Ground Truth</i> e (b) máscara segmentada pela SegNet treinada anteriormente	24

Lista de Tabelas

5.1	Comparação entre imagens redimensionadas e originais.	22
5.2	Média dos valores de SSIM entre métodos de interpolação.	23
5.3	Desvio padrão dos valores de SSIM entre métodos de interpolação.	23
5.4	Resultados Jaccard.	24
5.5	Resultados SSIM.	25

Lista Abreviaturas e Siglas

CPU *Central Processing Unit.* 20

GCP *Google Cloud Platform.* 19, 23

GPU *Graphics Processing Unit.* 20, 23

JPEG *Joint Photographic Experts Group.* 19

PNG *Portable Network Graphics.* 19

RGB *Red Green Blue.* 19

SSIM *Structural Similarity.* viii, 5, 6, 14, 15, 18, 21–24

Capítulo 1

Introdução

O câncer de pele é a forma mais comum de câncer no mundo, com uma estimativa de 165.580 casos de câncer de pele do tipo não melanoma e 6.260 casos do tipo melanoma (INCA, 2018). O melanoma surge nas células melanocíticas e possui alta possibilidade de se propagar para outras partes do corpo (metástase), e portanto, apresenta maior taxa de mortalidade. O não melanoma pode apresentar tumores de diversos tipos, como por exemplo os carcinomas basocelular e espinocelular, e possui menor taxa de mortalidade se comparado ao melanoma.

Uma das ferramentas utilizadas por dermatologistas para a detecção de melanomas em estágios iniciais é o mnemônico ferramental "ABCDE", onde cada letra representa uma característica que quando combinadas podem indicar se uma lesão de pele é um melanoma ou uma lesão benigna, sendo elas Assimetria, Borda, Cor, Diâmetro e Evolução. Estas características são avaliadas visualmente pelo dermatologista e a lesão é diagnosticada como suspeita de ser ou não um melanoma (Tsao et al., 2015).

O ISIC, do inglês *The International Skin Imaging Collaboration*¹, é um esforço internacional para melhorar o diagnóstico de lesões de pele por meio de um acervo de imagens dermatoscópicas, especializado em gerar imagens de lesões, catalogadas e categorizadas. O ISIC promove uma competição com o objetivo de usar tal acervo para produzir um método que possibilite o diagnóstico automático de melanomas. Uma das etapas da competição é a etapa de segmentação de lesões de pele em imagens.

Segmentação de imagens é uma tarefa responsável por detectar região de interesse e gerar uma máscara com a representação dos *pixels* desta região. Os métodos de segmentação de lesões de pele podem ser classificados em várias categorias, tais como: baseados em região, baseados em contornos ativos, baseados em bordas, baseados em limiarização e computação suave (classificação de *pixels* usando técnicas de inteligência artificial, incluindo redes neurais, lógica *fuzzy* e computação evolucionar)(Oliveira et al. (2016); Celebi et al. (2009)). A máscara gerada pela segmentação pode ser utilizada para avaliar o critério de borda do mnemônico

¹ <https://challenge2018.isic-archive.com/>. Acesso em: 30 de Novembro de 2019.

"ABCDE", pelo médico ou por meio de softwares, sendo assim, útil para o diagnóstico do câncer de pele tipo melanoma.

Algoritmos de aprendizagem profunda têm sido muito utilizados para segmentação de imagens recentemente (Mishra e Daescu (2017); Qian et al. (2018); Baghersalimi et al. (2019)). Estes algoritmos geram modelos que são treinados utilizando um conjunto de imagens de treinamento, que consistem de pares de imagens das lesões, como na Figura 1.1(a), e imagens monocromáticas correspondentes, que foram previamente demarcadas por um dermatologista. Nas imagens monocromáticas uma das cores corresponde a região da lesão (objeto de interesse), chamadas de *Ground Truth*, representada em branco na Figura 1.1(b). Além do conjunto de treinamento, deve-se selecionar também um conjunto de teste, que segue a mesma estrutura do conjunto de treinamento, porém com imagens distintas, que não foram usadas no processo de treinamento. Este conjunto é utilizado após o treinamento do modelo e serve para testar quão eficiente foi o treinamento.

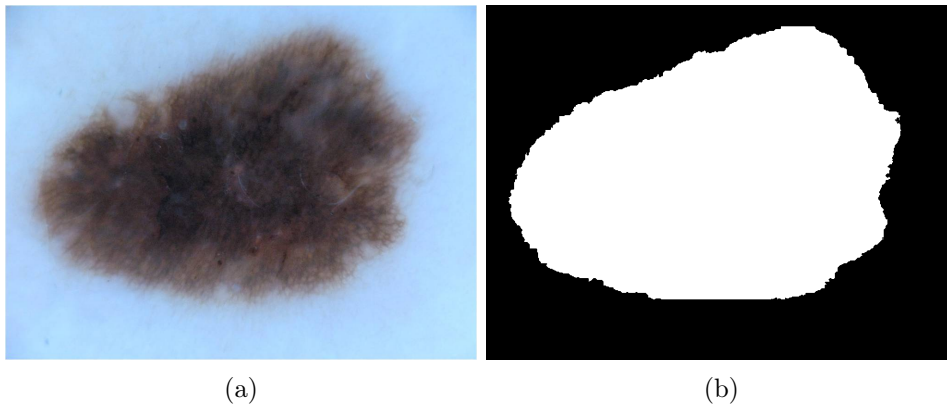


Figura 1.1: Exemplo de pares do ISIC: (a) uma lesão de pele e (b) sua respectiva máscara *Ground Truth*.

Redes neurais convolucionais, como a SegNet (Badrinarayanan et al., 2017), muitas vezes precisam de utilizar imagens padronizadas para realizar o treinamento, pois sua estrutura não foi desenvolvida para acomodar imagens de diferentes resoluções (Badrinarayanan et al., 2017). Portanto, as imagens do conjunto de treinamento e suas respectivas máscaras *Ground Truth*, devem ser redimensionadas para todas possuírem o mesmo tamanho. Para realizar tal redimensionamento, são utilizados métodos de interpolação (Getreuer, 2011), que aproximam valores para os novos *pixels* da imagem redimensionada, com base nos *pixels* da imagem original.

1.1 Justificativa

O uso de softwares baseados em modelos de aprendizagem profunda para a detecção automática de lesões é uma possível alternativa para auxiliar o dermatologista no diagnóstico

precoce (Esteva et al., 2017) e assim acelerar o processo de início de um tratamento, caso necessário. Para a criação de um modelo baseado em aprendizagem profunda, o conjunto de imagens de treinamento pode passar antes por um pré-processamento, onde várias transformações podem ser aplicadas às imagens com o objetivo de auxiliar no processo de treinamento.

Um tipo de pré-processamento que pode ser necessário, é o redimensionamento das imagens para uma resolução padrão, uma vez que algumas arquiteturas de redes convolucionais requerem a padronização das imagens de entrada. Esse redimensionamento pode ser realizado por diversas razões, tais como: conformar ao padrão de resolução determinado pelo modelo; adequar às limitações de hardware, como limite de memória volátil do sistema; e alcançar um menor tempo de processamento. Porém, a redução ou ampliação de imagens é um processo que introduz a possibilidade de descaracterização da imagem, como a criação ou destruição de padrões, remoção de detalhes finos e interferência entre cores próximas, que pode reduzir a precisão da segmentação da lesão de pele. Sendo assim, é desejável que o mínimo de informações sejam perdidas durante o processo de redimensionamento, conseqüentemente a escolha de um método de interpolação adequado é uma decisão importante para este fim. Este trabalho busca comparar métodos de interpolação por meio de experimentos computacionais.

1.2 Objetivos geral e específicos

Neste trabalho, temos como objetivo geral avaliar os impactos dos métodos de interpolação de imagens na segmentação de lesões de pele utilizando a rede neural convolucional SegNet, e como objetivos específicos:

- Estudar diferentes métodos de interpolação;
- Comparar as características das imagens geradas pelos métodos de interpolação.

1.3 Método do trabalho

Para analisar os impactos dos métodos de interpolação na segmentação de lesões de pele, cada método de interpolação selecionado será utilizado para o redimensionamento das imagens do acervo ISIC. As imagens resultantes do processo de redimensionamento serão utilizadas para o treinamento e construção do seu respectivo modelo que serão considerados para os testes. Os resultados das avaliações de cada modelo serão comparados, e então será analisado se existe uma diferença significativa entre os resultados obtidos pelos modelos.

1.4 Organização do trabalho

Este trabalho está organizado da seguinte forma: O Capítulo 2 apresenta uma breve revisão de literatura, introduzindo trabalhos relacionados que possuem relevância para o presente

trabalho. O Capítulo 3 apresenta a fundamentação teórica do trabalho e o funcionamento dos métodos que serão utilizados no mesmo. O Capítulo 4 explica a metodologia utilizada na execução dos experimentos. O Capítulo 5 apresenta os resultados dos testes preliminares, bem como os resultados das redes treinadas e a comparação de seus resultados. E por fim, o Capítulo 6 apresenta uma conclusão e ideias para possíveis trabalhos futuros.

Capítulo 2

Revisão de Literatura

Em um estudo sobre o impacto de diversos ruídos em uma imagem, tais como desfoque, ruído, contraste e compressão, Dodge e Karam (2016) obtiveram resultados comparativos que apontam que imagens com desfoque ou borradas tendem a interferir na qualidade dos resultados de classificação, podendo perder até 20% de acurácia com um kernel gaussiano de tamanho 2. Tais resultados foram obtidos por meio das redes neurais convolucionais treinadas usando o *dataset* ImageNet, sendo elas: VGG-CNN-S, GoogleNet e VGG16 (Dodge e Karam, 2016). Foi apurado que as redes testadas são extremamente sensíveis a efeitos de desfoque, mesmo que moderados. Os autores especulam que tal sensibilidade pode ser explicada pela redução de texturas que a rede utiliza para identificar objetos, levando a uma queda significativa no número de padrões reconhecíveis pela rede.

Koziarski e Cyganek (2018) abordaram os efeitos de baixa resolução em reconhecimento de padrões utilizando aprendizagem profunda. Os autores consideraram imagens com variados níveis de resolução da original até um oitavo da mesma. Para cada nível de resolução foram treinadas utilizando o *dataset* ImageNet três diferentes redes neurais convolucionais, sendo elas: AlexNet, VGGNet e ResNet. Os resultados obtidos apontaram que a diminuição da resolução pode afetar fortemente o reconhecimento de padrões, podendo apresentar mais de 30% de perda de acurácia nos resultados, como foi observado na VGGNet, que teve uma perda de aproximadamente 30% de acurácia nos resultados, saindo de 90% com a imagem original, para aproximadamente 60% de acurácia com um oitavo da resolução original.

Zhou Wang et al. (2004) desenvolveu uma métrica de semelhança entre imagens chamada Índice de similaridade Estrutural (*Structural Similarity* (SSIM)), onde imagens são comparadas utilizando as suas informações estruturais, de forma a reduzir o impacto de ruídos como sal e pimenta e artefatos de compressão nos valores de similaridade. Dessa forma é possível obter métricas referentes à qualidade de uma imagem de forma mais próxima à que o ser humano a faria, sendo assim mais relevante para imagens de origem natural.

Uma camada para redes neurais, chamada SSIMLayer, que leva em conta o SSIM, foi proposta por Abobakr et al. (2018). Os autores apontaram experimentos que indicam como

redes neurais podem ser facilmente afetadas por ruídos imperceptíveis para humanos, gerando resultados errôneos com alto grau de confiabilidade. Uma rede neural convolucional utilizando a camada SSIMLayer e outra rede utilizando a mesma arquitetura, porém sem a camada SSIMLayer adicionada, foram treinadas utilizando o *dataset* CIFAR (Krizhevsky, 2009) e depois os resultados dos testes comparados. Ao final dos experimentos Abobakr et al. (2018) concluíram que o uso de uma camada de análises de SSIM pode aumentar a precisão das redes, reduzindo o efeito de ruídos nas imagens. Nos testes foi obtido até 40% mais de acurácia nas redes treinadas, dependendo do nível de ruído da imagem a ser avaliada.

Öztireli e Gross (2015) propuseram uma abordagem utilizando técnicas de otimização para manter o SSIM ao reduzir as dimensões das imagens. A abordagem permitiu manter detalhes estruturais e finos, evitando introdução de ruídos como padrão moiré (um tipo padrão de interferência que surge em estruturas repetitivas) e similares. Utilizando testes com usuários, esta nova abordagem alcançou um melhor desempenho ao ser comparado com outros métodos, sendo escolhido em mais que 50% das vezes por participantes da pesquisa. Além disso, a abordagem apresentou um custo computacional muito baixo, sendo mais rápido que o método bicúbico, ao ser aplicado no *MSRA Salient Object Database* (Wang et al., 2017).

Badrinarayanan et al. (2017) publicaram um artigo em que propuseram um modelo de segmentação utilizando aprendizagem profunda, a SegNet, uma rede convolucional profunda para segmentação de imagens. A SegNet é uma versão modificada da rede VGG16, em que descarta as camadas totalmente conectadas para reduzir a quantidade de parâmetros e o tempo de treino. A rede foi testada no *dataset CamVid* de imagens de ruas, com o intuito de segmentar e identificar carros, pedestres e postes. O *dataset* CamVid possui imagens de cenários de trânsito, consistindo de 367 imagens de treino e 233 imagens de teste e seus respectivos *ground truths*. Todas as imagens possuem resolução de 360×480 e foram utilizadas em sua resolução original. O resultado dos testes comparativos mostra a SegNet com melhor acurácia que as redes FCN, Deeplab-LargeFOV e DeconvNet, obtendo até 90% de acurácia global, e com menor tamanho de rede e menor tempo de treino e inferência.

Para estudar o uso de redes neurais profundas para a classificação de lesões de pele, Esteva et al. (2017) utilizaram a rede GoogleNet Inception (Szegedy et al., 2015) treinada em um *dataset* de imagens clínicas constituído de 127.463 imagens de treinamento e validação, e 1.942 imagens de teste, coletadas de um conjunto de 18 repositórios de livre acesso. Dessas imagens, 3.374 são dermatoscópicas. As imagens utilizadas foram redimensionadas para uma resolução de 299×299 , sem menção do método de interpolação utilizado. Para avaliar a qualidade da rede, foram comparados os resultados da rede e as respostas de 21 dermatologistas com o *ground truth* das lesões. Os resultados demonstram a rede obtendo melhor acurácia de classificação que a obtida pelos dermatologistas envolvidos nas comparações.

Capítulo 3

Fundamentação Teórica

Neste capítulo são introduzidos o funcionamento de alguns métodos de interpolação de imagens, bem como o funcionamento da rede SegNet para a segmentação de imagens. Métodos de avaliação de interpolação e segmentação de imagens também são abordados neste capítulo.

3.1 Métodos de interpolação

Ao se realizar uma transformação geométrica em uma imagem, como por exemplo redimensionamento, muitas vezes *pixels* na imagem resultante, não terão um correspondente direto na imagem original. Este *pixel* sem correspondente direto na imagem original não irá conter dados sobre sua cor, e portanto não poderá ser representado na imagem como visto na Figura 3.1. Para obter-se uma imagem completa após uma transformação geométrica, é utilizado um método de interpolação, que irá realizar o preenchimento dos *pixels* com valores aproximados, usando como base os valores dos *pixels* da imagem original, como visto na Figura 3.2.

Existe uma vasta gama de métodos de interpolação de imagens, com cada um tendo seus casos de uso e apresentando resultados com características distintas. Nesta seção, os métodos de interpolação a serem utilizados neste trabalho serão introduzidos com uma breve descrição de seu funcionamento e casos de uso. Para demonstração da aplicação dos métodos, foi utilizado um recorte da imagem Lena¹, e um recorte da mesma imagem, redimensionada utilizando os métodos de interpolação. A partir da imagem da Lena foi possível obter uma visualização melhor dos efeitos dos métodos.

3.1.1 Vizinho mais próximo

O método mais ingênuo e simples de interpolação é o vizinho mais próximo (*Nearest-Neighbor*), que tal qual o nome sugere, busca na imagem original o vizinho mais próximo ao *pixel* de destino na imagem redimensionada. Este método não é recomendado para operações

¹<https://www.ece.rice.edu/~wakin/images/>. Acesso em: 31 de novembro de 2019.

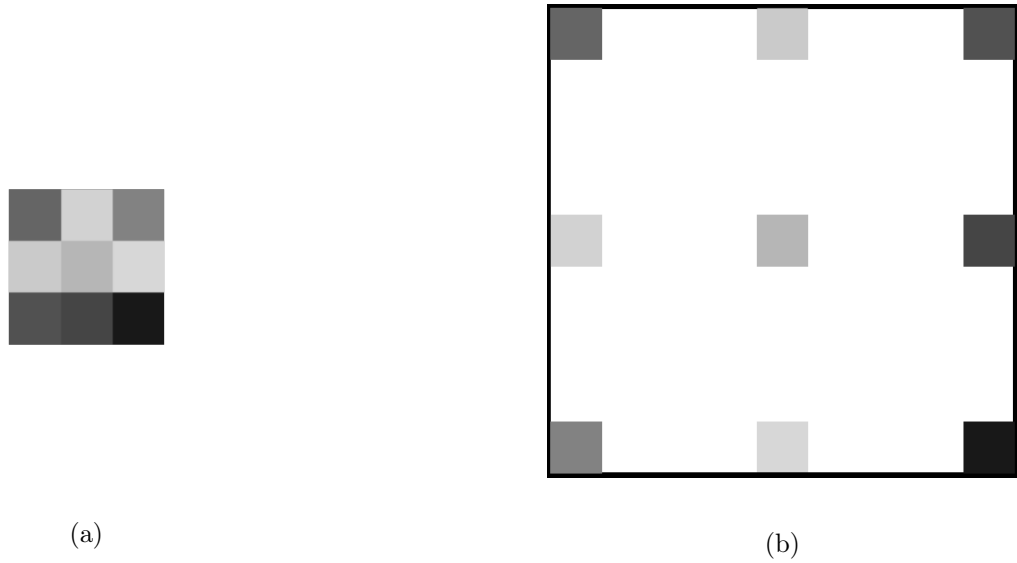


Figura 3.1: Exemplo interpolação: (a) imagem original e (b) imagem redimensionada sem interpolação.

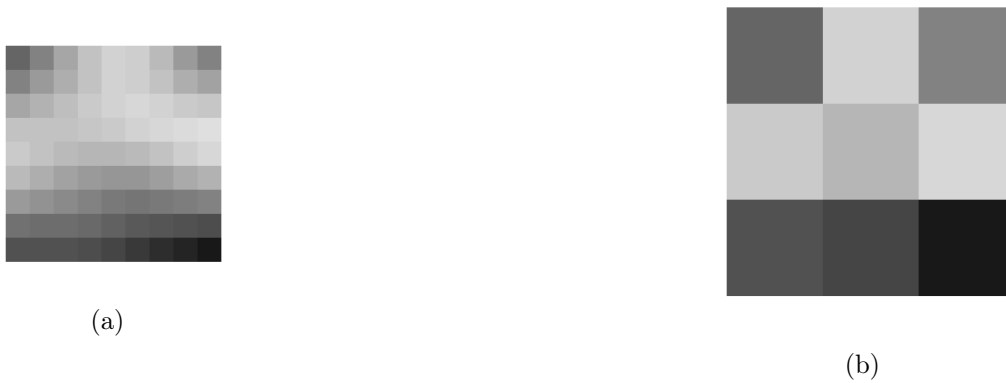


Figura 3.2: Exemplos de métodos de interpolação: (a) interpolação bicúbica e (b) interpolação vizinho mais próximo.

de interpolação em que se deseja reduzir a resolução da imagem, uma vez que pode gerar perda de detalhes muito facilmente, produzindo um efeito similar a de um filtro de passa-baixa (Getreuer, 2011). O vizinho mais próximo pode ser calculado pelo *kernel* $K(x, y)$ e pela equação de interpolação $P(t)$ de acordo com as Equações 3.1 e 3.2, respectivamente:

$$K(x, y) = P(x)P(y), \quad (3.1)$$

$$P(t) = \begin{cases} 1 & -1/2 \leq t \leq 1/2 \\ 0 & \text{caso contrário} \end{cases}, \quad (3.2)$$

onde x e y são as posições dos *pixels* a serem interpolados, um exemplo do resultado da aplicação deste método pode ser visto na Figura 3.3.

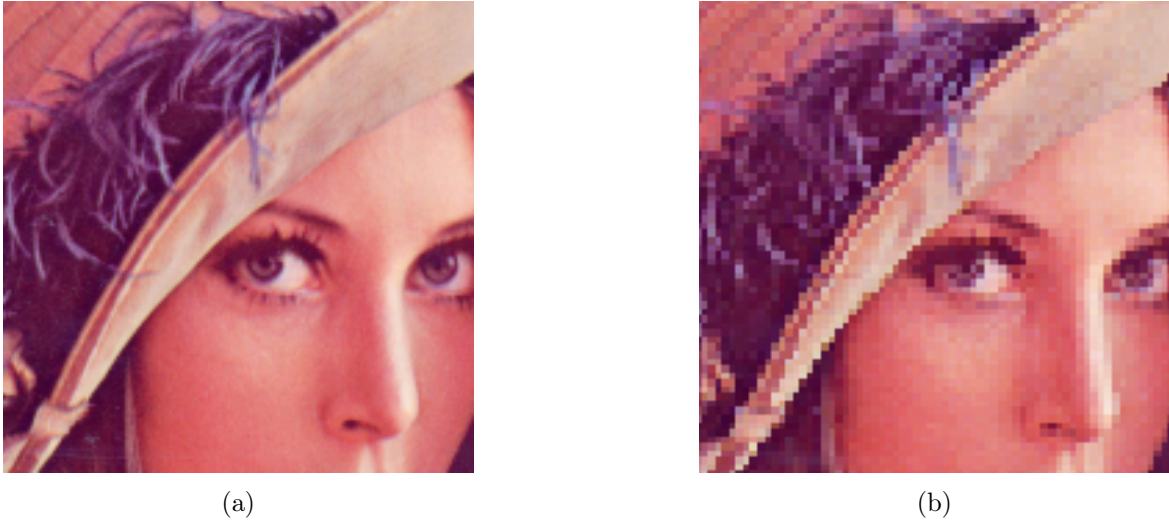


Figura 3.3: Exemplo da interpolação vizinho mais próximo: (a) imagem original e (b) imagem redimensionada.

3.1.2 Bilinear

O método de interpolação bilinear é simples e rápido, utilizado com muita frequência em operações de redimensionamento de imagens em várias aplicações (Getreuer, 2011). O funcionamento da interpolação bilinear consiste em interpolar linearmente os *pixel* de ambos os eixos na imagem original para encontrar o valor do *pixel* na imagem redimensionada. Desta forma, a imagem resultante do redimensionamento retém mais informações da imagem original. No entanto, as bordas são suavizadas, pois os *pixels* vizinhos irão afetar o valor dos *pixels* nas bordas. Isto resulta em uma imagem borrada e com bordas menos nítidas. O método bilinear pode ser calculado pela equação de interpolação $P(x, y)$ de acordo com a Equação 3.3, respectivamente:

$$\begin{aligned}
 P(x, y) \approx & \frac{(x_2 - x)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} Q_{11} + \frac{(x - x_1)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} Q_{21} + \\
 & + \frac{(x_2 - x)(y - y_1)}{(x_2 - x_1)(y_2 - y_1)} Q_{12} + \frac{(x - x_1)(y - y_1)}{(x_2 - x_1)(y_2 - y_1)} Q_{22},
 \end{aligned}
 \tag{3.3}$$

onde x e y são as posições do *pixel* a ser interpolado, x_i e y_j são as posições dos *pixels* conhecidos, e Q_{ij} são os valores dos *pixels* conhecidos. Um exemplo do resultado da aplicação deste método pode ser visto na Figura 3.4.

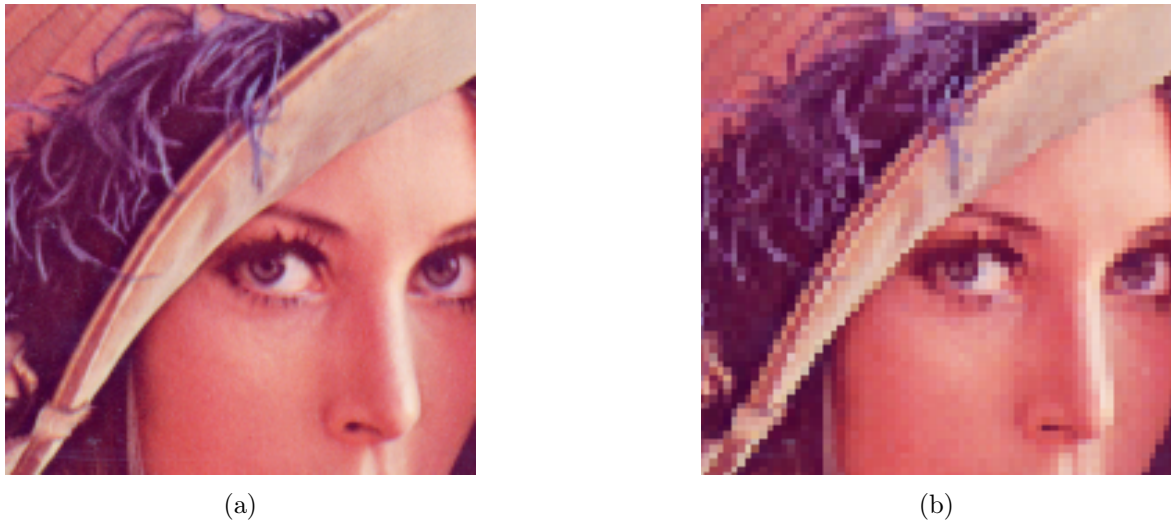


Figura 3.4: Exemplo da interpolação bilinear: (a) imagem original e (b) imagem redimensionada.

3.1.3 Bicúbico

De forma similar ao método de interpolação bilinear, o método bicúbico utiliza de interpolação entre *pixels* vizinhos (Getreuer, 2011). No entanto, com o intuito de tentar prever melhor as cores dos *pixels* resultantes da interpolação, é gerado um polinômio cúbico entre os pontos nos dois eixos da imagem original. O *pixel* resultante da imagem redimensionada é calculado com os dois polinômios e a posição do *pixel* resultante. Devido à natureza dos polinômios cúbicos este método preserva detalhes de maneira melhor que o método bilinear, porém tende a gerar mais artefatos de sobre-estimação ou sub-estimação, podendo causar bordas mais expressivas onde deveriam ser suaves, ou apresentar artefatos como por exemplo o padrão *móire* (Rajarampollu e Mankar, 2017). O método bicúbico pode ser calculado pelo *kernel* $K(x, y)$ e pela equação de interpolação $P(x)$ de acordo com as Equações 3.4 e 3.5, respectivamente:

$$K(x, y) = P(x)P(y), \quad (3.4)$$

$$P(x) = \begin{cases} (a + 2) * |x|^3 - (a + 3)|x|^2 + 1 & 0 \leq |x| < 1 \\ a * |x|^3 - 5a|x|^2 + 8a|x| - 4a & 1 \leq |x| < 2, \\ 0 & 2 \leq |x| \end{cases}, \quad (3.5)$$

onde x e y são as posições dos *pixels* a serem interpolados e a um valor entre -0.5 e -0.75 . Um exemplo do resultado da aplicação do método bicúbico pode ser visto na Figura 3.5.

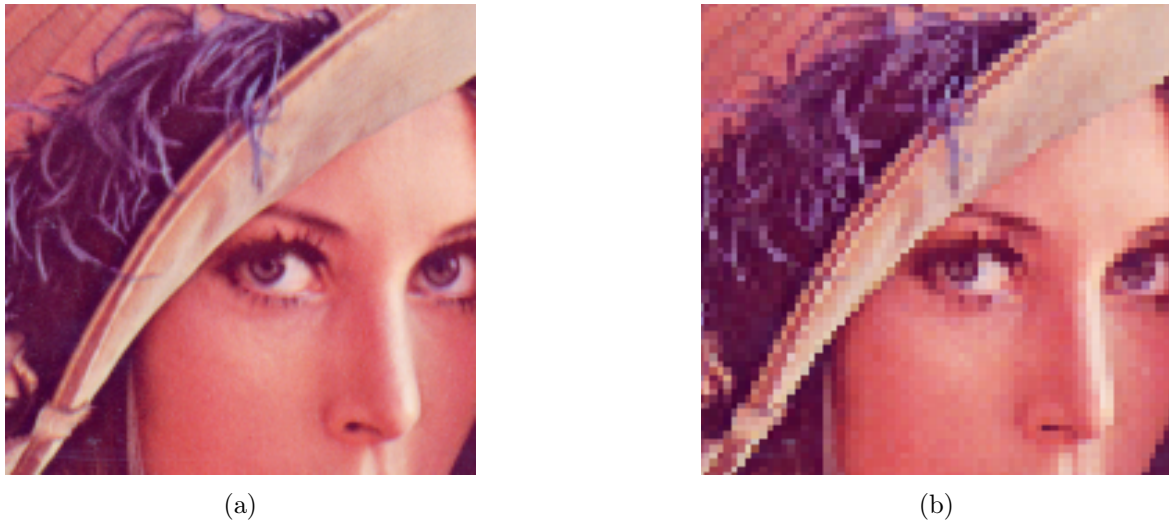


Figura 3.5: Exemplo da interpolação bicúbico: (a) imagem original e (b) imagem redimensionada.

3.1.4 Lanczos

O filtro lanczos utiliza uma Equação matemática que consiste do produto da função sinc (seno cardinal) normalizada e uma constante a , chamada de janela de lanczos (Parsania e Virparia, 2016). Esta função é aplicada em ambos os eixos da imagem original para obter o valor do *pixel* na imagem redimensionada. Este método é o mais complexo dentre os abordados, sendo necessárias múltiplas operações trigonométricas para se alcançar o resultado, tornando assim também o método mais lento (Parsania e Virparia, 2016). Em certos casos pode gerar artefatos circulares ao redor de bordas muito definidas. O método Lanczos pode ser calculado pelo *kernel* $K(x, y)$ e pela equação de interpolação $P(x)$ de acordo com as Equações 3.6 e 3.7, respectivamente:

$$K(x, y) = P(x)P(y), \quad (3.6)$$

$$P(x) = \begin{cases} \text{sinc}(x) * \text{sinc}(x/a) & -a < x < a \\ 0 & \text{caso contrário} \end{cases}, \quad (3.7)$$

onde x e y são as posições dos *pixels* a serem interpolados e a é a constante da janela de Lanczos. Um exemplo da aplicação do método pode ser visualizado na Figura 3.6.

3.1.5 Area sampling

O *area sampling* é um método de interpolação para redução de imagens, para a utilização do método é necessário a utilização de uma janela, cujo tamanho será definido pelo método de acordo com a proporção pela qual deseja-se redimensionar a imagem. Cada *pixel* da imagem

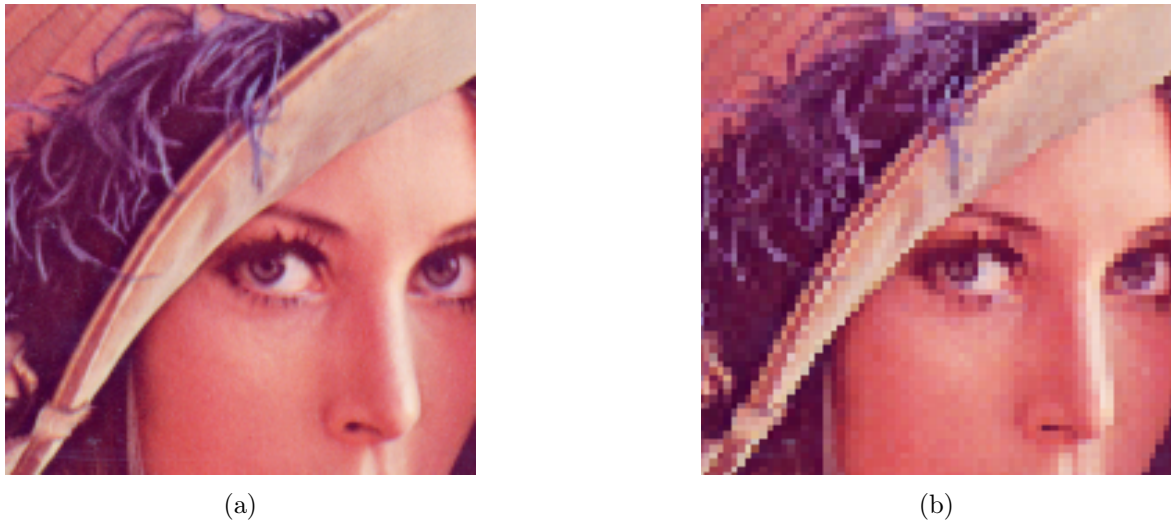


Figura 3.6: Exemplo da interpolação lanczos: (a) imagem original e (b) imagem redimensionada.

resultante é calculado por meio de uma média ponderada entre os *pixels* da imagem original abrangidos pela respectiva janela. Um exemplo da aplicação do método pode ser visualizado na Figura 3.7.

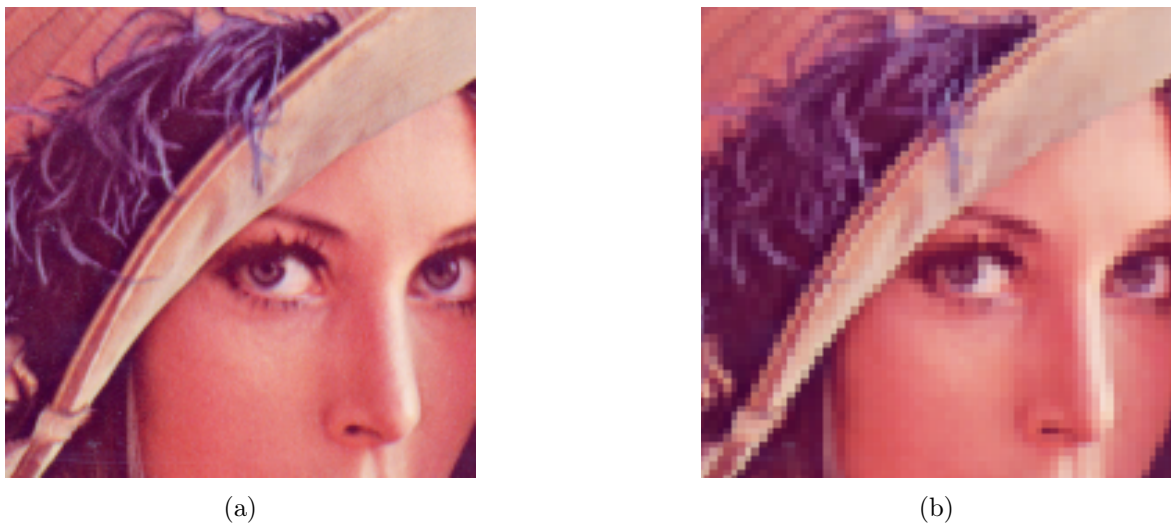


Figura 3.7: Exemplo da interpolação *area sampling*: (a) imagem original e (b) imagem redimensionada.

3.2 SegNet

Redes neurais convolucionais têm, nos últimos anos, sido aplicadas em problemas de segmentação de imagens, sendo uma técnica poderosa, capaz de extrair regiões de uma imagem

de forma automática, como por exemplo a rede SegNet (Badrinarayanan et al., 2017). Para tal, uma rede neural convolucional necessita de um banco de dados com diversas imagens para ser treinada e assim atingir um nível de confiabilidade alto em seus resultados (Badrinarayanan et al., 2017). Outros exemplos de redes convolucionais de segmentação são a *Full resolution Convolutional network* (FrCN) (Al-masni et al., 2018), AlexNet, GoogLeNet e ResNet (Garcia-Garcia et al., 2017).

SegNet é uma rede convolucional profunda *encoder-decoder*, proposta por Badrinarayanan et al. (2017), que pode ser aplicada à problemas de segmentação. Esta rede consegue alcançar melhores taxas de precisão global em comparação para problemas de segmentação com outras, como *Fully Convolutional Network* (FCN) e DeconvNet (Badrinarayanan et al., 2017), e ao mesmo tempo utilizando menos memória do sistema para executar.

A SegNet é estruturada utilizando uma rede *encoder* com 13 camadas seguida de uma rede *decoder* com 13 camadas e no final 1 (uma) camada de classificação *pixel a pixel*. Sua estrutura pode ser vista na Figura 3.8.

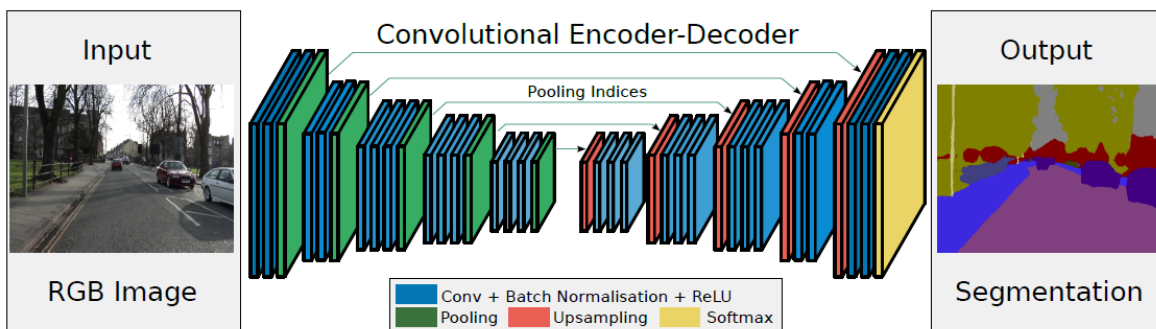


Figura 3.8: Estrutura da Segnet (Badrinarayanan et al., 2017).

3.2.1 Encoder

A SegNet foi desenvolvida de modo a reduzir o custo computacional de sua execução, descartando as camadas totalmente conectadas, que são camadas onde todos nós estão conectados com todos nós da camada seguinte. Tais camadas são utilizadas em redes como a VGG16 (Badrinarayanan et al., 2017), que possui 13 camadas de convolução, que são camadas que aprendem características visuais, e 3 camadas totalmente conectadas. Ao descartar as camadas totalmente conectadas, a SegNet reduz o número de parâmetros na sua rede *encoder*, e conseqüentemente diminui parte do custo computacional.

Após a execução das convoluções, operações realizadas entre um *pixel* e seus vizinhos, os mapas de características gerados são normalizados (valores serão ajustados entre um intervalo de 0 a 1, utilizando um operador ReLU). No mapa resultante é executado um *max pooling*, operação onde os maiores valores de ativação são escolhidos e os outros descartados, com uma

janela 2×2 , sem sobreposição. O resultado deste *max pooling* tem suas dimensões reduzidas por um fator de 2.

Apesar das sequências de *max pooling* e de redução de dimensão permitirem a rede se tornar mais invariante por translação, o processo também reduz a resolução espacial, o que leva a uma redução na definição das fronteiras entre objetos no mapa de características. Por esta razão é necessário adquirir e armazenar as informações das fronteiras, que é feito antes da redução da dimensão do mapa de características e armazenado para ser utilizado na rede *decoder*. Esta estrutura permite uma redução no uso de memória ao custo de uma pequena perda de precisão em comparação com redes que utilizam camadas totalmente conectadas.

3.2.2 Decoder

Cada camada de convolução da rede *encoder* possui uma camada equivalente na rede *decoder*. Nesta, os mapas de características da camada têm suas dimensões ampliadas utilizando as informações das fronteiras do *encoder* correspondente, gerando então um mapa de características esparsos, do qual um mapa de características denso é gerado por um *decoder filter bank*. A última camada *decoder* da rede produz um mapa de características com K canais (ou camadas), onde K é o número de objetos nas imagens a serem segmentadas. Este mapa é então processado por um classificador *softmax*, que classifica cada *pixels* individualmente, e retorna um mapa probabilístico de K canais, onde a predição de qual objeto aquele *pixel* pertence é o objeto com maior probabilidade para o *pixel* em questão.

3.3 Métodos de avaliação

Nesta seção serão apresentados os métodos utilizados para avaliar a qualidade da segmentação realizada pela SegNet nas imagens da etapa de teste em comparação com as máscaras *ground truth* da mesma. O *Jaccard Index* é a métrica utilizada para definir a colocação dos participantes do desafio ISIC. O SSIM foi escolhido por ter sido utilizado durante os testes preliminares para comparar as imagens resultantes da etapa de redimensionamento. Outra métrica considerada foi a *Sorensen Dice*, também recomendada pelo ISIC, porém foi descartada por ser equivalente ao *Jaccard Index*, sendo possível derivar diretamente uma a partir da outra.

3.3.1 Jaccard Index

Jaccard Index, ou em português índice Jaccard, é um método para encontrar a similaridade entre dois conjuntos, sendo computado pela divisão da cardinalidade da intersecção dos dois conjuntos, dividido pela cardinalidade de sua união. Esta operação produz um valor entre zero e um, correspondente à similaridade dos dois conjuntos, com 1 sendo totalmente iguais, e 0 sendo totalmente dissimilares, (Ni wattanakul et al., 2013) como visto na Equação 3.8,

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}, \quad (3.8)$$

onde A é a imagem segmentada e B a imagem original, o valor de $J(A,B)$ é o índice Jaccard.

3.3.2 SSIM

Proposta por Zhou Wang et al. (2004), o SSIM é uma métrica que leva em consideração a estrutura de uma imagem, e não apenas a quantidade de *pixels* iguais entre elas. Desta forma é possível se obter uma avaliação de proximidade entre imagens com mais relevância ao formato dos objetos, ou no formato da máscara de segmentação, como é o caso deste estudo. Este método, assim como o anterior, produz um valor entre zero e um, sendo 1 duas imagens totalmente iguais, e 0 sendo totalmente dissimilares. A métrica o SSIM pode ser calculado da seguinte maneira:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (3.9)$$

$$C_1 = (L * 0.01)^2, \quad (3.10)$$

$$C_2 = (L * 0.03)^2, \quad (3.11)$$

onde x e y são as imagens a serem avaliadas, μ_x é a média de x , μ_y é a média de y , σ_x^2 é a variância de x , σ_y^2 é a variância de y , σ_{xy} é a covariância de x e y . L é o maior valor que um pixel pode assumir dentro das imagens avaliadas, em imagens 8 bits esse valor seria 255 por canal.

Capítulo 4

Metodologia

Neste capítulo são detalhadas as etapas que foram realizadas neste trabalho para a execução dos experimentos, cujos os resultados são apresentados no Capítulo 5. Uma visão geral das etapas pode ser vista na Figura 4.1. Cada etapa é descrita com detalhes nas próximas seções, que estão divididas da seguinte maneira:

- Redimensionamento de imagens;
- Segmentação de imagens;
- Avaliação dos resultados.

4.1 Redimensionamento de imagens

As imagens usadas neste trabalho possuem dimensões e proporções entre seus eixos de coordenadas inconsistentes ao longo de toda a base de imagens. Para resolver este problema, foi necessário padronizar o tamanho das imagens que foram usadas, pois a arquitetura da rede SegNet trabalha apenas com grupos de imagens que possuem o mesmo tamanho. Nesta etapa, foram aplicados os métodos de interpolação apresentados no Capítulo 3, vizinho mais próximo, bilinear, bicúbico, lanczos e *area sampling*. Estes métodos foram escolhidos por serem parte da biblioteca OpenCV, uma biblioteca de código aberto para visão computacional. Para cada método foram aplicados os seguintes passos:

- Reduzir a resolução para a definida de 256 *pixels* na horizontal, mantendo as proporções originais, utilizando o método de interpolação da atual execução, para cada imagem.
- Acrescentar *pixels* pretos acima e abaixo de cada imagem, de forma a igualar à dimensão vertical da imagem à dimensão horizontal, sem afetar a proporção original da imagem;

Ao fim do pré-processamento, as imagens estarão em um formato amigável para a rede SegNet, que será tratada na próxima seção.

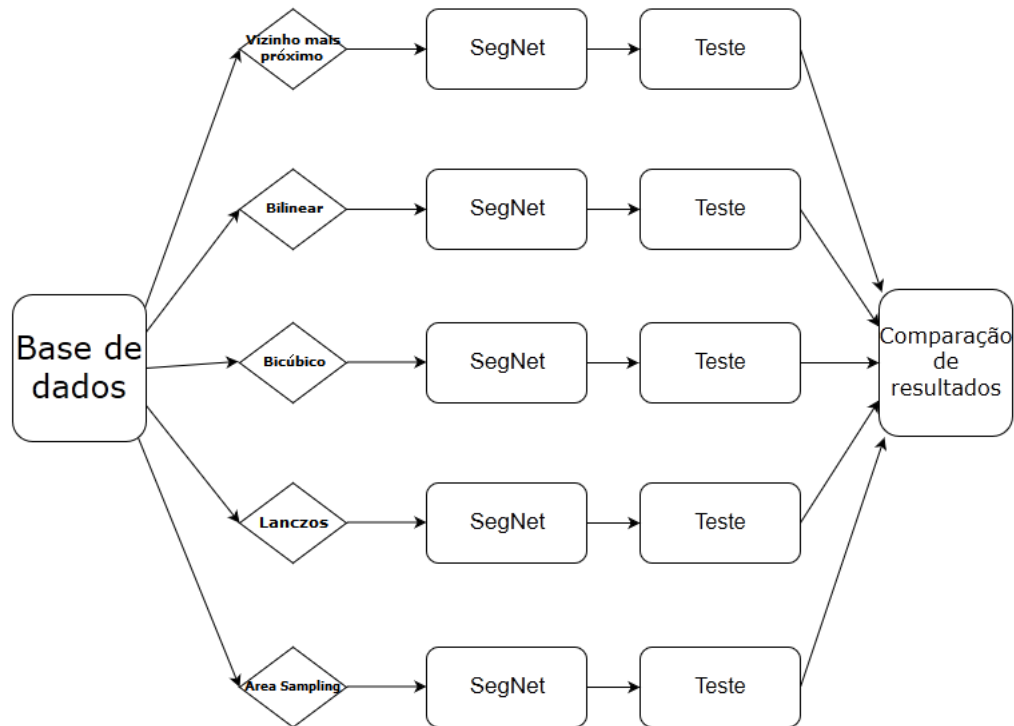


Figura 4.1: Visão geral das etapas que serão realizada neste trabalho.

4.2 Segmentação de Imagens

A segmentação de imagens de lesões de pele consiste de gerar uma máscara de segmentação com as regiões de interesse a partir de uma imagem de lesão de pele. Para a etapa de segmentação, foi utilizada a implementação da SegNet pela biblioteca Caffe, disponibilizada no GitHub¹. Para a utilização desta implementação, foi considerado um tutorial de instalação².

Para a utilização da SegNet é necessário que ela seja treinada a segmentar as regiões de interesse, de forma que ela aprenda a generalizar o procedimento de segmentação para o tipo de imagem desejada. Para o treinamento, a SegNet requer um conjunto de pares de imagens de treinamento e as imagens segmentadas correspondentes (*Ground Truth*). Das imagens resultantes da etapa de redimensionamento, cada método foi processado um por vez, de acordo com os passos descritos no Capítulo seguinte.

Após o treinamento, foi executada a etapa de testes, na qual um conjunto de pares de imagens de teste e suas imagens segmentadas correspondentes (*Ground Truth*) selecionadas pelo ISIC para este fim foram segmentadas pela rede recém treinada. Esta etapa retorna as máscaras de segmentação geradas pela rede, e que em seguida foram utilizadas para avaliar o

¹<https://github.com/TimoSaemann/caffe-segnet-cudnn5>. Acesso em: 31 de novembro de 2019.

²<http://mi.eng.cam.ac.uk/projects/segnet/tutorial.html>. Acesso em: 31 de maio de 2019.

desempenho da rede utilizando as métricas propostas na próxima subseção.

4.3 Avaliação dos resultados

Os resultados obtidos nas etapas de redimensionamento e segmentação foram avaliados considerando as métricas apresentadas no capítulo anterior, sendo elas Jaccard e SSIM. Para a comparação entre os resultados da etapa de redimensionamento foi utilizado apenas a métrica SSIM, onde cada imagem foi comparada com sua correspondente no original, e também entre cada método. Para avaliar a etapa de segmentação cada imagem resultante foi comparada com o *ground truth* e os resultados das métricas coletados. Então foram calculados os valores de média e desvio padrão das métricas Jaccard e SSIM, entre todas imagens resultantes de cada método de interpolação para avaliar se existe uma diferença significativa entre os resultados obtidos. Após a obtenção desses valores, foi possível avaliar quais métodos obtiveram os melhores resultados e quais obtiveram os piores.

Capítulo 5

Experimentos

Neste capítulo serão apresentados os resultados obtidos pelos testes preliminares e os experimentos realizados, bem como uma avaliação de seus resultados.

5.1 Bases de imagens

As imagens utilizadas para os testes preliminares foram obtidas a partir da base de dados de treinamento do ISIC 2018, composta por 2594 pares de imagens dermatoscópicas e seus respectivos *ground truths*. (Codella et al. (2017); Tschandl et al. (2018b)). Para o treinamento e teste as imagens foram obtidas a partir da base de dados do ISIC 2017 (Codella et al. (2017)). Esta base é composta por pares de imagens dermatoscópicas e seus respectivos *ground truth*, sendo 2000 pares para treinamento, 150 pares de validação, e 600 pares de testes. Ambas bases possuem imagem com resoluções entre 767×576 e 6688×4459 . Os pares de validação não são utilizados devido à forma com a qual a SegNet foi implementada. A base de imagens do ISIC 2017 foi utilizada pois a base mais recente, ISIC 2018, não disponibilizou as imagens *Ground Truth* da etapa de testes, dificultando a tarefa de avaliar a qualidade da segmentação. As imagens originais estão no formato JPEG (do inglês, *Joint Photographic Experts Group*) e no espaço de cores RGB (do inglês, *Red Green Blue*) e as imagens *ground truth* no formato PNG (do inglês, *Portable Network Graphics*) monocromático. Quatro exemplos de tais imagens podem ser vistos na Figura 5.1.

5.2 Ambiente de desenvolvimento

Para a realização dos experimentos, foi utilizado o Google *Cloud Platform* (GCP)¹. Google Cloud Platform é um serviço de instanciamento de máquinas virtuais na nuvem oferecido pela Google, onde é possível se instanciar uma máquina com configurações definidas pelo usuário,

¹<https://cloud.google.com/>

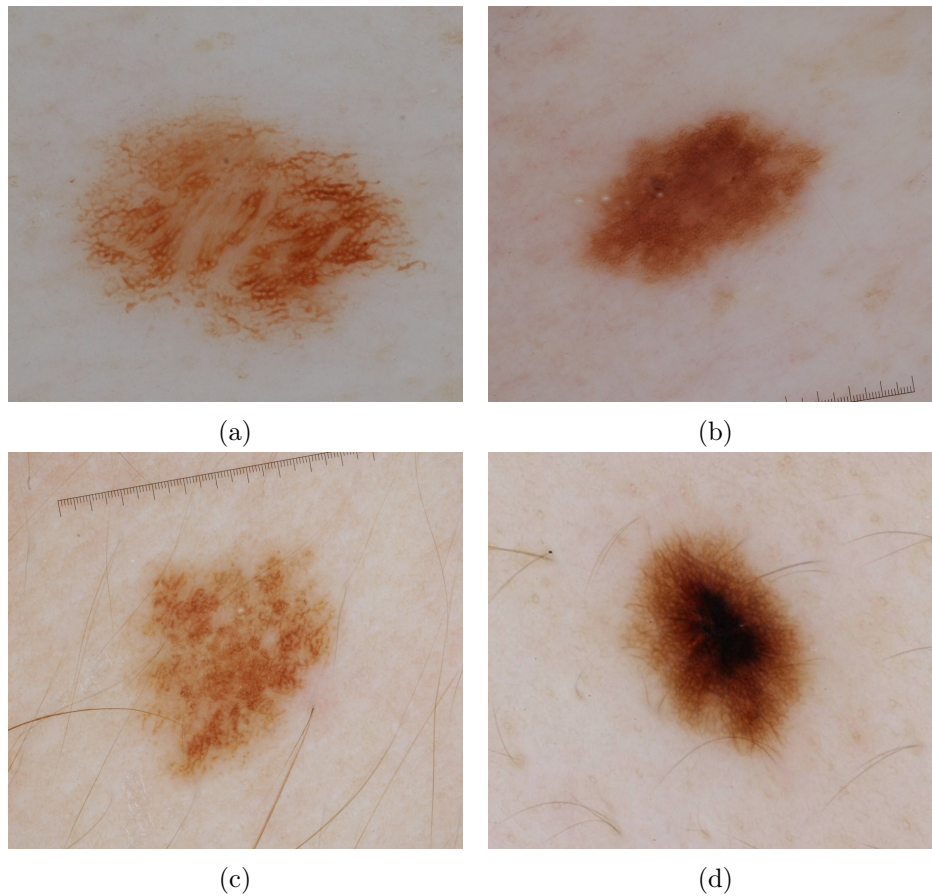


Figura 5.1: Exemplos de imagens dermatoscópicas (Codella et al. (2017); Tschandl et al. (2018a)).

selecionando partes como *Central Processing Unit (CPU)*, *Graphics Processing Unit (GPU)* e Memória, para utilizar como desejar.

Para armazenar os dados necessários para os experimentos, assim como os seus resultados, foi utilizado o Google Drive disponibilizado para os alunos da Universidade Federal de Ouro Preto (UFOP), que possui armazenamento ilimitado e pode ser acessado pela Google Cloud Platform de forma rápida, por fazer parte da infraestrutura da Google. Esta sinergia entre o Google Drive e o Google Cloud Platform foi útil, pelo tamanho do banco de dados utilizado e o volume de transferência de dados que ocorreu entre a máquina virtual e o armazenamento.

5.3 Hardware

Para os experimentos utilizando a SegNet foi instanciada no Google Cloud Platform uma máquina munida de CPU Xeon, 8Gbs de memória volátil e uma GPU Nvidia k80, com 12Gb de memória volátil dedicada. Tais configurações foram escolhidas com base na quantidade de memória necessária tal qual calculada pela SegNet durante sua inicialização.

Para a realização dos testes parciais foi utilizado uma máquina com um processador Phenom II x4 970, 12Gb de memória volátil e uma Amd Radeon RX 570 8Gb, esta máquina também foi utilizada para o redimensionamento das imagens utilizadas para o treinamento.

5.4 Estruturação dos arquivos

O ISIC disponibiliza três conjuntos de pares de imagens e *ground truths*, sendo o primeiro destinado à etapa de treinamento, o segundo para a etapa de validação, e o último para a etapa de teste. Visto que a SegNet não utiliza de um *dataset* de validação externo, este conjunto não será utilizado nestes experimentos. Após serem adquiridos pelo site do ISIC os pares de treinamento foram movidos para um diretório chamado *Dataset*, dentro deste diretório as imagens de treinamento originais foram movidas para o diretório *Train*, enquanto as imagens *Ground Truth* foram movidas para o diretório *TrainAnnot*. Os pares designados para teste, foram movidos para o diretório *Dataset/Test*, e as imagens *Ground Truth* foram movidas para o diretório *Dataset/TestAnnot*. Durante o processo de movimentar as imagens foi gerado um arquivo de texto para os conjuntos de treinamento e teste, com o caminho para cada imagem no formato [Caminho para *Imagem*][Caminho para *Ground Truth*], de forma a cada imagem estar associada a seu *Ground Truth*.

5.5 Testes Preliminares

Nesta seção são apresentados alguns resultados de testes preliminares, que auxiliaram na exemplificação da hipótese levantada neste trabalho. As imagens utilizadas foram obtidas a partir da base de dados do ISIC 2018 (Codella et al. (2017); Tschandl et al. (2018b)). Os resultados da etapa de redimensionamento foram avaliados por meio de dois testes iniciais. O primeiro teste foi feito comparando a imagem obtida pela aplicação de cada método de redimensionamento com a imagem original, e o segundo comparando as imagens obtidas pelo redimensionamento entre si. Para ambos os experimentos, foi utilizado a métrica SSIM, para representar os valores de semelhança entre as imagens.

5.5.1 Comparação com imagem original

As imagens redimensionadas foram comparadas com as originais, utilizando a métrica SSIM implementada na biblioteca *skimage.measure* para Python. Para tal, foram selecionadas aleatoriamente 727 imagens originais da base de dados ISIC para serem comparadas. Estas imagens foram reduzidas utilizando cada um dos métodos de interpolação apresentados anteriormente, um por vez. Em seguida, as imagens foram ampliadas novamente para a resolução da original utilizando o método vizinho mais próximo, com o objetivo de preservar as características da imagem reduzida. Desta forma, foi possível compará-las as imagens originais, por

meio do valor de SSIM. Na Tabela 5.1 são demonstrados os resultados de SSIM definidos para cada método, considerando a média e o desvio padrão de todas as imagens selecionadas.

Analisando estes resultados, o método *area sampling* se destaca, exibindo o maior valor de SSIM e conseqüentemente maior semelhança com a imagem original, além de um menor desvio padrão (valores em negrito na Tabela 5.1). Por outro lado, o método vizinho mais próximo exibiu a maior diferença entre a imagem redimensionada e a original, além de maiores valores de desvio padrão. Com estes valores, espera-se que os resultados do treinamento da rede SegNet irão ser mais precisos ao utilizar o método *area sampling* e menos precisos ao se utilizar vizinho mais próximo. Já no caso da utilização do bilinear, do bicúbico ou do lanczos, espera-se obter resultados intermediários entre os dois métodos citados anteriormente.

Tabela 5.1: Comparação entre imagens redimensionadas e originais.

-	Vizinho mais próximo	Bilinear	Bicúbico	Lanczos	<i>Area Sampling</i>
Média	0.824	0.889	0.883	0.882	0.903
Desvio Padrão	0.076	0.048	0.050	0.051	0.045

5.5.2 Comparação com imagens redimensionadas

Os resultados obtidos nos testes apresentados anteriormente foram utilizados para avaliar a similaridade entre os métodos, e avaliar quais demonstraram maior expectativa de resultados expressivos se comparados com os outros. Na Tabela 5.2, são exibidos os valores da média dos índices de SSIM dos métodos nas linhas pelos métodos nas colunas, de forma que o valor que coincide a linha de um método com a coluna de outro, é a média SSIM computada entre eles. Nos campos com valor 1 as imagens são idênticas, e portanto todos campos onde os métodos são comparados com si mesmo, irão possuir o valor 1, como por exemplo na linha do Vizinho mais próximo com a coluna Vizinho mais próximo. Nesta Tabela 5.2 pode ser visto que dentre todos métodos, o vizinho mais próximo é o que apresenta menores valores da média de SSIM ao ser comparado com todos outros métodos. Portanto, este método pode apresentar uma expectativa de resultados dissemelhantes aos outros métodos quando utilizado para treinar a rede SegNet. Outro dado a ser notado é o valor alto da média de SSIM entre os métodos lanczos e bicúbico, e o pequeno desvio padrão dos dados de SSIM coletados entre estes métodos, como demonstrado na Tabela 5.3. Isso demonstra que estes métodos produzem resultados similares ao longo de todas imagens, gerando grande expectativa de que os resultados do treinamento das respectivas redes, também sejam similares.

5.6 Treinamento e Teste

Os conjuntos de imagens gerados na etapa de redimensionamento foram utilizados para treinar a rede SegNet, cada conjunto gerando um modelo, considerando a base de dados

Tabela 5.2: Média dos valores de SSIM entre métodos de interpolação.

	Vizinho mais próximo	Bilinear	Bicúbico	Lanczos	<i>Area sampling</i>
Vizinho mais próximo	1	0.861	0.855	0.855	0.894
Bilinear	0.861	1	0.988	0.987	0.950
Bicúbico	0.855	0.988	1	0.992	0.942
Lanczos	0.855	0.987	0.992	1	0.941
<i>Area sampling</i>	0.894	0.950	0.941	0.941	1

Tabela 5.3: Desvio padrão dos valores de SSIM entre métodos de interpolação.

	Vizinho mais próximo	Bilinear	Bicúbico	Lanczos	<i>Area sampling</i>
Vizinho mais próximo	1	0.065	0.067	0.067	0.051
Bilinear	0.065	1	0.005	0.006	0.028
Bicúbico	0.067	0.005	1	0.003	0.030
Lanczos	0.067	0.006	0.003	1	0.030
<i>Area sampling</i>	0.050	0.028	0.030	0.031	1

do ISIC 2017 (Codella et al., 2017). A rede foi treinada utilizando as imagens no diretório *Train* e *Ground Truth*. Esta etapa é iniciada por um comando que consiste do caminho para o executável, compilado durante a instalação, e os argumentos “train”, para executar o treinamento da rede, “-gpu 0”, que determina o uso da GPU disponibilizada pelo GCP, e “-solver /SegNet/Models/segnet_solver.prototxt”, que determina qual estrutura da SegNet será utilizada. Após executado o comando, a rede inicia o treinamento, e exibe os valores de acurácia a cada 20 épocas de treinamento como definido pela implementação da SegNet. Ao chegar ao limite de 20 mil épocas, o treinamento foi encerrado pela rede e o modelo final armazenado em disco. O limite de 20 mil épocas foi definido durante testes preliminares no qual a rede foi treinada e avaliada, e foi apurado que 20 mil épocas proporciona resultados comparáveis à períodos mais longos de treino, sem perdas significativas de qualidade. Cada rede levou cerca de 14 horas para treinar totalmente em uma máquina com as configurações previamente descritas.

O conjunto de teste foi utilizado para avaliar todas redes geradas na etapa de Treinamento. Idealmente seria utilizado as imagens de teste em sua resolução original, porém devido a limitações da implementação utilizada pela biblioteca Caffe, as imagens em resolução original extrapolaram o limite de variáveis inteiras utilizadas pela arquitetura. Sendo assim, o conjunto de testes utilizado para testar todas redes foi redimensionado para a resolução 256×256 , igual ao utilizado para o treinamento, utilizando o método de interpolação Vizinho mais próximo, por este método não criar novos valores para os *pixels*, apenas replicando os valores já existentes. Este método foi utilizado para o teste de todos os modelos. Um exemplo de uma máscara *ground truth* e sua respectiva imagem segmentada pela rede pode ser vista na Figura 5.2.

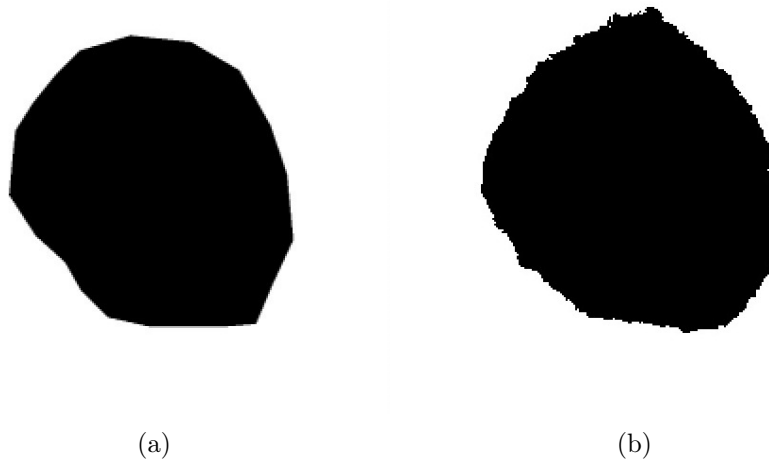


Figura 5.2: Um exemplo da qualidade da segmentação: (a) máscara *Ground Truth* e (b) máscara segmentada pela SegNet treinada anteriormente

5.7 Resultados

Após serem geradas as máscaras de segmentação do conjunto de testes, as mesmas foram comparadas com as respectivas máscaras *ground-truth* utilizando as métricas definidas anteriormente, Jaccard Index e SSIM. Os resultados foram adicionados a uma planilha e a média e desvio padrão, de cada métrica por rede foram calculados.

Após a execução dos testes preliminares na Seção 5.5, foi esperado que nestes experimentos, dentre todos os métodos apresentados, o método *Area Sampling* obtivesse os melhores resultados, e que o Vizinho mais próximo obtivesse os piores resultados. Como é possível ver nas Tabelas 5.4 e 5.5, os resultados obtidos são bem próximos, porém exibe um pequeno crescimento da qualidade dos resultados em relação à complexidade dos métodos de interpolação, com exceção do *Area Sampling* que apesar de ser menos complexo que Lanczos, conseguiu o melhor resultado na maior parte valores computados, tendo resultados ligeiramente piores que o Bicúbico na média do SSIM. Estes resultados se mostram similares aos resultados dos testes preliminares, mesmo que de forma menos expressiva, indicando que o método de interpolação utilizado para redimensionar as imagens de teste gera um impacto na qualidade das máscaras de segmentação geradas pelo modelo.

Tabela 5.4: Resultados Jaccard.

	Vizinho mais próximo	Bilinear	Bicubico	Lanczos	<i>Area</i>
Média	0,8425	0,8430	0,8464	0,8439	0,8472
Desvio Padrão	0,1432	0,1447	0,1405	0,1429	0,1400

É possível especular que a maior disparidade entre os métodos encontrados nos testes preliminares seja devida à métrica SSIM utilizada, que leva em conta toda a imagem e não

Tabela 5.5: Resultados SSIM.

	Vizinho mais próximo	Bilinear	Bicubico	Lanczos	<i>Area</i>
Média	0,8050	0,8055	0,8113	0,8061	0,8107
Desvio Padrão	0,1488	0,1509	0,1450	0,1478	0,1445

apenas o contorno da imagem. Portanto, levando em conta detalhes internos da lesão no seu cálculo, e sendo um indicativo de que este estudo pode ser expandido no futuro para incluir a etapa de segmentação de características locais das lesões, com possibilidade de revelar diferenças mais expressivas que as atualmente encontradas.

Capítulo 6

Conclusão

Neste trabalho, foram apresentados e comparados os impactos de diferentes métodos de interpolação de imagens, nos resultados da segmentação de uma lesão de pele em imagens dermatoscópicas, utilizando a rede neural convolucional SegNet. A relevância deste estudo se dá devido à grande letalidade do câncer de pele do tipo melanoma se não diagnosticado em estágio inicial, podendo tal diagnóstico ser auxiliado por técnicas computacionais de segmentação da lesão, visto que as bordas são um indicativo do tipo da lesão.

Ao analisar os resultados dos experimentos realizados no Capítulo 5, onde a rede SegNet foi treinada utilizando bases de imagens com diferentes métodos de interpolação usados para o redimensionamento, foi possível verificar que existe diferença entre os resultados dos métodos. Esta diferença, apesar de pequena, se mostrou congruente com os resultados dos experimentos preliminares, onde o mesmo comportamento foi notado, porém com maior expressividade das diferenças entre os métodos. Com isso, é possível concluir que, mesmo que pequeno, existe um impacto a ser considerado com os métodos de interpolação utilizados no pré-processamento das bases de dados a serem utilizadas para a tarefa de segmentação de lesões de pele, e que dentre os métodos de interpolação testados, o *Area Sampling* é o mais recomendado para esta tarefa.

Após o desenvolvimento do trabalho e a análise dos resultados foi possível levantar a hipótese de que os métodos de interpolação possam também ter impacto relevante na segmentação de características internas de uma lesão, uma vez que estas são menores em tamanho e mais suscetíveis a serem descartadas durante o processo de redimensionamento por métodos ingênuos de interpolação.

6.1 Trabalhos futuros

Com o aprendizado adquirido ao longo deste trabalho e com as informações consequentes dos resultados dos experimentos, novas teorias e questões foram levantadas, e assim esta última seção do trabalho serve o propósito de citar possíveis áreas a serem consideradas para trabalhos

futuros. A comparação dos métodos *Area Sampling*, Bilinear e Vizinho mais próximo na tarefa de segmentação de características locais das lesões de pele, possui potencial para exibir resultados mais expressivos que os encontrados neste trabalho. Opcionalmente implementar e utilizar o método "*Perceptually Based Downscaling*" (Öztireli e Gross, 2015) para comparar com os resultados aqui gerados e também para características locais, pois este tende a preservar pequenas características de imagens.

Referências Bibliográficas

- Abobakr, A.; Hossny, M. e Nahavandi, S. (2018). Ssimlayer: Towards robust deep representation learning via nonlinear structural similarity. *CoRR*, abs/1806.09152.
- Al-masni, M. A.; Al-antari, M. A.; Choi, M.-T.; Han, S.-M. e Kim, T.-S. (2018). Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. *Computer Methods and Programs in Biomedicine*, 162:221 – 231.
- Badrinarayanan, V.; Kendall, A. e 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.
- Baghersalimi, S.; Bozorgtabar, B.; Schmid-Saugeon, P.; Ekenel, H. e Thiran, J.-P. (2019). Dermonet: densely linked convolutional neural network for efficient skin lesion segmentation. *EURASIP Journal on Image and Video Processing*, 2019.
- Celebi, M. E.; Iyatomi, H.; Schaefer, G. e Stoecker, W. (2009). Lesion border detection in dermoscopy images. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, 33:148–53.
- Codella, N. C. F.; Gutman, D.; Celebi, M. E.; Helba, B.; Marchetti, M. A.; Dusza, S. W.; Kalloo, A.; Liopyris, K.; Mishra, N. K.; Kittler, H. e Halpern, A. (2017). Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (ISIC). *CoRR*, abs/1710.05006.
- Dodge, S. e Karam, L. (2016). Understanding how image quality affects deep neural networks. pp. 1–6.
- Esteva, A.; Kuprel, B.; Novoa, R. A.; Ko, J.; Swetter, S. M.; Blau, H. M. e Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542:115 EP –.

- Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V. e Rodríguez, J. G. (2017). A review on deep learning techniques applied to semantic segmentation. *CoRR*, abs/1704.06857.
- Getreuer, P. (2011). Linear Methods for Image Interpolation. *Image Processing On Line*, 1:238–259.
- INCA (2018). *Estimativa 2018: incidência de câncer no Brasil*. Instituto Nacional de Câncer José Alencar Gomes da Silva. Coordenação de Prevenção e Vigilância.
- Koziarski, M. e Cyganek, B. (2018). Impact of low resolution on image recognition with deep neural networks: An experimental study. *International Journal of Applied Mathematics and Computer Science*, 28:735–744.
- Krizhevsky, A. (2009). Learning multiple layers of features from tiny images. Technical report.
- Mishra, R. e Daescu, O. (2017). Deep learning for skin lesion segmentation. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1189–1194.
- Niwattanakul, S.; Singthongchai, J.; Naenudorn, E. e Wanapu, S. (2013). Using of jaccard coefficient for keywords similarity.
- Oliveira, R. B.; Filho, M. E.; Ma, Z.; Papa, J. P.; Pereira, A. S. e Tavares, J. M. R. (2016). Computational methods for the image segmentation of pigmented skin lesions: A review. *Computer Methods and Programs in Biomedicine*, 131:127 – 141.
- Öztireli, A. C. e Gross, M. (2015). Perceptually based downscaling of images. *ACM Trans. Graph.*, 34(4):77:1–77:10.
- Parsania, P. e Virparia, D. (2016). A comparative analysis of image interpolation algorithms. *IJARCCCE*, 5:29–34.
- Qian, C.; Liu, T.; Jiang, H.; Wang, Z.; Wang, P.; Guan, M. e Sun, B. (2018). A detection and segmentation architecture for skin lesion segmentation on dermoscopy images. *CoRR*, abs/1809.03917.
- Rajrapollu, P. e Mankar, V. (2017). Bicubic interpolation algorithm implementation for image appearance enhancement. *www.ijcst.com*, 8:23–26.
- Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V. e Rabinovich, A. (2015). Going deeper with convolutions. In *Computer Vision and Pattern Recognition (CVPR)*.

- Tsao, H.; Olazagasti, J.; M Cordoro, K.; Brewer, J.; Taylor, S.; Bordeaux, J.; Chren, M.-M.; J Sober, A.; Tegeler, C.; Bhushan, R. e Smith Begolka, W. (2015). Early detection of melanoma: Reviewing the abcdes. *Journal of the American Academy of Dermatology*, 72.
- Tschandl, P.; Rosendahl, C. e Kittler, H. (2018a). The ham10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5.
- Tschandl, P.; Rosendahl, C. e Kittler, H. (2018b). The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions. *CoRR*, abs/1803.10417.
- Wang, J.; Jiang, H.; Yuan, Z.; Cheng, M.-M.; Hu, X. e Zheng, N. (2017). Salient object detection: A discriminative regional feature integration approach. *International Journal of Computer Vision*, 123(2):251–268.
- Zhou Wang; Bovik, A. C.; Sheikh, H. R. e Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

Siglas

CPU *Central Processing Unit.* 20

GCP *Google Cloud Platform.* 19, 23

GPU *Graphics Processing Unit.* 20, 23

JPEG *Joint Photographic Experts Group.* 19

PNG *Portable Network Graphics.* 19

RGB *Red Green Blue.* 19

SSIM *Structural Similarity.* viii, 5, 6, 14, 15, 18, 21–24