

Treinando Rede Neural Profunda com Divisão Proporcional de Imagens para Segmentação de Estruturas da Retina

Pedro Victor de A. Fonseca, Alexandre Carvalho Araújo, João Dallyson S. de Almeida
Geraldo Braz Júnior

Universidade Federal do Maranhão (UFMA) Caixa Postal – 65080-805 – São Luís – MA
pedro.vaf@discente.ufma.br, {alexandrearaujo, jdallyson, geraldo}@nca.ufma.br

Abstract. *With the increase in the number of pathologies related to the human eye, the segmentation of the cup and the optical disc has become the main object of study by experiments linked to Deep Learning, aiming to improve the classification of the same structures allowing a better identification. This study proposes an approach for cup and optic disc segmentation that combines the proportional image splitting technique, concerning the segmentation area, and the U-Net network architecture with ResNet-34 encoder. The proposed approach showed promising results, achieving 96% of Dice in the optic disc segmentation in the RIM-ONE and DRISHTI-GS datasets and 90% and 85% of Dice in the optic cup segmentation in the DRISHTI-GS and RIM-ONE datasets, respectively.*

Resumo. *Com o aumento da quantidade de patologias relacionadas ao olho humano, a segmentação da escavação e do disco óptico se tornaram principais objetos de estudo por parte de experimentos interligados a Deep Learning, visando um aperfeiçoamento na classificação das mesmas estruturas possibilitando uma melhor identificação. Este estudo propõe uma abordagem de segmentação da escavação e do disco óptico combinando a técnica de divisão proporcional de imagem, em relação a área de segmentação, e a arquitetura de rede U-Net com encoder da ResNet-34. A abordagem proposta apresentou resultados promissores, alcançando 96% de Dice na segmentação do disco nos datasets RIM-ONE e DRISHTI-GS e 90% e 85% de Dice na segmentação das escavação nos datasets DRISHTI-GS e RIM-ONE, respectivamente.*

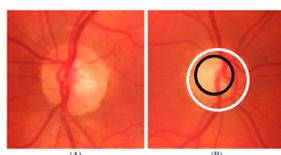
1. Introdução

O olho humano é um dos órgãos essenciais para a compreensão de mundo, composto de diversas estruturas e pela sua importância, diversos estudos são realizados para averiguar as patologias que afetam o globo ocular visando um melhor entendimento sobre o mesmo. Um dos exemplos mais predominantes são a do Glaucoma e da Catarata que representam uma grande quantidade de casos que acabaram levando pacientes a perda parcial ou total da visão [Collaborators and Study 2021, Bourne et al. 2021], com destaque ao Glaucoma que é previsto como causa principal de cegueira daqui a algumas décadas devido a sua capacidade de não ser uniforme e ser silenciosa, com expectativa de 118,8 milhões de pessoas afetadas por esta patologia [Wu et al. 2022]. Os exames realizados neste órgão comumente são realizados por especialistas que verificam se há alguma anormalidade nas áreas da escavação ou do disco óptico. Devido a um esforço necessário por ação humana para identificação destes problemas, algoritmos que atuam com a segmentação e

classificação relacionados as estruturas do olho tem sido desenvolvidos e utilizados com objetivo de auxiliar os especialistas [Lima et al. 2018, Silva et al. 2018, Moreira et al. 2021].

A segmentação da escavação e do disco óptico pode se provar uma tarefa desafiadora até mesmo para redes neurais mais refinadas, especialmente a da escavação óptica devido ao tamanho, tonalidade e a presença de vasos sanguíneos na região que acabam dificultando a classificação pixel a pixel de tais estruturas [Yu et al. 2019] que podem ser vistas na Figura 1. Devido a esses detalhes, diversas técnicas de processamento de imagem tem sido utilizadas, exemplos a serem citados incluem: a extração de uma região de interesse (ROI) que destaca uma parte da imagem original contendo o disco e a escavação óptica [Yu et al. 2019, Lima et al. 2020, Sevastopolsky 2017] e a extração do canal verde da imagem visando um realce das mesmas estruturas [Lima et al. 2020]. Outro fator de relevância é a utilização de arquiteturas utilizando *Deep Learning* que proporcionaram a geração de modelos de aprendizado próximos ao estado-da-arte.

Figura 1. (A) representa a área contendo o disco e escavação óptica enquanto que (B) delimita essas áreas. O círculo maior é o disco e o menor é a escavação.



Adaptado de [Yu et al. 2019]

Zilly et al. (2017) propôs uma arquitetura baseada em um aprendizado composto de conjuntos de camadas de convolução, baseando-se fortemente em técnicas de entropia com objetivo de maximizar a eficiência da rede e diminuir sua complexidade. Tal arquitetura se mostrou eficiente atingindo 97% de coeficiente de *Dice* na segmentação do disco óptico. Sevastopolsky (2017) utilizou a arquitetura modificada da U-Net para segmentar as mesmas estruturas, partindo da segmentação do disco e extraíndo uma ROI que foi utilizada na segmentação da escavação. Al-Bander et al. (2018) propuseram a utilização da arquitetura DenseNET em conjunto com uma rede convolucional para a segmentação das mesmas estruturas e detecção do glaucoma com treinamentos utilizando o dataset ORIGA apresentando bons resultados nos datasets DRISHTI-GS e RIM-ONE. Afolabi et al. (2019) utilizaram a U-Net padrão com objetivo de segmentar o disco e a escavação óptica nos datasets do DRISHTI-GS e RIM-ONE. Yu et al. (2019) utilizaram a U-Net em conjunto com as camadas pré-treinadas das ResNet-34 para a obtenção de resultados robustos nos datasets DRISHTI-GS, RIM-ONE e RIGA utilizando a técnica de extração de uma ROI facilitando assim a segmentação e identificação de glaucoma.

Contudo, mesmo os trabalhos citados obtendo bons resultados, ainda há espaço para melhoria no que diz respeito a segmentação de áreas pequenas. Nesse contexto, propomos uma nova abordagem de segmentação do disco e escavação da retina utilizando a técnica de divisão proporcional de imagens com base na região de segmentação com uso da arquitetura U-Net com encoder ResNet-34 apresentada por Yu et al. (2019). O método proposto inclui uma etapa inicial de separação proporcional das imagens utilizadas no treino/validação do modelo, em função das áreas do disco e escavação óptica, buscando

um balanceamento na quantidade de amostras dispostas em cada conjunto de imagens visando melhorar a generalização do modelo de segmentação produzido.

2. Materiais e Método proposto

Esta seção apresenta informações referentes ao conjunto de imagens utilizadas, o método proposto, a configuração dos experimentos e as métricas de avaliação utilizadas.

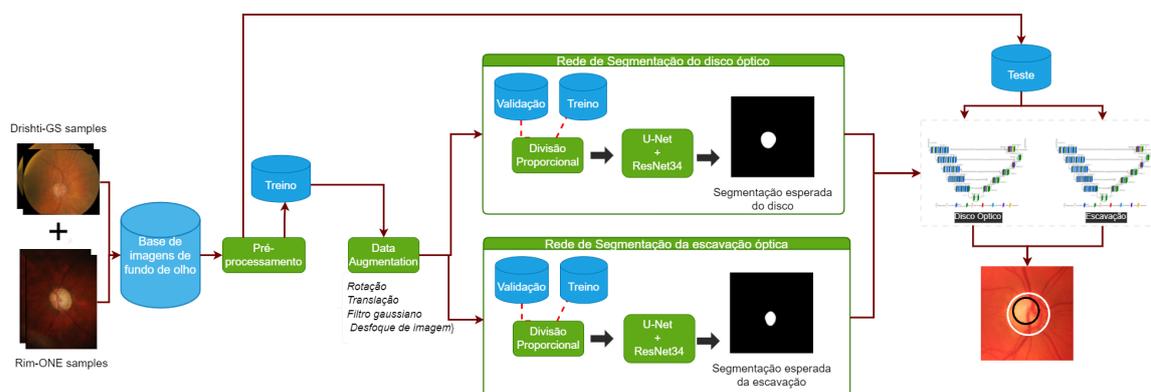
2.1. Base de Imagens

Nos experimentos foram selecionados os datasets públicos de imagens fundo de olho DRISHTI-GS [Sivaswamy et al. 2015] e RIM-ONE [Fumero et al. 2015], com DRISHTI-GS apresentando 101 imagens em sua totalidade de proporção 2896x1944 pixels e com um *field-of-view* de 30 graus, enquanto que o RIM-ONE apresenta 159 imagens de proporção 2144x1424 pixels. Cada imagem foi utilizada sem cortes, sem extração de ROI, diferenciando-se, assim, dos trabalhos utilizados como referência. As bases foram utilizadas em conjunto para o processo de treino.

2.2. Método proposto

O método proposto está organizado nas seguinte etapas: pré-processamento, *Data Augmentation*, divisão proporcional, arquitetura utilizada e treino/teste da arquitetura. A Figura 2 ilustra as etapas do método proposto. Na etapa de pré-processamento as imagens foram redimensionadas para a proporção de 256 x 256 pixels. Devido ao redimensionamento, as imagens perderam o nível de detalhe quando comparado com as dimensões originais, o que gerou deformação nas estruturas visadas para segmentação.

Figura 2. Etapas do método proposto.



2.2.1. Data augmentation

Com objetivo de aumentar o conjunto de imagens previamente obtido, novas imagens foram geradas realizando operações de rotação, alteração de contraste e desfoque, através da ferramenta Albumentation [Buslaev et al. 2020]. Este gerador movimentava as imagens

em ângulos de 45° em um limite de escala de 0.5, gira as imagens de forma vertical e horizontal em uma probabilidade de 50% e também nessa mesma probabilidade aplica filtros e efeitos de iluminação como o *ruído gaussiano*. Importante ressaltar que devido ao valor de 50% estabelecido previamente, cada imagem pode ou não ser modificada com todas as operações determinadas.

Para a definição de um conjunto de teste para o dataset da RIM-ONE, das 159 imagens originalmente contidas no mesmo dataset, um percentual de 20% dessas imagens foram destinadas para a criação de um conjunto de teste, totalizando 30 imagens da RIM-ONE para avaliação pós-treino e 129 imagens do mesmo dataset para treinamento. De forma geral, 129 amostras da RIM-ONE em conjunto com 50 imagens do DRISHTI-GS, totalizando 179 imagens, prosseguiram para serem distribuídas em 10 *folds* que foram utilizados durante a técnica *k-folding* que visa garantir um processo de validação mais confiável e robusto [Rodriguez et al. 2009]. Após a repartição, cada *fold* passou por uma etapa de *Data Augmentation* resultando, ao término do procedimento, um total de 4.180 imagens somando todas as amostras em cada *fold*. Importante ressaltar que os datasets de teste permaneceram inalterados sem *Data Augmentation*.

2.2.2. Divisão proporcional

Os datasets utilizados apresentam para cada imagem as delimitações do disco e da escavação óptica que podem ser utilizadas para treinamento, com destaque para as máscaras do DRISHTI-GS que são compostas de diversas marcações feitas por especialistas. Para fins de treino, para cada imagem do DRISHTI-GS foi selecionada do conjunto de delimitações uma que representasse a máscara de forma com que a menor região fosse escolhida, tal critério foi o padrão utilizado para todas as outras imagens, assim trabalhando com as marcações de um mesmo especialista. Já o dataset da RIM-ONE há apenas uma marcação para cada imagem.

A técnica de divisão proporcional foi utilizada para melhorar a distribuição de imagens entre os datasets de treino e validação, visando evitar um desbalanceamento entre imagens que possuem uma região de interesse grande em comparação com as pequenas. Inicialmente todo o conjunto de imagens disponíveis para treino é percorrido com o objetivo de obter o valor média de pixels brancos que referenciam a região de interesse contida nas máscaras, após isso, cada imagem é categorizada como “grande região de interesse” ou “pequena região de interesse” de acordo com sua quantidade de pixels brancos quando comparado com a média geral. Antes da atribuição ao respectivo dataset (treino ou validação) verifica-se a quantidade de imagens contidas no mesmo de forma com que não haja desbalanceamento excessivo para apenas um tipo de região de interesse. A distribuição perfeita seria no caso 50% de imagens com grande região e 50% com pequena região, contudo tal tarefa é muito difícil de ser realizada devido a disposição original dos datasets do DRISHTI-GS e RIM-ONE que são originalmente desbalanceados. Com esta observação, a divisão proporcional, para os datasets utilizados, apresenta uma variação de 20% com relação ao balanceamento perfeito no pior caso de distribuição.

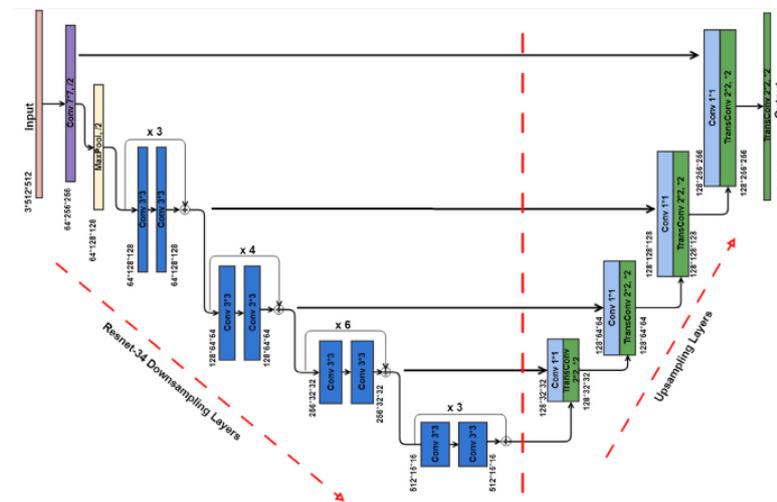
Para cada subconjunto de imagens distribuídas utilizando esta técnica, é retornado uma lista contendo imagens da forma mais balanceada possível, possibilitando um aprendizado mais diversificado e estável, contribuindo para a melhor generalização do modelo.

2.2.3. Arquitetura da Rede

A arquitetura da rede neural utilizada foi a U-Net [Ronneberger et al. 2015] modificada com o *encoder path* baseado na ResNet-34, como proposto em [Yu et al. 2019]. A estrutura padrão da U-net se faz presente nas camadas de decodificação *up-sampling* enquanto que a de *down-sampling* apresenta o modelo da ResNet-34 contendo uma camada convolucional 7x7 e 64 filtros. Após isso há 4 blocos residuais contendo cada bloco duas camadas convolucionais 3x3 apresentando *shortcuts* entre cada conexão.

Cada bloco residual apresenta, respectivamente, 3 conexões no primeiro bloco, 4 no segundo, 6 no terceiro e 3 no último, com cada bloco de 3 conexões possuindo *Batch Normalization*, além de um aumento exponencial na quantidade de filtros por um fator de 2, assim iniciando com 64 filtros e terminando em 512 no último bloco. Cada bloco de decodificação contém uma camada transposta de convolução 2x2 obtida da camada anterior com um stride de 2, concatenado com uma convolução 1x1 para cada respectivo bloco contido na seção de *down-sampling*. O tensor já concatenado passa pelo processo de *batch normalization* antes de prosseguir para a próxima camada de decodificação. Por fim, na última camada ocorre a última convolução transposta que retorna como saída uma imagem com as mesmas dimensões da imagem de entrada. A Figura 3 demonstra a arquitetura completa da rede neural.

Figura 3. Arquitetura da rede neural utilizada.



Adaptada de [Yu et al. 2019]

2.3. Experimentos

No treinamento da rede definiu-se a entropia cruzada binária ponderada como função *loss* visando atuar com uma melhor proporção dos pesos contidos em cada imagem e sua respectiva máscara, atribuindo para cada classe (0 e 1) um peso correspondente, tal função é descrita pela Equação 1. A escolha de tal função possibilitou evitar desbalanceamento de classes que poderiam ocorrer em qualquer outra função de *loss*, além de que existe suporte para a mesma em bibliotecas amplamente utilizadas quando se tratado de *deep*

learning [Pedregosa et al. 2011].

$$Loss = \left(\alpha \times \mathbf{t} \times \mathbf{p} + \alpha \times \left((\mathbf{t} - 1) \times (\mathbf{p} - 1) \right) \right) - (\alpha - 1)(1)$$

Onde α permanece no valor negativo de -1, “ t ” representa o valor alvo, a máscara real da imagem de acordo com especialistas, e “ p ” é uma imagem de predição que foi gerada pela rede. Foram utilizadas como métricas de avaliação, o coeficiente de *Dice* (D) descrito na Equação 2, que visa medir a similaridade entre duas amostras, no contexto deste trabalho, a similaridade entre as predições geradas e os valores de referência, e a métrica de *Jaccard* (J) (Equação 3) que é calculada pela divisão dos pixels da predição com relação a verdadeira máscara da imagem.

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

$$JaccardScore = \frac{TP}{TP + FP + FN} \quad (3)$$

Na qual TP representa pixels corretamente apontados como sendo da escavação ou disco, FP representa pixels erroneamente classificados como escavação ou disco e FN pixels classificados de forma errada como não sendo disco ou escavação. Cada *fold* foi treinado por 30 épocas com um batch size de valor 4 em conjunto com uma taxa de aprendizado estabelecida em 1e-4 com decaimento de 0.5 a cada 10000 passos, além disto, a técnica de Early Stopping foi utilizada com a paciência definida para 3 épocas [Li et al. 2020]. Todos os experimentos foram realizados em um hardware configurado com placa de vídeo NVIDIA GeForce GTX 1660 TI com 6 GB de RAM.

3. Resultados e Discussão

Esta seção descreve os resultados obtidos com os treinamentos da arquitetura proposta na segmentação da escavação e do disco óptico nos datasets de teste do DRISHTI-GS e da RIM-ONE, através da avaliação dos modelos utilizando métricas previamente definidas na seção anterior. O dataset do DRISHTI-GS apresentava, inicialmente, 50 imagens para treino e 51 para testes enquanto que o RIM-ONE possuía 159 imagens sem divisão, após os processos de augmentation a quantidade de imagens para o dataset do DRISHTI-GS era de 1150 enquanto da RIM-ONE 3030 imagens.

Com a arquitetura definida, experimentos foram realizados utilizando duas outras arquiteturas distintas, mais especificamente a U-Net modificada com a Resnet-50 e a U-Net integrada com a MobileNetV2. Essas arquiteturas foram escolhidas para avaliar como a metodologia proposta performava em uma rede residual mais profunda e com mais camadas (Resnet-50) e em uma arquitetura com poucos parâmetros para treinamento (MobileNetV2). A Tabela 1 e a Tabela 2 mostram as métricas obtidas por cada rede na tarefa de segmentação do disco e da escavação óptica nos datasets de teste do DRISHTI-GS e da RIM-ONE, respectivamente, com e sem uso da técnica de divisão proporcional.

Com as métricas dispostas, o modelo proposto da U-Net com a Resnet-34 utilizando a divisão proporcional, provou-se ser o mais efetivo para a segmentação no dataset de teste do DRISHTI-GS quando comparado com outras arquiteturas, em especial quando

Tabela 1. Segmentação do disco e escavação óptica
Dataset de teste do DRISHTI-GS. (*) representa com divisão proporcional aplicada

Arquitetura	Disco		Escavação	
	Dice score	IOU score	Dice score	IOU score
U-Net com Resnet-50	95,4	92,2	89,0	78,9
U-Net com Resnet-50 (*)	96,0	93,1	90,2	80,4
U-Net com MobileNetV2	94,3	91,1	88,0	77,3
U-Net com MobileNetV2 (*)	95,0	91,0	89,2	79,1
U-Net com Resnet-34	96,0	92,2	89,5	79,9
U-Net com Resnet-34 (*)	96,1	92,4	90,4	81,6

comparado com os valores obtidos para a escavação óptica. Importante ressaltar que para cada arquitetura, as métricas obtidas para a segmentação de ambas as estruturas foram maiores em treinamentos com a divisão proporcional aplicada.

Tabela 2. Segmentação do disco e escavação óptica
Dataset de teste da RIM-ONE. (*) representa com divisão proporcional aplicada

Arquitetura	Disco		Escavação	
	Dice score	IOU score	Dice score	IOU score
U-Net com Resnet-50	96,0	92,0	84,6	72,5
U-Net com Resnet-50 (*)	96,0	92,4	85,0	73,3
U-Net com MobileNetV2	95,4	91,0	81,9	67,5
U-Net com MobileNetV2 (*)	96,0	91,1	83,4	68,7
U-Net com Resnet-34	95,2	91,3	85,1	73,4
U-Net com Resnet-34 (*)	96,3	91,8	85,4	73,6

O modelo proposto apresenta resultados positivos na segmentação de ambas as estruturas com destaque para a métrica de *dice* obtida para o disco óptico, importante notar que o mesmo se destaca na segmentação da escavação. Novamente, a aplicação da divisão proporcional se mostra benéfica com relação as métricas obtidas para todas as arquiteturas.

Após a confirmação de eficiência da U-Net modificada com a Resnet-34, novos experimentos foram realizados com a mesma arquitetura. A Tabela 3 mostra os valores de *Dice* e *IOU* (*Jaccard Score*) para cada *fold* de validação que contém imagens do dataset do DRISHTI-GS e RIM-ONE durante o treinamento da técnica *k-fold* usando marcações do disco e da escavação óptica, apresentando ao final uma média das métricas alcançadas.

Obtendo como média de métricas 94% para coeficiente de Dice e 90% para o de Jaccard no treinamento para o disco óptico pode-se notar uma consistência nos resultados, dentre estes, o maior valor obtido por *fold* foi de 97% para Dice e 94% para IOU visto no *fold* de número nove.

Já para o treinamento com a escavação óptica, obteve-se ao valor de 88% para a métrica de Dice enquanto que a de Jaccard obteve 80%. Importante ressaltar que o maior valor obtido durante o treinamento foi visto no *fold* de número oito.

Tanto para segmentação da escavação quanto para o disco óptico, o modelo pré-treinado da U-Net com o backbone da Resnet-34 foi utilizado, modificando apenas o conjunto de imagens dispostas para treinamento e validação. A Tabela 4 demonstra os

Tabela 3. Resultados da técnica *K-fold* para o disco e escavação óptica usando divisão proporcional
Treino com a arquitetura da U-net com Resnet-34

N° do <i>fold</i>	Disco		Escavação	
	Dice score	IOU score	Dice score	IOU score
1	0,94	0,87	0,79	0,70
2	0,92	0,85	0,82	0,72
3	0,95	0,90	0,88	0,81
4	0,96	0,94	0,89	0,81
5	0,97	0,93	0,90	0,83
6	0,94	0,90	0,91	0,81
7	0,96	0,94	0,91	0,82
8	0,93	0,89	0,92	0,85
9	0,97	0,94	0,91	0,84
10	0,94	0,92	0,90	0,84
Média	0,94	0,90	0,88	0,80
Desvio \pm	0,08	0,09	0,03	0,05

resultados obtidos no dataset de teste do DRISHTI-GS em conjunto com os trabalhos utilizados como referência, ressaltando que o arranjo de teste utilizado contém 50 imagens que são disponibilizados por padrão pelo próprio dataset.

Tabela 4. Segmentação do disco e escavação óptica
Dataset de teste do Drishti-GS

Método	Disco		Escavação	
	Dice score	IOU score	Dice score	IOU score
[Zilly et al. 2017]	97,3	91,4	87,1	85,0
[Sevastopolsky 2017]	-	-	85	75
[Al-Bander et al. 2018]	94,9	90,4	82,8	71,1
[Yu et al. 2019]	96,4	94,2	88,7	78,0
[Afolabi et al. 2019]	-	-	95	79
Método proposto	96,1	92,4	90,4	81,6

Na segmentação do disco, nosso método alcança resultados próximos aos dos trabalhos anteriores que obtiveram suas respectivas métricas para a segmentação da estrutura com destaque para a obtenção de valores superiores para a escavação quando comparados com os resultados obtidos por Yu et al. (2019) utilizando-se a mesma arquitetura. Quanto a análise das métricas referentes a segmentação da escavação óptica, o modelo proposto conseguiu ultrapassar o valor de 90% no coeficiente de Dice aproximando-se ainda mais do melhor valor disposto por Afolabi et al. (2019) enquanto que está a quatro pontos percentuais abaixo do maior valor de Jaccard obtido por Zilly et al. (2017).

A Figura 4 demonstra a segmentação do disco de uma imagem contida no dataset de teste do DRISHTI-GS utilizando o modelo da U-Net com a ResNet-34 que obteve as métricas de 96,1% para Dice e 92,4% no dataset de teste descrito na tabela anterior. Por ser composta, geralmente, de uma área de interesse grande, a segmentação do disco não oferece um grande desafio para as redes neurais convolucionais.

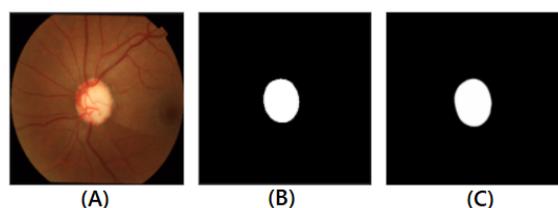


Figura 4. Segmentação do disco óptico. (A) Imagem original, (B) máscara real e (C) predição

A segmentação da escavação pode ser vista na Figura 5, tarefa esta que o modelo proposto obteve 90,4% de Dice e 81,6% de IOU. Devido a diversos fatores já descritos, a segmentação da escavação se prova difícil para as mais diversas arquiteturas.

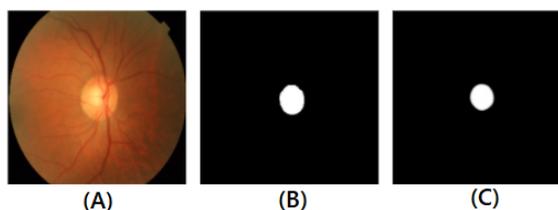


Figura 5. Segmentação da escavação óptica. (A) Imagem original, (B) máscara real e (C) predição

Utilizando o modelo pré-treinado da U-Net com a Resnet-34 para avaliação no dataset de teste da RIM-ONE obtêm-se as métricas dispostas na Tabela 5. O mesmo modelo obteve 96,1% de coeficiente de Dice e 93,3% de IOU para o disco óptico durante o treinamento e 87,3% de Dice e 75% de IOU para a escavação.

Tabela 5. Segmentação do disco e escavação óptica
Dataset de teste da RIM-ONE

Método	Disco		Escavação	
	Dice score	IOU score	Dice score	IOU score
[Zilly et al. 2017]	94,2	89,0	82,4	80,2
[Sevastopolsky 2017]	95	89	82	69
[Al-Bander et al. 2018]	90,3	82,8	69,0	55,6
[Yu et al. 2019]	96,1	92,5	84,4	74,2
[Afolabi et al. 2019]	96,1	88,3	89,0	76,2
Método proposto	96,3	91,8	85,4	73,6

O modelo proposto apresenta a maior métrica na segmentação do disco óptico quando são comparados em função do *Dice*. Em relação a escavação óptica, as métricas obtidas ficam abaixo dos maiores valores apresentados na tabela, arredondando os valores temos que o modelo fica abaixo em quatro pontos percentuais para *Dice* e sete pontos para Jaccard. Importante ressaltar que o método proposto apresenta métricas maiores de *Dice* do disco e escavação óptica quando comparado com os resultados obtidos por Yu et al. (2019) que utilizou a mesma arquitetura. As Figuras 6 e 7 demonstram, visualmente,

a segmentação do disco e da escavação óptica de imagens pertencentes ao dataset de teste da RIM-ONE.

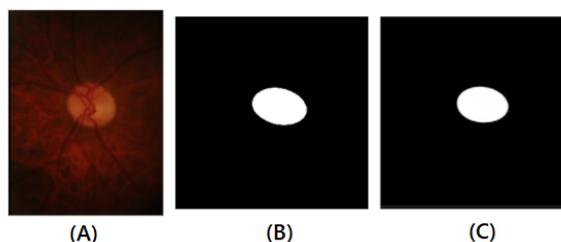


Figura 6. Segmentação do disco óptico. (A) Imagem original, (B) máscara real e (C) predição

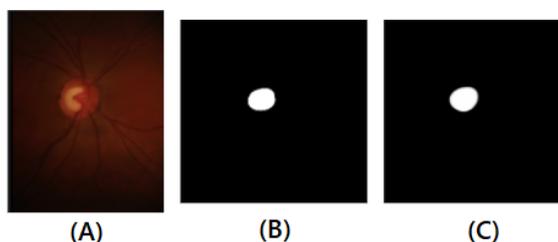


Figura 7. Segmentação da escavação óptica. (A) Imagem original, (B) máscara real e (C) predição

4. Conclusão

O trabalho apresentado visou demonstrar o uso da técnica de divisão proporcional buscando aperfeiçoar métricas previamente existentes quando posto em prática a segmentação do disco e da escavação óptica nos datasets do DRISHTI-GS e RIM-ONE. Diversos experimentos utilizando o dataset do DRISHTI-GS e da RIM-ONE foram conduzidos visando melhores resultados, tais experimentos foram realizados com arquiteturas e configurações de *Data Augmentation* distintas. Os resultados dispostos nesse trabalho foram reunidos de acordo com os melhores parâmetros que possibilitaram a maior métrica encontrada pelo modelo proposto até o momento.

De forma geral, os experimentos conduzidos com o modelo da U-Net modificada com a ResNet-34, em conjunto com a divisão proporcional, demonstraram através dos resultados apresentados que um processo de aprendizagem sólido e distribuído com relação as amostras auxiliam na obtenção de métricas positivas quanto a segmentação da escavação e disco óptico. Importante ressaltar a comparação das métricas obtidas pelo método proposto com os valores obtidos por Yu et al. (2019), que utilizou a mesma arquitetura para segmentar o disco e a escavação óptica, onde o nosso método apresentou resultados superiores da métrica *Dice* para a escavação e disco óptico.

Como trabalhos futuros sugere-se a ampliação do conjunto de imagens adicionando novas bases de imagens como ORIGA e REFUGE. Pretende-se, ainda, investigar novas arquiteturas de deep learning como a utilização de uma rede híbrida da U-Net com a EfficientNet que demonstra apresentar resultados promissores em tarefas de segmentação [Baheti et al. 2020] e, em conjunto, um modelo que utiliza também a U-Net

com arquitetura constituída de bottlenecks hierárquicas que se prova útil para tarefas de multi-segmentação [Tang et al. 2021]. Além disto, a utilização de novas tecnologias e bibliotecas que visem aperfeiçoar parâmetros como o HyperOpt são cogitadas em futuros experimentos.

5. Agradecimentos

Os autores desejam expressar seus agradecimentos ao apoio financeiro concedido pela CAPES (Finance code 001), CNPQ e FAPEMA para o desenvolvimento desta pesquisa e ao Núcleo de Computação Aplicada (NCA-UFMA) pelo suporte técnico e científico.

Referências

- Afolabi, O., Nelwamondo, F., and Mabuza, G. (2019). Segmentation of optic cup and disc for diagnosis of glaucoma on retinal fundus images.
- Al-Bander, B., Williams, B. M., Al-Nuaimy, W., Al-Tae, M. A., Pratt, H., and Zheng, Y. (2018). Dense fully convolutional segmentation of the optic disc and cup in colour fundus for glaucoma diagnosis. *Symmetry*, 10(4).
- Baheti, B., Innani, S., Gajre, S. S., and Talbar, S. N. (2020). Eff-unet: A novel architecture for semantic segmentation in unstructured environment. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1473–1481.
- Bourne, R., Steinmetz, J. D., Flaxman, S., Briant, P. S., Taylor, H. R., Resnikoff, S., Casson, R. J., Abdoli, A., Abu-Gharbieh, E., Afshin, A., et al. (2021). Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the global burden of disease study. *The Lancet global health*, 9(2):e130–e143.
- Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., and Kalinin, A. A. (2020). Alumentations: Fast and flexible image augmentations. *Information*, 11(2).
- Collaborators, G. and Study, V. (2021). Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to vision 2020: the right to sight: an analysis for the global burden of disease study. 9.
- Fumero, F., Sigut, J., Alayón, S., and González-Hernández, González de la Rosa, M. (2015). Interactive tool and database for optic disc and cupsegmentation of stereo and monocular retinal fundus images.
- Li, M., Soltanolkotabi, M., and Oymak, S. (2020). Gradient descent with early stopping is provably robust to label noise for overparameterized neural networks. In Chiappa, S. and Calandra, R., editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 4313–4324. PMLR.
- Lima, A., Maia, L. B., dos Santos, P. T. C., Junior, B. J., de Almeida, J. D., and de Paiva, A. C. (2018). Evolving convolutional neural networks for glaucoma diagnosis. In *Anais do XVIII Simpósio Brasileiro de Computação Aplicada à Saúde*. SBC.
- Lima, A. A., de Carvalho Araújo, A. C., de Moura Lima, A. C., de Sousa, J. A., de Almeida, J. D. S., de Paiva, A. C., and Júnior, B. J. (2020). Mask overlaying: a deep

- learning approach for individual optic cup segmentation from fundus image. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 99–104. IEEE.
- Moreira, J. M. M., de Almeida, J. D. S., Junior, B. J., and de Paiva, A. C. (2021). Detecção de glaucoma usando redes em cápsula. In *Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 188–199. SBC.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Rodriguez, J. D., Perez, A., and Lozano, J. A. (2009). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE transactions on pattern analysis and machine intelligence*, 32(3):569–575.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.
- Sevastopolsky, A. (2017). Optic disc and cup segmentation methods for glaucoma detection with modification of u-net convolutional neural network. *CoRR*, abs/1704.00979.
- Silva, M. G., Pessoa, A. C., de Almeida, J. D., Junior, B. J., and de Paiva, A. C. (2018). Diagnóstico do glaucoma em imagens de retinografia usando variantes de padroes locais binários. In *Anais do XVIII Simpósio Brasileiro de Computação Aplicada à Saúde*. SBC.
- Sivaswamy, J., Krishnadas, S., Chakravarty, A., Gopal, D., Joshi, G., Ujjwal, and Syed, T. (2015). Jsm biomedical imaging data papers keywords @bullet @bullet @bullet nerve head @bullet @bullet disk a comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. *JSM Biomedical imaging data papers*.
- Tang, S., Qi, Z., Granley, J., and Beyeler, M. (2021). U-net with hierarchical bottleneck attention for landmark detection in fundus images of the degenerated retina. *Lecture Notes in Computer Science*, page 62–71.
- Wu, Y., Szymanska, M., Hu, Y., Fazal, M. I., Jiang, N., Yetisen, A. K., and Cordeiro, M. F. (2022). Measures of disease activity in glaucoma. *Biosensors and Bioelectronics*, 196:113700.
- Yu, S., Xiao, D., Frost, S., and Kanagasingam, Y. (2019). Robust optic disc and cup segmentation with deep learning for glaucoma detection. *Computerized Medical Imaging and Graphics*, 74:61–71.
- Zilly, J., Buhmann, J. M., and Mahapatra, D. (2017). Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. *Computerized Medical Imaging and Graphics*, 55:28–41. Special Issue on Ophthalmic Medical Image Analysis.