

Um novo método baseado em detector de dois estágios para segmentação de instância de lesões retinianas usando o modelo Mask R-CNN e a biblioteca Detectron2

Marcelo Dias¹, Carlos Santos^{2,3}, Marilton Aguiar³, Daniel Welfer⁴,
Alejandro Pereira¹, Marcelo Ribeiro¹

¹Universidade Federal de Pelotas (UFPEL) – Pelotas, RS – Brasil

²Instituto Federal de Educação, Ciência e Tecnologia Farroupilha (IFFAR) –
Alegrete, RS – Brasil

³Programa de Pós-Graduação em Computação, Universidade Federal de Pelotas
(UFPEL) – Pelotas, RS – Brasil

⁴Programa de Pós-Graduação em Ciência da Computação, Universidade Federal
de Santa Maria (UFSM) – Santa Maria, RS – Brasil

Email: {marcelo.sdias,marilton,adspereira,mcribeiro}@inf.ufpel.edu.br,
carlos.santos@iffarroupilha.edu.br, daniel.welfer@ufsm.br

Abstract. *The development of computerized methods to perform the instance segmentation of retinal lesions such as microaneurysms, hemorrhages, soft exudates, and hard exudates plays a crucial role in the early diagnosis of Diabetic Retinopathy. This paper proposes a method based on a Mask R-CNN neural network architecture to perform instance segmentation of lesions associated with Diabetic Retinopathy. In addition, this network was trained, adjusted, and tested using different publicly available diabetic retinopathy datasets. Moreover, we implemented this network using the Detectron2 and OpenCV libraries. The proposed method reached in the DDR (Dataset for Diabetic Retinopathy), using the Stochastic Gradient Descent optimizer (SGD), the mean Average Precision (mAP) of 0.2664 for the limit of Intersection Over Union of 0.5 in the validation step.*

Resumo. *O desenvolvimento de métodos computacionais para realizar a segmentação de instância de lesões retinianas como microaneurismas, hemorragias, exsudatos algodinosos e exsudatos duros desempenha um papel fundamental no diagnóstico precoce da Retinopatia Diabética. Este artigo propõe um método baseado no modelo Mask R-CNN para realizar a segmentação de instância de lesões associadas à Retinopatia Diabética. A arquitetura de rede neural profunda foi treinada, ajustada e testada usando diferentes conjuntos de dados de retinopatia diabética disponíveis publicamente. Além disso, a rede neural foi implementada usando as bibliotecas Detectron2 e OpenCV. O método proposto alcançou no DDR (Dataset for Diabetic Retinopathy) e com otimizador SGD (Stochastic Gradient Descent) o mAP (mean Average Precision) de 0,2664 para o limite de Interseção sobre União de 0,5 na etapa de validação.*

1. Introdução

A Retinopatia Diabética (RD) é geralmente identificada por meio de exames oftalmológicos que visam identificar lesões retinianas, incluindo microaneurismas (MA), hemorragias (HE), exsudatos algodinosos (SE) e exsudatos duros (EX). No entanto, o rastreamento precoce da RD pelos métodos tradicionais é um desafio devido à escassez de profissionais e recursos para atender as crescentes demandas, principalmente em regiões mais pobres [Vocaturu and Zumpano 2020].

Um método computacional é capaz de prever com precisão a presença destas lesões e auxiliar no diagnóstico e rastreamento da RD. Na literatura, soluções baseadas em detectores de dois estágios têm sido apresentadas para auxiliar na identificação da doença, mas ainda há limitações nos resultados apresentados por esses trabalhos, principalmente na precisão associada à identificação de objetos muito pequenos nas imagens de fundo [Santos et al. 2023].

Nesse contexto, este artigo descreve um novo método para segmentação de instâncias de lesões retinianas associadas à RD. A detecção com segmentação de instância de lesões de fundo ainda é um problema pouco explorado e a principal contribuição deste trabalho é apresentar um novo método baseado em aprendizado profundo e métodos de processamento digital de imagens com o propósito de realizar de forma mais precisa a detecção de pequenas lesões presentes no tecido retiniano.

2. Trabalhos Relacionados

A revisão de literatura realizada identificou trabalhos relacionados que utilizam aprendizado profundo para detecção de lesões retinianas associadas à RD, e são brevemente relatados nesta seção. Suas principais características estão sumarizadas na Tabela 1.

No trabalho de [Porwal et al. 2020] são apresentados os resultados de modelos de aprendizado profundo usados para a classificação, detecção e segmentação de objetos em imagens de fundo. A principal contribuição deste trabalho foi a disponibilização do conjunto de imagens públicas IDRiD (*Indian Diabetic Retinopathy Image Dataset*). No desafio de segmentação, as equipes participantes exploraram arquiteturas como U-Net [Ronneberger et al. 2015] e Mask R-CNN [He et al. 2020]. Convém destacar que no desafio de detecção apenas o Disco Óptico e a Fóvea foram detectados.

Já em [Li et al. 2019] apresentou-se um novo conjunto de dados de RD chamado DDR (*Dataset for Diabetic Retinopathy*) e avaliaram-se modelos de aprendizado profundo para classificação, detecção e segmentação de lesões retinianas. Os resultados apresentados pelos autores na segmentação semântica de microlesões, como no caso dos microaneurismas, demonstraram a dificuldade dos modelos utilizados em identificar objetos pequenos nas imagens de fundo, uma vez que estas lesões possuem uma pequena cobertura de *pixels*.

O trabalho proposto por [Dai et al. 2021] apresentou um sistema para classificar RD e detectar lesões de fundo. O estudo teve algumas limitações, como: 1) uso exclusivo de um conjunto de dados de RD privado para realizar o treinamento dos modelos de aprendizado profunda, usando o conjunto de dados de RD público Kaggle eyePACS somente para avaliar a sub-rede responsável pela classificação de RD; e, 2) a sub-rede que detecta as lesões foi testada apenas no conjunto de dados privado usado pelos autores devido à

ausência de lesões rotuladas no conjunto de dados público usado nos experimentos.

[Shenavarmasouleh et al. 2021] propõem uma arquitetura para detecção de lesões de fundo composta por dois módulos. Um módulo para a detecção das lesões, e outro módulo para classificar a gravidade da RD. Os autores avaliaram a solução proposta com base nas métricas de *Intersection over Union (IoU)* e *mean Average Precision (mAP)*. Para realizar a detecção das lesões foi utilizado o modelo Mask R-CNN e o conjunto de dados público de RD e-Ophtha. Na detecção das lesões a solução proposta foi avaliada somente no conjunto de dados e-Ophtha, obtendo para o limiar de *IoU* de 0,5 *mAP* de 0,4780 na etapa de validação e *mAP* de 0,4370 na etapa de teste. O trabalho proposto limitou-se a realizar somente a detecção de exsudatos e microaneurismas. Além disso, não foram apresentados os resultados de *AP* por tipo de lesão.

Tabela 1. Trabalhos relacionados.

Artigo	Conjunto de Dados	Modelo	Métricas de Desempenho	Limitações
[Porwal et al. 2020]	IDRiD	Mask R-CNN	IoU = 0,9338	Somente o Disco Óptico e a Fóvea foram detectados
[Li et al. 2019]	DDR	DeepLab-v3 HED	mAP = 0,1702 mAP = 0,3317	Baixo desempenho na detecção de MA and SE
[Dai et al. 2021]	Privado	Mask R-CNN	AUC = 0,954 AUC = 0,901 AUC = 0,941 AUC = 0,967	Conjunto de Dados utilizado para treinamento e validação é privado
[Shenavarmasouleh et al. 2021]	e-Ophtha	Mask R-CNN	mAP = 0,4780	Somente exsudatos e microaneurismas foram detectados

Intersection over Union (IoU) *mean Average Precision (mAP)* *Area Under the Curve (AUC)*
Exsudatos Duros (EX) Microaneurismas (MA) Exsudatos Algodonosos (SE) Hemorragias (HE)

Com base nos trabalhos apresentados nesta Seção, considera-se a utilização de aprendizado profundo e técnicas de processamento digital de imagens para a implementação de um método que aprimore a detecção das lesões de fundo, principalmente das microlesões, como no caso dos microaneurismas. Com o método proposto pretende-se, portanto, realizar a segmentação de instâncias das lesões retinianas associados à Retinopatia Diabética, incluindo Exsudatos Duros, Microaneurismas, Exsudatos Algodonosos e Hemorragias.

3. Materiais e Métodos

O método proposto foi desenvolvido com base na arquitetura Mask R-CNN [He et al. 2020], conforme ilustrado no diagrama de blocos mostrado na Figura 1. Para a construção da arquitetura foi utilizada a biblioteca de código aberto Detectron2 [Wu et al. 2019, Hong et al. 2021, Amerikanos and Maglogiannis 2022].

Mask R-CNN é um modelo capaz de detectar e segmentar instâncias de objetos. Este modelo estende a arquitetura de detecção de objetos Faster R-CNN [He et al. 2020], adicionando uma estrutura paralela para prever máscaras de segmentação de objetos. A segmentação de instância combina tarefas de detecção de objetos, cujo objetivo é classificar e localizar objetos individualmente usando uma caixa delimitadora e também localizar cada *pixel* de cada objeto detectado na imagem.

Essa arquitetura funciona em duas etapas. A primeira etapa consiste em usar uma *Region Proposal Network (RPN)* [Ren et al. 2017, Zhao et al. 2019] para selecionar as

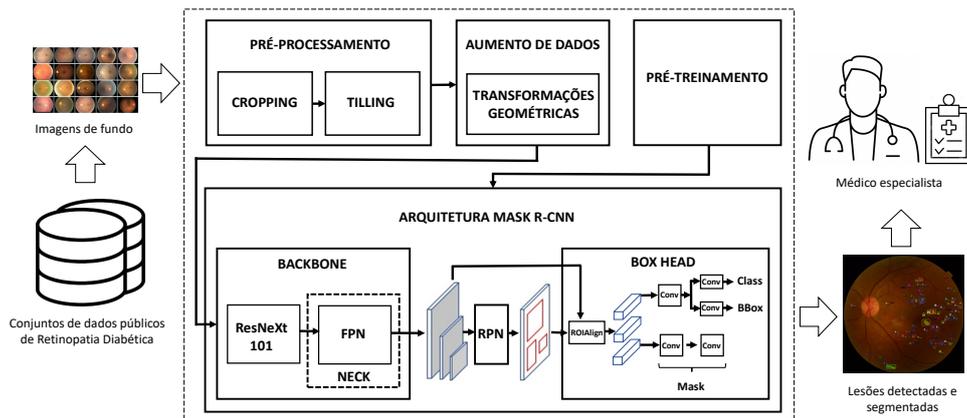


Figura 1. Diagrama de blocos do método proposto. Primeiramente, as imagens são enviadas para o bloco de Pré-processamento para aplicação de *Cropping* e *Tilling*. Em seguida, as imagens são transferidas para o bloco de Aumento de Dados para aplicação de Transformações Geométricas nas imagens pré-processadas. Após a realização do aumento de dados, as imagens são enviadas para treinamento da rede neural profunda. Porém, antes disso uma etapa de pré-treinamento é realizada para ajustar os pesos da rede neural. Por fim, são apresentadas as lesões de fundo detectadas e segmentadas em nível de *pixel*.

caixas delimitadoras (BBox) dos objetos candidatos. A segunda etapa visa classificar as caixas candidatas, refinar as caixas e prever as máscaras dos objetos (Mask).

Com o método proposto para segmentação de instância foi possível realizar uma detecção mais precisa das lesões de fundo, além de proporcionar ao médico especialista uma melhor visualização da extensão exata de cada lesão detectada, a fim de auxiliar no diagnóstico da doença. A seguir, são detalhadas as etapas que compõem a metodologia adotada.

3.1. Conjunto de Dados e Pré-Processamento das Imagens

Dois conjuntos de dados públicos com imagens do fundo foram usados para realizar os experimentos, o conjunto de dados DDR [Li et al. 2019] – que possui 757 imagens rotuladas – e o conjunto de dados IDRiD [Porwal et al. 2020] – que possui 81 imagens rotuladas. Ambos *datasets* têm *Ground Truth* para as lesões MA, HE, SE e EX em nível de *pixel*.

Assim como os trabalhos propostos por [El abadi and Hammod 2014] e [Alyoubi et al. 2021], realizou-se uma etapa de pré-processamento para *cropping* parcial do plano de fundo preto das imagens da retina, a fim de minimizar a geração de falsos positivos durante a detecção das lesões. No caso das imagens de fundo apenas os *pixels* da retina possuem informações significativas, o restante é considerado plano de fundo. Para encontrar a circunferência da retina utilizou-se o método de *Hough Gradient 2-1 Hough Transform* (21HT) [Yuen et al. 1990]. Após a localização da retina, transformou-se sua circunferência em seu retângulo equivalente com as coordenadas x_{min} e y_{min} – referentes à posição superior esquerda na imagem – e as coordenadas x_{max} e y_{max} – referentes à posição inferior direita, ilustrado na Figura 2. A área externa ao retângulo em vermelho corresponde à área de corte realizada na imagem de fundo.

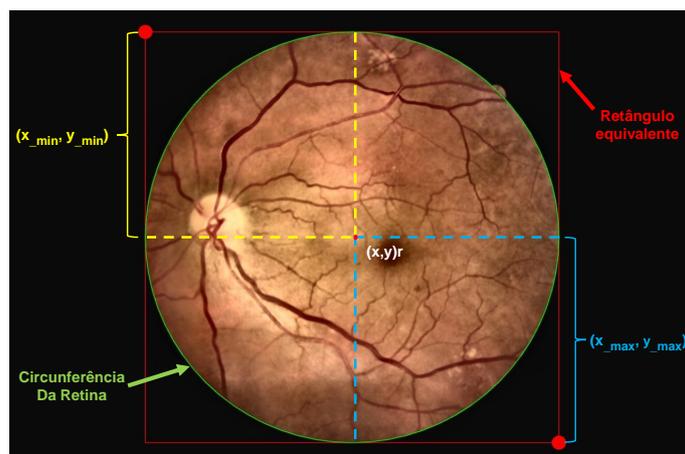


Figura 2. Transformação da circunferência da retina em seu retângulo equivalente por meio das coordenadas x_{min} e y_{min} (posição superior esquerda), e as coordenadas x_{max} e y_{max} (posição inferior direita).

A última etapa da etapa de pré-processamento é a realização de *Tiling*, no qual as imagens originais são cortadas em blocos (*tiles*). Após a redução da resolução das imagens ao serem utilizadas na camada de entrada da rede neural, algumas microlesões praticamente desapareceram, em função do tamanho diminuto destas lesões. Portanto, o aumento da resolução das imagens utilizadas na camada de entrada da rede neural proporciona um campo receptivo maior em torno das lesões e, com isto, um número maior de características ficam acessíveis para que as camadas convolucionais possam aprendê-las. Assim, a aplicação do método de *Tiling* tem por objetivo proporcionar um aumento na precisão do método proposto associado a uma extração de características mais eficaz das lesões de fundo.

Assim como o trabalho proposto por [Unel et al. 2019], o método de *Tiling* foi implementado com tamanho de blocos das sub-imagens resultantes organizados de acordo o tamanho da imagem original disponível no conjunto de dados. Foi realizada a construção de blocos estáticos de tamanho 2×2 , e com uma área de sobreposição entre estes blocos de 15%, a fim de preservar as lesões ao longo dos limites destes blocos, e evitar a perda de informações proveniente da criação dos *tiles*.

3.2. Aumento de Dados

A quantidade limitada de lesões rotuladas disponíveis em conjuntos de dados públicos restringe a quantidade de características extraídas pela rede neural profunda durante o estágio de treinamento. Devido a estes problemas foi realizada uma etapa de aumento de dados *on-the-fly*, no qual o carregador de dados da Detectron2 aplica os métodos de aumento *Resize Shortest Edge* e *Random Flip* nas imagens pré-processadas na etapa anterior. Após realizar o aumento de dados, o método proposto foi treinado e ajustado para realizar a segmentação de instância das lesões de fundo.

3.3. Arquitetura da Rede Neural Profunda

A arquitetura da rede neural que compõe o método proposto é basicamente constituída por três módulos principais, sendo um *Backbone*, um *Region Proposal Network* (RPN) e

um *Box Head* (cabeça de ROI), conforme ilustrado no diagrama de blocos apresentado na Figura 3.

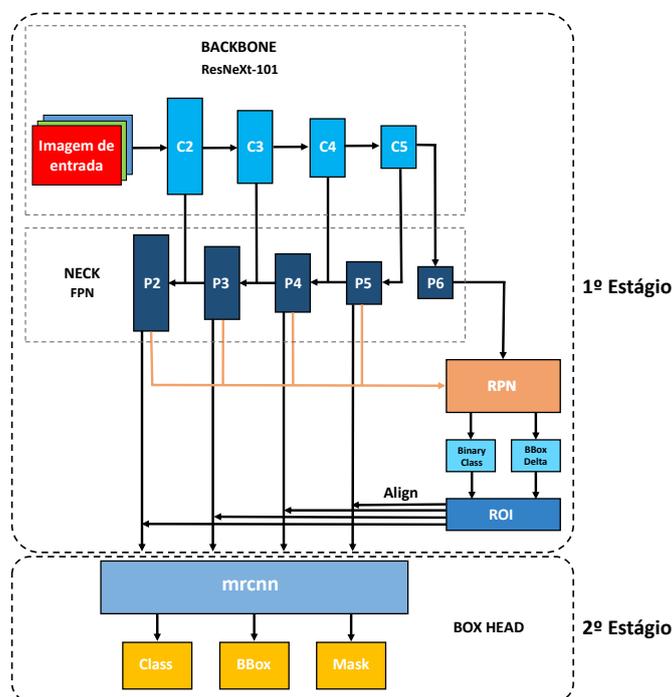


Figura 3. Diagrama de blocos da arquitetura Mask R-CNN que compõe o método proposto. No primeiro estágio, a arquitetura possui um módulo de *Backbone* composto por uma rede neural profunda ResNeXt-101 e um módulo de *Neck* composto por uma FPN (*Feature Pyramid Network*). Os mapas de características das imagens de fundo são extraídas pelo *Backbone* (C2, C3, C4 e C5) e encaminhadas para as camadas da FPN (P2, P3, P4, P5 e P6), que são integradas com um módulo RPN, responsável pela geração dos valores de saída de classificação (*Binary Class*) e regressão de caixas delimitadoras (*BBox Delta*). Para realizar a tarefa de segmentação de forma mais eficaz é utilizada uma camada *ROI Align*, que preserva a localização espacial dos objetos. O segundo estágio da arquitetura é composto por um módulo *Box Head*, que recebe as regiões selecionadas pelo classificador ROI e gera as *bounding boxes* (*BBox*) e as máscaras de segmentação (*Mask*) para estas regiões.

Backbone é uma rede neural convolucional convencional, cujo objetivo é extrair as características das lesões nas imagens de fundo. Em seguida, a arquitetura possui um módulo RPN, cuja tarefa é inspecionar todo o FPN do *Backbone* de cima para baixo, para propor regiões que possam conter lesões nas imagens da retina. Deve-se notar que o RPN detecta regiões com base em características multiescala. O último módulo da arquitetura é o *Box Head*, responsável por recortar e interpolar os mapas de características das regiões propostas geradas no RPN. Para fazer as inferências da lesões de fundo é aplicada a supressão de não-máximos, em que as caixas delimitadoras com pontuação abaixo de um limiar pré-definido sejam descartadas.

3.4. Pré-treinamento

O método proposto se utiliza de aprendizado por transferência para pré-treinar a arquitetura da rede neural. Os pesos foram pré-treinados no conjunto de dados COCO (*Common*

Objects in Context) [Lin et al. 2014] e utilizados para inicializar os pesos da arquitetura de rede neural usada no método proposto. Após este pré-treinamento a saída do modelo proposto foi modificada para realizar a segmentação de instância das lesões retinianas, preservando os pesos das camadas superiores. O treinamento envolve quatro etapas principais, a saber: 1) As camadas iniciais da arquitetura são pré-treinadas com os pesos do conjunto de dados COCO; 2) As últimas camadas são cortadas e substituídas por novas camadas; 3) As novas camadas adicionadas são ajustadas no conjunto de dados de RD; e, 4) Após o ajuste fino das camadas finais da arquitetura, toda a rede neural é retreinada, para que pequenos ajustes sejam realizados nos pesos de toda a arquitetura.

3.5. Treinamento e Ajuste-fino do Modelo

Para realização dos experimentos, os conjuntos de dados foram particionados em conjuntos de Treino, Validação e Teste, em uma proporção de 50:20:30, respectivamente. O ajuste fino do método proposto, realizado no conjunto de Validação, visa otimizar os hiperparâmetros para obter resultados mais precisos na detecção e segmentação das lesões de fundo. O melhor ajuste de hiperparâmetros é apresentado na Tabela 2.

Tabela 2. Hiperparâmetros ajustados na etapa de validação.

Hiperparâmetro	Valor
<i>Anchor Sizes</i>	(8, 16, 32, 64, 128)
<i>Images per Batch</i>	2
<i>Learning Rate</i>	0,001
<i>Max Iterations</i>	100.000
<i>Momentum</i>	0,937
<i>Score Thresh Test</i>	0,25
<i>Number of Workers</i>	4
Otimizador	SGD
<i>ROI Heads Batch Size per Image</i>	1.024
<i>RPN Batch Size per Image</i>	512
<i>Test Detections per Image</i>	256
<i>Weight Decay</i>	0,0005

O método proposto foi treinado com *Number of Workers* igual a 4, *Max Iterations* igual a 100.000 e *Learning Rate* de 0,001. O tamanho das âncoras de célula foi definido como (8, 16, 32, 64, 128). Para reduzir o sub-ajuste do modelo na etapa de treinamento. O método proposto foi regularizado por meio do aumento de dados realizado, em que foi possível criar imagens artificiais e aumentar a quantidade de exemplos das lesões rotuladas.

Para minimizar a probabilidade de sobre-ajuste da rede neural, principalmente devido ao desequilíbrio no número de exemplos das diferentes classes de lesões investigadas foi aplicada a técnica de regularização L2 [Goodfellow et al. 2020], por meio da introdução de *Weight Decay* durante a etapa de treinamento do modelo.

4. Resultados experimentais e Discussões

Para realizar o treinamento e ajuste do método proposto foi utilizado o conjunto de dados DDR. O conjunto de dados IDRiD também foi usado para avaliar a capacidade de

generalização do método proposto. A avaliação foi realizada tanto na detecção (BBox) quanto na segmentação (Mask) das lesões de fundo, sendo adotada a métrica *Intersection over Union (IoU)*.

Conforme descrito nas Tabelas 3 e 4, o método proposto foi comparado com diferentes modelos que utilizam redes neurais profundas de última geração. A Tabela 3 apresenta os resultados obtidos com a métrica *mAP* para o limite de *IoU* igual a 0,5 e com otimizador SGD, usando o conjunto de dados DDR. O método proposto usando o *Backbone* ResNeXt-101-FPN obteve a melhor precisão nas tarefas de detecção e segmentação nos experimentos realizados, conforme indicado em negrito, com um *mAP* de 0,2604.

Tabela 3. Resultados obtidos nas tarefas de detecção e segmentação de lesões de fundo com a métrica *mAP* para o limite de *IoU* de 0,5 no conjunto de validação do conjunto de dados DDR e com otimizador SGD.

	Modelos	Backbone	<i>mAP</i>
BBox	Mask R-CNN	ResNet-50 C4	0,1325
	Mask R-CNN	ResNet-50 C5-dilated	0,1368
	Mask R-CNN	ResNet-50 FPN	0,1737
	Mask R-CNN	ResNet-101 C4	0,1746
	Mask R-CNN	ResNet-101 C5-dilated	0,1678
	Mask R-CNN	ResNet-101 FPN	0,1909
	Mask R-CNN	ResNeXt-101 FPN 32×8d	0,2124
	Método Proposto	ResNeXt-101 FPN 32×8d	0,2664
Mask	Mask R-CNN	ResNet-50 C4	0,1617
	Mask R-CNN	ResNet-50 C5-dilated	0,1328
	Mask R-CNN	ResNet-50 FPN	0,1795
	Mask R-CNN	ResNet-101 C4	0,1624
	Mask R-CNN	ResNet-101 C5-dilated	0,1603
	Mask R-CNN	ResNet-101 FPN	0,1848
	Mask R-CNN	ResNeXt-101 FPN 32×8d	0,2205
	Método Proposto	ResNeXt-101 FPN 32×8d	0,2604

O método proposto também foi avaliado no conjunto de dados IDRiD, conforme apresentado na Tabela 4. A Figura 4 apresenta os gráficos de perda e *Average Precision* obtidos durante o treinamento e validação do método proposto com o otimizador SGD. É possível verificar que o método proposto alcançou maior precisão na detecção de Exsudatos Algodonosos e menor precisão na detecção de Microaneurismas.

Para avaliar a precisão do método proposto foram calculadas as métricas *mAP* e *AP* com 10 limiares de *IoU* de 0,5 na etapa de validação do *dataset* DDR para cada tipo de lesão, conforme mostrado na Tabela 5. Os valores de *AP@0,5* na detecção das lesões retinianas foram de 0,2779, 0,3309, 0,1607 e 0,2961, para EX, SE, MA e HE, respectivamente, conforme apresentado na Tabela 5. Para entender melhor os resultados obtidos a Figura 5 apresenta um exemplo de segmentação de instância realizada pelo método proposto em uma imagem de fundo do *dataset* DDR.

Percebe-se que o método proposto obteve resultados promissores na segmentação de instância das lesões de fundo investigadas, mesmo com a presença de múltiplas lesões com tamanhos e formas variáveis. Pelo fato do modelo Mask R-CNN possuir um módulo

Tabela 4. Resultados obtidos nas tarefas de detecção e segmentação de lesões de fundo com a métrica mAP para o limite de IoU de 0,5 no conjunto de validação do *dataset* IDRiD e com otimizador SGD.

Modelos		Backbone	mAP
BBox	Mask R-CNN	ResNet-50 C4	0,2254
	Mask R-CNN	ResNet-50 C5-dilated	0,1789
	Mask R-CNN	ResNet-50 FPN	0,2050
	Mask R-CNN	ResNet-101 C4	0,2120
	Mask R-CNN	ResNet-101 C5-dilated	0,2065
	Mask R-CNN	ResNet-101 FPN	0,1656
	Mask R-CNN	ResNeXt-101 FPN 32×8d	0,2188
	Método Proposto	ResNeXt-101 FPN 32×8d	0,3460
Mask	Mask R-CNN	ResNet-50 C4	0,1921
	Mask R-CNN	ResNet-50 C5-dilated	0,1645
	Mask R-CNN	ResNet-50 FPN	0,2138
	Mask R-CNN	ResNet-101 C4	0,2161
	Mask R-CNN	ResNet-101 C5-dilated	0,1841
	Mask R-CNN	ResNet-101 FPN	0,1789
	Mask R-CNN	ResNeXt-101 FPN 32×8d	0,2050
	Método Proposto	ResNeXt-101 FPN 32×8d	0,3210

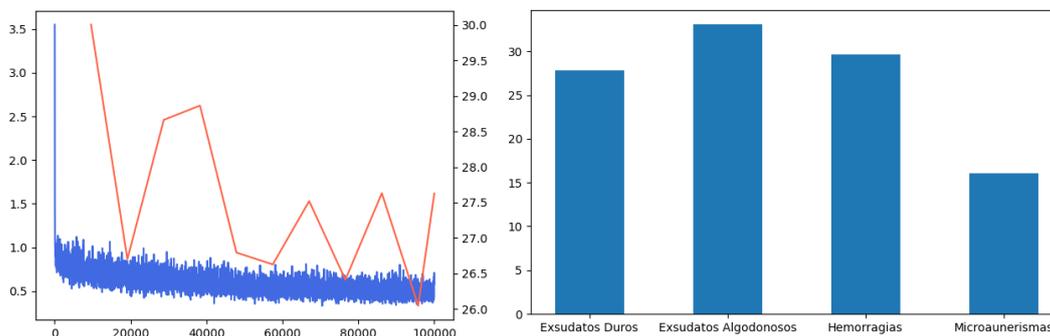


Figura 4. Gráficos de Perda e *Average Precision* obtidos durante o treinamento e validação do método proposto. (a) Perda total (curva em azul) no treinamento e AP de BBox para o limite de IoU de 0,5 no conjunto de validação do *dataset* DDR. (b) AP de BBox para o limite de IoU de 0,5 para as lesões de fundo no conjunto de validação do *dataset* DDR.

Tabela 5. Resultados obtidos com detecção (BBox) e segmentação (Mask) pelo método proposto com AP e mAP com 10 limiares de IoU de 0,5 no conjunto de validação do conjunto de dados DDR.

Modelo		AP				mAP
		EX	SE	MA	HE	
Método Proposto	BBox	0,2779	0,3309	0,1607	0,2961	0,2664
	Mask	0,2698	0,3357	0,1674	0,2687	0,2604

Average Precision (AP) *mean Average Precision* (mAP) Exsudatos Duros (EX)
Exsudatos Algodonosos (SE) Microaneurismas (MA) Hemorragias (HE)

5. Conclusões

Este artigo apresentou um método baseado em um detector de dois estágios para segmentação de instância de lesões associadas à RD. A arquitetura da rede neural foi construída com base no modelo Mask R-CNN e implementado por meio da biblioteca de código aberto Detectron2. Para treinamento e avaliação do método proposto foram utilizados os conjuntos de dados de RD públicos, DDR e IDRiD.

Os *datasets* foram divididos em conjunto de Treino, Validação e Teste em uma proporção de 50:20:30, respectivamente. Com o limite *IoU* de 0,5, a segmentação de instância das lesões obteve *mAP* de 0,2664 na etapa de validação e *mAP* de 0,1600 na etapa de teste. O método proposto alcançou *AP@0.5* de 0,2779 para Exsudatos Duros, 0,3309 para Exsudatos Algodonosos, 0,1607 para Microaneurismas e 0,2961 para Hemorragias. Os resultados apresentados neste trabalho indicam que a segmentação de instância de lesões de fundo é extremamente difícil porém pode auxiliar no diagnóstico médico.

Como trabalhos futuros, pretende-se desenvolver soluções que combinem diferentes contextos, como a classificação, detecção e segmentação de lesões associadas à RD, a fim de alcançar resultados mais precisos e também proporcionar ao médico especialista uma quantidade maior de informações para auxiliá-lo na realização do diagnóstico da RD.

Agradecimentos

Este estudo foi financiado parcialmente pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, e pelo Programa Institucional de Bolsas de Iniciação Científica - PROBIC/FAPERGS.

Referências

- Alyoubi, W. L., Abulkhair, M. F., and Shalash, W. M. (2021). Diabetic retinopathy fundus image classification and lesions localization system using deep learning. *Sensors*, 21(11).
- Amerikanos, P. and Maglogiannis, I. (2022). Image analysis in digital pathology utilizing machine learning and deep neural networks. *Journal of Personalized Medicine*, 12(9).
- Dai, L., Wu, L., Li, H., Cai, C., Wu, Q., Kong, H., Liu, R., Wang, X., Hou, X., Liu, Y., Long, X., Wen, Y., Lu, L., Shen, Y., Chen, Y., Shen, D., Yang, X., Zou, H., Sheng, B., and Jia, W. (2021). A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nature Communications*, 12(1).
- El abbadi, N. and Hammoud, E. (2014). Automatic early diagnosis of diabetic retinopathy using retina fundus images enas hamood al-saadi-automatic early diagnosis of diabetic retinopathy using retina fundus images. *EUROPEAN ACADEMIC RESEARCH*, 2.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11):139–144.
- He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2020). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):386–397.
- Hong, A., Lee, G., Lee, H., Seo, J., and Yeo, D. (2021). Deep learning model generalization with ensemble in endoscopic images. *CEUR Workshop Proceedings*, 2886:80–89.

- Li, T., Gao, Y., Wang, K., Guo, S., Liu, H., and Kang, H. (2019). Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. *Information Sciences*, 501:511–522.
- Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8693 LNCS(PART 5):740–755.
- Porwal, P., Pachade, S., Kokare, M., Deshmukh, G., Son, J., Bae, W., Liu, L., Wang, J., Liu, X., Gao, L., Wu, T. B., Xiao, J., Wang, F., Yin, B., Wang, Y., Danala, G., He, L., Choi, Y. H., Lee, Y. C., Jung, S. H., Li, Z., Sui, X., Wu, J., Li, X., Zhou, T., Toth, J., Baran, A., Kori, A., Chennamsetty, S. S., Safwan, M., Alex, V., Lyu, X., Cheng, L., Chu, Q., Li, P., Ji, X., Zhang, S., Shen, Y., Dai, L., Saha, O., Sathish, R., Melo, T., Araújo, T., Harangi, B., Sheng, B., Fang, R., Sheet, D., Hajdu, A., Zheng, Y., Mendonça, A. M., Zhang, S., Campilho, A., Zheng, B., Shen, D., Giancardo, L., Quellec, G., and Mériaudeau, F. (2020). IDRiD: Diabetic Retinopathy – Segmentation and Grading Challenge. *Medical Image Analysis*, 59.
- Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351:234–241.
- Santos, C., Aguiar, M., Welfer, D., Dias, M., Pereira, A., Ribeiro, M., and Belloni, B. (2023). A new approach for fundus lesions instance segmentation based on mask r-cnn x101-fpn pre-trained architecture. *IEEE Access*, 11:43603–43618.
- Shenavarmasouleh, F., Mohammadi, F. G., Amini, M. H., Taha, T., Rasheed, K., and Arabnia, H. R. (2021). Drdrv3: Complete lesion detection in fundus images using mask r-cnn, transfer learning, and lstm.
- Unel, F. O., Ozkalayci, B. O., and Cigla, C. (2019). The power of tiling for small object detection. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 582–591.
- Vocaturro, E. and Zumpano, E. (2020). The contribution of AI in the detection of the Diabetic Retinopathy. *Proceedings - 2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2020*, pages 1516–1519.
- Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., and Girshick, R. (2019). Detectron2. <https://github.com/facebookresearch/detectron2>.
- Yuen, H., Princen, J., Illingworth, J., and Kittler, J. (1990). Comparative study of hough transform methods for circle finding. *Image and Vision Computing*, 8(1):71–77.
- Zhao, Z.-Q., Zheng, P., tao Xu, S., and Wu, X. (2019). Object detection with deep learning: A review.