

Improved Detection of Fundus Lesions Using YOLOR-CSP Architecture and Slicing Aided Hyper Inference

Alejandro Pereira, Carlos Santos, Marilton Aguiar, Daniel Welfer, Marcelo Dias, and Marcelo Ribeiro

Abstract—Diabetes affects millions of people worldwide, and diabetic retinopathy is a complication of diabetes. Brazil is the sixth in the world in the incidence of diabetes and has the highest numbers in Latin America with 15.7 million adults affected by this condition. Typically diabetic retinopathy is identified by lesions such as hard and soft exudates, microaneurysms, and vitreous hemorrhages. Early diagnosis of these lesions is essential to prevent the progression of the disease to severe stages and the consequent loss of vision. As the disease diagnosis is based on image analysis, it is possible to use deep learning models to detect artifacts in the retina. This article proposes a new method that uses a YOLOR-CSP architecture combined with the Slicing Aided Hyper Inference framework to detect fundus lesions. The proposed method was trained, adjusted, and evaluated using the DDR dataset, obtaining a mAP equal to 38.08%. The proposed method achieved values of AP equal to 40.90%, 46.60%, 26.10%, and 38.70% for hard exudates, soft exudates, microaneurysms, and vitreous hemorrhages, respectively, surpassing similar studies found in the literature.

Index Terms—diabetes mellitus, diabetic retinopathy, fundus image, lesions detection, deep learning.

I. INTRODUÇÃO

A Retinopatia Diabética (RD) está entre as mais comuns complicações microvasculares do diabetes e uma das maiores causas de perda irreversível de visão [1]. Segundo dados estatísticos do censo demográfico de 2000, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), a primeira causa de deficiência entre 24,5 milhões de brasileiros é a visual, representando 48,1% do total, sendo portanto um problema de saúde pública nacional [2]. É possível classificar a RD de acordo com a presença de determinados tipos de lesões na retina como os Exsudatos Duros e os Exsudatos Algodonosos, Microaneurismas e Hemorragias no fundo do olho [3].

Diante deste cenário, este artigo tem como principal contribuição apresentar um método capaz de detectar com maior precisão essas lesões de fundo associadas à RD por meio de uma arquitetura de rede neural profunda de última geração chamada YOLOR-CSP, cuja premissa é combinar conhecimento

explícito e implícito para obter uma rede unificada capaz de solucionar múltiplas tarefas em uma única vez [4]. Além disso, o método proposto utiliza o *framework* SAHI (*Slicing Aided Hyper Inference*) na etapa de pré-processamento, a fim de aumento de o problema associado à detecção de microlesões da retina.

II. FUNDAMENTAÇÃO E TRABALHOS RELACIONADOS

A arquitetura da YOLOR-CSP possui como base uma YOLOv4-CSP [5]. Os diferentes blocos contidos na arquitetura são responsáveis pela forma com que a YOLOR-CSP obtém o conhecimento explícito. Assim, a principal contribuição do modelo YOLOR (*You Only Learn One Representation*) é a introdução de módulos para extrair o conhecimento implícito e integrá-lo com o conhecimento explícito para obtenção de uma representação unificada. A Fig. 1 ilustra o modelo arquitetural da rede neural que compõe a abordagem proposta. A estrutura é composta por três blocos principais: *Backbone*, *Neck* e *Head*.

O *Backbone* é responsável pela extração de características das imagens. A entrada do *Backbone* foi configurada para receber imagens com tamanho de $640 \times 640 \times 3$ no espaço de cores RGB (*Red*, *Green*, *Blue*). Além disso, o *Backbone* é composto por uma sequência de seis camadas convolucionais com *Downsampling*. Este processo permite que as características presentes na imagem sejam compactadas e apenas as características principais sejam extraídas, proporcionando um nível maior de abstração das imagens.

O bloco *Neck* permite que as características extraídas no *Backbone* sejam combinadas e encaminhadas para a realização das detecções. Este bloco é formado por módulos CSP responsáveis por conectar as camadas frontais e posteriores da rede, a fim de melhorar a integração das diferentes partes da arquitetura e reduzir o tamanho do modelo.

Por fim, o bloco *Head* é responsável por fazer a previsão final do modelo. Na arquitetura YOLOR-CSP este bloco é formado por três cabeças de detecção: a camada YOLO P3, responsável pela detecção de pequenos objetos; a camada YOLO P4, responsável por detectar objetos médios; e, a camada YOLO P5, encarregada pela detecção de objetos grandes.

O *framework* SAHI, proposto por Fatih Akyon et al. [6], é uma solução de código aberto que fornece um processo genérico de inferência e ajuste fino auxiliado por fracionamento. O processo adotado pela ferramenta consiste, basicamente, em dividir a imagem de entrada em diversas janelas que se

Alejandro Pereira, Marcelo Dias and Marcelo Ribeiro Ciência da Computação, Universidade Federal de Pelotas, Pelotas, Brazil. (e-mail: adspereira@inf.ufpel.edu.br).

Carlos Santo Centro de Informática, Instituto Federal Farroupilha, Alegrete, Brazil. (e-mail: carlos.santos@iffarroupilha.edu.br).

Carlos Santo and Marilton Aguiar Programa de Pós-Graduação em Computação, Universidade Federal de Pelotas, Pelotas, Brazil. (e-mail: marilton@inf.ufpel.edu.br).

Daniel Welfer Departamento de Computação Aplicada, Universidade Federal de Santa Maria, Santa Maria, Brazil. (e-mail: daniel.welfer@ufsm.br).

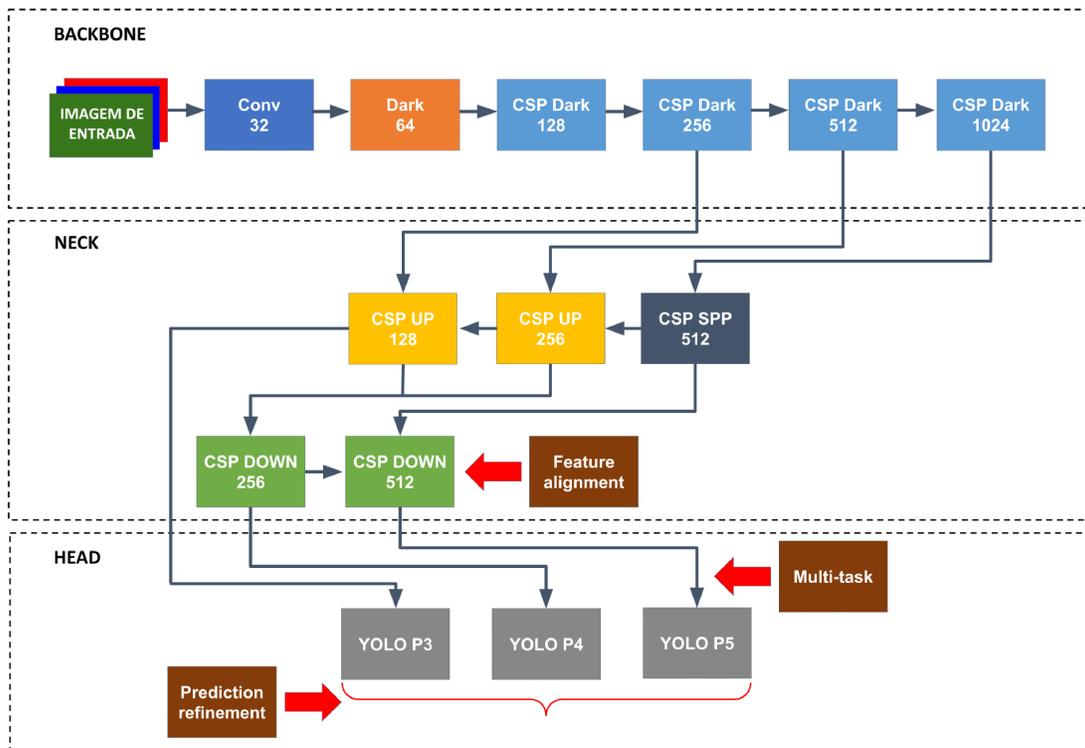


Fig. 1. Diagrama de blocos da arquitetura da rede neural que compõe a abordagem proposta para a detecção de lesões de fundo. A estrutura é dividida em três blocos principais: *Backbone*, *Neck* e *Head*.

sobrepõem entre si. Assim, o método produz uma visibilidade maior para pequenos objetos.

Em Li et al. [7] foram avaliados os modelos de aprendizado profundo Faster R-CNN (*Faster Regions with Convolutional Neural Network features*) [8], SSD (*Detector MultiBox Single Shot*) [9] e YOLO (*You Only Look Once*) [10] para detecção de lesões associadas à RD. Os resultados demonstraram que os modelos baseados em aprendizado profundo possuem dificuldade para detectar lesões retinianas, principalmente no caso dos Microaneurismas, onde foi obtido um *AP* (*Average Precision*) inferior a 1%.

Santos et al. [11] também propuseram uma abordagem para detecção de lesões de fundo utilizando a arquitetura YOLOv4 [5]. Os resultados apresentados superaram os dos trabalhos relacionados encontrados na literatura. O modelo proposto ainda apresentou baixa precisão na detecção de Microaneurismas, com um *AP* de 1,93%.

Santos et al. [12] apresentaram outra abordagem para detecção de lesões de fundo baseada no modelo de rede neural profunda YOLOv5 [13]. Este modelo também é um detector de estágio único assim como a YOLOv4. Os autores também aplicaram o método de *Tiling*, em que as imagens são cortadas em blocos a fim de reduzir a perda de informações das lesões proveniente da diminuição da resolução das imagens quando utilizadas para treinamento da rede neural profunda. Os melhores resultados apresentados neste trabalho foram *mAP* (*mean Average Precision*) de 26,3% no conjunto de validação e 15% no conjunto de teste do *dataset* DDR (*Dataset for Diabetic Retinopathy*).

Este trabalho apresenta um método inovador capaz de de-

tecar lesões de fundo por meio de técnicas de processamento de imagens, aumento de dados e uma rede neural profunda implementada a partir da arquitetura YOLOR-CSP. O método proposto tem por finalidade aumentar a precisão na identificação das lesões retinianas associadas à RD, proporcionando um meio mais eficaz para diagnóstico precoce da doença.

III. MATERIAIS E MÉTODOS

Para a realização dos experimentos foi utilizado no ambiente de *hardware* um equipamento com processador AMD Ryzen 5 2600X Six-Core @ 12x 3.6GHz, com 16GB de memória RAM e uma GPU NVIDIA TITAN Xp com 12 GB de VRAM. A abordagem proposta utilizou como base a arquitetura YOLOR-CSP para realizar a detecção das lesões de fundo. Além disso, para melhorar a precisão da detecção das lesões foram utilizadas etapas de pré-processamento, aumento de dados e de pré-treinamento.

Conforme o diagrama de blocos apresentado na Fig. 2, as imagens do conjunto de treino primeiramente são passadas para o bloco de pré-processamento. Neste bloco, realiza-se a eliminação parcial do plano de fundo preto das imagens. Em seguida, o *framework* SAHI é aplicado por meio do método *COCO slice*. Na sequência as imagens pré-processadas são transferidas para o bloco de aumento de dados. Nesta etapa, novas imagens são criadas artificialmente para serem utilizadas na camada de entrada da rede neural. Por fim, as imagens são utilizadas para o treinamento da rede neural. Porém, antes disso, uma etapa de pré-treinamento é realizada com pesos disponibilizados no repositório oficial da YOLOR [14].

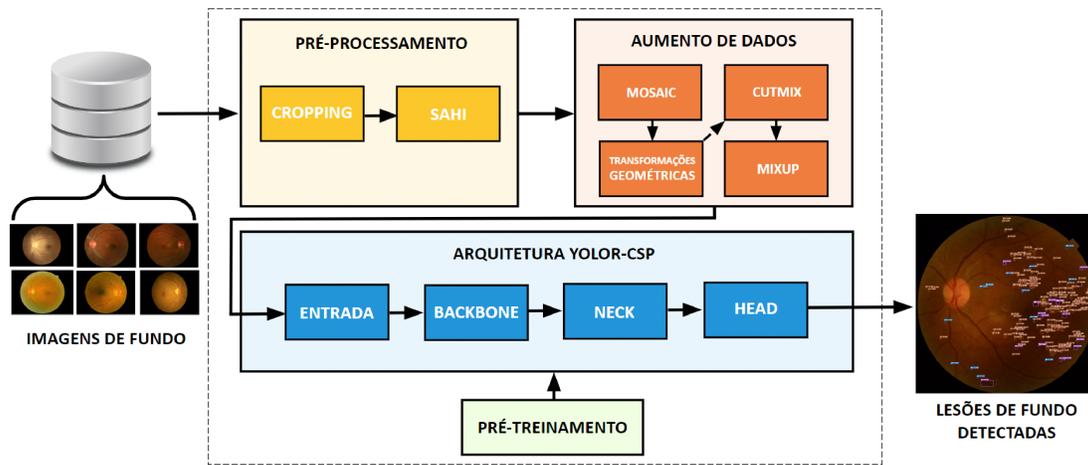


Fig. 2. Diagrama de blocos da abordagem proposta para detecção de lesões de fundo.

A. Conjunto de Dados

Nos experimentos foram utilizados os conjuntos de dados públicos DDR e o *Indian Diabetic Retinopathy Image Dataset* (IDRiD).

O conjunto de dados DDR é composto por 13.673 imagens coloridas do fundo do olho das quais 757 possuem anotações das caixas delimitadoras em termos de *pixel*. O conjunto de dados IDRiD [15] contém 516 imagens onde foram realizadas anotações em termos de *pixel* das lesões típicas relacionadas a RD e também do disco óptico, sendo disponibilizadas 81 imagens com as anotações das lesões retinianas: Microaneurismas (MA), Hemorragias (HE), Exsudatos Duros (EX) e Exsudatos Algodonosos (SE).

B. Pré-Processamento

No método proposto, a primeira etapa de pré-processamento é o *Cropping* parcial do plano de fundo preto das imagens da retina. A próxima etapa de pré-processamento foi utilizar o *framework* SAHI como alternativa ao método de *Tiling* para aumento de o problema associado à detecção das microlesões da retina, como no caso dos microaneurismas.

C. Aumento de Dados

Embora os conjuntos de dados DDR e IDRiD possuam imagens da retina com alta resolução e anotações em termos de *pixel*, possuem uma pequena quantidade de lesões rotuladas para fins de treinamento de redes neurais profundas. Para mitigar este problema foi realizada uma etapa de aumento de dados *on-the-fly*.

O primeiro método de aumento utilizado foi o *CutMix*. Este método é a combinação das técnicas de *MixUp* e *DropOut*, e consiste em remover um bloco de uma imagem origem, resultando em uma área sem informação de intensidade luminosa e de cor sobre a imagem. Após, esta área recebe um bloco de mesmas dimensões extraído de outra imagem. Outro método utilizado foi o *Mosaic*, que combina quatro sub-imagens aleatórias de proporções diferentes para criar uma nova imagem. Dentre as vantagens de utilizar este método está a possibilidade de reduzir o tamanho do lote durante o

treinamento da rede neural pois cada imagem do lote possui a combinação de quatro diferentes imagens. Também foi utilizada a técnica de *MixUp*, responsável por criar uma nova imagem a partir da combinação de outras duas. Basicamente, calcula-se a média a partir da combinação ponderada de duas imagens. O método também realiza a combinação das anotações para obter as anotações da imagem resultante. Por fim, foi realizado o aumento de dados por meio de transformações geométricas. Foram aplicadas três transformações: *Horizontal flip left-right* com probabilidade de 0,5, *Translation* numa fração de até 0,5 *pixels* da imagem em qualquer direção e *Scale* com variação de 0,8 até 1,2.

D. Treinamento e Ajuste da Arquitetura

Como mencionado, para realizar o treinamento do método proposto foram utilizadas as imagens e anotações dos conjuntos de dados públicos DDR [7] e IDRiD [15]. Para realizar os experimentos foi utilizado o método de divisão dos conjuntos de dados em Treino, Validação e Teste em uma proporção de 50:20:30, respectivamente. A Tabela I apresenta a quantidade de imagens em cada conjunto de dados, bem como o número de imagens utilizadas para o Treinamento, Validação e Teste.

TABELA I
TOTAL DE IMAGENS EM CADA CONJUNTO DE DADOS E A QUANTIDADE DE IMAGENS USADAS NAS ETAPAS DE TREINAMENTO, VALIDAÇÃO E TESTE.

| Dataset | Treinamento | Validação | Teste | Total de imagens |
|---------|-------------|-----------|-------|------------------|
| DDR | 383 | 149 | 225 | 757 |
| IDRiD | 42 | 17 | 22 | 81 |

O modelo foi treinado por 8.000 épocas e com tamanhos de lotes igual a 16. A taxa de aprendizagem foi de 0,01 e a taxa de *Momentum* de 0,937. Os tamanhos das âncoras assim como os demais hiperparâmetros do modelo foram ajustados com base nos resultados obtidos com a realização dos experimentos. Na Tabela II apresenta-se o melhor ajuste obtido para o método proposto.

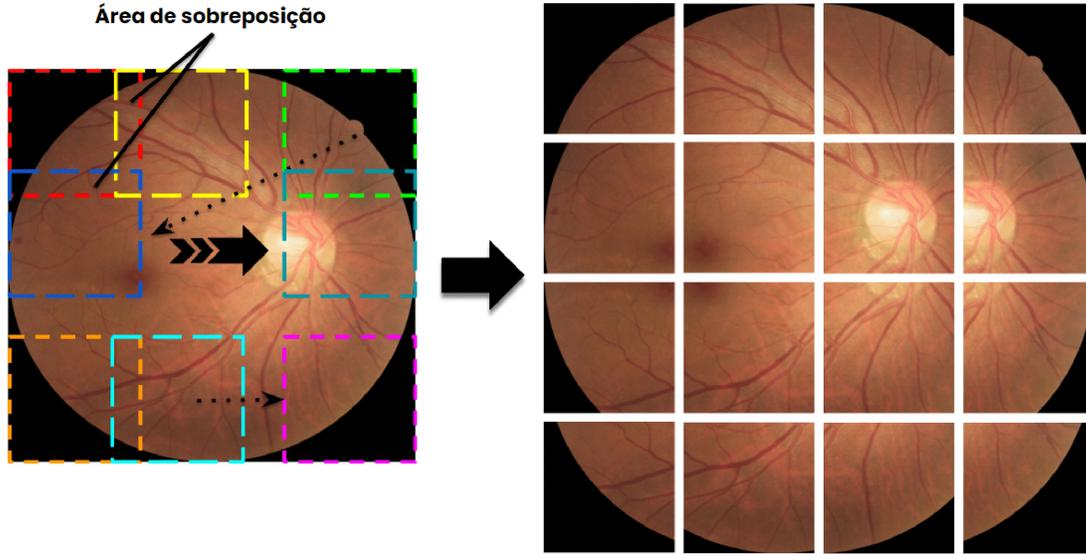


Fig. 3. Processo de *slice* realizado pelo *framework* SAHI, com fatias de tamanho 512×512 *pixels* e área de sobreposição de 20%. À esquerda, uma imagem de fundo do conjunto de dados DDR e, à direita, exemplos de sub-imagens criadas a partir da imagem original usando o método *slice* COCO disponível no *framework*.

TABELA II

AJUSTE DOS HIPERPARÂMETROS UTILIZADOS NO MÉTODO PROPOSTO.

| Hiperparâmetros | Valor |
|-----------------------|---|
| <i>Dropout</i> | 10% |
| Função de Ativação | SiLU |
| Limiar <i>IoU</i> NMS | 0,45 |
| Limite de Confiança | 0,25 |
| <i>Momentum</i> | 0,937 |
| Número de Épocas | 8.000 |
| Otimizador | SGD e Adam |
| Tamanho do Lote | 16 |
| Taxa de Aprendizagem | 0,01 |
| <i>Weight Decay</i> | 0,0005 |
| Âncoras iniciais | (10, 13), (16, 30), (33, 23) - P3 (30, 61), (62, 45), (59, 119) - P4 (116, 90), (156, 198), (373, 326) - P5 |
| Âncoras ajustadas | (3, 3), (4, 4), (7, 7) - P3 (10, 10), (15, 15), (23, 28) - P4 (33, 24), (44, 49), (185, 124) - P5 |

E. Métricas de Desempenho

A avaliação de um modelo é realizada de acordo com seu desempenho preditivo obtido em um conjunto de dados. A métrica *IoU* (*Intersection Over Union*) avalia o grau de sobreposição entre as caixas delimitadoras do *Ground Truth* e as caixas delimitadoras previstas pelo modelo. Nos experimentos foram utilizadas as métricas *AP* e *mAP* para o limite de *IoU* de 0,5 para fins de comparação com trabalhos correlatos encontrados na literatura que também utilizaram este limite para avaliar a precisão na detecção de lesões retinianas. A métrica *IoU* varia de 0 a 1, onde 0 indica nenhuma sobreposição entre as caixas e 1 indica uma sobreposição perfeita entre a caixa predita e o *Ground Truth*. A *IoU* é obtida da Equação 1, onde

gt se refere à caixa delimitadora do *Ground Truth* e *pd* a caixa delimitadora predita.

$$IoU = \frac{\text{Area Overlap}}{\text{Area Union}} = \frac{\text{Area}(gt \cap pd)}{\text{Area}(gt \cup pd)} \quad (1)$$

Já a métrica *AP* é definida pela área sob a curva de Precisão \times Revocação, também conhecida como curva PR. A precisão avalia a porcentagem de previsões corretas sobre a classe positiva. Já a revocação avalia dentre todas as situações da classe positiva como valor esperado, quantas estão corretas. Outra métrica comumente utilizada para avaliar modelos de detecção de objetos é a *mAP*, cujo cálculo é realizado através da média dos *AP* de todas as classes de objetos, conforme apresentado na Equação 2.

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (2)$$

IV. RESULTADOS

O modelo proposto foi comparado com diferentes modelos no estado da arte que realizam a detecção de objetos em estágio único. A Tabela III apresenta os modelos utilizados nos experimentos, bem como os resultados obtidos com as métricas *AP* e *mAP* para o limite de *IoU* de 0,5 com otimizador SGD (*Stochastic Gradient Descent*). O trabalho proposto utilizando a arquitetura YOLOR-CSP com *Cropping* e SAHI alcançou a melhor precisão na detecção das lesões de fundo nos experimentos realizados no conjunto de validação do *dataset* DDR, conforme destacado em negrito, com um *mAP* de 0,3808 e valores de *AP* iguais a 0,4090, 0,4660, 0,2610 e 0,3870, para EX, SE, MA e HE, respectivamente.

A Fig. 4 apresenta o gráfico referente à curva PR obtida durante a etapa de validação no conjunto de dados DDR para o limite de *IoU* de 0,5 e com o otimizador SGD. No gráfico o eixo *x* representa a Revocação e o eixo *y* a Precisão. No espaço da curva PR quanto mais próximo do canto superior

TABELA III

RESULTADOS OBTIDOS COM OTIMIZADOR SGD EM COMPARAÇÃO AOS TRABALHOS RELACIONADOS COM AS MÉTRICAS AP E mAP PARA O LIMITE DE IoU DE 0,5 NO CONJUNTO DE VALIDAÇÃO DO CONJUNTO DE DADOS DDR.

| Modelos | AP | | | | mAP |
|-------------------------------------|---------------|---------------|---------------|---------------|---------------|
| | EX | SE | MA | HE | |
| SSD [7] | 0 | 0,0227 | 0 | 0,0007 | 0,0059 |
| YOLO [7] | 0,0039 | 0 | 0 | 0,0101 | 0,0035 |
| YOLOv5+SGD+ <i>Tiling</i> [12] | 0,2290 | 0,3280 | 0,1050 | 0,3330 | 0,2490 |
| YOLOX-S [16] | 0,0909 | 0,3379 | 0,0636 | 0,1562 | 0,1622 |
| PPYOLOE-S [17] | - | - | - | - | 0,2030 |
| YOLOv5-S (r4.0) [13] | 0,1050 | 0,1990 | 0,0348 | 0,2020 | 0,1352 |
| YOLOv5-S (r5.0) [13] | 0,1140 | 0,2660 | 0,0415 | 0,2050 | 0,1566 |
| YOLOv5-S (r6.0) [13] | 0,1360 | 0,2600 | 0,0584 | 0,2200 | 0,1686 |
| YOLOv5-S (r6.1) [13] | 0,1270 | 0,2590 | 0,0495 | 0,2310 | 0,1666 |
| YOLOv5-S (r6.2) [13] | 0,1330 | 0,2540 | 0,0598 | 0,2150 | 0,1655 |
| YOLOR-CSP [18] | 0,2440 | 0,2990 | 0,1950 | 0,2880 | 0,2565 |
| YOLOv7-tiny-SiLU [19] | 0,1120 | 0,1370 | 0,0189 | 0,1480 | 0,1040 |
| Método inovador desenvolvido | 0,4090 | 0,4660 | 0,2610 | 0,3870 | 0,3808 |

direito a curva estiver melhor será o desempenho preditivo do modelo. Os valores representados no gráfico estão relacionados ao método inovador desenvolvido (indicada em negrito) na Tabela III.

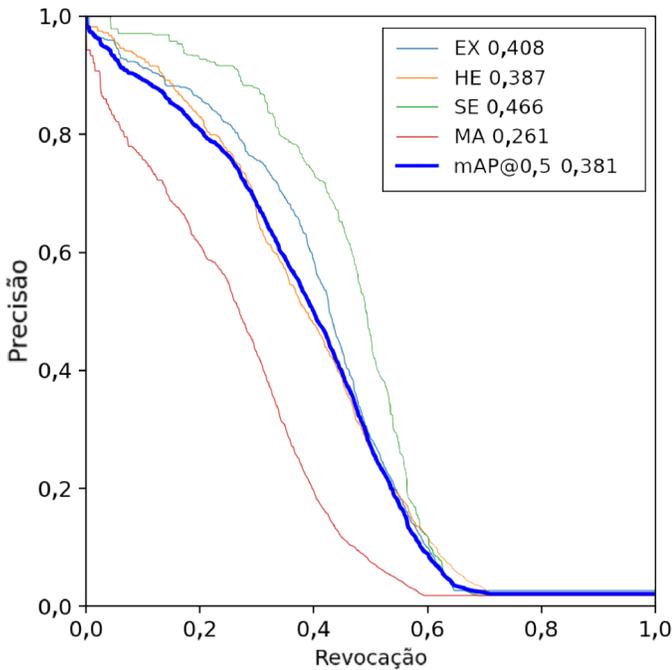


Fig. 4. Curva PR para o limite de IoU de 0,5 obtido durante a etapa de validação com otimizador SGD no *dataset* DDR.

É possível verificar que o trabalho proposto apresentou um desempenho superior em prever SE (curva na cor verde), EX (curva na cor azul claro) e HE (curva na cor laranja), onde foram obtidos valores de AP iguais a 0,466, 0,408 e 0,387, respectivamente. Por outro lado, o desempenho menos satisfatório obtido foi na detecção dos MA (curva na cor vermelha), com um valor de AP igual a 0,261. Além disso, a curva na cor azul marinho representa o mAP obtido por todas as classes, cujo valor obtido foi igual a 0,381.

Na Tabela IV são apresentados os resultados obtidos com as

métricas AP e mAP para o limite de IoU de 0,5 com otimizador SGD no conjunto de teste do *dataset* DDR. Obtiveram-se resultados superiores aos demais modelos, alcançando mAP de 0,222 e valores de AP iguais a 0,329 para EX, 0,203 para SE, 0,138 para MA e 0,218 para HE.

TABELA IV

RESULTADOS OBTIDOS COM OTIMIZADOR SGD EM COMPARAÇÃO AOS TRABALHOS RELACIONADOS COM AS MÉTRICAS AP E mAP PARA O LIMITE DE IoU DE 0,5 NO CONJUNTO DE TESTE DO CONJUNTO DE DADOS DDR.

| Modelos | AP | | | | mAP |
|-------------------------------------|---------------|---------------|---------------|---------------|---------------|
| | EX | SE | MA | HE | |
| SSD [7] | 0,0002 | 0 | 0,0001 | 0,0056 | 0,0015 |
| YOLO [7] | 0,0012 | 0 | 0 | 0,0109 | 0,0030 |
| YOLOv5+SGD+ <i>Tiling</i> [12] | 0,2100 | 0,1380 | 0,0530 | 0,1710 | 0,1430 |
| YOLOX-S [16] | 0,0909 | 0,1926 | 0,0909 | 0,0909 | 0,1163 |
| PPYOLOE-S [17] | - | - | - | - | 0,1000 |
| YOLOv5-S (r4.0) [13] | 0,1130 | 0,0789 | 0,0204 | 0,1110 | 0,0808 |
| YOLOv5-S (r5.0) [13] | 0,1210 | 0,0871 | 0,0180 | 0,1100 | 0,0840 |
| YOLOv5-S (r6.0) [13] | 0,1270 | 0,0866 | 0,0340 | 0,1290 | 0,0942 |
| YOLOv5-S (r6.1) [13] | 0,1200 | 0,0932 | 0,0272 | 0,1200 | 0,0901 |
| YOLOv5-S (r6.2) [13] | 0,1240 | 0,0693 | 0,0333 | 0,1240 | 0,0877 |
| YOLOR-CSP [18] | 0,2450 | 0,0711 | 0,0769 | 0,2100 | 0,1508 |
| YOLOv7-tiny-SiLU [19] | 0,1100 | 0,0654 | 0,0185 | 0,0673 | 0,0653 |
| Método inovador desenvolvido | 0,3280 | 0,2030 | 0,1400 | 0,2190 | 0,2225 |

Na Fig. 5 é apresentado o gráfico da curva PR obtido durante a etapa de teste no conjunto de dados DDR com o otimizador SGD para o limite de IoU de 0,5. Os valores traçados no gráfico estão relacionados ao trabalho proposto (destacado em negrito) na Tabela IV.

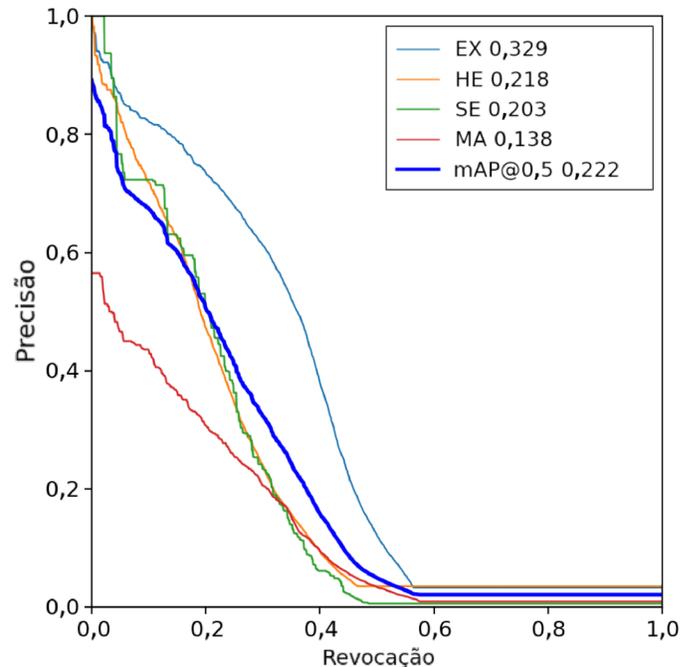


Fig. 5. Curva PR para o limite de IoU de 0,5 obtido durante a etapa de teste com otimizador SGD no *dataset* DDR.

Ao compará-lo com o comportamento apresentado na Fig. 4, é possível observar algumas divergências. Na etapa de validação o modelo proposto detectou SE com maior precisão porém, na etapa de teste, o melhor resultado obtido foi na

predição de EX (curva na cor azul claro). Por outro lado, a maior dificuldade em ambas etapas continuou sendo em prever MA (curva na cor vermelha).

Também foram realizados experimentos com o otimizador Adam. Na Tabela V são apresentados os resultados obtidos com as métricas AP e mAP para o limite de IoU de 0,5 no conjunto de validação do *dataset* DDR. Mais uma vez o trabalho proposto obteve resultados superiores aos demais modelos, alcançando um mAP de 0,3218 e valores de AP com iguais a 0,3760, 0,3620, 0,2340 e 0,3150, para EX, SE, MA e HE, respectivamente.

TABELA V

RESULTADOS OBTIDOS COM OTIMIZADOR ADAM EM COMPARAÇÃO AOS TRABALHOS RELACIONADOS COM AS MÉTRICAS AP E mAP PARA O LIMITE DE IoU DE 0,5 NO CONJUNTO DE VALIDAÇÃO DO CONJUNTO DE DADOS DDR.

| Modelos | AP | | | | mAP |
|-------------------------------------|---------------|---------------|---------------|---------------|---------------|
| | EX | SE | MA | HE | |
| SSD [7] | 0 | 0,0227 | 0 | 0,0007 | 0,0059 |
| YOLO [7] | 0,0039 | 0 | 0 | 0,0101 | 0,0035 |
| YOLOv5+Adam+Tiling [12] | 0,2240 | 0,3650 | 0,1110 | 0,3520 | 0,2630 |
| YOLOX-S [16] | 0,1689 | 0,5183 | 0,0909 | 0,3355 | 0,2784 |
| PPYOLOE-S [17] | - | - | - | - | 0,2040 |
| YOLOv5-S (r4.0) [13] | 0,1200 | 0,2030 | 0,0565 | 0,1840 | 0,1409 |
| YOLOv5-S (r5.0) [13] | 0,0099 | 0,3360 | 0,0188 | 0,0551 | 0,1050 |
| YOLOv5-S (r6.0) [13] | 0,1360 | 0,2170 | 0,0759 | 0,2360 | 0,1662 |
| YOLOv5-S (r6.1) [13] | 0,1440 | 0,2010 | 0,0664 | 0,2070 | 0,1546 |
| YOLOv5-S (r6.2) [13] | 0,1370 | 0,1760 | 0,0721 | 0,2040 | 0,1473 |
| YOLOR-CSP [18] | 0,1790 | 0,2810 | 0,1750 | 0,2580 | 0,2233 |
| YOLOv7-tiny-SiLU [19] | 0,0710 | 0,1140 | 0,0225 | 0,1290 | 0,0841 |
| Método inovador desenvolvido | 0,3760 | 0,3620 | 0,2340 | 0,3150 | 0,3218 |

A Tabela VI apresenta os resultados obtidos com as métricas AP e mAP para o limite de IoU de 0,5 com otimizador Adam no conjunto de teste do *dataset* DDR. O trabalho proposto obteve valores de AP iguais a 0,2980, 0,1070, 0,1150 e 0,2020, respectivamente nas classes de EX, SE, MA e HE e um mAP de 0,1805, superando os resultados obtidos pelos demais modelos.

TABELA VI

RESULTADOS OBTIDOS COM OTIMIZADOR ADAM EM COMPARAÇÃO AOS TRABALHOS RELACIONADOS COM AS MÉTRICAS AP E mAP PARA O LIMITE DE IoU DE 0,5 NO CONJUNTO DE TESTE DO CONJUNTO DE DADOS DDR.

| Modelos | AP | | | | mAP |
|-------------------------------------|---------------|---------------|---------------|---------------|---------------|
| | EX | SE | MA | HE | |
| SSD [7] | 0,0002 | 0 | 0,0001 | 0,0056 | 0,0015 |
| YOLO [7] | 0,0012 | 0 | 0 | 0,0109 | 0,0030 |
| YOLOv5+Adam+Tiling [12] | 0,2210 | 0,1570 | 0,0553 | 0,1840 | 0,1540 |
| YOLOX-S [16] | 0,1445 | 0,1793 | 0,0909 | 0,0909 | 0,1264 |
| PPYOLOE-S [17] | - | - | - | - | 0,0980 |
| YOLOv5-S (r4.0) [13] | 0,1200 | 0,0373 | 0,0251 | 0,1140 | 0,0741 |
| YOLOv5-S (r5.0) [13] | 0,0223 | 0,0464 | 0,0251 | 0,0405 | 0,0336 |
| YOLOv5-S (r6.0) [13] | 0,1410 | 0,0799 | 0,0344 | 0,1300 | 0,0963 |
| YOLOv5-S (r6.1) [13] | 0,1460 | 0,0487 | 0,0328 | 0,1350 | 0,0906 |
| YOLOv5-S (r6.2) [13] | 0,1410 | 0,0459 | 0,0353 | 0,1270 | 0,0873 |
| YOLOR-CSP [18] | 0,1920 | 0,0927 | 0,0769 | 0,2120 | 0,1434 |
| YOLOv7-tiny-SiLU [19] | 0,0599 | 0,0560 | 0,0195 | 0,0480 | 0,0459 |
| Método inovador desenvolvido | 0,2980 | 0,1070 | 0,1150 | 0,2020 | 0,1805 |

O trabalho proposto apresentou resultados superiores aos demais trabalhos de mesmo propósito encontrados na literatura.

Ao comparar os resultados obtidos com os otimizadores SGD e Adam foi possível perceber que ao utilizar o otimizador SGD a abordagem proposta obteve uma maior precisão na detecção das lesões de fundo. A Tabela VII apresenta o tempo médio de inferência para detectar as lesões de fundo por imagem no conjunto de dados DDR nas etapas de validação e teste em comparação ao trabalho proposto por Santos et al. [12].

TABELA VII

TEMPO MÉDIO DE INFERÊNCIA POR IMAGEM NO CONJUNTO DE DADOS DDR NAS ETAPAS DE VALIDAÇÃO E TESTE EM COMPARAÇÃO A TRABALHOS DE MESMO PROPÓSITO ENCONTRADOS NA LITERATURA.

| Modelos | Tempo de Inferência (ms) | |
|-----------------------------------|--------------------------|-------|
| | Validação | Teste |
| YOLOv5+Adam+Tiling [12] | 5,5 | 7,5 |
| YOLOv5+SGD+Tiling [12] | 4,6 | 5,9 |
| Método inovador desenvolvido+SGD | 16,4 | 15,8 |
| Método inovador desenvolvido+Adam | 16,2 | 15,7 |

Em Santos et al. [12] o menor tempo médio de inferência por imagem foi obtido com o otimizador SGD e *Tiling*, com 4,6 ms. Já o trabalho proposto, o menor tempo médio de inferência foi obtido com o otimizador Adam, com 16,2 ms no conjunto de validação e de 15,7 ms no conjunto de teste. Além disso, os resultados obtidos pelo trabalho proposto utilizando o otimizador SGD foram bastante similares aos obtidos com Adam, chegando ao tempo de 16,4 ms no conjunto de validação e de 15,8 ms no conjunto de teste. Apesar do trabalho proposto por Santos et al. [12] ser cerca de 3 vezes mais rápido quando comparado ao nosso trabalho, o método proposto demonstrou ser mais preciso na detecção das lesões de fundo.

Também foram realizados experimentos com o conjunto de dados público de RD IDRiD [15] a fim de avaliar a precisão da abordagem proposta em diferentes conjuntos de dados. Na Tabela VIII são apresentados os resultados obtidos com as métricas AP e mAP para o limite de IoU de 0,5 no conjunto de validação do *dataset* IDRiD. Durante esta etapa o melhor resultado obtido foi utilizando o otimizador SGD, com mAP de 0,4453 e valores de AP iguais a 0,4350, 0,5470, 0,3960 e 0,4030, para EX, SE, MA e HE, respectivamente.

TABELA VIII

RESULTADOS OBTIDOS COM AS MÉTRICAS AP E mAP PARA O LIMITE DE IoU DE 0,5 NO CONJUNTO DE VALIDAÇÃO DO CONJUNTO DE DADOS IDRiD.

| Modelos | AP | | | | mAP |
|-----------------------------------|---------------|---------------|---------------|--------|---------------|
| | EX | SE | MA | HE | |
| Método inovador desenvolvido+SGD | 0,4350 | 0,5470 | 0,3960 | 0,4030 | 0,4453 |
| Método inovador desenvolvido+Adam | 0,4330 | 0,3790 | 0,3970 | 0,4030 | 0,4030 |

Durante a etapa de teste, utilizando o conjunto de dados IDRiD, o melhor resultado obtido foi utilizando o otimizador Adam, atingindo um mAP de 0,4170 e valores de AP iguais a 0,5110, 0,4360, 0,3570 e 0,3640, respectivamente, nas classes de EX, SE, MA e HE, conforme apresentado a Tabela IX.

TABELA IX

RESULTADOS OBTIDOS COM AS MÉTRICAS AP E mAP PARA O LIMITE DE IoU DE 0,5 NO CONJUNTO DE TESTE DO CONJUNTO DE DADOS IDRiD.

| Modelos | AP | | | | mAP |
|-----------------------------------|---------------|---------------|---------------|---------------|---------------|
| | EX | SE | MA | HE | |
| Método inovador desenvolvido+SGD | 0,5170 | 0,4110 | 0,3950 | 0,3220 | 0,4113 |
| Método inovador desenvolvido+Adam | 0,5110 | 0,4360 | 0,3570 | 0,3640 | 0,4170 |

Os resultados obtidos durante as etapas de validação e teste usando o conjunto de dados IDRiD apresentaram uma maior precisão na detecção das lesões retinianas, superando os resultados obtidos no conjunto de dados DDR. Estes resultados validam a capacidade de generalização do método inovador proposto para a detecção de lesões de fundo associadas à RD.

V. CONCLUSÕES

Este artigo apresentou um modelo de rede neural convolucional baseado em um detector de estágio único no estado da arte para detectar lesões de fundo causadas pela RD. A etapa de aumento de dados possibilitou melhorar a precisão do modelo, mesmo treinando a rede neural com tamanho de lotes menores. Os melhores resultados foram obtidos utilizando o otimizador SGD combinado com os métodos *Cropping* e *Slice COCO do framework SAHI*, alcançando um mAP de 0,3808 na etapa de validação e de 0,2225 na etapa de teste para o limite de IoU de 0,5. Na etapa de validação no conjunto de dados DDR o método proposto alcançou um AP de 40,9% para Exsudatos Duros, 46,6% para Exsudatos Algodonosos, 26,1% para Microaneurismas e de 38,7% para Hemorragias. Além disso, no conjunto de dados IDRiD com o otimizador SGD, foi obtido um mAP de 0,4453 na etapa de validação e de 0,4113 na etapa de teste. Também foram alcançados valores de AP iguais a 43,50% para Exsudatos Duros, 54,70% para Exsudatos Algodonosos, 39,60% para Microaneurismas e de 40,30% para Hemorragias. Ademais, a área de saúde ocular e visual deve ser considerada uma das prioridades do Sistema Único de Saúde (SUS), pois sua manutenção e aprimoramento requerem esforços significativos, principalmente da inclusão de novas tecnologias para o atendimento regular à população. Os resultados apresentados neste artigo demonstram que é possível identificar com precisão lesões causadas pela RD por meio de um modelo computacional e, portanto, acreditamos que este estudo poderá contribuir de maneira significativa para melhorar a eficácia do atendimento e tratamento prestados aos pacientes assistidos pelo SUS.

AGRADECIMENTOS

Este estudo foi financiado parcialmente pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Código de Financiamento 001.

REFERÊNCIAS

- [1] P. Riordan-Eva and E. T. Cunningham, *Vaughan & Asbury's General Ophthalmology, 19th Edition*, 19th ed. McGraw-Hill Education/Medical, Oct. 2017.

- [2] I. B. de Geografia e Estatística (IBGE), “Censo 2000,” <https://www.ibge.gov.br/estatisticas/sociais/administracao-publica-e-participacao-politica/9663-censo-demografico-2000.html>, 2000, [Online]; accessed 06-Mar-2023].
- [3] A. A. Alghadyan, “Diabetic retinopathy - An update,” *Saudi Journal of Ophthalmology*, vol. 25, no. 2, pp. 99–111, 2011. [Online]. Available: <http://dx.doi.org/10.1016/j.sjopt.2011.01.009>
- [4] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, “You only learn one representation: Unified network for multiple tasks,” 2021. [Online]. Available: <https://arxiv.org/abs/2105.04206>
- [5] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “Yolov4: Optimal speed and accuracy of object detection,” 2020.
- [6] F. C. Akyon, S. O. Altinuc, and A. Temizel, “Slicing aided hyper inference and fine-tuning for small object detection,” in *Proceedings of the International Conference on Image Processing (ICIP)*. IEEE, Oct. 2022. [Online]. Available: <https://doi.org/10.1109/icip46576.2022.9897990>
- [7] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, and H. Kang, “Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening,” *Information Sciences*, vol. 501, pp. 511–522, 2019. [Online]. Available: <https://doi.org/10.1016/j.ins.2019.06.011>
- [8] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-CNN: Towards real-time object detection with region proposal networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017. [Online]. Available: <https://doi.org/10.1109/tpami.2016.2577031>
- [9] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “Ssd: Single shot multibox detector,” in *Proceedings of European Conference on Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 21–37.
- [10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, Las Vegas, NV, USA, 27–30 June 2016, 2016, pp. 779–788.
- [11] C. Santos, M. S. De Aguiar, D. Welfer, and B. Belloni, “Deep neural network model based on one-stage detector for identifying fundus lesions,” in *2021 International Joint Conference on Neural Networks (IJCNN)*, Shenzhen, China, 18–22 July 2021, 2021, pp. 1–8.
- [12] C. Santos, M. Aguiar, D. Welfer, and B. Belloni, “A new approach for detecting fundus lesions using image processing and deep neural network architecture based on yolo model,” *Sensors*, vol. 22, no. 17, 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/17/6441>
- [13] J. Glenn, “YOLOv5 releases,” <https://github.com/ultralytics/yolov5/releases/>, 2022, [Online; accessed 04-September-2022].
- [14] “<https://github.com/wongkinyiu/yolor>.”
- [15] P. Porwal, S. Pachade, M. Kokare, G. Deshmukh, J. Son, W. Bae, L. Liu, J. Wang, X. Liu, L. Gao *et al.*, “IDRiD: Diabetic Retinopathy – Segmentation and Grading Challenge,” *Medical Image Analysis*, vol. 59, 2020.
- [16] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, “Yolox: Exceeding yolo series in 2021,” 2021.
- [17] S. Xu, X. Wang, W. Lv, Q. Chang, C. Cui, K. Deng, G. Wang, Q. Dang, S. Wei, Y. Du, and B. Lai, “Pp-yolox: An evolved version of yolo,” 2022. [Online]. Available: <https://arxiv.org/abs/2203.16250>
- [18] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, “You only learn one representation: Unified network for multiple tasks,” 2021.
- [19] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” 2022. [Online]. Available: <https://arxiv.org/abs/2207.02696>



Alejandro Pereira is an undergraduate student in Computer Science at the Federal University of Pelotas (UFPEL). From 2021 to 2022 he participated in scientific initiation programs funded by CNPq, working on the design of cyber-physical systems for pest management. His research has focused on the areas of Artificial Intelligence, especially using computer vision to solve health problems, and also in the area of wireless sensor networks applied to pest management.



Carlos Santos Ph.D. candidate at the Postgraduate Program in Computing, Federal University of Pelotas (UFPEL). Received the B.Sc. in Informatics from University of the Region of Campanha (URCAMP), in 2010 and his M.Sc. in Electrical Engineering from the Federal University of Pampa (UNIPAMPA), in 2017. Currently, he is a Professor at Federal Institute of Education, Science and Technology Farroupilha (IFFAR). His current research interests include Applications of Artificial Intelligence in Diagnostic Imaging, Computer Aided Detection (CAD) and Computer Aided Diagnosis (CADx).



Marilton Aguiar Associate Professor at the Federal University of Pelotas (UFPEL) in the scope of the Undergraduate Courses in Computer Science and Engineering and the Graduate Program in Computing. Director of the Center for Technological Development at the Federal University of Pelotas. His research has focused on applications of Artificial Intelligence in solving environmental and health problems.



Daniel Welfer Professor of the Department of Applied Computing of the Federal University of Santa Maria (UFSM). Currently, he is also a permanent professor at the Graduate Program in Computer Science (PPGCC). Specializing in the fields of image processing, artificial intelligence and software engineering.



Marcelo Dias is an undergraduate student of Computer Science at Federal University of Pelotas (UFPEL). He has participated in undergraduate research programs (PROBIC/FAPERGS) funded by government agencies. His research interests are in the areas of Artificial Intelligence, using image detection to diagnostic.



Marcelo Ribeiro is an undergraduate student of Computer Science at the Federal University of Pelotas (UFPEL). His research interests are in the areas of Artificial Intelligence, with focus on object detection and synthetic data generation.