



**UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIAS
COORDENAÇÃO DE CIÊNCIA DA COMPUTAÇÃO
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

ANDRÉ FELIPE CARVALHO PINHEIRO

**DETECÇÃO AUTOMÁTICA DO DISCO ÓPTICO EM
RETINOGRÁFIAS UTILIZANDO YOLO**

**São Luis
2019**

ANDRÉ FELIPE CARVALHO PINHEIRO

**DETECÇÃO AUTOMÁTICA DO DISCO ÓPTICO EM
RETINOGRÁFIAS UTILIZANDO YOLO**

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

Orientador: Prof. Dr. João Dallyson Sousa de Almeida

São Luis
2019

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

Carvalho Pinheiro, André Felipe.

Detecção Automática do Disco Óptico em
Retinografias Utilizando Yolo / André Felipe Carvalho
Pinheiro. - 2019.

56 p.

Orientador(a): João Dallyson Sousa de Almeida.
Monografia (Graduação) - Curso de Ciência da
Computação, Universidade Federal do Maranhão, São Luís,
2019.

1. Detecção. 2. Disco Óptico. 3. Redes Neurais. 4.
Retinografia. 5. Yolo. I. Sousa de Almeida, João
Dallyson. II. Título.

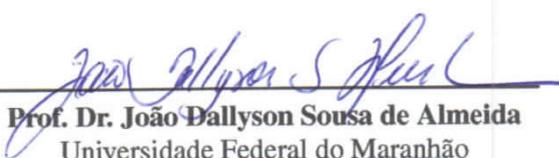
ANDRÉ FELIPE CARVALHO PINHEIRO

**DETECÇÃO AUTOMÁTICA DO DISCO ÓPTICO EM
RETINOGRÁFIAS UTILIZANDO YOLO**

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

Data da Defesa: 10/01/2019

Banca Examinadora


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


Prof. Dr. Geraldo Braz Junior
Universidade Federal do Maranhão
Examinador 1


Prof. Ms. Giovanni Lucca França da Silva
Universidade Federal do Maranhão
Examinador 2

São Luis
2019

À minha mãe e meu pai pela confiança e investimento, a minha filha Laura, pelo futuro.

AGRADECIMENTOS

À minha família, em especial minha mãe Eron por ter me dado a oportunidade de sempre estudar durante minha juventude mesmo diante das adversidades, meu pai Eurípedes por sempre exigir que eu desse meu melhor, meus tios e tias Erany, Erinalva, Eralda, Eroilza, Odylo, Walter, Erineu por terem cuidado de mim como um filho durante minha infância e terem ajudado a formar meu caráter, aos meus irmãos que espero seguirem meu exemplo de dedicação, a minha namorada Larissa pelo apoio nos momentos mais difíceis e a mais nova integrante que se juntará a família em fevereiro, com muito amor, minha filha Laura.

Ao Colégio Universitário e todos os professores que lá tive, por terem contribuído não só com minha formação escolar, mas como cidadão e me fazerem além de sonhar alto, saber que posso atingir meus objetivos.

À Universidade Federal do Maranhão e seu corpo Docente que mesmo diante de todos os problemas me proporcionaram um bom ambiente acadêmico e a mais do que isso foram os que mais contribuíram para a construção do meu senso crítico, visão de mundo e convicções, além de proporcionar o acesso e a produção de Ciência, esta que se tornou um dos meus principais *hobbies*.

Ao DEINF e professores, mas em especial ao meu orientador Dallyson por ter acreditado e investido em mim durante os últimos anos que me proporcionou ter um bom currículo acadêmico. Aos professores Geraldo e Luciano pelo seu rigor, que me reprovaram em disciplinas mas com isso me fizeram dar o melhor de mim nos períodos seguintes e com isso aprender que sou capaz. Ao professor Anselmo por proporcionar a participação em projetos durante minha vida acadêmica e por suas inesquecíveis aulas e didática com os cofos, as quais foram essenciais para minha formação. A todos estes profissionais, declaro a minha mais profunda, sincera admiração e inspiração.

Aos meus amigos da turminha, sem estes os últimos 6 anos não teriam sido vividos com a mesma intensidade que foram: Ricardo, Gabriel, Alex, Schönfinkel, Rodrigo, Márcio, Júlia, Tárcio, Hugo, Allyson, Geovane, Robherson, Leonardo, Wellington, Arlem, muito obrigado.

"Em algum lugar, alguma coisa incrível está esperando para ser conhecida."
(Carl Sagan)

RESUMO

A visão é um dos sentidos mais importantes e mais utilizado dentre os seres vivos, nos seres humanos permitindo perceber o mundo ao seu redor através da luminosidade em um complexo processo entre as membranas que envolvem os olhos: esclera, coroide e retina. As células responsáveis por transformar a informação luminosa captada pelo olho em impulsos elétricos para o cérebro através da retina chamadas de fotorreceptoras. Em uma parte da retina estas células não estão presentes, neste ponto cego é onde todas as fibras neuronais do nervo óptico se unem, responsáveis por transmitir os impulsos elétricos entre a retina e o cérebro, este lugar é denominado de Disco Óptico (DO). Patologias como diabetes e o glaucoma podem manifestar-se na retina e o glaucoma como a segunda maior causa de cegueira no mundo. As causas ainda não são totalmente conhecidas, mas o principal fator para o desenvolvimento desta patologia é a alta pressão intraocular que pode ser causada principalmente por lesões na retina por consequência do diabetes ou da idade elevada. Uma das maneiras de realizar o diagnóstico do glaucoma é através da retinografia, que é o exame feito da retina para analisar estruturas como a mácula, fóvea, vasos sanguíneos e o DO. O campo que envolve o uso de ferramentas computacionais para auxílio ou acompanhamento do diagnóstico médico na oftalmologia por meio de imagens digitais de retina podem prover informações sobre mudanças patológicas. A detecção e o tratamento precoce dos vários tipos de retinopatias são cruciais para evitar perda da visão. Atentos a esta problemática, este estudo objetiva propor um novo método computacional para detectar automaticamente a região do DO em retinografias, realizando o enquadramento da menor área possível, isso possibilita que doenças como glaucoma e retinopatia diabética possam ser automaticamente diagnosticadas em etapas posteriores com melhor acuidade. Para tanto, pretende-se utilizar a rede neural convolucional profunda especializadas em detecção de objetos: YOLO, obtendo resultado de 98,14% de acerto em uma base de 3428 imagens.

Palavras-chaves: Detecção, Retinografia, Disco Óptico, Redes Neurais, Yolo.

ABSTRACT

Vision is one of the most important and most used senses among human beings allowing us to perceive the world around us through the luminosity in a complex process among the membranes that surround the eyes: sclera, choroid and retina. The cells in charge of transforming the luminous information captured by the eye into electrical impulses to the brain through the retina are called photoreceptors. In a part of the retina these cells are not present, in this blind spot all neuronal fibers of the optic nerve connect and are responsible for transmitting the electrical impulses between the retina and the brain, this place is denominated Optical Disc (DO). Pathologies such as diabetes and glaucoma can manifest in the retina and glaucoma as the second leading cause of blindness in the world. The causes are not yet fully understood, but the main factor for the development of this pathology is the high intraocular pressure that can be caused mainly by retinal damage due to diabetes or advanced age. One way to diagnose glaucoma is through retinal examination, which is the retinal exam used to analyze structures such as the macula, fovea, blood vessels, and the optic disc (OD). The field involving the use of computational tools to aid or monitor medical diagnosis in ophthalmology through digital retinal images may provide information on pathological changes. Detection and early treatment of various types of retinopathies are crucial to avoid loss of vision. This study aims to propose a new computational method to automatically detect the DO region in retinographies, making it possible to diagnose diseases such as glaucoma and diabetic retinopathy in later stages with better acuity . In order to do so, we intend to use the deep convolutional neural network specialized in object detection: YOLO, obtaining the result of 98.14% accuracy in a base of 3428 images.

Key-words: Detection, Fundus, Optic Disk, Neural Networks , Yolo.

LISTA DE FIGURAS

Figura 1 – Olho humano.	14
Figura 2 – Retinografia de um olho Glaucomatoso	15
Figura 3 – Ilustração de uma RNA de única camada.	17
Figura 4 – Ilustração de uma RNA MLP.	18
Figura 5 – Ilustração de uma convolução de um filtro sobre uma imagem.	19
Figura 6 – Representação de um filtro max-pooling sobre uma imagem.	20
Figura 7 – Arquitetura da LeNet.	20
Figura 8 – Arquitetura da Yolo.	22
Figura 9 – Cálculo baseado na âncora.	22
Figura 10 – Computação da IoU.	24
Figura 11 – Resultados obtidos por Silva <i>et al.</i> (2016) comparados com outros estudos.	28
Figura 12 – Arquitetura de Niu <i>et al.</i> (2017).	28
Figura 13 – Resultados do Algoritmo <i>Saliency Map</i> em Niu <i>et al.</i> (2017).	29
Figura 14 – Resultados positivos de Niu <i>et al.</i> (2017).	29
Figura 15 – (a) Região original do DO. (b) Imagem normalizada de (a). (c) Branqueamento de (b). (d) Não região do DO. (e) Imagem normalizada de (d). (f) Branqueamento de (e). (g) Imagem média.	30
Figura 16 – Arquitetura proposta por Alghamdi <i>et al.</i> (2016).	30
Figura 17 – Arquitetura proposta por Xu <i>et al.</i> (2017).	31
Figura 18 – A esquerda uma imagem normal, a direita uma imagem de paciente com glaucoma.	34
Figura 19 – Imagem da base MESSIDOR.	34
Figura 20 – Imagem da base STARE.	35
Figura 21 – Imagem da base HRF.	35
Figura 22 – Imagem da base Drive.	36
Figura 23 – Imagem da base Drishti.	37
Figura 24 – Imagem da base Drions.	37
Figura 25 – Imagem da base Drions.	38
Figura 26 – Imagem da base de Alipour <i>et al.</i> (2012).	38
Figura 27 – Imagem da base de JSIEC.	39
Figura 28 – Fluxo da Metodologia.	39
Figura 29 – Diagrama detalhando a separação das bases de treino e teste.	43
Figura 30 – Gráfico comparativos dos resultados obtidos por experimento e base utilizada.	44
Figura 31 – Resultado da marcação de do <i>bouding box</i> de Niu <i>et al.</i> (2017).	45
Figura 32 – Resultado da marcação de do <i>bouding box</i> de Alghamdi <i>et al.</i> (2016).	46
Figura 33 – Erros do Experimento 1 na base STARE.	47
Figura 34 – Erros do Experimento 1 na base JSIEC.	48
Figura 35 – Piores resultados em termos de IoU.	49

LISTA DE TABELAS

Tabela 1 – Bases de imagens utilizadas por Silva <i>et al.</i> (2016)	26
Tabela 2 – Comparação dos resultados das duas arquiteturas em Niu <i>et al.</i> (2017).	29
Tabela 3 – Resultados de Alghamdi <i>et al.</i> (2016).	31
Tabela 4 – Resultados de Xu <i>et al.</i> (2017).	32
Tabela 5 – Bases de imagens utilizadas neste trabalho.	33
Tabela 6 – Arquitetura da CNN utilizada.	41
Tabela 7 – Comparação dos resultados obtidos deste trabalho com o método de (a) (SILVA <i>et al.</i> , 2016)	45
Tabela 8 – Comparação dos resultados obtidos com o método de Niu <i>et al.</i> (2017).	46
Tabela 9 – Comparação com os resultados de Alghamdi <i>et al.</i> (2016).	46
Tabela 10 – Comparação com os resultados de Xu <i>et al.</i> (2017).	47
Tabela 11 – Avaliação dos resultados em termos de IoU.	48
Tabela 12 – (a) Alghamdi <i>et al.</i> (2016), (b) Xu <i>et al.</i> (2017), (c) Niu <i>et al.</i> (2017)	49

LISTA DE ABREVIATURAS E SIGLAS

CNN	Convolutional Neural Network.
DO	Disco Óptico.
IoU	Intersection Over Union.
MLP	Multilayer Perceptron.
PDO	<i>Pixel</i> do Disco Óptico
RNA	Rede Neural Artificial.
SVM	Support Vector Machine.
YOLO	You Only Look Once.

SUMÁRIO

1	INTRODUÇÃO	14
1.1	OBJETIVOS	15
1.1.1	Objetivo Geral	15
1.1.2	Objetivos Específicos	15
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	Redes Neurais Artificiais	17
2.2	Redes Neurais Convolucionais	19
2.3	You Only Look Once v2	21
2.4	Métricas de Validação	23
2.4.1	Intersect Over Union	24
2.4.2	Métrica Padrão em Detecção do DO	24
3	TRABALHOS RELACIONADOS	25
3.1	Métodos que Utilizam Processamento de Imagens	25
3.2	Métodos que Utilizam CNN	28
4	METODOLOGIA	33
4.1	Bases Utilizadas	33
4.1.1	Rim-One	33
4.1.2	MESSIDOR	34
4.1.3	STARE	35
4.1.4	HRF	35
4.1.5	DRIVE	36
4.1.6	Drishti	36
4.1.7	Drions	36
4.1.8	DIARETDB0 E DIARETDB1	37
4.1.9	Alipour Dataset	38
4.1.10	JSIEC Dataset	38
4.2	Método Proposto	39
4.2.1	Marcação da Região do DO	40
4.2.2	Treinamento da Rede	40
5	RESULTADOS E DISCUSSÃO	43
5.0.1	Experimentos	43
5.1	Comparação com técnicas clássicas de processamento de imagem	44
5.2	Comparação com técnicas de Deep Learning	45

6	CONCLUSÃO	50
----------	------------------	-----------

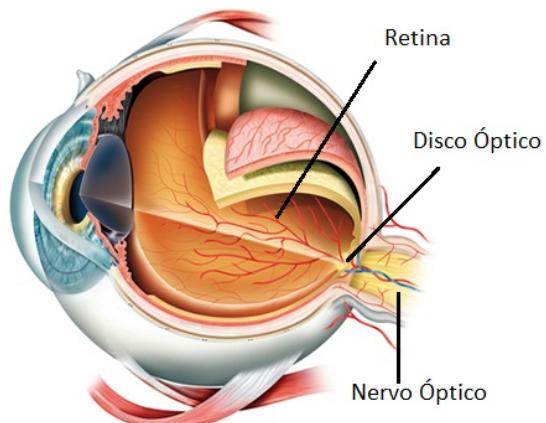
REFERÊNCIAS	51
--------------------	-----------

1 INTRODUÇÃO

A visão é um dos sentidos mais importantes e mais utilizado dentre os seres vivos, nos seres humanos permitindo perceber o mundo ao seu redor através da luminosidade em um complexo processo entre as membranas que envolvem os olhos: esclera, coroide e retina, esta última é a camada mais interna do olho, também chamada de fundo de olho, responsável por captar a luz transformando-a em sinal elétrico para ser interpretado pelo cérebro que cria uma representação perceptual deste impulso (MORAES, 2018).

As células responsáveis por transformar esta informação luminosa em impulsos elétricos são chamadas de fotorreceptoras que são classificadas em dois tipos: cones e bastonetes, sendo elas responsáveis por perceber as cores (cones) e tonalidades de claro e escuro (bastonetes). Em uma parte da retina não há células fotorreceptoras, consequentemente um ponto cego, este é onde todas as fibras neuronais do nervo óptico se unem, responsáveis por transmitir (Figura 1) os impulsos elétricos entre a retina e o cérebro, este lugar é denominado de Disco Óptico (DO).

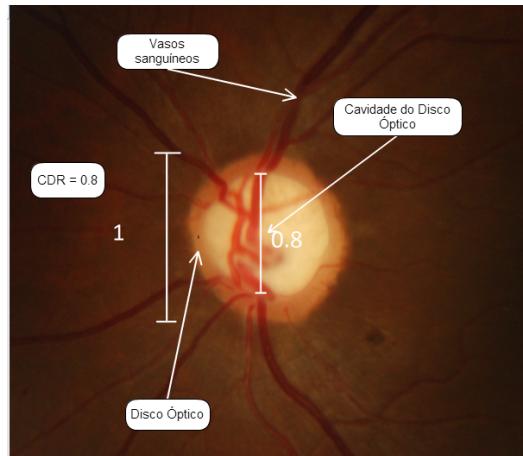
Figura 1 – Olho humano.



Fonte: Adaptado de Moraes (2018).

Nas imagens de fundo de olho, patologias como diabetes e o glaucoma podem ser diagnosticadas com auxílio da localização do DO, que possui normalmente formato oval e diâmetro médio de 1,5 mm, podendo variar de acordo com a população e o eixo vertical tende a ser entre 5 a 10% maior que o horizontal. Já a cavidade do Disco Óptico (CDO) localiza-se centralmente dentro do DO e também tem forma arredondada e o maior eixo horizontal. A razão entre os diâmetros verticais da CDO e do DO (CDR) é usada para indicar um fator de risco glaucomatoso. Um olho normal possui CDR igual ou inferior a 0,5, qualquer valor superior a esse já pode ser considerado como forte indicio de glaucoma, pois menos de 7% dos olhos normais possuem essa característica (PARANHOS-JUNIOR; OMI; PRATA-JUNIOR, 2009), como o da Figura 2.

Figura 2 – Retinografia de um olho Glaucomatoso



Fonte: Elaborado pelo autor.

É de imensa importância que o diagnóstico de patologias relacionadas a retina como glaucoma ou diabetes seja realizado o mais previamente possível a fim de evitar as consequências mais graves como a cegueira. Sob esta ótica faz-se necessário métodos eficientes, rápidos e precisos para o seu diagnóstico, o que passa necessariamente pela análise do Disco Óptico (DO) e sua cavidade (CDO).

Trabalhos que envolvem a utilização de imagens de fundo de olho e *Convolutional Neural Networks* (CNN) como Xu *et al.* (2017) utiliza a arquitetura da VGG (SIMONYAN; ZISSERMAN, 2014), esta arquitetura originalmente classifica imagens inteiras, mas o autor propõe uma abordagem em que remove-se as camadas totalmente conectadas e classifica *pixel a pixel* como pertencente ou não na região do DO, usando as bases MESSIDOR, STARE e ORIGA (ZHANG *et al.*, 2010).

Este trabalho é justificável por oferecer um meio de facilitar o diagnóstico destas patologias através da localização do DO, o que na maioria dos casos de diagnóstico por retinografias é a primeira etapa, seja para eliminar falsos positivos na segmentação ou localização de exsudatos para diagnóstico de retinopatia diabética ou para análise da CDO no diagnóstico do glaucoma.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

O objetivo deste trabalho é detectar a região do DO em imagens de retinografia.

1.1.2 Objetivos Específicos

- Utilizar 11 bases de imagens de retinografias publicamente disponíveis.

- Realizar a marcação manual da região do DO nas imagens sem marcação.
- Treinar uma arquitetura de rede neural convolucional profunda capaz de localizar a região do DO.
- Comparar os resultados aos apresentados no estado da arte.
- Disponibilizar a rede treinada para detectar o DO em retinografias.

2 FUNDAMENTAÇÃO TEÓRICA

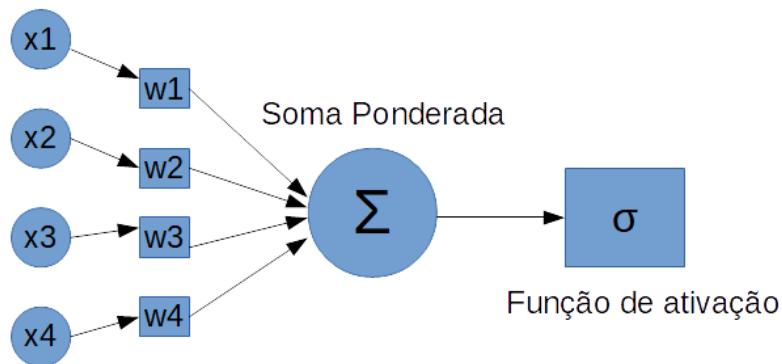
Este capítulo aborda a fundamentação teórica utilizada para atingir os objetivos deste trabalho, serão abordados os conceitos de técnicas de *deep learning*, que atualmente são estado da arte em problemas de detecção de objetos, para a localização da região do DO (REDMON *et al.*, 2016).

2.1 Redes Neurais Artificiais

Um dos passos mais importantes para compreender o funcionamento das redes neurais artificiais (RNA) é entender como funciona o seu treinamento, e deste processo é necessário discorrer sobre a utilização do Gradiente Descendente para a correção dos pesos de uma rede pois este é o principal algoritmo de minimização de erros em RNAs. A Figura 3 ilustra o funcionamento de uma RNA simples de uma camada, onde para cada padrão de entrada que é inserido na rede $x = x_1, x_2, \dots, x_n$ multiplicado pelos seus respectivos n pesos w_1, w_2, \dots, w_n espera-se uma saída S da rede para que esse padrão seja igual a d , sendo assim obtemos:

$$d = S = \sum_{i=0}^N w_i x_i \quad (2.1)$$

Figura 3 – Ilustração de uma RNA de única camada.



Fonte: Elaborado pelo autor.

Widrow e Hoff (1960) propuseram então uma função de erro para avaliar o desempenho da rede, de forma que:

$$E = E(d - S) = E(d - \sum_{i=0}^N w_i x_i) = E(w_0, w_1, \dots, w_n) = E(\vec{w}) \quad (2.2)$$

A função E é uma função dos pesos da rede, pois os valores da saída desejada d e entradas x é dado de antemão pelo problema, sendo assim a tarefa de treinamento da rede

consiste em um processo de busca pelo espaço de pesos que produza o menor erro possível. Widrow e Hoff (1960) propuseram então uma função de erro simples e popular em diversas áreas, o erro quadrático médio, a média da soma dos erros para cada padrão submetido a rede:

$$E = \frac{1}{2M} \sum_{p=1}^M \left[d_p - \sum_{i=0}^N w_i x_i \right]^2 \quad (2.3)$$

Onde M é o número de padrões, d o valor esperado e P o padrão. O resultado desta função é sempre positivo ou nulo, e quanto mais ela se aproxima de 0, melhor o desempenho da rede. O objetivo do treinamento então é buscar os valores dos coeficientes que possam minimizar esta função de erro. Esta é a parte onde aplica-se o gradiente descendente, que é uma ferramenta padrão de otimização de funções iterativamente, partindo do princípio de que a derivada de uma função em um ponto é a inclinação da reta tangente naquele ponto, ao derivar a função de erro em relação ao peso w_i , obtemos a direção em que essa função cresce com a variação do peso w_i , como o objetivo é minimizar o erro, então toma-se a direção oposta deste crescimento, daí obtemos a regra delta:

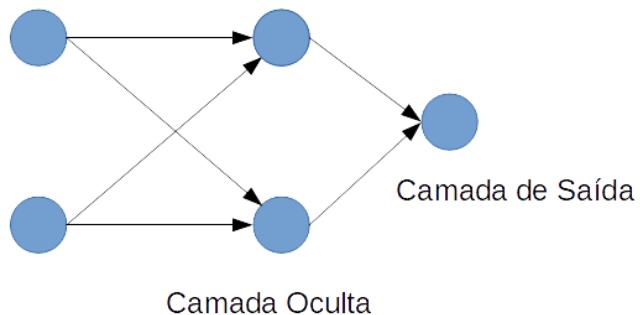
$$\delta = w_i + \alpha(d - S)x_i \quad (2.4)$$

A parcela $(d - S)x_i$ é a derivada da função de erro multiplicada por α que é a taxa de aprendizagem que regula o ajuste total da alteração do peso, proporcionando uma descida leve de gradiente, evitando que a alteração ultrapasse um mínimo local.

RNAs são capazes de aprender padrões através de impulsos de entrada, no entanto em sua forma mais básica, Perceptron (ROSENBLATT, 1958) não seria possível resolver problemas de natureza mais complexa dado seu caráter não linear. No entanto as arquiteturas com mais de um Perceptron e camadas ocultas (Figura 4) foram capazes de sanar esta limitação, solucionando problemas de natureza não linear.

Figura 4 – Ilustração de uma RNA MLP.

Camada de Entrada



Camada Oculta

Fonte: Elaborado pelo autor.

Este tipo de arquitetura é chamado de *Multilayer Perceptron* (MLP), onde a saída de cada neurônio fica assim responsável por "aprender" a reconhecer cada tipo de padrão de entrada e propagar este sinal para as camadas posteriores.

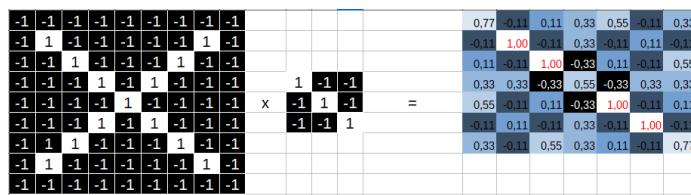
2.2 Redes Neurais Convolucionais

Como demonstrado na seção anterior, as MLPs são boas em reconhecer padrões, no entanto seu desempenho depende da qualidade dos dados de entrada da rede, estes dados precisam ter características bem definidas para que a rede possa discriminá-las bem cada classe, a fim de descobrir uma maneira mais eficaz de extrair características de dados brutos com grande poder discriminatório LeCun *et al.* (1998) introduziu então o conceito de *Convolutional Neural Networks* (CNN), o que viria nos anos seguintes revolucionar todo o campo de Machine Learning obtendo resultado de estado da arte em diversas aplicações de classificação de imagens como Pascal VOC Challenge (EVERINGHAM *et al.*, 2010). Baseado no conceito de convolução de um *kernel* sobre uma imagem, o resultado desta operação seria capaz de expressar as características denotadas por este *kernel* na imagem.

Por definição uma convolução é uma operação matemática em duas funções (f e g) produzindo uma terceira, em processamento de imagens esta operação consiste em uma resposta de um impulso expresso pela função *kernel* ou a localização na imagem da presença daquele padrão apresentado pelo *kernel*.

A Figura 5 demonstra a aplicação de um *kernel* que expressa característica de uma linha diagonal em uma imagem que representa um "X", o resultado é chamado de mapa de ativação e representa em quais lugares da imagem essa linha diagonal está presente. Para a definição de um filtro convolucional são definidos quatro parâmetros: tamanho do filtro, stride, zero-padding e número de filtros.

Figura 5 – Ilustração de uma convolução de um filtro sobre uma imagem.



Fonte: Elaborado pelo autor.

O stride é o tamanho do deslocamento que o *kernel* fará ao percorrer a imagem (geralmente 1), zero-padding é o preenchimento das bordas da imagem com zeros. A dimensão da saída de um filtro convolucional é dada por:

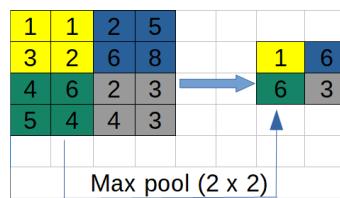
$$\frac{N + 2 * padding - F}{stride} + 1 \quad (2.5)$$

Onde N é o tamanho da entrada, F a dimensão do filtro. Exemplificando, para uma imagem de entrada de $32 \times 32 \times 3$ e 10 filtros de tamanho 5×5 obtemos:

$$\frac{2+2*1-5}{1+10} = 32, \quad 32 \times 32 \times 10 \text{ mapas de ativação.}$$

Além das camadas convolucionais são propostos outros dois elementos para CNN: pooling e camada completamente conectada. A camada de pooling consiste em aplicar uma função como média, moda, máximo, mínimo, dentre outras a fim de reduzir a dimensionalidade da entrada e adicionar uma pequena invariância a rotação e translação. No contexto das CNNs o mais comumente utilizado é o max-pooling, que dado um tamanho de *kernel* aplicado em uma imagem, o resultado da convolução é o valor máximo dos *pixels* de vizinhança (Figura 6). A camada completamente conectada é uma MLP, enquanto as camadas convolucionais tem o papel de aprender quais os padrões das entradas submetidas através dos filtros, a MLP tem o papel de classificar a entrada de acordo com os mapas de ativação resultantes.

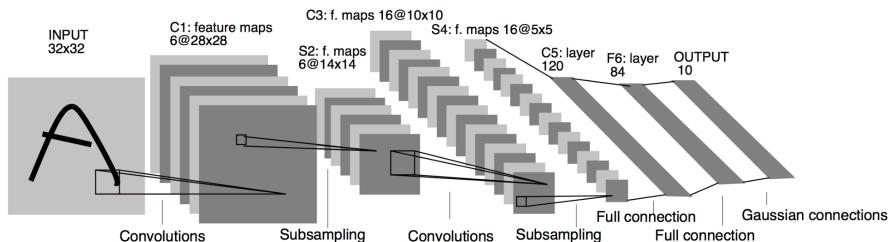
Figura 6 – Representação de um filtro max-pooling sobre uma imagem.



Fonte: Elaborado pelo autor.

A Figura 7 apresenta a arquitetura da LeNet proposta por LeCun *et al.* (1998), a imagem de entrada possui dimensão 32×32 , seguida por dois blocos de convolução e pooling, a MLP responsável pela classificação possui duas camadas de 120 e 80 neurônios cada.

Figura 7 – Arquitetura da LeNet.



Fonte: Adaptado de LeCun *et al.* (1998).

A fim de introduzir a não linearidade, é utilizada uma função de ativação para a saída de cada camada, esta função é a transformação não linear do sinal de entrada, o valor mapeado representa o nível de ativação de cada neurônio, caso o sinal deva ser propagado ou não pela rede. Dentre as funções de ativação as mais comuns são:

- Tangente hiperbólica: $\text{Tanh}(x) = \frac{e^{2x}-1}{e^{2x}+1}$

- ReLu (*Rectified Linear Units*): $f(x) = \max(0, x)$
- Sigmoidal: $f(x) = \frac{1}{1+e^{-x}}$
- Softmax: $\sigma(z)_j = \frac{e^{z_j}}{\sum_{i=1}^K e^{z_i}}$ para $j = 1, \dots, K$, onde z designa o vetor dos neurônios de saída.

Como visto na Seção 2.1 é preciso calcular o gradiente da função de erro, através da regra da cadeia calcula-se a derivada de cada função, portanto a função de ativação necessariamente deve ser derivável. A mais utilizada é a ReLu devido a sua simplicidade, pois para entradas menores que zero o sinal é simplesmente cancelado, o sinal não é propagado, tornando a rede esparsa e simplificando a computação. A função de ativação *Softmax* é um caso especial, ela é uma função Sigmoidal capaz de mapear as entradas para a probabilidade de pertencimento de cada classe do problema.

2.3 You Only Look Once v2

A tarefa de detecção de objetos difere um pouco da classificação tradicional de imagens pois refere-se a localizar as coordenadas de um ou mais objetos dentro de uma imagem. Neste sentido o *You Only Look Once* (*YoLo*) (REDMON *et al.*, 2016), do inglês: você só olha uma vez, se trata de um método de detecção de objetos que usa uma CNN, podendo ser trocada de acordo com a necessidade da aplicação. O nome deve-se ao fato de a imagem ser submetida apenas uma vez pela rede inteira para realizar a predição, diferentemente dos métodos baseados em *sliding window*¹ e *region proposal*².

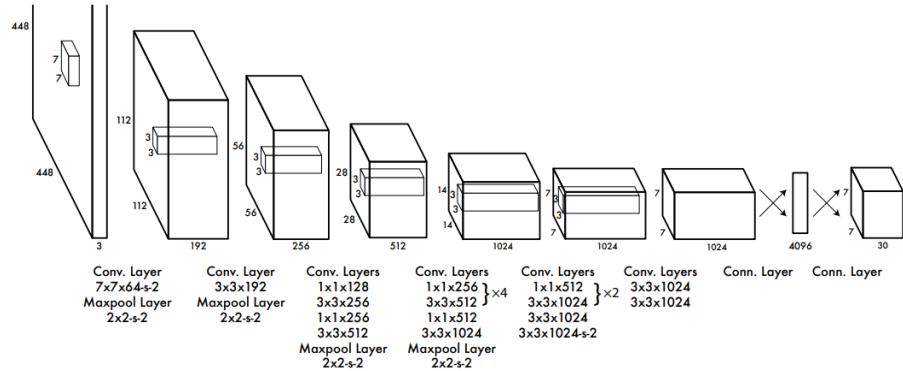
A YoLo destaca-se por inovar com três principais estratégias para utilizar uma CNN como detector invés de um simples classificador, sendo elas: (1) divisão da imagem em *grid*, (2) função de erro, (3) uso de âncoras. Sua arquitetura (Figura 8) é baseada na CNN *Inception* (SZEGEDY *et al.*, 2017), responsável por realizar a classificação, enquanto as camadas finais são substituídas para realizar a regressão das coordenadas do *bouding box* resultante.

O primeiro passo consiste em dividir a imagem em um *grid* de tamanho $S \times S$, segundo a proposta do artigo original este valor é fixado como 7, esta etapa é fundamental para a codificação da saída da rede. Cada célula deste *grid* é responsável por realizar a predição de B *bouding boxes*, por esta razão as camadas finais tem dimensão $7 \times 7 \times (B * 5 + C)$. Cada tensor de saída contém os valores: $p_c, b_x, b_y, b_h, b_w, c_1, c_2, \dots, c_n$, onde p_c indica se na célula S existe um objeto, os valores de b indicam as coordenadas do *bouding box* e c_n indica o *score* de confiança para determinada classe de objeto (C).

¹ Método onde utiliza-se uma estratégia de janela que desliza sobre a imagem *pixel a pixel* a fim de gerar entradas para um classificador decidir se trata-se ou não de um objeto.

² Métodos genéricos que utilizem qualquer estratégia para gerar regiões que sejam fortes candidatas a objetos para um classificador, como *Selective Search* (UIJLINGS *et al.*, 2013).

Figura 8 – Arquitetura da Yolo.



Fonte: Adaptado de Redmon *et al.* (2016)

O score de confiança é dado por:

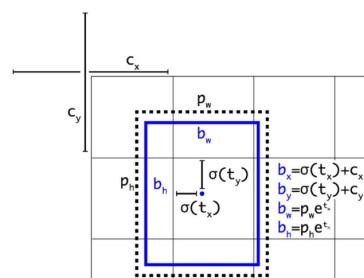
$$Pr(\text{Classe}_i | \text{Objeto}) * Pr(\text{Objeto}) * IOU_{pred}^{truth} = Pr(\text{Classe}_i) * IOU_{pred}^{truth} \quad (2.6)$$

Este score reflete tanto a probabilidade de haver um objeto quanto a qualidade do *bouding box* predito. Na proposta original a saída é mapeada para um tensor $7 \times 7 \times 30$, utilizando duas âncoras (dois objetos por célula S) e 20 classes de objetos.

A Figura 9 demonstra como são calculadas as coordenadas do *bouding box* a partir dos valores preditos e das âncoras pré estabelecidas, onde cada célula é responsável por detectar um objeto para da âncora, sendo t_x, t_y, t_w, t_h as coordenadas do *bouding box* em relação a célula S, c_x, c_y são os valores do tamanho da largura e da altura de cada célula S, $e^{t_w}, p_h e^{t_h}$ são os valores de largura e altura da âncora, calcula-se as coordenadas em relação a imagem por:

$$\begin{aligned} b_x &= \sigma(t_x) + c_x \\ b_y &= \sigma(t_y) + c_y \\ b_w &= p_w e^{t_w} \\ b_h &= p_h e^{t_h} \end{aligned} \quad (2.7)$$

Figura 9 – Cálculo baseado na âncora.



Fonte: Adaptado de Redmon *et al.* (2016)

Outra característica que diferencia a arquitetura da YoLo é a sua função de erro que é a soma quadrática composta pelo erro de classificação, erro de localização e erro de confiança. Se um objeto é detectado, o erro de classificação em cada célula é o erro quadrado das probabilidades condicionais de classe para cada classe:

$$\begin{aligned}
 & \sum_{i=1}^{S^2} \mathbb{1}_i^{obj} \sum_{c \in classes} (p_i(c) - \hat{p}_i(c))^2 \\
 & + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \\
 & + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] \\
 & + \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{obj} (C_i - \hat{C}_i)^2 \\
 & + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{noobj} (C_i - \hat{C}_i)^2
 \end{aligned} \tag{2.8}$$

onde:

- $\sum_{i=1}^{S^2}$ denota que temos que somar o erro para cada célula S do *grid*
- $\sum_{j=0}^B$ denota que temos que somar o erro para cada âncora B
- $\mathbb{1}_i^{obj}$ é 1 quando temo um objeto na célula S e a âncora B é responsável por este objeto, 0 caso contrário.
- $\mathbb{1}_{ij}^{noobj}$ é 1 quando não temos um objeto na célula S e a âncora B é responsável por este objeto, 0 caso contrário.
- $\mathbb{1}_i^{obj}$ é 1 quando há um objeto na célula S , 0 caso contrário.
- λ são as constantes que definem o peso de cada erro.
- Os termos p, x, y, w, h referem-se as coordenadas e valores de confiança preditas e as desejadas.

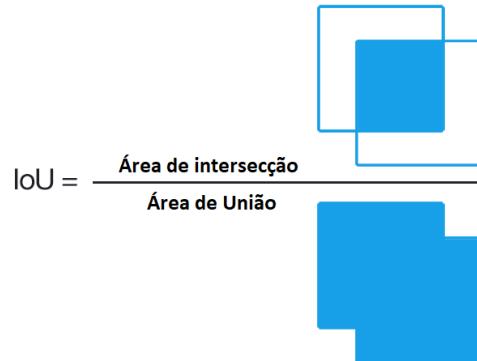
2.4 Métricas de Validação

Esta Seção aborda sobre duas métricas que são utilizadas na validação dos resultados deste trabalho, sendo uma delas padrão na validação de detecção de objetos e a outra é utilizada em trabalhos relacionados a detecção do disco óptico.

2.4.1 Intersect Over Union

Intersect Over Union (IoU) é a medida padrão e mais comum utilizada para avaliar a qualidade de detecção de um *bouding box*, na YoLo ela é utilizada juntamente com uma probabilidade condicional de um objeto em um *bouding box* para definir o score de confiança de saída da rede (Equação 2.6). Trata-se da área de intersecção dividida pela área de união entre o *bouding box* predito e o *bouding box* de ground truth, sua computação é melhor ilustrada através da Figura 10. Este tipo de métrica penaliza muito qualquer resultado que esteja minimamente diferente do *bouding box* de ground truth, portanto costuma-se considerar que valores acima de 0,5 sejam suficientes para classificar uma predição como correta, valores acima de 0,7 são considerados excelentes, próximos de 0,9 podem ser considerados resultados perfeitos para um observador humano.

Figura 10 – Computação da IoU.



Fonte: Elaborado pelo autor.

2.4.2 Métrica Padrão em Detecção do DO

No trabalho de Tobin *et al.* (2007) propõem-se um método de simples verificação de desempenho para algoritmos de detecção do DO, comparando a coordenada estimada pelo algoritmo com a marcação feita pelo especialista na imagem original através da seguinte desigualdade:

$$\sqrt{(x_{DO} - x_E)^2 + (y_{DO} - y_E)^2} \leq 1R \quad (2.9)$$

Onde (x_{DO}, y_{DO}) são as coordenadas do centro do DO detectado pelo algoritmo e (x_E, y_E) são as coordenadas do centro marcado pelo especialista e R é o raio do DO marcado pelo especialista. O resultado é dado como positivo caso a desigualdade seja verdadeira.

3 TRABALHOS RELACIONADOS

Neste capítulo abordaremos alguns trabalhos da literatura que tem relação com os objetivos deste estudo, a fim de estabelecer parâmetros de qualidade para os resultados obtidos, além de apresentar abordagens anteriores para resolução dos problemas propostos.

3.1 Métodos que Utilizam Processamento de Imagens

O trabalho de Silva *et al.* (2016) visa detectar o DO através da estratégia de comitê de algoritmos com votação ponderada, essa técnica consiste em analisar o resultado de diversos algoritmos e definir dentre eles qual pode se aproximar mais do melhor resultado. Silva *et al.* (2016) aponta três principais abordagens que os principais trabalhos da literatura tendem a seguir, além de suas problemáticas, sendo elas: (1) baseado em características locais como o maior nível de intensidade dos *pixels* na região do DO, mas este método tende a funcionar bem apenas em imagens de retinas saudáveis; (2) baseado em casamento de padrões, busca o DO baseado em propriedades de cor, textura e forma, procurando na imagem a região mais similar a esse modelo, no entanto tende a não funcionar em imagens com grandes lesões; (3) abordagem geográfica baseando-se na informação dos vasos, visto que eles se originam no DO, no entanto é uma operação de difícil realização. Silva *et al.* (2016) propôs então um comitê utilizando 5 metodologias que serão descritas no decorrer desta Seção.

Akram *et al.* (2010) é o primeiro método do comitê, consiste na aplicação de um filtro da média no canal G do modelo de cores RGB, a região de interesse do DO é considerada a parte de maiores níveis de intensidade da imagem, então aplica-se uma transformada circular de Hough (DUDA; HART, 1971). Os testes foram feitos nas bases DRIVE (STAAL *et al.*, 2004), DIARETDB0 (KAUPPI *et al.*, 2006) e DIARETDB1 (KÄLVIÄINEN; UUSITALO, 2007). A avaliação de sucesso ou erro foi feita por observação de especialista, obtendo acurácia de 96,7%.

Rajaput, Reshma e Sidramappa (2011) aplicaram uma equalização de histograma no canal vermelho do modelo de cores RGB, imagem é invertida e as áreas de mínima intensidade são identificadas utilizando a transformada H. Foi definido empiricamente o valor de h (altura limite) como 20 *pixels*. Obtém-se uma imagem binária com os *pixels* brancos representando as regiões de mínima intensidade da imagem original, as regiões de mínimo são *pixels* conectados e que possuem o mesmo valor de intensidade cujos *pixels* da borda externa possuem um valor de intensidade mais alto. Para eliminar falsas regiões localizadas, os autores aplicaram uma abertura morfológica com um elemento estruturante no formato de disco. Por fim, a média de intensidades de cor é calculada para identificar e selecionar as regiões com menor média de intensidade como a região do DO. Os testes foram realizados em 33 imagens da base DRIVE, porém, não há relato no trabalho sobre a quantidade de erros e acertos do método de detecção do DO.

Sekar e Nagarajan (2012), propõem-se um método de localização do DO com o uso

de histograma e técnicas de agrupamento. Primeiramente determina-se as regiões mais claras agrupando-se os *pixels* mais claros da banda vermelha da imagem. Dentre as regiões, três são selecionadas, agora no canal verde e utilizando três métodos: método da máxima diferença; método da máxima variância; e, método do filtro Gaussiano passa baixa. Baseado no histograma na banda azul dos três candidatos selecionados, seleciona-se uma sub-janela em cada um dos três candidatos. O centro da sub-janela que possuir o maior número de *pixels* claros é escolhido como centro do DO. O sucesso foi de 99,50%, no entanto os autores não descreveram a metodologia usada para calcular este resultado.

Em Punnolil (2013) o centro do DO é detectado utilizando as informações das arcadas vasculares superiores e inferiores da retina. Para detecção do DO, utilizou-se um operador de fechamento morfológico no canal verde da imagem com um elemento estruturante octogonal, eliminando assim os vasos da região do DO. Após, utilizou-se um operador de vizinhança em coluna e a imagem resultante foi binarizada, por fim, o centro da região resultante da imagem binarizada é dado como o centro do DO. Os testes foram realizados nas bases DRIVE, STARE e DIARETDB0, totalizando 251 imagens, no entanto não há informações no trabalho sobre a sua acurácia.

Zubair, Yamin e Khan (2013) aplicaram uma equalização de histograma (GONZALEZ, 2009) no canal verde da imagem RGB, seguida de uma operação morfológica de erosão para remover os candidatos que não são DO, ao fim aplica-se uma dilatação morfológica com elemento estruturante próximo ao tamanho do DO, com os candidatos obtidos escolhe-se como DO aquele que obtiver maior contraste em relação aos outros. Os testes foram realizados na base MESSIDOR e os autores relataram 96,65% de acurácia.

A Tabela 1 contém as bases de imagens utilizadas por Silva *et al.* (2016), a base STARE possui 397 imagens, no entanto utilizaram apenas 103 destas.

Tabela 1 – Bases de imagens utilizadas por Silva *et al.* (2016)

Base	Quantidade
Messidor	1200
STARE	103
Drishti-GS	101
DRIONS	110
ARIA	116
Total	1630

Fonte: Adaptado de Silva *et al.* (2016).

A avaliação dos pesos dos algoritmos do comitê proposto pelo autor consistem em calcular a taxa de sucesso para cada um dos cinco métodos, selecionando uma porcentagem das imagens aleatoriamente para treinamento, o experimento foi realizado com dois grupos de teste, um com 20% e outro com 40% das imagens de cada base. A taxa de sucesso T_i é determinada

por:

$$T_i = \frac{\text{Total de sucessos}}{\text{Total de falhas}} \quad (3.1)$$

Silva *et al.* (2016) utilizou a métrica descrita na Seção 2.4.2, usando R como sendo o raio máximo de cada base. A taxa de sucesso (T_i) foi adotada como peso de cada algoritmo i . A acurácia final foi avaliada como sendo a média dos resultados nos dois testes realizados com 20% e 40% das imagens.

O método de votação para detecção do DO funciona em 4 principais passos:

- Uma janela W_i quadrada com lado de comprimento igual ao raio do DO é centralizada em cada *pixel* da imagem
- As saídas dos algoritmos que tenham detectado DO nesta janela, são computadas em TS_w .

$$TS_w = \sum_i^{\forall i \in W} T_i \quad (3.2)$$

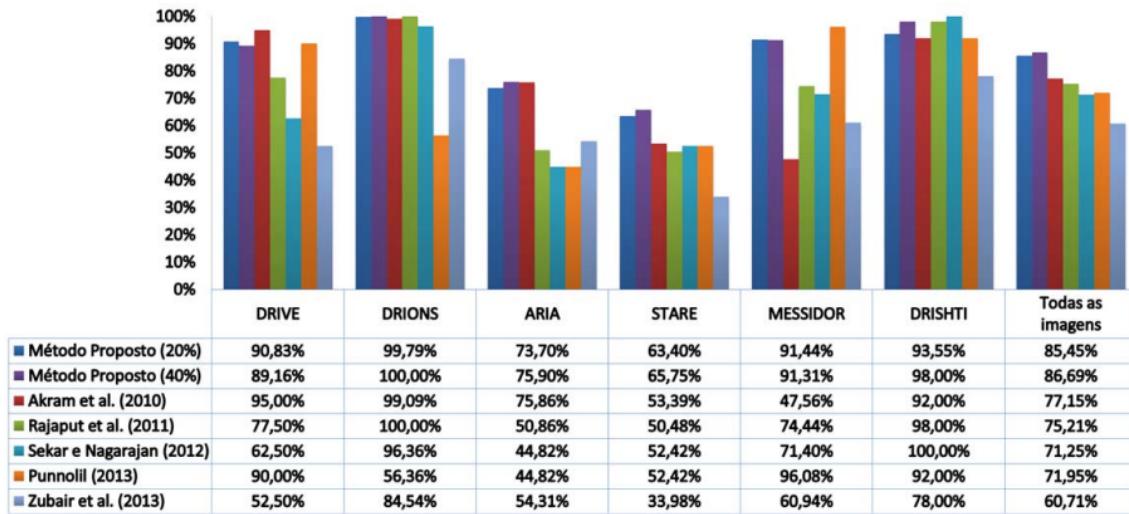
- Em caso de empate escolhe-se a região que obtiver marcação do algoritmo com melhor taxa de sucesso é escolhido.
- Calcular o centro do DO através da equação:

$$x' = \frac{\sum_i^{\forall i \in W} T_i x_i}{\sum_{j=1}^5 T_j} \quad (3.3)$$

$$y' = \frac{\sum_i^{\forall i \in w} T_i y_i}{\sum_{j=1}^5 T_j} \quad (3.4)$$

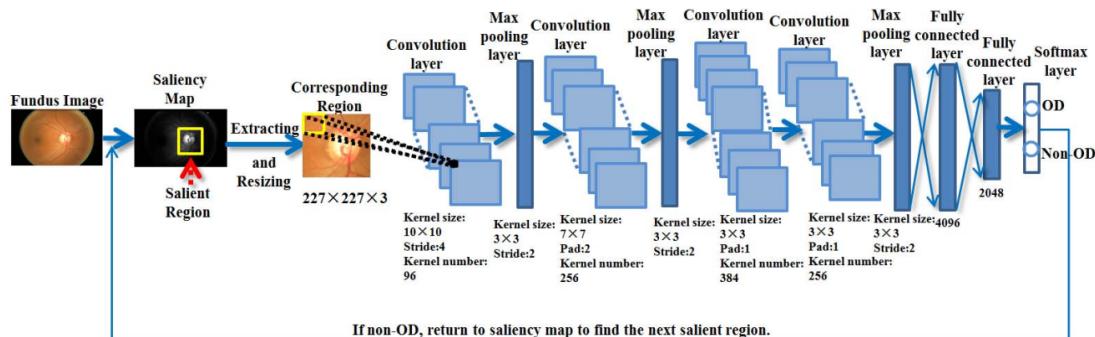
O Figura 11 demonstra os resultados do método aplicado comparando com os cinco algoritmos propostos na utilização do comitê. Observa-se que o resultado obtido por ele é superior aos outros algoritmos apenas em duas bases.

Através da análise destes resultados podemos observar que obtiveram um bom acerto na maioria das bases, 89,09% como média geral. No entanto nota-se que o desempenho foi insatisfatório em pelo menos duas bases, isso demonstra que o método não é capaz de obter bons acertos em diferentes bases.

Figura 11 – Resultados obtidos por Silva *et al.* (2016) comparados com outros estudos.Fonte: Adaptado de Silva *et al.* (2016)

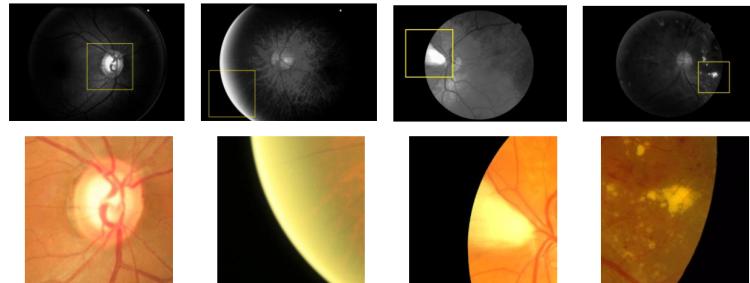
3.2 Métodos que Utilizam CNN

Além das técnicas clássicas de processamento de imagens há trabalhos que utilizam redes neurais convolucionais para detecção da região do DO, o trabalho de Niu *et al.* (2017) propõe uma método que integra uma seleção de regiões prévias chamado de *Saliency Map* (ACHANTA *et al.*, 2009) a uma CNN para classificar estas regiões em DO ou não DO. Esta técnica é capaz de tornar um objeto com maiores níveis de intensidade mais destacado de acordo com as características da visão humana, com alta taxa de contraste, em situações ideais podendo detectar a região do DO, mas tende a falhar quando a imagem submetida houver ocorrência de patologias como retinopatia diabética. A Figura 12 demonstra a arquitetura da CNN utilizada além do fluxo de sua metodologia.

Figura 12 – Arquitetura de Niu *et al.* (2017).Fonte: Adaptado de Niu *et al.* (2017).

A Figura 13 demonstra o resultado do *Saliency Map* para quatro imagens diferentes, sendo que três apresentam algum padrão de anormalidade, o que justifica o uso da CNN para classificação dessas regiões.

Figura 13 – Resultados do Algoritmo *Saliency Map* em Niu *et al.* (2017).



Fonte: Adaptado de Niu *et al.* (2017)

Niu *et al.* (2017) não especifica se a arquitetura de rede foi inspirada em alguma outra clássica, no entanto para validação dos seus resultados foram comparados substituindo esta rede convolucional pela Alexnet (Krizhevsky; Sutskever; Hinton, 2012). Os resultados comparativos obtidos estão descritas na Tabela 2. Para validação foram usadas as bases ORIGA, separando 500 imagens para treino e 150 para teste, e MESSIDOR com 800 imagens para treino e 400 no teste. O no entanto não informa qual metodologia foi utilizada para validar se uma região foi marcada corretamente ou não, ou até mesmo avaliar o quanto precisa ser esta marcação. A Figura 14 demonstra alguns dos seus resultados positivos, onde pode-se observar que o *bouding box* extrapola a região do DO o que certamente acarretaria em um baixo IoU.

Figura 14 – Resultados positivos de Niu *et al.* (2017).



Fonte: Adaptado de Niu *et al.* (2017).

Tabela 2 – Comparaçao dos resultados das duas arquiteturas em Niu *et al.* (2017).

	ORIGA	MESSIDOR	Todas imagens
Alexnet	99,33%	98,25%	98,79%
Niu <i>et al.</i> (2017)	99,33%	98,75%	99,04%

Fonte: Adaptado de Niu *et al.* (2017).

A proposta de Alghamdi *et al.* (2016) também utiliza uma CNN em dois estágios para localizar a região do DO e classificá-la em normal, suspeita ou anormal. Sua abordagem propõem um pré-processamento da imagem antes do treinamento que consiste em normalizar a imagem

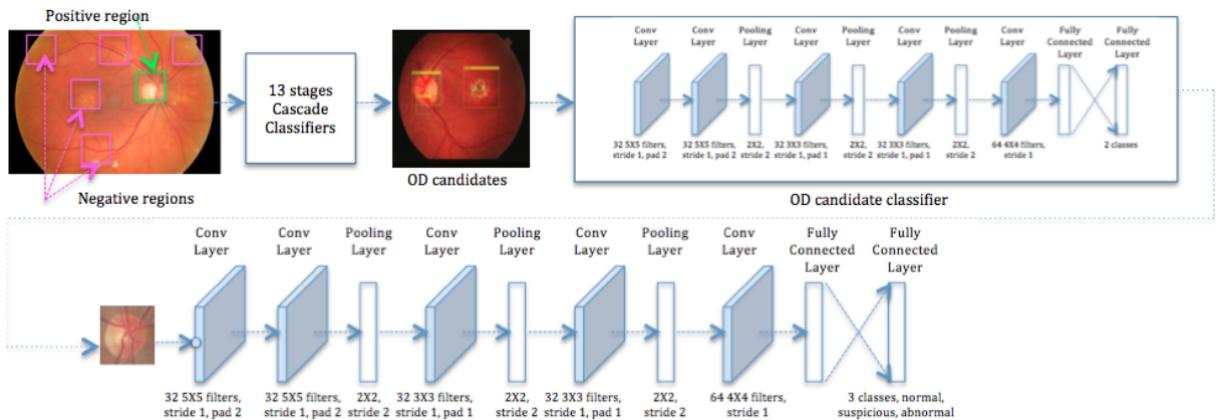
subtraindo uma imagem média da original e dividindo-a pelo desvio padrão e após isso aplica uma técnica de branqueamento não especificada, a fim de normalizar características de escala tornando a modelagem mais fácil. A etapa de pré processamento pode ser visualizada na Figura 15. Para gerar as regiões candidatas a DO o Alghamdi *et al.* (2016) utilizam o detector de objetos Viola e Jones (2001), as regiões geradas são submetidas a primeira CNN que classifica em OD e não OD, uma segunda CNN classifica em normal, suspeito e anormal. O fluxo de seu método pode ser observado na Figura 16.

Figura 15 – (a) Região original do DO. (b) Imagem normalizada de (a). (c) Branqueamento de (b). (d) Não região do DO. (e) Imagem normalizada de (d). (f) Branqueamento de (e). (g) Imagem média.



Fonte: Adaptado de: Alghamdi *et al.* (2016).

Figura 16 – Arquitetura proposta por Alghamdi *et al.* (2016).



Fonte: Adaptado de Alghamdi *et al.* (2016).

Para validação dos resultados foram utilizadas as bases DRIVE, STARE, DIARETDB1 e MESSIDOR públicas, enquanto as bases HAPIEE, KENYA e PAMDI são coletados de estudos de base populacional em colaboração com o Moorfields Eye Hospital Reading Centre, totalizando 5781 imagens. O autor não especifica como avaliou se uma marcação foi realizada corretamente, seus resultados obtidos estão descritos na Tabela 3.

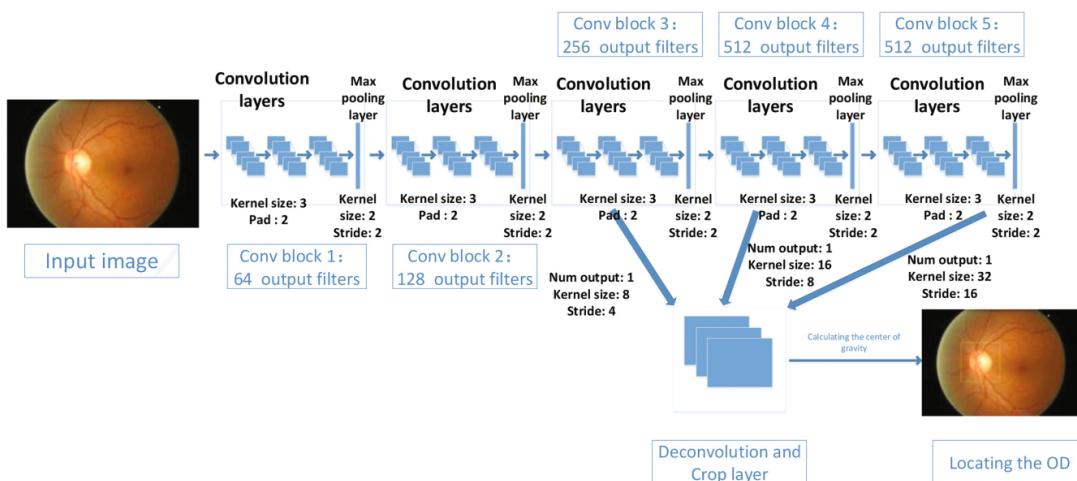
O autor não revela quantas imagens foram utilizadas para treino ou teste, no entanto se considerarmos o acerto de cada base proporcionalmente ao total de imagens a acurácia final é de 97,86% ou considerando uma média simples do resultado total a acurácia final seria de 96,60%.

Tabela 3 – Resultados de Alghamdi *et al.* (2016).

Base	Imagens	Acurácia
DRIVE	40	100,00%
DIARETDB1	89	98,88%
MESSIDOR	1200	99,20%
STARE	402	86,71%
KENYA	1125	99,53%
HAPPIE	1951	98,36%
PAMDI	907	98,13%
KSSH	67	92,00%
Total	5781	

Fonte: Adaptado de Alghamdi *et al.* (2016).

Xu *et al.* (2017) utiliza uma estratégia diferente de CNN para detectar o DO, dentre os pesquisados, seu método é o único que propõem uma classificação *pixel a pixel* da imagem, o que intuitivamente poderia resultar alto custo computacional no entanto o autor relata um tempo de processamento de até 0,93 segundos por imagem. Isto deve-se a uma etapa de pré processamento da entrada que consiste em: no conjunto de treinamento as imagens são subtraídas pela média e então a imagem é redimensionada para 400×600 . A imagem é submetida por inteiro a uma versão modificada da VGG que consiste na remoção das camadas completamente conectadas, as saídas dos três últimos blocos convolucionais são submetidas a um bloco de-deconvolucionar de corte (Figura 17). A última camada da rede tem como saída a probabilidade de cada *pixel* pertencer a classe DO ou não DO, calcula-se então o centro de gravidade de todos os *pixels* com probabilidade maior que 0,9. O DO é centralizado neste centro de gravidade e o *bouding box* é definido como 1/4 da imagem. Seus resultados obtidos estão descritos na Tabela 4.

Figura 17 – Arquitetura proposta por Xu *et al.* (2017).

Fonte: Adaptado de Xu *et al.* (2017).

Tabela 4 – Resultados de Xu *et al.* (2017).

Base	Acurácia
MESSIDOR	99,43%
STARE	89,30%
ORIGA	89,00%
Todas	96,30%

Fonte: Adaptado de Xu *et al.* (2017).

Através destes trabalhos apresentados podemos observar que de maneira geral os métodos que utilizam CNNs tem desempenho superior, pois todos apresentam resultados superiores a 90% de acerto enquanto os que utilizam técnicas clássicas de processamento de imagens tem acertos que variam entre 60% a 86%, são muito dependentes de parâmetros e por isso tendem a não funcionar bem em imagens de bases que não foram utilizadas nos testes, este problema tende a não ocorrer nos métodos que utilizam CNN visto seu maior poder de generalização, no entanto estes são dependentes de muitas imagens para poder fazer esta generalização, o que pode nem sempre estar disponível.

4 METODOLOGIA

Este capítulo apresenta os materiais e método que foram utilizados para o desenvolvimento deste trabalho, como bases de imagens, software e método proposto.

4.1 Bases Utilizadas

Esta seção apresenta as bases utilizadas para a validação da metodologia proposta, assim como suas características. A Tabela 5 contém as bases utilizadas e quantas imagens cada uma possui.

Tabela 5 – Bases de imagens utilizadas neste trabalho.

Base	Quantidade
Alipour Dataset (ALIPOUR <i>et al.</i> , 2012)	60
DIARETDB0 (KAUPPI <i>et al.</i> , 2006)	130
DIARETDB1 (KÄLVIÄINEN; UUSITALO, 2007)	89
DRIONS (CARMONA <i>et al.</i> , 2008)	110
Drishti-GS (SIVASWAMY <i>et al.</i> , 2014)	101
DRIVE (STAAL <i>et al.</i> , 2004)	40
HRF (ODSTRCLIK <i>et al.</i> , 2013)	45
Messidor (DECENCIÈRE <i>et al.</i> , 2014)	1200
Rim One (FUMERO <i>et al.</i> , 2011)	169
STARE (HOOVER; KOUZNETSOVA; GOLDBAUM, 2000)	397
JSIEC Dataset (LINCHUAN, 2018)	1087
Total	3428

Fonte: Elaborado pelo autor.

4.1.1 Rim-One

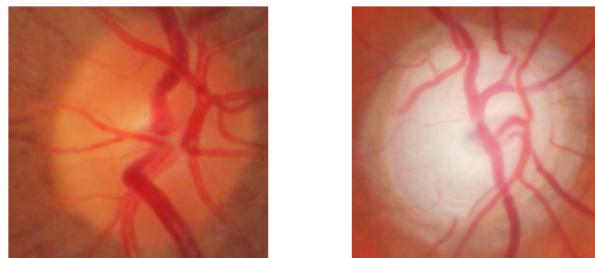
Esta base é composta por 169 imagens ONH (*Optic Nerve Head*) obtidas a partir de 169 imagens de fundo de olho diferentes. O ONH é a parte mais identificável da retina após o corte da imagem com a região de interesse. As retinografias foram capturados em três hospitais espanhóis: Hospital Universitario de Canarias, Hospital Clínico San Carlos e Hospital Universitario Miguel Servet. Foram compiladas imagens de diferentes fontes médicas a fim de garantir um conjunto de imagens heterogêneo.

A câmera usada para capturar esses imagens é uma câmera de fundo Nidek AFC-210 com um corpo de uma Canon EOS 5D Mark II de 21,1 megapixels. A Figura 18 demonstra dois exemplos de DO em imagens desta base. As imagens da base foram divididas nos seguintes subconjuntos pelos especialistas:

- Olho normal (não glaucomatoso): 118 imagens.

- Glaucoma moderado: 14 imagens.
- Glaucoma precoce: 12 imagens.
- Glaucoma profundo: 14 imagens.
- Hipertensão Ocular (OHT): 11 imagens.

Figura 18 – A esquerda uma imagem normal, a direita uma imagem de paciente com glaucoma.



Fonte: Adaptado de Fumero *et al.* (2011).

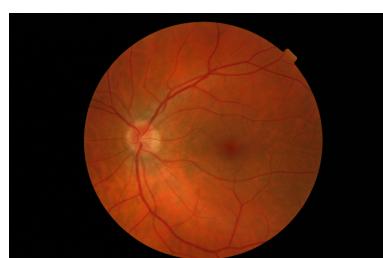
Esta base foi criada através da colaboração de cinco especialistas no domínio do glaucoma desses hospitais, sendo quatro oftalmologistas e um optometrista. Estes escolheram pacientes para amostras de glaucoma e os voluntários para amostras de olhos saudáveis e a imagem final com seu diagnóstico associado (FUMERO *et al.*, 2011).

4.1.2 MESSIDOR

As 1200 imagens da base Messidor foram adquiridas por 3 departamentos oftalmológicos usando uma câmera 3CCD de vídeo colorido em um retinógrafo não midriático TRC NW6 da Topcon com um campo de visão de 45°. As imagens foram capturadas usando 8 bits por plano de cor em 1440×960 , 2240×1488 ou 2304×1536 pixels (DECENCIÈRE *et al.*, 2014).

A base MESSIDOR destaca-se das restantes principalmente por sua grande variedade de tipos de imagens, há diversas imagens repletas de exsudados, o que torna a tarefa de detecção do DO mais difícil devido ao alto índice de falso positivo nessa imagens, a exemplo a Figura 19

Figura 19 – Imagem da base MESSIDOR.



Fonte: Decencière *et al.* (2014)

4.1.3 STARE

Iniciado nos EUA, nas universidades da Califórnia e San Diego em 1975 o projeto STARE (STructured Analysis of the Retina). Esta base possui 402 imagens capturadas com uma câmera TopCon TRV-50 com uma angulação de 35º e dados clínicos que foram fornecidos pelo Shiley Eye Center, na Universidade da Califórnia, em San Diego, e pelo Veterans Administration Medical Center, em San Diego (HOOVER; KOUZNETSOVA; GOLDBAUM, 2000).

O projeto contém diagnósticos para as imagens, além da marcação dos vasos sanguíneos para 40 imagens e marcação do centro do DO para 80 imagens. A Figura 20 é exemplo de uma das imagens desta base.

Figura 20 – Imagem da base STARE.

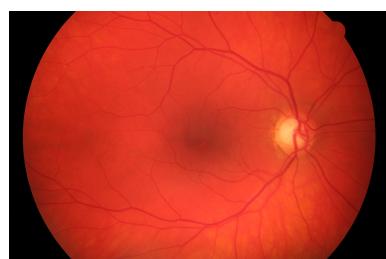


Fonte: Hoover, Kouznetsova e Goldbaum (2000).

4.1.4 HRF

O base de imagens pública HRF (High Resolution Fundus) contém imagens de pacientes saudáveis, 15 imagens de pacientes com retinopatia diabética e 15 imagens de pacientes glaucomatosos. São disponibilizadas também imagens binárias com a marcação dos vasos sanguíneos feitas por especialista. As imagens foram capturadas com uma câmera de fundo Canon CR-1 com um campo de visão de 45º e diferentes configurações de aquisição. A Figura 21 é uma das imagens com glaucoma da base HRF de dimensões 3544×2336 (ODSTRCILIK *et al.*, 2013).

Figura 21 – Imagem da base HRF.



Fonte: Odstrcilik *et al.* (2013).

4.1.5 DRIVE

Esta base foi criada a partir de um programa de rastreio de retinopatia diabética na Holanda, 400 indivíduos entre 25 e 90 anos de idade foram avaliados, no entanto apenas 40 imagens foram escolhidas, dentre estas 33 são normais e 7 possuem sinais de retinopatia diabética (Figura 22). As imagens foram adquiridas usando uma câmera Cannon 3CCD não midriática com campo de visão de 45º (STAAL *et al.*, 2004).

A base foi dividida em 20 imagens de treino e 20 de teste, para cada uma delas há marcações manuais dos vasos sanguíneos feitas por especialistas que foram treinados por um oftalmologista experiente.

Figura 22 – Imagem da base Drive.



Fonte: Staal *et al.* (2004).

4.1.6 Drishti

A base Drishti-GS, composta por 101 imagens e diagnóstico adquiridas no Aravind Eye Hospital de Madurai, na Índia. Estas possuem marcações do DO e da CDO realizadas por 4 oftalmologistas com diferentes níveis de experiência clínica. Todas as imagens foram adquiridas com DO centrado, com campo de visão de 30 graus e resolução máxima de 2040 × 1760, os pacientes tinham idade entre 40 e 80 anos com número aproximado entre homens e mulheres (SIVASWAMY *et al.*, 2014).

4.1.7 Drions

Esta base possui 110 imagens coloridas da retina coletadas no Serviço de Oftalmologia do Hospital Miguel Servet, Saragoça (Espanha). A média de idade dos pacientes foi de 53 anos sendo 46,2% do sexo masculino e 53,8% do sexo feminino e todos eram de etnia caucasiana. 23,1% dos pacientes tinham glaucoma simples crônico e 76,9% hipertensão ocular. As imagens foram adquiridas com uma câmara de fundo de olho analógica colorida, aproximadamente centrada na ONH e armazenadas em formato de slide. Para ter as imagens em formato digital,

Figura 23 – Imagem da base Drishti.



Fonte: Sivaswamy *et al.* (2014)

elas foram digitalizadas usando um scanner de alta resolução HP-PhotoSmart-S20, formato RGB, resolução de 600x400 e 8 bits / pixel (CARMONA *et al.*, 2008). As imagens desta base possuem a menor resolução dentre as usadas neste trabalho, a exemplo a Figura 24.

Figura 24 – Imagem da base Drions.



Fonte: Carmona *et al.* (2008).

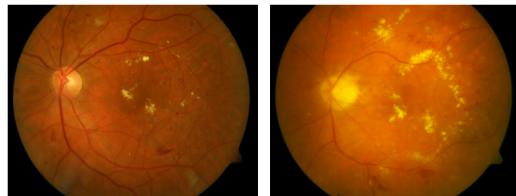
4.1.8 DIARETDB0 E DIARETDB1

Estas bases pertencem a um banco de dados público para o *benchmarking* de detecção de retinopatia diabética a partir de imagens de fundo de olho. O principal objetivo do projeto foi definir um banco de dados e um protocolo de teste que pode ser usado para avaliar os métodos de detecção de retinopatia diabética. Entretanto estas imagens podem também ser utilizadas para detecção do DO, pois na maioria das imagens este está visível, além de serem imagens de boa qualidade.

As imagens foram capturadas usando a mesma câmera de fundo digital de campo de visão de 50 graus com configurações variadas de imagens. A base 1 consiste em 89 imagens de fundo de cor, das quais 84 contêm pelo menos sinais leves de retinopatia diabética, e 5 são considerados normais, que não contêm sinais de retinopatia diabética segundo todos os especialistas que participaram a avaliação. Já a base 0 contém 130 imagens, das quais 20 são normais e 110 contêm sinais da retinopatia diabética (exsudatos duros, exsudatos moles, microaneurismas, hemorragias e neovascularização) (KAUPPI *et al.*, 2006; KÄLVIÄINEN; UUSITALO, 2007).

A Figura 25 demonstra duas imagens, a esquerda da DIARETDB0 e a esquerda DIARETDB1 a direita, nota-se que possuem um bom número de incidências de exsudatos

Figura 25 – Imagem da base Drions.



Fonte: Kauppi *et al.* (2006), Kälviäinen e Uusitalo (2007)

4.1.9 Alipour Dataset

Esta base é utilizada no trabalho de Alipour *et al.* (2012). Contendo 75 imagens, sendo 45 normais e 30 anormais, no entanto no link disponibilizado para download só há 25 imagens normais e 35 anormais. A base é utilizada para análise da zona avascular foveal para graduação da gravidade de retinopatia com base na Transformada de Wavelet. As imagens possuem resolução de 720×576 e possuem DO parcialmente visível em algumas destas imagens (Figura 26).

Figura 26 – Imagem da base de Alipour *et al.* (2012).



Fonte: Alipour *et al.* (2012).

4.1.10 JSIEC Dataset

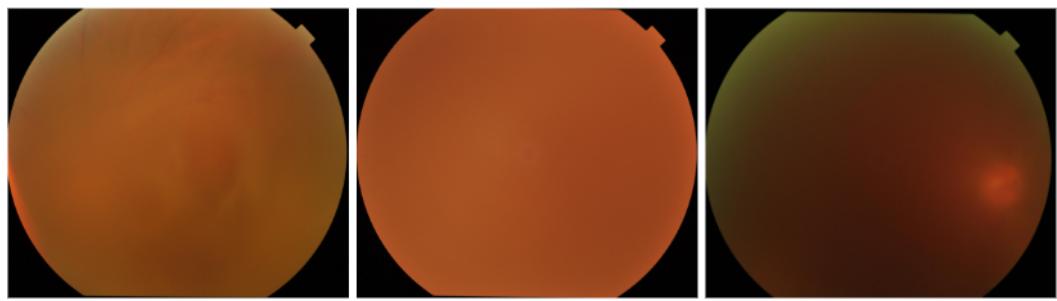
Esta base possui 1087 imagens de fundo de olho divididas em 37 categorias e são provenientes do Centro de Visão Internacional Shantou Conjunto (JSIEC), localizado na cidade de Shantou, província de Guangdong, China. Esta base é um subconjunto do total de 209494 imagens. Os direitos autorais dessas imagens pertencem a JSIEC e podem ser usados livremente para qualquer finalidade.

Estas imagens foram postadas por um usuário chamado Linchuan em uma plataforma voltada ao compartilhamento de datasets e competições de Machine Learning chamada Kaggle

(LINCHUAN, 2018), no entanto estas foram as únicas informações dadas por este usuário sobre a base e não foram encontradas mais informações sobre esta na literatura.

Dentre as categorias das imagens há pelo menos 2 com imagens extremamente borradass ou turvas, como na Figura 27, além de 25 imagens com DO não visível 27, no entanto optou-se por utilizar estas imagens a fim de melhorar o desempenho da rede para testar sua generalização quanto a verdadeiros positivos.

Figura 27 – Imagem da base de JSIEC.



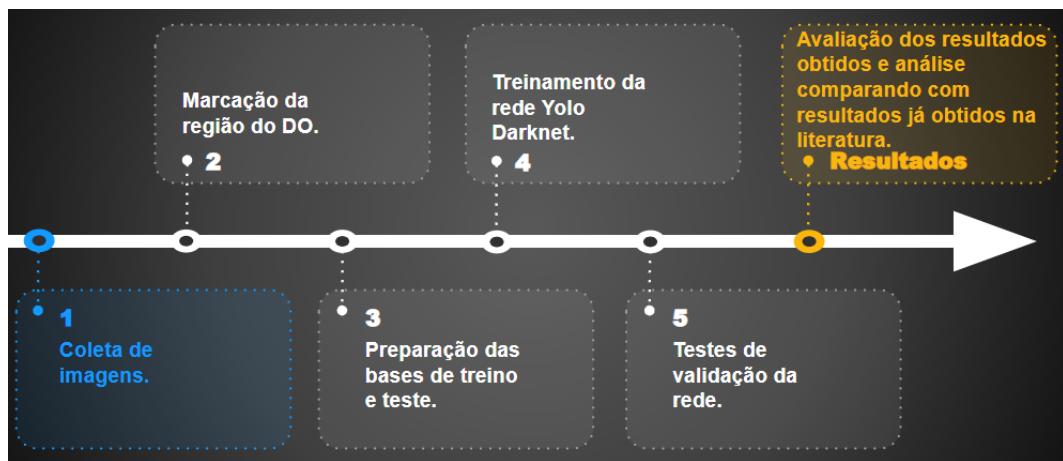
Fonte: Linchuan (2018).

4.2 Método Proposto

Esta seção apresenta o método proposto para alcançar os objetivos deste trabalho. A Figura 28 resume as etapas adotadas.

A primeira etapa refere-se a pesquisa bibliográfica referente as bases de imagens públicas disponíveis as quais já foram abordadas na Seção 4.1

Figura 28 – Fluxo da Metodologia.



Fonte: Elaborado pelo autor.

4.2.1 Marcação da Região do DO

Dentre as bases selecionadas para o estudo apenas três possuem marcação do DO (Rim One, Drishti e DRIONS), somando 380 imagens. A fim de ampliar o número total de imagens utilizadas por este trabalho utilizou-se o software de marcação de *bouding boxes* Yolo Mark (AB, 2016) para marcar a região do DO nas imagens das bases restantes.

Para fazer as anotações no formato requerido pela implementação da rede utilizada, colocam-se as imagens em um diretório do software, a partir daí o mesmo consegue identificar as imagens que serão marcadas, então para cada imagem que é aberta pelo software cria-se um arquivo no formato .txt com o mesmo nome da imagem que está sendo marcada. Para cada marcação realizada é escrita uma linha contendo as coordenadas relativas do *bouding box* selecionado no seguinte formato:

$$\text{classe } x_{cbb} \text{ } y_{cbb} \text{ } altura_{bb} \text{ } largura_{bb}$$

onde classe refere-se a classe do objeto marcado, para este fim utiliza-se apenas uma classe (0), já a variável (x_{cbb}, y_{cbb}) refere-se as coordenadas do valor relativo do centro do *bouding box* em relação ao tamanho da imagem, este valor é dado por:

$$x_{cbb} = \frac{x_{absoluto}}{\text{largura_imagem}} \quad (4.1)$$

$$y_{cbb} = \frac{y_{absoluto}}{\text{altura_imagem}} \quad (4.2)$$

$$\text{largura}_{bb} = \frac{\text{largura_absoluta}}{\text{largura_imagem}} \quad (4.3)$$

$$\text{altura}_{bb} = \frac{\text{altura_absoluta}}{\text{altura_imagem}} \quad (4.4)$$

onde os valores absolutos são os valores das coordenadas do *bouding box*. Ao todo foram marcadas 3428 imagens, tanto em imagens que continham ou não o DO no campo de visão.

4.2.2 Treinamento da Rede

Para realização do treinamento nos dois experimentos foi utilizada a rede YoLo v2 (Seção 2.3). A arquitetura da CNN utilizada foi a proposta no trabalho original Redmon *et al.* (2016), já mencionada na Figura 8. A Tabela 6 descreve a arquitetura utilizada com todas as suas camadas, os blocos convolucionais seguidos de Maxpool são inspirados na arquitetura da CNN *Inception* (SZEGEDY *et al.*, 2017), utilizando o conceito de *Network in Network* (LIN; CHEN; YAN, 2013).

Tabela 6 – Arquitetura da CNN utilizada.

Tipo	Filtro	Tamanho/Stride	Saída
Convolucional	32	3×3	224×224
Maxpool		2×2	112×112
Convolucional	64	3×3	112×112
Maxpool		2×2	56×56
Convolucional	128	3×3	56×56
Convolucional	64	1×1	56×56
Convolucional	128	3×3	56×56
Maxpool		2×2	28×28
Convolucional	256	3×3	28×28
Convolucional	128	1×1	28×28
Convolucional	256	3×3	28×28
Maxpool		2×2	14×14
Convolucional	512	3×3	14×14
Convolucional	256	1×1	14×14
Convolucional	512	3×3	14×14
Convolucional	256	1×1	14×14
Convolucional	512	3×3	14×14
Maxpool		2×2	7×7
Convolucional	1024	3×3	7×7
Convolucional	512	3×3	7×7
Convolucional	1024	3×3	7×7
Convolucional	512	1×1	7×7
Convolucional	1024	3×3	7×7
Convolucional	1000	1×1	7×7
Avgpool		Global	1000
Softmax			

Fonte: Adaptado de (REDMON *et al.*, 2016).

Como este trabalho é baseado na detecção de apenas um objeto foram realizadas algumas alterações na arquitetura a fim de otimizar o processo de treinamento. As âncoras foram alteradas para refletir os tamanhos próximos do DO. O tensor de saída foi alterado para as dimensões $7 \times 7 \times 30$, isto justifica-se pela imagem ser dividida em um grid de tamanho 7×7 , como saída da rede, para cada célula do grid há o resultado da predição da localização do *bouding box*, se a âncora for responsável pela detecção do objeto e a confiança dessa detecção: $(p_{obj}, x_c, y_c, w, h, c) * 5$ âncoras.

A taxa de aprendizado foi configurada como 10^{-3} , as imagens de entrada foram redimensionadas para o tamanho 416×416 , o batch de treinamento foi de 64 imagens. Durante o treinamento são utilizadas algumas técnicas de *data augmentation*¹ como cortes randômicos, rotações, mudanças de matiz, saturação, exposição.

¹ Técnica para realizar o aumento de bases de dados, no contexto de visão computacional serve para introduzir exemplos com diferentes escalas, rotação ou iluminação.

Para experimento 1 a rede foi treinada por 26000 épocas, para realização dos testes foi utilizado o conjunto de pesos da época 25300, já o experimento 2 foi treinado por 22500 épocas, e os pesos utilizados para o teste foram da época 21800.

5 RESULTADOS E DISCUSSÃO

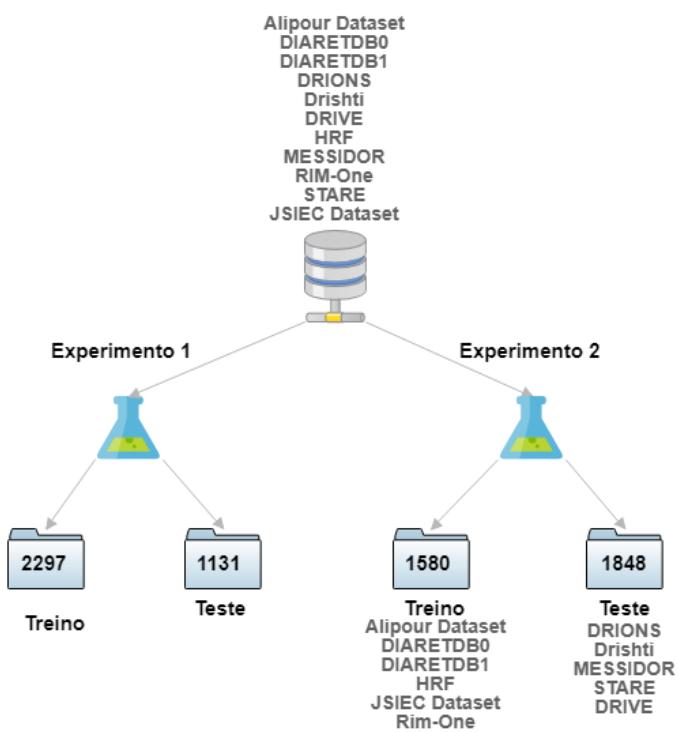
Este capítulo descreve os resultados obtidos por este trabalho, bem como sua metodologia de validação e comparações com outros estudos da literatura.

5.0.1 Experimentos

Para validar a metodologia todas as imagens foram convertidas para o formato JPG por motivos de implementação do *framework* utilizado e foram realizados dois experimentos com diferentes bases de treino e teste visando fazer uma melhor comparação com os resultados dos trabalhos relacionados citados. Durante o treinamento e teste as imagens são redimensionadas para o tamanho 416×416 devido a necessidade de adaptar a imagem ao tamanho dos filtros da rede.

No primeiro experimento foram selecionadas aleatoriamente 33% das imagens para teste, totalizando 1131 imagens de teste e 2297 imagens de treino. No segundo experimento as imagens da base de teste foram as mesmas imagens das bases usadas no trabalho de Silva *et al.* (2016) (Tabela 1), exceto pela base ORIGA, a qual não foi encontrada para download publicamente. O restante das imagens foram utilizadas na base de treinamento. Sendo assim, obteve-se 1848 imagens de teste e 1580 imagens de treino, ou seja, 52,1% para base de treino e 53,9% para base de teste, assim como ilustrado na Figura 29.

Figura 29 – Diagrama detalhando a separação das bases de treino e teste.



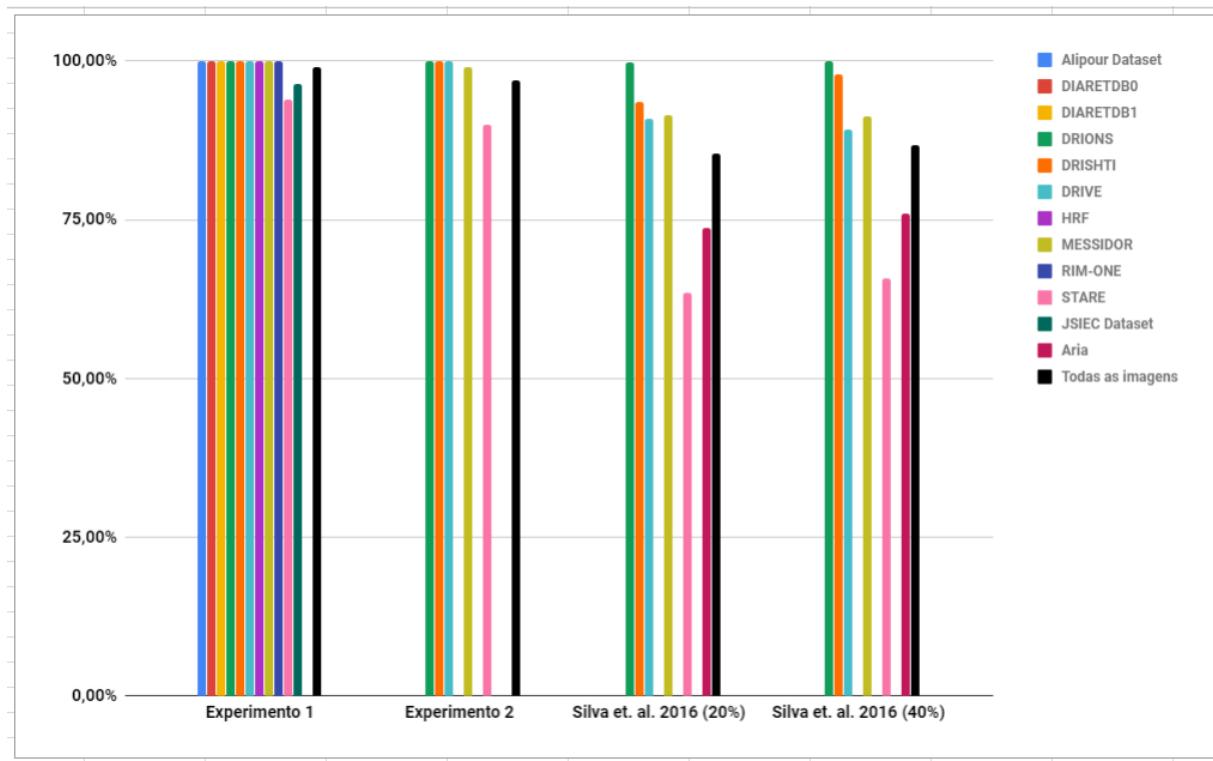
Fonte: Elaborado pelo autor.

5.1 Comparação com técnicas clássicas de processamento de imagem

O gráfico da Figura 30 descreve os resultados obtidos pela metodologia proposta neste estudo (Experimento 1 e 2) comparando com o método de Silva *et al.* (2016), dividido em quatro setores horizontais, cada um representando um experimento de cada estudo. Observa-se que os experimentos propostos são superiores a 90% em todas as bases utilizadas, enquanto o trabalho comparado possui um resultado médio de no máximo 86,69%.

Para validação dos resultados foi utilizado o mesmo método do trabalho citado, descrito na equação 2.9, considerando-se o centro do DO como o centro do *bouding box* e o raio sendo a altura do *bouding box*.

Figura 30 – Gráfico comparativos dos resultados obtidos por experimento e base utilizada.



Fonte: Elaborado pelo autor.

A Tabela 7 expõem em mais detalhes os resultados apresentados no gráfico da Figura 30. Os espaços preenchidos com '-' indicam em qual base e método não foi possível realizar os testes. Este resultado também foi obtido através da métrica descrita na Equação 2.9.

Nota-se que o pior desempenho do método proposto se deu nas bases MESSIDOR, STARE e JSIEC, as quais foram as únicas bases em que o acerto foi menor que 100% (3 das 11 utilizadas). Tal resultado pode se justificar devido a grande variedade de imagens nestas bases, além das várias patologias presentes, principalmente na STARE e JSIEC, esta última é dividida

Tabela 7 – Comparação dos resultados obtidos deste trabalho com o método de (a) (SILVA *et al.*, 2016)

Base	Experimento 1	Experimento 2	(a) (20%)	(a) (40%)
DRIONS	100,00%	100,00%	99,79%	100,00%
DRISHTI	100,00%	100,00%	93,55%	98,00%
DRIVE	100,00%	100,00%	90,83%	89,16%
MESSIDOR	100,00%	99,00%	91,44%	91,31%
STARE	93,93%	89,92%	63,40%	65,75%
Aria	-	-	73,70%	75,90%
Alipour Dataset	100,00%	-	-	-
DIARETDB0	100,00%	-	-	-
HRF	100,00%	-	-	-
DIARETDB1	100,00%	-	-	-
RIM-ONE	100,00%	-	-	-
JSIEC Dataset	96,36%	-	-	-
Todas	98,14%	97,00%	85,45%	86,69%

Fonte: Elaborado pelo autor.

em 37 categorias e em diversas delas o DO não está visível, o mesmo caso da STARE onde em quase 1/3 das imagens o DO é parcialmente ou totalmente não visível o que foi a causa do erro em mais de 50% das imagens no Experimento 1

5.2 Comparação com técnicas de Deep Learning

Comparando com métodos que utilizam CNNs, a proposta do presente trabalho também se mostra superior, tanto por utilizar mais bases de dados quanto por obter resultados superiores como observado na Tabela 8. Através de análise empírica também pode-se observar que a qualidade do *bouding box* marcado também é superior, como demonstrado na Figura 31 o *bouding box* resultante do método de Niu *et al.* (2017) extrapola a verdadeira região do DO em pelo menos 40% em largura e altura, o que resulta em um baixo IoU.

Figura 31 – Resultado da marcação de do *bouding box* de Niu *et al.* (2017).



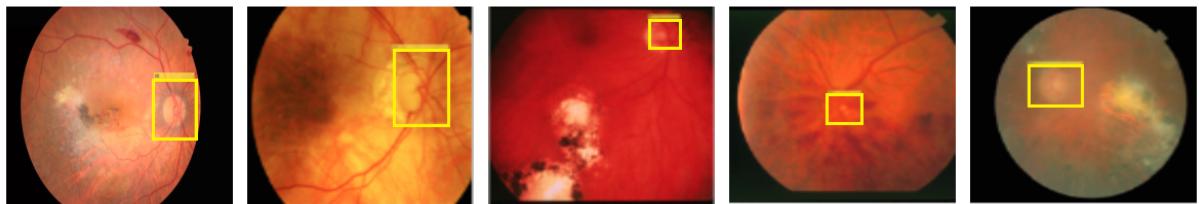
Fonte: Adaptado de Niu *et al.* (2017).

Tabela 8 – Comparaçao dos resultados obtidos com o método de Niu *et al.* (2017).

Método	ORIGA	MESSIDOR	N. Imagens
Experimento 1	-	100,00%	396
Experimento 2	-	99,00%	1200
Niu <i>et al.</i> (2017)	99,33%	98,75%	600

Fonte: Elaborado pelo autor.

Já o estudo de Alghamdi *et al.* (2016) apresenta maior quantidade de imagens, no entanto a maior parte delas não pertence a bases públicas, impossibilitando a comparação com seu resultado por completo, no entanto a Tabela 9 faz uma comparação entre os resultados obtidos com as bases em comum. Já a Figura 32 demonstra os *bouding boxes* encontrados, sendo possível observar que os *bouding boxes* marcados por este método é superior ao método citado anteriormente e compara-se a qualidade do obtido por este estudo.

Figura 32 – Resultado da marcação de do *bouding box* de Alghamdi *et al.* (2016).

Fonte: Adaptado de Alghamdi *et al.* (2016).

Tabela 9 – Comparaçao com os resultados de Alghamdi *et al.* (2016).

	Alghamdi <i>et al.</i> (2016)	Experimento 1
DRIVE	100,00%	100,00%
MESSIDOR	99,20%	100,00%
STARE	86,71%	93,93%
DIARETDB1	98,88%	100,00%
Todas	96,30%	98,14%

Fonte: Elaborado pelo autor.

O método de Xu *et al.* (2017) possui resultado superior na base MESSIDOR, no entanto em seu experimento foram utilizadas apenas 575 imagens desta base para treino e 526 para teste, totalizando 1101 das 1200 disponíveis, isso somado a baixa qualidade dos *bouding boxes* marcados devido a seu tamanho pré estabelecido, tornam o método proposto neste trabalho superior.

As Figuras 33 e 34 demonstram alguns dos casos de erro do Experimento 1, a partir daí pode-se observar que tais erros devem-se principalmente a má qualidade da imagem, alguma

Tabela 10 – Comparação com os resultados de Xu *et al.* (2017).

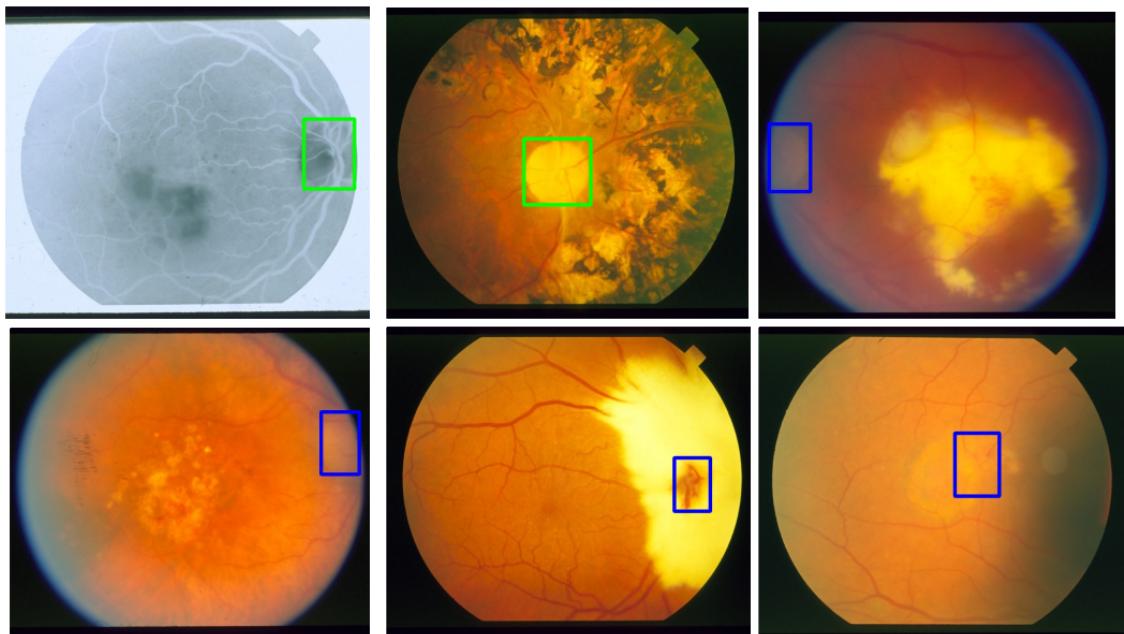
	Xu <i>et al.</i> (2017)	Experimento 1
MESSIDOR	99,43%	100,00%
STARE	89,30%	93,33%
Todas	96,30%	98,14%

Fonte: Elaborado pelo autor.

grave patologia ou ao DO não visível. Os *bouding boxes* marcados em azul são referentes a saída gerada pela rede, já os verdes são os marcados como *ground-truth*.

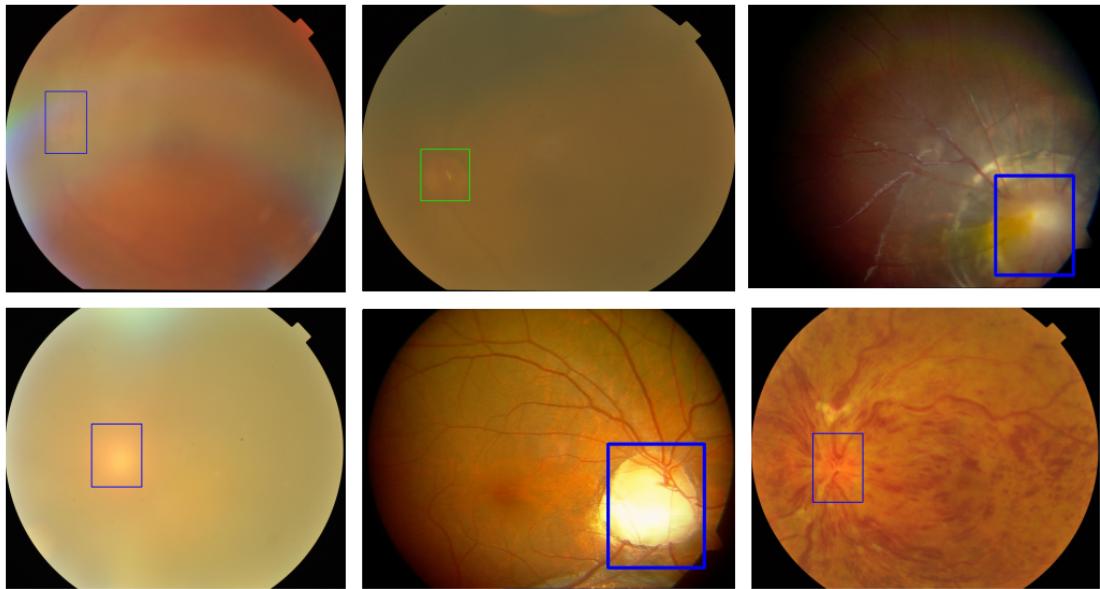
Os resultados apresentados até agora são baseados na Equação 2.9. No entanto a métrica mais comum na detecção de objetos é a IoU (descrito na Seção 2.4.1), esta métrica é também utilizada como Score de predição no artigo original da YoLo (REDMON *et al.*, 2016), como já mencionado esta métrica penaliza qualquer desvio do *bouding box* marcado no *ground truth*, empiricamente pode-se levar em consideração que IoU de 0,55 pode ser considerado um resultado satisfatório para análise, visto essa afirmação, a Tabela 11 demonstra os erros obtidos para determinados níveis de acerto no IoU.

Figura 33 – Erros do Experimento 1 na base STARE.



Fonte: Elaborado pelo autor.

Figura 34 – Erros do Experimento 1 na base JSIEC.



Fonte: Elaborado pelo autor.

A utilização da IoU como métrica para validação dos resultados torna este trabalho rigoroso e compatível com os demais trabalhos da literatura, IoU superior a 0,90 indica um *bouding box* praticamente perfeito, nível que foi atingido em 30,20% das imagens. Já a média geral do IoU no Experimento 1 foi de 75,73%, no Experimento 2 a média foi de 71,19%

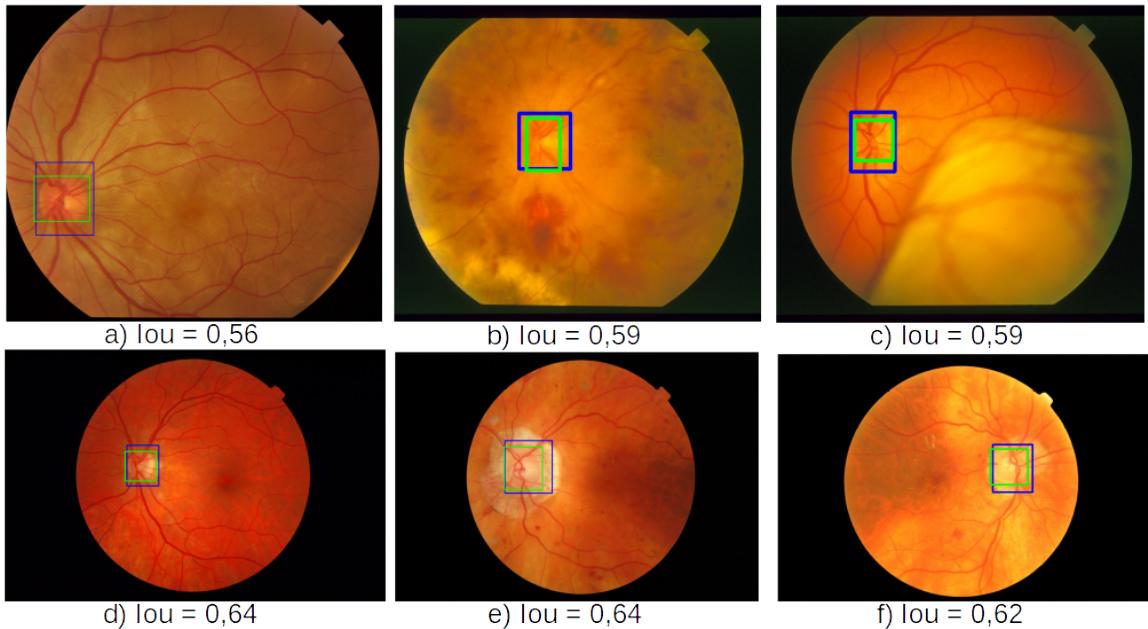
Tabela 11 – Avaliação dos resultados em termos de IoU.

IoU maior que	Ac. Experimento 1	Ac. Experimento 2
0,55	97,43%	92,57%
0,60	96,81%	88,72%
0,70	92,21%	68,00%
0,80	65,69%	30,09%
0,90	42,81%	05,81%
Média	75,73%	71,19%

Fonte: Elaborado pelo autor.

A Figura 5.2 demonstra o resultado em termos de IoU em alguns dos piores resultados obtidos em imagens classificadas corretamente. Tomando-se o intervalo de IoU entre 0,55 e 0,65 dentre as imagens classificadas corretamente, apenas 23 de 1110 obtiveram este resultado, com IoU abaixo de 0,55 foram classificadas 8 imagens.

Figura 35 – Piores resultados em termos de IoU.



Fonte: Elaborado pelo autor.

A Tabela 12 compara os resultados obtidos pelos Experimentos 1 e 2 com os demais estudos apresentados que utilizam CNNs para detecção do DO. O experimento proposto mostrou-se superior a todos os outros apresentados em todas as bases que foram comparadas, assim como uma média geral, utilizando a métrica da Seção 2.4.2.

Tabela 12 – (a) Alghamdi *et al.* (2016), (b) Xu *et al.* (2017), (c) Niu *et al.* (2017)

Base	Experimento 1	Experimento 2	(a)	(b)	(c)
MESSIDOR	100,00%	99,00%	99,20%	99,43%	98,75%
STARE	93,93%	89,92%	86,71%	89,30%	
DRIVE	100,00%	100,00%	100,00%		
DIARETDB1	100,00%			98,88%	
ORIGA					99,33%
DRIONS	100,00%	100,00%			
DRISHTI	100,00%	100,00%			
Alipour Dataset	100,00%		-		
DIARETDB0	100,00%		-		
HRF	100,00%		-		
RIM-ONE	100,00%		-		
JSIEC Dataset	96,36%		-		
Todas	98,14%	97,00%	93,30%	96,30%	99,04%

Fonte: Elaborado pelo autor.

6 CONCLUSÃO

Este trabalho apresentou um método para detectar o Disco Óptico em imagens de fundo de olho utilizando técnicas de *deep learning*, além de um conjunto de marcações para a região do DO em 3428 imagens de fundo de olho para a maior parte das bases públicas disponíveis e utilizadas na literatura.

A detecção do DO foi obtida utilizando a arquitetura da CNN YoLo com algumas modificações em seus parâmetros de saída, A localização foi obtida com sucesso em 98,14% das imagens, superior aos resultados encontrados nos trabalhos relacionados e até mesmo testando em um conjunto maior de imagens.

Por utilizar uma grande variedade de bases com diversos tipos de patologia ou saudáveis esta metodologia mostrou-se robusta e capaz de generalizar seu poder discriminatório, mesmo em imagens que não possuem DO no campo de visão, que eram a minoria das imagens. Os resultados obtidos foram de 100% de acerto em quase todas as bases testadas, apenas nas duas com o maior número de imagens de má qualidade houve erro, além de provar-se melhor que os trabalhos já conhecidos que utilizam técnicas clássicas de processamento de imagens ou *deep learning*.

Este método mostrou-se promissor para ser utilizado em diversos fins como uma pré etapa no diagnóstico do glaucoma através da análise do DO e da CDO, exclusão de falsos positivos na detecção de exsudatos para o diagnóstico de retinopatia diabética entre outros.

Em trabalhos futuros pretende-se formatar este trabalho em forma de artigo a fim de submete-lo em periódico de relevância nacional, além de utilizar o método de detecção proposto para realizar o diagnóstico do glaucoma com uma CNN como classificador, além de detecção de exsudatos para diagnóstico de retinopatia diabética, bem como a ampliação das bases com diagnóstico para estas patologias, devido a demanda de grande volume de dados para treinamento de CNNs. As marcações realizadas neste trabalho devem ser submetidas a revisão de um ou mais especialistas antes de serem disponibilizadas à comunidade.

REFERÊNCIAS

- AB, A. **Yolo Mark**. 2016. Acessado em 28/11/2018 as 16:43. Disponível em: <https://github.com/AlexeyAB/Yolo_mark>. 40
- ACHANTA, R.; HEMAMI, S.; ESTRADA, F.; SUSSTRUNK, S. Frequency-tuned salient region detection. In: IEEE. **Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on**. [S.I.], 2009. p. 1597–1604. 28
- AKRAM, M. U.; KHAN, A.; IQBAL, K.; BUTT, W. H. Retinal images: optic disk localization and detection. In: SPRINGER. **International Conference Image Analysis and Recognition**. [S.I.], 2010. p. 40–49. 25
- ALGHAMDI, H. S.; TANG, H. L.; WAHEEB, S. A.; PETO, T. Automatic optic disc abnormality detection in fundus images: a deep learning approach. University of Iowa, 2016. 9, 10, 29, 30, 31, 46, 49
- ALIPOUR, S. H. M.; RABBANI, H.; AKHLAGHI, M.; DEHNAVI, A. M.; JAVANMARD, S. H. Analysis of foveal avascular zone for grading of diabetic retinopathy severity based on curvelet transform. **Graefe's Archive for Clinical and Experimental Ophthalmology**, Springer, v. 250, n. 11, p. 1607–1614, 2012. 9, 33, 38
- CARMONA, E. J.; RINCÓN, M.; GARCÍA-FEIJÓÓ, J.; CASA, J. M. Martínez-de-la. Identification of the optic nerve head with genetic algorithms. **Artificial Intelligence in Medicine**, Elsevier, v. 43, n. 3, p. 243–259, 2008. 33, 37
- DECENCIÈRE, E.; ZHANG, X.; CAZUGUEL, G.; LAY, B.; COCHENER, B.; TRONE, C.; GAIN, P.; ORDONEZ, R.; MASSIN, P.; ERGINAY, A. *et al.* Feedback on a publicly distributed image database: the messidor database. **Image Analysis & Stereology**, v. 33, n. 3, p. 231–234, 2014. 33, 34
- DUDA, R. O.; HART, P. E. **Use of the Hough transformation to detect lines and curves in pictures**. [S.I.], 1971. 25
- EVERINGHAM, M.; GOOL, L. V.; WILLIAMS, C. K.; WINN, J.; ZISSERMAN, A. The pascal visual object classes (voc) challenge. **International journal of computer vision**, Springer, v. 88, n. 2, p. 303–338, 2010. 19
- FUMERO, F.; ALAYÓN, S.; SANCHEZ, J.; SIGUT, J.; GONZALEZ-HERNANDEZ, M. Rim-one: An open retinal image database for optic nerve evaluation. In: IEEE. **Computer-Based Medical Systems (CBMS), 2011 24th International Symposium on**. [S.I.], 2011. p. 1–6. 33, 34
- GONZALEZ, R. C. **Digital image processing**. [S.I.]: Pearson Education India, 2009. 26
- HOOVER, A.; KOUZNETSOVA, V.; GOLDBAUM, M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. **IEEE Transactions on Medical imaging**, IEEE, v. 19, n. 3, p. 203–210, 2000. 33, 35
- KÄLVIÄINEN, R.; UUSITALO, H. Diaretldb1 diabetic retinopathy database and evaluation protocol. In: CITESEER. **Medical Image Understanding and Analysis**. [S.I.], 2007. v. 2007, p. 61. 25, 33, 37, 38

- KAUPPI, T.; KALESNYKIENE, V.; KAMARAINEN, J.-K.; LENSU, L.; SORRI, I.; UUSITALO, H.; KÄLVIÄINEN, H.; PIETILÄ, J. Diaretdb0: Evaluation database and methodology for diabetic retinopathy algorithms. **Machine Vision and Pattern Recognition Research Group, Lappeenranta University of Technology, Finland**, v. 73, 2006. 25, 33, 37, 38
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2012. p. 1097–1105. 29
- LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, IEEE, v. 86, n. 11, p. 2278–2324, 1998. 19, 20
- LIN, M.; CHEN, Q.; YAN, S. Network in network. **arXiv preprint arXiv:1312.4400**, 2013. 40
- LINCHUAN. **1000 fundus images with 37 categories**. 2018. Acessado em 29/11/2018 as 18:16. Disponível em: <<https://www.kaggle.com/linchundan/fundusimage1000>>. 33, 39
- MORAES, P. L. **Visão, um dos cinco sentidos**. 2018. Acessado em 28/12/2018. Disponível em: <<https://mundoeducacao.bol.uol.com.br/biologia/visao.htm>>. 14
- NIU, D.; XU, P.; WAN, C.; CHENG, J.; LIU, J. Automatic localization of optic disc based on deep learning in fundus images. In: **IEEE. Signal and Image Processing (ICSIP), 2017 IEEE 2nd International Conference on**. [S.l.], 2017. p. 208–212. 9, 10, 28, 29, 45, 46, 49
- ODSTRCILIK, J.; KOLAR, R.; BUDAI, A.; HORNEGGER, J.; JAN, J.; GAZAREK, J.; KUBENA, T.; CERNOSEK, P.; SVOBODA, O.; ANGELOPOULOU, E. Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. **IET Image Processing**, IET, v. 7, n. 4, p. 373–383, 2013. 33, 35
- PARANHOS-JUNIOR, A.; OMI, C.; PRATA-JUNIOR, J. Sociedade brasileira de glaucoma: Iii consenso brasileiro de glaucoma primário de ângulo aberto. **Sao Paulo: BestPoint**, p. 77–96, 2009. 14
- PUNNOLIL, A. A novel approach for diagnosis and severity grading of diabetic maculopathy. In: **IEEE. Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on**. [S.l.], 2013. p. 1230–1235. 26
- RAJAPUT, G.; RESHMI, B.; SIDRAMAPPA, C. Automatic localization of fovea center using mathematical morphology in fundus images. **International Journal of Machine Intelligence**, Bioinfo Publications, 49, Vighnahar Shopping Complex Kharghar, Navi Mumbai 410210 India, v. 3, n. 4, 2011. 25
- REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: Unified, real-time object detection. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 779–788. 17, 21, 22, 40, 41, 47
- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological review**, American Psychological Association, v. 65, n. 6, p. 386, 1958. 18

- SEKAR, G. B.; NAGARAJAN, P. Localisation of optic disc in fundus images by using clustering and histogram techniques. In: **IEEE. Computing, Electronics and Electrical Technologies (ICCEET), 2012 International Conference on.** [S.I.], 2012. p. 584–589. 25
- SILVA, R. R. V. e; ARAUJO, F. H. D. de; SANTOS, L. M. R. dos; VERAS, R. d. M. S.; MEDEIROS, F. N. S. de. Optic disc detection in retinal images using algorithms committee with weighted voting. **IEEE Latin America Transactions**, IEEE, v. 14, n. 5, p. 2446–2454, 2016. 9, 10, 25, 26, 27, 28, 43, 44, 45
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014. 15
- SIVASWAMY, J.; KRISHNADAS, S.; JOSHI, G. D.; JAIN, M.; TABISH, S.; UJJWAFT, A. Drishti-gs: Retinal image dataset for optic nerve head (ohn) segmentation. In: **IEEE. Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on.** [S.I.], 2014. p. 53–56. 33, 36, 37
- STAAL, J.; ABRÀMOFF, M. D.; NIEMEIJER, M.; VIERGEVER, M. A.; GINNEKEN, B. V. Ridge-based vessel segmentation in color images of the retina. **IEEE transactions on medical imaging**, IEEE, v. 23, n. 4, p. 501–509, 2004. 25, 33, 36
- SZEGEDY, C.; IOFFE, S.; VANHOUCKE, V.; ALEMI, A. A. Inception-v4, inception-resnet and the impact of residual connections on learning. In: **AAAI**. [S.I.: s.n.], 2017. v. 4, p. 12. 21, 40
- TOBIN, K. W.; CHAUM, E.; GOVINDASAMY, V. P.; KARNOWSKI, T. P. Detection of anatomic structures in human retinal imagery. **IEEE transactions on medical imaging**, IEEE, v. 26, n. 12, p. 1729–1739, 2007. 24
- UIJLINGS, J. R.; SANDE, K. E. V. D.; GEVERS, T.; SMEULDERS, A. W. Selective search for object recognition. **International journal of computer vision**, Springer, v. 104, n. 2, p. 154–171, 2013. 21
- VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: **IEEE. Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on.** [S.I.], 2001. v. 1, p. I–I. 30
- WIDROW, B.; HOFF, M. E. Adaptive switching circuits. In: **MIT PRESS. Neurocomputing: foundations of research.** [S.I.], 1960. 17, 18
- XU, P.; WAN, C.; CHENG, J.; NIU, D.; LIU, J. Optic disc detection via deep learning in fundus images. In: **Fetal, Infant and Ophthalmic Medical Image Analysis.** [S.I.]: Springer, 2017. p. 134–141. 9, 10, 15, 31, 32, 46, 47, 49
- ZHANG, Z.; YIN, F. S.; LIU, J.; WONG, W. K.; TAN, N. M.; LEE, B. H.; CHENG, J.; WONG, T. Y. Orig-a-light: An online retinal fundus image database for glaucoma analysis and research. In: **IEEE. Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE.** [S.I.], 2010. p. 3065–3068. 15
- ZUBAIR, M.; YAMIN, A.; KHAN, S. A. Automated detection of optic disc for the analysis of retina using color fundus image. In: **IEEE. Imaging systems and techniques (IST), 2013 IEEE international conference on.** [S.I.], 2013. p. 239–242. 26