

# Glaucoma diagnostic using nerve fiber layer thickness images

## Diagnóstico do glaucoma a partir de imagens de espessuras da camada de fibras nervosas

BRAGA, Samira J.\*; GOMI, Edson S.\*

\*Escola Politécnica da Universidade de São Paulo

**Abstract**—Abstract here.**Keywords**— word 1; word 2.**Resumo**— É necessária a inserção do resumo para artigo escrito em Português.**Palavras-chave**— palavra 1; palavra 2.**Classificação**— Mestrado**Categoria**— Iniciante

### I. INTRODUÇÃO

Glaucoma é uma neuropatia óptica crônica multifatorial e de lenta progressão que causa perda de campo visual. Pesquisa realizada por Quigley e Broman mostrou que glaucoma é a segunda maior causa de cegueira no mundo e estimaram que atingirá 79,6 milhões de pessoas até 2020 [1]. O glaucoma caracteriza-se pela perda da camada de fibras nervosas no olho, o que pode causar cegueira se não for tratada corretamente [2]. O dano na camada de fibras nervosas acontece antes de alteração no campo visual do paciente e, por isso, o diagnóstico precoce é um fator importante para evitar a perda da visão [3].

O objetivo deste trabalho é investigar se é possível fazer o diagnóstico do glaucoma a partir das imagens da espessura da camada de fibras nervosas por meio de uma rede convolucional.

### II. DIAGNÓSTICO DE GLAUCOMA

Atualmente, o diagnóstico do glaucoma é feito com uma combinação de exame estrutural, por meio de tomografia de coerência óptica OCT (Optical Coherence Tomography) [4], e exame funcional, por meio da perimetria computadorizada SAP (Standard Automated Perimetry) [5]. Os exames estruturais avaliam a camada de fibras nervosas para identificar alterações na sua espessura e os exames funcionais avaliam o campo visual do paciente em busca de áreas de perda da visão.

O exame OCT utiliza o princípio da interferometria luminosa para medir as espessuras das estruturas intraoculares. Ao realizar uma varredura, o equipamento emite feixes de laser infravermelho e mede o tempo que a luz leva para ser refletida. Em cada estrutura atravessada, uma parte dessa luz é refletida

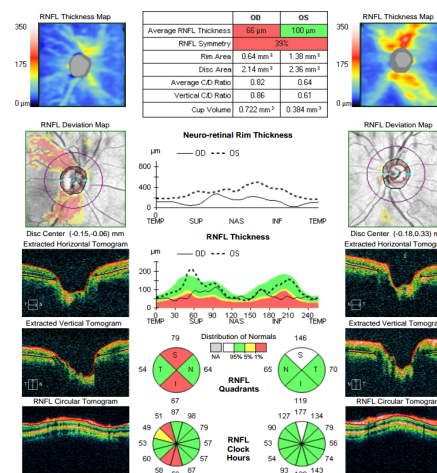


Figura 1. Saída de um exame de OCT.

de volta. O cálculo da espessura é baseado na diferença entre o feixe de luz de referência e a luz refletida [6].

Na figura 1 vemos a saída gerada pelo equipamento Cirrus HD-OCT da Carl Zeiss Meditec Inc. As imagens no topo do relatório apresenta o mapa das espessuras da camada de fibras nervosas. As áreas em azul indicam espessuras próximas a 0, as áreas amarelas indicam espessuras até 175 microns e as áreas vermelhas indicam espessuras entre 175 e 350 microns. O disco central em cinza é a fóvea, ponto cego do olho de onde sai o nervo óptico. Abaixo do mapa de espessuras está o mapa de desvio que mostra as áreas onde a espessura está diferente da população normal, de acordo com um banco de dados normativo do software do equipamento. Também nesse gráfico está representada a linha circular de onde são retirados os pontos para a representação da espessura nos gráficos na parte de baixo da saída. Esses gráficos mostram a espessura ao longo da circunferência, com centro no nervo óptico. A partir dos pontos dessa circunferência, são gerados os gráficos de quadrantes e horas de relógio, que representam as espessuras médias em cada região [7].

Nenhum exame fornece resultados definitivos da situação do paciente, necessitando assim da avaliação de um profissional

oftalmologista especialista em glaucoma para conclusão do resultado. Esta análise não é padronizada e pode variar de acordo com o julgamento do oftalmologista [8]. É possível utilizar classificadores de aprendizagem de máquina para fazer o diagnóstico de glaucoma para auxiliar na tomada de decisão. Um classificador de rede neural profunda foi utilizado com dados de exame de campo visual para detectar estágios iniciais de glaucoma [9]. Dados de OCT e SAP em conjunto também foram utilizados para treinar diversos tipos de classificadores [7], [10].

### III. REDES NEURAIAS CONVOLUCIONAIS

Redes neurais profundas estão sendo amplamente utilizadas em diversos domínios para classificação e detecção de objetos, principalmente em visão computacional e reconhecimento e processamento de linguagem natural. Essas redes consistem em várias camadas conectadas que aprendem a reconhecer padrões nos dados apresentados, sejam eles imagens ou sons [11]. O principal tipo de rede utilizado é a rede neural convolucional (Convolutional Neural Network, CNN). As CNNs mostraram ser excelentes ferramentas no processamento de imagens, inclusive na área médica, sendo utilizadas para classificação, detecção de objetos e segmentação. Esse sucesso deve-se principalmente à utilização de grandes quantidades de exemplos de treinamento, que permitem que a rede aprenda a reconhecer características a partir dos dados brutos das imagens [12].

As CNNs são compostas de várias camadas conectadas, sendo os principais tipos as convolucionais, pooling e as camadas totalmente conectadas. As camadas convolucionais servem como filtros para identificar as características das imagens, como curvas, bordas e cores. As saídas das camadas convolucionais são mapas de características identificadas na imagem de entrada. A convolução é feita utilizando uma máscara (ou kernel) que desliza por todos os pixels da imagem multiplicando os valores do pixel pelos valores da máscara. Essa multiplicação é feita em toda a imagem com diferentes máscaras, gerando assim as saídas dos diferentes filtros em cada camada de convolução.

Após a identificação dessas características, a camada de pooling agrupa as características e reduz a resolução da imagem. Também utilizando uma máscara, a operação de pooling irá calcular o valor máximo em cada região da imagem, agrupando características similares em um mesmo pixel. Na figura 2 é possível acompanhar o agrupamento de características e a redução da resolução da imagem ao longo de 5 camadas convolucionais e pooling.

As camadas totalmente conectadas recebem todas as características identificadas nas imagens e fazem a classificação final. A saída dessa camada será um vetor com o número de classes a serem classificadas, cada classe terá um valor de probabilidade sendo o maior valor a classe identificada para a imagem. O desenho de uma configuração de rede convolucional profunda pode ser visto na figura 3. Várias camadas de convolução e pooling utilizando ativações por ReLU (Rectified Linear Unit) são seguidas por duas ou mais

camadas totalmente conectadas. O aumento da quantidade de camadas mostrou-se eficiente no aumento da acurácia de classificação das redes convolucionais [13].

#### A. Transfer Learning

Para treinar uma CNN é necessário ter um conjunto muito grande de exemplos de treinamento. Redes como GoogLeNet, VGG, AlexNet, foram treinadas com datasets como o ImageNet com 14 milhões de imagens distribuídas em 1000 categorias para obter bons resultados [15]. No entanto, imagens médicas são muito mais difíceis de se obter, principalmente com anotações de classes, devido ao tempo para realizar os estudos e para anotar todos os exemplos [12].

Uma alternativa para a utilização de redes profundas com um dataset menor é a utilização do transfer learning. Essa técnica consiste em usar uma rede treinada com um dataset grande de imagens em um domínio mais amplo e transferir esse conhecimento para um domínio mais específico. Dessa forma, ao invés de utilizar inicialização randômica dos parâmetros da rede, utiliza-se os pesos de uma rede já treinada. A utilização de uma rede pré-treinada reduz o tempo de treinamento e a necessidade de um dataset muito grande para o novo objetivo [16].

A aplicação do transfer learning pode ser feita em diversos contextos onde haja dificuldade de obter dados para treinamento. Em um survey, Shao *et al.* [17] relatam diversos usos de transfer learning para treinamento de redes com diferentes domínios alvo. Em oftalmologia, essa técnica mostrou-se eficiente para classificação de retinopatia diabética [18] e degeneração macular [19]. Nestes trabalhos foram utilizadas redes pré-treinadas com o dataset ImageNet, o que mostra que é possível utilizar o conhecimento de um domínio de origem mais amplo em uma tarefa específica.

#### B. Redes pré-treinadas

Entre as diversas redes convolucionais já treinadas com o ImageNet, foi escolhida a rede VGG16 para utilização neste experimento. Essa rede foi utilizada para classificação de degeneração macular com imagens de OCT em [19]. Os resultados apresentados por Lee *et al.* são promissores, obtendo área sob a curva ROC de 97,46%.

A rede VGG16 foi utilizada no ImageNet Large Scale Visual Recognition Challenge (ILSVRC) em 2014, uma competição de reconhecimento de imagens utilizando redes neurais profundas. A rede é composta ao todo de 21 camadas com ativações ReLU, sendo 13 camadas convolucionais, 5 max pooling e 3 camadas totalmente conectadas no final, duas com 4096 saídas e a última com 1000 saídas de classificação. Diferente de outras arquiteturas, o VGG16 utiliza filtros menores nas camadas de convolução, o que diminui o número de pesos necessários mas ainda com um poder discriminativo alto. [13].

### IV. EXPERIMENTOS E RESULTADOS

#### A. Dataset

O dataset original foi obtido com o departamento de oftalmologia da Unicamp. O dataset consiste de imagens de

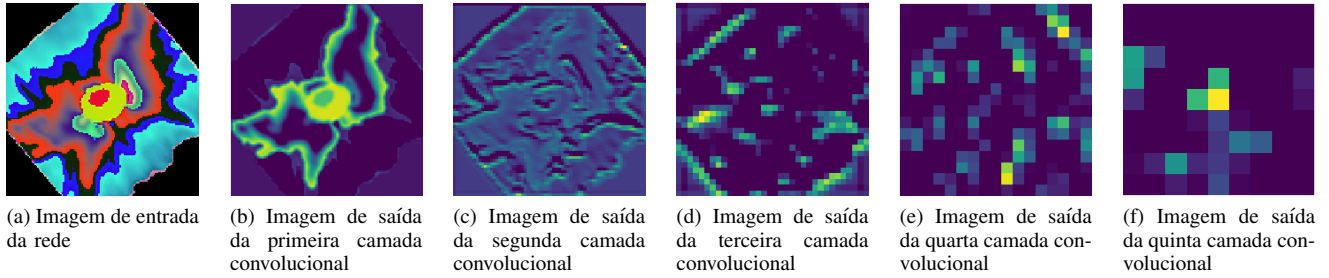


Figura 2. Exemplos de saídas de cada camada convolucional ao longo da rede.

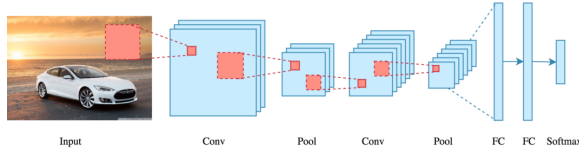


Figura 3. Esquema de camadas de uma rede neural convolucional [14].

OCT com tamanho 136x136 de 56 olhos com glaucoma e 66 olhos normais, totalizando 122 pacientes. Os gráficos de espessura de fibras nervosas foram obtidos através da extração das imagens do PDF do exame. Foram selecionados para o experimento somente os olhos de pacientes que foram manualmente classificados por especialistas.

Para a separação do dataset em treino e validação, foram selecionados 20% de olhos normais e 20% de olhos com glaucoma para validação, e o restante para treino, totalizando 98 imagens de treino e 24 para validação. As imagens selecionadas para teste não estão presentes no dataset de treino, para que o algoritmo possa classificar imagens independentes do dataset de treino.

Para evitar overfitting, foi empregada uma técnica para aumentar o número de exemplos a partir das imagens no dataset de treino. Cada imagem foi rotacionada 100 vezes em ângulos aleatórios entre 0 e 360 graus, gerando assim um dataset de treino com 9800 imagens. As imagens de validação não foram rotacionadas.

### B. Pré-processamento

Para utilização do transfer learning, foi necessário fazer a subtração do pixel médio. O pixel médio é calculado sobre todas as imagens do dataset de treino utilizando a equação 1, onde  $P_c$  é o valor médio por canal,  $M$  e  $N$  são as dimensões das imagens de entrada (largura e altura, respectivamente),  $I$  é a quantidade total de imagens no dataset de treino,  $p_{mni}$  é o valor do pixel na posição  $m,n$  na imagem  $i$ . O valor do pixel médio é calculado em cada canal RGB da imagem, gerando assim um vetor de 3 posições com um valor médio por canal.

$$P_c = \frac{1}{M * N * I} \sum p_{mni} \quad (1)$$

Onde houveram falhas na aquisição da imagem gerando áreas escuras, como no exemplo da figura 4, pixels com valores RGB próximos ao preto foram substituídos pelo valor de preto

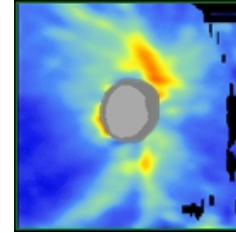


Figura 4. Exemplo de imagem onde houve falha na aquisição, gerando as áreas pretas no lado direito.

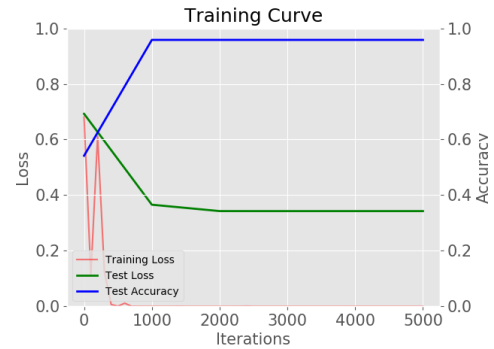


Figura 5. Acurácia e erro de treino e validação da rede VGG16 com transfer learning.

absoluto RGB (0, 0, 0) \*\*para que não tenham influência sobre a decisão do classificador\*\*.

### C. Resultados com transfer learning

Neste experimento, utilizamos a mesma arquitetura da rede VGG16, alterando a saída da última camada totalmente conectada para duas saídas, correspondente às duas classes a serem classificadas: normal e glaucoma. As imagens do Imagenet tinham a resolução de 224x224 pixels, enquanto que as imagens de OCT têm resolução de 136x136 pixels, por isso a camada de entrada da rede também foi alterada para a resolução das imagens do dataset utilizado.

Os pesos pré-treinados foram carregados para inicialização apenas das camadas convolucionais. As três últimas camadas totalmente conectadas foram iniciadas com valores aleatórios de uma distribuição normal com desvio padrão 0.01.

O treinamento foi realizado em todas as camadas da rede, utilizando o gradiente descendente estocástico por 5000 iterações, com mini batches de 15 imagens. Os parâmetros de momentum e weight decay foram definidos como 0.9 e 0.0005, respectivamente. A taxa de aprendizagem inicial foi de 0.001. A cada 1000 iterações a taxa de aprendizagem foi diminuída utilizando a equação 2.

$$base\_lr * \gamma^{\lfloor \frac{iter}{step} \rfloor} \quad (2)$$

Onde  $base\_lr$  é a taxa de aprendizagem inicial,  $\gamma$  é um parâmetro do Caffe definido com o valor 0.1,  $iter$  é o número da iteração atual e  $step$  é um parâmetro definido como 1000.

A validação do modelo foi feita utilizando um dataset de 24 imagens não vistas pelo algoritmo durante a fase de treinamento. Foi obtida acurácia final de 95.8%. O gráfico na figura 5 mostra a evolução dos valores de erro e acurácia durante o processo de treinamento da rede. A validação foi feita a cada 1000 iterações. É possível identificar a estabilização da acurácia após 1000 iterações, quando o valor de erro do treinamento chega próximo à zero.

Os experimentos foram realizados em um servidor com duas GPUs Nvidia Quadro K5200 com 8GB de memória cada. O framework de deep learning utilizado para definição da rede e treinamento em todos os experimentos foi o Caffe, desenvolvido em C++ pela universidade de Berkeley [20].

## V. DISCUSSÃO

A principal barreira neste estudo foi a quantidade limitada de dados, o que levou à utilização de técnicas para aumentar o dataset artificialmente. Ainda não foi possível determinar se as imagens rotacionadas geraram impacto no aprendizado da rede. Os resultados obtidos sugerem que houve overfitting no dataset de treinamento, sendo necessário o teste em um dataset diferente.

## VI. CONCLUSÃO

### REFERÊNCIAS

- [1] QUIGLEY, H. A. The number of people with glaucoma worldwide in 2010 and 2020. *British Journal of Ophthalmology*, BMJ, v. 90, n. 3, p. 262–267, mar 2006. Disponível em: <<http://dx.doi.org/10.1136/bjo.2005.081224>>.
- [2] QUIGLEY, H. A. Glaucoma. *The Lancet*, v. 377, n. 9774, p. 1367–1377, 2011.
- [3] MALIK, R.; SWANSON, W. H.; GARWAY-HEATH, D. F. 'structure-function relationship' in glaucoma: past thinking and current concepts. *Clinical & Experimental Ophthalmology*, Wiley-Blackwell, v. 40, n. 4, p. 369–380, apr 2012. Disponível em: <<http://dx.doi.org/10.1111/j.1442-9071.2012.02770.x>>.
- [4] WOLLSTEIN, G. et al. Optical coherence tomography longitudinal evaluation of retinal nerve fiber layer thickness in glaucoma. *Arch Ophthalmol*, v. 123, p. 464–471, 2005.
- [5] CHAUHAN, B. C.; DRANCE, S. M.; DOUGLAS, G. R. The use of visual field indices in detecting changes in the visual field in glaucoma. *Investigative Ophthalmology & Visual Science*, v. 31, n. 3, p. 512, 1990. Disponível em: <<http://dx.doi.org/>>.
- [6] HUANG, D. et al. Optical coherence tomography. *Science*, American Association for the Advancement of Science, v. 254, n. 5035, p. 1178–1181, 1991. ISSN 0036-8075. Disponível em: <<http://science.sciencemag.org/content/254/5035/1178>>.

- [7] DIAS, M. *Geracao de Populacoes Artificiais para Treinamento de Clasificadores de Apoio ao diagnostico de Glaucoma*. Tese (Doutorado) — Escola Politécnica da USP, 2009.
- [8] KROESE, M.; BURTON, H. Primary open angle glaucoma. the need for a consensus case definition. *J Epidemiol Community Health*, v. 57, p. 752–754, 2003.
- [9] ASAKA, R. et al. Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. *Ophthalmology*, Elsevier BV, v. 123, n. 9, p. 1974–1980, sep 2016. Disponível em: <<http://dx.doi.org/10.1016/j.ophtha.2016.05.029>>.
- [10] BOWD, C. et al. Bayesian machine learning classifiers for combining structural and functional measurements to classify healthy and glaucomatous eyes. *Investigative Ophthalmology & Visual Science*, v. 49, n. 3, p. 945, 2008. Disponível em: <<http://dx.doi.org/10.1167/iov.07-1083>>.
- [11] LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, may 2015. Disponível em: <<http://dx.doi.org/10.1038/nature14539>>.
- [12] GREENSPAN, H.; GINNEKEN, B. van; SUMMERS, R. M. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, v. 35, n. 5, p. 1153–1159, May 2016. ISSN 0278-0062.
- [13] SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014. Disponível em: <<http://arxiv.org/abs/1409.1556>>.
- [14] DERTAT, A. *Applied Deep Learning - Part 4: Convolutional Neural Networks*. Disponível em: <<https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2>>.
- [15] RUSSAKOVSKY, O. et al. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, v. 115, n. 3, p. 211–252, 2015.
- [16] Tan, C. et al. A Survey on Deep Transfer Learning. *ArXiv e-prints*, ago. 2018.
- [17] SHAO, L.; ZHU, F.; LI, X. Transfer learning for visual categorization: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, v. 26, n. 5, p. 1019–1034, May 2015. ISSN 2162-237X.
- [18] LI, X. et al. Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification. In: *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. [S.l.: s.n.], 2017. p. 1–11.
- [19] LEE, C. S.; BAUGHMAN, D. M.; LEE, A. Y. Deep learning is effective for classifying normal versus age-related macular degeneration oct images. *Ophthalmology Retina*, v. 1, n. 4, p. 322 – 327, 2017. ISSN 2468-6530. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2468653016301749>>.
- [20] JIA, Y. et al. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.