

Glaucoma Diagnosis Using Nerve Fiber Layer Thickness Images

Diagnóstico do Glaucoma Usando Imagens de Espessura da Camada de Fibras Nervosas

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

*Escola Politécnica da Universidade de São Paulo

Abstract—Glaucoma is a cronic disease of slow progression and is characterized by the loss of the nerve fiber layer in the eye, if not treated it can lead to irreversible blindness. The goal of this paper is to investigate the possibility of diagnosing glaucoma from nerve fiber layer thickness images using a convolutional neural network. Using VGG16 and transfer learning technique, the net was initialized using pretrained weights for training and glaucoma classification on a dataset of images of nerve fiber layer thickness. Initial results show that the use of OCT images in glaucoma diagnosis is viable, however, high loss on validation dataset suggests overfitting. One hypothesis for this result is the limited number of images in the training dataset.

Keywords— Glaucoma Diagnosis, Optical Coherence Tomography, Convolutional Neural Network, Transfer Learning

Resumo— Glaucoma é uma doença crônica e de lenta progressão que caracteriza-se pela perda da camada de fibras nervosas no olho, se não tratada pode levar à cegueira irreversível. 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 neural convolucional. Utilizando a rede VGG16 com técnica de transfer learning, foi feita a inicialização dos pesos pré-treinados para treinamento e classificação de glaucoma em um dataset de imagens de espessura da camada de fibras nervosas. Os resultados iniciais mostram que a utilização de imagens de OCT no diagnóstico de glaucoma é viável, porém, o alto erro no dataset de validação indica overfitting. Uma das hipóteses para o resultado é a pequena quantidade de imagens de treinamento.

Palavras-chave— Diagnóstico de Glaucoma, Tomografia de Coerência Óptica, Redes Neurais Convolucionais

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].

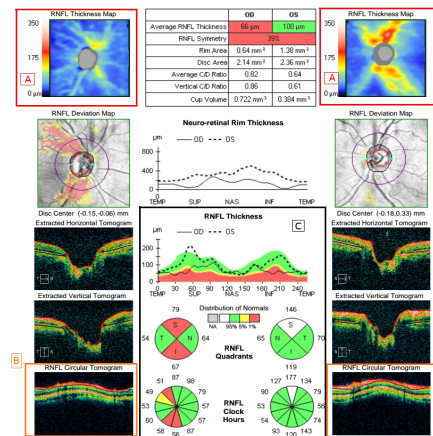


Figura 1. Saída de um exame de OCT.

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 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 [7]. Na região A estão

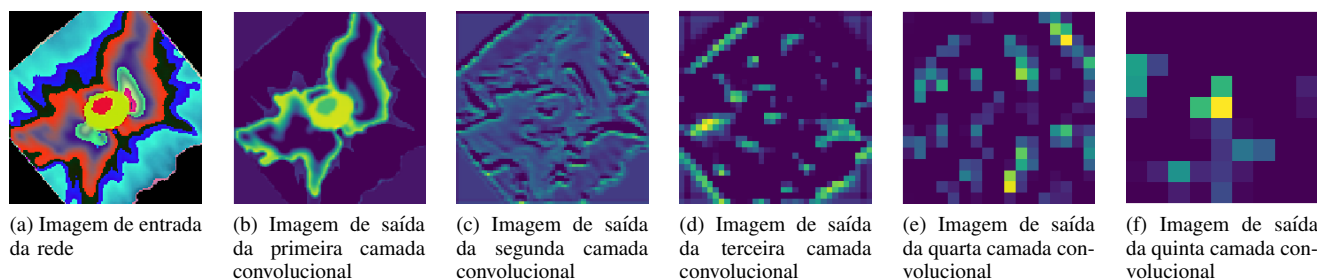


Figura 2. Exemplos de saídas de cada camada convolucional ao longo da rede.

as imagens de espessuras da camada de fibras nervosas. Essas espessuras são representadas como um mapa de calor, onde espessuras próximas à 0 são apresentadas em azul, do verde até o amarelo e laranja são espessuras até 175 microns e de vermelho até aproximar-se do branco espessuras de 175 à 350 microns. O disco central em cinza é a fóvea, ponto cego de onde sai o nervo óptico. Na região B, são exibidas as imagens tomográficas da camada de fibras nervosas em um corte circular em volta da fóvea. A região C exibe gráficos de espessura na região circular ao redor da fóvea, também chamada de mapa TSNIT (Temporal, Superior, Nasal, Inferior, Temporal). Além do gráfico, são exibidos também as médias de espessura por quadrantes e horas de relógio. As medidas de espessuras são comparadas com o banco de dados normativo do aparelho para indicar se a medida está dentro de 95% da população normal (em verde), dentro de apenas 5% da população (em amarelo), ou em apenas 1% da população (em vermelho) [8].

O estudo de Gracitelli *et al.* mostra que os parâmetros de média de espessura por quadrantes ou a média global da espessura da camada de fibras nervosas é amplamente utilizado na análise para diagnóstico do glaucoma [9]. No entanto, nenhum exame fornece resultados definitivos da situação do paciente, sendo necessária a 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 [10]. É possível utilizar classificadores de aprendizagem de máquina para auxiliar na tomada de decisão do oftalmologista. Um classificador de rede neural profunda foi utilizado com dados de exame de campo visual para detectar estágios iniciais de glaucoma [11]. Diferentemente deste estudo, Asaoka *et al.* utilizaram dados do SAP, com uma arquitetura de rede com apenas 4 camadas. Dados de OCT e SAP em conjunto também foram utilizados para treinar diversos tipos de classificadores [12], [13]. Em geral, equipamentos oftalmológicos possuem seus próprios algoritmos para análise e classificação, mas o estudo de Kwokleung Chan *et al.* mostra que algoritmos de classificação são mais eficientes [14].

III. REDES NEURAIS CONVOLUCIONAIS

Redes neurais profundas estão sendo amplamente utilizadas 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, sejam eles imagens ou sons [15]. O principal tipo de rede utilizado é a rede neural convolucional CNN (Convolutional Neural Network). 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 [16].

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. 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 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 uma CNN pode ser vista 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 CNNs [17].

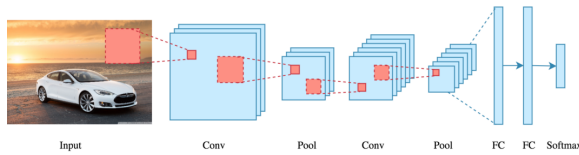


Figura 3. Esquema de camadas de uma rede neural convolucional [18].

A. Transfer Learning

Para treinar uma CNN é necessário ter um conjunto muito grande de exemplos para treinamento e teste. As redes VGG, por exemplo, foram treinadas utilizando o dataset ILSVRC-2012 [19] que contém 1000 classes e foi dividido em 3 conjuntos: 1.3 milhões de imagens para treino, 50 mil imagens para validação e 100 mil imagens para teste [17]. 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 [16].

Uma alternativa para a utilização de CNNs com um dataset menor é a utilização do *transfer learning*. Esta 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, 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 [20].

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.* [21] 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 [22] e degeneração macular [23]. 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 no estudo de Lee *et al.* [23]. Os resultados apresentados 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 CNNs. A rede é composta de 21 camadas com ativações ReLU, sendo 13 camadas convolucionais, 5 max pooling e 3 camadas totalmente conectadas, duas com 4096 saídas e a última com 1000 saídas de classificação. Diferente de outras arquiteturas, o VGG16 utiliza kernels menores nas camadas de convolução, o que diminui o número de pesos necessários mas ainda com um poder discriminativo alto. [17].

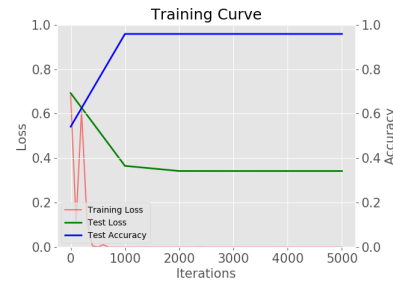


Figura 4. Acurácia e erro de treino e validação no experimento.

IV. EXPERIMENTOS E RESULTADOS

A. Dataset

O dataset original foi obtido com o departamento de oftalmologia da Unicamp. O dataset consiste de imagens de OCT com tamanho 136x136 de 56 olhos com glaucoma e 66 olhos normais, totalizando 122 pacientes. 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.

Para evitar overfitting, foi empregada 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 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 e 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, gerando um vetor de 3 posições com um valor médio por canal.

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

C. Resultados com transfer learning

Neste experimento foi utilizada 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. Os pesos pré-treinados foram carregados para inicialização apenas das camadas convolucionais. As três últimas camadas totalmente conectadas foram inicializadas 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

iterrações, 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 e foi diminuída a cada 1000 iterrações utilizando a equação 2, onde $base_lr$ é a taxa de aprendizagem inicial, γ é um parâmetro definido com o valor 0.1, $iter$ é o número da iterração atual e $step$ é um parâmetro definido como 1000.

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

A validação do modelo foi feita utilizando um dataset de 24 imagens. A acurácia final obtida foi de 95.8%. O gráfico na figura 4 mostra a evolução dos valores de erro e acurácia durante o processo de treinamento. A validação foi feita a cada 1000 iterrações. É possível identificar a estabilização da acurácia após 1000 iterrações, 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 [24].

V. DISCUSSÃO

A rede VGG16 foi escolhida para esse experimento por ter obtido bons resultados na classificação de outras doenças oftalmológicas [22], [23], porém, outras arquiteturas também podem ser utilizadas. Como trabalho futuro será feita a comparação dos resultados obtidos neste trabalho com outras CNNs.

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.

VI. CONCLUSÃO

Neste trabalho foram realizados experimentos com a CNN VGG16 para classificação de pacientes com glaucoma e normais a partir de imagens de espessuras da camada de fibras nervosas. Para diminuir o tempo de treinamento e permitir o uso de um dataset reduzido, foi empregada a técnica de *transfer learning* para inicializar os pesos a partir de uma rede pré-treinada com ImageNet.

Os resultados iniciais mostram que é viável o diagnóstico do Glaucoma a partir das imagens de OCT. Infelizmente os resultados obtidos até o momento mostram um alto erro no dataset de validação. Uma das hipóteses para explicar o resultado é que ocorreu overfitting devido à pequena quantidade de imagens disponíveis no dataset atual. A próxima etapa desta pesquisa será realizar novos experimentos com um conjunto estendido de imagens que está em preparação.

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.
- [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.
- [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.
- [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.
- [7] Carl Zeiss Meditec. *Cirrus HD-OCT*. 2016. Disponível em: <<https://www.zeiss.com/meditec/int/products/ophthalmology-optometry/glaucoma/diagnostics/oct/oct-optical-coherence-tomography/cirrus-hd-oct.html>>.
- [8] AREF, A.; BUDENZ, D. Spectral domain optical coherence tomography in the diagnosis and management of glaucoma. *Ophthalmic Surgery Lasers and Imaging Retina*, Slack Incorporated, v. 41 Suppl, 11 2010. ISSN 2325-8160.
- [9] GRACITELLI, C. P. B.; ABE, R. Y.; MEDEIROS, F. A. Spectral-domain optical coherence tomography for glaucoma diagnosis. *The Open Ophthalmology Journal*, Bentham Open, v. 9, n. PMC4460228, p. 68–77, mar. 2015. ISSN 1874-3641.
- [10] 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.
- [11] ASAOKA, 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.
- [12] 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.
- [13] SILVA, F. R. et al. Sensitivity and specificity of machine learning classifiers for glaucoma diagnosis using Spectral Domain OCT and standard automated perimetry. *Arquivos Brasileiros de Oftalmologia*, scielo, v. 76, p. 170 – 174, 06 2013. ISSN 0004-2749.
- [14] CHAN, T.-W. L. K. et al. Comparison of machine learning and traditional classifiers in glaucoma diagnosis. *IEEE Transactions On Biomedical Engineering*, 2002.
- [15] LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, may 2015.
- [16] 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.
- [17] SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.
- [18] DERTAT, A. *Applied Deep Learning - Part 4: Convolutional Neural Networks*. 2017. Disponível em: <<https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2>>.
- [19] RUSSAKOVSKY, O. et al. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, v. 115, n. 3, p. 211–252, 2015.
- [20] Tan, C. et al. A Survey on Deep Transfer Learning. *ArXiv e-prints*, ago. 2018.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] JIA, Y. et al. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.