

Rede Inception V3 Voltada a Identificação do Glaucoma: Comparação entre Métodos de Otimização

Athyrson M. Ribeiro¹, Francisco de Paula S. Araújo Junior¹

¹Universidade Estadual do Piauí (UESPI)
Av. Nossa Senhora de Fátima, S/N, Bairro de Fátima, CEP: 64202-220

athyrsonmr@gmail.com, pjhatata@hotmail.com

Abstract. *Among the possible uses of Convolutional Neural Networks (RNCs) is the aid in the early diagnosis of several diseases, including glaucoma, a pathology that causes damage to the optic nerve and can lead to permanent loss of vision. Glaucoma is the second most common cause of blindness in the world. In this work five different optimization methods performed in an RNC implemented with the Inception V3 architecture aimed at the identification of glaucoma through images were compared. The results show that, between the evaluated methods, the recently proposed method Adabound achieves the best results during and after the training process.*

Resumo. *Dentre os usos possíveis de Redes Neurais Convolucionais (RNCs) destaca-se o auxílio no diagnóstico precoce de diversas doenças, dentre elas o glaucoma, patologia que provoca danos ao nervo óptico e pode levar a perda permanente da visão. O glaucoma é a segunda doença que mais causa cegueira no mundo. Neste trabalho foi realizada uma comparação entre cinco métodos de otimização aplicados em uma RNC implementada com a arquitetura Inception V3 voltada a identificação de glaucoma através de imagens. É observado que, dentre os métodos testados, o recém proposto método Adabound obteve melhores resultados durante e após o treinamento da rede.*

1. Introdução

Glaucoma é o nome dado a uma classe de doenças que afetam o nervo óptico causando danos. Esta é a segunda patologia que mais causa perda de visão no mundo (Kingman, 2004). A perda de visão causada pelo glaucoma é irreversível, porém com o tratamento precoce, o avanço desta doença pode ser significativamente atrasado. Sob a vista disso, tecnologias que auxiliam na identificação do glaucoma através de imagens do nervo óptico já vem sendo utilizadas por alguns profissionais da área, e grandes bancos de dados vem sendo construídos.

Redes Neurais Convolucionais (RNCs) são soluções estado da arte para a uma enorme quantidade de problemas. Dentre os usos possíveis dessa tecnologia destaca-se o auxílio no diagnóstico precoce de diversas doenças, dentre elas o glaucoma. As RNCs se apresentam em diversas arquiteturas e escolha de parâmetros. Dentre os parâmetros que mais influenciam no desempenho de uma rede neural está a escolha do otimizador, função que tem como objetivo diminuir o erro entre os resultados obtidos por uma rede

em comparação com os resultados desejados. De modo geral, quanto menor o erro obtido por uma rede neural, maior a acurácia dos resultados e conseqüentemente melhor o desempenho da rede.

Esse trabalho almeja avaliar a influência da escolha do otimizador na acurácia de uma RNC voltada a identificação do glaucoma através de imagens do disco óptico, comparando cinco métodos de otimização e verificando qual dentre os métodos testados contribui mais para o aumento da acurácia da rede. Para este trabalho foram escolhidos os seguintes métodos de otimização para comparação de desempenho, conhecidos como Adam, Adagrad, RMSprop, SGD com *momentum* e Adabound.

Uma das principais contribuições deste trabalho é a validação do desempenho do método Adabound (Luo, Xiong, Liu, & Sun, 2019) na classificação de imagens em duas categorias (imagens normais e imagens com glaucoma). Adabound é um método de otimização proposto recentemente, e por este motivo ainda não foi amplamente testado. Esse método é indicado por seus autores como um método que apresenta ótimo desempenho.

2. Referencial Teórico

2.1. Redes Neurais Convolucionais

De acordo com Sousa (2017) uma rede neural artificial (RNA) pode ser definida como um conjunto de unidades de processamento ligadas entre si através de uma grande quantidade de conexões. Assim como nas redes biológicas, as unidades de processamento das RNAs são os neurônios, elementos básicos de sua constituição e as conexões entre os neurônios são conhecidas por sinapses artificiais. Redes neurais convolucionais (RNCs) são uma versão de redes neurais artificiais que possui uma arquitetura capaz de realizar diversos tratamentos nos dados de entrada. Essas adaptações ajudam a rede a obter uma maior eficácia no tempo consumido pela etapa de treinamento e uma maior acurácia na classificação dos dados.

2.1.1. Arquitetura Inception V3

A arquitetura Inception V3, proposta por Szegedy et al. (2015), tem como diferencial a presença de módulos chamados *inception*. Segundo Leonardo e Faria (2017) esses módulos são os extratores de características convolucionais que tem como função aprender com poucos parâmetros representações ricas de informação. Tradicionalmente uma camada convolucional aprende filtros em um espaço 3D por meio das dimensões altura, largura e dimensão de canal. A idéia por trás do bloco Inception é facilitar o processo de mapeamento entre canais e correlações espaciais ao fatorar explicitamente a série de operações que examinam de forma independente essas correlações, ao contrário do modelo tradicional em que um único kernel é responsável por essas tarefas.

2.2. Métodos de Otimização

Segundo Taqi, Awad, Al-Azzo, & Milanova (2018), uma rede neural precisa que as variáveis de cada camada sejam alteradas de forma que esta funcione melhor no processo de classificação. Para esta tarefa é importante que o desempenho da rede seja

continuamente medido através da comparação entre o resultado obtido e o resultado esperado. Sendo o objetivo da otimização minimizar o resultado da função *cross-entropy* (entropia cruzada), que é sempre positivo e se torna zero quando o resultado obtido é exatamente o mesmo resultado esperado.

3. Metodologia

Para o treinamento da RNC utilizada neste trabalho foram utilizadas imagens de cinco bancos de imagens públicos disponibilizados na internet que contém imagens obtidas de pacientes diagnosticados com glaucoma e imagens obtidas de pacientes diagnosticadas sem essa doença. Os bancos de imagens selecionados foram RIM-ONE R1, RIM-ONE R2 e RIM-ONE R3, HRF e Drishti. Os dois primeiros bancos são os únicos que contém apenas imagens do nervo óptico, sendo que o banco RIM-ONE R1 possui 40 imagens diagnosticadas com glaucoma e 118 normais, e o banco RIM-ONE R2 possui 200 imagens diagnosticadas com glaucoma e 255 normais. O banco RIM-ONE R3 contém imagens do olho diagnosticado do paciente e possui 74 imagens diagnosticadas com glaucoma e 85 normais. O banco HRF é composto por imagens do fundo do olho do paciente, e contém 15 imagens diagnosticadas com glaucoma, 15 imagens normais e 15 imagens com retinopatia diabética, que foram desconsideradas para este trabalho. O banco Drishti possui também imagens do fundo do olho do paciente, sendo 70 imagens diagnosticadas com glaucoma e 31 imagens normais.

Neste trabalho todos os arquivos foram armazenados no Google Drive e para o treinamento da RNC foi utilizada a plataforma Google Colaboratory, também conhecida por Colab, um ambiente de desenvolvimento gratuito em nuvem baseado em Jupyter notebooks que conta com aceleração por GPU e integração fácil com o Google Drive. O framework Pytorch 1.1.0, idealizado para aprendizado de máquina, escrito em Python, e com suporte a API CUDA foi utilizado para o treinamento dos modelos. Para o aumento de dados foi utilizada a biblioteca Keras, também escrito em Python, a qual possui funções dedicadas ao processamento de imagens.

Para a avaliação dos métodos de otimização estudados neste trabalho a arquitetura Inception V3 foi selecionada empiricamente. Inception V3 é uma das arquiteturas já pré-implementadas pelo framework Pytorch que apresenta melhor desempenho, de acordo com os dados apresentados na própria documentação do Pytorch. Para o treinamento da rede foram utilizadas apenas imagens do nervo óptico. Foi implementado um script para recortar apenas essa região das imagens dos bancos RIM-ONE R3, HRF e Drishti. Esse processo não foi necessário nos bancos RIM-ONE R1 e RIM-ONE R2, pois ambos contém apenas imagens dessa região. Após o recorte, as imagens dos bancos foram todas reunidas, e organizadas de acordo com sua classificação. Em seguida, para a geração de nova imagens a partir das originais foram aplicadas aleatoriamente distorções como variação do brilho, rotação e espelhamento horizontal e vertical das imagens, utilizando o script para aumento de dados desenvolvido. A base foi aumentada de 1050 para 9335 imagens. Devido ao fato de que as imagens dos bancos foram disponibilizadas em formatos de compressão diferentes, após a aplicação do script de aumento de dados, todas as imagens foram convertidas para o formato JPG e em seguida foram embaralhadas. Finalmente, 75% de todas as

imagens obtidas foram separadas para o treinamento, 20% para validação, e 5% para teste. A quantidade de imagens para cada fase foi decidida empiricamente.

Antes de serem enviadas para o treinamento as imagens passaram por um pré-processamento disponibilizado pelo framework Pytorch, no qual as imagens foram redimensionadas para 299 pixels, que é o tamanho de imagem com o qual o modelo Inception trabalha, e em seguida foi realizada uma normalização com média [0.485, 0.456, 0.406] para os canais vermelho, azul e verde respectivamente e desvio padrão de [0.229, 0.224, 0.225] para respectivamente os mesmo canais de cores, em todas as imagens. O valor para cada canal foi escolhido empiricamente.. A rede neural convolucional Inception V3 utilizada para o treinamento foi implementada com as configurações padrões disponibilizadas pelo framework Pytorch.

4. Resultados

Assim como no trabalho de Taqi et al. (2018), cada sessão de treinamento rodou por 200 épocas. Os testes com cada otimizador foram realizados utilizando 0.001 como tamanho de passo inicial, valor escolhido empiricamente. Com exceção deste, todos os parâmetros padrões de cada otimizador foram mantidos. Pode ser observado nos gráficos das figuras 4.1 que o otimizador Adabound foi o que obteve os melhores resultados durante a maioria das épocas. Sendo também o método que demonstrou o desempenho mais estável, ao contrário dos métodos Adagrad e RMSprop que apresentaram desempenho ótimo em algumas épocas, porém em muitas outras tiveram um desempenho inferior àqueles obtidos pelos próprios em épocas recentes anteriores.

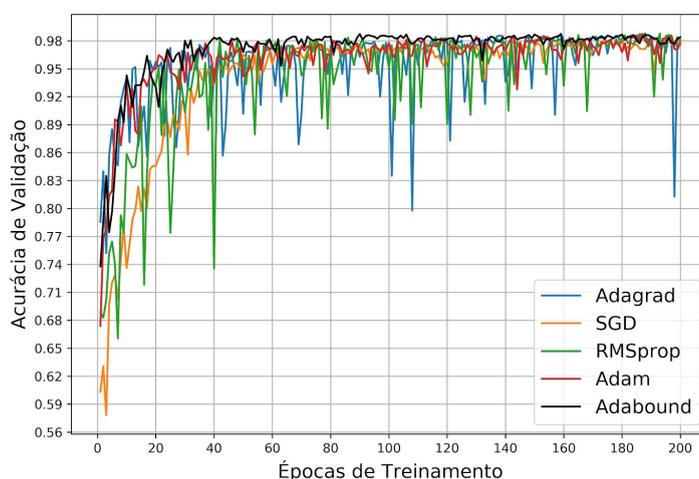


Figura 4.1. Gráfico com o histórico da acurácia de validação ao longo do treinamento da rede para cada algoritmo. Fonte: Autoria Própria.

Após cada modelo da rede ser treinado por 200 épocas foi realizado um teste de classificação para cada modelo, na qual se utilizaram as 535 imagens que haviam sido separadas anteriormente para teste. Através dos resultados obtidos foi possível gerar uma matriz de confusão para cada modelo, que é uma tabela que relaciona os resultados esperados com os resultados obtidos pelo modelo, após treinamento, ao classificar as imagens reservadas para validação. Com a matriz de confusão foi possível extrair

métricas como a acurácia (ACC), sensibilidade (SEN), especificidade (ESP), valor preditivo positivo (VPP) e negativo (VPN) para cada otimizador na classificação de imagens com glaucoma e normais. A Tabela 1 contém os valores obtidos por cada otimizador (Os melhores resultados para cada métrica estão em negrito).

Pode-se notar ao examinar a Tabela 1 que o otimizador Adabound teve os melhores resultados nas métricas acurácia, especificidade e valor preditivo positivo. Esse otimizador foi ainda o melhor em classificar corretamente imagens com glaucoma, superando o otimizador Adam, um dos otimizadores mais utilizados atualmente no treinamento de redes neurais. O otimizador Adam também alcançou ótimos resultados em todos os avaliadores de desempenho. Adagrad e SGD, a contrário, obtiveram os piores resultados em todas as métricas de avaliação, ao passo que o método RMSprop obteve o melhor desempenho na classificação correta de imagens normais.

Tabela 1. Desempenho de cada otimizador após 200 épocas de treinamento

	ACC	SEN	ESP	VPP	VPN
SGD	96,1%	96,2%	96,0%	96,5%	95,6%
RMSprop	98,3%	99,0%	97,6%	97,9%	98,8%
Adagrad	96,6%	97,9%	95,2%	95,9%	97,5%
Adam	98,5%	98,3%	98,8%	98,9%	98,0%
Adabound	98,7%	98,3%	99,2%	99,3%	98,0%

A Figura 4.1 apresenta o gráfico comparativo de acurácia de validação. Porém optou-se agora por exibir apenas o desempenho dos otimizadores que obtiveram os melhores resultados nas métricas da Tabela 1. Pode-se verificar com esse gráfico que os otimizadores Adam e Adabound apresentaram um desempenho mais estável que o método RMSprop. Tendo este apresentado um desempenho inferior aos demais durante grande parte das épocas de treinamento. Esse gráfico também deixa claro o desempenho ligeiramente superior do otimizador Adabound em relação aos demais, ao longo de todo o treinamento.

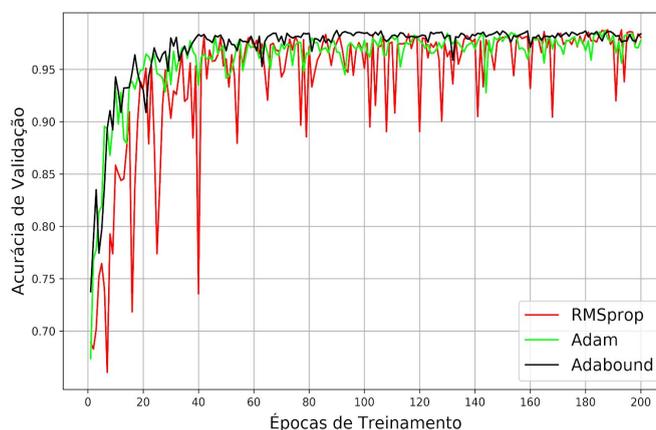


Figura 4.2. Gráficos da Acurácia de Validação para os Otimizadores Adam, RMSprop e Adabound. Fonte: Autoria Própria.

5. Considerações Finais

A partir dos resultados obtidos por esse trabalho pode-se verificar que a escolha do método de otimização tem um grande impacto no comportamento da rede. Verifica-se que o método Adabound, o mais recentemente proposto dentre os métodos testados neste trabalho, se destacou positivamente, tanto em estabilidade como em acurácia. Não se pode afirmar, apenas com os resultados obtidos neste trabalho, qual o melhor método de otimização para treinamento de RNCs. Alguns métodos são mais adequados para lidar com alguns tipos de dados do que outros. Entretanto, pode-se concluir que o método Adabound tem grande potencial para problemas envolvendo classificação de imagens. Os resultados obtidos neste trabalho podem servir, não apenas como uma base para futuros algoritmos de detecção de glaucoma através de imagens, mas também como um pontapé inicial na construção de modelos de RNCs.

Referencias

- Hinton, G., Srivastava, N., & Swersky, K. (2012). Neural networks for machine learning lecture 6a overview of mini-batch gradient descent. Cited on, 14, 8.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Kingman, S. (2004). Glaucoma is second leading cause of blindness globally. Bulletin of the World Health Organization, 82, 887-888.
- Leonardo, M. M., & Faria, F. A. (2017). Um Sistema de Reconhecimento de Espécies de Moscas-das-Frutas. Revista Eletrônica de Iniciação Científica em Computação, 17(2).
- Luo, L., Xiong, Y., Liu, Y., & Sun, X. (2019). Adaptive gradient methods with dynamic bound of learning rate. arXiv preprint arXiv:1902.09843.
- SOUSA, F. M. M. M. D. (2017). Uma Otimização do Perceptron Multicamadas Utilizando Aproximação Quadrática e Método de Ponto Proximal.
- Sutskever, I., Martens, J., Dahl, G., & Hinton, G. (2013). On the importance of initialization and momentum in deep learning. In International conference on machine learning (pp. 1139-1147).
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).
- Taqi, A. M., Awad, A., Al-Azzo, F., & Milanova, M. (2018). The impact of multi-optimizers and data augmentation on TensorFlow convolutional neural network performance. In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (pp. 140-145). IEEE.