

Otimizando o diagnóstico automatizado de glaucoma a partir de imagens de fundo de olho

Lucas Mayer Ceschini¹, Lucas Micol Policarpo¹,
Vinicius Facco Rodrigues¹, Rodrigo da Rosa Righi¹, Gabriel de Oliveira Ramos¹

¹PPG em Computação Aplicada – Universidade do Vale dos Sinos - São Leopoldo/RS

ceschini.lucas@gmail.com, {lmpolicarpo, vfrdrigues, rrrighi, gdoramos}@unisinos.br

Resumo. *Glaucoma é a principal causa mundial de perda irreversível de visão. Afim de viabilizar a implantação de uma ferramenta de diagnóstico de glaucoma para a clínica médica, um trabalho base foi selecionado e otimizado. Ao unificar duas redes de segmentação reduzimos o tempo de processamento em 24,24%, e adicionando uma segunda rede de classificação direta aumentamos a sensibilidade do modelo em 3%, em comparação com o trabalho base.*

1. Introdução

Glaucoma é uma doença ocular responsável pelo maior número de perda irreversível de visão do mundo, sendo a segunda maior causa de cegueira, ficando atrás somente da catarata [Mantravadi and Vadhar 2015]. Os tipos mais comuns de glaucoma são o glaucoma de ângulo aberto e o glaucoma de ângulo fechado. O glaucoma de ângulo fechado é precedido de um súbito aumento na pressão intraocular, causando uma dor aguda. Já o glaucoma de ângulo aberto se desenvolve de forma silenciosa e indolor, apresentando como primeiros sintomas a perda da visão periférica, em um estágio da doença onde já não há mais volta [Mantravadi and Vadhar 2015]. Por conta disso, este tipo de glaucoma é o mais comum e o mais perigoso, sendo responsável por até 80% dos casos [Mantravadi and Vadhar 2015].

A avaliação de imagens de fundo de olho obtidas a partir de ferramentas de fundoscopia possibilita a identificação da escavação responsável por ampliar o ângulo do copo do disco óptico, a característica mais comum do glaucoma de ângulo aberto [Ahn et al. 2018]. É possível classificar imagens de fundo de olho como normais ou glaucomatosas de forma automatizada a partir de modelos de Deep Learning. Porém, normalmente estes modelos possuem arquiteturas robustas e complexas, que necessitam de um elevado poder computacional para convergirem. Clínicas médicas dificilmente terão os equipamentos necessários para executar estes sistemas.

A proposta de [Civit-Masot et al. 2020] apresenta modelos mais leves, que tem como propósito serem implantados na clínica médica. Além disso, tem o intuito de auxiliar o médico, em conjunto com abordagens de Inteligência Artificial (IA) explicável para aumentar a confiança do usuário no sistema. Este trabalho propõem uma remodelagem na arquitetura de [Civit-Masot et al. 2020]. Tendo como objetivo melhorar as previsões da rede de classificação, através da adição de uma segunda rede, igualmente leve. E também, simplificar a arquitetura, unificando as redes de segmentação do copo e disco, tornando-a mais leve e rápida para uma implantação embarcada.

2. Proposta

A arquitetura do modelo base consiste em uma rede de classificação direta baseada em MobileNetV2 [Sandler et al. 2018], e duas redes de segmentação, uma para o disco e outra para o copo do olho, ambas baseadas em um modelo U-net [Ronneberger et al. 2015] modificado. A arquitetura proposta neste estudo consiste em uma rede de classificação baseada em *Ensemble Learning* das arquiteturas MobileNetV2 e EfficientNet, e uma rede de segmentação unificada para o disco e copo. Ambas arquiteturas e o fluxo de execução são apresentadas na Figura 1. As cores em destaque representam as áreas correlatas entre os dois modelos.

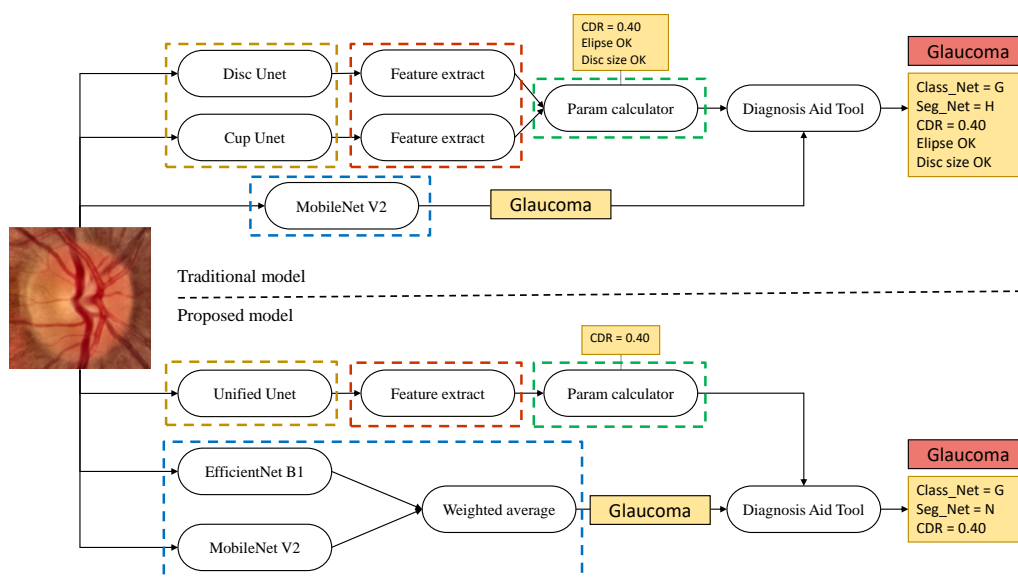


Figura 1. Arquitetura do modelo base e do modelo proposto.

As contribuições deste estudo são: (i) a adição de uma segunda rede de classificação leve, EfficientNetB1 [Tan and Le 2019], que é agrupada com a rede já existente e que resulta em um aumento de 3% na sensibilidade, reduzindo assim expressivamente a quantidade de falsos negativos, perigosos para o ambiente clínico; e (ii) a união de duas redes de segmentação, uma para o disco e outra para o copo, resultando na mesma eficácia porém simplificando o modelo, indo de quatro milhões de parâmetros para dois milhões, assim reduzindo o tempo de diagnóstico.

3. Experimentos e Resultados

Nesta seção, são descritos os experimentos realizados e os resultados obtidos em cada etapa do processo de implementação, detalhando alguns processos de implementação, métricas e hiper-parâmetros utilizados. A adição da segunda rede de classificação e seu posterior agrupamento com a rede original tem o propósito de melhorar as métricas de acurácia, *Area Under the Curve* (AUC) e principalmente sensibilidade. Seguindo para a união das redes de segmentação, onde o objetivo é simplificar a arquitetura e tornar o armazenamento mais leve e o processamento mais ágil.

Afim de melhorar as previsões da rede de classificação, foi proposto adicionar uma segunda rede, igualmente leve e com o mesmo propósito de ser aplicável em um

sistema embarcado. Com isso, a partir da sugestão presente no trabalho base, foi selecionado a EfficientNetB1 [Tan and Le 2019], Composta de pouco menos que oito milhões de parâmetros, é considerado um modelo leve e capaz de prover resultados tão eficazes quanto outras arquiteturas mais robustas, por exemplo, VGG ou Inception. A rede foi treinada com os mesmos hiper-parâmetros que o classificador do artigo base, e seus resultados se mostraram promissores, pois mesmo não aumentando a AUC, aumentou a sensibilidade de 0,9140 para 0,9462. A Tabela 1 compara os resultados. Isso quer dizer que o número de casos falsos negativos foi expressamente reduzido, mesmo com um aumento no número de falsos positivos, ainda assim a contribuição é válida, pois falsos negativos acarretam em perda irreversível de visão, já falsos positivos acarretam em exames adicionais desnecessários.

Tabela 1. Métricas MobileNetV2, EfficientNetB1 e do seu agrupamento

Métrica	MobileNetV2	EfficientNetB1	Modelo Agrupado
AUC	0,9567	0,9394	0,9512
Acurácia	0,8897	0,8345	0,8897
Especificidade	0,8461	0,6538	0,7884
Sensitividade	0,9140	0,9355	0,9462

Com intuito de simplificar a arquitetura e torná-la mais leve para uma possível implantação embarcada, as redes de segmentação foram unificadas, para segmentar tanto o disco como o copo. Devido ao aumento de complexidade, a rede precisou de mais épocas durante o treinamento para atingir os mesmos resultados, indo de 100 para 300. Ainda assim, é evidente que uma rede unificada é capaz de alcançar os mesmos resultados que as redes de segmentação binárias, porém de uma forma mais ágil. A Tabela 2 apresenta um comparativo entre as métricas das redes tradicionais e da rede unificada. Foram medidas as métricas DICE e Intersection over Union(IoU). Percebe-se os resultados comparáveis porém removendo completamente uma das redes neurais, com isso, foi possível reduzir o custo de armazenamento e processamento do subsistema de segmentação pela metade. Isso pode se provar valioso para aqueles interessados em implantar o sistema em computadores com configurações medianas.

Tabela 2. Métricas da segmentação do disco + copo e rede unificada

Métrica	Rede disco + Rede copo	Rede unificada
IoU	0,9382	0,9455
DICE	0,9671	0,9502

Por último, foi realizado uma análise de desempenho entre as duas arquiteturas. Ambas foram executadas de fim a fim com as mesmas entradas e dispondo do mesmo hardware, um servidor com processador i7 9750hz, placa de video nvidia 1660TI com 6GB de memória integrada, e memória RAM de 2666mhz e com 8GB. Foram feitas 50 execuções dos dois modelos para obter resultados com alta confiabilidade. A Tabela 3 apresenta os resultados obtidos. As cores no nome da coluna são referentes as regiões destacadas na Figura 1. Como visto na Tabela 3, o modelo proposto possui um menor tempo médio para geração de diagnóstico, majoritariamente por conta da simplificação da Unet. Ao unificar as redes de segmentação, foi possível reduzir o tempo de processamento

do subsistema de segmentação em 24,24%. Por outro lado, o *ensemble* das duas redes de classificação direta aumentaram o tempo de processamento do subsistema de classificação em três segundos. Porém isso não é crítico, pois o gargalo do processamento são as redes de segmentação, e ao reduzir o seu tempo, o tempo de processamento total também foi reduzido. Além do mais, o processo de treinamento é muito mais demorado nas redes de segmentação se comparado com a rede de classificação, aumentando ainda mais o valor dessa redução.

Tabela 3. Comparativo do tempo de execução das arquiteturas

	Unet (s)	Feature extract (s)	Param calculator (ns)	Mobile Net (s)	Tempo total (s)
Modelo tradicional	31.656	0,0165	700	1.633	33.699
Modelo proposto	20.504	0,0281	800	4.911	25.605

4. Conclusão

As contribuições deste trabalho foram a adição de uma segunda rede de classificação e a união das redes de segmentação. A primeira contribuição foi capaz de aumentar a sensibilidade, reduzindo ainda mais o número de falsos negativos e assim garantindo ainda mais valor para seu uso na clínica. A segunda contribuição fez com que a arquitetura fosse simplificada, ocupando menos espaço de armazenamento e utilizando menos recursos de processamento, sem perder acurácia, tendo como única consequência a necessidade de mais tempo de treinamento inicial. O resultado final é um modelo que mantém a precisão da proposta tradicional, ocupa menos espaço em disco, tornando-se viável a utilização em uma ferramenta embarcada, e que também possui um menor tempo de execução e geração de diagnóstico.

Referências

- Ahn, J. M., Kim, S., Ahn, K.-S., Cho, S.-H., Lee, K. B., and Kim, U. S. (2018). A deep learning model for the detection of both advanced and early glaucoma using fundus photography. *PLoS ONE*, 13.
- Civit-Masot, J., Domínguez-Morales, M. J., Vicente-Díaz, S., and Civit, A. (2020). Dual machine-learning system to aid glaucoma diagnosis using disc and cup feature extraction. *IEEE Access*, 8.
- Mantravadi, A. and Vadhar, N. (2015). Glaucoma. *Primary Care - Clinics in Office Practice*, 42(3).
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520.
- Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *36th International Conference on Machine Learning, ICML 2019*, 2019-June.