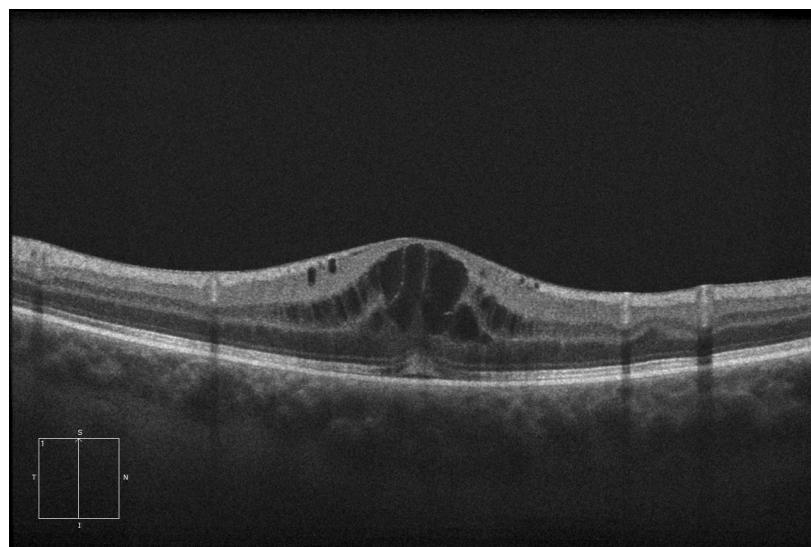


PRÁCTICA II

Pablo Chantada Saborido & José Romero Conde



Índice

1. Introducción	3
2. O noso formulamento	6
3. Conclusiós	8

1. Introducción

Esta práctica supuxo todo un reto para nós. O conxunto de datos foi especialmente difícil polos seguintes motivos:

- **Escaseza de datos.** Acostumados a miles (MNIST) ou centos (SmartPorts) de imaxes, contar con só decenas delas supuxo unha dificultade no problema a resolver. Isto débese a que a nosa rede ten que aprender moito de cada imaxe, e saber extrapolalar o aprendido a imaxes que nunca viu. Nada sínxelo.
- **Alta variabilidade.** Se foran poucos datos pero a realidade fose sempre moi parecida, non habería tanto problema, a cuestión é que dunha imaxe a outra pode haber pouco que teñan en común. As hai que a parede ocular é fina e ten unha fendedura, as hai sen fendedura mais cun gran vaso, as hai sen fendedura e con múltiples vasos... isto é para nós un problema porque o fluxo óptico *existe* dun xeito distinto en cada imaxe de OCT. Polo tanto, a nosa rede debe aprender todas esas variacións con poucos exemplos.
- **Imperfección da supervisión.** Aínda que poderíamos ter feito un AutoEncoder para ciorarnos dunha boa representación *latente*, limitámonos ao uso das máscaras para propagar o sinal de erro. Ao non seren perfectas e consistentes as etiquetas (zonas dun certo nivel de gris rodeadas dun capilar, en unhas máscaras representábase o capilar e noutras non), non podemos esperar que o noso algoritmo o sexa. Ademais, está baseándose exclusivamente en exemplos mentres que un médico razoa e delibera en base aos seus coñecementos teóricos e de dominio. A nosa rede non pode facer tal cousa.

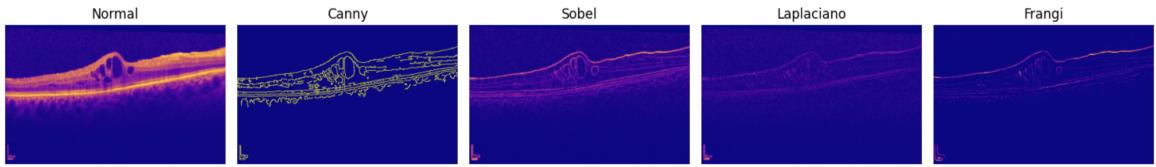


Arriba pode verse un exemplo que nos chamou a atención. Na área rodeada cun círculo violeta, na imaxe orixinal pódense ver tres *paredes*, en cambio o GT só respectou a da dereita de todo e a do medio suixerina pero non a rematou, e a da esquerda, o GT nin a asomou. En cambio, a nosa rede si tivo en conta a primeira parede. Este exemplo mostra que o GT non é perfecto e que perseguiño pode ofrecer bons resultados pero só ata un punto. Por outro lado, no hospital cando for a usarse a rede, igual non importa se segmenta ben esa parede, nós non podemos sabelo pero unha revisión dun blog médico [3] suxire que probablemente ese nivel de detalle non sexa importante.

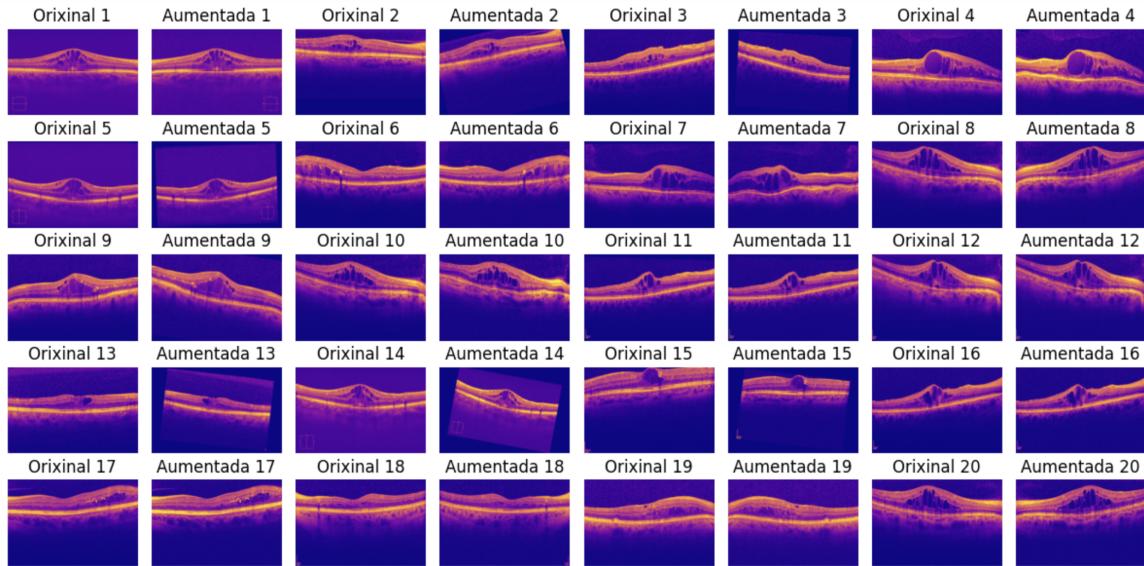
Para afrontar o problema, por tanto, armámonos cunha serie de técnicas e trucos aprendidos nesta asignatura, en Aprendizaxe Profundo e mais en Principios de Visión por Computador. Son os seguintes:

- **Canles adicionais de entrada.** A rede ten que aprender que é o fluxo óptico desde cero, sen saber que é un bordo ou un círculo. Non lle pasa o mesmo aos médicos, que cando empezan na oftalmoloxía xa teñen un adestrado sistema de percepción visual. Polo tanto, para axilizar este proceso, ademais da imaxe en branco e negro \mathcal{I} (unha canle) decidimos acompañala dos $\phi_i(\mathcal{I})$ onde os ϕ_i son algoritmos de procesado de imaxe. Inicialmente escollemos Canny [1], Sobel, Laplaciano e Frangi [2]. Do último tiñamos altas esperanzas por estar tamén orientado ao

ámbito médico mais áinda despois dunha considerable procura de hiperparámetros decidimos descartalo para, finalmente, só quedarnos con Canny. Unha xustificación desta decisión é a imaxe de abaxo.

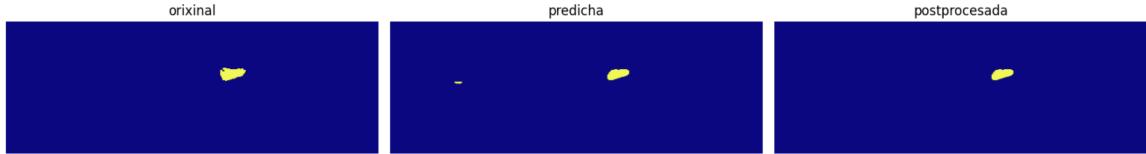


- **Aumento de datos.** Fundamentalmente baseámonos nas suxestións do propio artigo da UNet [6], é dicir: transformacións afíns e elásticas. Coas primeiras foi relativamente doado atopar hiperparámetros apropiados pero as segundas foi máis difícil, en cambio (ao estaren deseñadas para un contexto de segmentación médica) ofreceron moi bons resultados. Finalmente atopamos $\alpha = 500$ e $\sigma = 20$ axeitados. Ademais, para estas dúas transformacións, ao seren relativamente *agresivas*, atopamos que é mellor non transformar sempre (para darréllas á rede uns poucos exemplos inalterados e aprenda deles). En concreto aplicamos cada unha das delas cunha probabilidade de 0.7, polo tanto, para unha imaxe, a probabilidade de ser alterada por ambas transformacións é de ≈ 0.5 . Ademais destas dúas, aplicamos (malia observar pouca diferenza) volteos horizontais, deformacións na cor (*color jitter*), variacións no enfoque e ruído gaussiano aditivo. O efecto da totalidade das transformacións sobre un subconjunto das imaxes do conxunto de datos vese abaxo.



Se ben pode parecer *leve*, aumentos de datos más agresivos non vían a luz da converxencia. Podemos dicir que a configuración de hiperparámetros do aumento de datos está fortemente baseada na experimentación. Comentar que tamén probamos a recortar as imaxes centrando a máscara como forma de aumento de datos pero encontramos que empeoraba o rendemento; ese tipo de cousas poden ser útiles para tarefas de clasificación pero como a segmentación depende da escala en concreto e no conxunto de datos sempre era relativamente a mesma escala, facer grandes variacións nese sentido pode (como temos visto) empeorar o rendemento da rede.

- **Posprocesado das máscaras.** Aínda que non diferenciables, e por tanto non contribuían ao sinal de erro do adestramento do algoritmo, decidimos aplicar unha serie de transformacións sobre as máscaras para achegarse más ás reais. Pulindo detalles de xeito que, se un oftalmólogo tivera que usar o noso sistema, certas impurezas corrixibles non o molestarián. En concreto, axudámonos do operador morfolóxico de peche.

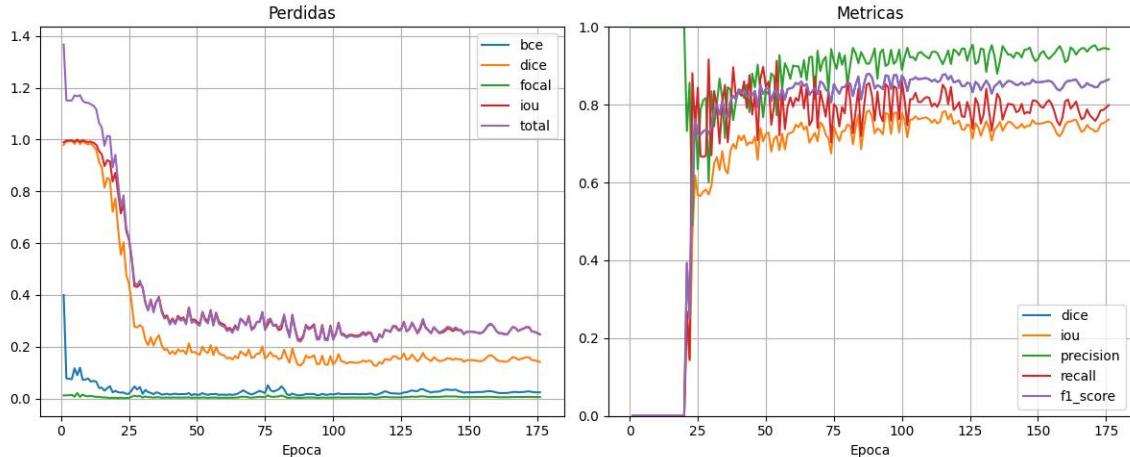


Comentar, non obstante, que por non ser o obxectivo da práctica o resultado final senón o adestramento da rede en si mesma, limitámonos a unha primeira idea de só aplicar peche, mais de seguro poderían implementarse máis e mellores alternativas.

- **Trucos no adestramento da rede.** Para asegurar converxencia e acelerar o adestramento usamos Dropout [7] como regularización e AdamW [5] como optimizador que tamén implementa regularización a través dos pesos. Estes dous elementos permitiron que se propagase axeitadamente o sinal de erro na rede, que nalgúns experimentos fixemos moi grande. Adicionalmente, adicando un subconjunto dos datos a *validación*, implementamos *EarlyStopping* con paciencia na perda en validación e máis un xestor do paso de aprendizaxe que, tamén, tiña paciencia coa perda en validación. Estes dous componentes fixeron dinámicos parámetros que, de seren estáticos e sempre iguais, non permitirían aprender con éxito ás redes porque as redes más grandes precisan (en xeral) máis épocas, e un único paso de aprendizaxe pode estar ben pero non adoita explotar todo o potencial dunha rede dada. Como última consideración, implementamos unha función de perda combinada, do seguinte xeito:

$$\mathcal{L} = \alpha_{BCE} \times \mathcal{L}_{BCE} + \alpha_{FOCAL} \times \mathcal{L}_{FOCAL} + \alpha_{DICE} \times \mathcal{L}_{DICE} + \alpha_{IOU} \times \mathcal{L}_{IOU}$$

onde nós fixamos $\alpha_{BCE} = 0,7$, $\alpha_{FOCAL} = 0,2$, $\alpha_{DICE} = 0,4$, $\alpha_{IOU} = 0,7$. Eses coeficientes fixáronse a partir de observar as primeiras gráficas de adestramento. Abaixo pódese ver unha gráfica coa función de perda aquí descrita.



2. O noso formulamento

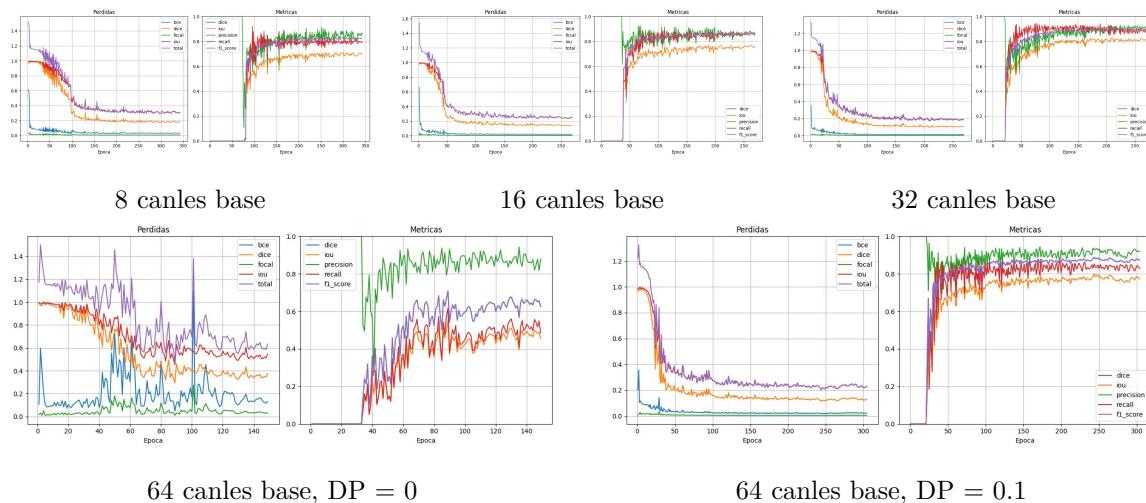
Despois de ter feito a práctica anterior, recoñecemos as vantaxes dun bo formulamento inicial, tanto no nivel conceptual e máis de *visión* como no nivel técnico e máis de *programación*. É por isto que desde un primeiro momento formulamos a práctica como un todo, por tanto, non fixemos primeiro un *Baseline* e logo experimentos e melloras sobre iso; en vez diso, desde un primeiro momento formulamos na parte de programación a interface necesaria para, logo, facer experimentos e atopar que combinación ofrecen os resultados. Comentamos agora, aspectos nos que fixemos probas e as nosas conclusións:

- **Tamaño das imaxes.** As primeiras redes que adestramos (poderíamos consideralas en parte, *baselines*) empregaban imaxes relativamente pequenas (por custo computacional) e moi rectangulares (200 píxeles por 500), porque observando as imaxes vimos que tiñan un ratio moi pronunciado. Isto pódese comprobar cos seguintes comandos de UNIX (no directorio das imaxes):

```
$ identify * | awk '{print $3}' | awk -F'x' '{suma += $1/$2; n++} END {print
"Razón ancho-alto media:", suma/n}'
>> Razón ancho-alto media: 2.6163
```

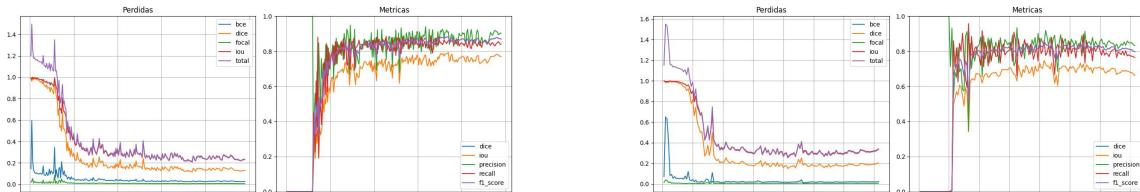
Posteriormente decatámonos de que para que a rede percibira propiamente as paredes do Fluxo Óptico, estas non podían ter de ancho 2 ou 3 píxeles, entón, malia seren más computacionalmente custosos, os modelos adestrados con imaxes más grandes ofreceron mellores resultados. Tamén decatámonos de que malia seren as imaxes moi rectangulares, a máscara sempre estaba no centro e non pagaba a pena o bordo, entón, malia saber que as imaxes eran moi rectangulares, a nosa decisión final foi de 400 por 500.

- **Profundidade e canles base da UNet.** Despois da lectura do artigo [6], decatámonos de que a UNet, máis que un modelo en concreto, é unha concepción que pode instanciarse no xeito que o fixeron os autores, é dicir, 3 niveis de profundidade 64 canles de saída na primeira convolución, pero que facilmente pode mutar a outras configuracións. Abaixo probamos a adestrar a rede coa mesma configuración variando só as canles de entrada (4 primeiras imaxes) e vemos que, (como se suxeriu xa no 2012 [4]) canto máis grande é a rede mellores resultados. Tamén aproveitamos para mostrar o efecto do Dropout [7] (recomendado no artigo orixinal [6]), que é o único que varía nas dúas imaxes de abaxo.



Os mesmos resultados poden dicirse da profundidade, ademais, a profundidade ten que ver co tamaño do campo receptivo, o cal tamén encontramos benficioso.

- **Aumento de datos.** Definir o aumento de datos como o temos feito tamén pode considerarse unha extensión sobre o *baseline*. Abaixo móstranse dous adestramentos da mesma rede (64 canles base, 3 niveis de profundidade na UNet e Dropout de 0.05) e pódese ver que o aumento de datos esencialmente permite adestramentos máis longos, nos que a aprendizaxe ata converxencia sexa maior.



Con aumento de datos

Sen aumento de datos

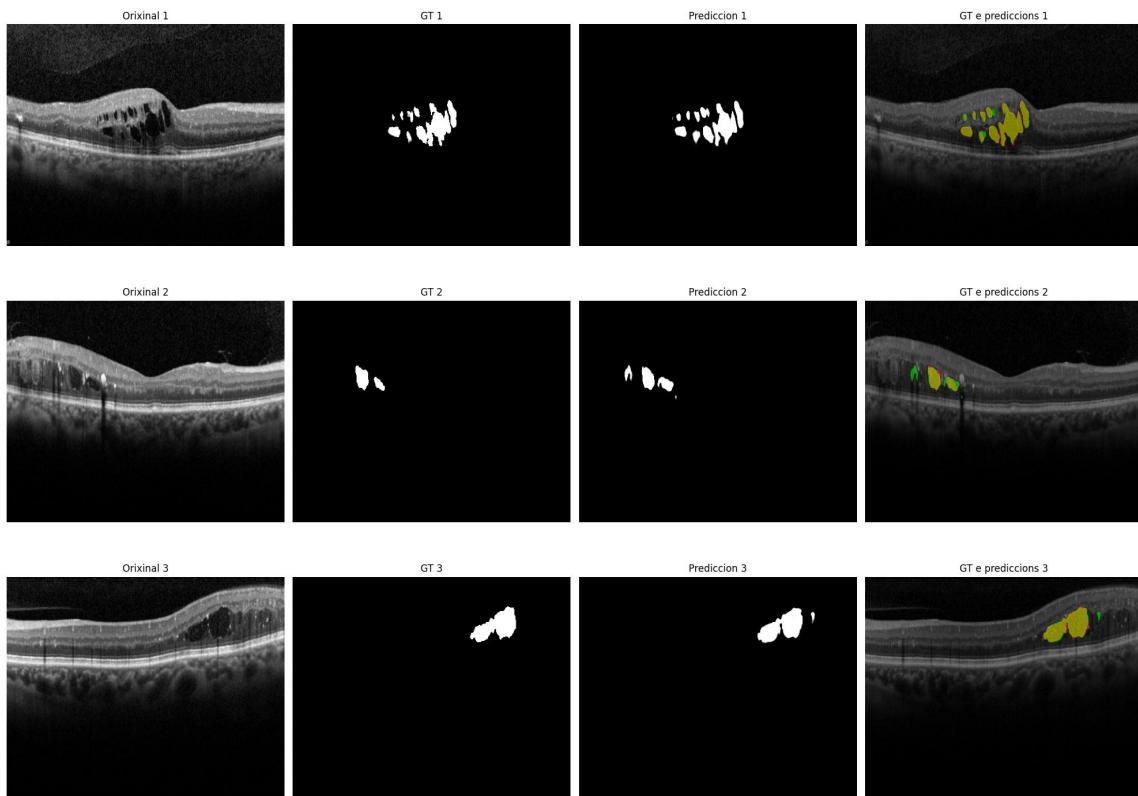


Figura 1: Resultados dunha execución con aumento de datos, 32 canles iniciais, 4 niveis de profundidade, ningún Dropout e un DICE en proba de 0.9 (Para proba usáronse 5 imaxes, tendo destinadas outras 5 a validación e 40 ao adestramento).

3. Conclusións

En primeiro lugar, o problema da segmentación en imaxes médicas presenta desafíos específicos que requieren dunha aproximación diferente dos casos aos que estamos acostumados. A escaseza de datos, a alta variabilidade entre imaxes e a imperfección das etiquetas foron os principais obstáculos que tivemos que afrontar.

A arquitectura UNet demostrou ser moi eficaz para esta tarefa, especialmente cando se configura adecuadamente en termos de profundidade e número de canles. Nos nosos experimentos, puidemos comprobar que (por desfortuna para a nosa limitada capacidade de cómputo) as redes máis grandes obteñen mellores resultados, o cal está en liña co observado noutros traballos do campo.

O aumento de datos foi unha técnica fundamental para mellorar o rendemento do modelo, especialmente as transformacións elásticas, que xustamente foron propostas para segmentación de imaxes médicas ca UNet. Esta técnica permitiuños estender o conxunto de adestramento e mellorar a capacidade de xeneralización da rede ante variacións que sabemos que son comúns en texidos biolóxicos.

A combinación de diferentes funcións de perda proporcionou unha supervisión máis certa e permitiu obter mellores resultados que utilizando unha única métrica. Unha investigación más exhaustiva podería usar xestores dos hiperparámetros α_i co tempo, pero os nosos resultados, para o tempo e alcance da práctica, resultaron ser suficientes.

A incorporación de canles adicionais de entrada, como o filtro Canny, permitiu á rede aprender características relevantes do fluxo óptico sen necesidade de partir desde cero, o que acelerou o proceso de aprendizaxe e mellorou os resultados finais. En outras palabras, permitiuños non ter que empezar de cero.

En definitiva, este traballo demostra que as redes neuronais profundas, especificamente a arquitectura UNet con modificacións apropiadas, poden acadar resultados moi satisfactorios na segmentación do fluxo óptico en imaxes OCT, mesmo en escenarios con datos limitados e alta variabilidade. Os nosos mellores modelos acadaron un coeficiente DICE de 0.9 no conxunto de proba, o que suxire un alto grao de concordancia coas anotacións médicas, sobretodo considerando que seguramente un médico non consiga un 1 (e posiblemente un oftalmólogo de prácticas consiga menos ca nos) e mostra o potencial desta aproximación para aplicacións clínicas reais.

Referencias

- [1] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.
- [2] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever. Multiscale vessel enhancement filtering. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI'98*, pages 130–137. Springer, 1998.
- [3] <https://eyeguru.org/essentials/interpreting-octs/>.
- [4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 25, pages 1097–1105. Curran Associates, Inc., 2012.
- [5] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2019.
- [6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Springer, 2015.
- [7] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.