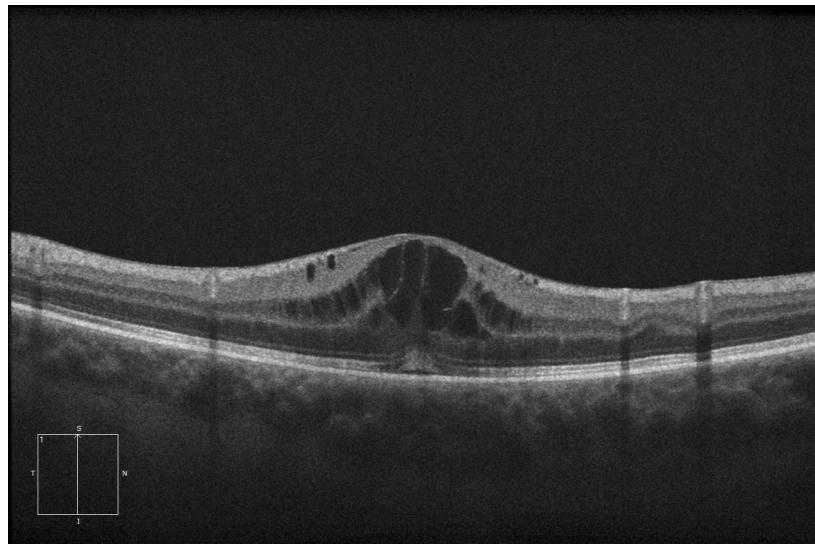


PRÁCTICA II

Pablo Chantada Saborido & José Romero Conde 430



Índice

1. Introdución	3
2. O noso planteamento	6
3. Conclusiόns	7

1. Introdución

Esta práctica supuso todo un reto para nos. 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 extrapolar o aprendido a imaxes que nunca viu. Nada sixelo.
- **Alta variabilidade.** Se foran poucos datos pero a realidade fosse sempre moi parecida, non habería tanto problema, a cuestión é que de unha imaxe a outra pode haber pouco que teñan en común. As hai que a parede ocular é fina e ten unha fendidura, as hai sen fendidura mais con un gran vaso, as hai sen fendidura e con múltiples vasos... esto é para nos 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.
- **Imperfectude da supervisión.** Aínda que poderíamos ter feito un AutoEncoder para cencionarnos de unha boa representación *latente*, limitámonos ó uso das máscaras para propagar a sinal de error. Ó non seren perfectas e consistentes as etiquetas (zoas 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 mentras que un médico razona e delibera en base os seus coñecementos teóricos e de domino. A nosa Rede non pode facer tal cousa.

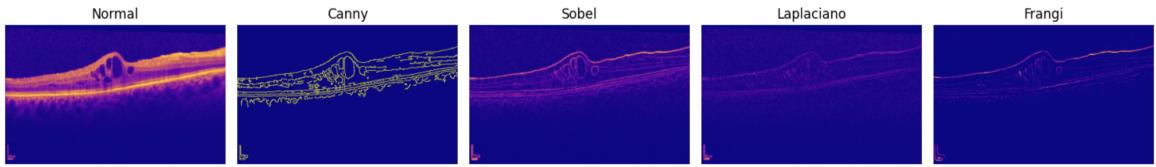


Arriba pode verse un exemplo que nos chamou a atención. Na area rodeada cun circulo violeta, na imaxe orixinal pódese ver tres *paredes*, en cambio o GT só respetou á da dereita de todo e a do medio a suixeriu pero non a remetou e a da esquerda, o GT nin a asomou. En cambio a nosa Rede, sí tivo en conta a primeira parede. Este exemplo mostra que o GT non é perfecto e que, perseguiilo pode ofrecer bons resultados pero só ate un punto. Por outro lado, no hospital cando fora a usarse a Rede, igual non importa se segmenta ben esa parede, nos non podemos sabelo pero una revisión dun blog médico [?] suxire que probablemente ese nível de detalle non sexa importante.

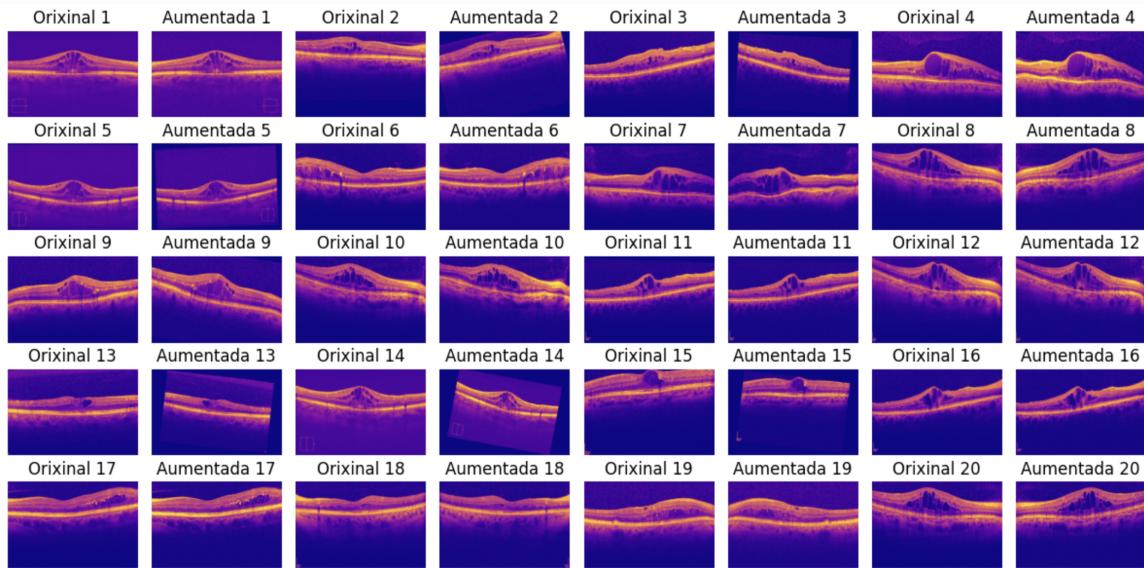
Para afrotalo problema, por tanto, armamosnos con unha serie de técnicas e trucos aprendidos nesta asignatura, en Aprendizaxe Profundo e mais en Principios de Visión por Computador. Son os seguintes:

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

ó ámbito médico máis aínda despois dunha considerable búsqueda de hiperparámetros decidimos descartalo para, finalmente, só quedarnos con Canny. Unha xustificación desta decisión é a imaxe de abaxo.

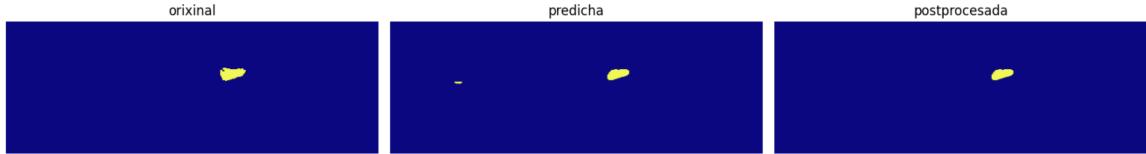


- **Aumento de datos.** Fundamentalmente baseámonos nas suxerencias do propio artigo da UNet [?] é dicir: transformacións afíns e elásticas. Cas primeiras foi realtivamente doado atopar hiperparámetros apropiados pero as segundas foi máis difícil, en cambio (ó estaren diseñadas para un contexto de segmentación médica) ofreceron moi bons resultados. Finalmente atopamos $\alpha = 500$ e $\sigma = 20$ axeitados. Ademáis, para estas dúas transformacións, ó seren realativamente *agresivas*, atopamos que é millor non transformar sempre (para darlle á Rede uns poucos exemplos inalterados e aprenda deles). En concreto aplicamos cada unha das cunha probabilidade de 0,7, polo tanto, para unha imaxe, a probabilidade de seren alterada por ambas transformacións é de $\approx 0,5$. Ademáis destas dous, aplicamos (malia observando pouca diferencia) volteos horizontais, deformacións na cor (*color jitter*), variacións no enfoque e ruído gausiano aditivo. O efecto da totalidade das transformacións sobre un subconjunto das imaxes do conxunto de datos vése abaxo.



Se ben pode parecer *leve*, aumentos de datos más agresivos non veían a luz da converxencia. Podemos dicir que a configuración de hiperparámetros do Aumento de Datos está fortemente baseada na experimentación. Comentár que tamén probamos a recortar as imaxes contrando 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 realtivamente a mesma escala, facer grandes variacións nese sentido pode (como temos visto) empeorar o rendemento da Rede.

- **PostProcesado das máscaras.** Aínda que non-diferenciables, e por tanto non contribúían á sinal de error do adestramento do algoritmo, decidimos aplicar unha serie de transformacións sobre as máscaras para acercarse más ás reais. Pulindo detalles de xeito que, se un oftalmólogo tivera que usar o noso sistema, certas impurezas corrixibles non o molestarían. En concreto nos axudamos do operador morfolóxico de peche.

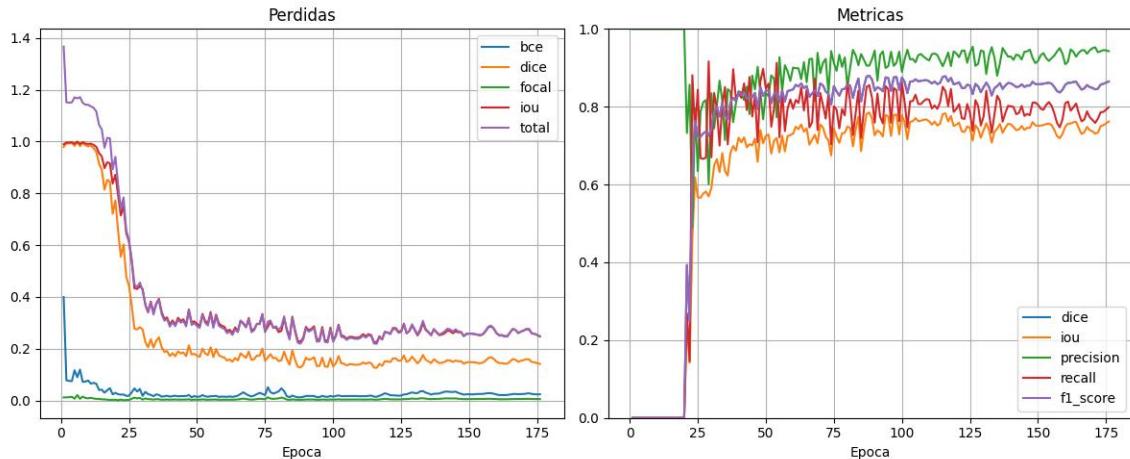


Comentar, non obstante, que por non ser o obxectivo da práctica o resultado final senón o adestramento da Rede en sí mesma, limitámonos a unha primeira idea de só aplicar peche, máis de seguro poderíanse implementar máis e mellores alternativas.

- **Trucos no adestrado da Rede.** Para asegurar converxencia e acelerar o adestramento usamos Dropout [?] como regularización e AdamW [?] como optimizador que tamén implementa regularización a través dos pesos. Estos dous elementos permitiron que se propagase axeitadamente a sinal de error na Rede, que nalgúns experimentos fixemos moi grande. Adicionalmente, adicando un subconjunto dos datos a *validación*, implementamos *EarlyStopping* con paciencia na pérdida en validación e máis un xestor do paso de aprendizaxe que, tamén, tiña paciencia ca pérdida en validación. Estos dous compoñentes fixeron dinámicos parámetros que, de ser estáticos e sempre iguais, non permitirían aprender con éxito ás Redes porque as Redes máis 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 pérdida 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 nos fixamos $\alpha_{BCE} = 0,7$, $\alpha_{FOCAL} = 0,2$, $\alpha_{DICE} = 0,4$, $\alpha_{IOU} = 0,7$. Esos coeficientes fixáronse a partires de observar as primeiras gráficas de adestramento. Abaixo pódese ver unha grafica ca función de pérdida aquí descrita.



2. O noso planteamento

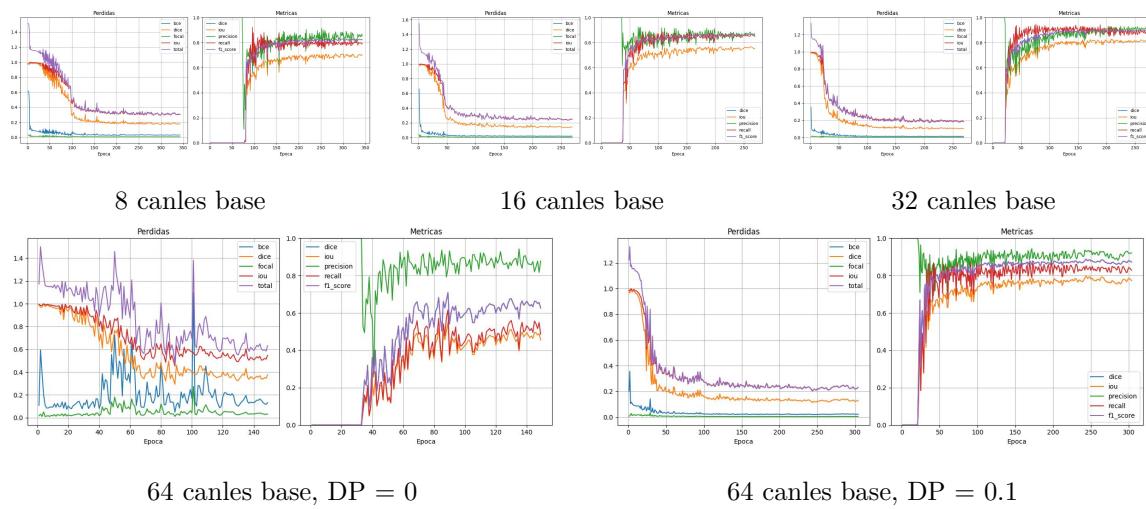
Despois de ter feito a práctica anterior, reconocemos das vantaxes dun bo planteamento inicial, tanto no nivel conceptual e máis de *visión* como no nivel técnico e máis de *programación*. É por isto que dende un primeiro momento planteamos a práctica como un todo, por tanto, non fixemos primeiro un *Baseline* e logo experimentos e melloras sobre iso; en vez, dende un primeiro momento planteamos na parte de programación a interfaz 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 compacional) e moi rectangulares (200 pixeles 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 Razon  
ancho-alto media:", suma/n}'  
>Razon ancho-alto media: 2.6163
```

Posteriormente decatámonos que para que a Rede percibira propiamente as paredes do Fluxo Óptico, estas non podían ter de ancho 2 ou 3 pixeles, entón, malia seren más computacionalmente costosos, os modelos adestrados con imaxes máis grandes ofreceran mellores resultados. Tamén decatámonos de que malia seren as imaxes moi rectangulares, a mascara sempre estaba no centro e non pagaba a pena o borde, 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 ó artigo [?], decatámonos 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 fácilmente pode mutar a outras configuracións. Abaixo probamos a adestrar a Rede ca mesma configuración variando só os canles de entrada (4 primeiras imaxes) e vemos que, (como suixeruse xa no 2012 [?]) canto maior é a Rede mellores resultados. Tamén aprobeitamos para mostrar o efecto do Dropout [?] (recomendado no artigo orixinal [?]), que é o único que varía nas dúas imaxes de abaxo.



3. Conclusións

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.