Cómo entrenar bien un baseline:

Experiencias y recomendaciones desde la aplicación de deep learning en oftalmología

Dr. José Ignacio Orlando

Grupo Yatiris. Instituto PLADEMA / CONICET / UNICEN

@ignaciorlando

José Ignacio Orlando

Inteligencia artificial
en medicina

↓

Investigación aplicada

Resolver un problema

↙ ↘

Modelo que ya
existe pero nadie usó

Problema nuevo

Modelo que
no existía

Un modelo nuevo!

Un modelo nuevo!

Contribución metodológica al área

¿Cómo lo validamos?

benchmark
protolo de evaluación
(dataset + métricas)

baselines
modelos para resolver
un problema similar

El baseline

Adecuado para el problema
Identificar las alternativas que cualquiera usaría

El mejor de todos
Si hay varios, elegimos el que tiene mejores benchmarks

Entrenado con las mismas oportunidades
Misma capacidad, misma cantidad de datos de entrenamiento, misma rigurosidad en la calibración de hiperparámetros

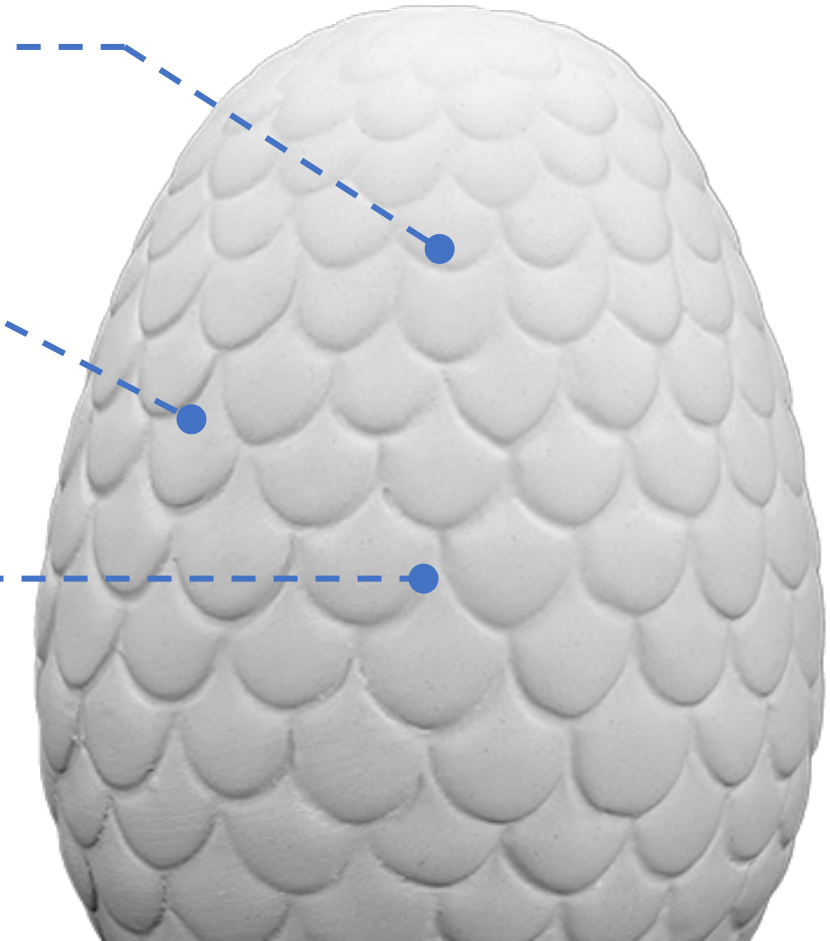


Image classification

Liu et al. CVPR 2022.

Vision Transformers como estado del arte

O sea que ConvNets nunca más?

A ConvNet for the 2020s

Zhuang Liu[1,2*] Hanzi Mao[1] Chao-Yuan Wu[1] Christoph Feichtenhofer[1] Trevor Darrell[2] Saining Xie[1†]

[1]Facebook AI Research (FAIR) [2]UC Berkeley

Code: https://github.com/facebookresearch/ConvNeXt

Abstract

The "Roaring 20s" of visual recognition began with the introduction of Vision Transformers (ViTs), which quickly superseded ConvNets as the state-of-the-art image classification model. A vanilla ViT, on the other hand, faces difficulties when applied to general computer vision tasks such as object detection and semantic segmentation. It is the hierarchical Transformers (e.g., Swin Transformers) that reintroduced several ConvNet priors, making Transformers practically viable as a generic vision backbone and demonstrating remarkable performance on a wide variety of vision tasks. However, the effectiveness of such hybrid approaches is still largely credited to the intrinsic superiority of Transformers, rather than the inherent inductive biases of convolutions. In this work, we reexamine the design spaces and test the limits of what a pure ConvNet can achieve. We gradually "modernize" a standard ResNet toward the design of a vision Transformer, and discover several key components that contribute to the performance difference along the way. The outcome of this exploration is a family of pure ConvNet models dubbed ConvNeXt. Constructed entirely from standard ConvNet modules, ConvNeXts compete favorably with Transformers in terms of accuracy and scalability, achieving 87.8% ImageNet top-1 accuracy and outperforming Swin Transformers on COCO detection and ADE20K segmentation, while maintaining the simplicity and efficiency of standard ConvNets.

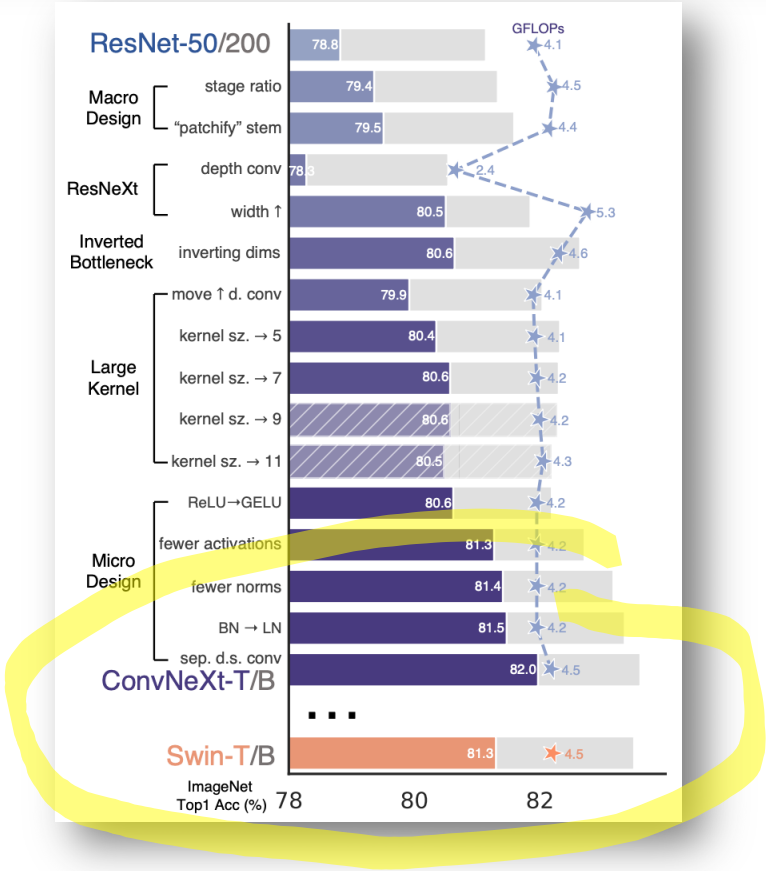


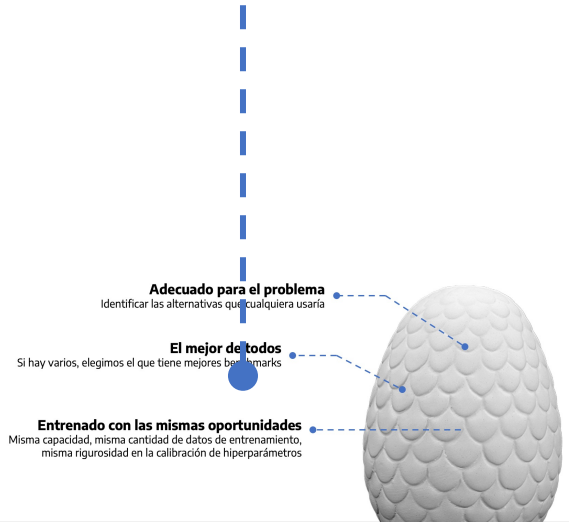
Image classification

Liu et al. CVPR 2022.

Vision Transformers como estado del arte

Realmente tenemos que abandonar las ConvNets?

NO! Hay que armarlas con las mismas herramientas que usan los mejores ViTs!



A ConvNet for the 2020s

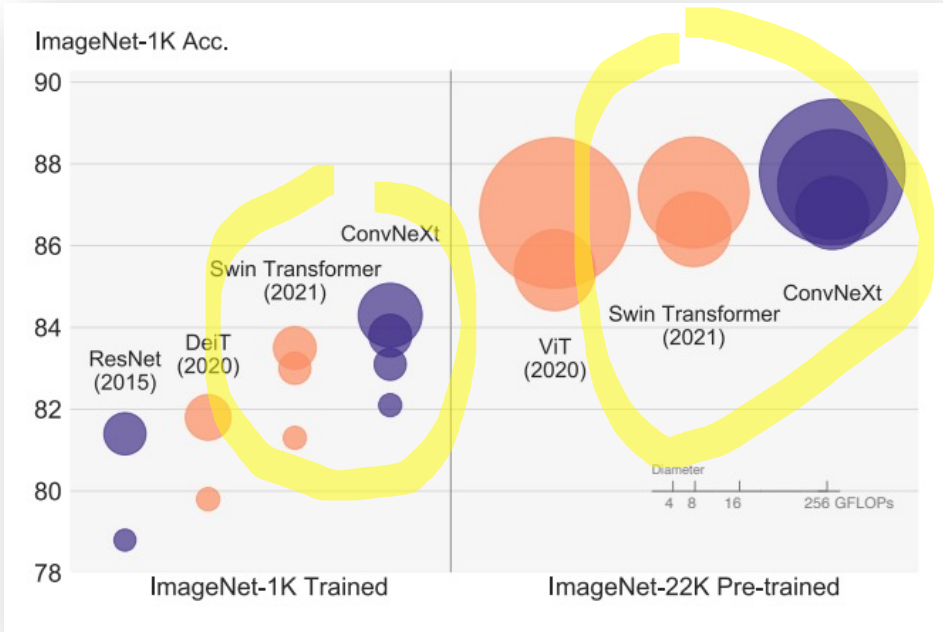
Zhuang Liu[1,2*] Hanzi Mao[1] Chao-Yuan Wu[1] Christoph Feichtenhofer[1] Trevor Darrell[2] Saining Xie[1†]

[1]Facebook AI Research (FAIR) [2]UC Berkeley

Code: https://github.com/facebookresearch/ConvNeXt

Abstract

The "Roaring 20s" of visual recognition began with the introduction of Vision Transformers (ViTs), which quickly superseded ConvNets as the state-of-the-art image classification model. A vanilla ViT, on the other hand, faces difficulties when applied to general computer vision tasks such as object detection and semantic segmentation. It is the hierarchical Transformers (e.g., Swin Transformers) that reintroduced several ConvNet priors, making Transformers practically viable as a generic vision backbone and demonstrating remarkable performance on a wide variety of vision tasks. However, the effectiveness of such hybrid approaches is still largely credited to the intrinsic superiority of Transformers, rather than the inherent inductive biases of convolutions. In this work, we reexamine the design spaces and test the limits of what a pure ConvNet can achieve. We gradually "modernize" a standard ResNet toward the design of a vision Transformer, and discover several key components that contribute to the performance difference along the way. The outcome of this exploration is a family of pure ConvNet models dubbed ConvNeXt. Constructed entirely from standard ConvNet modules, ConvNeXts compete favorably with Transformers in terms of accuracy and scalability, achieving 87.8% ImageNet top-1 accuracy and outperforming Swin Transformers on COCO detection and ADE20K segmentation, while maintaining the simplicity and efficiency of standard ConvNets.

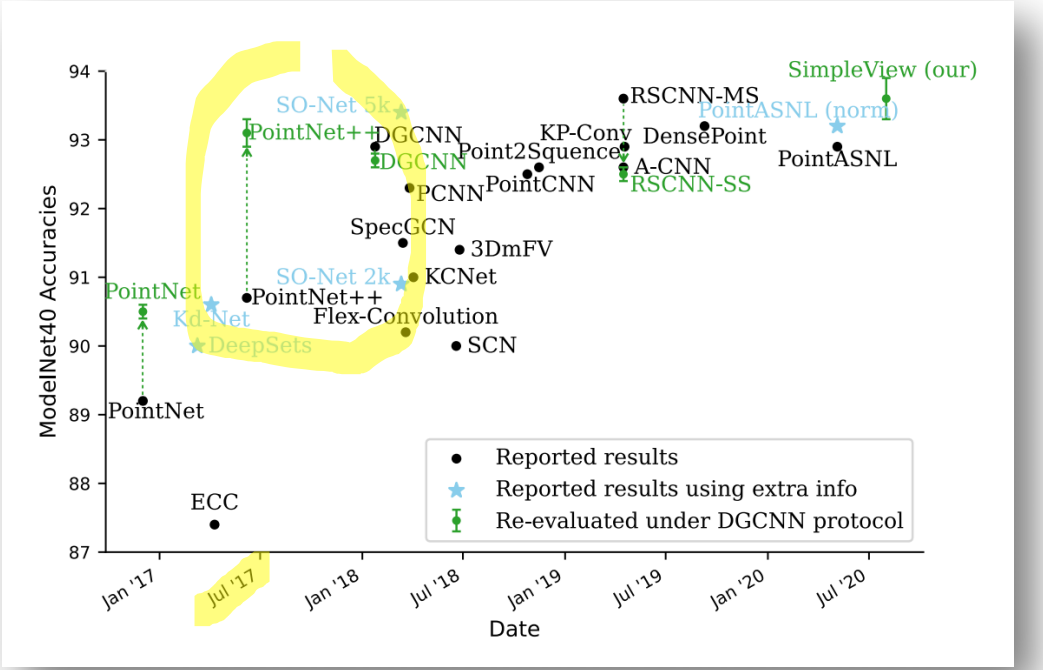


Revisiting Point Cloud Shape Classification with a Simple and Effective Baseline

Ankit Goyal [1] **Hei Law** [1] **Bowei Liu** [1] **Alejandro Newell** [1] **Jia Deng** [1]

Abstract

Processing point cloud data is an important component of many real-world systems. As such, a wide variety of point-based approaches have been proposed, reporting steady benchmark improvements over time. We study the key ingredients of this progress and uncover two critical results. First, we find that auxiliary factors like different evaluation schemes, data augmentation strategies, and loss functions, which are independent of the model architecture, make a large difference in performance. The differences are large enough that they obscure the effect of architecture. When these factors are controlled for, PointNet++, a relatively older network, performs competitively with recent methods. Second, a very simple projection-based method, which we refer to as SimpleView, performs surprisingly well. It achieves on par or better results than sophisticated state-of-the-art methods on ModelNet40 while being half the size of PointNet++. It also outperforms state-of-the-art methods on ScanObjectNN, a real-world point cloud benchmark, and demonstrates better cross-dataset generalization. Code is available at https://github.com/princeton-vl/SimpleView.



Point cloud shape classification

Goyal et al. ICML 2021.

Muchos modelos diferentes evaluados en el mismo benchmark

Realmente necesitamos complejizar más los modelos?

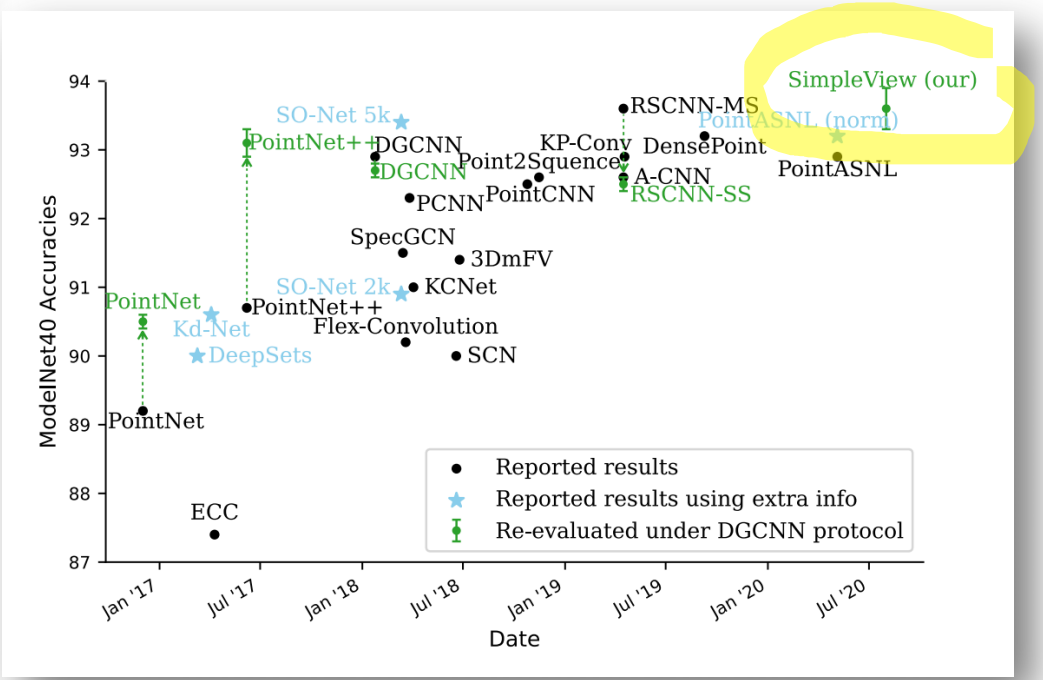
NO! Hay que entrenarlos bien!

Revisiting Point Cloud Shape Classification with a Simple and Effective Baseline

Ankit Goyal [1] **Hei Law** [1] **Bowei Liu** [1] **Alejandro Newell** [1] **Jia Deng** [1]

Abstract

Processing point cloud data is an important component of many real-world systems. As such, a wide variety of point-based approaches have been proposed, reporting steady benchmark improvements over time. We study the key ingredients of this progress and uncover two critical results. First, we find that auxiliary factors like different evaluation schemes, data augmentation strategies, and loss functions, which are independent of the model architecture, make a large difference in performance. The differences are large enough that they obscure the effect of architecture. When these factors are controlled for, PointNet++, a relatively older network, performs competitively with recent methods. Second, a very simple projection-based method, which we refer to as SimpleView, performs surprisingly well. It achieves on par or better results than sophisticated state-of-the-art methods on ModelNet40 while being half the size of PointNet++. It also outperforms state-of-the-art methods on ScanObjectNN, a real-world point cloud benchmark, and demonstrates better cross-dataset generalization. Code is available at https://github.com/princeton-vl/SimpleView.



Point cloud shape classification

Goyal et al. ICML 2021.

Muchos modelos diferentes evaluados en el mismo benchmark

Realmente necesitamos complejizar más los modelos?

NO! Hay que entrenarlos bien!

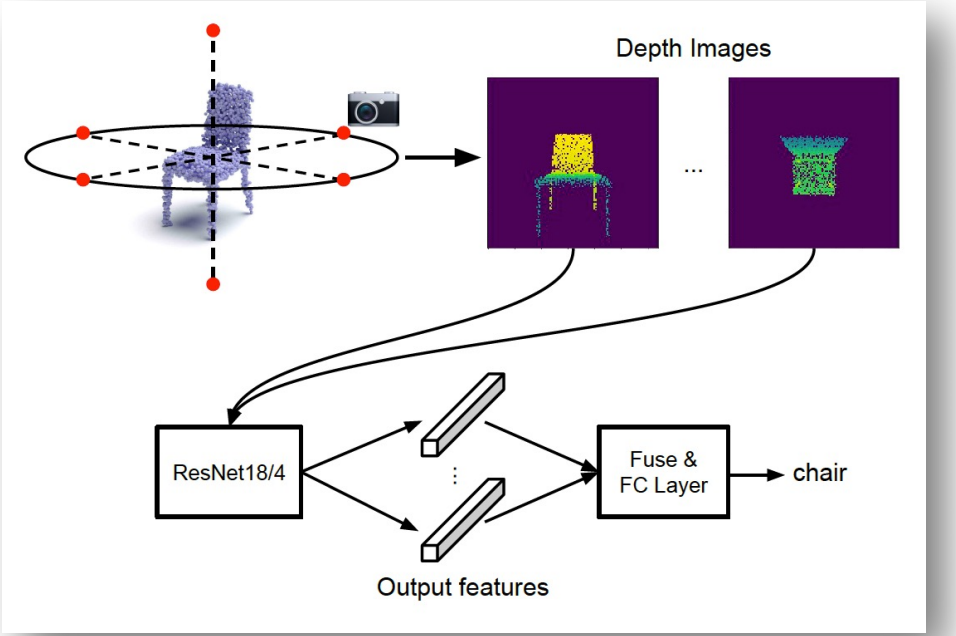
Modelos simples que funcionan mejor que modelos más complejos!

Revisiting Point Cloud Shape Classification with a Simple and Effective Baseline

Ankit Goyal [1] Hei Law [1] Bowei Liu [1] Alejandro Newell [1] Jia Deng [1]

Abstract

Processing point cloud data is an important component of many real-world systems. As such, a wide variety of point-based approaches have been proposed, reporting steady benchmark improvements over time. We study the key ingredients of this progress and uncover two critical results. First, we find that auxiliary factors like different evaluation schemes, data augmentation strategies, and loss functions, which are independent of the model architecture, make a large difference in performance. The differences are large enough that they obscure the effect of architecture. When these factors are controlled for, PointNet++, a relatively older network, performs competitively with recent methods. Second, a very simple projection-based method, which we refer to as SimpleView, performs surprisingly well. It achieves on par or better results than sophisticated state-of-the-art methods on ModelNet40 while being half the size of PointNet++. It also outperforms state-of-the-art methods on ScanObjectNN, a real-world point cloud benchmark, and demonstrates better cross-dataset generalization. Code is available at https://github.com/princeton-vl/SimpleView.



Point cloud shape classification

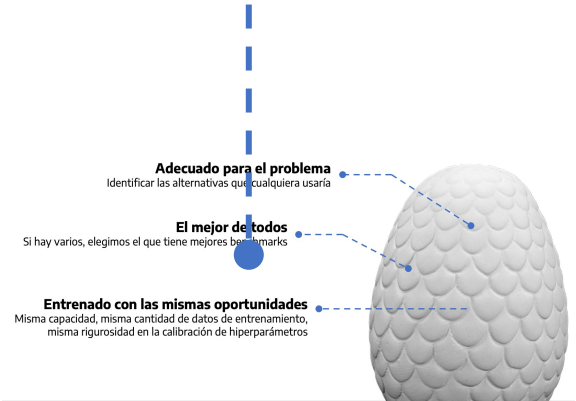
Goyal et al. ICML 2021.

Muchos modelos diferentes evaluados en el mismo benchmark

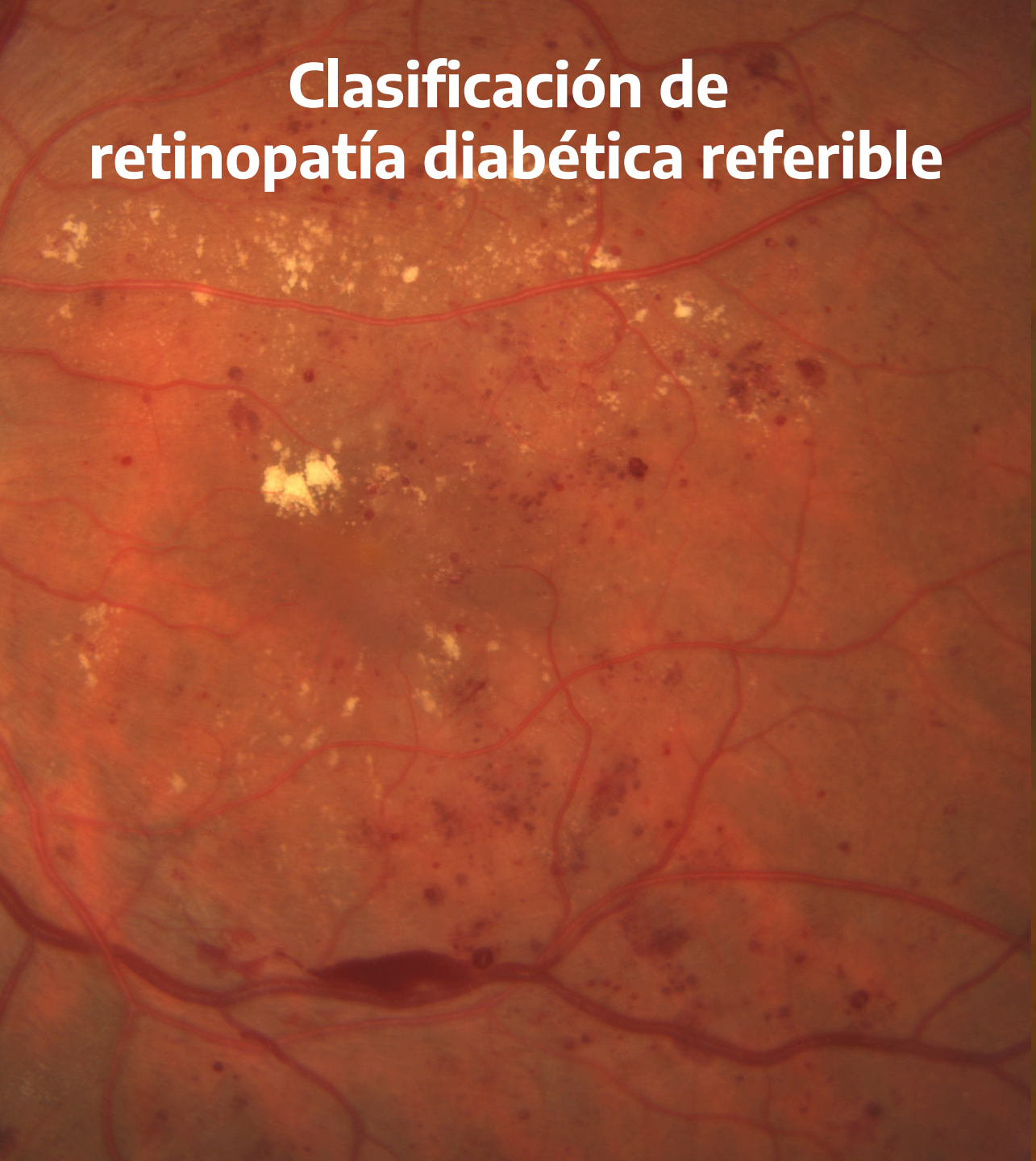
Realmente necesitamos complejizar más los modelos?

NO! Hay que entrenarlos bien!

Modelos simples que funcionan mejor que modelos más complejos!



¿Qué pasa en aplicaciones médicas?


Clasificación de
retinopatía diabética referible


Segmentación de
disco/copa óptica en glaucoma


Clasificación de
retinopatía diabética referible


Segmentación de
disco/copa óptica en glaucoma

Clasificación de retinopatía diabética referible

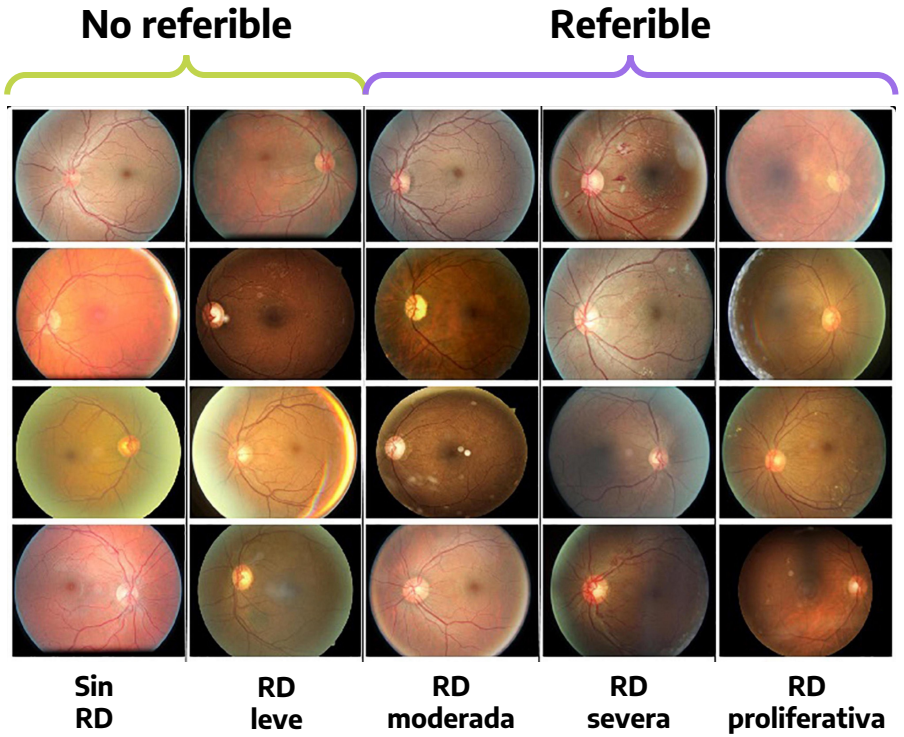
Problema binario de clasificación
No referible (clase 0) y referible (clase 1)

Múltiples datasets públicos disponibles
Más de 100.000 imágenes etiquetadas en diferentes conjuntos

Benchmarks específicos
Challenges diseñados con sus propios datos y métricas

Protocolos de evaluación diversos
Al publicarse los datos, dejan de respetarse los criterios aplicados

Muchos métodos diferentes
Revistas médicas → Modelos básicos
Revistas técnicas → Métodos complejos



Entrenemos un baseline

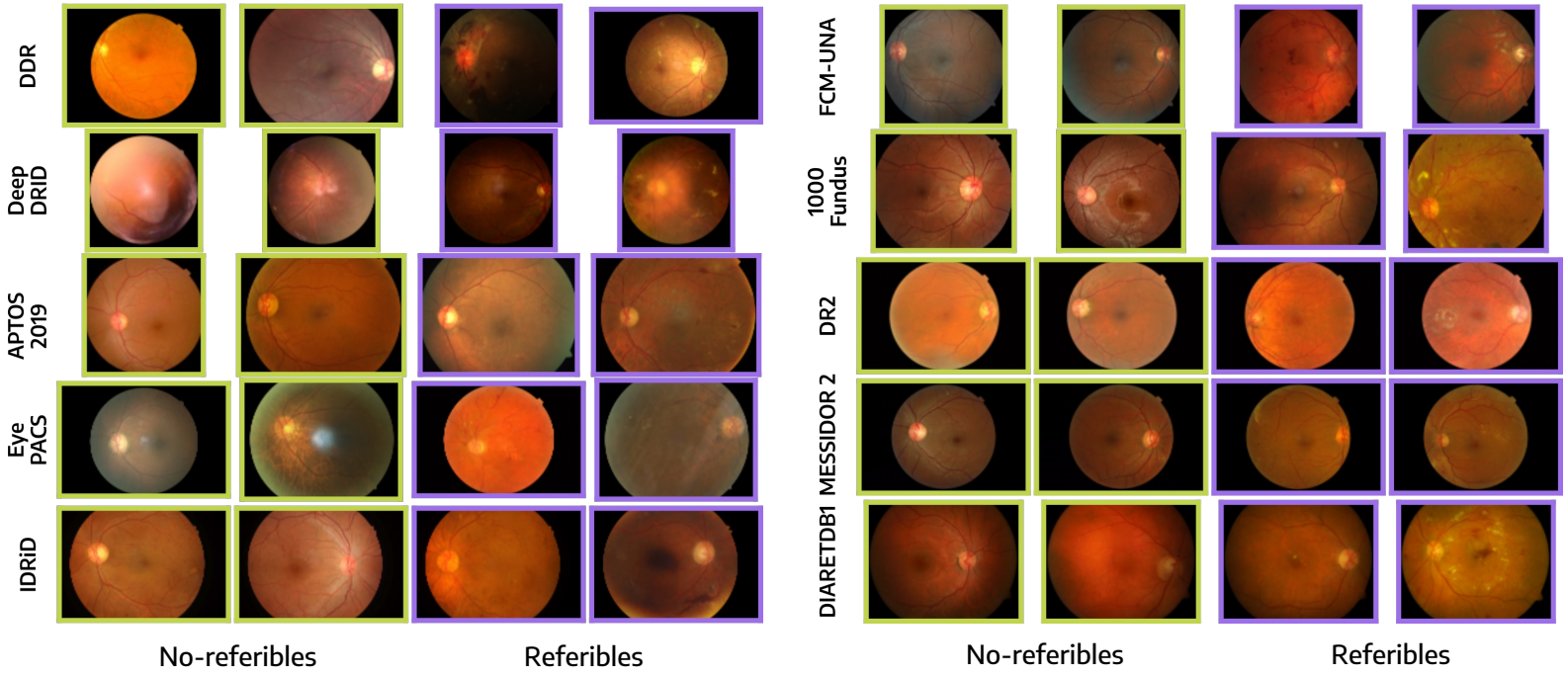
Clasificación de retinopatía diabética referible

(1) Identificamos todas las bases públicas con etiquetas de referibilidad

Particiones
¿Ya están divididas?
¿Qué patrones de uso sigue la literatura?

Imágenes
¿Qué resolución tienen?
¿Qué cámaras se utilizan?
¿Qué ángulo de FOV?
¿Midriática o no midriática?
¿Qué étnias específicas fueron escaneadas?

Etiquetas
¿Grados? ¿Binarias?



Castilla, T. Martínez, M. S. Leguía, M. Larrabide, I. and Orlando, J I.
"A ResNet is All You Need? Modeling A Strong Baseline for Detecting Referable Diabetic Retinopathy in Fundus Images"
18th International Symposium on Medical Information Processing and Analysis (SIPAIM 2022).

Clasificación de retinopatía diabética referible

(1) Identificamos todas las bases públicas con etiquetas de referibilidad

Particiones
¿Ya están divididas?
¿Qué patrones de uso sigue la literatura?

Imágenes
¿Qué resolución tienen?
¿Qué cámaras se utilizan?
¿Qué ángulo de FOV?
¿Midriática o no midriática?
¿Qué étnias específicas fueron escaneadas?

Etiquetas
¿Grados? ¿Binarias?

Dataset	Num. samples			Training	Validation	Test
	Non-referable DR	Referable DR	Total			
APTOS2019 *	2175	1487	3662	3662	0	0
DeepDRID[29]	900	700	1600	1200	0	400
DDR[8]	6896	5626	12522	6260	2503	3759
EyePACS †	71548	17154	88702	28098	7026	53576
IDRiD[28]	193	323	516	372	40	103
FCM-UNA[30]	191	566	757	0	0	757
1000Fundus[31]	56	88	144	0	0	144
DR2[4]	337	98	435	0	0	435
MESSIDOR 2[27]	1287	457	1744	0	0	1744
DIARETDB1[32]	43	46	89	0	0	89
Martínez (private)	454	30	484	0	0	484
HEC (private)	26	9	35	0	0	35
Total	89433	27735	117168	39592	9569	61526

Castilla, T. Martínez, M. S. Leguía, M. Larrabide, I. and Orlando, J I.
"A ResNet is All You Need? Modeling A Strong Baseline for Detecting Referable Diabetic Retinopathy in Fundus Images"
18th International Symposium on Medical Information Processing and Analysis (SIPAIM 2022).

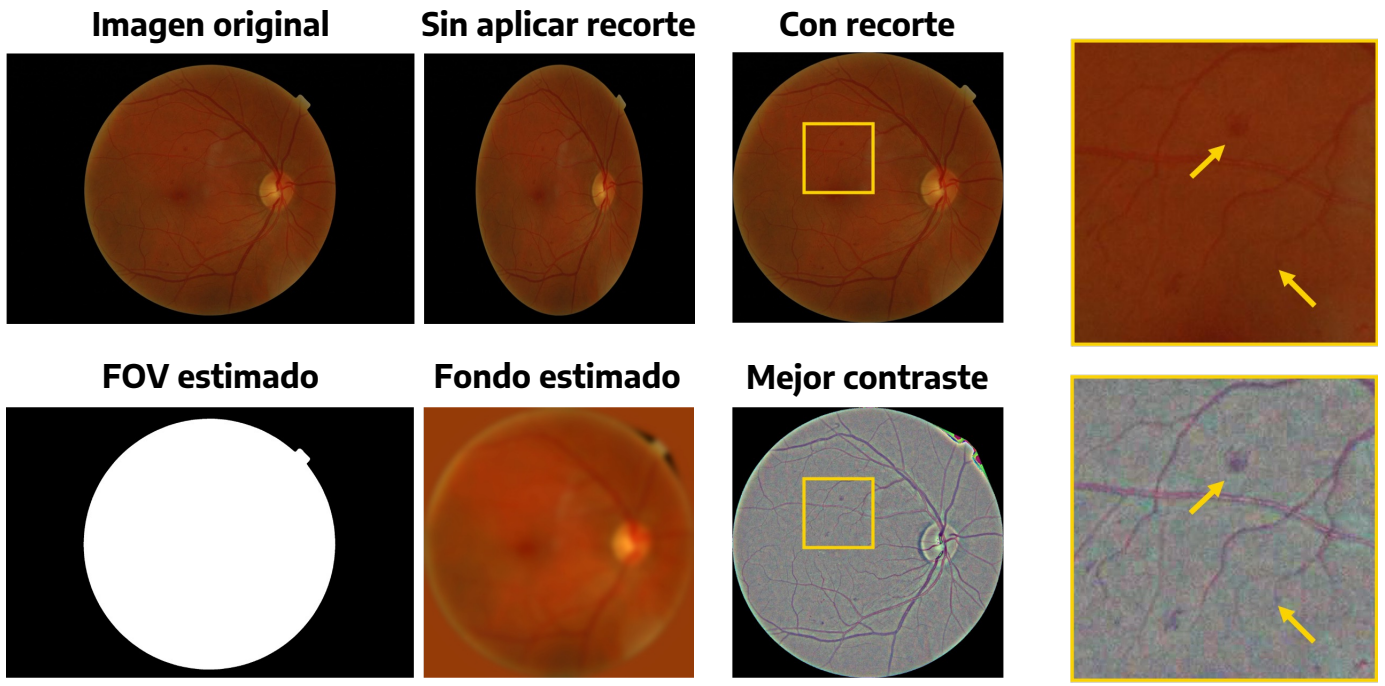
Clasificación de retinopatía diabética referible

(2) Preprocesamiento para integrar estudios de diferentes conjuntos

Recorte automático del FOV
Para evitar deformar las imágenes con artefactos poco realistas

Mejora de contraste
Para mejorar la identificación de las lesiones típicas de la RD

Resolución 512 x 512 px
Para evitar la desaparición de lesiones muy pequeñas



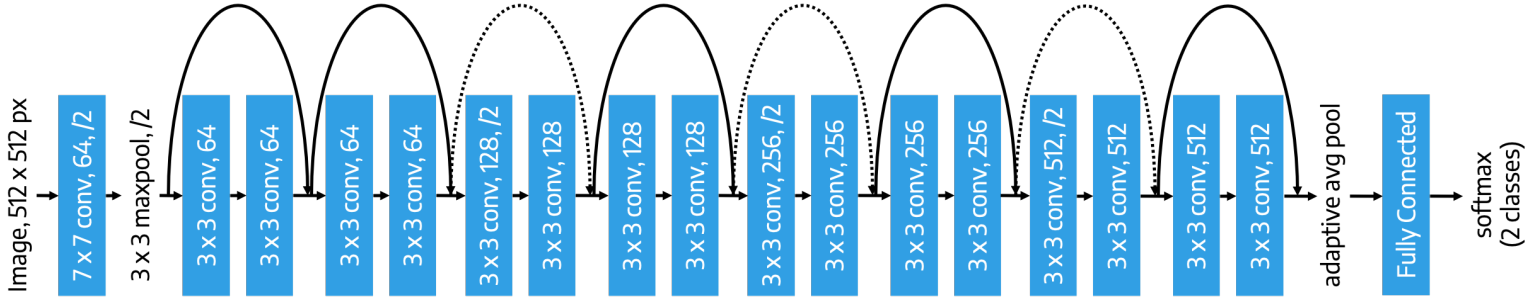
Castilla, T. Martínez, M. S. Leguía, M. Larrabide, I. and Orlando, J I.
"A ResNet is All You Need? Modeling A Strong Baseline for Detecting Referable Diabetic Retinopathy in Fundus Images"
18th International Symposium on Medical Information Processing and Analysis (SIPAIM 2022).

Clasificación de retinopatía diabética referible

(3) Entrenamiento de un modelo simple aplicando buenas prácticas

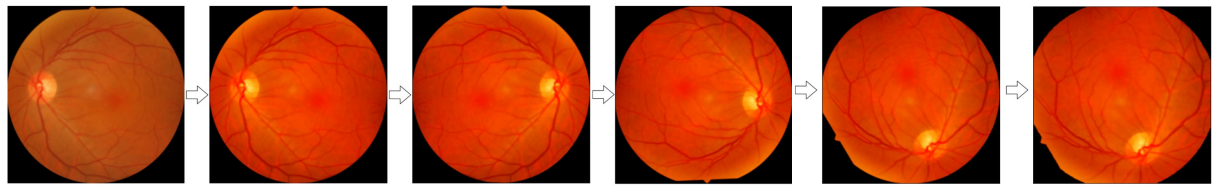
ResNet18

Pre-entrenada en ImageNet
Adaptada para clasificación binaria
Fine-tuning



Ajuste incremental del data augmentation

Forward selection de los parámetros de las estrategias de data augmentation



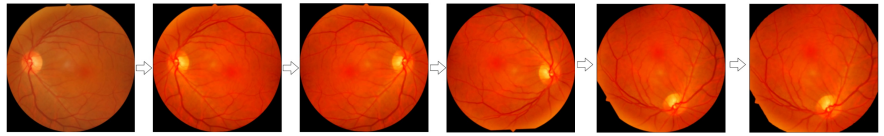
Castilla, T. Martínez, M. S. Leguía, M. Larrabide, I. and Orlando, J I.
"A ResNet is All You Need? Modeling A Strong Baseline for Detecting Referable Diabetic Retinopathy in Fundus Images"
18th International Symposium on Medical Information Processing and Analysis (SIPAIM 2022).

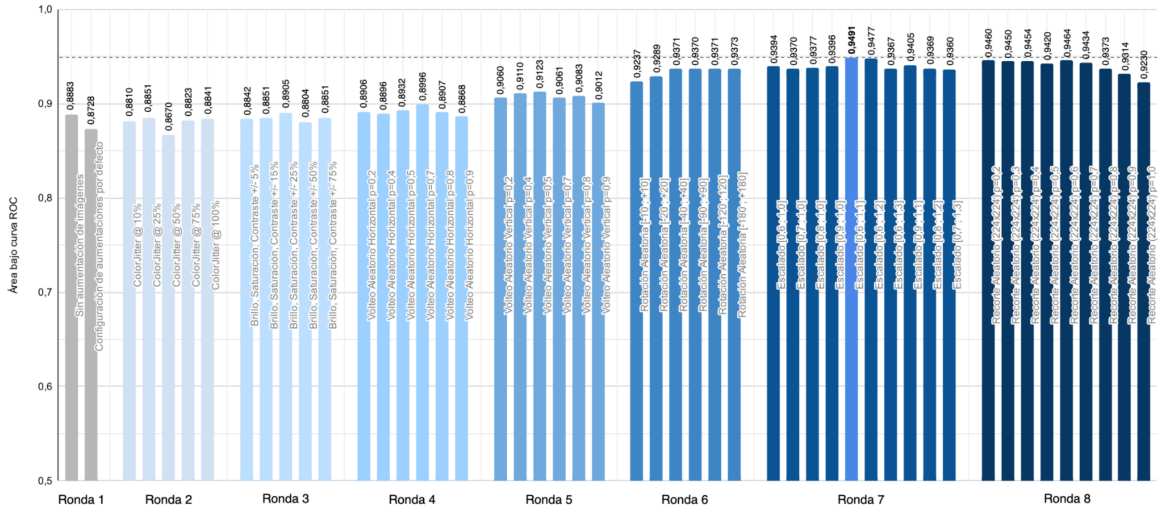
Clasificación de retinopatía diabética referible

(3) Entrenamiento de un modelo simple aplicando buenas prácticas

ResNet18
Pre-entrenada en Image-Net
Adaptada para clasificación binaria
Fine-tuning

Ajuste incremental del data augmentation
Forward selection de los parámetros de las estrategias de data augmentation





Castilla, T. Martínez, M. S. Leguía, M. Larrabide, I. and Orlando, J I.
"A ResNet is All You Need? Modeling A Strong Baseline for Detecting Referable Diabetic Retinopathy in Fundus Images"
18th International Symposium on Medical Information Processing and Analysis (SIPAIM 2022).

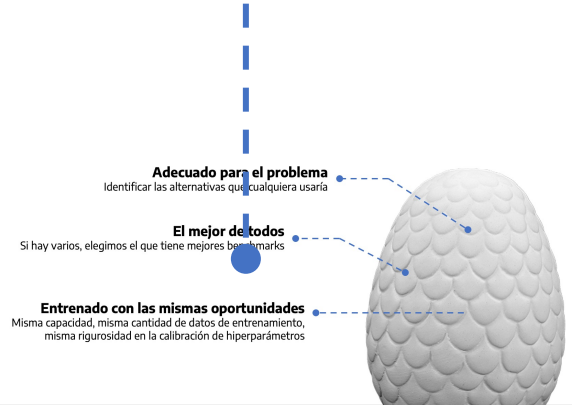
Clasificación de retinopatía diabética referible

(4) Evaluación sobre conjuntos y particiones estándar de la literatura + datos propios

Valores iguales o superiores a los de la literatura

Comparado con trabajos evaluados con iguales métricas y en iguales conjuntos

Modelo simple + datos + preprocesamiento



Per-image referable DR classification				
Test dataset	**Method**	**AUC (95% CI)**	**Se (95% CI)**	**Sp (95% CI)**
1000Fundus	**ResNet-18 (ours)**	**1.000** (1.000 - 1.000)	**1.000** (0.958 - 1.000)	**0.982** (0.906 - 0.997)
DDR	Zago *et al.* 2020[10]	0.833 (0.819 - 0.846)	-	-
	ResNet-18 (ours)	**0.965** (0.960 - 0.970)	**0.749** (0.728 - 0.769)	**0.978** (0.971 - 0.984)
DeepDRID	**ResNet-18 (ours)**	**0.959** (0.944 - 0.972)	**0.883** (0.828 - 0.922)	**0.868** (0.817 - 0.907)
DIARETDB1	**ResNet-18 (ours)**	**0.981** (0.956 - 0.999)	**0.957** (0.855 - 0.988)	**0.930** (0.814 - 0.976)
DR2	Pires *et al.* 2019[15]	0.963 (0.938 - 0.981)	-	-
	ResNet-18 (ours)	**0.974** (0.962 - 0.985)	**0.847** (0.763 - 0.905)	**0.961** (0.935 - 0.977)
EyePACS	Quellec *et al.* 2017[7]	0.944	-	-
	Pires *et al.* 2019[15]	0.946	-	-
	ResNet-18 (ours)	**0.951** (0.949 - 0.954)	**0.732** (0.723 - 0.740)	**0.979** (0.978 - 0.980)
FCM-UNA	**ResNet-18 (ours)**	**0.986** (0.980 - 0.992)	**0.882** (0.852 - 0.906)	**0.990** (0.963 - 0.997)
IDRiD	Zago *et al.* 2020[10]	0.796 (0.715 - 0.892)	-	-
	Hervella *et al.* 2022[12]	0.944	-	-
	ResNet-18 (ours)	**0.949** (0.914 - 0.980)	**0.828** (0.718 - 0.901)	**0.897** (0.764 - 0.959)
MESSIDOR 2	Gulshan *et al.* 2016[3]	**0.990** (0.986 - 0.995)	**0.961** (0.924 - 0.983)	0.939 (0.924 - 0.953)
	Gargeya *et al.* 2017[6]	0.940	0.930	0.870
	Voets *et al.* 2019[33]	0.853 (0.835 - 0.871)	0.818	0.687
	Zago *et al.* 2020[10]	0.944 (0.925 - 0.966)	0.900 (0.860 - 0.961)	0.870 (0.863 - 0.871)
	Li *et al.* 2022[19]	0.977 (0.974 - 0.981)	0.923 (0.917 - 0.925)	0.947 (0.937 - 0.954)
	ResNet-18 (ours)	0.973 (0.967 - 0.979)	0.895 (0.863 - 0.920)	**0.941** (0.927 - 0.953)
HEC	**ResNet-18 (ours)**	**0.961** (0.900 - 1.000)	**1.000** (0.610 - 1.000)	**0.862** (0.694 - 0.945)
Martínez	**ResNet-18 (ours)**	**0.955** (0.927 - 0.980)	**0.800** (0.627 - 0.905)	**0.934** (0.907 - 0.953)

Castilla, T. Martínez, M. S. Leguía, M. Larrabide, I. and Orlando, J I.
"A ResNet is All You Need? Modeling A Strong Baseline for Detecting Referable Diabetic Retinopathy in Fundus Images"
18th International Symposium on Medical Information Processing and Analysis (SIPAIM 2022).


Clasificación de
retinopatía diabética referible


Segmentación de
disco/copa óptica en glaucoma

Segmentación de disco/copa óptica en glaucoma

Problema multiclase/multilabel de segmentación
Disco óptico (clase 1), copa óptica (clase 2) y fondo (clase 0)

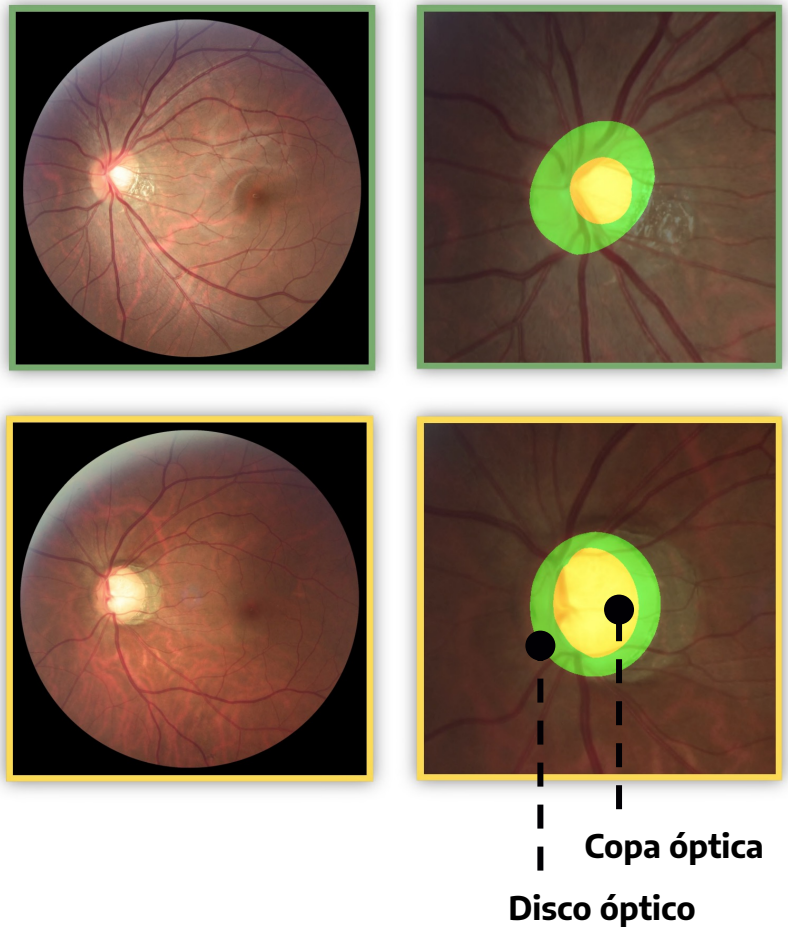
Múltiples datasets públicos disponibles
Más de 3.400 imágenes segmentadas en diferentes conjuntos

Benchmarks específicos
Challenges diseñados con sus propios datos y métricas

Protocolos de evaluación diversos
Al publicarse los datos, dejan de respetarse los criterios aplicados

Muchos métodos diferentes
Revistas médicas → Modelos básicos
Revistas técnicas → Métodos complejos

Entrenemos un baseline

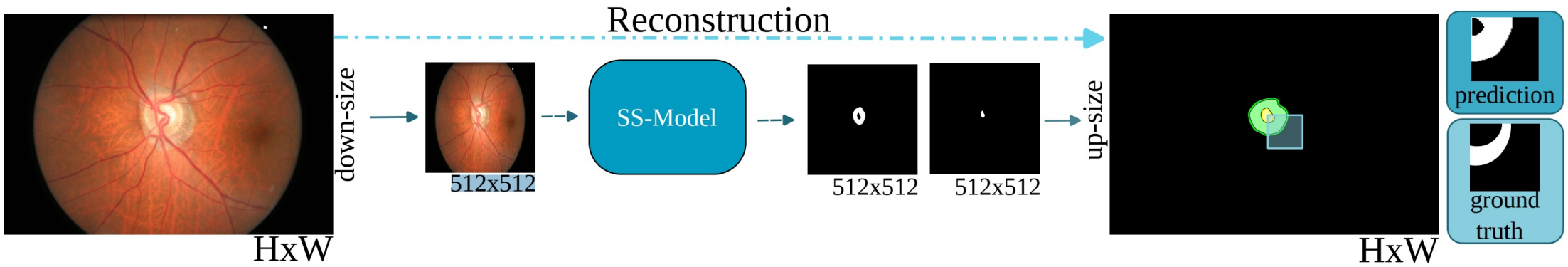


Segmentación de disco/copa óptica en glaucoma

Entrenemos un baseline ¿Cuál?

Una única etapa (Single Stage, SS)

U-Net multiclase estándar



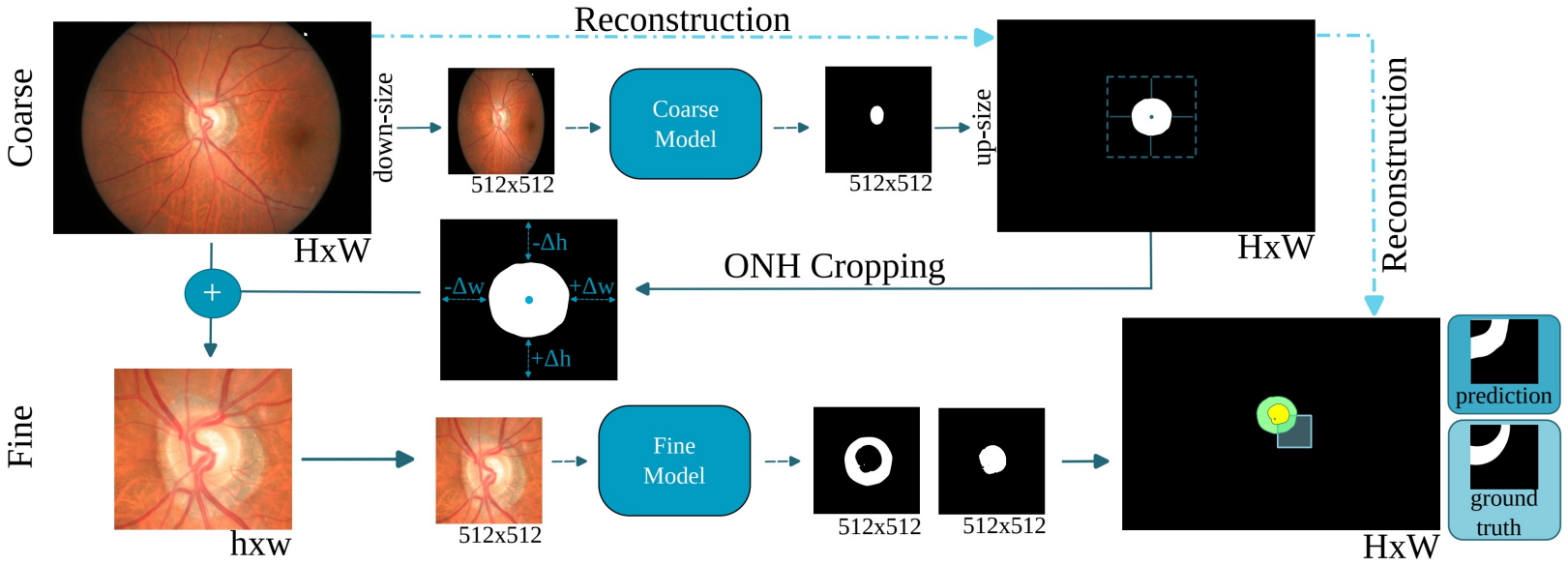
Moris, T. Dazeo, N. .... Larrabide, I. and Orlando, J I.
"Asessing Coarse-to-Fine Deep Learning Models for Optic Disc and Cup Segmentation in Fundus Images"
18th International Symposium on Medical Information Processing and Analysis (SIPAIM 2022).

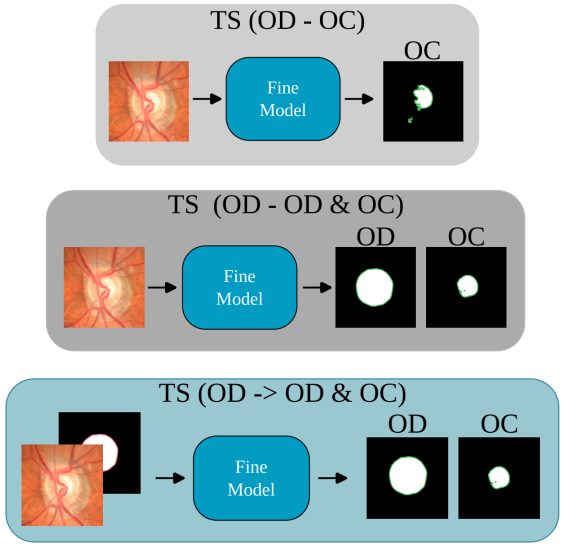
Segmentación de disco/copa óptica en glaucoma

Entrenemos un baseline ¿Cuál?

Dos etapas (Coarse-to-Fine, TS)

Dos U-Nets, una que segmenta OD en baja resolución y otra que segmenta OD/OC en alta resolución





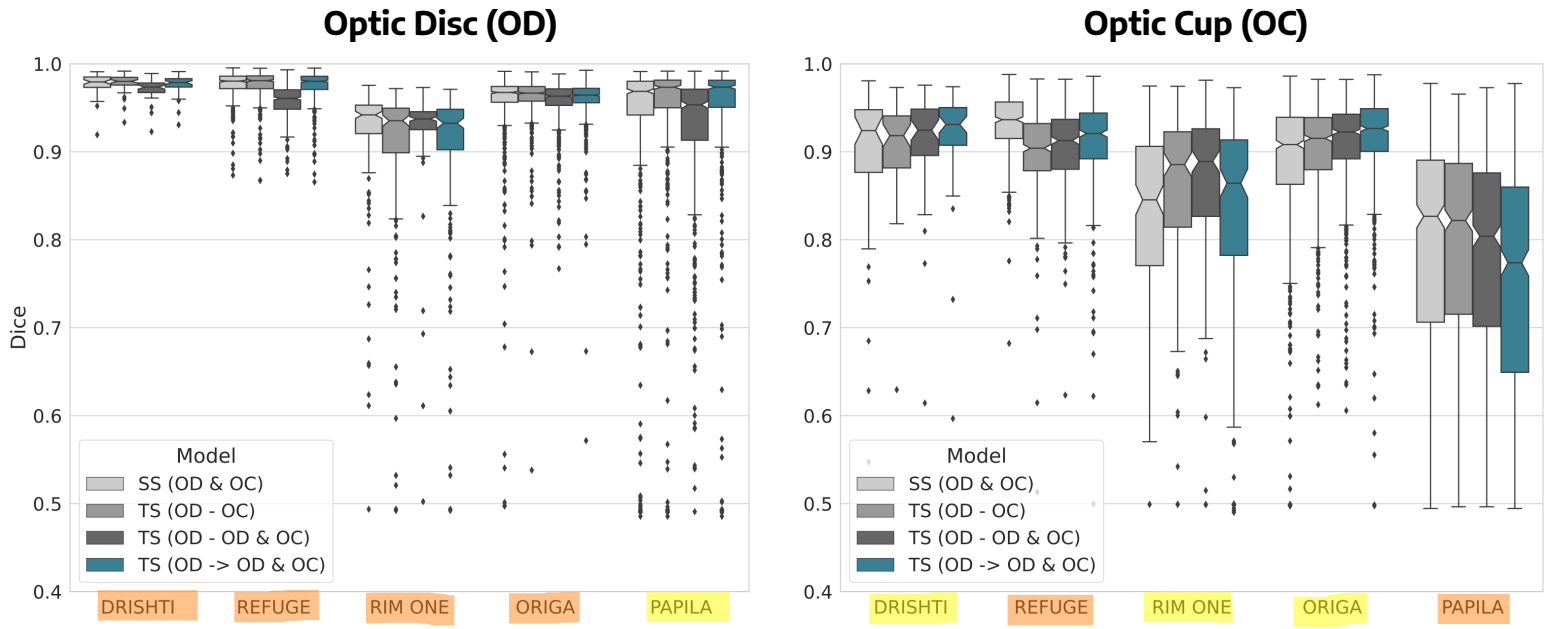
Moris, T. Dazeo, N. .... Larrabide, I. and Orlando, J I.
"Asessing Coarse-to-Fine Deep Learning Models for Optic Disc and Cup Segmentation in Fundus Images"
18th International Symposium on Medical Information Processing and Analysis (SIPAIM 2022).

Segmentación de disco/copa óptica en glaucoma

Evaluación sobre conjuntos y particiones estándar de la literatura

No necesariamente las dos etapas ayudan

Independientemente del diseño especifico que usemos de la conexión entre etapas



Moris, T. Dazeo, N. .... Larrabide, I. and Orlando, J I.
"Asessing Coarse-to-Fine Deep Learning Models for Optic Disc and Cup Segmentation in Fundus Images"
18th International Symposium on Medical Information Processing and Analysis (SIPAIM 2022).

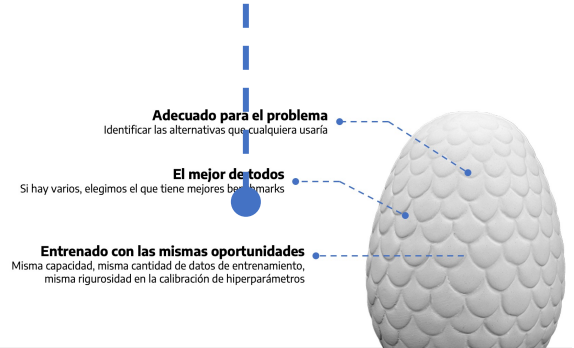
Segmentación de disco/copa óptica en glaucoma

Evaluación sobre conjuntos y particiones estándar de la literatura

No necesariamente las dos etapas ayudan

Independientemente del diseño especifico que usemos de la conexión entre etapas

Algunos modelos SS son mejores o iguales que el estado del arte!



Method	Optic disc (OD) segmentation				Optic cup (OC) segmentation			
	REFUGE	DRISHTI	RIM ONE v3	ORIGA	REFUGE	DRISHTI	RIM ONE v3	ORIGA
Al-Bandet *et al.* (2018)[11]	-	0.949	0.904	**0.965**	-	0.828	0.690	0.866
Team CUHKMED (REFUGE, 2019)[1]	0.960	-	-	-	0.883	-	-	-
Team Masker (REFUGE, 2019)[1]	0.946	-	-	-	0.884	-	-	-
Team BUCT (REFUGE, 2019)[1]	0.953	-	-	-	0.873	-	-	-
Wang *et al.* (2019)[9]	0.960	0.965	0.865	-	0.883	0.858	0.787	-
Shah *et al.* (2019)[8]	-	0.960	0.940	-	-	0.890	0.820	-
Tabassum *et al.* (2020)[10]	-	0.959	**0.958**	-	-	**0.924**	**0.862**	-
Liu *et al.* (2021)[7]	0.960	**0.978**	-	-	0.890	0.912	-	-
He *et al.* (2022)[19]	0.954	-	-	-	0.869	-	-	-
SS-model (REFUGE)	0.971 ± 0.015	0.957 ± 0.039	*0.841 ± 0.130*	*0.909 ± 0.123*	0.922 ± 0.033	0.899 ± 0.064	*0.773 ± 0.131*	*0.874 ± 0.119*
TS (OD → OD&OC) (REFUGE)	*0.974 ± 0.014*	*0.966 ± 0.024*	0.830 ± 0.134	0.894 ± 0.140	0.879 ± 0.071	*0.901 ± 0.063*	0.740 ± 0.168	0.851 ± 0.154
SS-model (DRISHTI)	0.881 ± 0.122	0.969 ± 0.025	*0.911 ± 0.065*	0.837 ± 0.186	0.795 ± 0.123	0.899 ± 0.073	0.764 ± 0.118	0.778 ± 0.165
TS (OD → OD&OC) (DRISHTI)	*0.917 ± 0.084*	*0.973 ± 0.018*	0.822 ± 0.059	*0.880 ± 0.159*	0.801 ± 0.131	0.899 ± 0.082	0.787 ± 0.121	0.809 ± 0.163
SS-model (multi-dataset)	**0.976 ± 0.016**	0.977 ± 0.012	0.919 ± 0.072	0.956 ± 0.049	**0.931 ± 0.035**	0.895 ± 0.090	0.829 ± 0.099	0.890 ± 0.079
TS (OD → OD&OC) (multi-dataset)	**0.976 ± 0.017**	0.977 ± 0.011	0.898 ± 0.094	0.959 ± 0.028	0.910 ± 0.055	0.917 ± 0.063	0.820 ± 0.136	**0.915 ± 0.056**

Moris, T. Dazeo, N. .... Larrabide, I. and Orlando, J I.
"Asessing Coarse-to-Fine Deep Learning Models for Optic Disc and Cup Segmentation in Fundus Images"
18th International Symposium on Medical Information Processing and Analysis (SIPAIM 2022).

¿Cómo entrenar bien a tu baseline?

¿Nuestra receta?

↓

¡La misma que deberías aplicar a tu modelo!

Los datos

Reunir la mayor cantidad de datos posibles
Mejor uso de la capacidad del modelo (large-scale era)

Establecer un criterio de integración de datos
No mezclarlos así nomás!

Generar un buen conjunto de validación
Para asegurar la efectividad del ajuste de hiperparámetros

Plantear conjuntos de test para evaluar diferentes escenarios
Recordar estudiar sesgos en los resultados!

Vas a publicar?

Entrená/evaluá también los baselines en las benchmarks
Respetando iguales particiones de entrenamiento / validación / test (más experimentos, sí!)

Publicá las particiones que creaste a partir de datos públicos
Para que otras personas puedan reproducir tus mismos experimentos

El entrenamiento

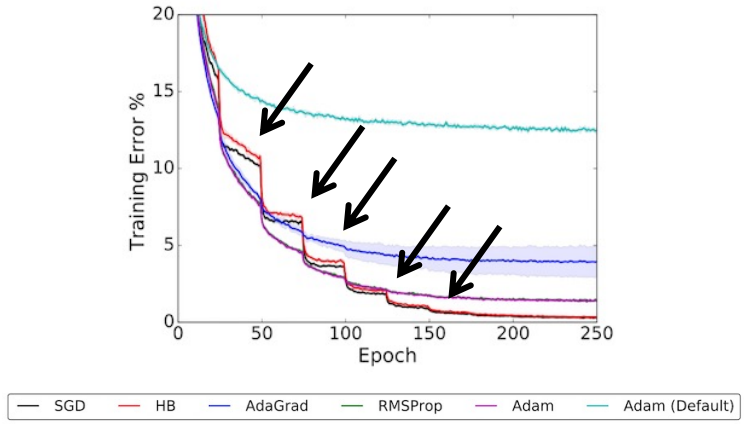
1. Lograr overfitting

Sí o sí la loss de entrenamiento a (casi) 0
Si no puedo llegar, tengo underfitting → mi baseline está mal entrenado

Buscar mejor combinación de learning rate + optimizador
Tabular combinaciones learning rate + optimizador vs. performance

¿Necesito un learning rate scheduler?
Si veo convergencia en la curva de aprendizaje muy lejos de cero, probar!

¿Necesito más capacidad?
Si learning rate + optimizador + scheduler no alcanzan, probablemente...



El entrenamiento

1. Lograr overfitting

Sí o sí la loss de entrenamiento a (casi) 0
Si no puedo llegar, tengo underfitting → mi baseline está mal entrenado

Buscar mejor combinación de learning rate + optimizador
Tabular combinaciones learning rate + optimizador vs. performance

¿Necesito un learning rate scheduler?
Si veo convergencia en la curva de aprendizaje muy lejos de cero, probar!

¿Necesito más capacidad?
Si learning rate + optimizador + scheduler no alcanzan, probablemente

Evaluar en training y validation set

No regularizar todavía!!

El entrenamiento

2. Reducir el overfitting

Ahora lo queremos bajar...
... pero sin romper todo

Ordenar qué políticas de data augmentation
Definir operaciones + estrategia tipo RandAugment
(nro. operaciones vs. fuerza)

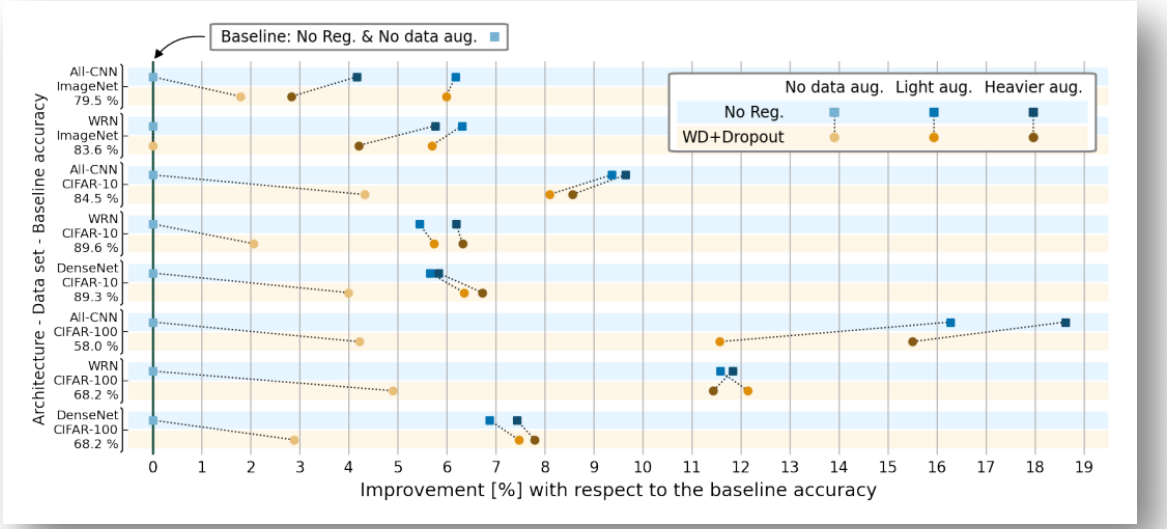
Weight decay?
Probar algunas tasas

Dropout?
Estudiar dónde conviene aplicarlo.

Data augmentation instead of explicit regularization

Alex Hernandez-Garcia*
Institute of Cognitive Science
University of Osnabrück
Germany
ahernandez@uos.de

Peter König
Institute of Cognitive Science
University of Osnabrück
Germany
pkoenig@uos.de



El entrenamiento

2. Reducir el overfitting

Ahora lo queremos bajar...
... pero sin romper todo

Ordenar qué políticas de data augmentation
Definir operaciones + estrategia tipo RandAugment
(nro. operaciones vs. fuerza)

Weight decay?
Probar algunas tasas

Dropout?
Estudiar dónde conviene aplicarlo.

Ver mejoras en validación
a costa de pérdida en training

Forward selection

Un modelo nuevo!

Contribución metodológica al área

¿Cómo lo validamos?

benchmark
protolo de evaluación
(dataset + métricas)

baselines
modelos para resolver
un problema similar

Slides acá!

¡Gracias por tu atención!

¿Tenés alguna pregunta?

jiorlando@pladema.exa.unicen.edu.ar

Cómo entrenar bien un baseline:

Experiencias y recomendaciones desde la aplicación de deep learning en oftalmología

Dr. José Ignacio Orlando

Grupo Yatiris. Instituto PLADEMA / CONICET / UNICEN.

@ignaciorlando

José Ignacio Orlando