

# VISIÓN POR COMPUTADOR

YURI VLADIMIR HUALLPA VARGAS

ABRIL 2023

## GENDER RECOGNITION (3 POINT)

Se ha comenzado el desarrollo de esta tarea utilizando el modelo proporcionado en el desafío "Identificación de modelos de automóviles" (CBGN), y se ha experimentado con la siguiente configuración de filtros (32>64>128), un optimizador SGD ( $lr=0.1$ ,  $decay=1e-6$ ) y 50 épocas. Dicha configuración ha logrado una precisión del 92% en el conjunto de desarrollo.

Dado el buen rendimiento del modelo, se ha llevado a cabo un aumento de datos y se ha agregado el early stopping para evitar el sobreajuste y el cómputo innecesario. Con estos cambios se ha alcanzado una precisión del 95,88%. Posteriormente, se ha probado con los siguientes learning rate scheduler (CosineDecay, ReduceLROnPlateau), donde CosineDecay con  $lr=0.1$  ha logrado una precisión del 96,50%.

Se han probado diferentes configuraciones de la capa MaxPool dentro del "CBGN", y la mejor configuración ha sido agregar uno después de cada convolución, con estos cambios el modelo ha alcanzado una precisión del 97,50%. Después de realizar numerosos cambios, se ha seguido la siguiente estrategia que ha permitido alcanzar el objetivo propuesto: dado que había un desequilibrio en los datos, primero se entrenó el modelo como en el paso previo, y luego se cambió el optimizador a AdamW, el learning rate scheduler a ReduceLROnPlateau con un learning rate de 0,0001 y finalmente se penalizó cada clase según el porcentaje de desequilibrio. Estas configuraciones han permitido alcanzar una precisión del **98,04%** de exactitud.

Para el modelo pequeño, únicamente se ha reducido la cantidad de filtros a 16>32>64, lo que ha dado lugar a un modelo con 77,201 parámetros, que ha logrado una precisión del **97,09%** de exactitud.

## CAR MODEL IDENTIFICATION WITH BI-LINEAR MODELS (5 POINTS)

Se ha adaptado el código número 2 proporcionado en clases y se han realizado una serie de experimentos con los modelos Resnet50 y VGG16. Se ha logrado obtener una exactitud del 71.68% en el conjunto de pruebas, y a continuación se describen los detalles de los experimentos.

En primer lugar, se experimentó con el modelo Resnet50 pre entrenado en el conjunto de datos "Imagenet". Para ello, se realizó el mismo preprocesamiento de datos y se eligieron diferentes bloques de convolución como salida (conv2\_block3\_out, conv3\_block4\_out, conv4\_block6\_out), siendo el bloque con mejor resultado "conv4\_block6\_out". Todos los experimentos se realizaron con el modelo Resnet50 congelado. Además, es necesario indicar que también se entrenó el modelo inicialmente sin congelar los pesos, pero se observaron problemas de desvanecimiento de gradiente.

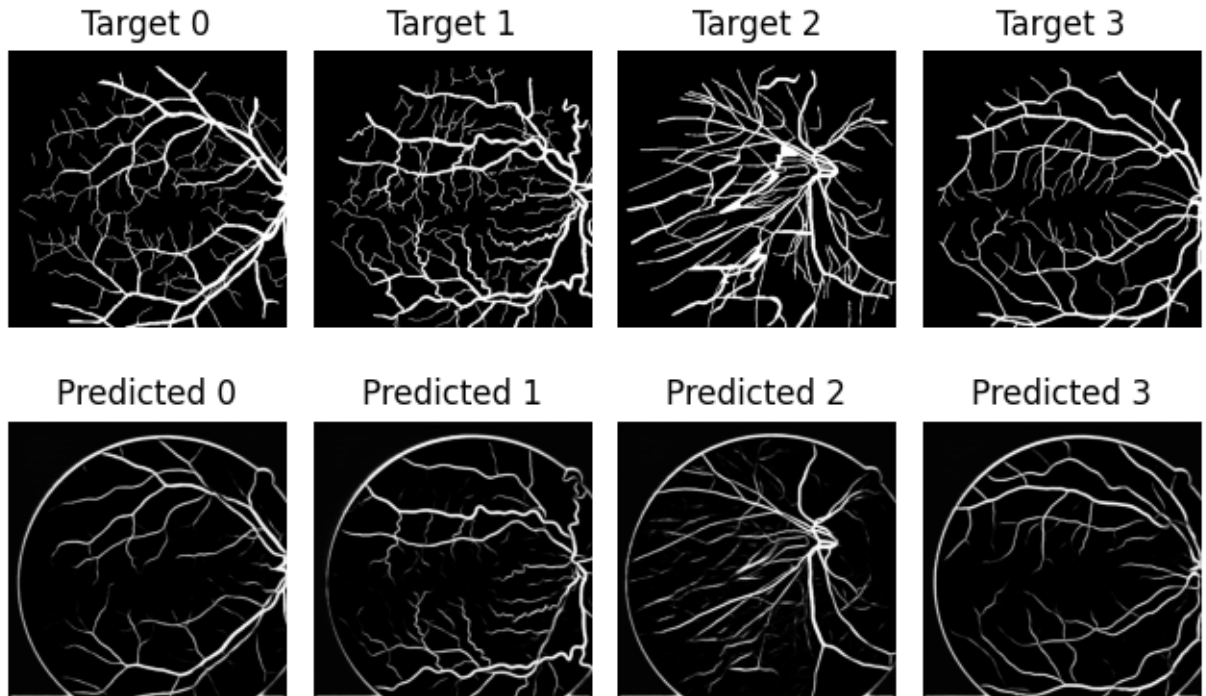
Los parámetros utilizados fueron: Lr = 0.1, 0.01 y 0.001. Optimizador = Adam, SGD, Adamw. Learning-scheduler = ReduceLROnPlateau y CosineDecay. Se logró conseguir una exactitud del **57.10%** con un Lr = 0.1, Optimizador = Adam y Early-stopping-val-loss = 5.

En segundo lugar, para experimentar con el modelo VGG16 se eligieron los siguientes bloques (block3\_conv3, block4\_conv3, block5\_conv3), y el bloque con mejor resultado y las mismas configuraciones realizadas en el primer experimento fue "block4\_conv3". Se logró conseguir una exactitud del **67.60%** en el conjunto de pruebas.

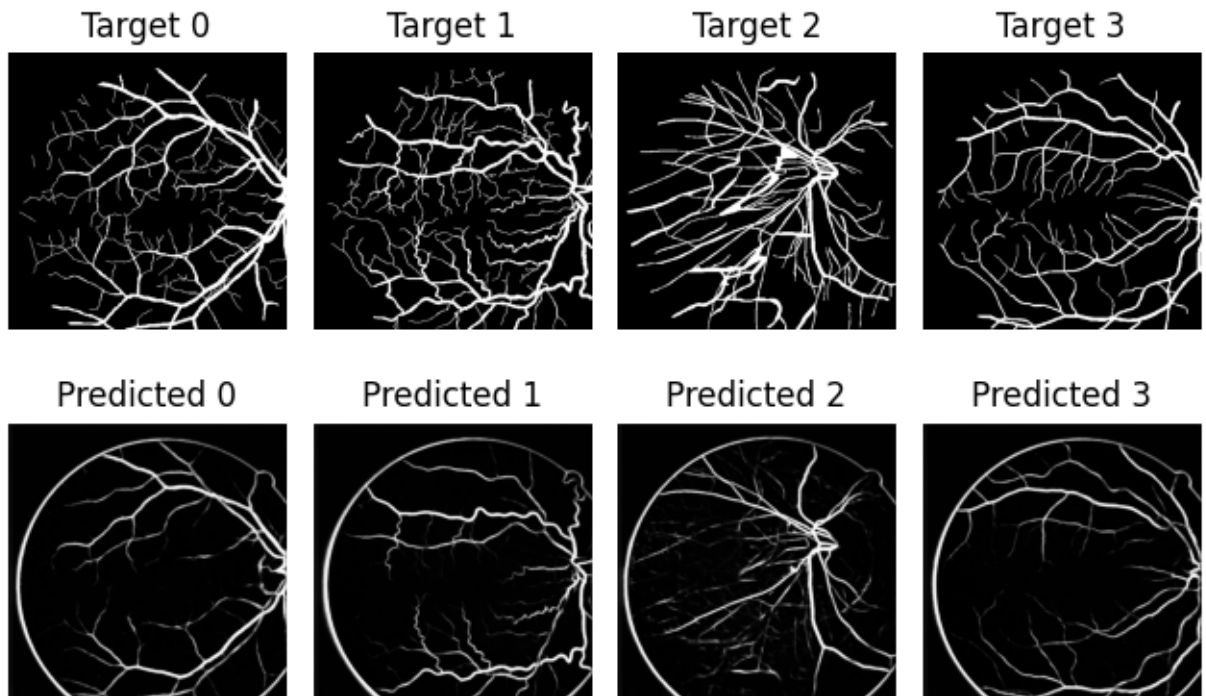
Finalmente, siguiendo las sugerencias dadas en clases, se descongeló el modelo VGG16 y se volvió a entrenar completamente, pero reduciendo el Lr a 0.0001, con lo que se alcanzó una exactitud del **71.68%**.

## IMAGE SEGMENTATION (4 POINTS)

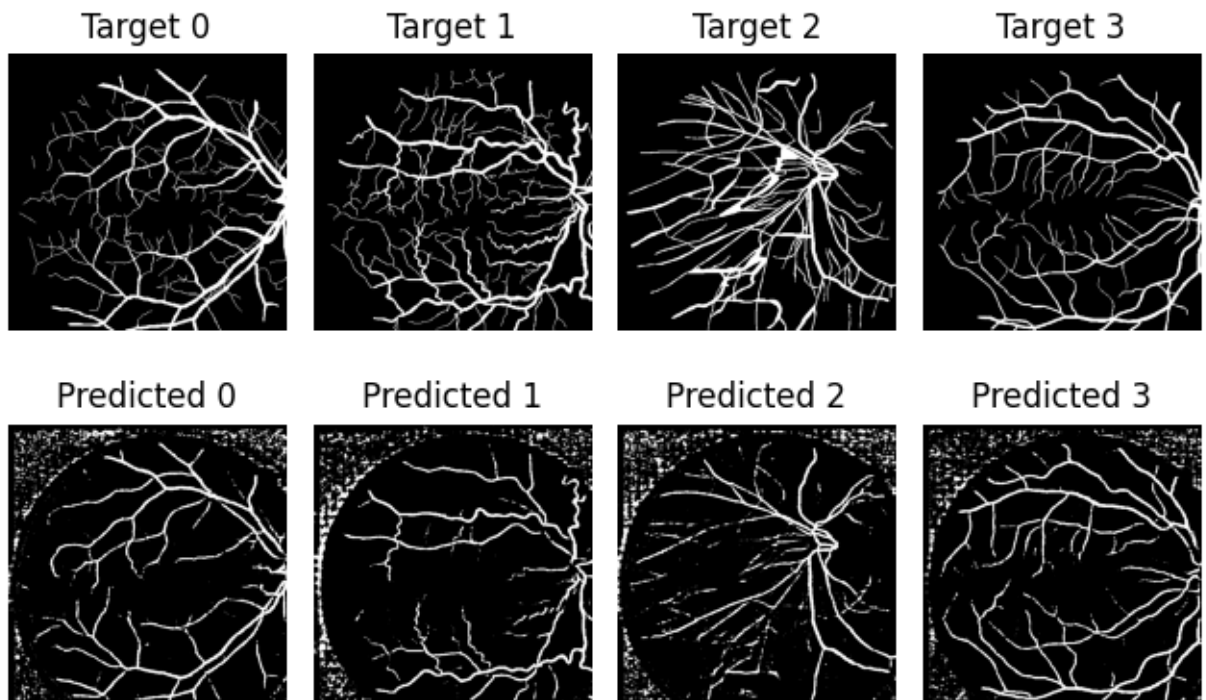
1. Se ha entrenado el modelo UNET sin pre entrenamiento y sin aplicar ninguna técnica de preprocesamiento. Después de 800 épocas, con una tasa de aprendizaje de 0.001 y utilizando el optimizador Adam, se obtuvo el siguiente resultado.



2. Resultados con aumento de datos, como zoom, desplazamiento vertical y horizontal, recorte, rotación, etc. Como se observa, no se ha conseguido cambios significativos respecto al experimento 1.



3. Se han probado una VGG16 y una Resnet50 como codificadores de la UNET. En primer lugar, se congelaron los pesos y no se observó una mejora significativa en ambos modelos. En segundo lugar, se realizó un preprocesamiento de las imágenes para adaptarlos al mismo tipo de entrada con las cuales los modelos pre entrenados fueron ajustados. En tercer lugar, se realizó un entrenamiento completo y se observó que la Resnet50 generaba mejores detalles, pero el conjunto de datos era insuficiente para ajustarlo adecuadamente. Por otro lado, el modelo VGG16 logró estimar imágenes similares al objetivo, aunque perdió algunos detalles en el proceso. Los resultados obtenidos con 200 épocas, Lr = 0.001 y optimizador = AdamW, se muestran a continuación.

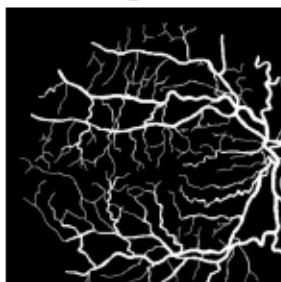


4. Por los resultados del paso anterior se consideró ajustar el learning rate a  $1e-4$  y se incrementó el número de épocas a 400:

Target 0



Target 1



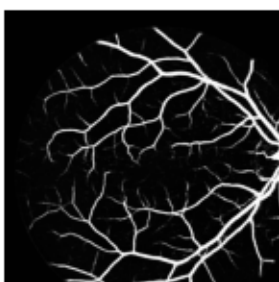
Target 2



Target 3



Predicted 0



Predicted 1



Predicted 2



Predicted 3

