

Trabajo de Fin de Grado:

“Estudio y comparativa de métodos de visualización de redes Deep Learning”

Grado en Ingeniería Electrónica Industrial y
Automática

Alumno: David Baquero Delgado
(alu0100489972@ull.edu.es)

Tutora: Silvia Alayón Miranda



RESUMEN

Este Trabajo de Fin de Grado tiene como objetivo la interpretación del comportamiento de varias redes convolucionales de aprendizaje profundo entrenadas para diagnosticar el glaucoma por medio de retinografías. Se estudiarán diferentes técnicas de visualización y diferentes arquitecturas de redes. El interés del estudio es intentar identificar qué partes de la imagen son relevantes en la decisión de cada red según cada método de interpretación, y ver si los resultados son coherentes entre métodos, entre redes, y con los criterios seguidos por el médico para diagnosticar.

ABSTRACT

The objective of this Final Degree is the interpretation of the behavior of several deep learning convolutional networks trained to diagnose glaucoma by means of retinographies. Different visualization techniques and different network architectures have been studied. The interest of the study is to try to identify which parts of the image are relevant in the decision of each network according to each interpretation method, and to see if the results are consistent between methods, between networks, and with the criteria followed by the doctor to diagnose.

Índice:



1. Introducción.	4
1.1.Descripción del problema médico	4
1.2.Objetivos del Trabajo de Fin de Grado	6
1.3. Estudios previos	7
1.4. Planificación del TFG	7
2. Fundamentos teóricos	9
2.1. Redes de Aprendizaje Profundo (Deep Learning)	9
2.2. Redes Convolucionales	11
2.3. Arquitecturas DL conocidas	13
2.3.1 VGG16	14
2.3.2 VGG19	14
2.3.3 MobileNet	15
2.3.4 MobileNet V2	16
2.3.5 DenseNet	17
2.4. Técnicas de Visualización	17
2.4.1 Grad-CAM	18
2.4.2 Score-Cam	19
2.4.3 Oclusión	20
3. Material y herramientas	22
3.1. Base de datos RIM-ONE DL	22
3.2. Herramientas	22
3.2.1 Google Colab	22
3.2.2 Tensor Flow	22
3.2.3 Keras	22
3.2.4 Keras-vis	23
3.2.5 Pandas	23
3.2.6 Open CV	23
4. Desarrollo Experimental	24
5. Análisis de Resultados	27
5.1. Organización e interpretación de resultados	27
5.2. Primer estudio: análisis de resultados de distintas técnicas de visualización para una misma red	30
5.2.1 VGG16	30
5.2.1.1 Resultados de Grad-CAM	30
5.2.1.2 Resultados de Score-CAM	34
5.2.1.3 Resultados del método de oclusión (OCL)	37
5.2.1.4 Comparativa de resultados de los tres métodos con la VGG16	41
5.2.2 VGG19	41



5.2.2.1 Resultados de Grad-CAM	41
5.2.2.2 Resultados de Score-CAM	45
5.2.2.3 Resultados de OCL	48
5.2.2.4 Comparativa de resultados de los tres métodos con la VGG19	52
5.2.3 MobileNet	52
5.2.3.1 Resultados de Grad-CAM	52
5.2.3.2 Resultados de Score-CAM	56
5.2.3.3 Resultados de OCL	59
5.2.3.4 Comparativa de resultados de los tres métodos con la MobileNet	63
5.2.4 MobileNetV2	63
5.2.4.1 Resultados de Grad-CAM	63
5.2.4.2 Resultados de Score-CAM	66
5.2.4.3 Resultados de OCL	70
5.2.4.4 Comparativa de resultados de los tres métodos con la MobileNetV2	73
5.2.5 DenseNet	73
5.2.5.1 Resultados de Grad-CAM	73
5.2.5.2 Resultados de Score-CAM	77
5.2.5.3 Resultados de OCL	81
5.2.5.4 Comparativa de resultados de los tres métodos con la DenseNet	84
5.3. Segundo estudio: análisis de resultados de todas las redes con una misma técnica de visualización	84
5.3.1 GradCam	85
5.3.2 Score-CAM	88
5.3.3 OCL	91
6. Conclusiones y líneas futuras	95
7. Conclusions and Future lines	96
8. Bibliografía	97



1. Introducción.

En los últimos años hemos visto cómo la Inteligencia Artificial está cada vez más presente en nuestras vidas, el potencial de esta tecnología es enorme pudiendo ser incorporado a cualquier ámbito del que se dispongan una gran cantidad de datos. El uso de esta tecnología puede despertar algunos debates éticos, pero es innegable el impacto positivo que pueden llegar a tener en nuestras vidas cierto tipo de aplicaciones, como por ejemplo las relacionadas con el ámbito médico.

En este Trabajo Fin de Grado (TFG) se estudia la aplicación de una técnica de Inteligencia Artificial, concretamente de las redes neuronales de Aprendizaje Profundo (Deep Learning - DL) en el diagnóstico de una enfermedad oftalmológica, el Glaucoma.

1.1.Descripción del problema médico

El glaucoma es una enfermedad crónica causada por una degeneración progresiva e irreversible de las fibras del nervio óptico. Esta enfermedad ocular afecta a más de 66 millones de personas en el mundo y es la segunda causa principal de ceguera [23].

La retinografía, que es una fotografía del fondo de ojo, es uno de los instrumentos más utilizados en el diagnóstico de enfermedades oftalmológicas, ya que en ella se pueden apreciar detalladamente el estado de las estructuras del ojo (figura 1).

La enfermedad produce un cambio estructural en la cabeza del nervio óptico, en la región elíptica llamada disco óptico o papila que contiene una depresión central llamada copa. En el glaucoma el agrandamiento de la copa óptica ocurre en todas las direcciones pero generalmente las primeras zonas afectadas son la región inferior seguida de la superior, mientras que las regiones temporal y nasal son las últimas en degenerar. Estos sectores se muestran en la figura 2. Esta regla es muy útil en la práctica clínica porque cada variación respecto a la norma es índice en daño glaucomatoso. A una reducción del grosor del borde siempre se acompaña un aumento de la excavación [7].

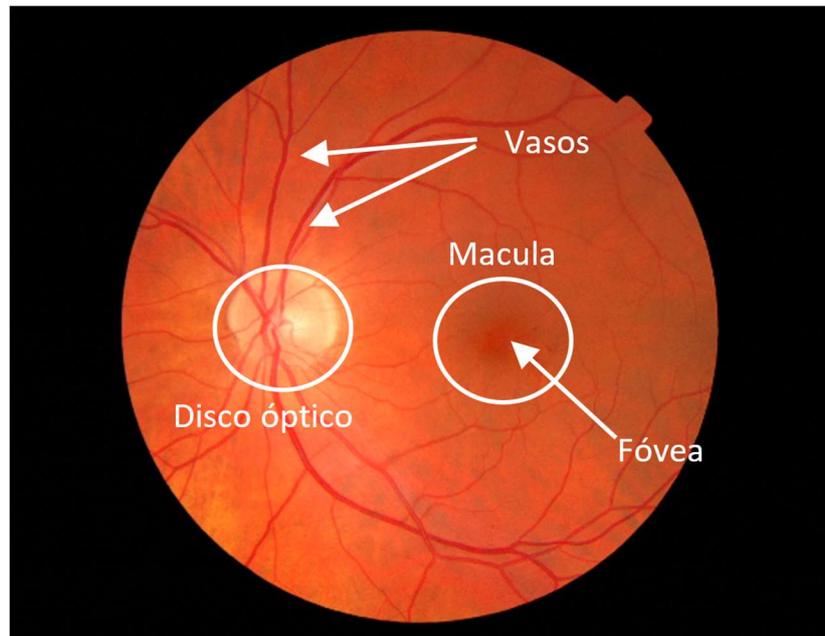


Figura 1. Retinografía: disco óptico, vasos y zona de la fóvea.

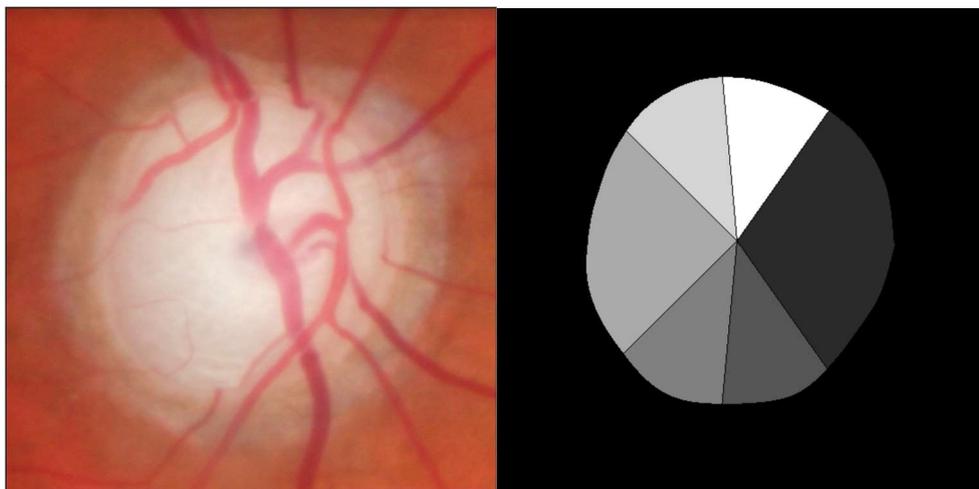


Figura 2. Disco óptico y sectores en los que se divide. El sector más oscuro es el nasal, el opuesto es el temporal, los dos de arriba son los superiores, y los de debajo los inferiores.

Un método estimativo para apreciar la enfermedad con una foto de fondo de ojo es, midiendo la relación del diámetro copa-disco (CDR) es la medida que se usa para cuantificar la deformación del disco óptico. Un valor alto de CDR indica la presencia de glaucoma. En la figura 3 se puede apreciar la diferencia en el tamaño de la copa en un ojo normal y en uno con glaucoma.

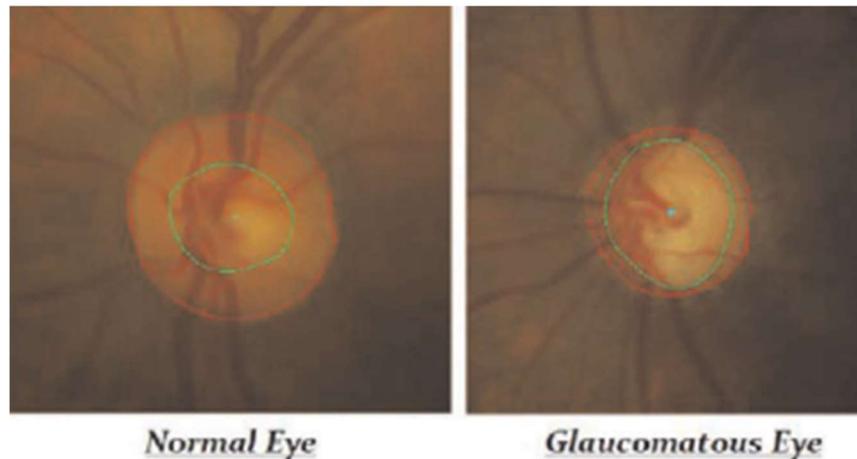


Figura 3. Ojo sano y ojo con glaucoma con la copa marcada. Imagen extraída de [30].

Las fotografías del fondo del ojo son ampliamente usadas para el diagnóstico del glaucoma. El estudio visual de estas imágenes es una operación muy subjetiva, y no siempre sencilla, sobre todo en los casos más incipientes de la enfermedad. Por este motivo, los métodos automatizados podrían reducir costos y efectuar predicciones rápidas y coherentes, ayudando al especialista en este problema concreto.

1.2. Objetivos del Trabajo de Fin de Grado

Uno de los mayores problemas de las Redes Neuronales es que son “cajas negras” para los investigadores. Con un buen entrenamiento son capaces de mostrar soluciones o predicciones excelentes, pero no se sabe qué procesos internos hacen que se comporten así, siendo este uno de sus principales puntos débiles. Saber el porqué, no solo ayudaría a una mejor comprensión del problema, sino también a dar una confiabilidad a los resultados.

El objetivo del proyecto se centra en el estudio de distintas técnicas ya existentes en la literatura científica para la visualización de redes neuronales DL entrenadas para la detección del glaucoma. Estas técnicas han sido utilizadas para visualizar el comportamiento de las redes cuando procesan imágenes naturales, pero no hay muchas publicaciones que estén relacionadas con el problema del diagnóstico del Glaucoma.

Las redes DL analizadas en este TFG han sido entrenadas por el grupo de investigación de Análisis de Imagen Médica de la Universidad de La Laguna, publicadas en [12]. Estas redes son capaces de detectar si un ojo tiene glaucoma analizando una retinografía recortada en torno al disco óptico.

Por medio de diferentes experimentos, se intentará descifrar en este TFG qué puntos o características de la imagen de fondo de ojo son relevantes a la hora de diagnosticar esta enfermedad.



1.3. Estudios previos

En los últimos años ha crecido el interés en la aplicación de las técnicas de Deep Learning sobre retinografías para ayudar al diagnóstico del glaucoma [5, 20, 32, 37]. Estos trabajos están enfocados al diagnóstico de la enfermedad, y a la segmentación de disco óptico y copa, haciendo hincapié en encontrar el diseño de red DL que mejore los resultados de trabajos anteriores. Pero aunque las redes CNN - DL hayan alcanzado un nivel alto en la detección del glaucoma, a menudo estas técnicas son criticadas porque todavía no es evidente cuáles son los patrones que permiten a las redes llegar a la predicción correcta.

Existen varios métodos de visualización propuestos en la literatura científica para intentar descifrar qué partes de la imagen son las más influyentes sobre la decisión de la red, y deducir de ahí qué patrones son los que aprende a distinguir. En el apartado 2.4 se comentarán algunos de estos métodos, por ser los estudiados en este TFG. En la aplicación de métodos DL en la problemática del glaucoma, muy pocos trabajos dedican esfuerzos a intentar visualizar el comportamiento interno de las redes. Los pocos que lo hacen suelen aplicar el método GradCam [26] en la parte final del trabajo, como curiosidad, sin realizar un análisis profundo [10, 13, 17]. El trabajo [35] es el único que hemos encontrado totalmente focalizado a analizar en qué zonas de la imagen retinográfica se fija una red DL (VGG16), estudiando las respuestas de las neuronas de las capas intermedias.

En general, analizando las conclusiones de estos trabajos, se cree que en la detección del glaucoma son importantes la copa óptica y la zona del anillo neuroretiniano. En este TFG queremos comprobar estas impresiones, además de estudiar qué sectores son los más determinantes.

1.4. Planificación del TFG

El presente TFG se organizó en diferentes fases:

- 1) Estudio de las Redes Neuronales DL y sus diferentes arquitecturas para distintas aplicaciones, sobre todo redes convolucionales (Convolutional Neural Network -CNN) especialmente diseñadas para el tratamiento de imágenes.
- 2) Comprobación de la correcta clasificación de las imágenes de control por parte de las diferentes redes y elección de las que formarán parte del estudio.
- 3) Estudio de los métodos de visualización publicados en la literatura.
- 4) Aplicación de los distintos métodos de visualización. De cada experimento se cogieron las redes y los métodos que obtuvieron mejores resultados.



- 5) Análisis de los resultados experimentales obtenidos.
- 6) Discusión y conclusiones.



2. Fundamentos teóricos

En este apartado se recogen los fundamentos teóricos más importantes de las técnicas utilizadas en este TFG: por una parte, se ofrece una descripción de las redes neuronales DL, y dentro de éstas, de las convolucionales (CNN), se presentan las arquitecturas más famosas, y por último, se comentan las técnicas de visualización más utilizadas en la literatura científica.

2.1. Redes de Aprendizaje Profundo (Deep Learning)

Durante las últimas dos décadas, debido a la mejora de los algoritmos, a los nuevos métodos de análisis, a los avances en hardware y a la gran cantidad de datos disponibles, el interés por las redes neuronales ha crecido significativamente, dando lugar a modelos cada vez más complejos (o más “profundos”, redes neuronales con mayor número de capas ocultas).

En 1958 Frank Rosenblatt propuso el Perceptrón [24], la unidad más sencilla de procesamiento en una red neuronal (una neurona), cuya salida es $y = Wx + b$, donde y representa la predicción, W los pesos y b el sesgo, parámetros que se ajustan durante el entrenamiento. La x es la característica (variable de entrada). El resultado pasará por una función de activación no lineal para producir un resultado de 0 o 1.

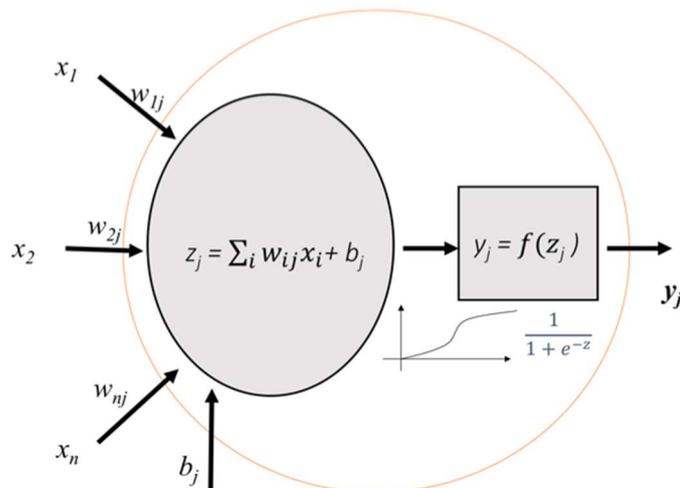


Figura 4. Perceptrón

Una red neuronal es la unión de varias neuronas artificiales conectadas entre sí en capas compartiendo información unas con otras hasta la salida (figura 5). Una vez definido un modelo, este tiene que ser entrenado con el objetivo de que los parámetros de las neuronas (pesos y sesgos) alcancen los valores ideales. Estos valores son los que hacen que la salida de la red sea la esperada ante la entrada que se introdujo.

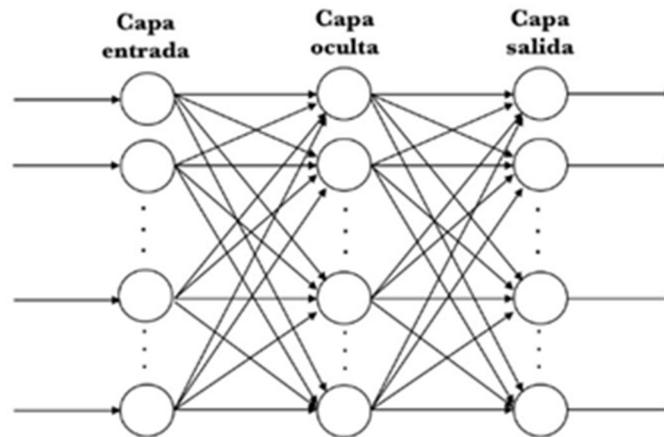


Figura 5. Red neuronal.

En el entrenamiento supervisado los datos de entrada, ejemplos etiquetados con anterioridad, son procesados por la red para ajustar los parámetros y minimizar la función *loss*. Esta función es un concepto importante del Deep Learning y representa cómo de mala ha sido una predicción. La fase de entrenamiento puede resultar fácil, pero la elección de unas variables incorrectas puede dar lugar a un sobreajuste (overfitting), situación en la que la red es incapaz de realizar predicciones de datos fuera del entrenamiento.

Una red de Deep Learning es una red neuronal con múltiples capas internas entre la capa de entrada y la capa de salida (figura 6), generando una jerarquía en la información y dando lugar a mayores niveles de abstracción. Su entrenamiento es un proceso costoso, debido al alto número de parámetros [33].

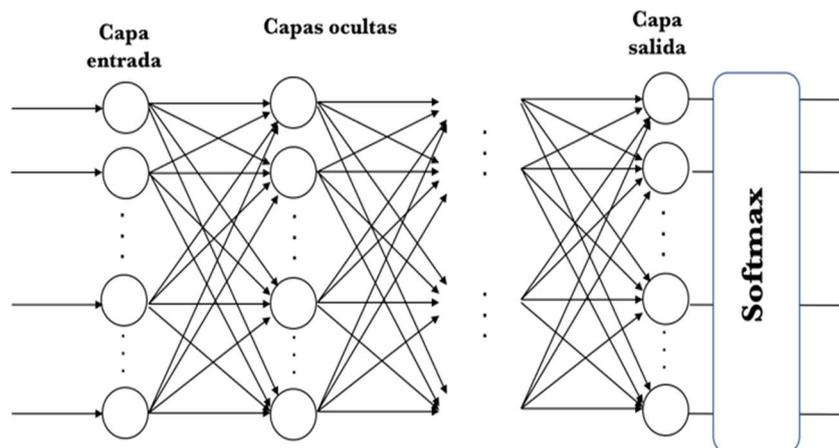


Figura 6. Red neuronal DL.

Existen diferentes tipos de redes neuronales DL. Para el procesamiento de imágenes, es de especial importancia un tipo denominado Red Convolutiva (CNN), que será descrita en la siguiente sección.



2.2. Redes Convolucionales

Una red neuronal Convolutiva (Convolutional Neural Network - CNN) es un tipo particular de red DL utilizada en la visión por computador. Las CNNs están compuestas por varias capas, una de entrada (*Input Layer*), otra de salida (*Output Layer*) y varias capas ocultas internas (*Hidden layer*). La entrada esperada para estas redes es una imagen.

Cada capa está formada por neuronas con diferentes pesos y sesgos que son capaces de aprender patrones espaciales. Esto hace que cada capa sea especialista en el reconocimiento de un patrón y a medida que nos adentramos en la red la complejidad de estos aumenta. Por ejemplo en una red entrenada para el reconocimiento facial las primeras capas serán las encargadas de reconocer los bordes, colores, etc., y las más internas en reconocer ojos, narices, hasta finalmente juntar todos esos patrones aprendidos en capas anteriores para saber con exactitud lo que es un rostro humano (figura 7).

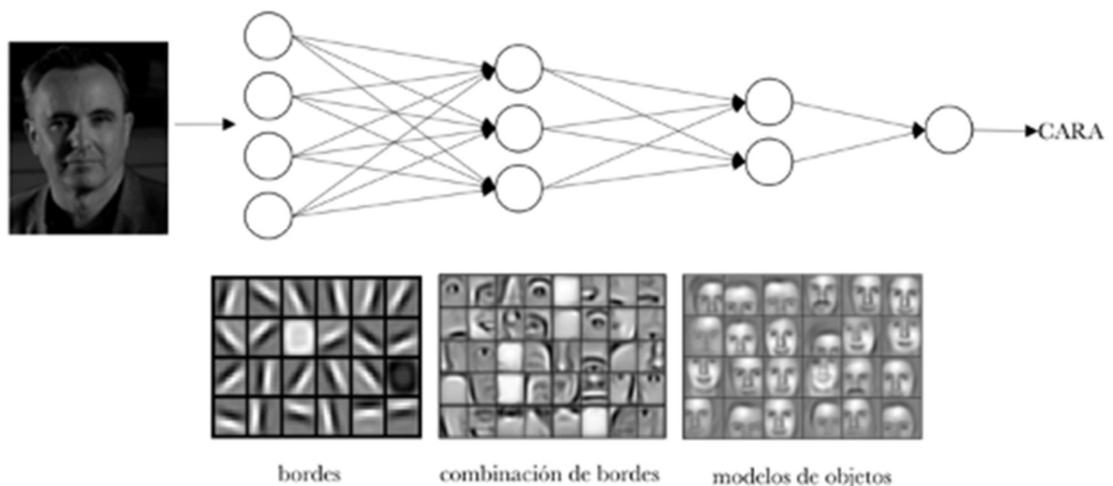


Figura 7. Red entrenada para distinguir caras. Imagen extraída de [33].

La operación de convolución es la principal característica de las redes CNN. A diferencia de las redes densas, que aprenden patrones globales, las redes convolucionales aprenden patrones locales en pequeñas ventanas. Esta propiedad hace que una vez aprendido un patrón pueda localizarlo en cualquier lugar de la imagen, algo que una conexión densa no podría conseguir. Además las capas convolucionales pueden aprender mediante jerarquías, es decir, una primera capa puede aprender elementos básicos y una segunda aprender patrones más complejos formados por esos elementos básicos aprendidos en capas anteriores.

Las capas convoluciones suelen operar sobre tensores de 3D, llamados *feature maps*, con ejes espaciales de altura y anchura (*height* y *width*), además de un eje de canal (*channels*) también llamado profundidad (*depth*). Para una imagen de color RGB, la dimensión del eje *depth* es 3, pues la imagen tiene tres canales: rojo, verde y azul. Para una imagen en blanco y negro, la dimensión del eje *depth*



es 1 (nivel de gris). La convolución opera sobre una capa de entrada y otra capa profunda u oculta (*hidden layer*). Esta conexión no es de todas las neuronas, la imagen de entrada es recorrida de izquierda a derecha y de arriba hacia abajo por una ventana de tamaño específico. Los píxeles o neuronas de esta ventana en cada posición son analizados por una neurona en la siguiente capa (figura 8).

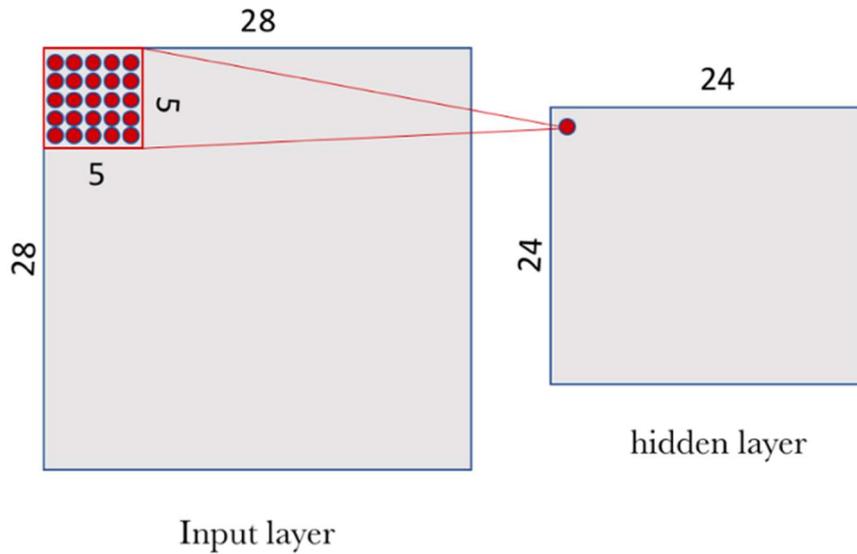


Figura 8. Operación de convolución [33].

Como comentamos anteriormente, cada neurona consta de un sesgo b y una matriz de pesos W (filtro o kernel). Siguiendo el ejemplo mostrado en la figura 8, la capa oculta tiene 24×24 neuronas, y a cada una le corresponde al producto escalar entre el filtro y las 25 neuronas encerradas en la ventana 5×5 de la capa de entrada. Lo importante de las capas convolucionales es que se utiliza el mismo sesgo y filtro en todas las neuronas de una misma capa.

Normalmente después de las capas convolucionales existen unas capas de “pooling”. Esta operación condensa la información de la capa anterior respetando la relación espacial. Un ejemplo se puede ver en la figura 9, donde aplicando una ventana 2×2 se reduce el tamaño de la capa a la mitad.

En una CNN se concatenan capas convolucionales y capas de pooling, de manera que a medida que se va aumentando el número de filtros se va disminuyendo la dimensión espacial.

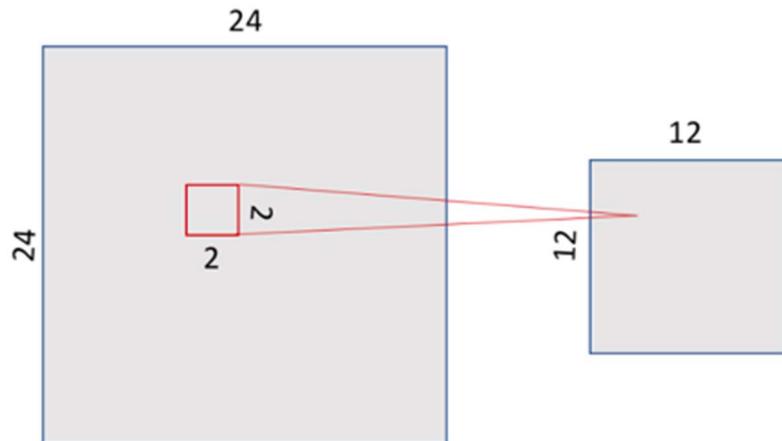


Figura 9. Operación de pooling. Imagen extraída de [33]

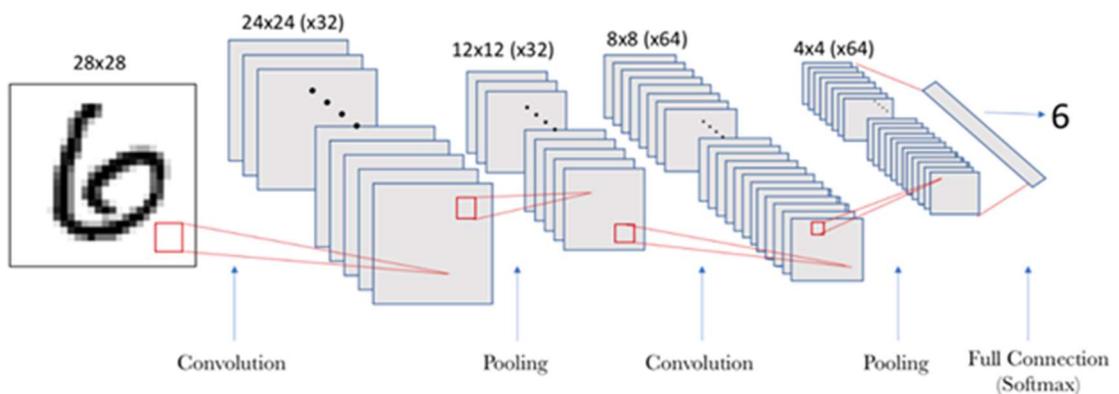


Figura 10. Estructura de una CNN. Imagen extraída de [33].

2.3. Arquitecturas DL conocidas

El avance en la visión por computador que se ha logrado con las redes CNN ha traído consigo en las últimas décadas mejoras en diferentes aspectos como, el diseño de las capas, funciones de activación y pérdida, regulación, optimización y cálculo rápido. Esto ha provocado la aparición de diferentes arquitecturas reconocidas por la comunidad con nombres específicos. Con frameworks como Keras [21] podemos importar estas redes pre-entrenadas, ajustadas para clasificar las imágenes naturales de la base de Imagenet [2], y aplicarlas sobre imágenes de diferente naturaleza (por la propiedad de *Transfer Learning*) reduciendo de forma considerable la cantidad de datos necesarios para el entrenamiento y los tiempos del mismo. A continuación, detallamos las utilizadas en este TFG.



2.3.1 VGG16

Esta red fue propuesta por K. Simonyan y A. Zisserman, fue uno de los modelos más famosos del congreso ILSVRC-2014 consiguiendo alcanzar una precisión del 92,7% con el conjunto de datos ImageNet. Muestra una arquitectura lineal, donde la entrada, una imagen RGB 224×224 , pasa por una serie de capas convolucionales aplicándose filtros de tamaño 3×3 (mínimo tamaño espacial, arriba-abajo, izquierda-derecha) con un desplazamiento de un píxel. La entrada en cada una de las capas siempre respeta la distribución espacial, y después de determinadas capas convolucionales les sigue una capa de max-pooling (5 en total). La agrupación máxima se realiza en ventanas 2×2 con 2 píxeles de desplazamiento. Por último los datos entran en una capa softmax 1×1 , originando una transformación lineal de los resultados. Todas las capas utilizan la función ReLu[16]. Esta arquitectura se muestra en la figura 11.

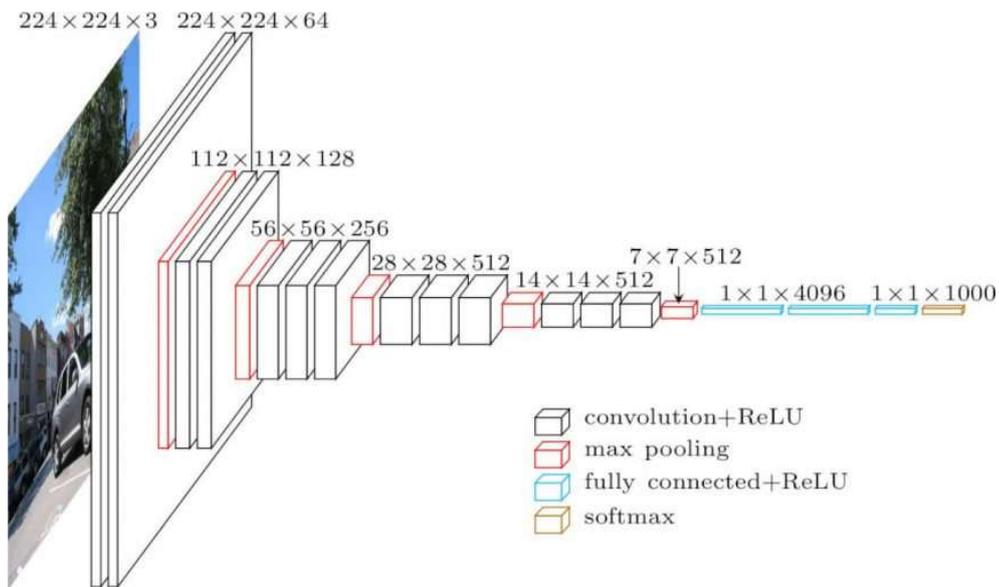


Figura 11. Arquitectura VGG16. Extraída de [16]

2.3.2 VGG19

La *VGG19* fue propuesta por los mismos investigadores que la *VGG16*. La única diferencia radica en poseer una mayor profundidad (figura 12). Con este aumento de la profundidad se logran mejores resultados de clasificación [16].

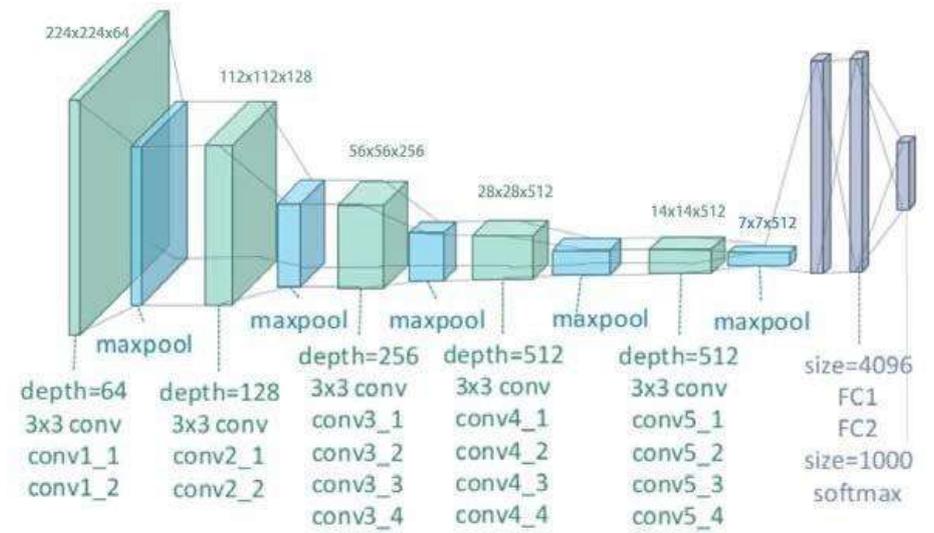


Figura 12. Arquitectura VGG19. Extraída de [8]

2.3.3 MobileNet

MobileNet es una arquitectura propuesta por Google pensada principalmente para aplicaciones de visión móviles. Por tanto, el objetivo final siempre fue optimizar la potencia de cómputo lo máximo posible para poder operar en este tipo de dispositivos. La arquitectura *MobileNet* se basa en factorizar las convoluciones tradicionales en dos tipos de capas, una primera capa convolucional *depthwise* y una capa convolucional 1x1 *pointwise*. Esta división permite reducir el tiempo de respuesta de la red y reducir el tamaño del modelo [19].

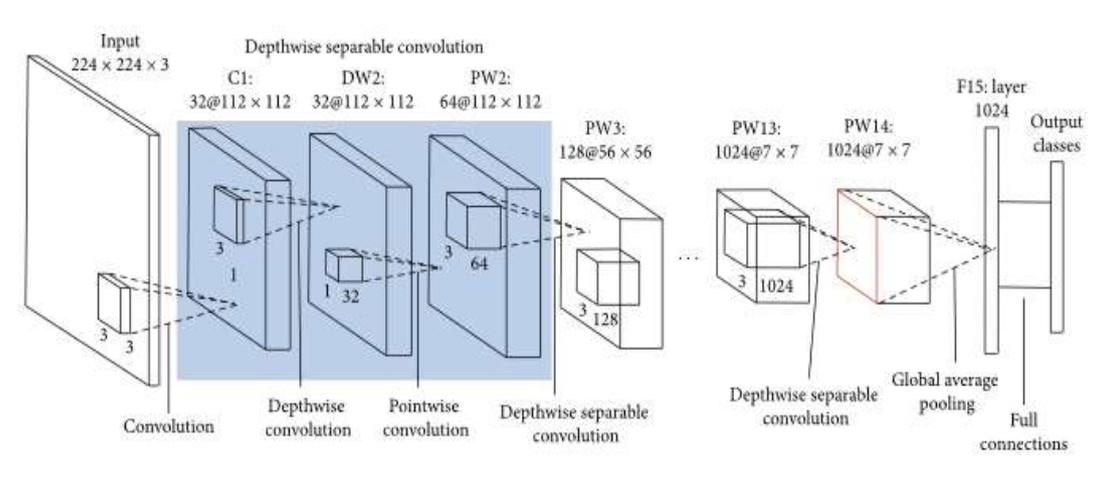


Figura 13. Arquitectura Mobilenet. Extraída de [19]

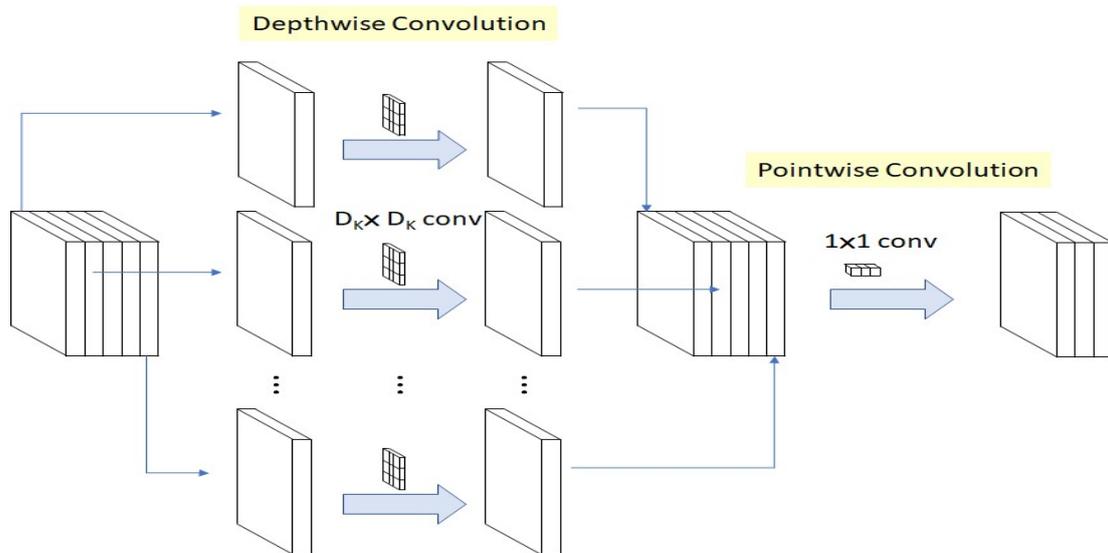


Figura 14. Operaciones Mobilenet. Extraída de [19]

2.3.4 MobileNet V2

La *MobileNet V2* es también creación de Google y es una evolución de la primera generación de *MobileNet*. Esta red requiere el 70% de los parámetros de su predecesora, y esta evolución ha traído consigo un menor tiempo de computación para obtener la misma precisión [18].

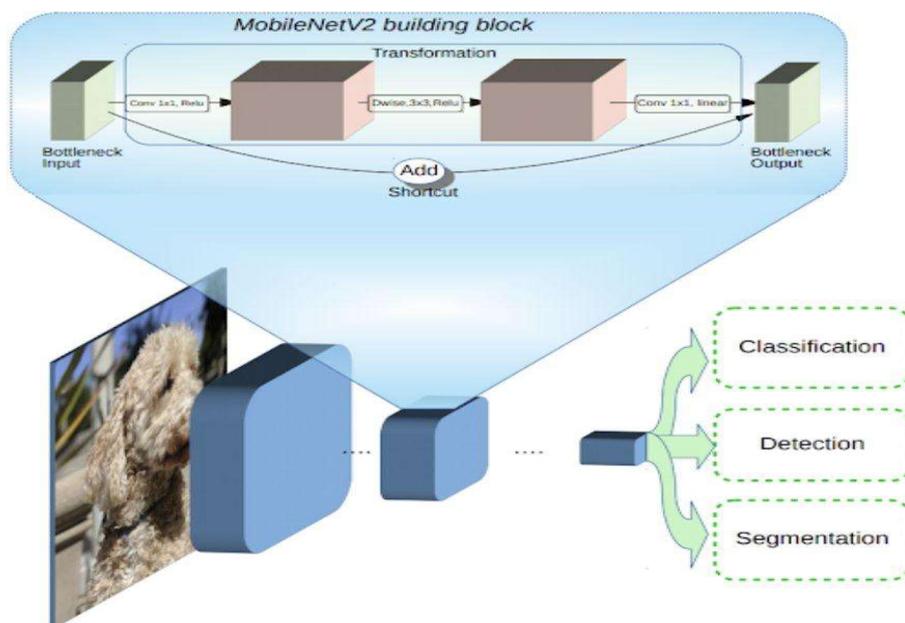


Figura 15. Esquema de las mejoras de MobilenetV2. Extraída de [25]



2.3.5 DenseNet

La *DenseNet* es la arquitectura más compleja utilizada en este TFG. Esta red tiene tres formas de tratar los mapas de características entre las diferentes capas convolucionales: *Standard Connectivity*, *Residual/Skip Connectivity*, y *Dense Connectivity*. Con estas operaciones cada capa puede obtener tanto información básica como información profunda de capas predecesoras. El tamaño del mapa de características es el mismo en cada capa para facilitar su fusión y existe una capa de agrupación entre los diferentes bloques para reducir el tamaño del mapa [19].

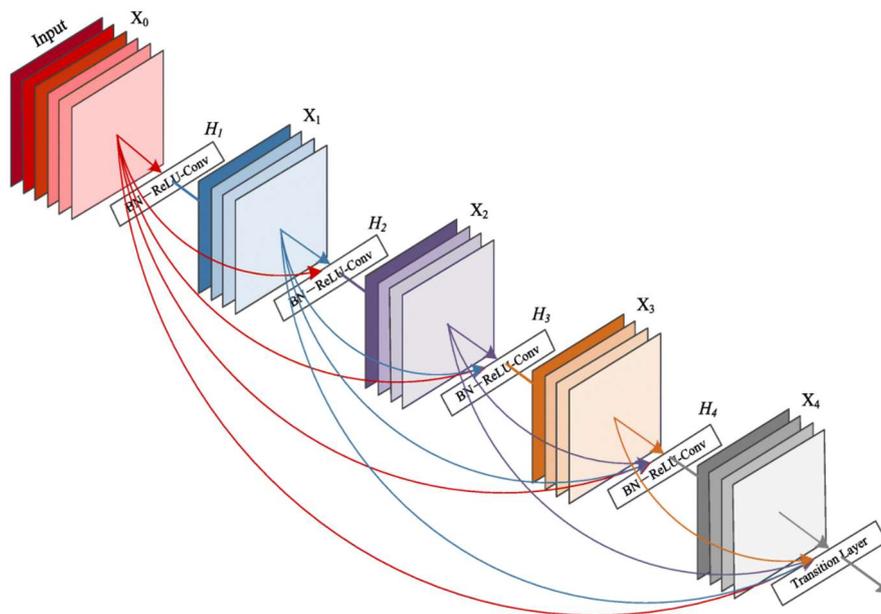


Figura 16. Densenet. Operaciones *Standard Connectivity*, *Residual/Skip Connectivity*, y *Dense Connectivity*. Extraída de [19]

2.4. Técnicas de Visualización

Como se ha comentado anteriormente, las CNN están cada vez más presentes en el análisis de imagen y han sido de las más investigadas en los últimos años, lo que ha dado lugar a avances espectaculares en su eficiencia. A pesar de su éxito, estas redes presentan un gran inconveniente: la falta de transparencia. Se sabe que funcionan muy bien, pero no es fácil comprender los motivos, ya que no ofrecen una explicación de las decisiones que toman internamente. Por esto normalmente se utilizan como cajas negras. Pero una explicación interna de su comportamiento es elemental en muchos ámbitos, entre ellos, en Medicina.

Para intentar explicar este comportamiento, en los últimos años han surgido diversos métodos de visualización. Los más populares, conocidos como “métodos de atribución”, tienen como objetivo identificar las características



(features) presentes en la imagen de entrada que dan lugar a la predicción final del modelo.

Los métodos basados en *back-propagation* normalmente calculan el gradiente (o alguna de sus variantes) de una salida en particular con respecto a la entrada usando *back-propagation* para derivar la contribución de las características, dando lugar a mapas que indican qué partes de la imagen de entrada son más relevantes, los *saliency maps* [29].

Por otra parte, los métodos de perturbación también intentan descubrir qué características de una imagen de entrada son las más influyentes en la decisión de la red. Parten de la suposición de que la contribución de una característica puede ser determinada midiendo cómo cambia la predicción cuando esa característica es alterada [35]. El gran inconveniente de estos métodos es su alto coste computacional, ya que es necesario que vayan perturbando la entrada de manera secuencial. Además, el modo en que se va a perturbar la imagen también puede cambiar mucho el resultado final, y no está claro cómo se puede hacer esto de manera óptima.

Tanto los métodos de interpretación basados en la perturbación como los basados en *back-propagation* ignoran las capas intermedias de la red, que podrían contener información importante sobre la interpretabilidad de la misma. Hay un tercer tipo de métodos que investigan lo que ocurre en las capas ocultas de la red, para intentar determinar qué características de la entrada son más relevantes. Dentro de estos métodos podemos citar a CAM [28], GradCAM [26] y ScoreCAM [36].

En este TFG se han realizado pruebas con diversas técnicas de interpretación. A continuación, se describirán con más detalle las que finalmente han sido seleccionadas.

2.4.1 Grad-CAM

La técnica de GradCAM (*Gradient-weighted Class Activation Mapping*) [26] tiene gran flexibilidad de uso, ya que puede ser implementada en gran cantidad de arquitecturas CNN sin necesidad de reentrenar la red ni cambiar su estructura. La técnica se basa en observar qué neuronas se activan y son más relevantes siguiendo la información del gradiente que fluye hasta ellas para un caso particular. Aunque este procedimiento puede ser aplicado a cualquier capa, se suele aplicar a la última capa convolucional por poseer mayor información.

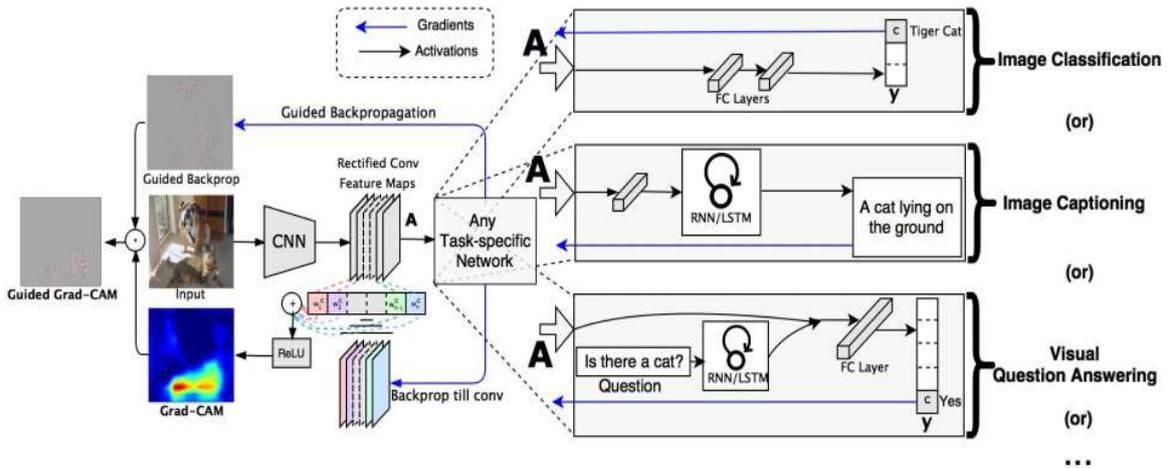


Figura 17. Esquema del algoritmo de Grad-Cam. Extraída de [26]

Para obtener los mapas, primero se calcula la puntuación del gradiente para una clase c , calculando la derivada parcial de una salida y^c (antes de la capa softmax) respecto el mapa de características de una determinada capa convolucional $A^k \rightarrow \frac{\partial y^c}{\partial A_{ij}^k}$. Estos gradientes calculados fluyen hacia atrás para obtener los pesos mas importantes de la neurona a_k^c :

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

Después de calcular los pesos para la clase objetivo c , se realiza una combinación ponderada de los mapas de activación, seguida de una unidad lineal rectificadora (ReLU).

$$L_{Grad-Cam}^c = ReLU(\sum_k \alpha_k^c A^k)$$

La importancia del ReLU reside en que permite diferenciar los píxeles importantes con más influencia positiva para aumentar y^c . Sin el uso de esta función los resultados son un 15,3% peores, mostrando una localización de la clase objetivo menos exacta y señalando otras zonas que no tienen interés en el problema.

2.4.2 Score-Cam

Los autores de Score-CAM [36] afirman que este método es mejor que el Grad-CAM porque, al no depender de los gradientes, evita ruidos irrelevantes y genera explicaciones más limpias y significativas. Es una mezcla entre el grad-CAM y los métodos de oclusión, en el sentido de que hace oclusión en la imagen de entrada, pero para ello utiliza los mapas de características (activation maps) de la última capa de convolución como máscaras para hacer esta oclusión (previamente ajustan el tamaño de estos feature maps al tamaño de la imagen original). De esta manera



obtienen una puntuación para cada mapa de características que indica la importancia de este mapa en la clase predicha finalmente. Estas puntuaciones se usan como pesos, y se hace la suma pesada de los mapas de características con estos pesos, para calcular el mapa de calor final.

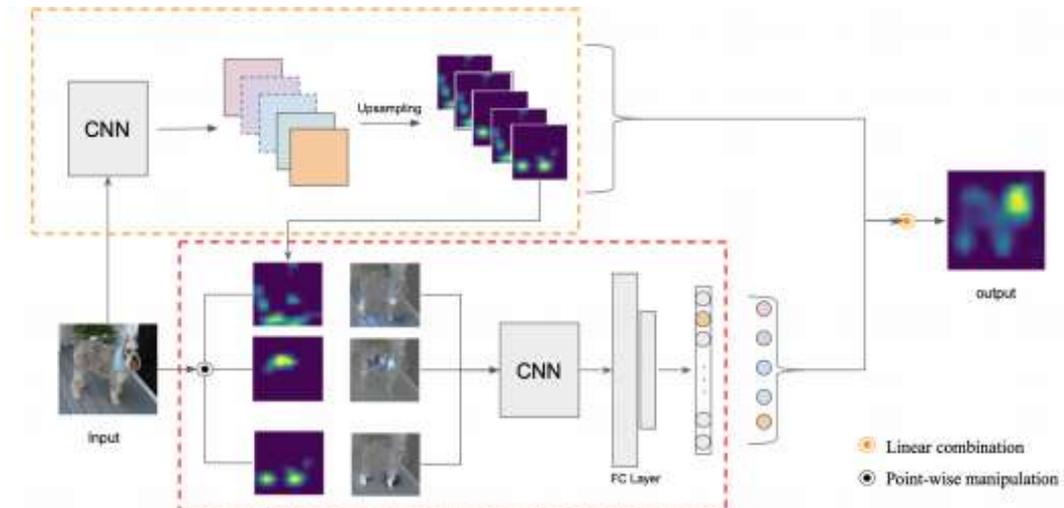


Figura 18. Esquema algoritmo Score-Cam. Imagen extraída de [36].

Considerando una capa de la CNN en un modelo f con una clase determinada c , con su correspondiente activación A , donde k es el k -ésimo canal de A_l quedando A_l^k , se define el Score-CAM como:

$$L_{ScoreCam}^c = ReLu\left(\sum_k \alpha_l^c A_l^k\right)$$

donde los pesos son $\alpha_l^c = C(A_l^k)$, con C siendo el coeficiente de puntuación de la correspondiente capa. La aplicación del ReLu en la combinación lineal tiene el mismo objetivo que en el método Grad-CAM.

2.4.3 Oclusión

Los métodos de oclusión son métodos de perturbación. Inicialmente fueron concebidos para saber si era importante el entorno de una imagen a la hora de hacer la predicción. Los resultados fueron negativos, demostrando que las CNN son capaces de discriminar su objetivo, ya que si el objeto a clasificar estaba ocluido la probabilidad de acierto bajaba considerablemente.

La implementación más sencilla de esta técnica consiste en tapar la imagen de entrada con una ventana de negro o de gris de tamaño x que la irá cubriendo mientras se monitorea la salida. Los resultados pueden ser descritos como parches que estimulan los distintos mapas de características. Este modo de oclusión por ventanas es el utilizado en este TFG, pero existen otros modos de



realizar perturbaciones sobre la imagen. Por ejemplo, en [9, 34] se proponen diversos métodos para aprender la máscara de perturbación más óptima



3. Material y herramientas

3.1. Base de datos RIM-ONE DL

En este TFG se han utilizado las imágenes de la base de datos RIM-ONE. RIM-ONE (**R**etinal **I**Mage database for **O**ptic **N**erve **E**valuation) es una base de imágenes de retinografías de libre acceso diseñada para evaluar y comparar métodos de segmentación de disco óptico y excavación. En su última versión ofrece 313 retinografías de ojos sanos y 174 de ojos con glaucoma para la evaluación de algoritmos de DL. Las imágenes se encuentran recortadas alrededor del disco óptico y etiquetadas por tres médicos especialistas [12]. La base está accesible en [39].

3.2. Herramientas

3.2.1 Google Colab

Se trata de un servicio de Google [14] en la nube inspirado en el Jupyter Notebook. Como este, se trata de un entorno de trabajo en el navegador que permite ejecutar y programar en Python. Al igual que otros productos de Google, los cuadernos pueden ser compartidos a otros usuarios. Los documentos resultantes son entornos dinámicos por lo que pueden ser modificados y ejecutados por otros usuarios desde cualquier lugar. Además cuentan con un servicio de acceso a GPUs de forma gratuita, esto es muy práctico ya que ofrece la posibilidad de adentrarse en el mundo del Machine Learning sin la necesidad de disponer de un ordenador de altas capacidades.

3.2.2 Tensor Flow

Es una librería de código abierto enfocada en Machine Learning [15]. Es un producto desarrollado por Google para crear y entrenar redes neuronales. La versión más reciente es la 2, aunque en este TFG se ha usado la primera.

3.2.3 Keras

Es un framework de Python que facilita la creación de modelos de *Deep Learning* [21]. Debido a su gran popularidad ha sido integrado en el núcleo de Tensor Flow. Su éxito se debe a la versatilidad y simplicidad para programar una red neuronal, ya que con pocas líneas se puede crear un modelo. Aparte es una buena herramienta para los desarrolladores permitiendo entrar en profundidad en las redes neuronales.



3.2.4 Keras-vis

Es una librería que funciona bajo Keras importando una serie de funciones [22]. Estas funciones son bastante útiles porque proporcionan métodos de visualización de filtros convolucionales, capas densas, técnicas de visualización, etc. Hay que destacar que la librería actualmente no se encuentra actualizada, por lo que solo funciona en la versión 1.0 de Tensor Flow.

3.2.5 Pandas

Es una biblioteca de python distribuida como software libre bajo la licencia BSD [1]. Programado junto a numpy es una herramienta para la manipulación y análisis de datos.

3.2.6 Open CV

Esta librería [4] fue desarrollada en C++ pero puede ser ejecutada bajo varios lenguajes como, por ejemplo, Python. Todas sus funciones están enfocadas en el tratamiento de imágenes por lo que es ampliamente utilizada en la visión artificial.



4. Desarrollo Experimental

La experimentación realizada en este TFG ha sido desarrollada en cuatro fases:

Fase 1: Elección de las redes neuronales DL

Como se ha mencionado anteriormente, para la investigación planteada en este TFG contamos con las imágenes retinográficas de ojos sanos y con glaucoma de la base RIM-ONE DL. Además, tenemos disponibles varias redes neuronales DL ya entrenadas con esta base de datos; *Xception*, *VGG16*, *VGG19*, *ResNet50*, *Inception V3*, *InceptionResNetV2*, *MobileNet*, *DenseNet21*, *NASNetMobile*, y *MobileNetV2*, publicadas en [12]. Esto nos permite centrar el trabajo en la parte de visualización de las redes ya entrenadas.

Lo primero que se hizo fue clasificar todas las imágenes de la base con estas redes. Ninguna red presenta un 100% de eficiencia, así que es necesario descartar las imágenes en las que se equivoca cada una, ya que el interés de este TFG es analizar en qué se fija la red cuando hace un diagnóstico correcto.

Todas las redes, excepto la *NasNetMobile*, clasificaron bien la mayoría de imágenes, motivo por el que fue descartada del estudio. Durante el mismo, otras redes fueron también descartadas: por ejemplo, la *ResNet50* se eliminó debido a que su arquitectura interna no permite la aplicación de técnicas de visualización dependientes del gradiente. Más adelante se tuvieron que descartar las redes de la familia *Inception* (*Xception*, *Inception V3* y *InceptionResnetV2*). El motivo fue que estas redes fueron entrenadas sobre RIM-ONE con imágenes de tamaño 224x224, pero sus autores indican que están entrenadas originalmente sobre ImageNet con imágenes de 256x256. A pesar de esto, los resultados fueron buenos, pero, no pudiendo asegurar que se respetara la correcta relación espacial en los mapas de características, no se continuó haciendo pruebas sobre ellas.

Finalmente se eligieron las cinco redes descritas en la sección 2.3: *VGG16*, *VGG19*, *MobileNet*, *MobileNetV2* y *DenseNet*.

Fase 2: Elección de las técnicas de visualización

Una vez elegidas las redes que forman parte del estudio, se comenzó a probar en ellas diferentes técnicas de visualización. Las técnicas finalmente seleccionadas fueron *Grad-CAM*, *Score-CAM* y *oclusión* (descritas en la sección 2.4).

Aparte de estas técnicas se experimentó con otras como [6, 29, 28, 31]. La decisión de no incluirlas en los experimentos finales se debe a la dificultad de interpretabilidad que muestran sus resultados. Los métodos descartados funcionan realmente bien cuando procesan imágenes naturales, según se puede apreciar en los artículos de estos autores. Pero en el caso de las retinografías la



interpretación es muy confusa. Estos métodos ofrecen mapas *saliency* muy ruidosos, que marcan píxeles aislados en vez de regiones de interés. Un ejemplo se muestra en la figura 19.



Figura 19. Mobilenetv2. Imagen de un ojo normal con la técnica ELPR[6]

Fase 3: Elección de código

El estudio realizado en la fase 2 sobre técnicas de visualización existentes implicó la búsqueda del código de los diferentes métodos. Es importante indicar que, para un mismo método, a menudo hay disponibles en la red muchos códigos de diferentes usuarios.

En el caso de Grad-CAM no existe código disponible directamente de los autores. Estudiamos 3 códigos diferentes que aseguraban implementar bien el método, y comparamos los resultados. Finalmente nos decidimos por el ofrecido en la librería Keras-Vis [22] por estar mejor documentado.

Con el Score-CAM se siguió el mismo procedimiento que en el caso anterior: se probaron 3 códigos diferentes y se compararon los resultados. Finalmente, el código [3] fue el que funcionó mejor.

El método de oclusión fue implementado con el código que se ofrece en [11].

Fase 4: Realización de los experimentos

Una vez escogidas las técnicas de visualización, los códigos que las implementan y las redes que forman parte del estudio, se procedió a introducir en cada red todas las imágenes de la base RIM-ONE DL que están bien clasificadas por dicha red. Los resultados obtenidos son los mapas que indican qué zonas de la imagen de entrada fueron más relevantes en la decisión final de la red, los denominados “mapas de calor”. Estos mapas de calor se asemejan a una imagen térmica que representa las “zonas calientes” de la imagen.



Como se mencionó en la descripción de estos métodos, para calcular el Grad-CAM y el Score-CAM es necesario utilizar la última capa convolucional de cada red. En las VGG esta capa es fácil de localizar, pero en redes más profundas el uso del comando `keras.utils.plot_model()` resulta muy útil, ya que muestra de forma gráfica la arquitectura de la red. A continuación se indica el nombre de las capas que se utilizaron para cada una de las redes:

- **VGG16** `block5_conv3`
- **VGG19** `block5_conv4`
- **MobileNet** `conv_pw_13`
- **MobileNetV2** `Conv_1`
- **Densenet** `conv5_block16_concat`

Hemos observado que elegir la capa adecuada es algo bastante crítico en estos métodos, pero con las capas anteriormente indicadas los resultados eran válidos. Sólo en el cálculo del Score-CAM sobre la *MobileNet* la capa seleccionada cambió, pasando a ser la penúltima capa convolucional (`conv_pw_12`). Cambiamos de capa al ver que los resultados con la última capa eran nulos (los mapas de calor que se obtenían estaban vacíos).

Con respecto al método de oclusión, se hicieron pruebas con ventanas de diferentes tamaños (20x20, 40x40, 60x60) con sus píxeles a 0 (en negro). Uno de los inconvenientes de esta técnica es que no se sabe a priori cuál es el tamaño de ventana más adecuado. Ventanas muy pequeñas, aparte de requerir más tiempo para la ejecución, pueden llegar a no ser suficientemente grandes para ocultar áreas de relevancia, dando lugar a mapas con escasa información. Por el contrario, con ventanas más grandes el tiempo de ejecución disminuye, pero puede originar mapas enfocados en zonas de la imagen sin interés, como por ejemplo, las esquinas. Por todo esto se decidió probar con tres tamaños de ventana diferentes, y finalmente se eligió el tamaño intermedio para realizar los experimentos. En la figura 20 se muestran mapas de calor obtenidos con distintos tamaños de ventana.

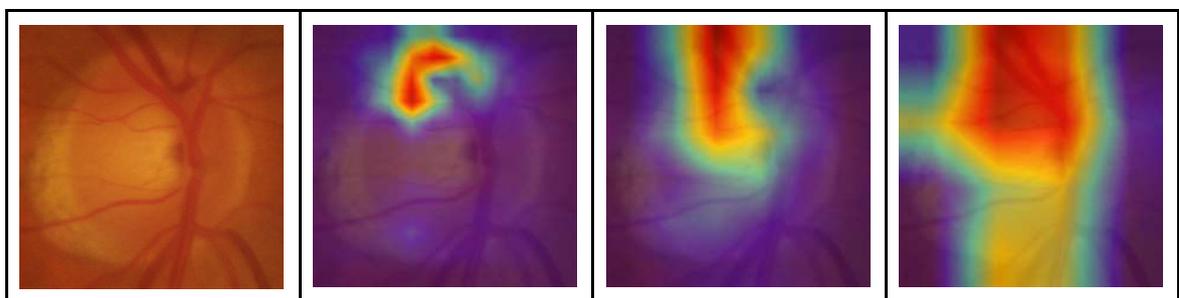


Figura 20. Resultados de la VGG16 .Representados de izquierda a derecha de tamaños pequeños a mayores.



5. Análisis de Resultados

5.1. Organización e interpretación de resultados

Una vez realizados todos los experimentos, se debe buscar una estrategia para poder analizar todos los resultados, debido al alto número de resultados obtenidos. Haciendo un cálculo aproximado sobre los experimentos realizados: con 5 redes, 3 técnicas de visualización diferentes, y las imágenes bien clasificadas por cada red (de media, 487 imágenes con más de un 90% de acierto), se tienen finalmente más de 6500 mapas de calor que analizar.

El principal objetivo de este TFG es averiguar qué características presentes en la imagen de un ojo son determinantes en la predicción de la red. El problema con el tipo de imágenes analizadas (retinografías) es la dificultad de relacionar la zona marcada de la imagen con un patrón característico fácil de interpretar. Cuando se analizan imágenes de perros, gatos, coches, etc., se puede ver de forma clara si su funcionamiento es acertado porque las estructuras en las que se fija tienen significado semántico para nosotros (una cara, las patas, las ruedas, etc), pero cuando no podemos hacer este tipo de relaciones, hay muchas suposiciones posibles.

Por todo esto, con el fin de hacer un análisis lo más resolutivo posible, se presentan dos análisis diferentes:

- Primero se estudiará, para una misma red, los mapas de calor generados con los diferentes métodos de visualización. Con este estudio se quiere saber si los métodos son coherentes entre sí, y marcan las mismas regiones en las imágenes, lo que indicaría que esas regiones son las más importantes para la decisión de esa red en particular.
- El segundo estudio será observar, para una misma técnica de visualización, lo que se obtiene con diferentes redes. La principal razón de este análisis es saber si existen zonas comunes de interés entre las redes.

Ambos estudios son estudios comparativos. Los mapas de calor obtenidos se umbralizan (umbral = 0.5), para trabajar con la parte más significativa de los mismos. En la figura 21 se muestra este proceso gráficamente.

Se analiza en qué regiones de la retinografía están situadas estas partes significativas de los mapas. Para ello disponemos de las imágenes de RIM-ONE DL segmentadas y etiquetadas por regiones: disco, copa, vasos y sectores (expuestos en la sección 1.1).

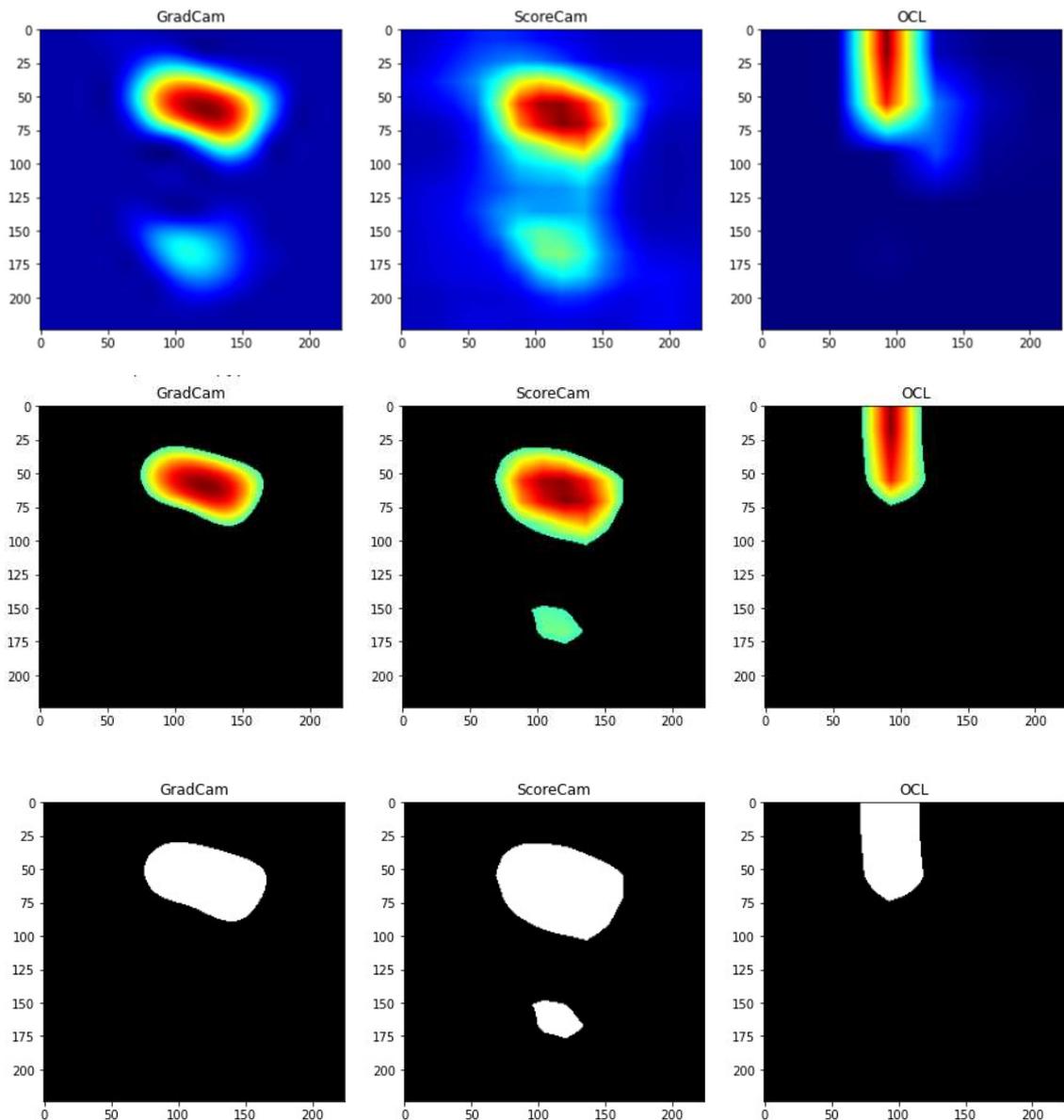


Figura 21. Mapas de calor obtenidos para una imagen con Grad-CAM, Score-CAM y oclusión, y proceso de umbralizado de los mismos.

Para calcular el grado de coincidencia entre el mapa de calor y una región segmentada de la retinografía, se hizo uso del índice de Jaccard o IoU (*Intersection over Union*) [27]. Esta técnica proporciona un valor o índice de similitud entre dos conjuntos, en este caso imágenes. Su explicación matemática es:

$$IoU = \frac{|A \cap B|}{|A \cup B|}$$

donde el numerador representa intersección de dos conjuntos y el denominador su unión (figura 22).

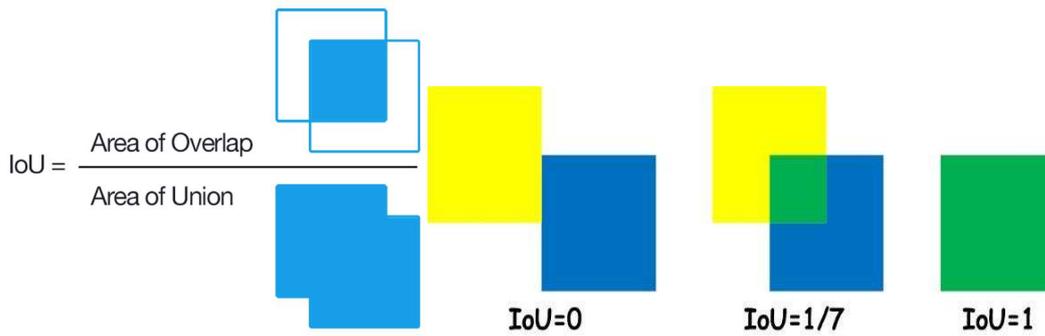


Figura 22. Explicación gráfica del índice IoU.[27]

El índice IoU se calcula entre todos los mapas umbralizados y las distintas regiones segmentadas en las retinografías. Para poder interpretar rápidamente los valores de IoU obtenidos, se plantea una representación en forma de histograma para cada una de las regiones consideradas (disco, copa, vasos y sectores). Estos histogramas estarán divididos en resultados para ojos con glaucoma y para ojos normales. En la figura 23 se presenta uno de estos histogramas como ejemplo. Es el resultado de comparar los mapas de calor de GradCAM obtenidos para una determinada red con la región “copa” para las imágenes normales de RIM-ONE DL clasificadas correctamente por esa red. En el eje x se representa el índice IoU, entre 0 y 1, espaciado en 5 intervalos. En el eje y se muestra el número de imágenes de ojos normales que presentan un valor de IoU correspondiente a ese intervalo. Como se puede observar en este caso, parece que la red se fija en más de 120 imágenes normales en la copa.

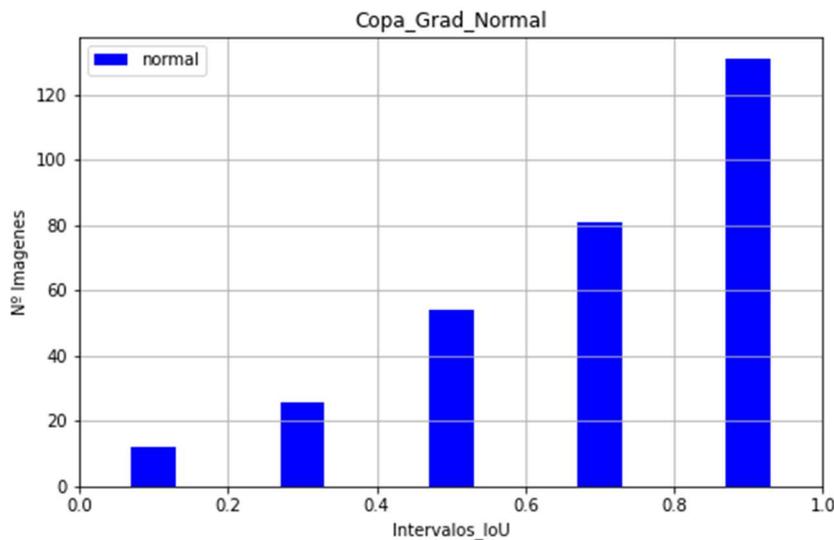


Figura 23. Ejemplo de histograma

Para simplificar la interpretación de los resultados, en los análisis se tomarán en cuenta los casos extremos, cuando el IoU esté por debajo de 0,20 (interpretando esa región como poco relevante) y cuando esté por encima de 0,80 (indicando una región muy relevante).



5.2. Primer estudio: análisis de resultados de distintas técnicas de visualización para una misma red

Se estudiarán los mapas de calor resultantes de cada técnica de visualización aplicados sobre una misma red. El interés es averiguar si los distintos métodos de interpretación empleados tienen coherencia entre sí, observando el valor de los índices IoU representados en los histogramas.

Los datos serán expuestos de la siguiente manera: primero se mostrarán algunas imágenes resultantes a modo ejemplo (por cuestión de espacio es imposible mostrarlas todas), y después se ofrecerán los histogramas obtenidos de IoU entre los mapas obtenidos y las regiones del ojo segmentadas. Los histogramas con barras rojas son los histogramas de las imágenes de glaucoma y los de barras azules son los correspondientes a ojos normales. Por último, se presentará una breve discusión de los resultados.

5.2.1 VGG16

5.2.1.1 Resultados de Grad-CAM

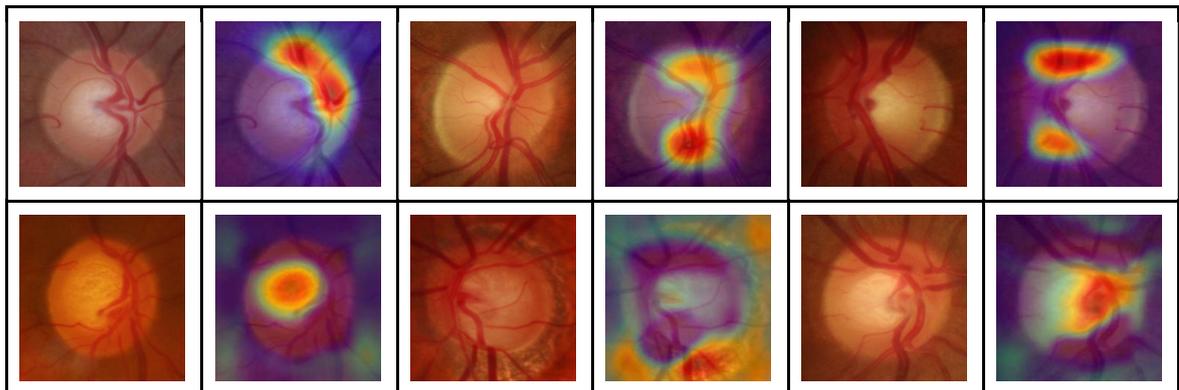


Figura 24. Imágenes de ojos sanos (primera fila) y ojos con glaucoma (segunda fila). Se muestra la imagen original y el mapa Grad-CAM obtenido con la red VGG16.

Histogramas IoU

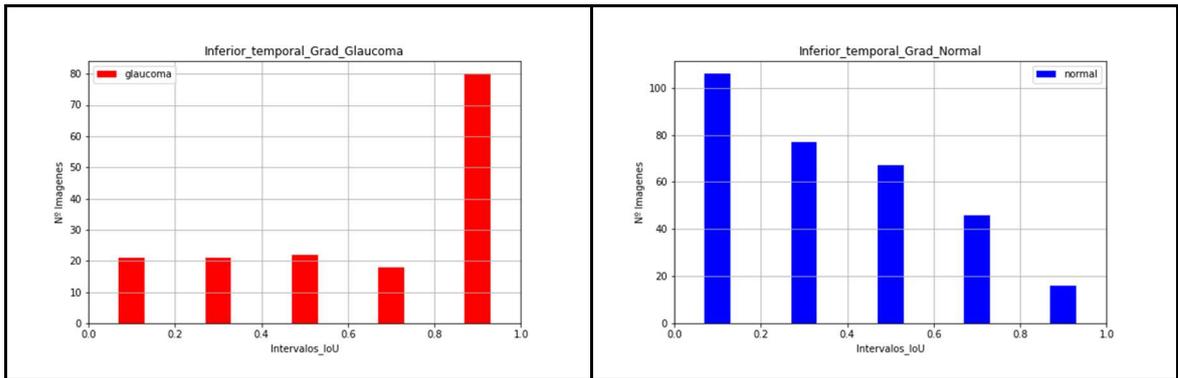


Figura 25. IoU entre mapas Grad-CAM y sector temporal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

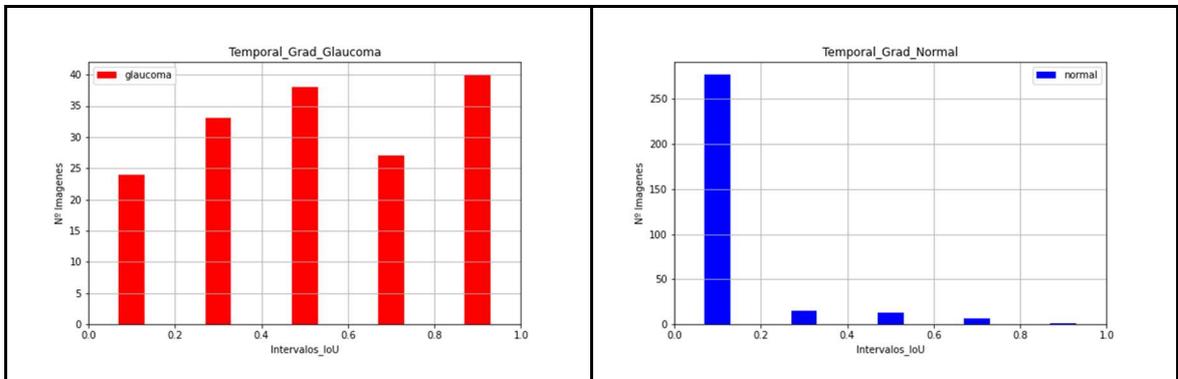


Figura 26. IoU entre mapas Grad-CAM y sector temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

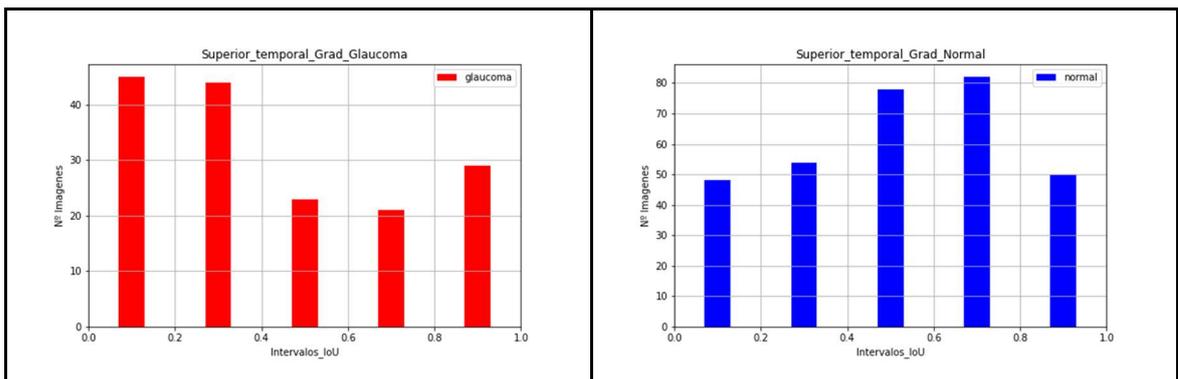


Figura 27. IoU entre mapas Grad-CAM y sector temporal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

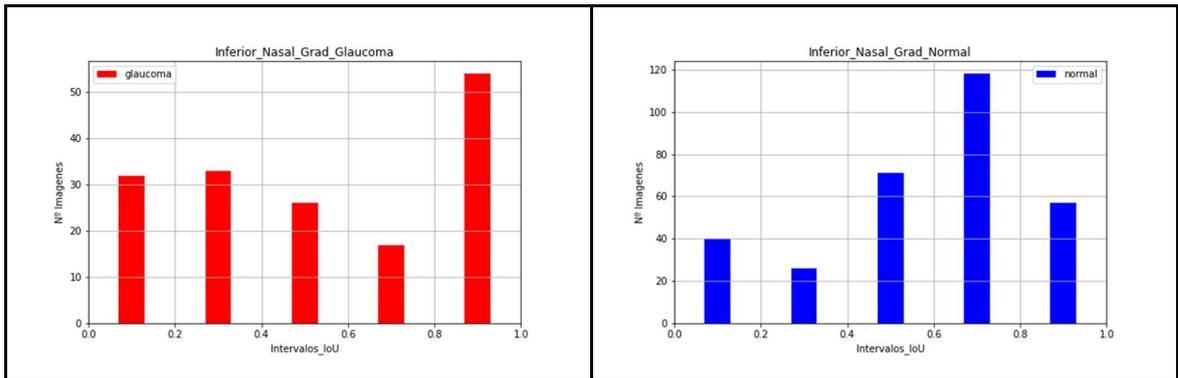


Figura 28. IoU entre mapas Grad-CAM y sector nasal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

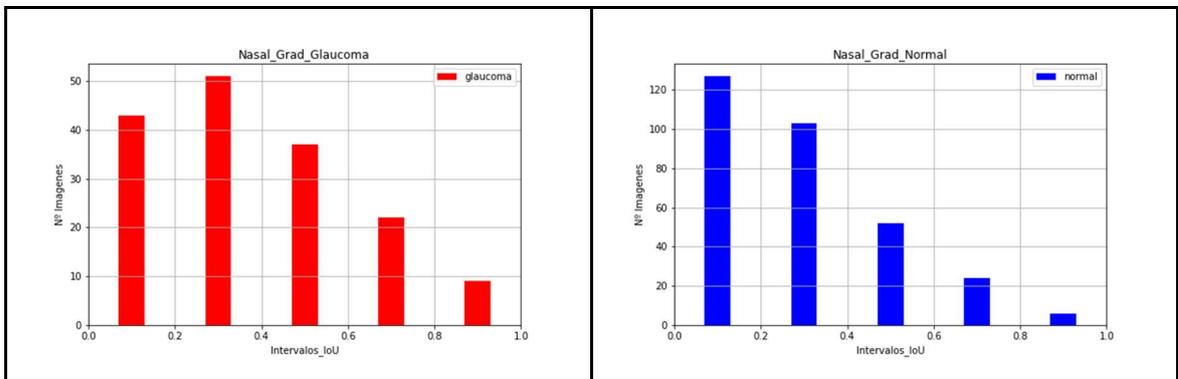


Figura 29. IoU entre mapas Grad-CAM y sector nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

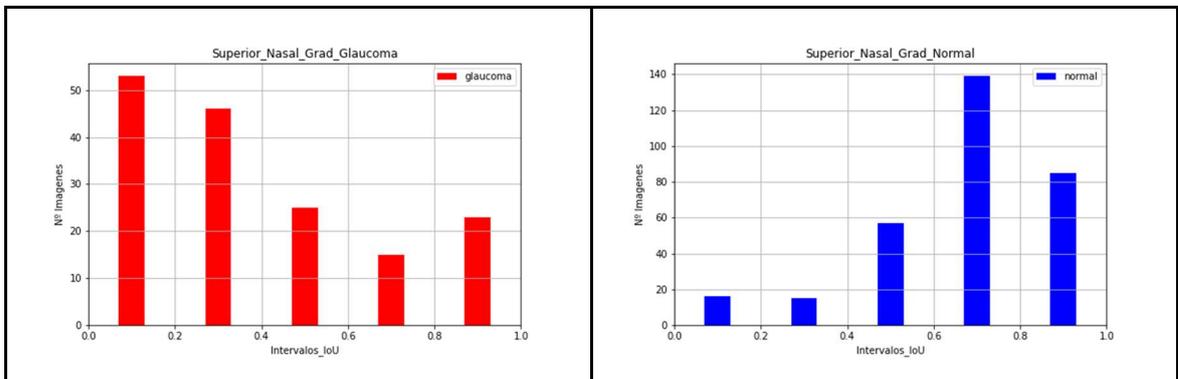


Figura 30. IoU entre mapas Grad-CAM y sector nasal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

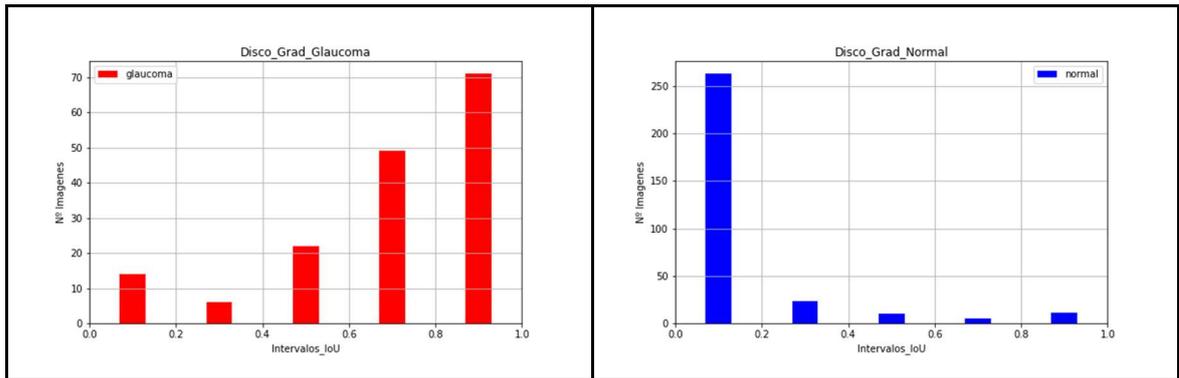


Figura 31. IoU entre mapas Grad-CAM y disco para ojos con glaucoma (izquierda) y ojos normales (derecha).

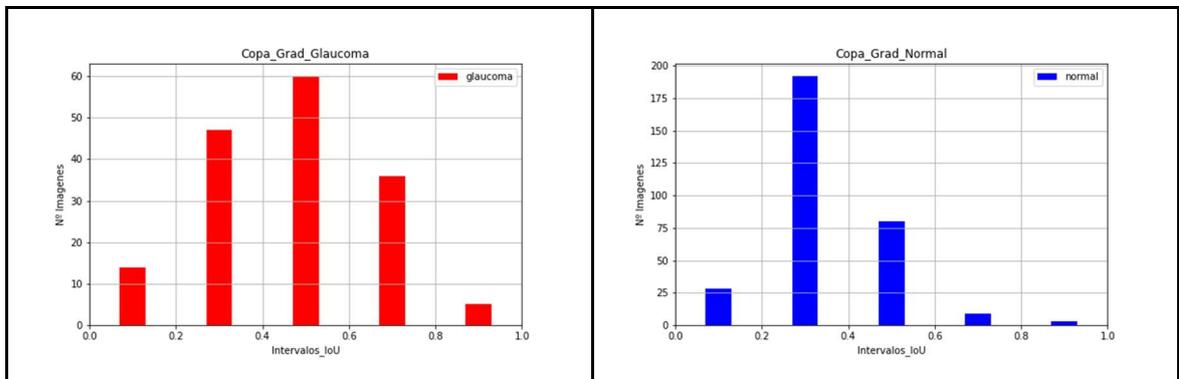


Figura 32. IoU entre mapas Grad-CAM y copa para ojos con glaucoma (izquierda) y ojos normales (derecha).

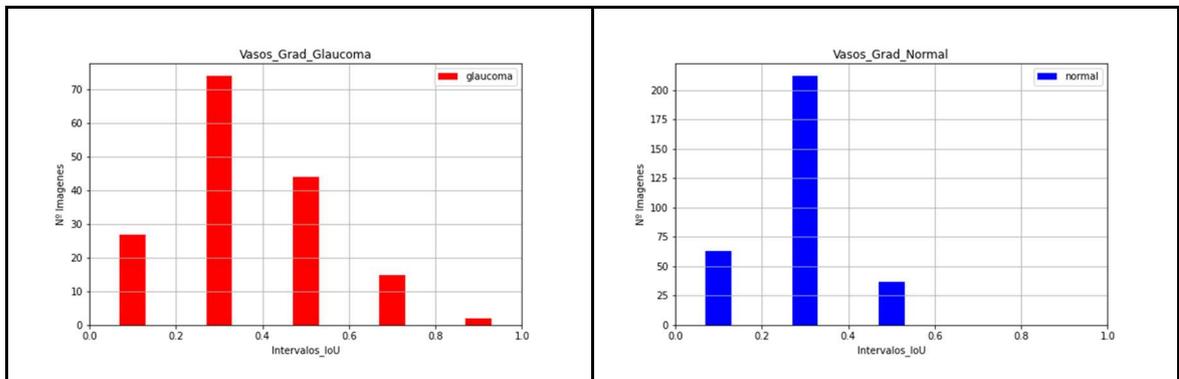


Figura 33. IoU entre mapas Grad-CAM y vasos para ojos con glaucoma (izquierda) y ojos normales (derecha).

Discusión de resultados de Grad-CAM

En el caso de los glaucomas, el IoU es alto para los sectores temporales, especialmente para el sector temporal inferior, y para el sector nasal inferior. Todo esto indica que la zona inferior de la imagen es relevante en caso de enfermedad. En cambio, esta zona no parece importante en caso de ojos normales.

Analizando las regiones del ojo, resulta relevante cómo el disco parece muy influyente en el caso de glaucoma, mientras que apenas es relevante en



imágenes normales. La importancia de la copa y los vasos es mucho menor para ambos tipos de imágenes.

5.2.1.2 Resultados de Score-CAM

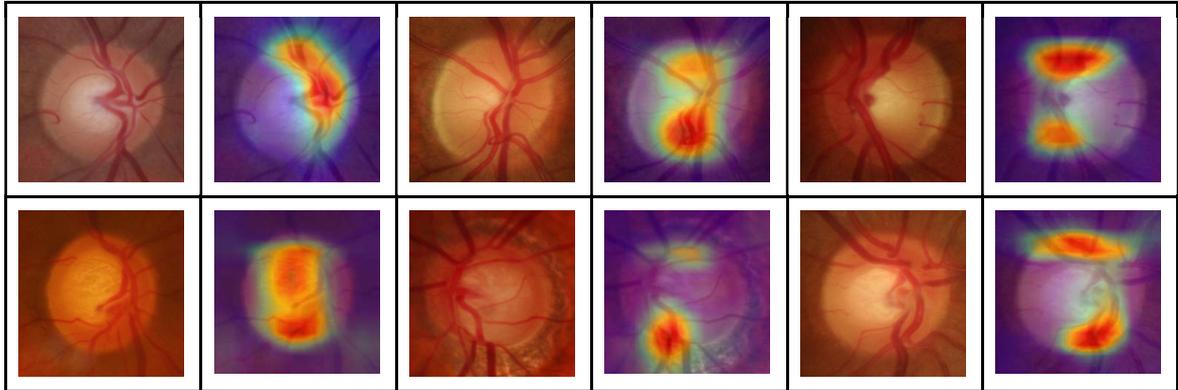


Figura 34. Imágenes de ojos sanos (primera fila) y ojos con glaucoma (segunda fila). Se muestra la imagen original y el mapa Score-CAM obtenido con la red VGG16.

Histogramas IoU

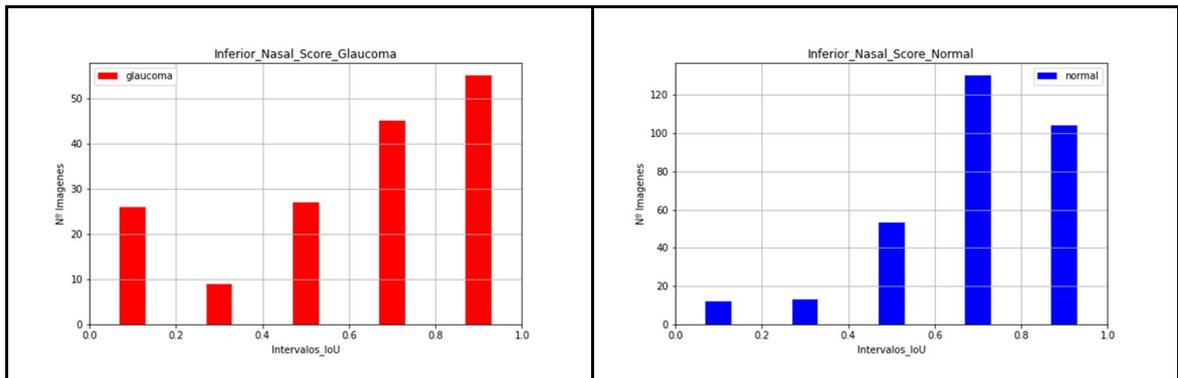


Figura 35. IoU entre mapas Score-CAM y sector nasal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

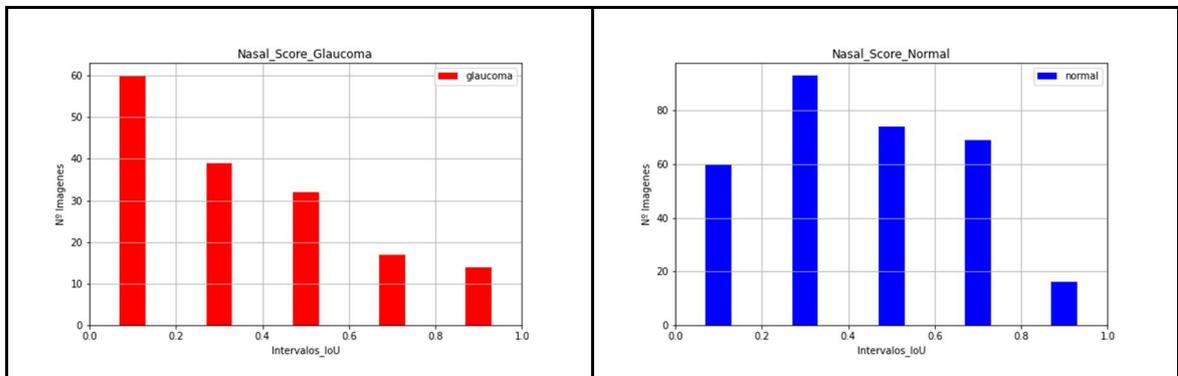


Figura 36. IoU entre mapas Score-CAM y sector nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

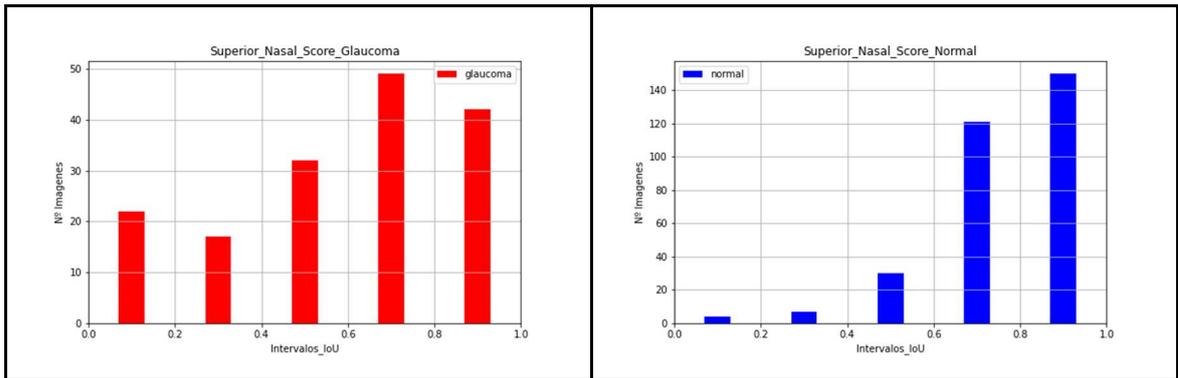


Figura 37. IoU entre mapas Score-CAM y sector nasal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

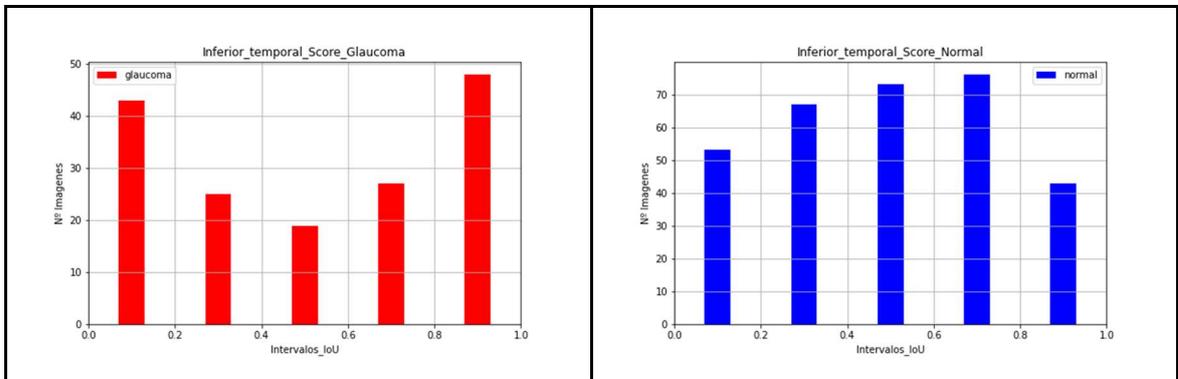


Figura 38. IoU entre mapas Score-CAM y sector temporal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

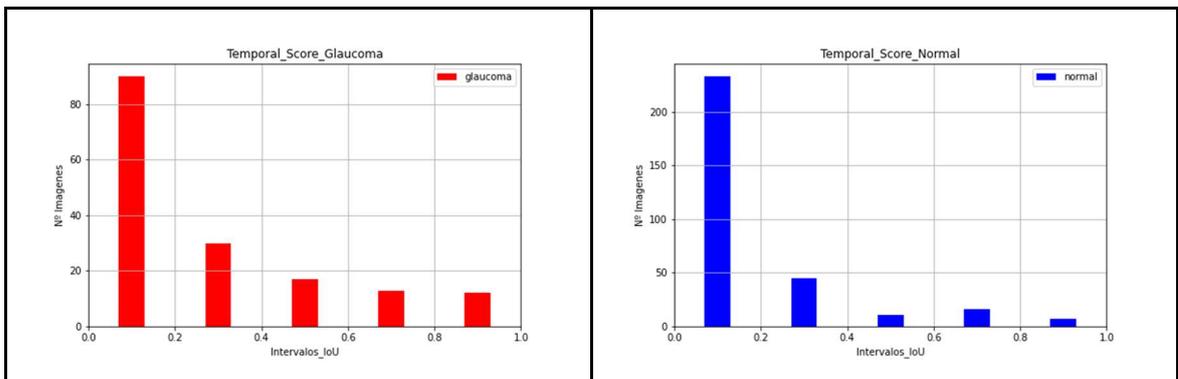


Figura 39. IoU entre mapas Score-CAM y sector temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

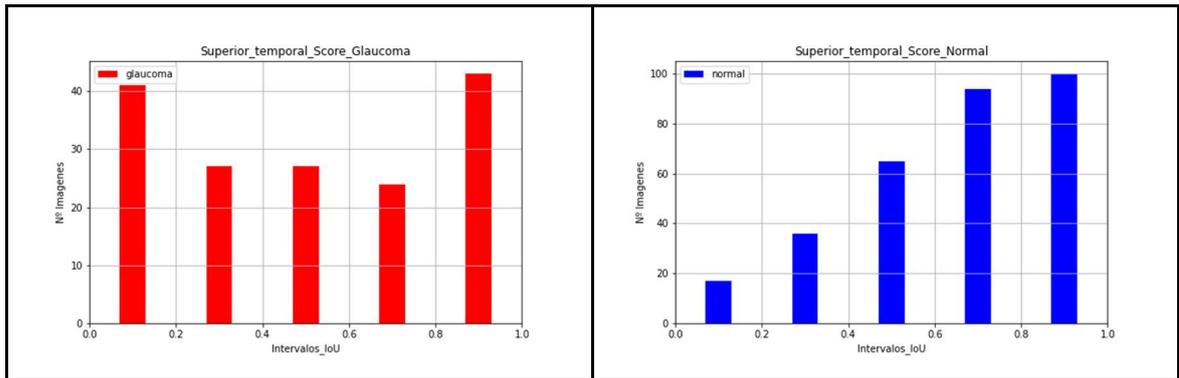


Figura 40. IoU entre mapas Score-CAM y sector temporal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

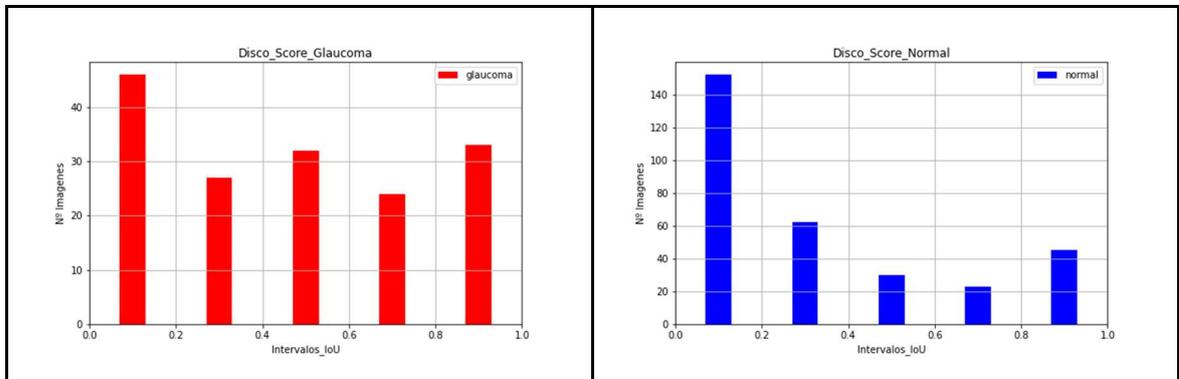


Figura 41. IoU entre mapas Score-CAM y disco para ojos con glaucoma (izquierda) y ojos normales (derecha).

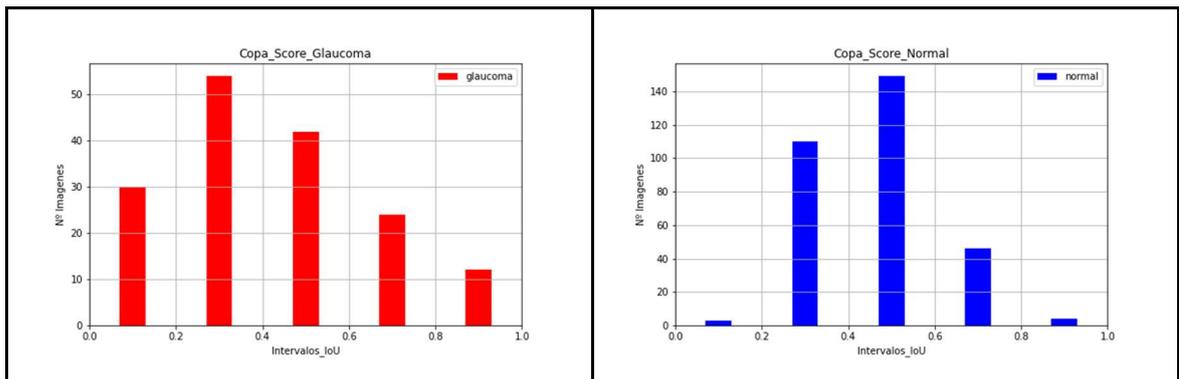


Figura 42. IoU entre mapas Score-CAM y copa para ojos con glaucoma (izquierda) y ojos normales (derecha).

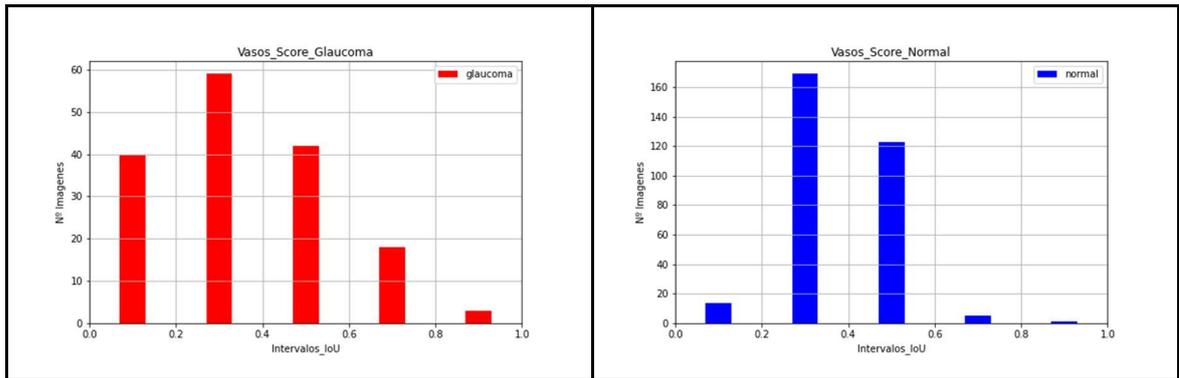


Figura 43. IoU entre mapas Score-CAM y vasos para ojos con glaucoma (izquierda) y ojos normales (derecha).

Discusión de resultados de Score-CAM

Los sectores nasales y temporales no parecen importantes para ningún tipo de imagen. En cambio, los nasales inferior y superior y los temporales inferior y superior son mucho más influyentes, tanto para ojos con glaucoma como para ojos normales. Por lo tanto, según este método la VGG16 se fija bastante en las partes alta y baja de la imagen para diagnosticar.

Analizando las regiones del ojo, el Score-CAM coincide con el Grad-CAM: el disco parece influyente en el caso de glaucoma, y apenas relevante en imágenes normales. La importancia de la copa y los vasos es, de nuevo, mucho menor para ambos tipos de imágenes

5.2.1.3 Resultados del método de oclusión (OCL)

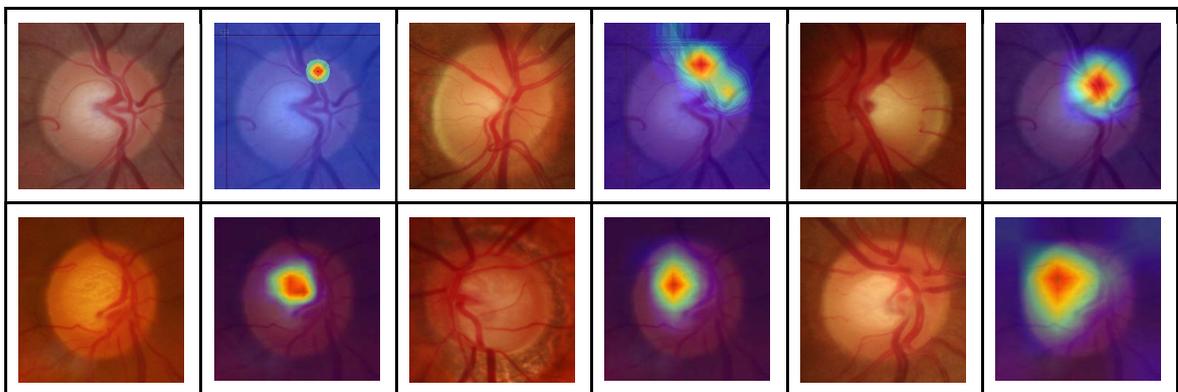


Figura 44. Imágenes de ojos sanos (primera fila) y ojos con glaucoma (segunda fila). Se muestra la imagen original y el mapa OCL obtenido con la red VGG16.



Histogramas IoU

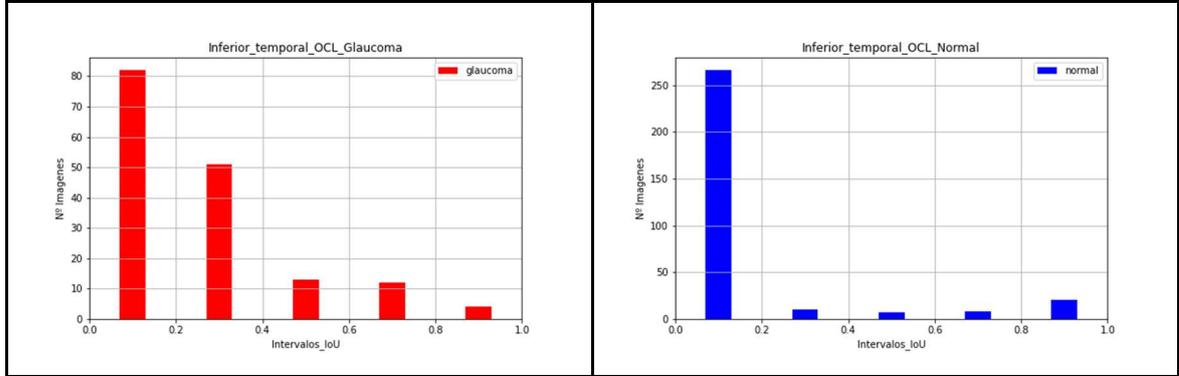


Figura 45. IoU entre mapas OCL y sector inferior temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

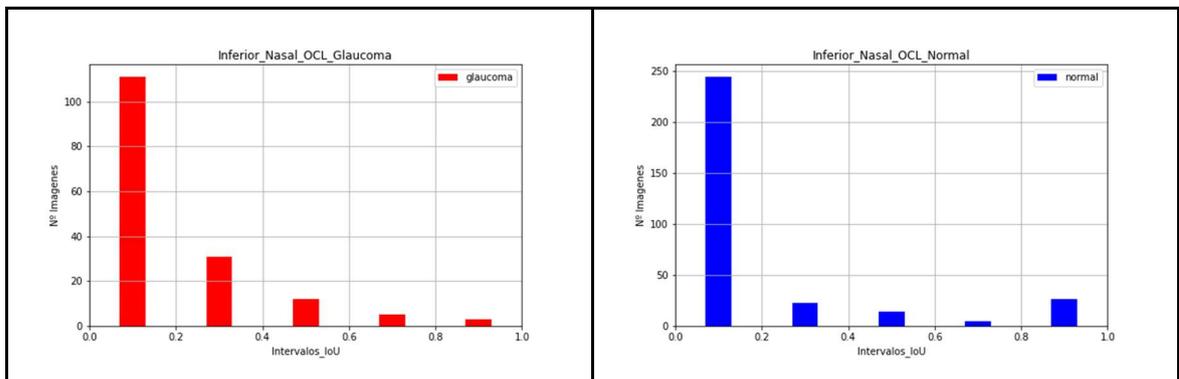


Figura 46. IoU entre mapas OCL y sector inferior nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

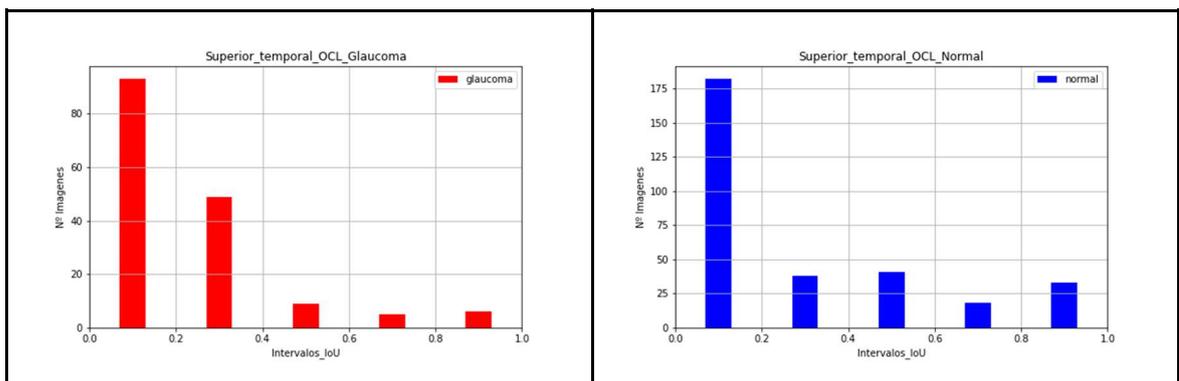


Figura 47. IoU entre mapas OCL y sector superior temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

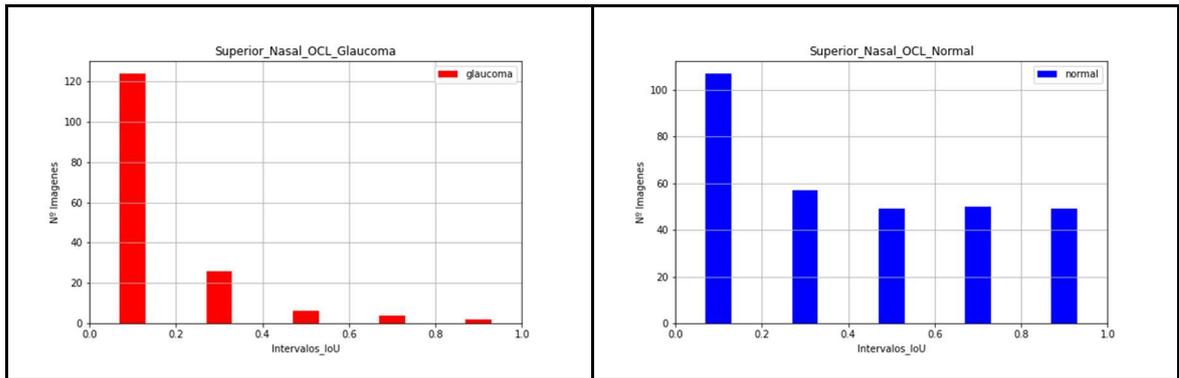


Figura 48. IoU entre mapas OCL y sector superior nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

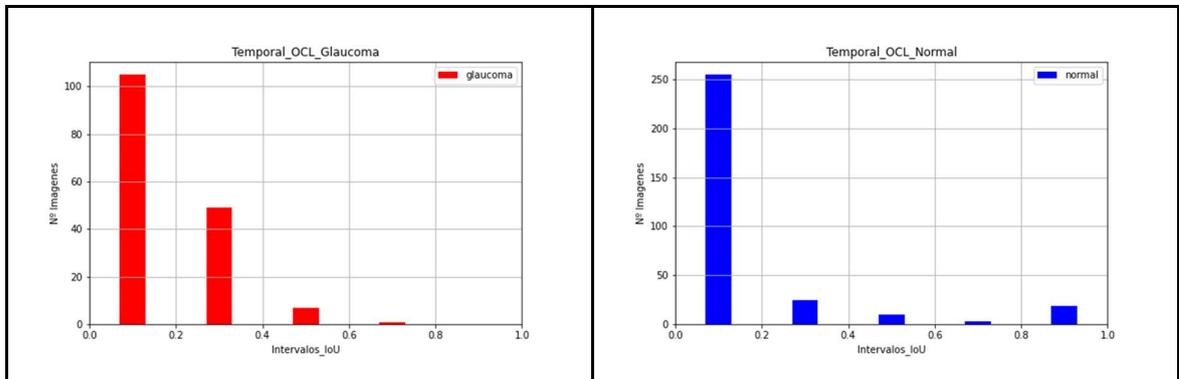


Figura 49. IoU entre mapas OCL y sector temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

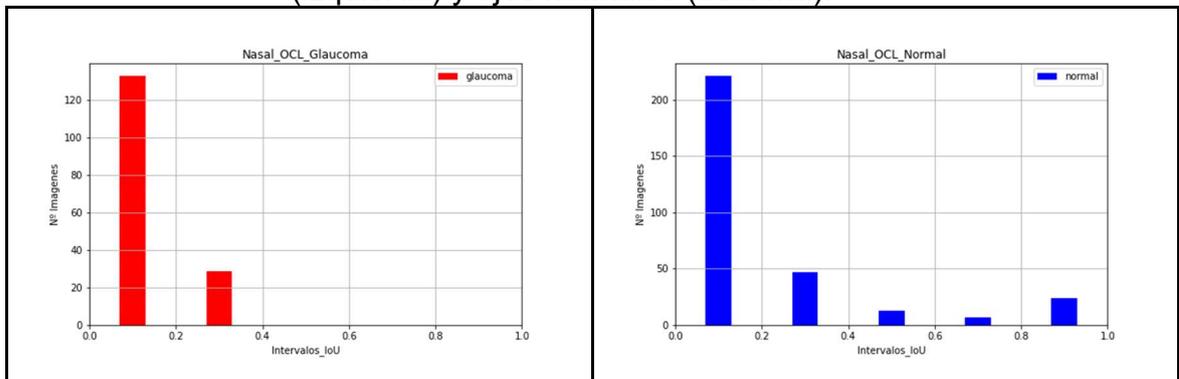


Figura 50. IoU entre mapas OCL y sector nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

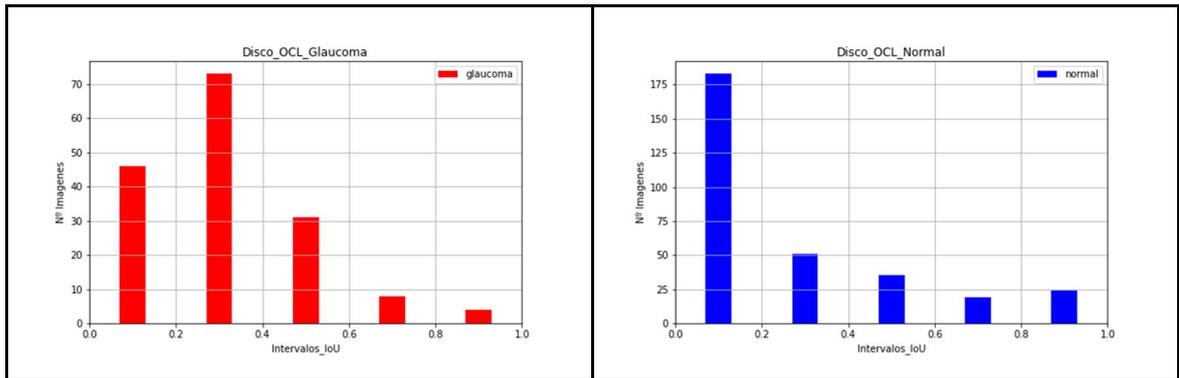


Figura 51. IoU entre mapas OCL y disco para ojos con glaucoma (izquierda) y ojos normales (derecha).

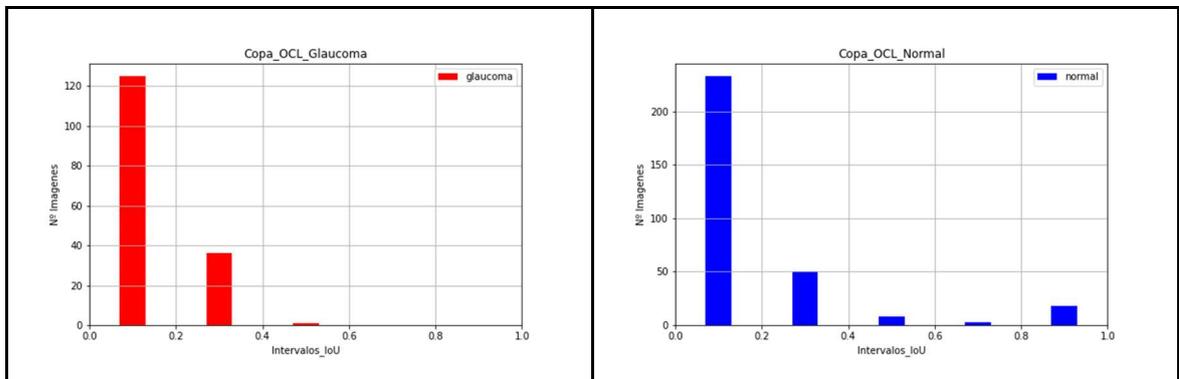


Figura 52. IoU entre mapas OCL y copa para ojos con glaucoma (izquierda) y ojos normales (derecha).

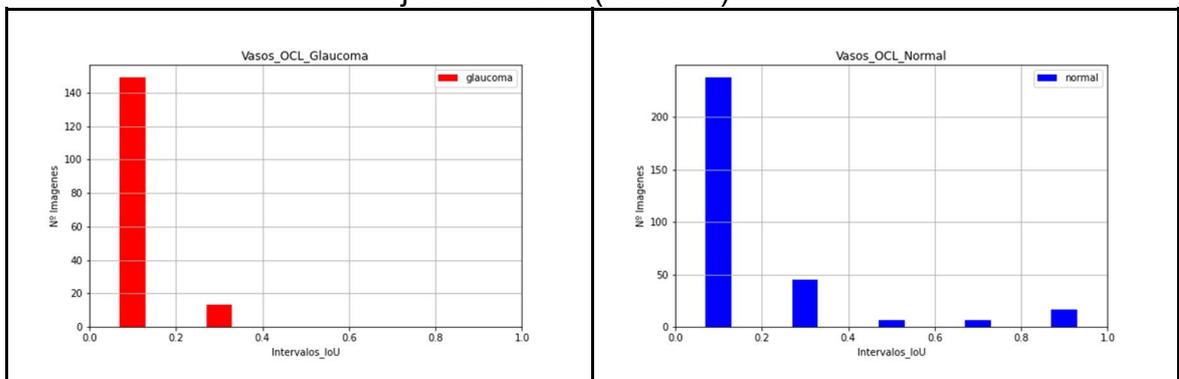


Figura 53. IoU entre mapas OCL y vasos para ojos con glaucoma (izquierda) y ojos normales (derecha).

Discusión de resultados del método de oclusión

A la hora de analizar estos resultados, es importante tener en cuenta que los mapas de calor que genera el método OCL son mucho más pequeños que los de los otros métodos. El tamaño de los mapas en este caso es altamente dependiente del tamaño de la ventana considerada. Al ser tan pequeños, los índices IoU disminuyen notablemente. De hecho, en todas las regiones el mayor porcentaje de IoU está por debajo del 0.20. Por ese motivo se pondrá énfasis en las regiones que presenten una distribución más repartida, entendiendo que son más importantes.



Para las imágenes normales, parece que el método indica que la zona superior es más relevante que la inferior. Para los glaucomas esta influencia es mucho más pequeña. Lo más representativo de este método sobre el resto de las regiones es la mayor influencia que tiene el disco en caso de glaucoma.

5.2.1.4 Comparativa de resultados de los tres métodos con la VGG16

Después de haber estudiado los resultados de todas las técnicas sobre la VGG16 se puede observar que los tres métodos coinciden en lo siguiente:

- En los sectores, los más importantes son los que están situados en la parte superior e inferior, sobre todo los inferiores para el caso de Glaucoma.
- Con respecto a las restantes regiones del ojo, la más determinante en los tres métodos para los dos tipos de imágenes es claramente el disco óptico.
- Los sectores nasales y temporales, y la copa y los vasos no parecen relevantes en la decisión de la VGG16.

5.2.2 VGG19

5.2.2.1 Resultados de Grad-CAM

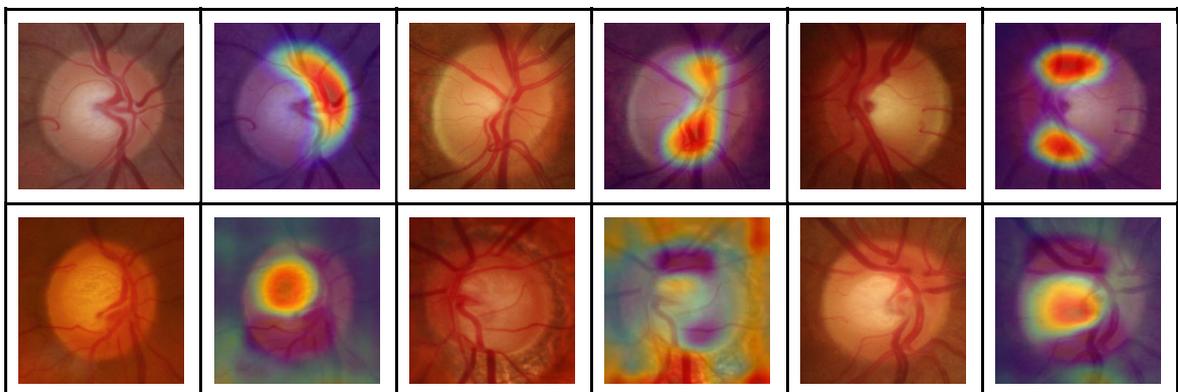


Figura 54. Imágenes de ojos sanos (primera fila) y ojos con glaucoma (segunda fila). Se muestra la imagen original y el mapa Grad-CAM obtenido con la red VGG19.

Histogramas IoU

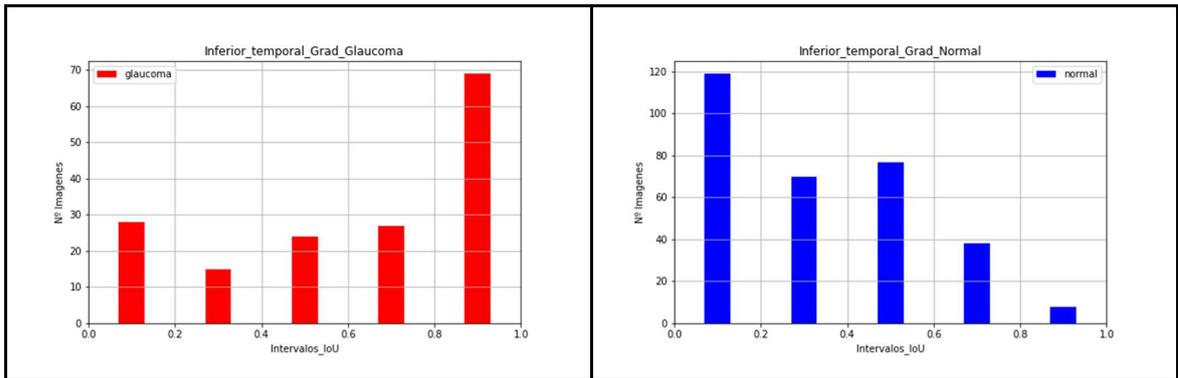


Figura 55. IoU entre mapas Grad-CAM y sector temporal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

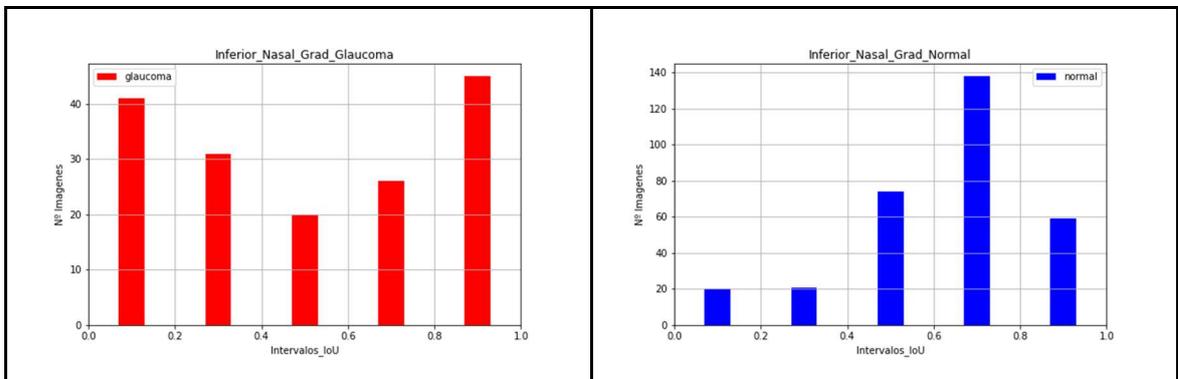


Figura 56. IoU entre mapas Grad-CAM y sector nasal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

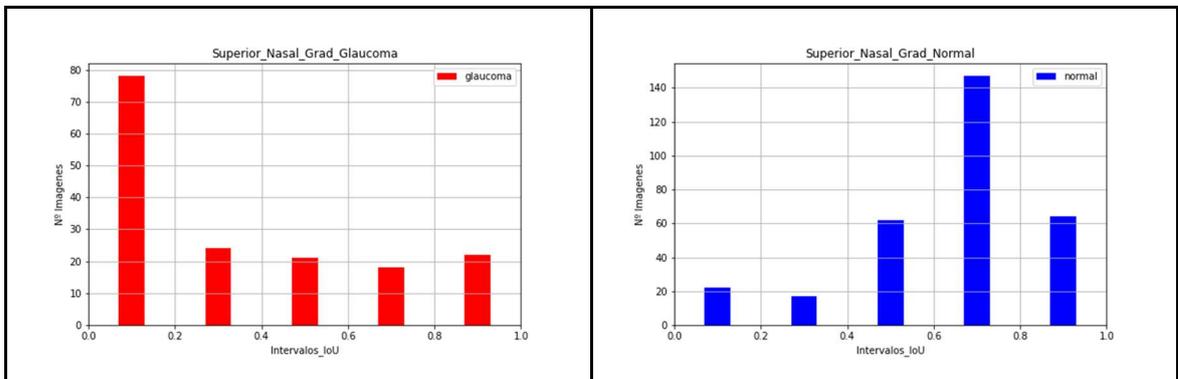


Figura 57. IoU entre mapas Grad-CAM y sector nasal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

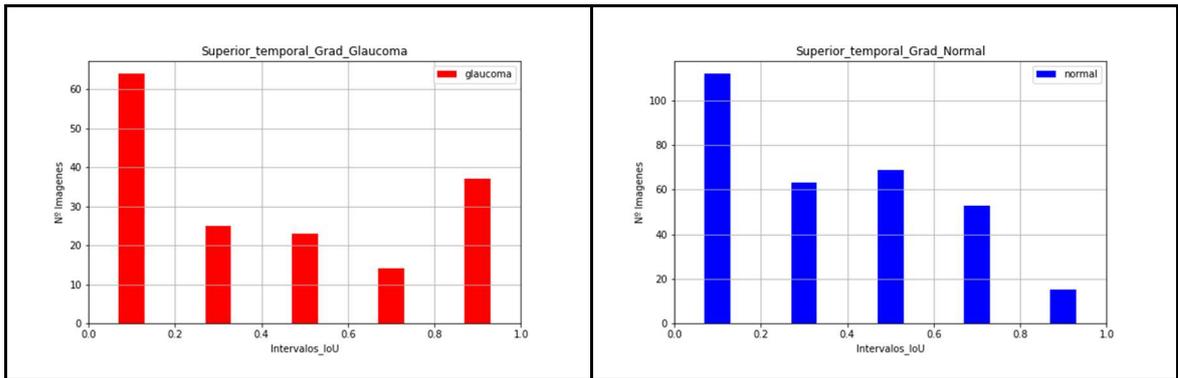


Figura 58. IoU entre mapas Grad-CAM y sector temporal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

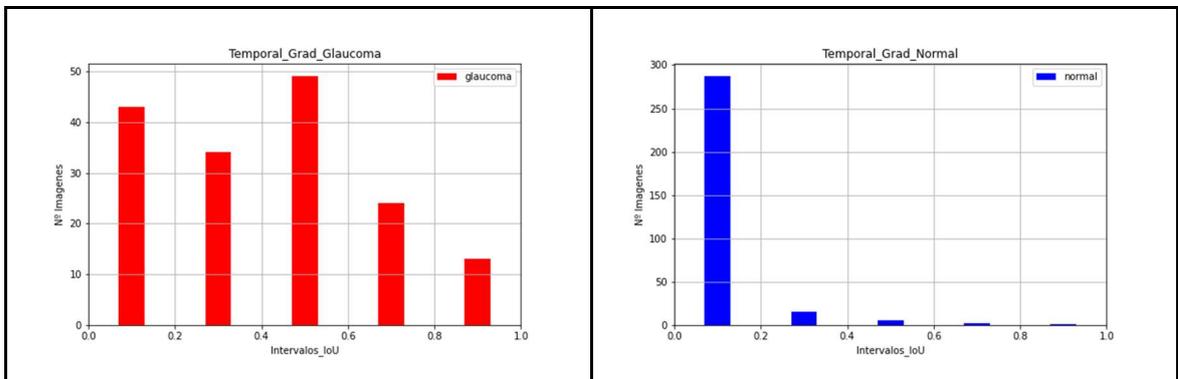


Figura 59. IoU entre mapas Grad-CAM y sector temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

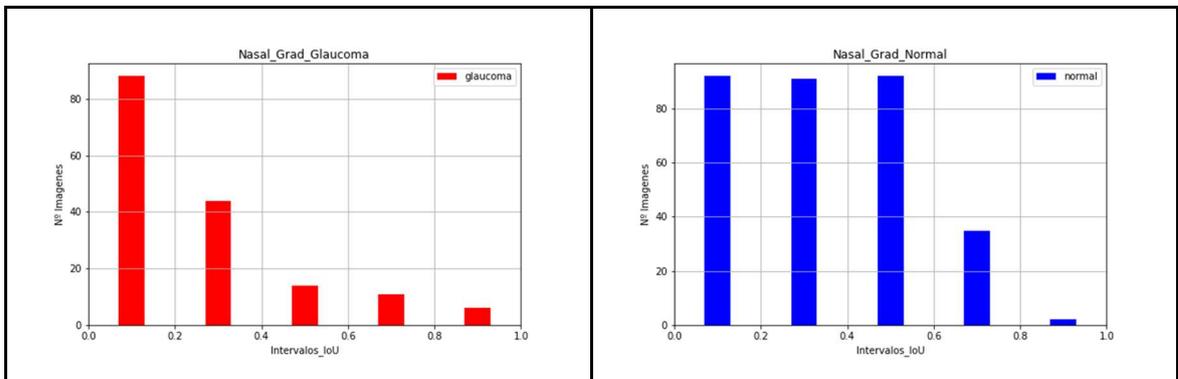


Figura 60. IoU entre mapas Grad-CAM y sector nasal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

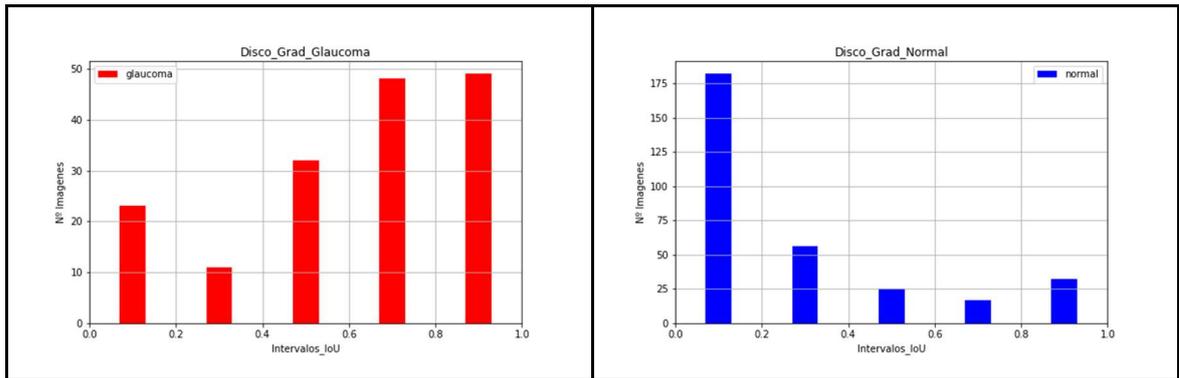


Figura 61. IoU entre mapas Grad-CAM y disco para ojos con glaucoma (izquierda) y ojos normales (derecha).

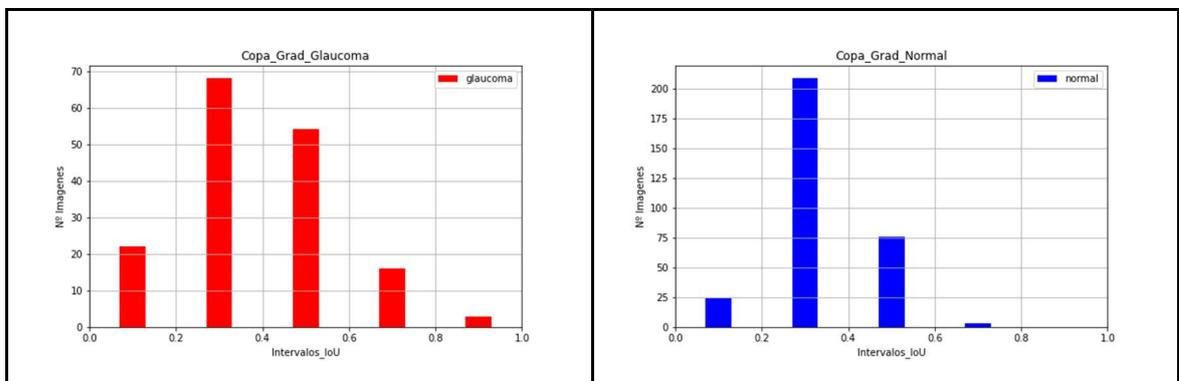


Figura 62. IoU entre mapas Grad-CAM y copa para ojos con glaucoma (izquierda) y ojos normales (derecha).

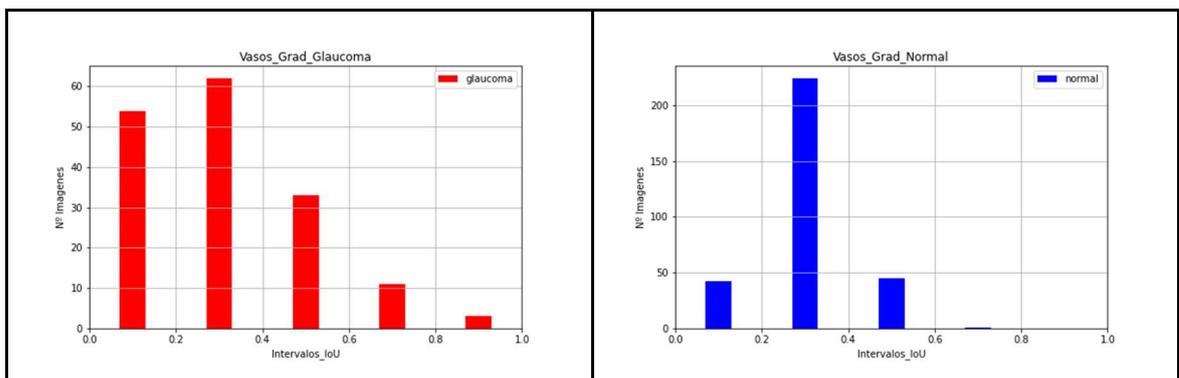


Figura 63. IoU entre mapas Grad-CAM y vasos para ojos con glaucoma (izquierda) y ojos normales (derecha).

Discusión de resultados de Grad-CAM

Los sectores inferiores (tanto nasal como temporal) son muy importantes para el diagnóstico del Glaucoma. En el caso de ojos normales esta zona es mucho menos relevante. Los sectores nasales y temporales parecen irrelevantes, tanto en ojos sanos como en ojos con glaucoma.



De las restantes regiones de la imagen, la única que tiene relevancia, y solo en el caso de Glaucoma, es el disco.

5.2.2.2 Resultados de Score-CAM

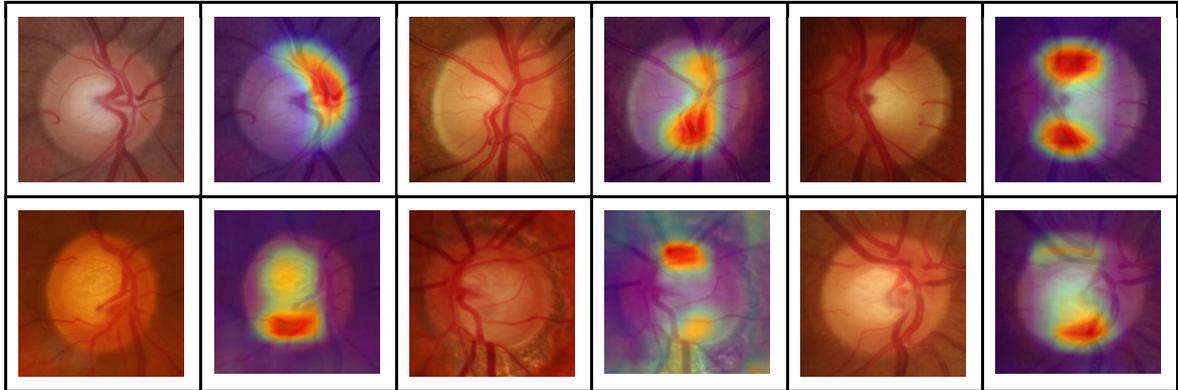


Figura 64. Imágenes de ojos sanos (primera fila) y ojos con glaucoma (segunda fila). Se muestra la imagen original y el mapa Grad-CAM obtenido con la red VGG19.

Histogramas IoU

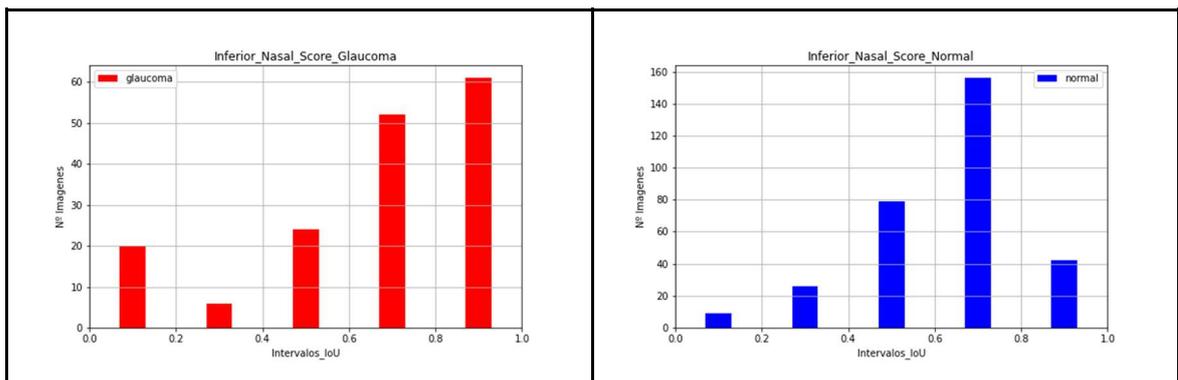


Figura 65. IoU entre mapas Score-CAM y sector nasal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

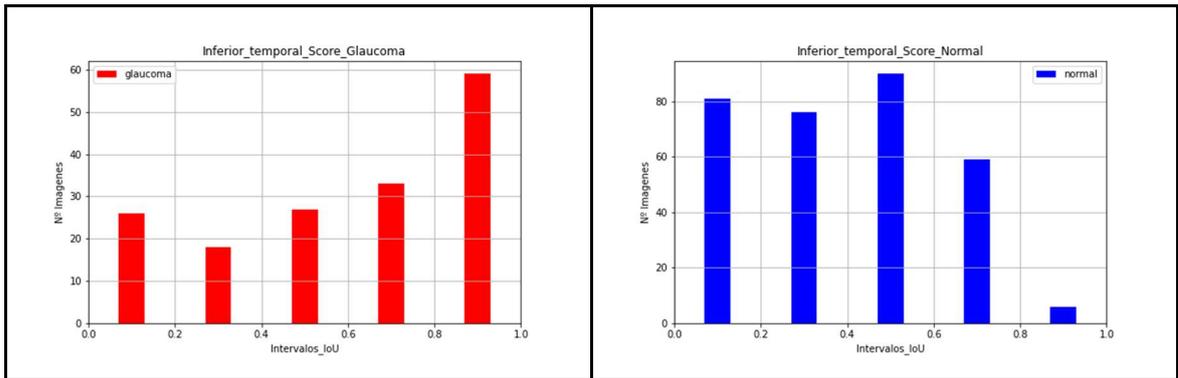


Figura 66. IoU entre mapas Score-CAM y sector temporal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

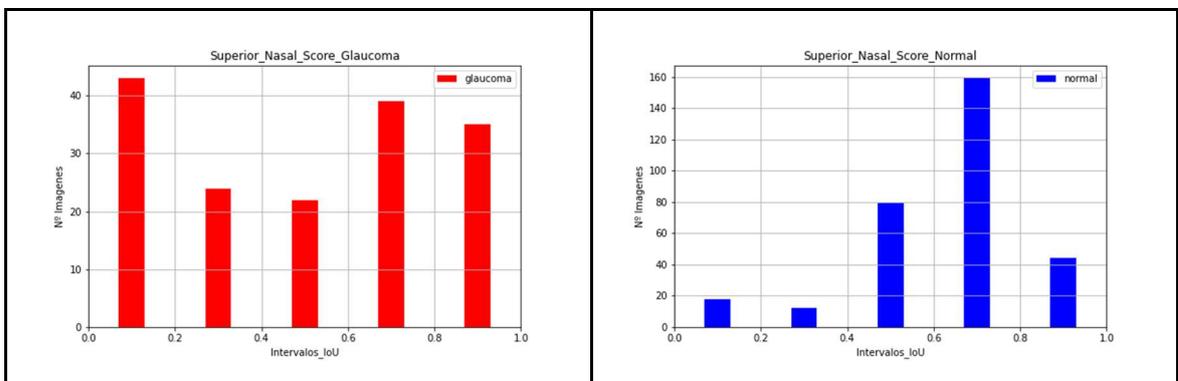


Figura 67. IoU entre mapas Score-CAM y sector nasal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

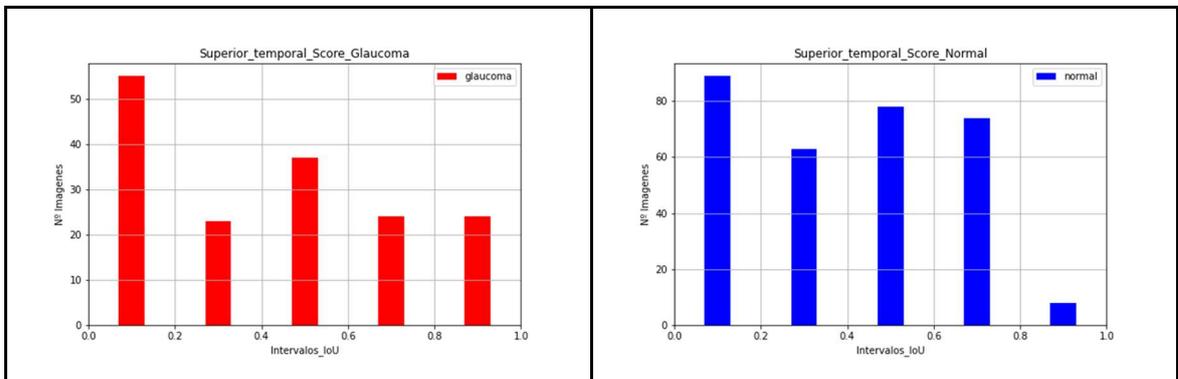


Figura 68. IoU entre mapas Score-CAM y sector temporal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

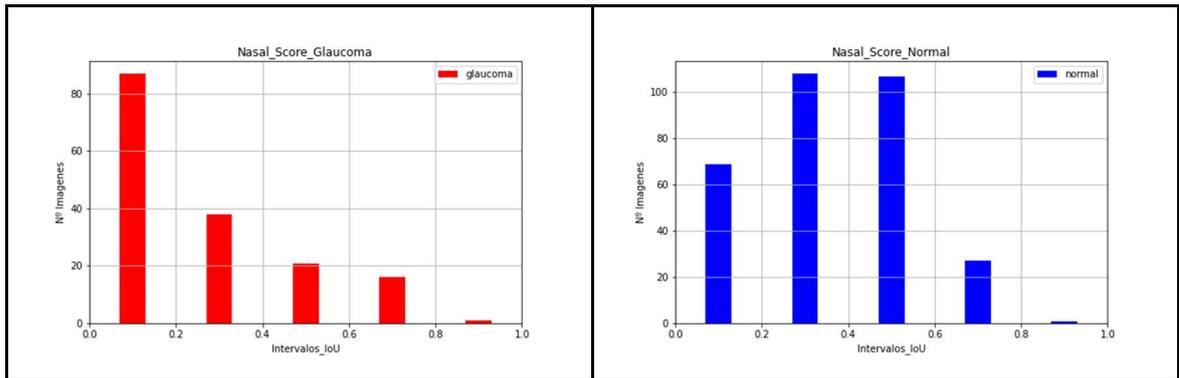


Figura 69. IoU entre mapas Score-CAM y sector nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

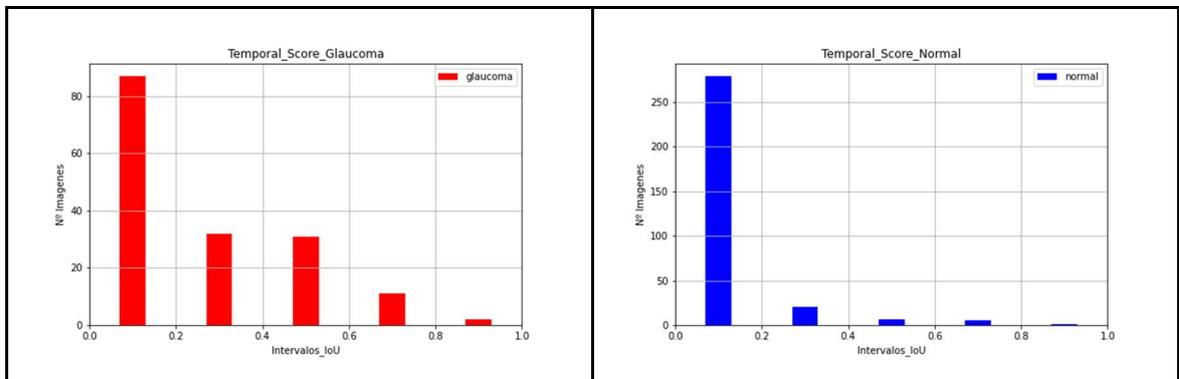


Figura 70. IoU entre mapas Score-CAM y sector temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

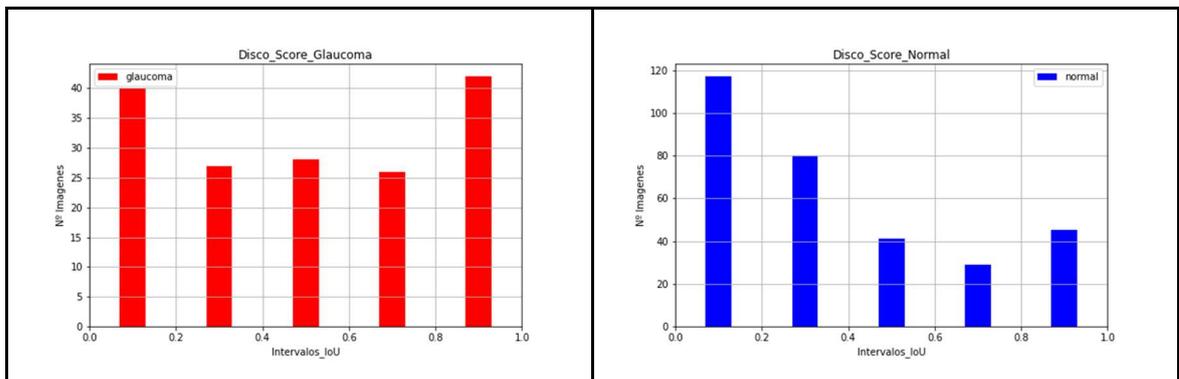


Figura 71. IoU entre mapas Score-CAM y disco para ojos con glaucoma (izquierda) y ojos normales (derecha).

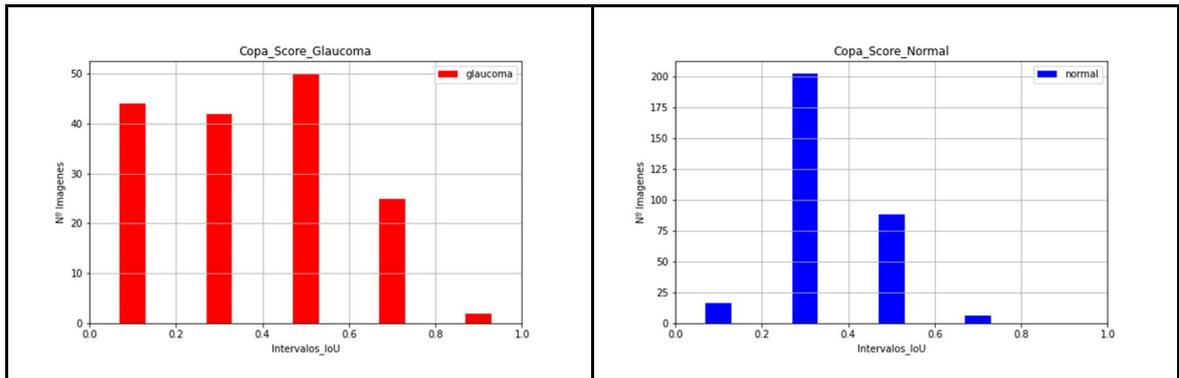


Figura 72. IoU entre mapas Score-CAM y copa para ojos con glaucoma (izquierda) y ojos normales (derecha).

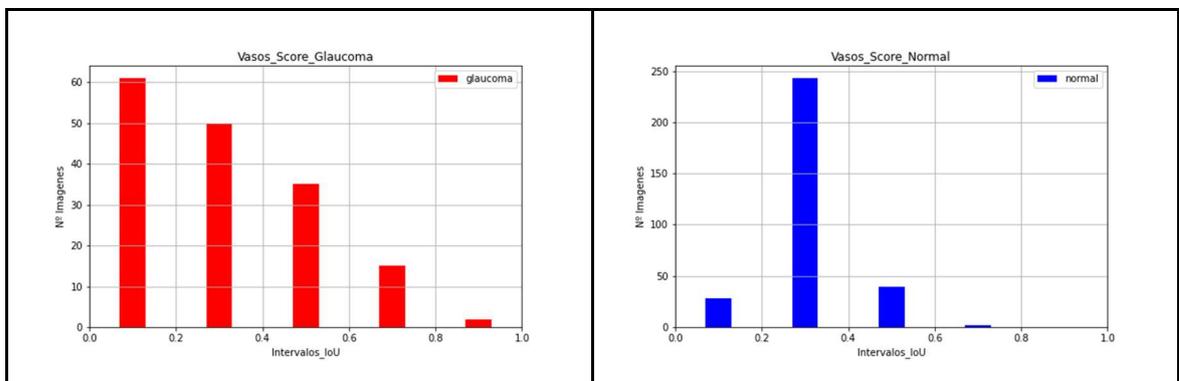


Figura 73. IoU entre mapas Score-CAM y vasos para ojos con glaucoma (izquierda) y ojos normales (derecha).

Discusión de resultados de Score-CAM

El Score-CAM coincide con lo reflejado por el Grad-CAM: los sectores inferiores (tanto nasal como temporal) son muy importantes para el diagnóstico del Glaucoma. En el caso de ojos normales esta zona es mucho menos relevante. Los sectores nasales y temporales parecen irrelevantes, tanto en ojos sanos como en ojos con glaucoma. Y, de nuevo, el disco es muy importante, sobre todo en las imágenes de glaucoma.

5.2.2.3 Resultados de OCL

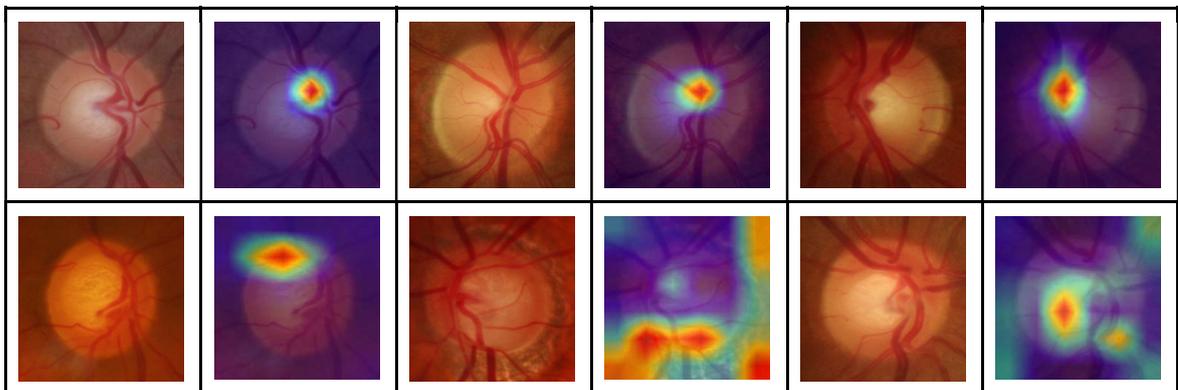




Figura 74. Imágenes de ojos sanos (primera fila) y ojos con glaucoma (segunda fila). Se muestra la imagen original y el mapa OCL obtenido con la red VGG19.

Histogramas IoU

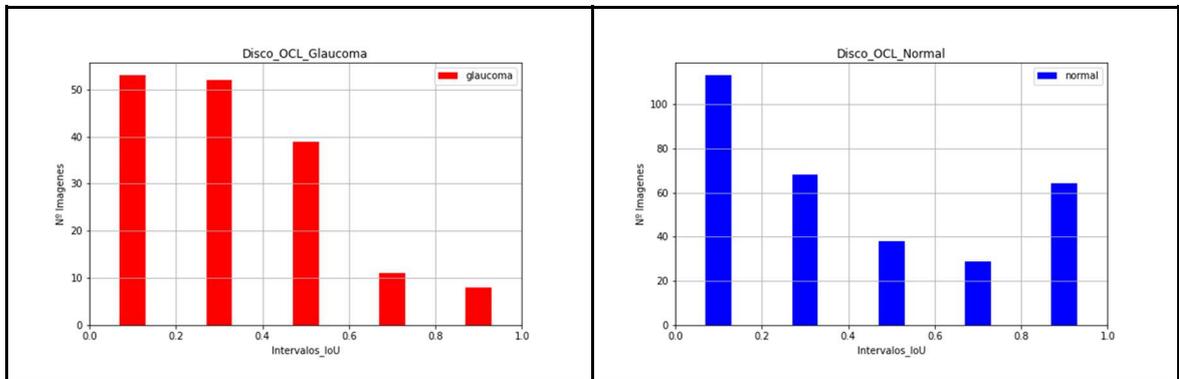


Figura 75. IoU entre mapas OCL y disco para ojos con glaucoma (izquierda) y ojos normales (derecha).

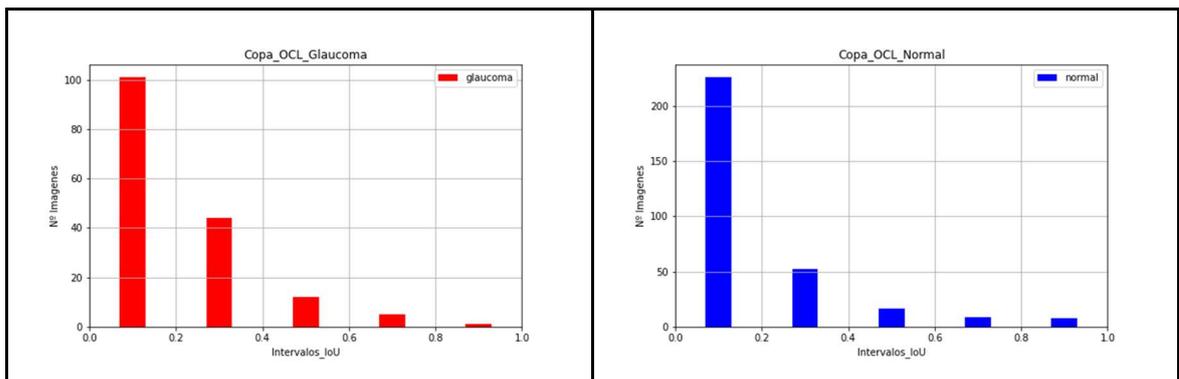


Figura 76. IoU entre mapas OCL y copa para ojos con glaucoma (izquierda) y ojos normales (derecha).

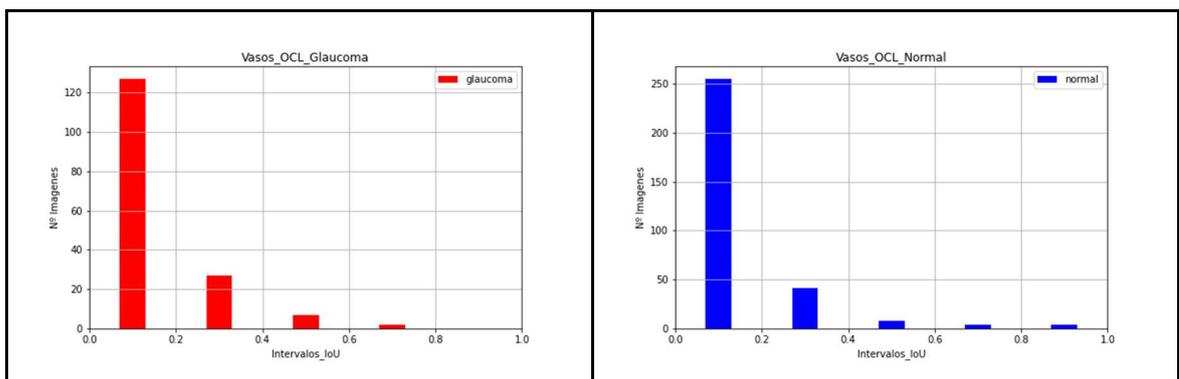


Figura 77. IoU entre mapas OCL y vasos para ojos con glaucoma (izquierda) y ojos normales (derecha).

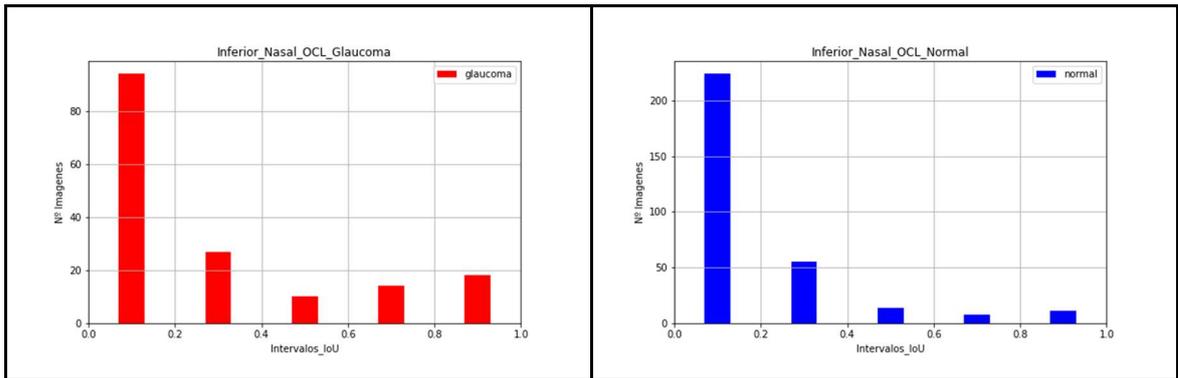


Figura 78. IoU entre mapas OCL y sector inferior nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

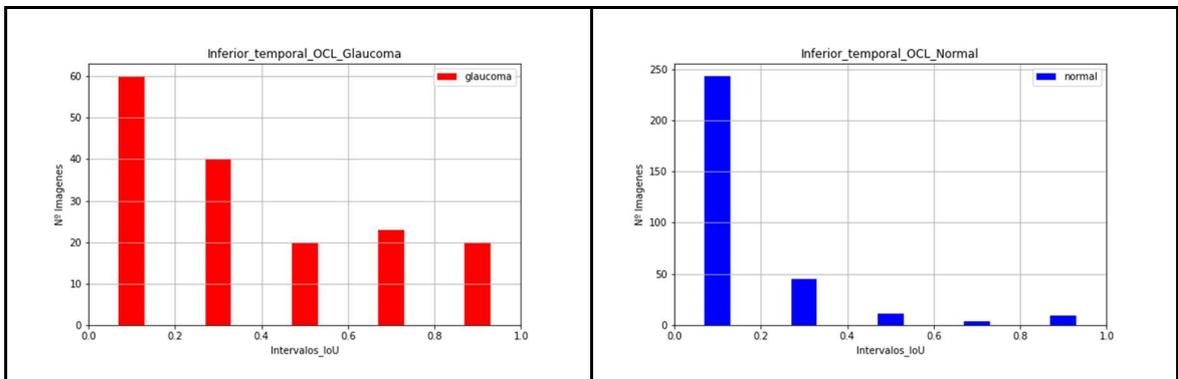


Figura 79. IoU entre mapas OCL y sector inferior temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

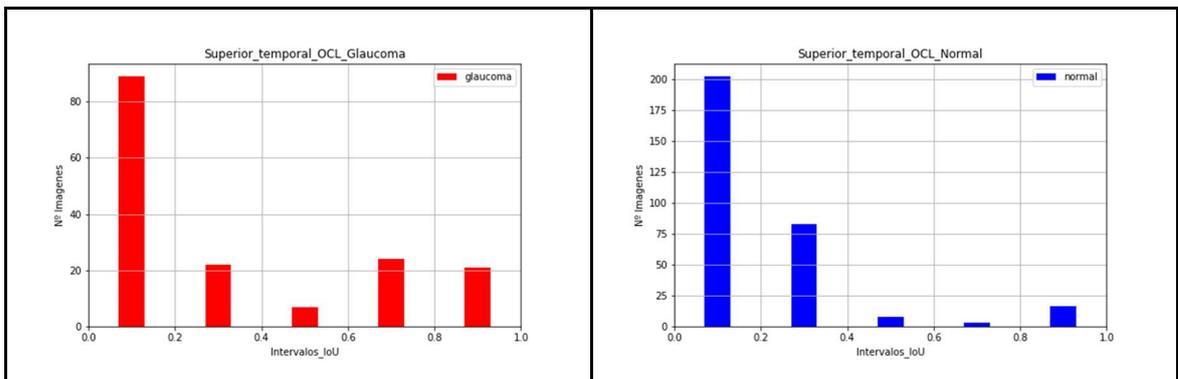


Figura 80. IoU entre mapas OCL y sector superior temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

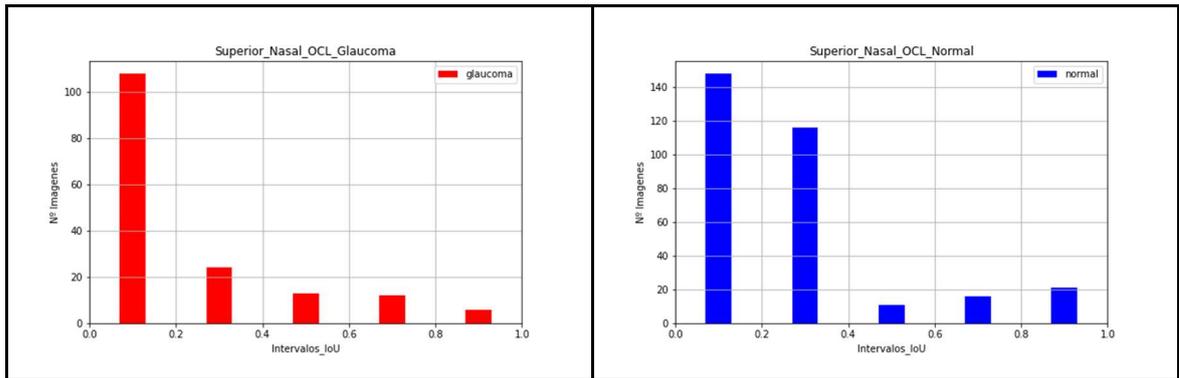


Figura 81. IoU entre mapas OCL y sector superior nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

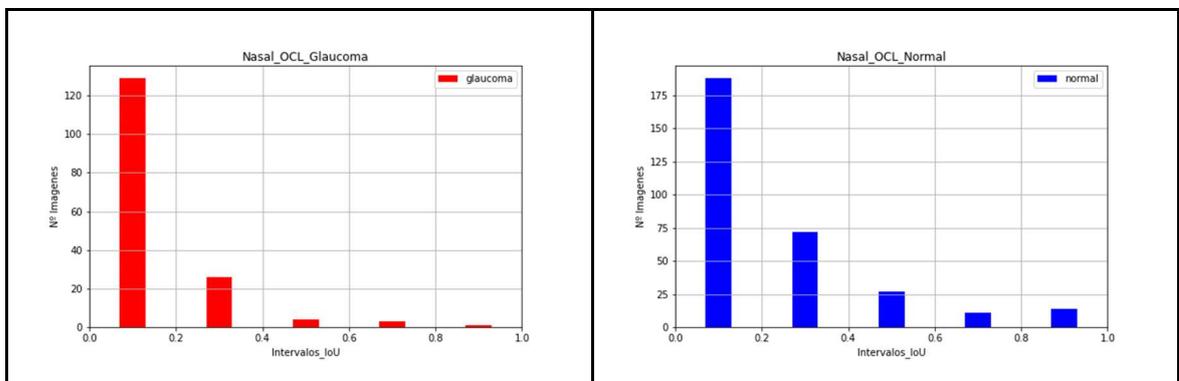


Figura 82. IoU entre mapas OCL y sector nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

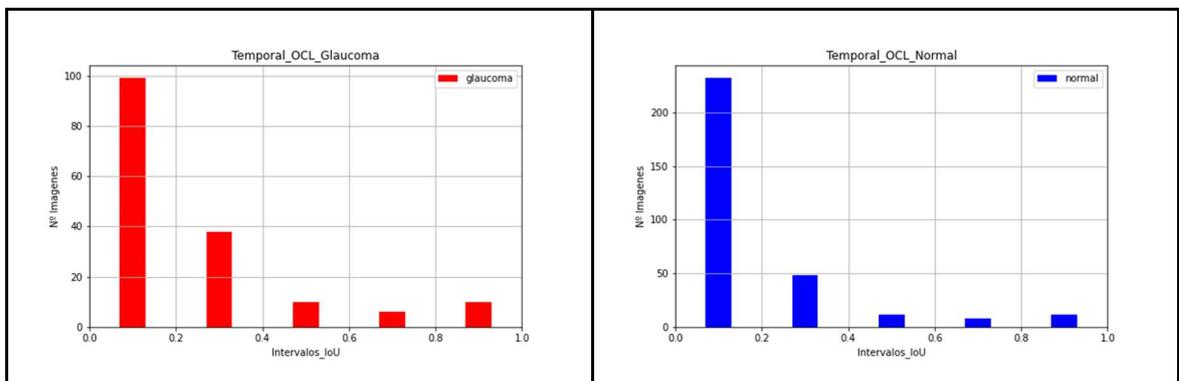


Figura 83. IoU entre mapas OCL y sector temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

Discusión de resultados de OCL

A pesar de que los mapas OCL son más pequeños, y por tanto los índices IoU también, se puede observar la misma tendencia que mostraban los dos métodos anteriores: los sectores nasal y temporal parecen irrelevantes, mientras que las zonas superior e inferior son más importantes. De nuevo, el disco es la región más analizada, sobre todo en imágenes de ojos normales.



5.2.2.4 Comparativa de resultados de los tres métodos con la VGG19

En el caso de VGG19 los tres métodos coinciden bastante entre ellos, y con lo observado en la red VGG16 (sección anterior):

- En los sectores, los más importantes son los que están situados en la parte superior e inferior, sobre todo los inferiores para el caso de Glaucoma.
- Con respecto a las restantes regiones del ojo, la más determinante en los tres métodos para los dos tipos de imágenes es claramente el disco óptico.
- Los sectores nasales y temporales, y la copa y los vasos no parecen relevantes en la decisión de la VGG16.

5.2.3 MobileNet

5.2.3.1 Resultados de Grad-CAM

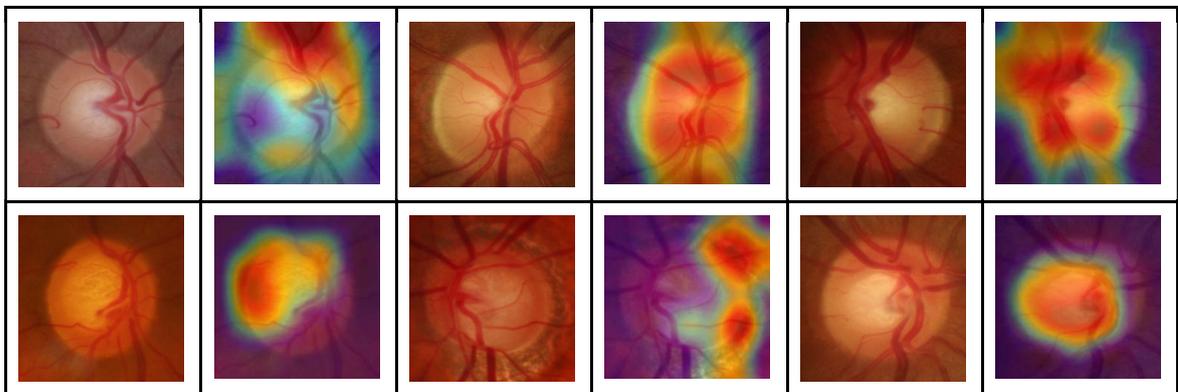


Figura 84. Imágenes de ojos sanos (primera fila) y ojos con glaucoma (segunda fila). Se muestra la imagen original y el mapa Grad-CAM obtenido con la red MobileNet.

Histogramas IoU

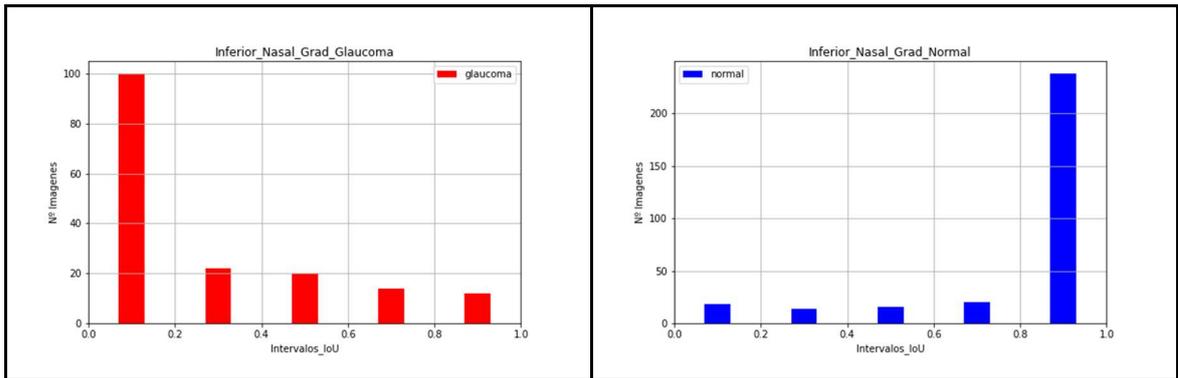


Figura 85. IoU entre mapas Grad-CAM y sector nasal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

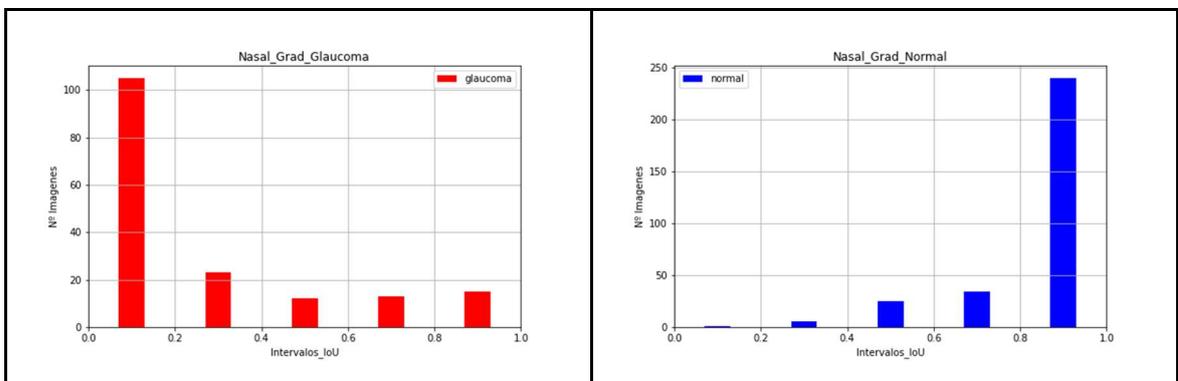


Figura 86. IoU entre mapas Grad-CAM y sector nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

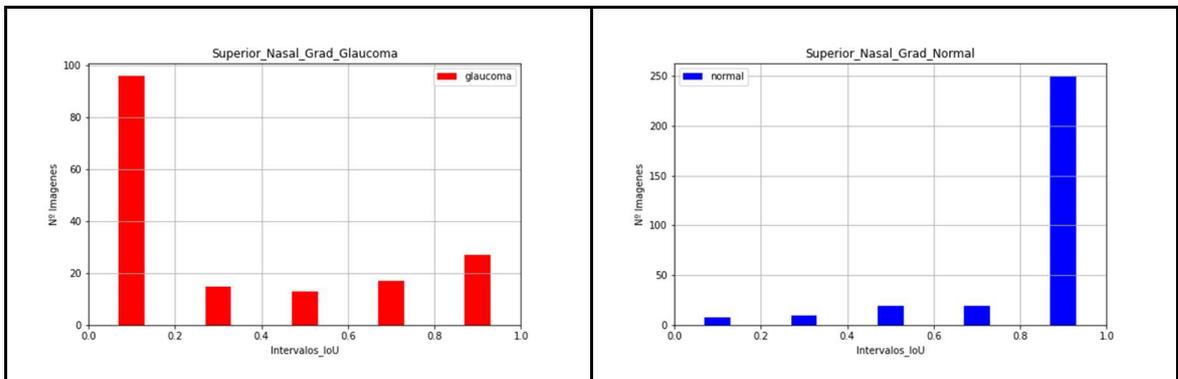


Figura 87. IoU entre mapas Grad-CAM y sector superior nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

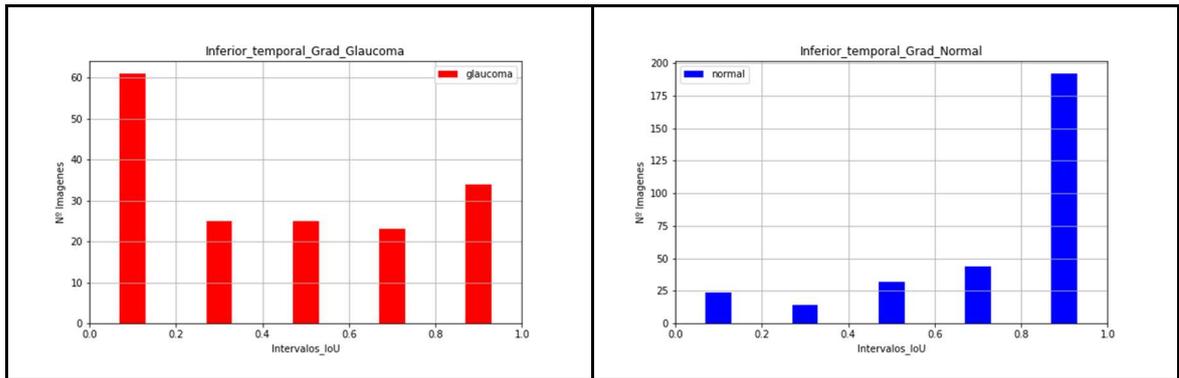


Figura 88. IoU entre mapas Grad-CAM y sector inferior temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

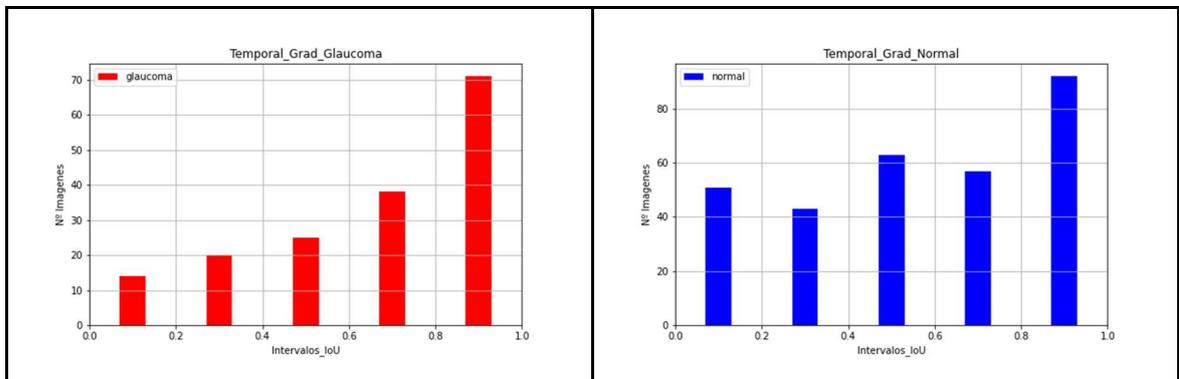


Figura 89. IoU entre mapas Grad-CAM y temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

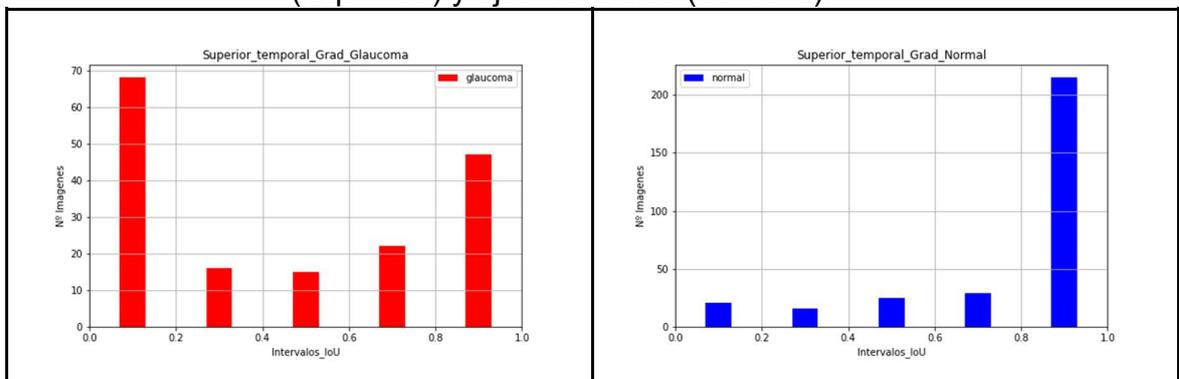


Figura 90. IoU entre mapas Grad-CAM y sector superior temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

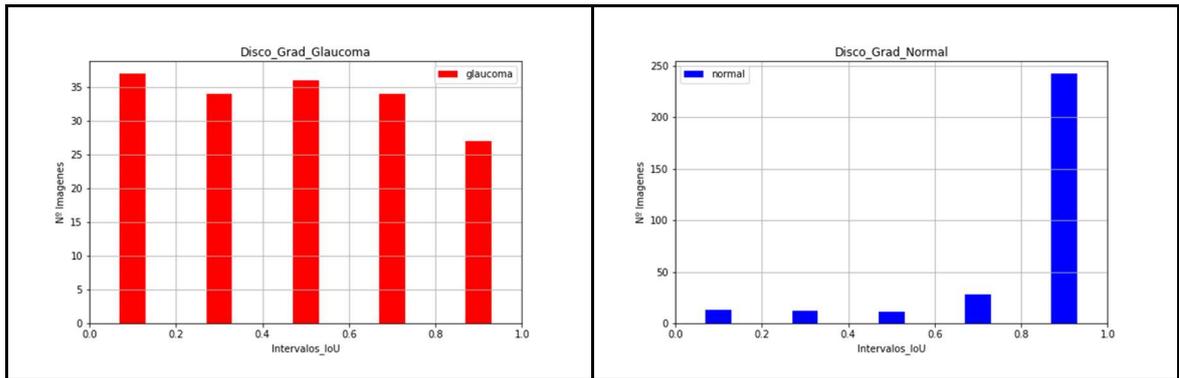


Figura 91. IoU entre mapas Grad-CAM y disco para ojos con glaucoma (izquierda) y ojos normales (derecha).

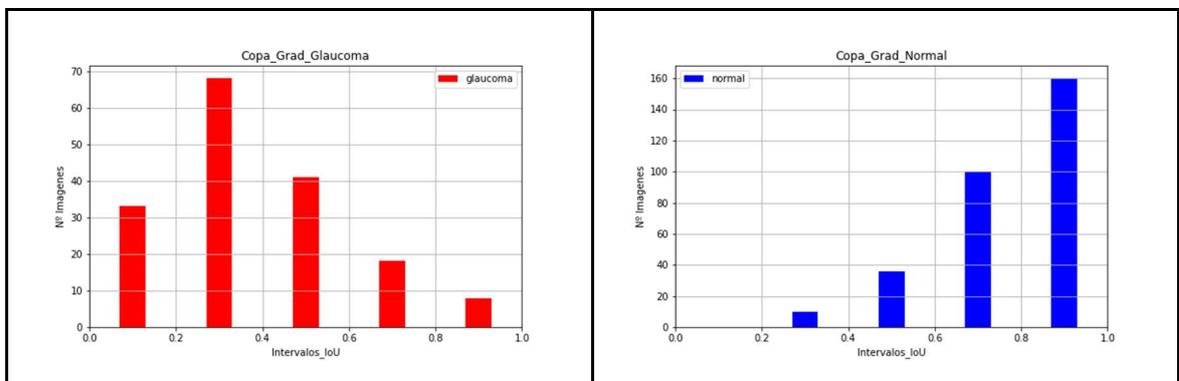


Figura 92. IoU entre mapas Grad-CAM y copa para ojos con glaucoma (izquierda) y ojos normales (derecha).

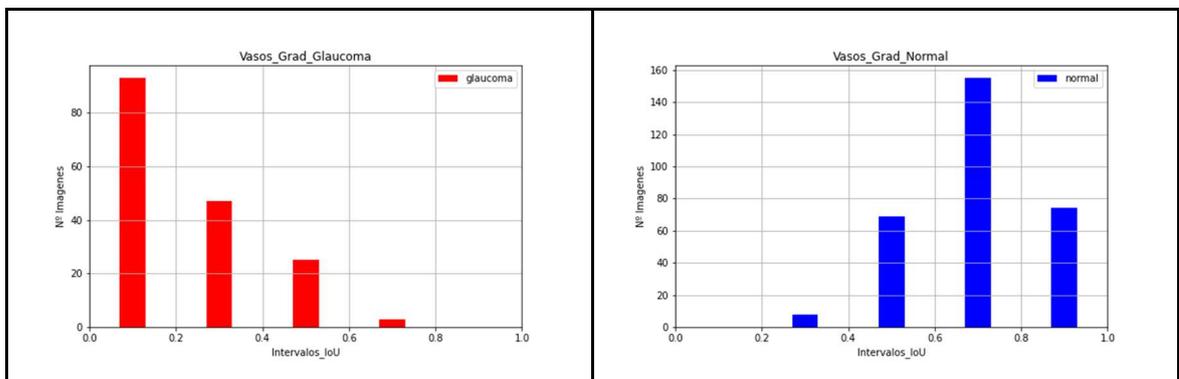


Figura 93. IoU entre mapas Grad-CAM y vasos para ojos con glaucoma (izquierda) y ojos normales (derecha).

Discusión de resultados de Grad-CAM

Todos los sectores nasales de la imagen parecen muy importantes para diagnosticar ojos sanos, aunque no tanto para glaucomas. Los sectores temporales también son relevantes para los ojos normales, y más que los nasales para los ojos con Glaucoma.



El disco, la copa e incluso los vasos son considerados por esta red cuando analiza las imágenes normales. Para el glaucoma la región más importante entre estas tres es el disco.

5.2.3.2 Resultados de Score-CAM

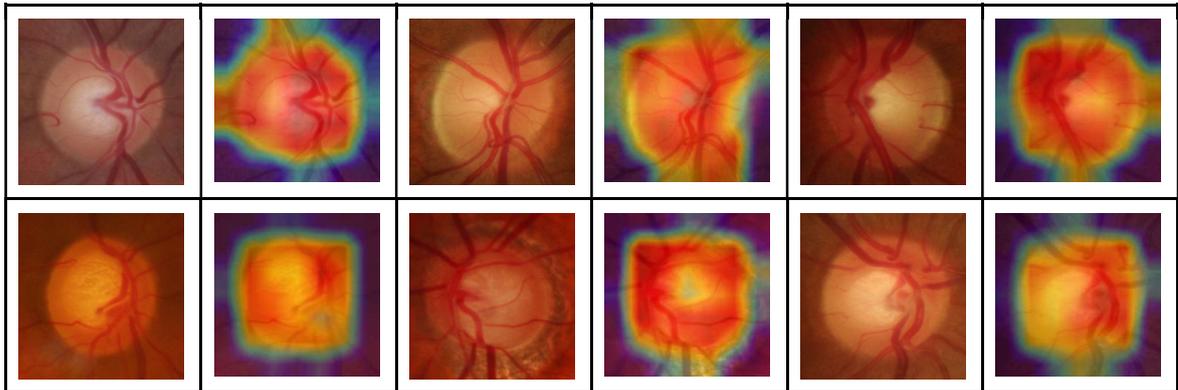


Figura 94. Imágenes de ojos sanos (primera fila) y ojos con glaucoma (segunda fila). Se muestra la imagen original y el mapa Score-CAM obtenido con la red MobileNet.

Histogramas IoU

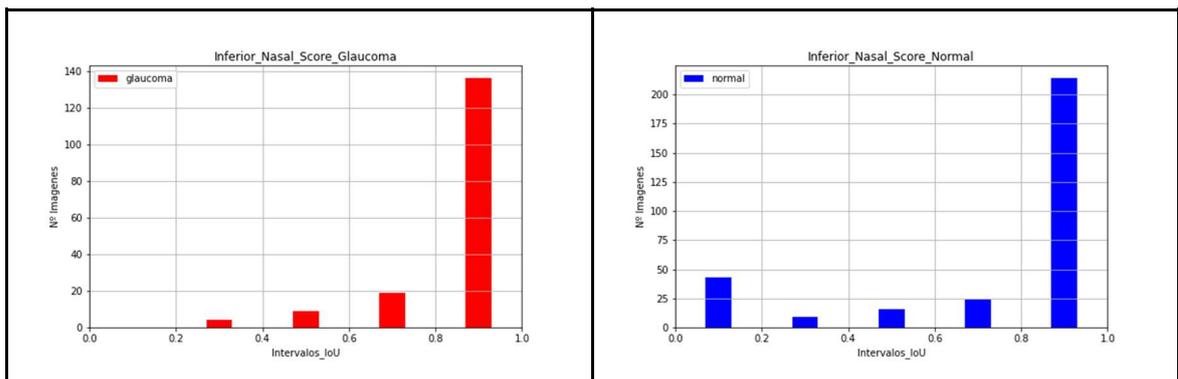


Figura 95. IoU entre mapas Score-CAM y sector nasal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

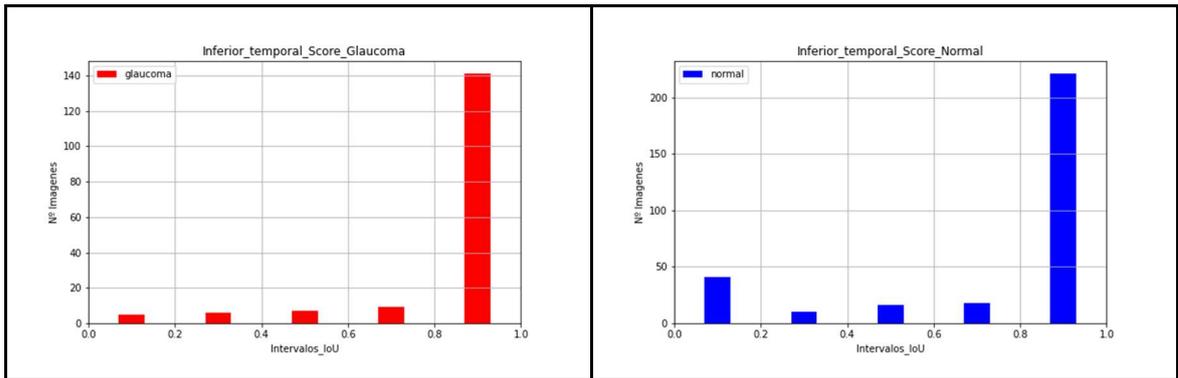


Figura 96. IoU entre mapas Score-CAM y sector temporal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

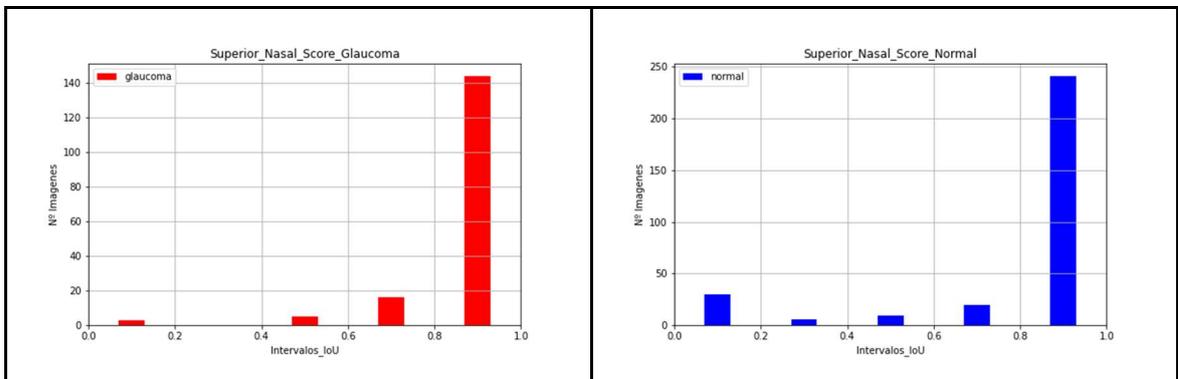


Figura 97. IoU entre mapas Score-CAM y sector nasal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

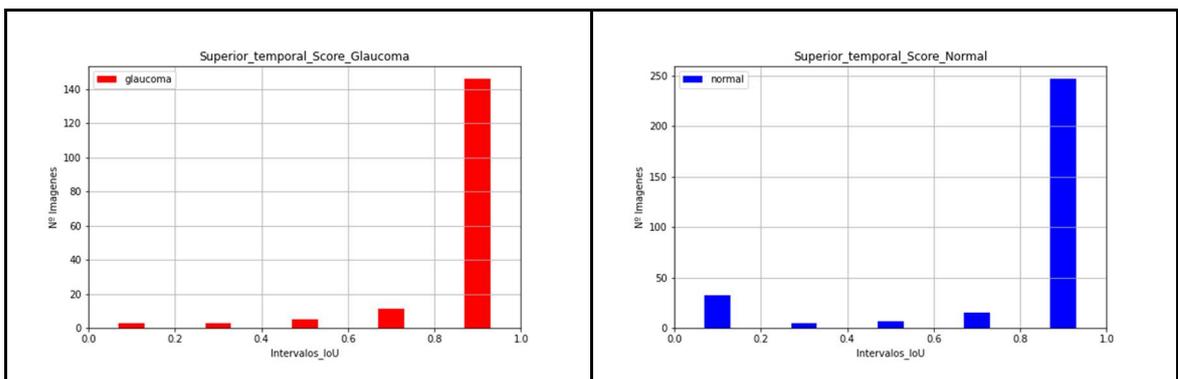


Figura 98. IoU entre mapas Score-CAM y sector temporal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

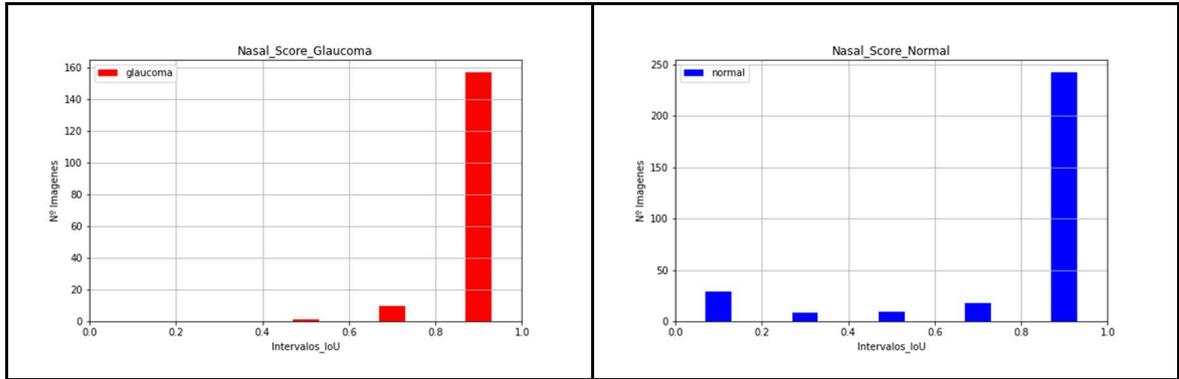


Figura 99. IoU entre mapas Score-CAM y sector nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

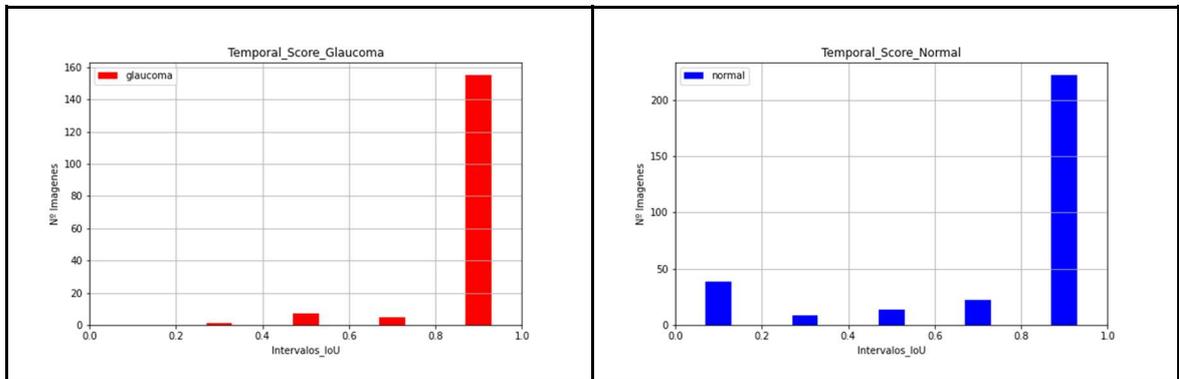


Figura 100. IoU entre mapas Score-CAM y sector temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

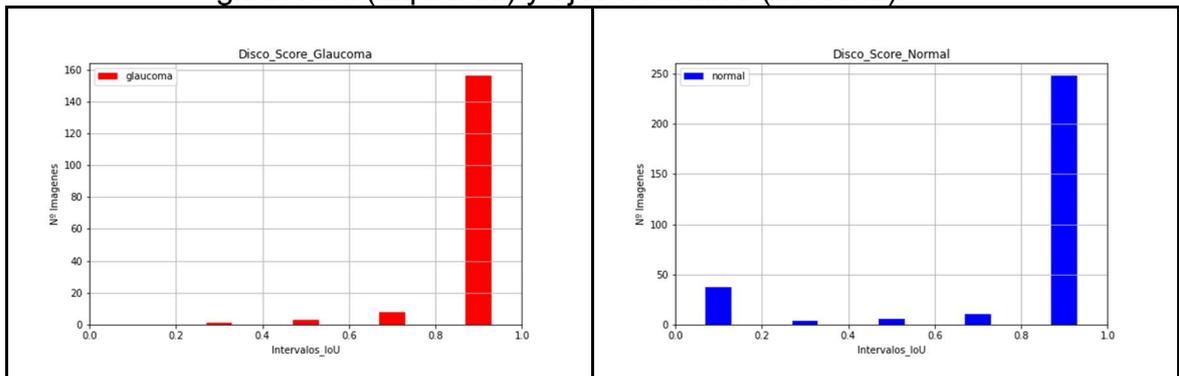


Figura 101. IoU entre mapas Score-CAM y disco para ojos con glaucoma (izquierda) y ojos normales (derecha).

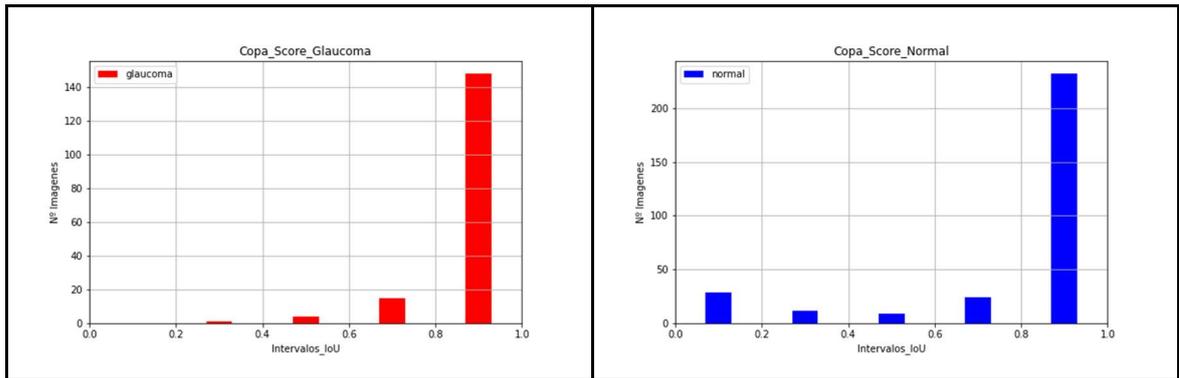


Figura 102. IoU entre mapas Score-CAM y copa para ojos con glaucoma (izquierda) y ojos normales (derecha).

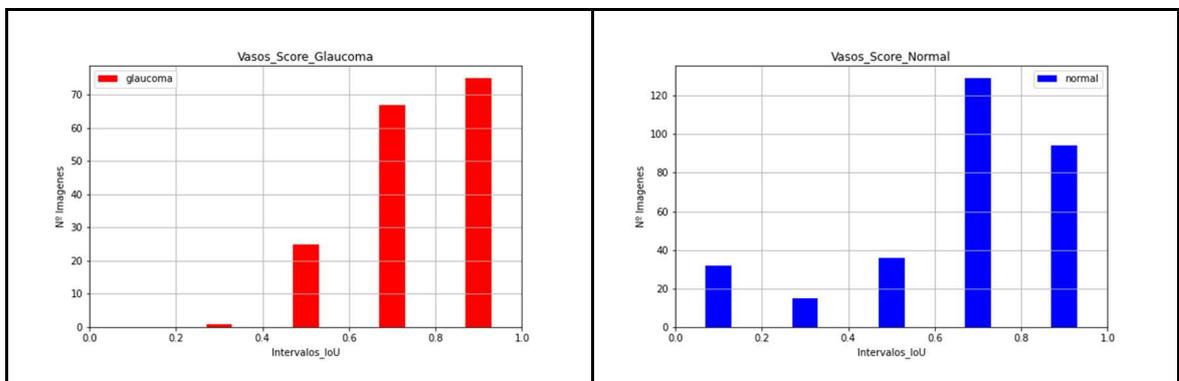


Figura 103. IoU entre mapas Score-CAM y vasos para ojos con glaucoma (izquierda) y ojos normales (derecha).

Discusión de resultados de Score-CAM

Todos los sectores nasales y temporales de la imagen parecen muy importantes para diagnosticar ojos sanos y con glaucoma. El disco, la copa y los vasos también son importantes en ambos casos.

5.2.3.3 Resultados de OCL

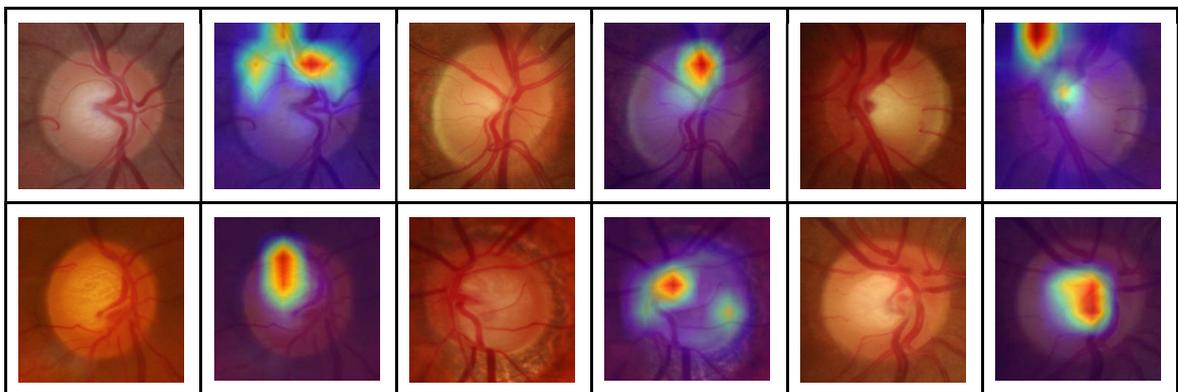




Figura 104. Imágenes de ojos sanos (primera fila) y ojos con glaucoma (segunda fila). Se muestra la imagen original y el mapa OCL obtenido con la red MobileNet.

Histogramas IoU

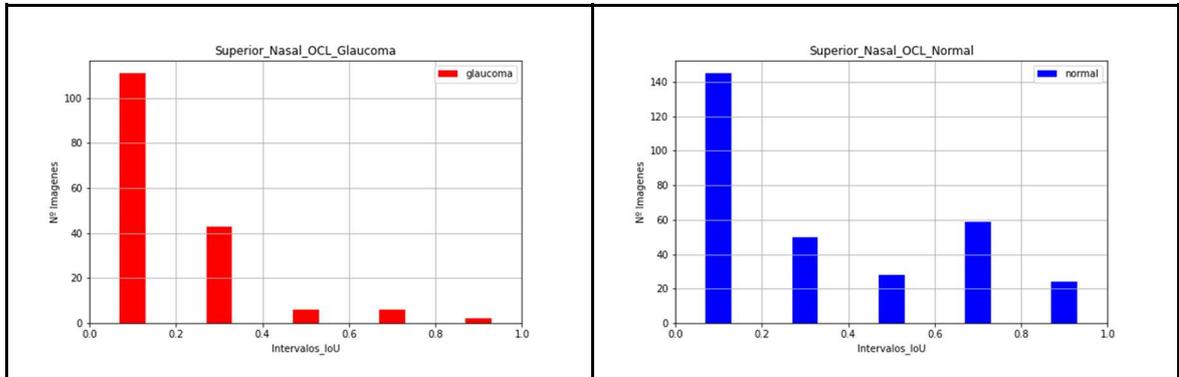


Figura 105. IoU entre mapas OCL y sector nasal superior para ojos con glaucoma (izquierda) y ojos normales (derecha)

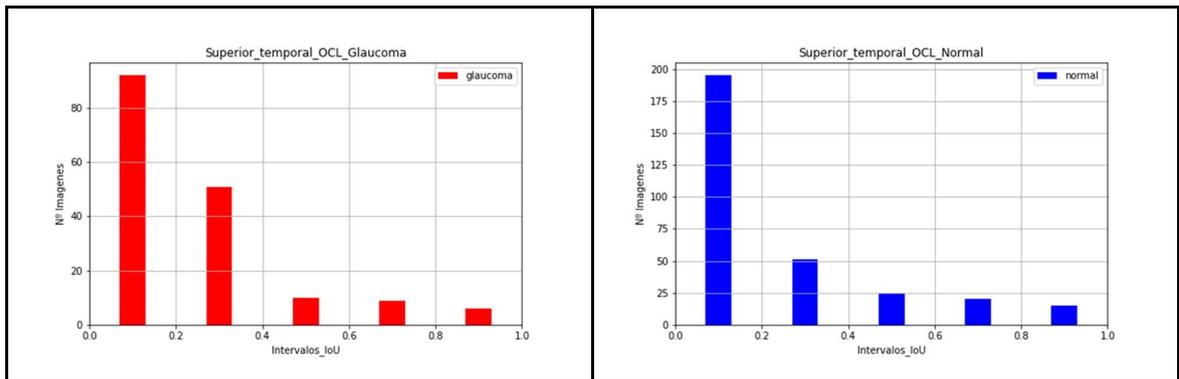


Figura 106. IoU entre mapas OCL y sector temporal superior para ojos con glaucoma (izquierda) y ojos normales (derecha)

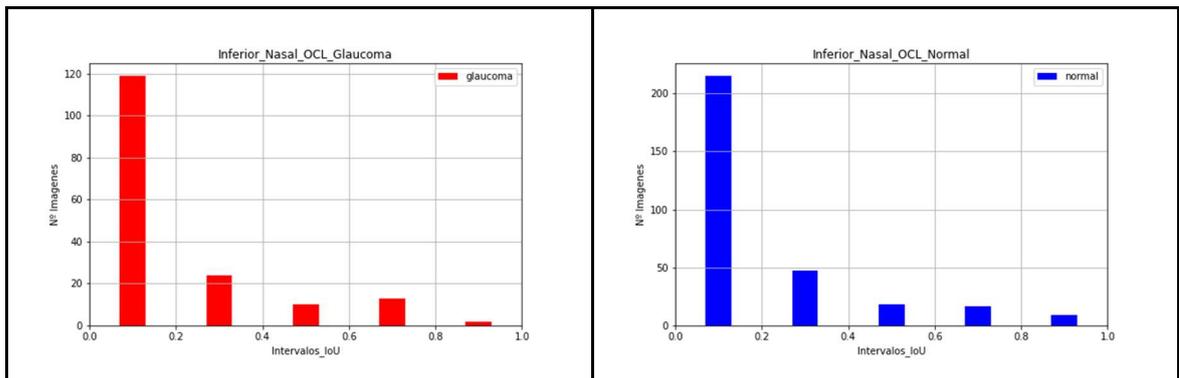


Figura 107. IoU entre mapas OCL y sector nasal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha)

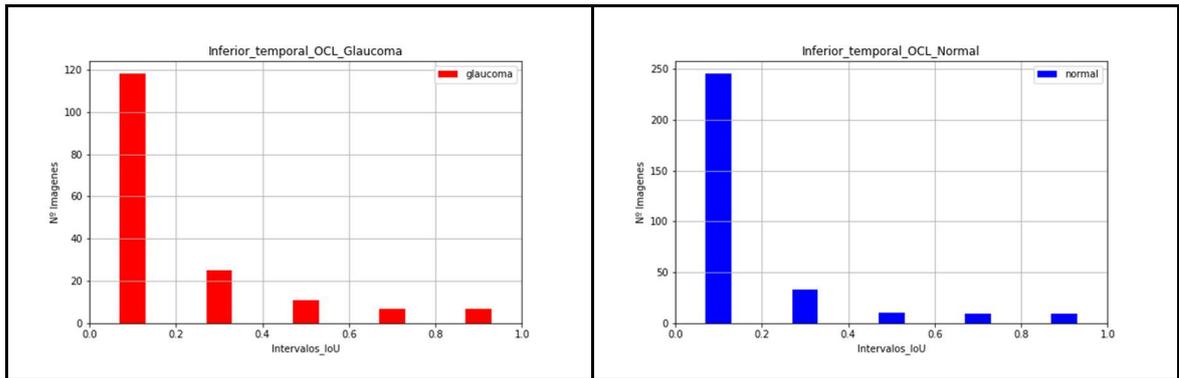


Figura 108. IoU entre mapas OCL y sector temporal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha)

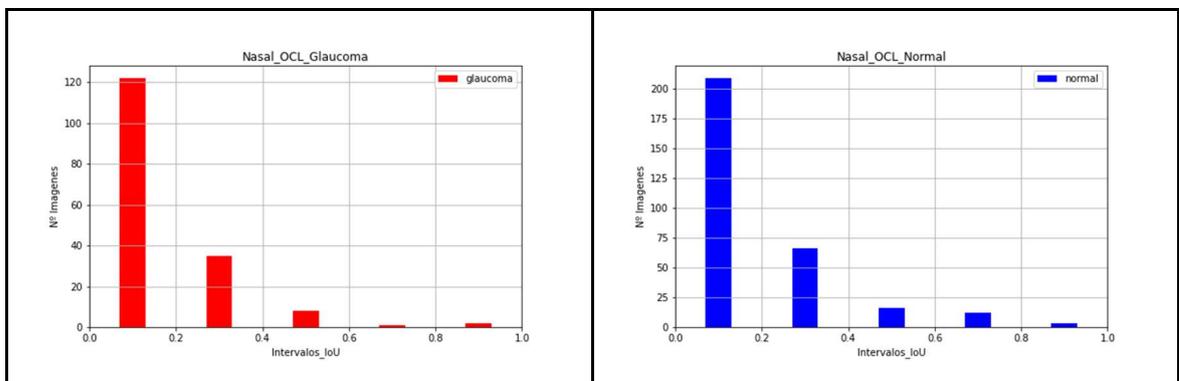


Figura 109. IoU entre mapas OCL y sector nasal para ojos con glaucoma (izquierda) y ojos normales (derecha)

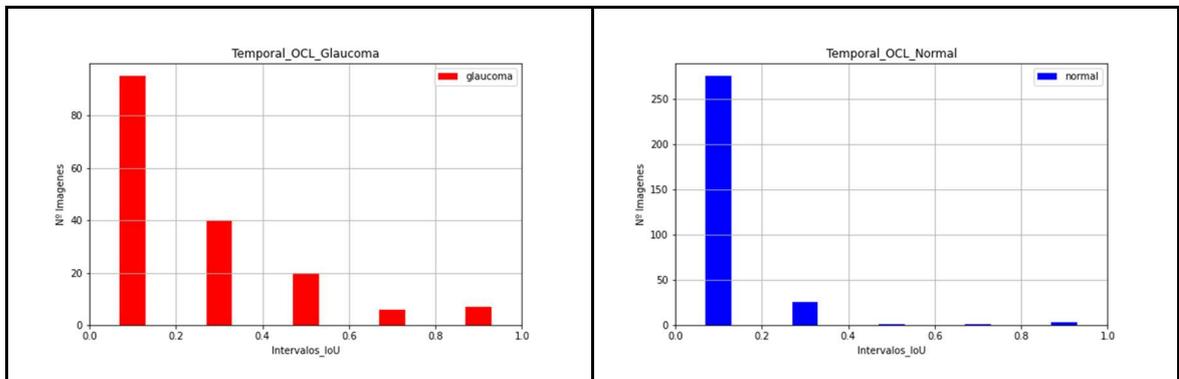


Figura 110. IoU entre mapas OCL y sector temporal para ojos con glaucoma (izquierda) y ojos normales (derecha)

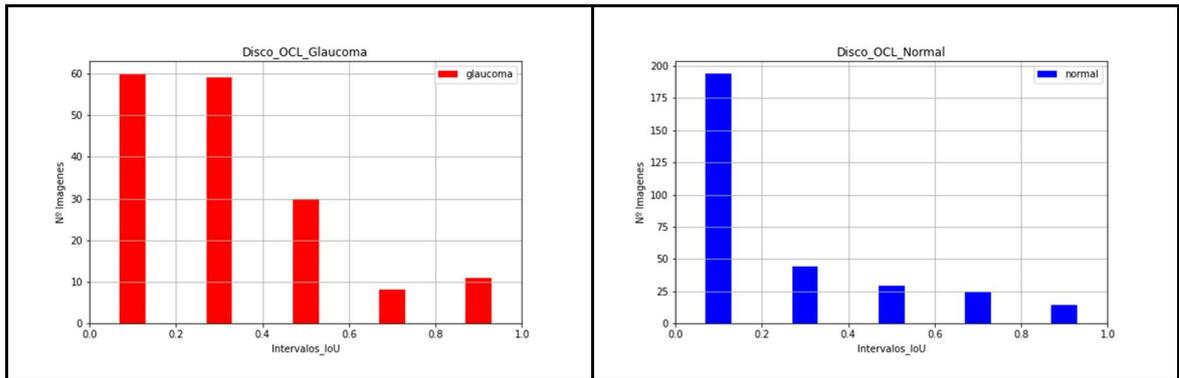


Figura 111. IoU entre mapas OCL y disco para ojos con glaucoma (izquierda) y ojos normales (derecha)

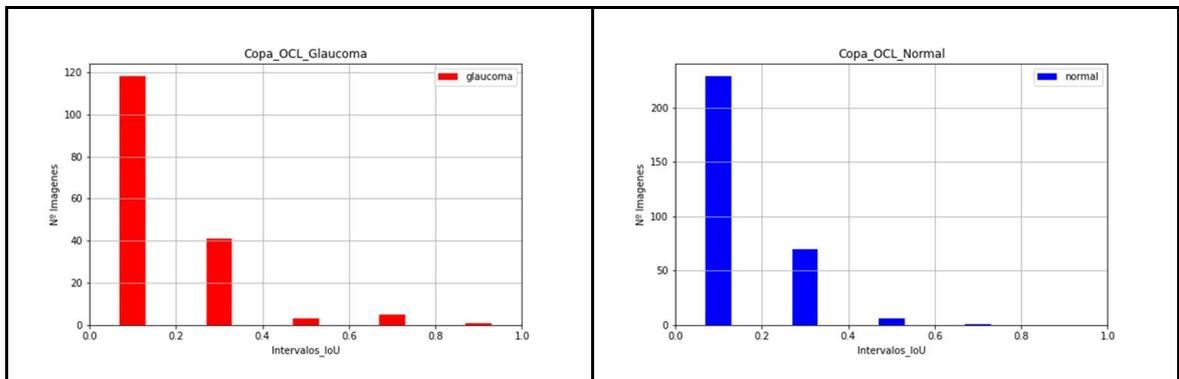


Figura 112. IoU entre mapas OCL y copa para ojos con glaucoma (izquierda) y ojos normales (derecha)

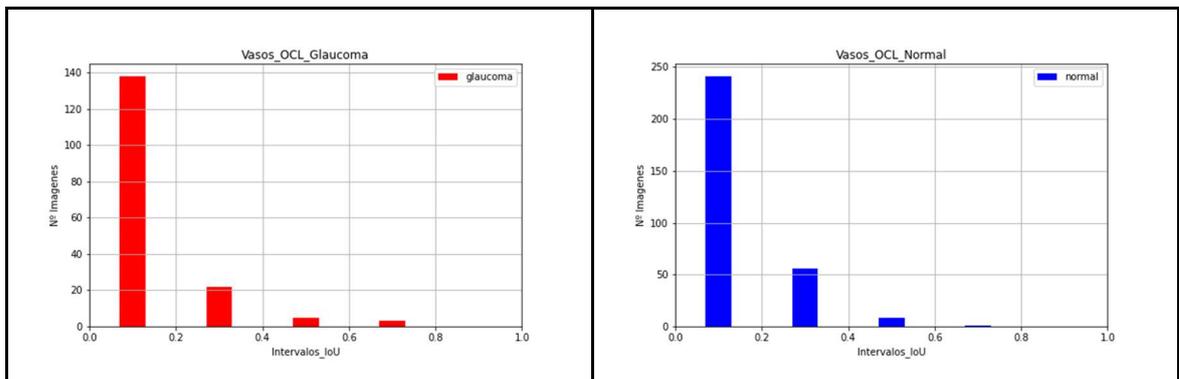


Figura 113. IoU entre mapas OCL y vasos para ojos con glaucoma (izquierda) y ojos normales (derecha)

Discusión de resultados de OCL

Los sectores superiores parecen ser los más relevantes, sobre todo para ojos sanos. El disco es la región más importante para los ojos con glaucoma.



5.2.3.4 Comparativa de resultados de los tres métodos con la MobileNet

Aunque las técnicas aplicadas no muestran resultados tan claros como con las redes anteriores, se puede destacar que parece que los sectores nasales y superiores son más influyentes para el diagnóstico de ojos sanos, y los sectores temporales para el diagnóstico de la enfermedad. El disco sigue siendo analizado en ambos casos.

5.2.4 MobileNetV2

5.2.4.1 Resultados de Grad-CAM

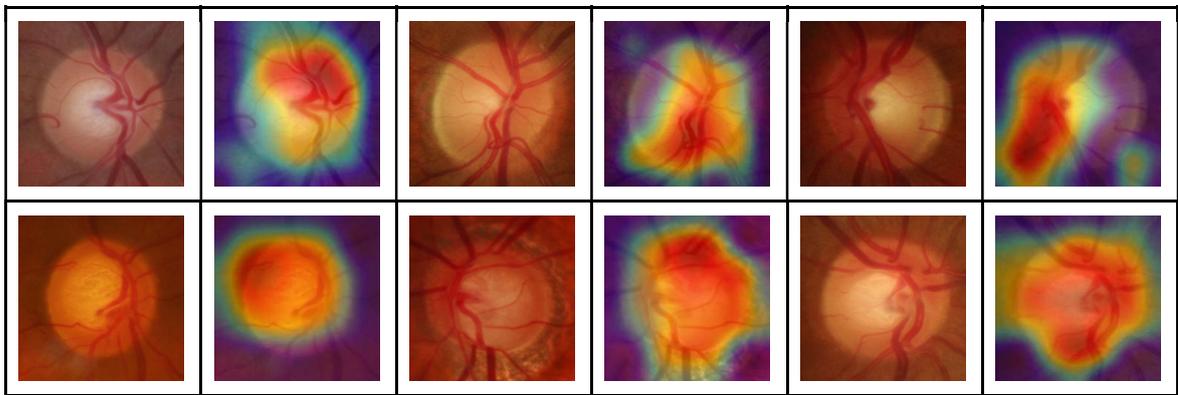


Figura 114. Imágenes de ojos sanos (primera fila) y ojos con glaucoma (segunda fila). Se muestra la imagen original y el mapa Grad-CAM obtenido con la red MobileNetV2.

Histogramas IoU

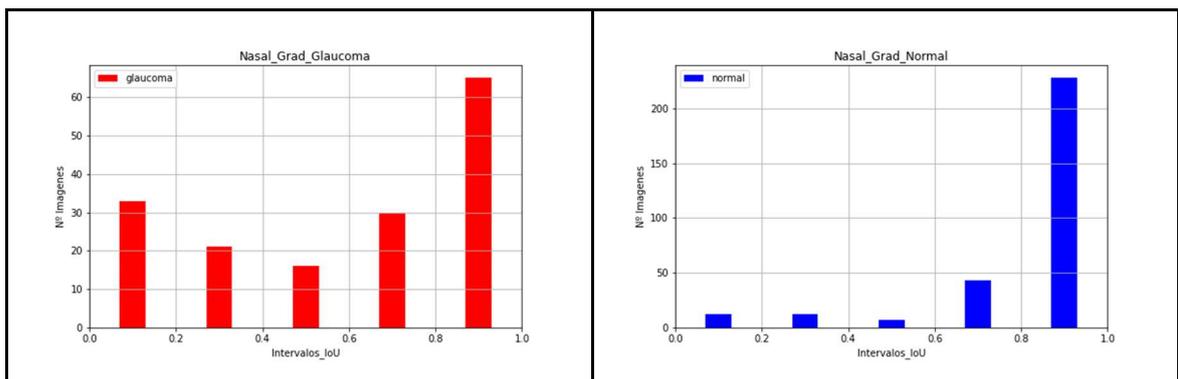


Figura 115. IoU entre mapas Grad-CAM y sector nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

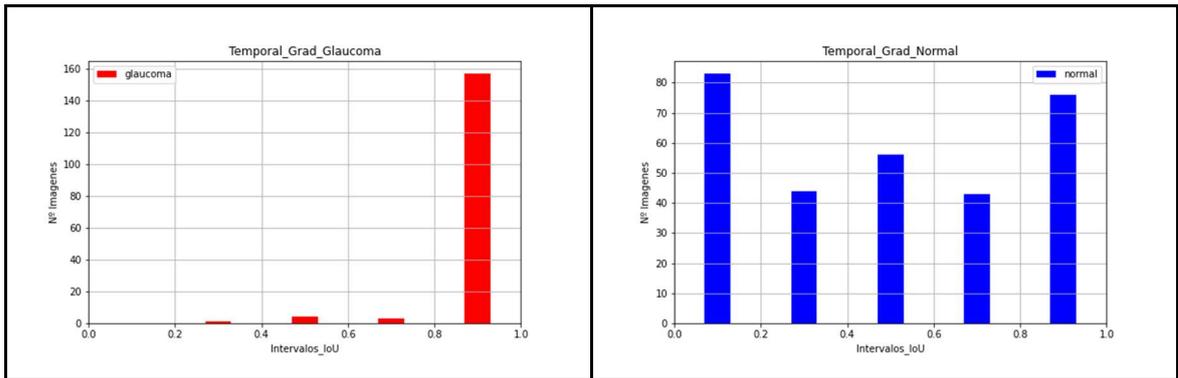


Figura 116. IoU entre mapas Grad-CAM y sector temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

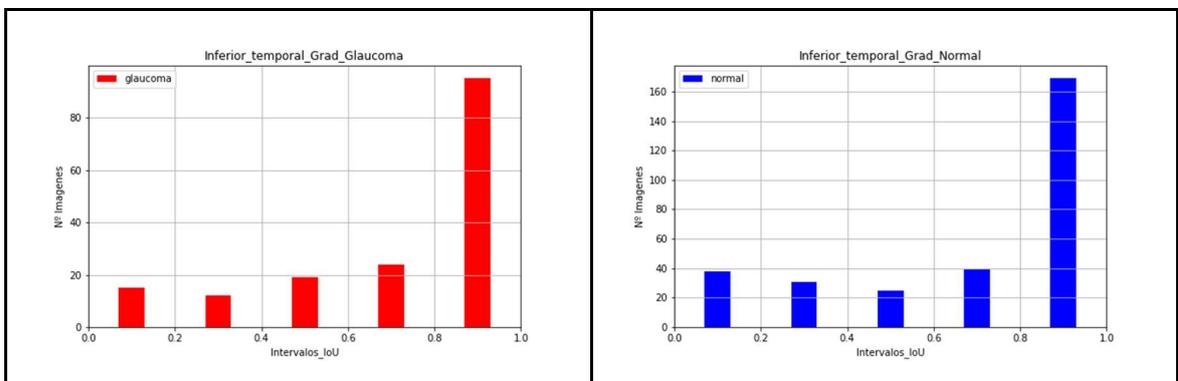


Figura 117. IoU entre mapas Grad-CAM y sector temporal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

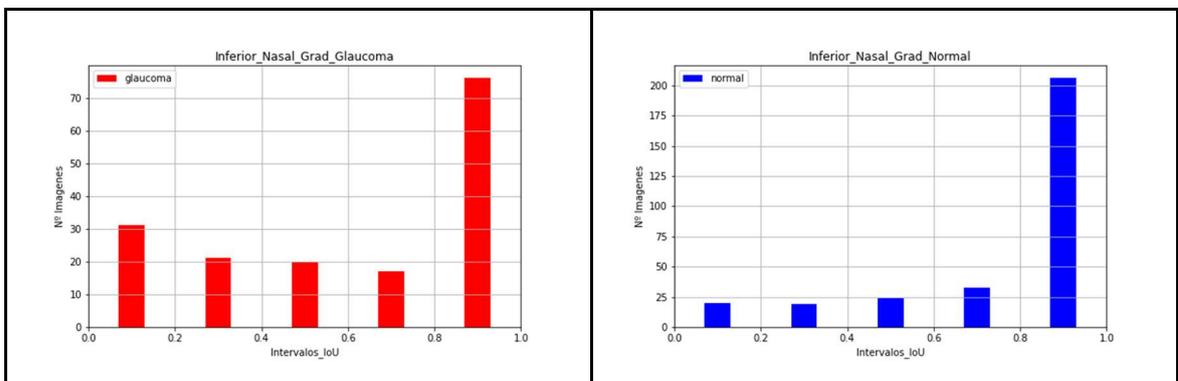


Figura 118. IoU entre mapas Grad-CAM y sector nasal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

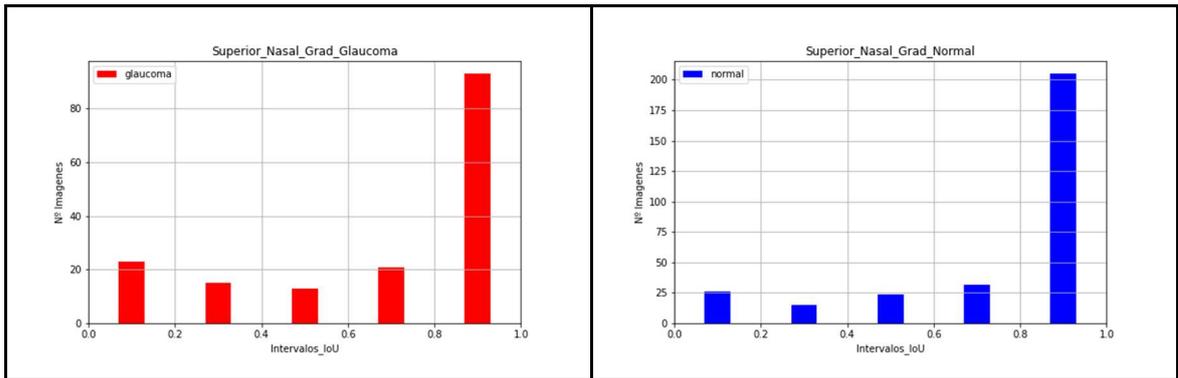


Figura 119. IoU entre mapas Grad-CAM y sector superior nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

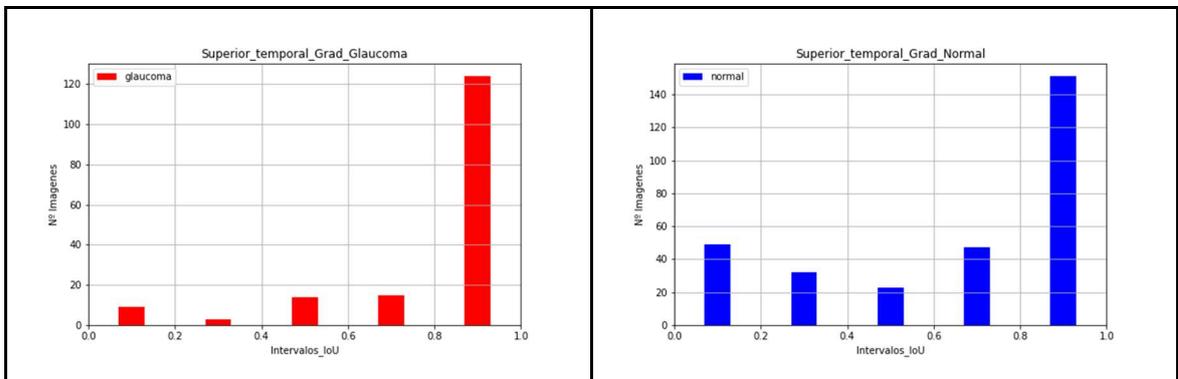


Figura 120. IoU entre mapas Grad-CAM y sector superior temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

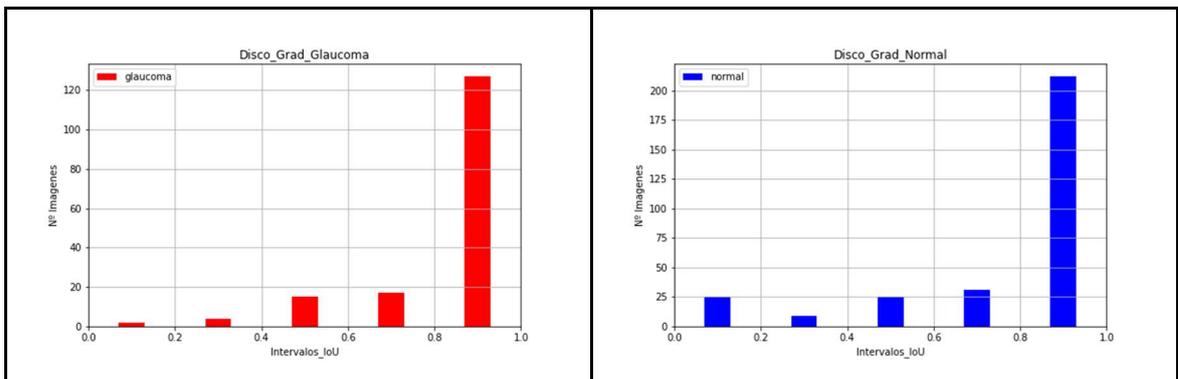


Figura 121. IoU entre mapas Grad-CAM y disco para ojos con glaucoma (izquierda) y ojos normales (derecha).

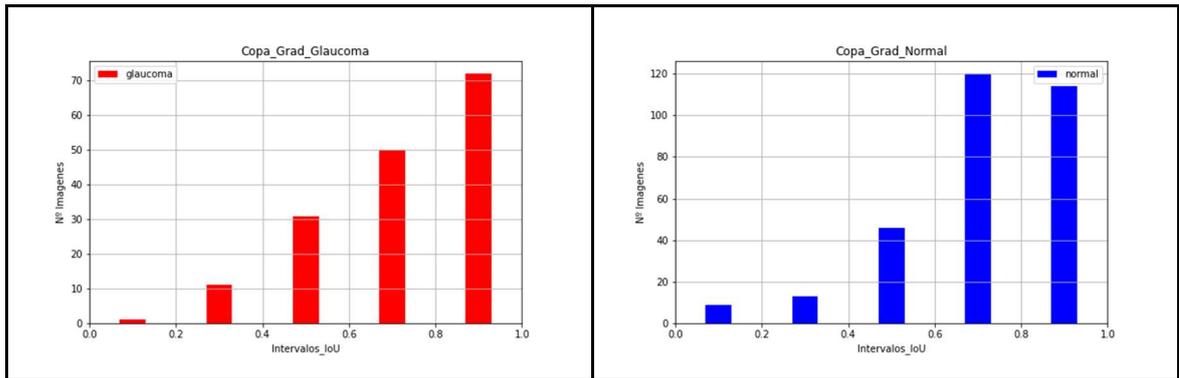


Figura 122. IoU entre mapas Grad-CAM y copa para ojos con glaucoma (izquierda) y ojos normales (derecha).

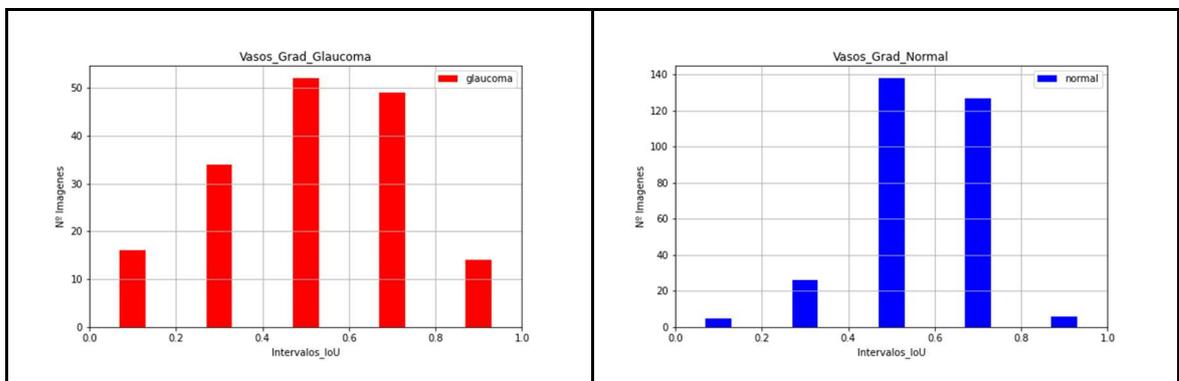


Figura 123. IoU entre mapas Grad-CAM y vasos para ojos con glaucoma (izquierda) y ojos normales (derecha).

Discusión de resultados de Grad-CAM

Todos los sectores nasales y temporales de la imagen parecen muy importantes para los ojos sanos y con glaucoma. El disco y la copa son también regiones muy relevantes para ambos casos.

5.2.4.2 Resultados de Score-CAM

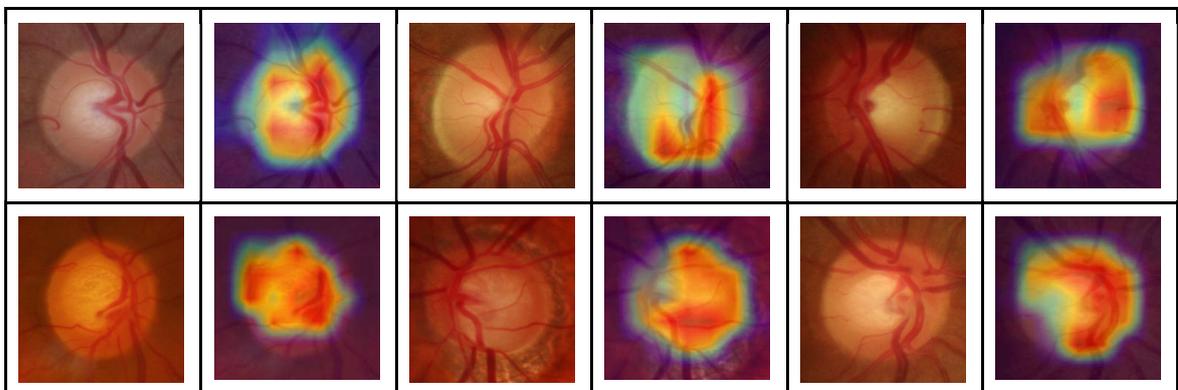


Figura 124. Imágenes de ojos sanos (primera fila) y ojos con glaucoma (segunda fila). Se muestra la imagen original y el mapa Grad-CAM obtenido con la red MobileNetV2.



Histogramas IoU

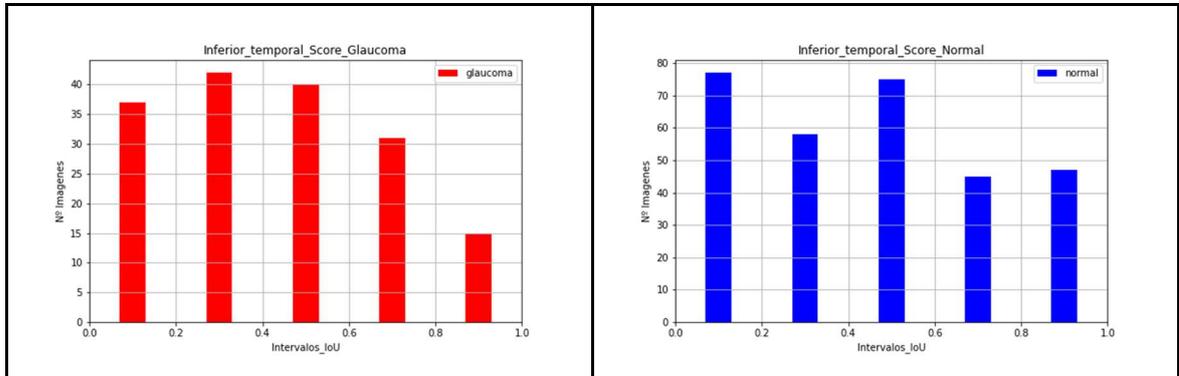


Figura 125. IoU entre mapas Score-CAM y sector temporal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

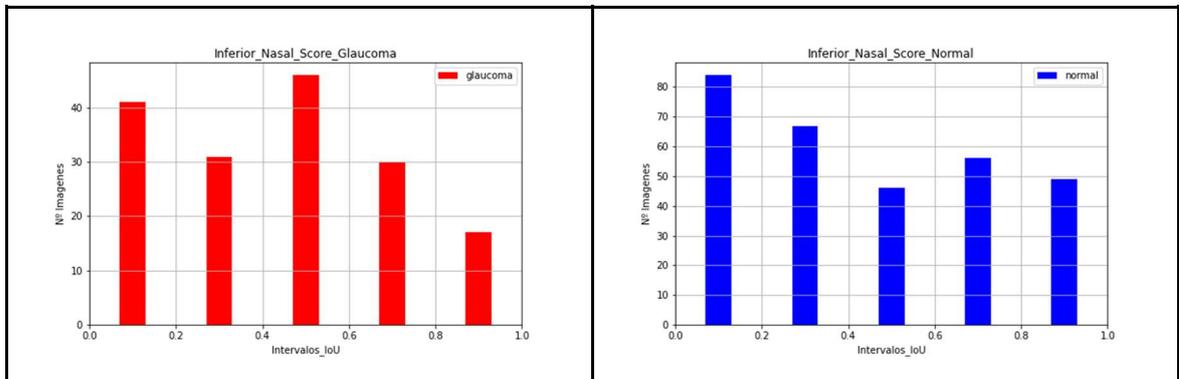


Figura 126. IoU entre mapas Score-CAM y sector nasal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

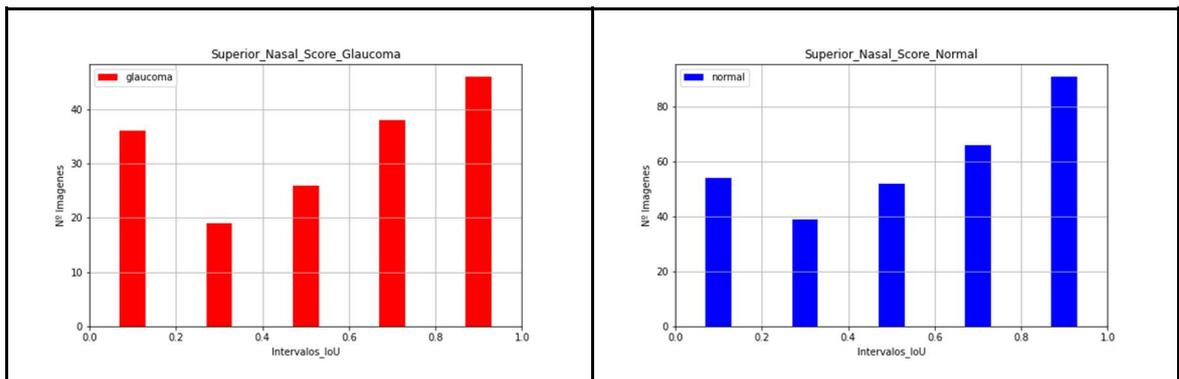


Figura 127. IoU entre mapas Score-CAM y sector nasal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

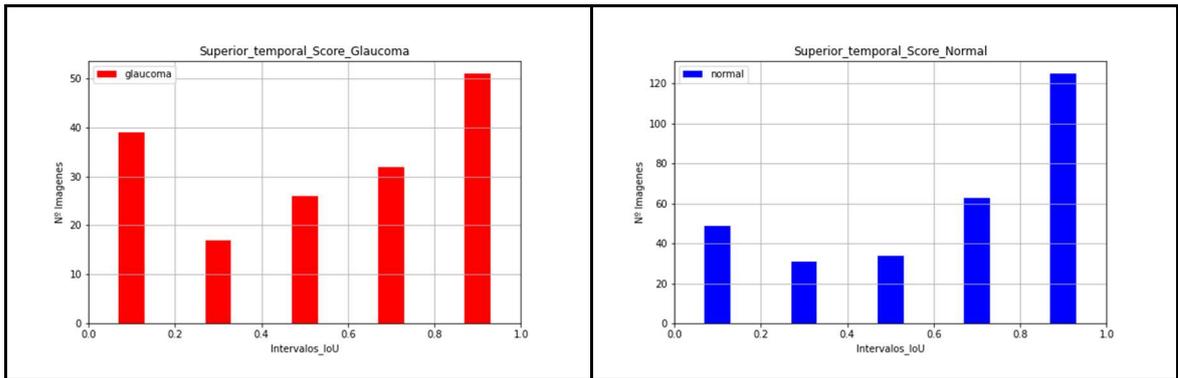


Figura 128. IoU entre mapas Score-CAM y sector temporal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

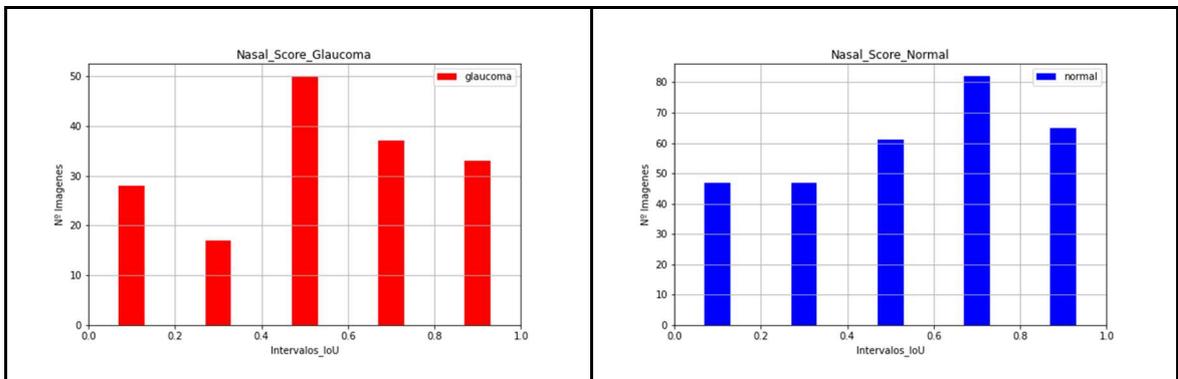


Figura 129. IoU entre mapas Score-CAM y sector nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

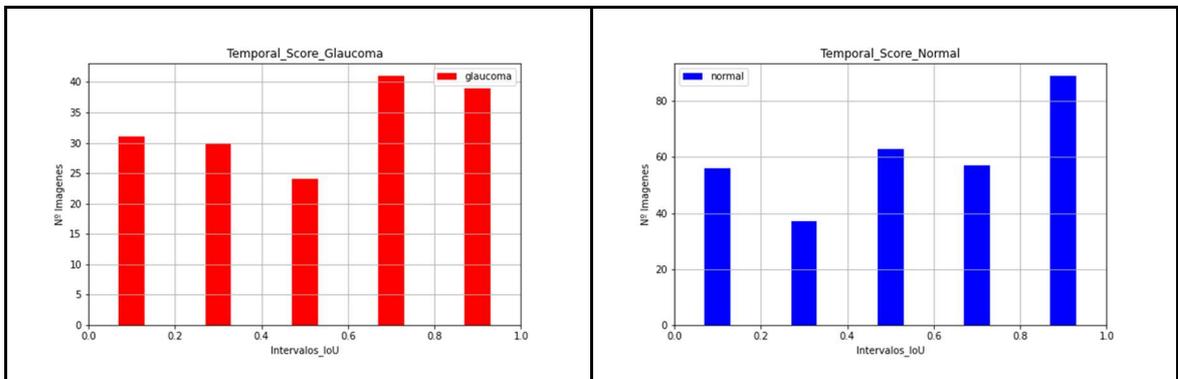


Figura 130. IoU entre mapas Score-CAM y sector temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

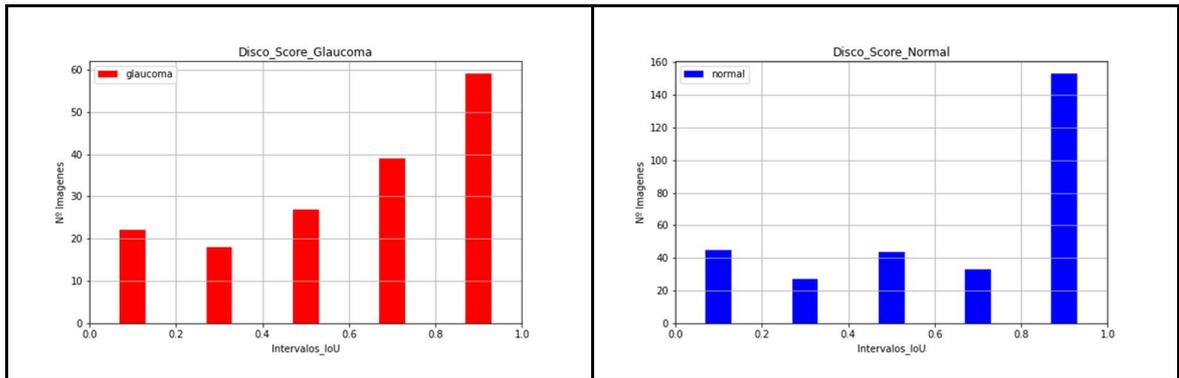


Figura 131. IoU entre mapas Score-CAM y disco para ojos con glaucoma (izquierda) y ojos normales (derecha).

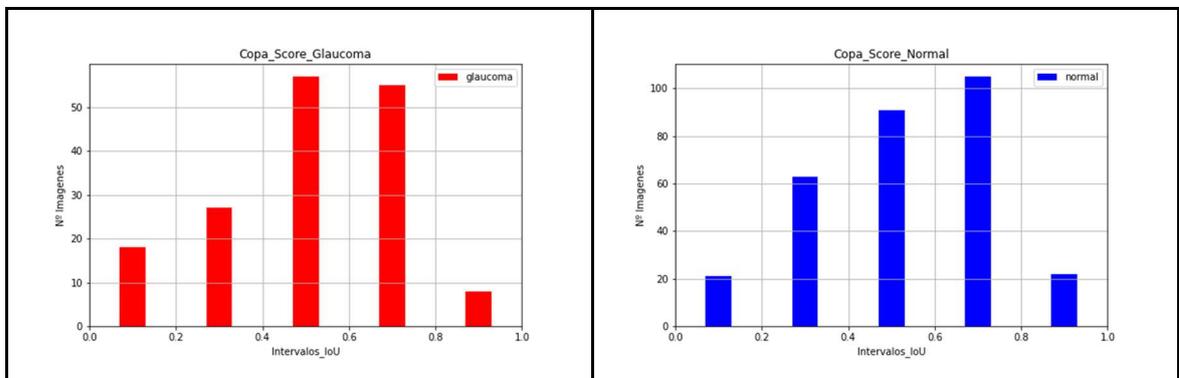


Figura 132. IoU entre mapas Score-CAM y copa para ojos con glaucoma (izquierda) y ojos normales (derecha).

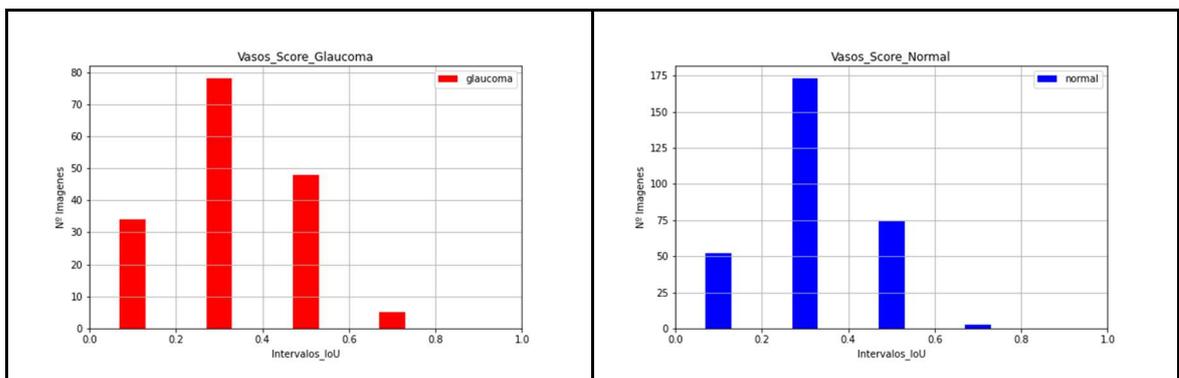


Figura 133. IoU entre mapas Score-CAM y vasos para ojos con glaucoma (izquierda) y ojos normales (derecha).

Discusión de resultados de Score-CAM

En los mapas Score-CAM los sectores superiores son más importantes que los inferiores, para ambos tipos de imágenes. Entre el sector nasal y temporal, el temporal parece más relevante.

Analizando las regiones, la región más destacada por los mapas es el disco, en mucha menor medida la copa, y los vasos son irrelevantes.



5.2.4.3 Resultados de OCL

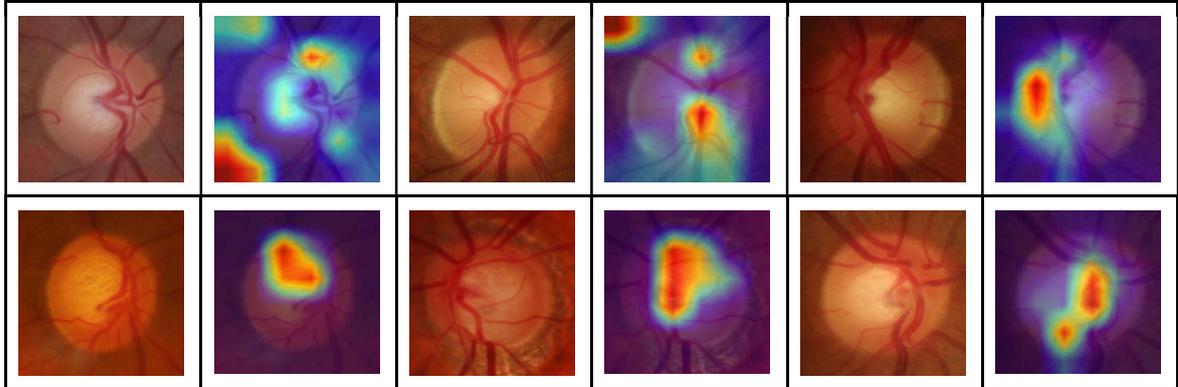


Figura 134. Imágenes de ojos sanos (primera fila) y ojos con glaucoma (segunda fila). Se muestra la imagen original y el mapa OCL obtenido con la red MobileNetV2.

Histogramas IoU

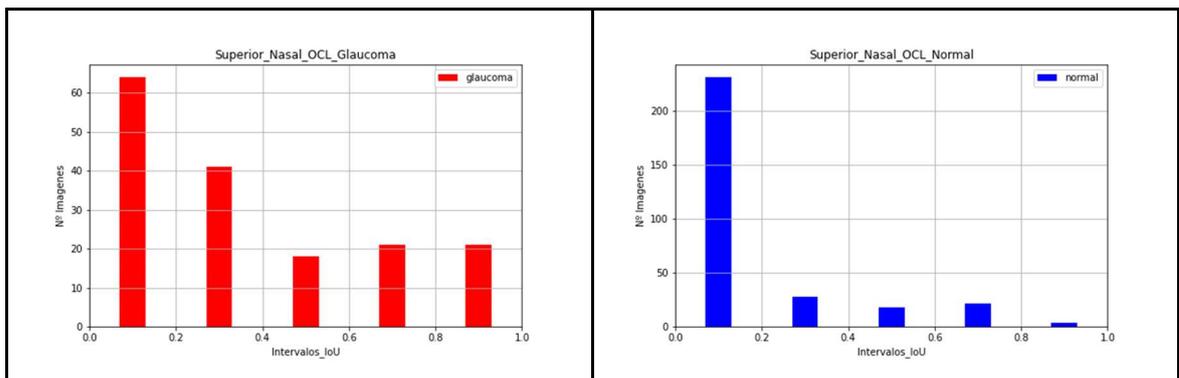


Figura 135. IoU entre mapas OCL y sector nasal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

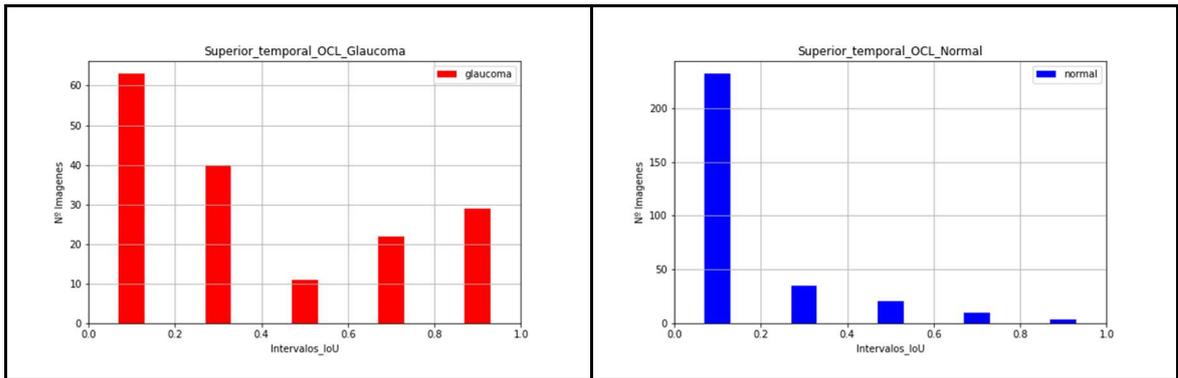


Figura 136. IoU entre mapas OCL y sector temporal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

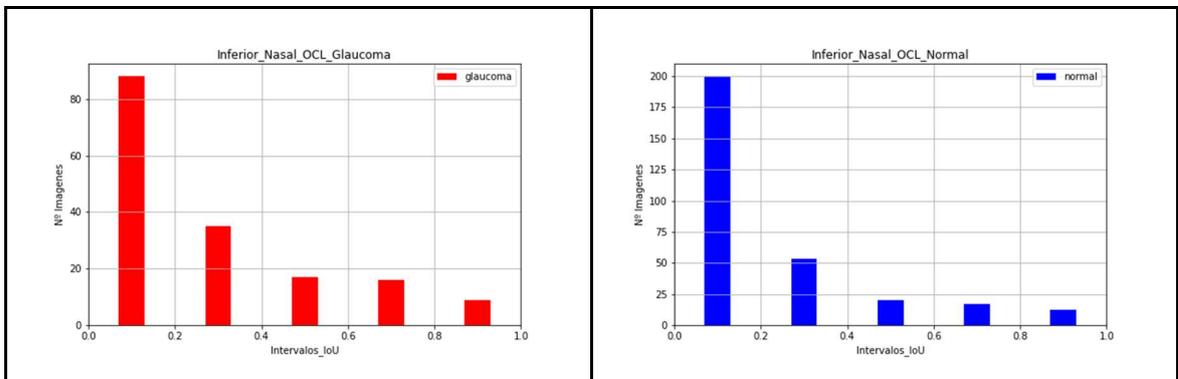


Figura 137. IoU entre mapas OCL y sector nasal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

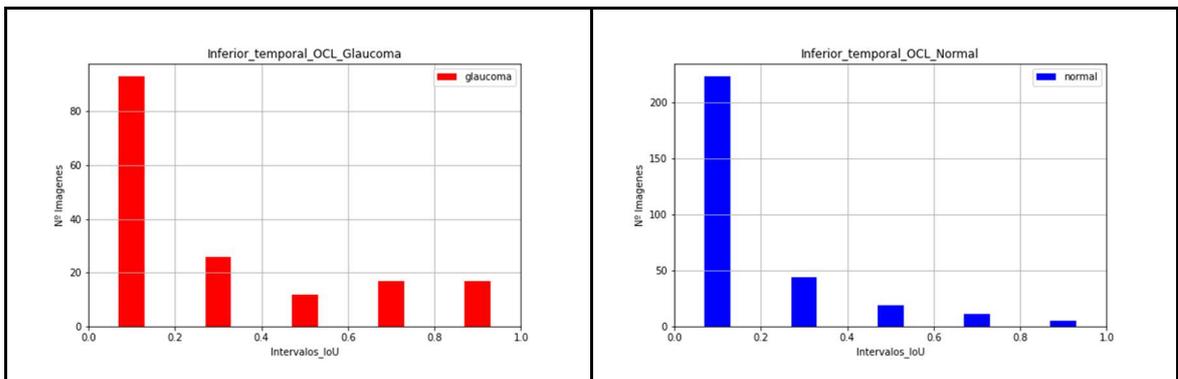


Figura 138. IoU entre mapas OCL y sector temporal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

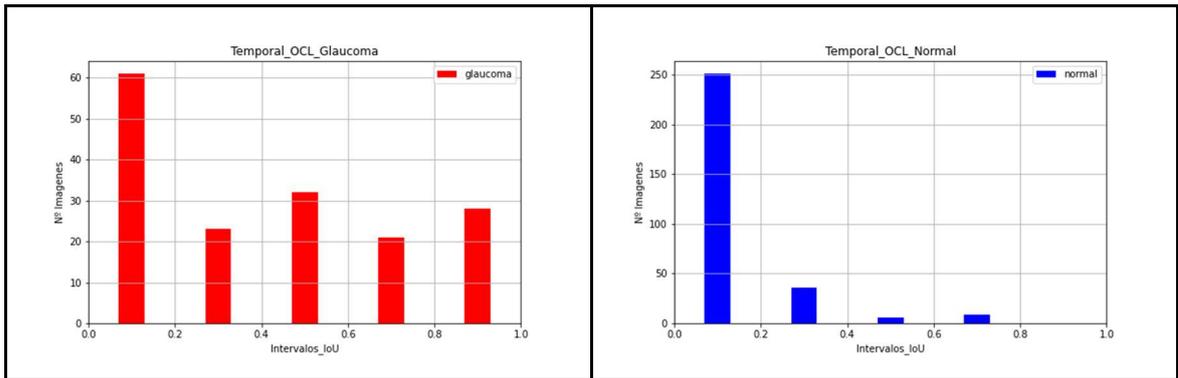


Figura 139. IoU entre mapas OCL y sector temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

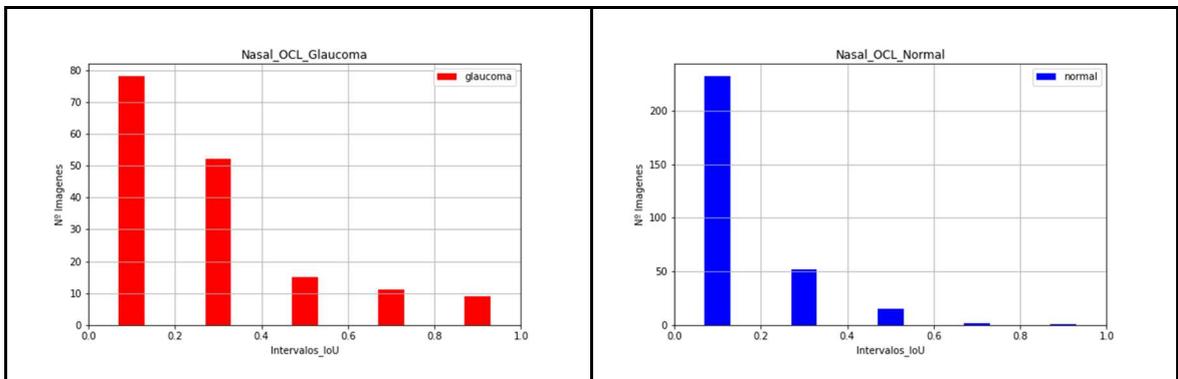


Figura 140. IoU entre mapas OCL y sector nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

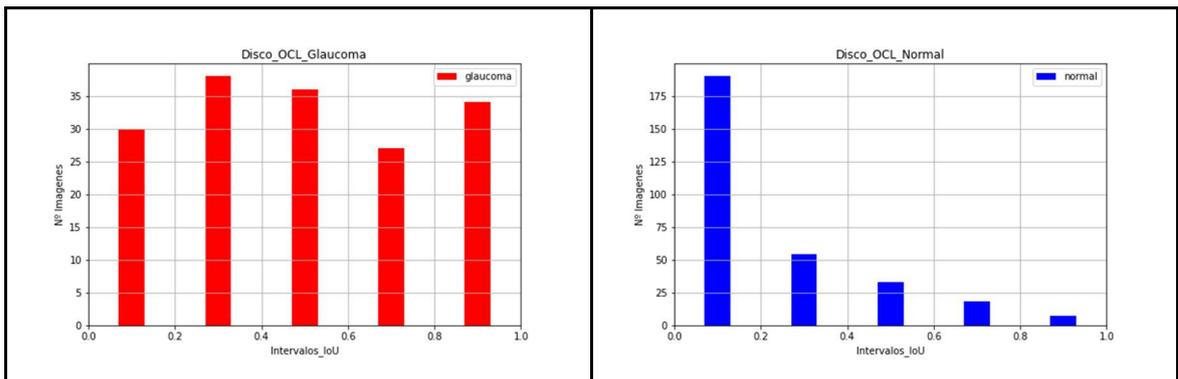


Figura 141. IoU entre mapas OCL y disco para ojos con glaucoma (izquierda) y ojos normales (derecha).

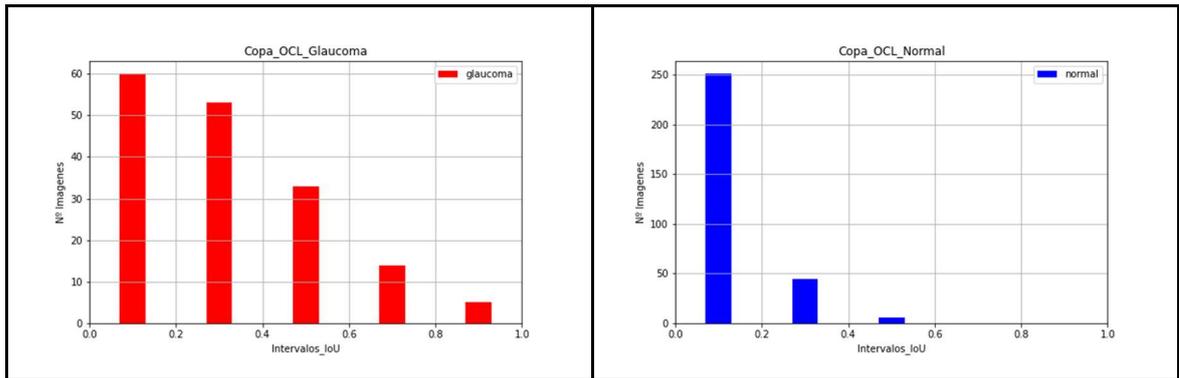


Figura 142. IoU entre mapas OCL y copa para ojos con glaucoma (izquierda) y ojos normales (derecha).

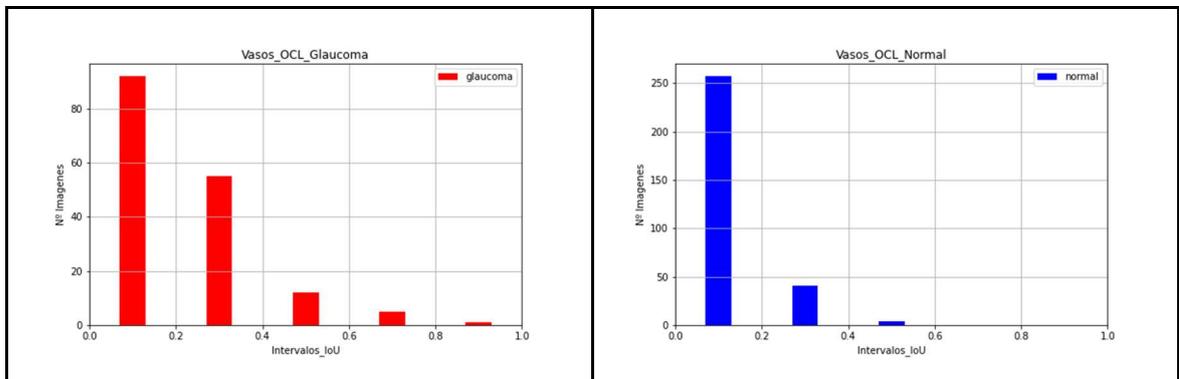


Figura 143. IoU entre mapas OCL y vasos para ojos con glaucoma (izquierda) y ojos normales (derecha).

Discusión de resultados de OCL

Los sectores superiores y el sector temporal son más relevantes en las imágenes de glaucoma. El disco también parece muy relevante en caso de glaucoma.

5.2.4.4 Comparativa de resultados de los tres métodos con la MobileNetV2

En general, se puede apreciar que los tres métodos indican que esta red se fija en los sectores superiores, en el temporal y en el disco.

5.2.5 DenseNet

5.2.5.1 Resultados de Grad-CAM

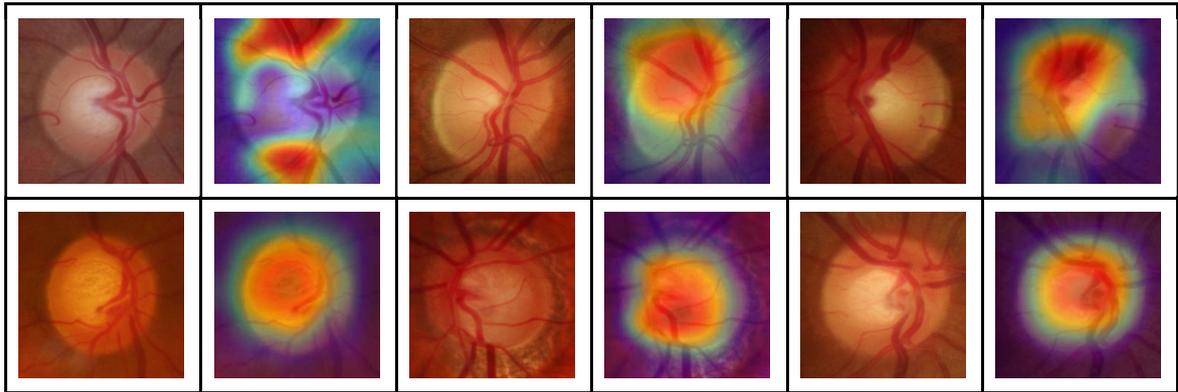


Figura 144. Imágenes de ojos sanos (primera fila) y ojos con glaucoma (segunda fila). Se muestra la imagen original y el mapa Grad-CAM obtenido con la red DenseNet.

Histogramas IoU

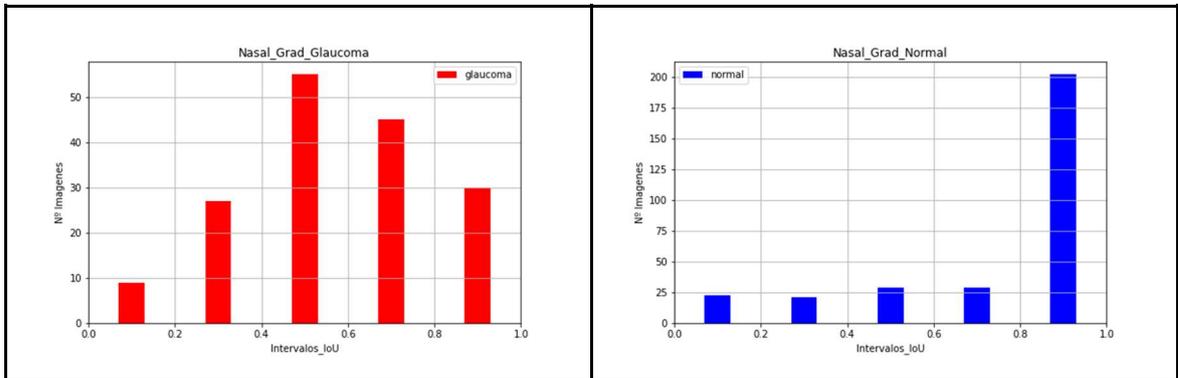


Figura 145. IoU entre mapas Grad-CAM y sector nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

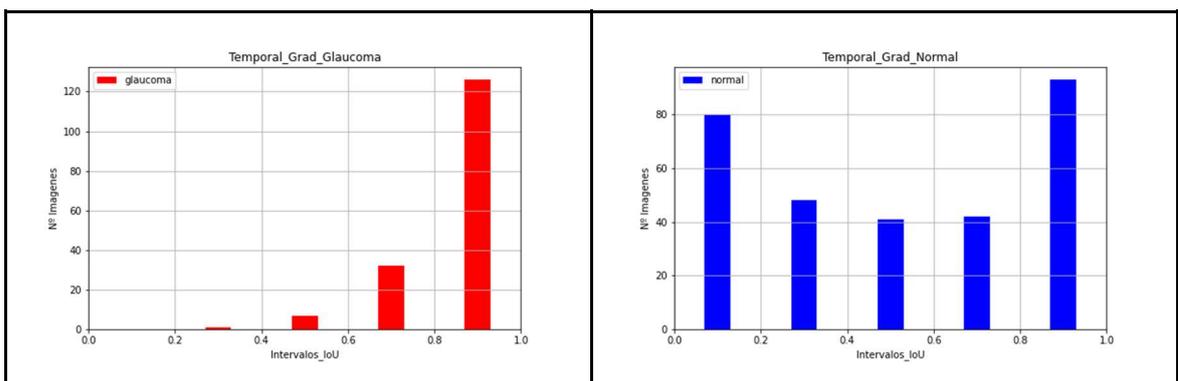


Figura 146. IoU entre mapas Grad-CAM y sector temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

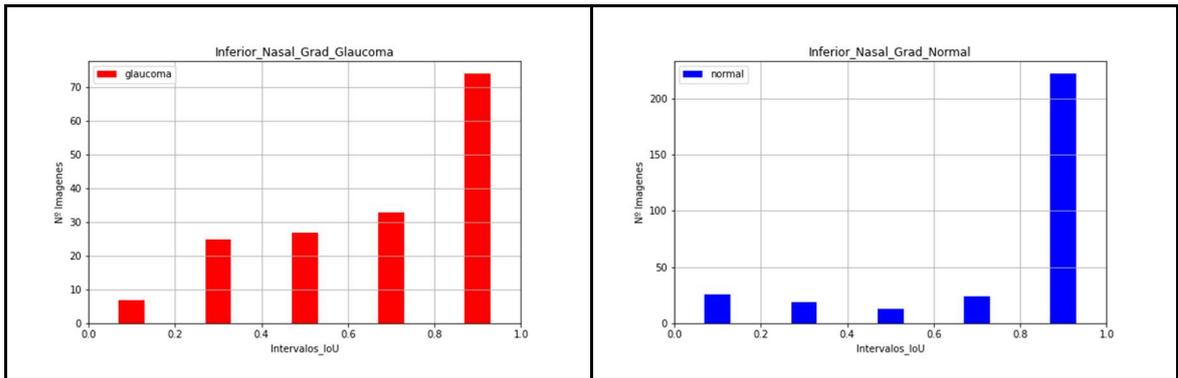


Figura 147. IoU entre mapas Grad-CAM y sector nasal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

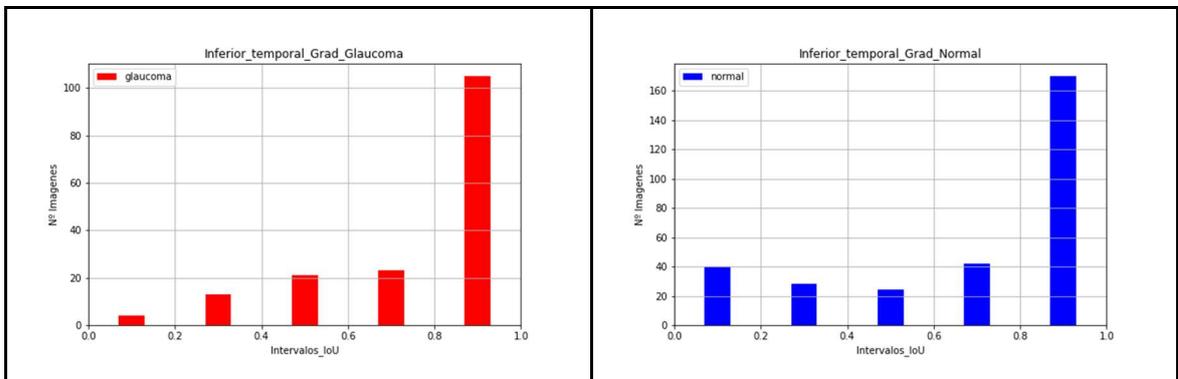


Figura 148. IoU entre mapas Grad-CAM y sector temporal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

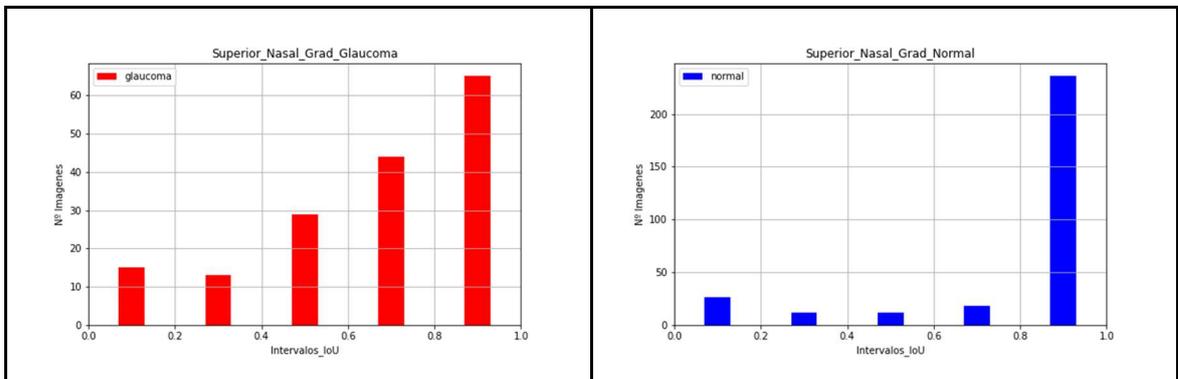


Figura 149. IoU entre mapas Grad-CAM y sector nasal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

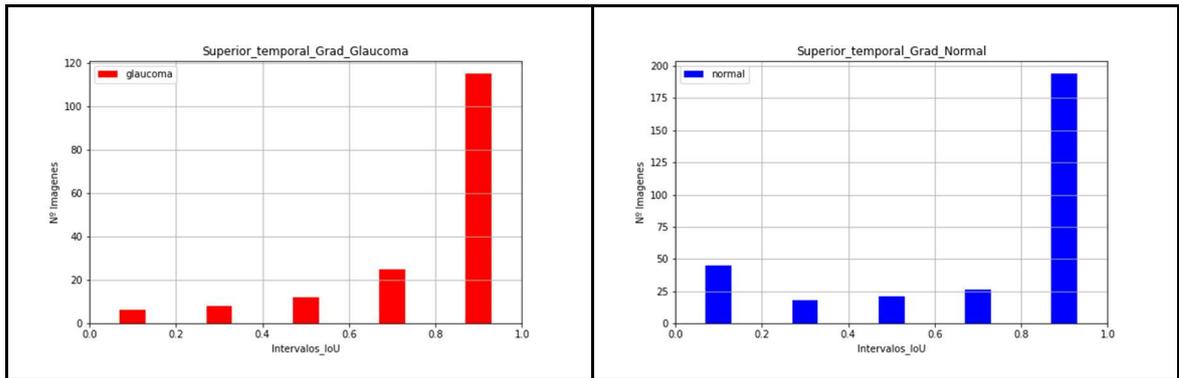


Figura 150. IoU entre mapas Grad-CAM y sector temporal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

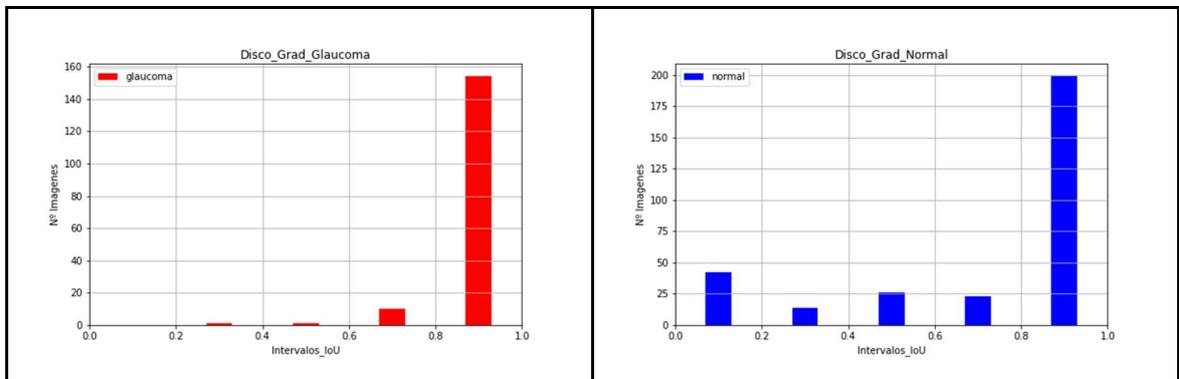


Figura 151. IoU entre mapas Grad-CAM y disco superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

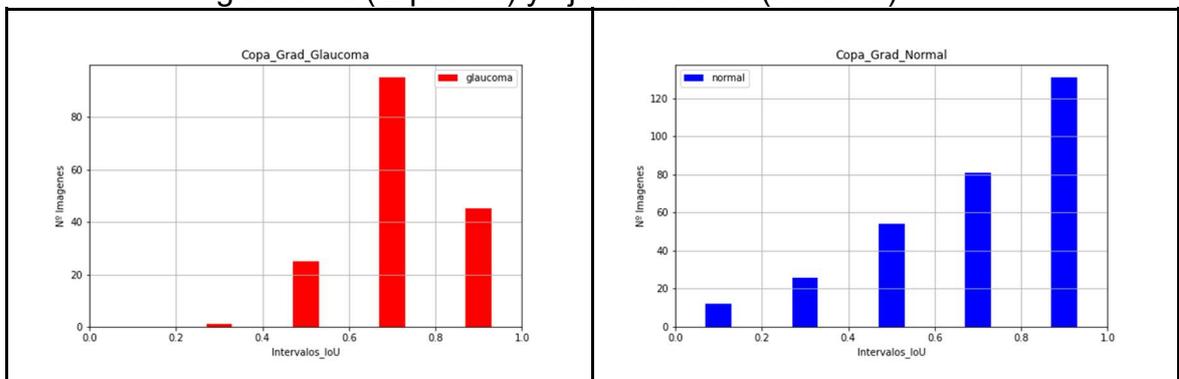


Figura 152. IoU entre mapas Grad-CAM y copa superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

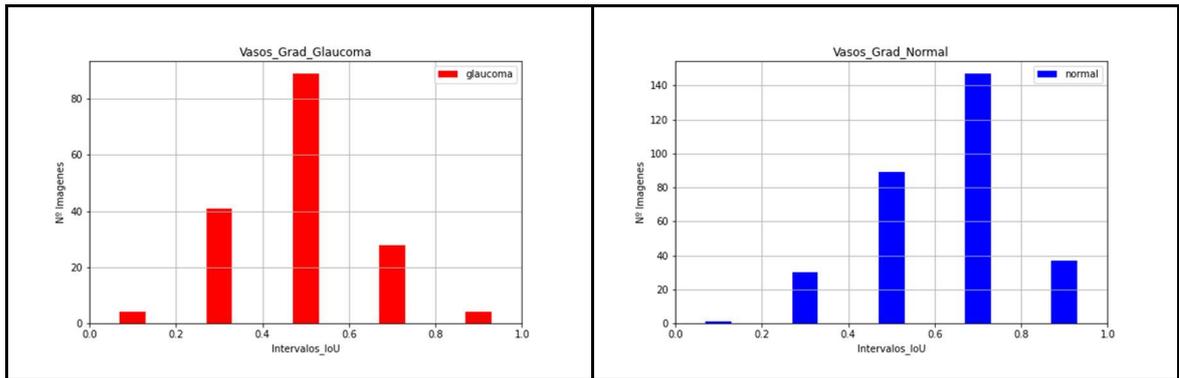


Figura 153. IoU entre mapas Grad-CAM y vasos para ojos con glaucoma (izquierda) y ojos normales (derecha).

Discusión de resultados de Grad-CAM

El sector temporal y el sector nasal parecen ser relevantes en los dos tipos de imágenes, del mismo modo que los sectores superiores e inferiores. Esto quiere decir que los mapas se centran en el disco óptico, fijándose en todos sus sectores. Esto se ve confirmado en los histogramas relativos al disco.

5.2.5.2 Resultados de Score-CAM

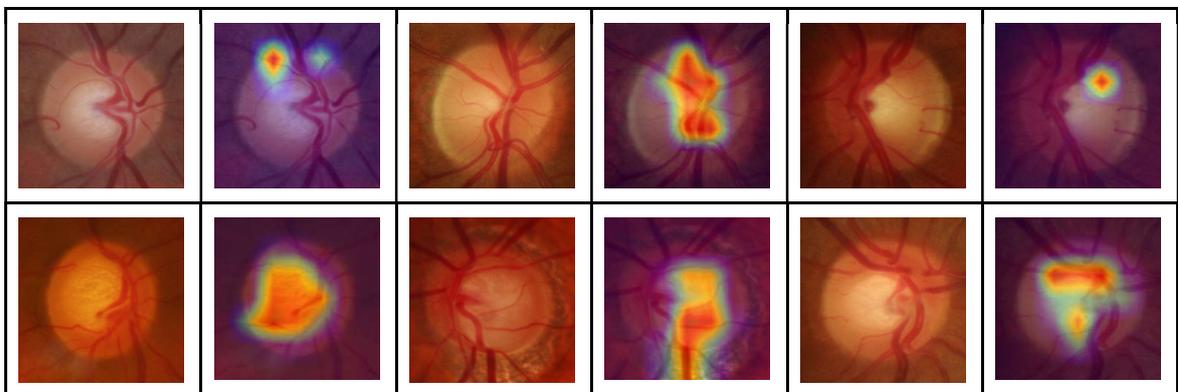


Figura 154. Imágenes de ojos sanos (primera fila) y ojos con glaucoma (segunda fila). Se muestra la imagen original y el mapa Score-CAM obtenido con la red DenseNet.



Histogramas IoU

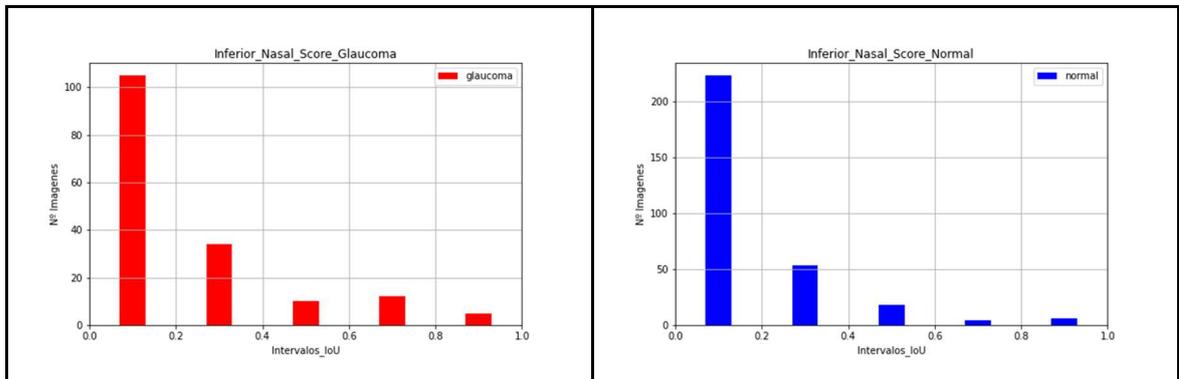


Figura 155. IoU entre mapas Score-CAM y sector nasal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

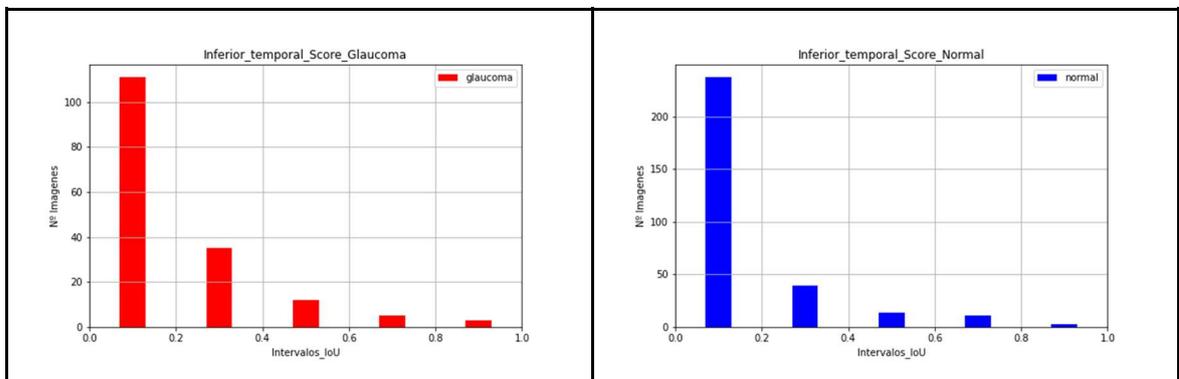


Figura 156. IoU entre mapas Score-CAM y sector temporal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

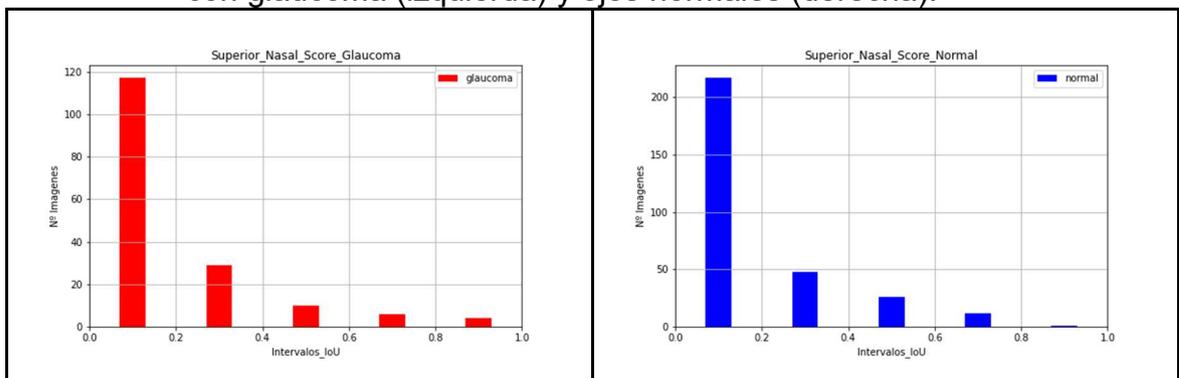


Figura 157. IoU entre mapas Score-CAM y sector nasal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

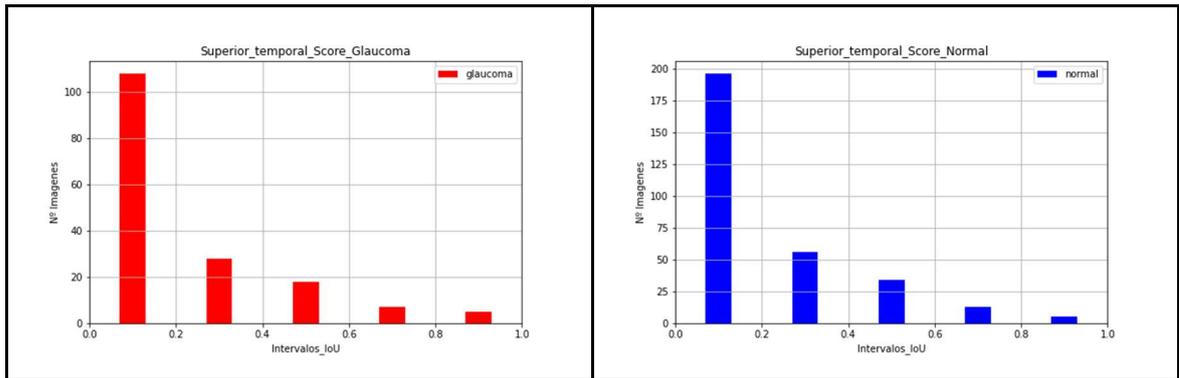


Figura 158. IoU entre mapas Score-CAM y sector temporal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

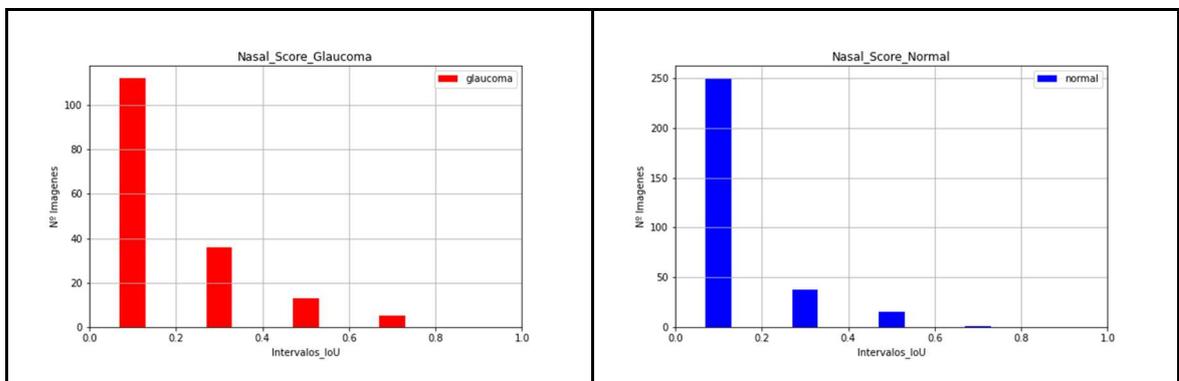


Figura 159. IoU entre mapas Score-CAM y sector nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

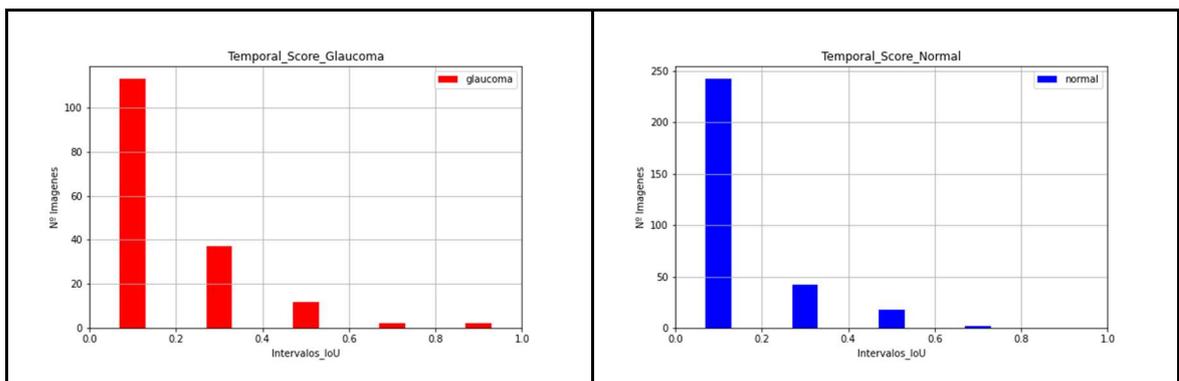


Figura 160. IoU entre mapas Score-CAM y sector temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

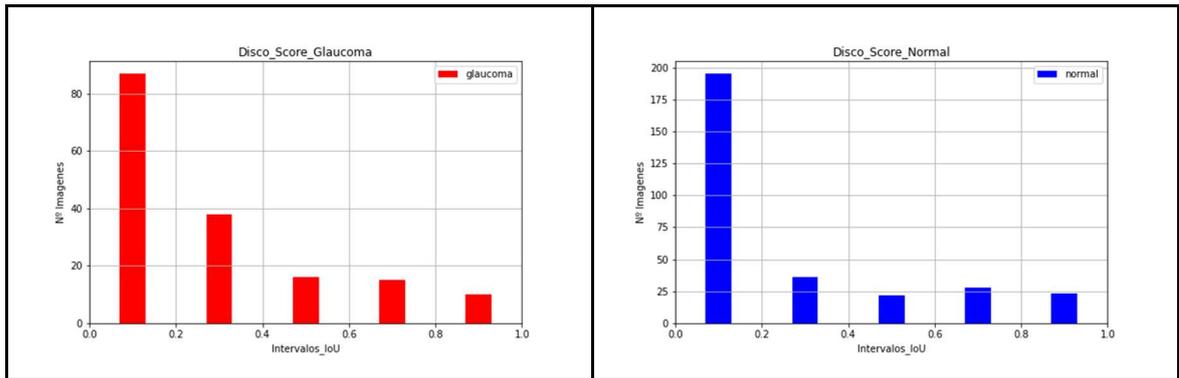


Figura 161. IoU entre mapas Score-CAM y disco para ojos con glaucoma (izquierda) y ojos normales (derecha).

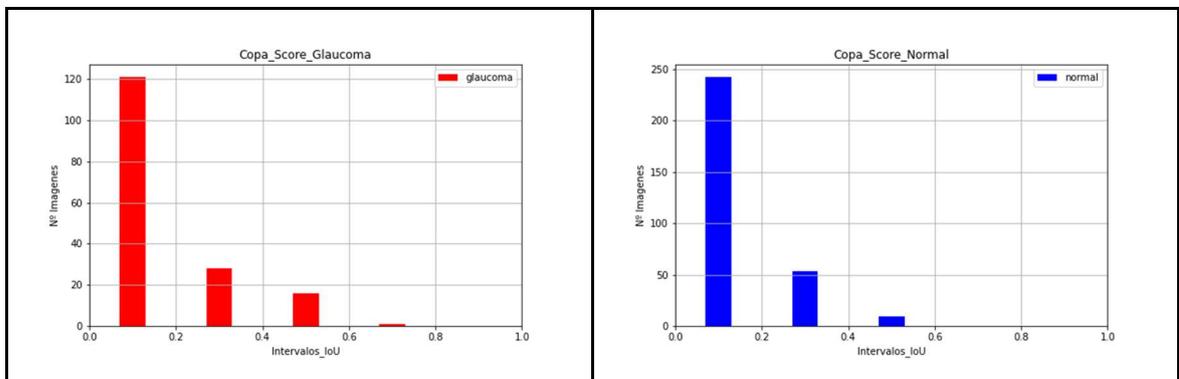


Figura 162. IoU entre mapas Score-CAM y copa para ojos con glaucoma (izquierda) y ojos normales (derecha).

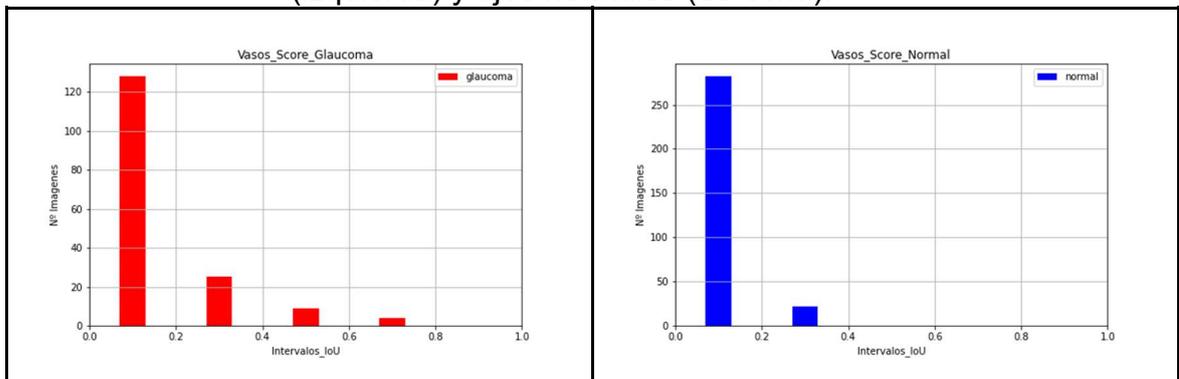


Figura 163. IoU entre mapas Score-CAM y vasos para ojos con glaucoma (izquierda) y ojos normales (derecha).

Discusión de resultados de Score-CAM

Hasta ahora, para todas las redes estudiadas, los mapas Grad-CAM y los mapas Score-CAM eran parecidos. Pero en el caso particular de esta red se puede observar que el Score-Cam se fija en regiones totalmente diferentes a las estudiadas por Grad-CAM.

En los mapas Score-CAM no hay ningún sector relevante, ni para imágenes normales ni para imágenes de glaucoma. Esto quiere decir que los mapas no se



centran en el disco óptico, sino fuera del mismo. Esto se ve confirmado en los histogramas relativos al disco.

5.2.5.3 Resultados de OCL

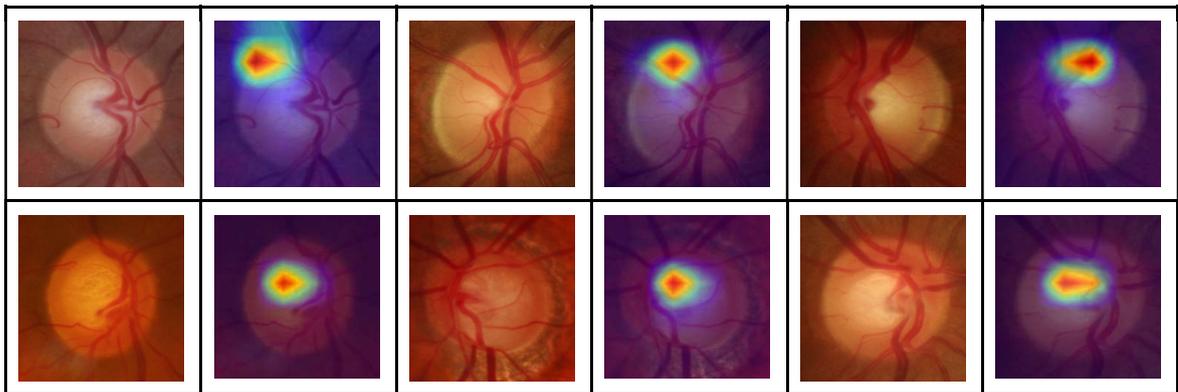


Figura 164. Imágenes de ojos sanos (primera fila) y ojos con glaucoma (segunda fila). Se muestra la imagen original y el mapa Score-CAM obtenido con la red DenseNet.

Histogramas IoU

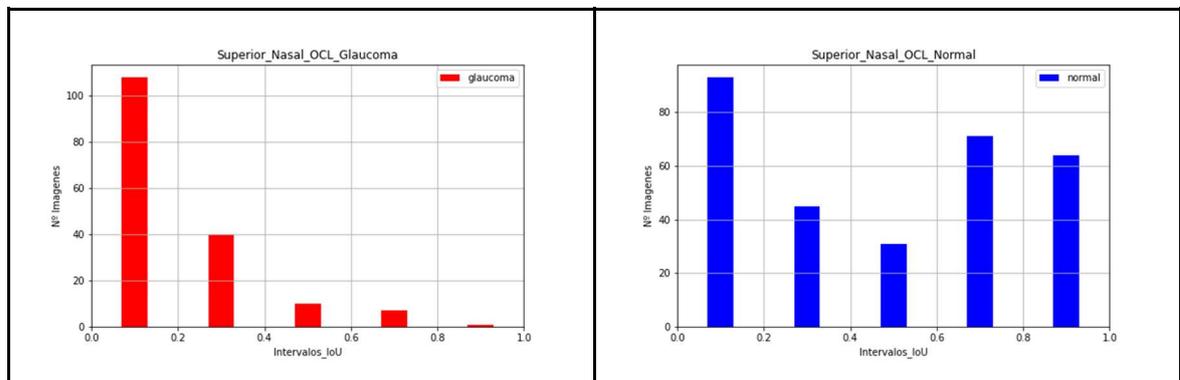


Figura 165. IoU entre mapas OCL y sector nasal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

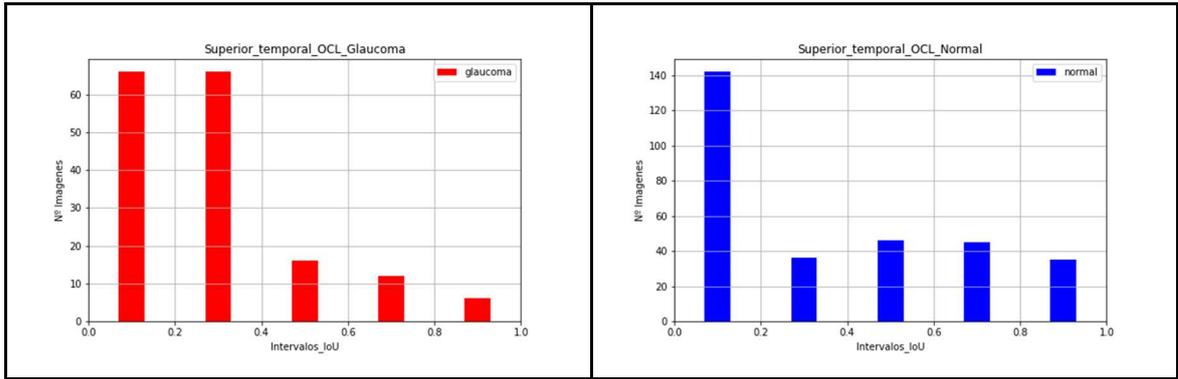


Figura 166. IoU entre mapas OCL y sector temporal superior para ojos con glaucoma (izquierda) y ojos normales (derecha).

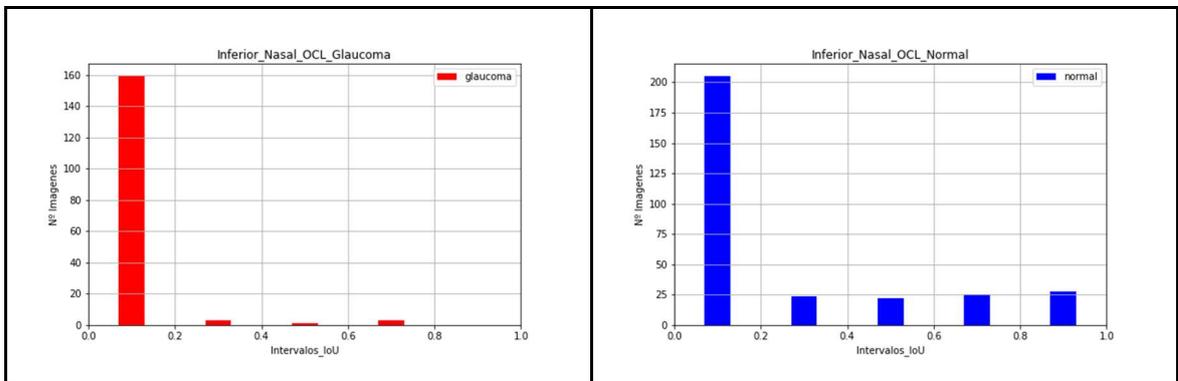


Figura 167. IoU entre mapas OCL y sector nasal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

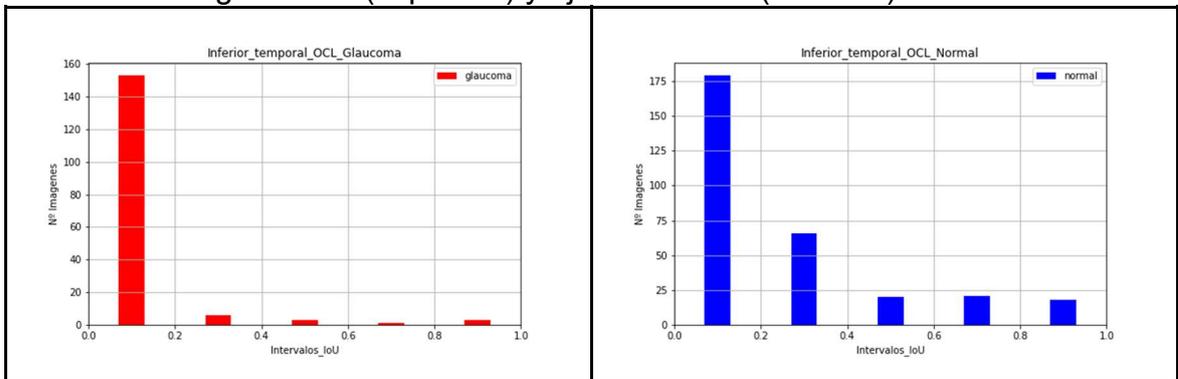


Figura 168. IoU entre mapas OCL y sector temporal inferior para ojos con glaucoma (izquierda) y ojos normales (derecha).

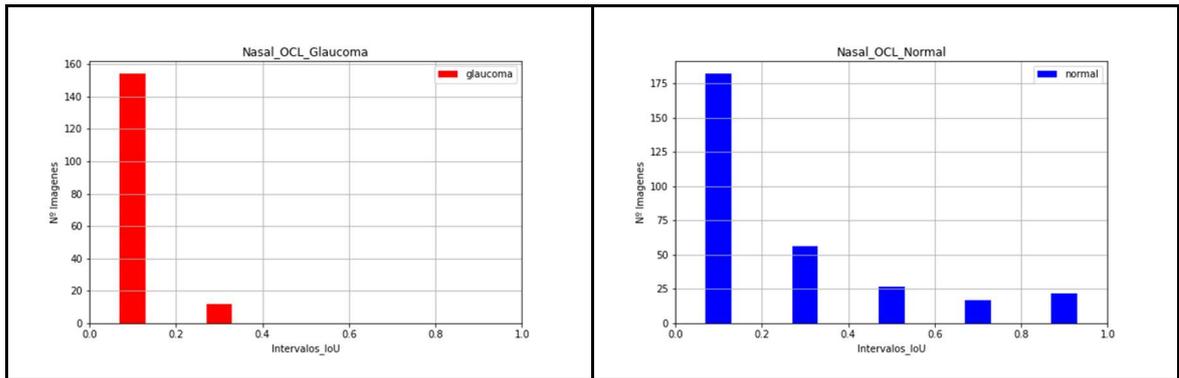


Figura 169. IoU entre mapas OCL y sector nasal para ojos con glaucoma (izquierda) y ojos normales (derecha).

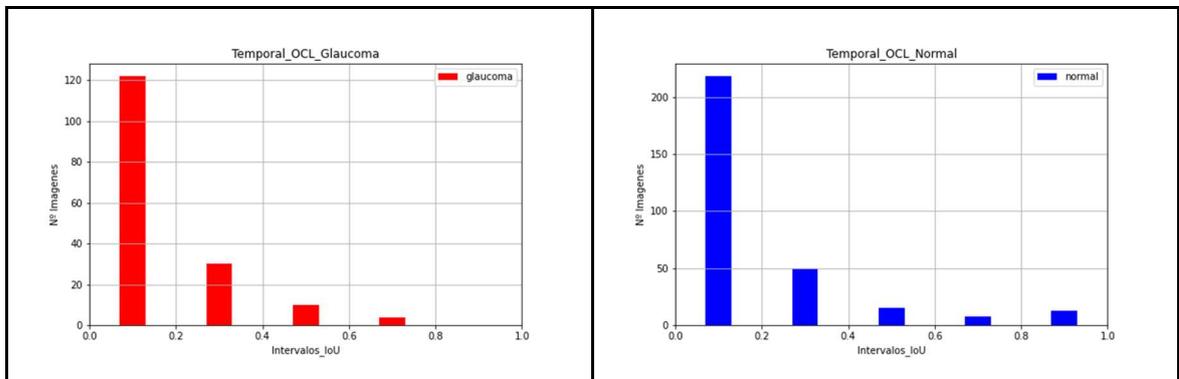


Figura 170. IoU entre mapas OCL y sector temporal para ojos con glaucoma (izquierda) y ojos normales (derecha).

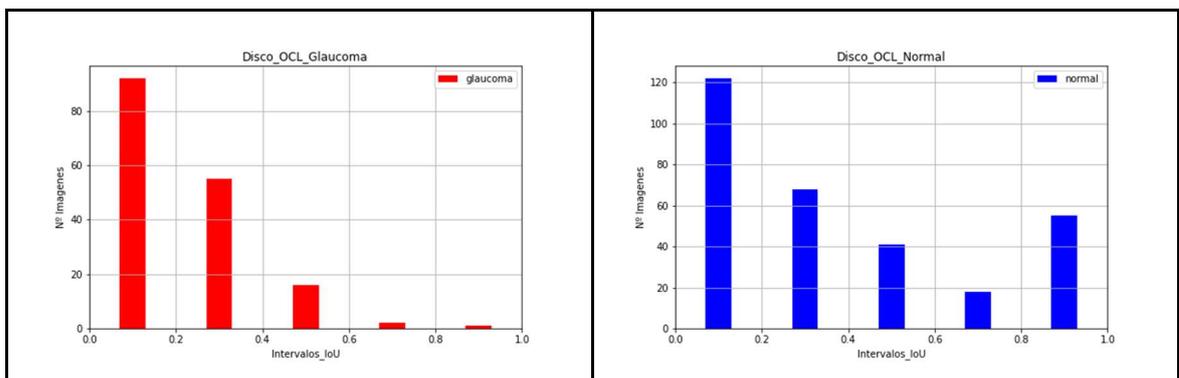


Figura 171. IoU entre mapas OCL y disco para ojos con glaucoma (izquierda) y ojos normales (derecha).

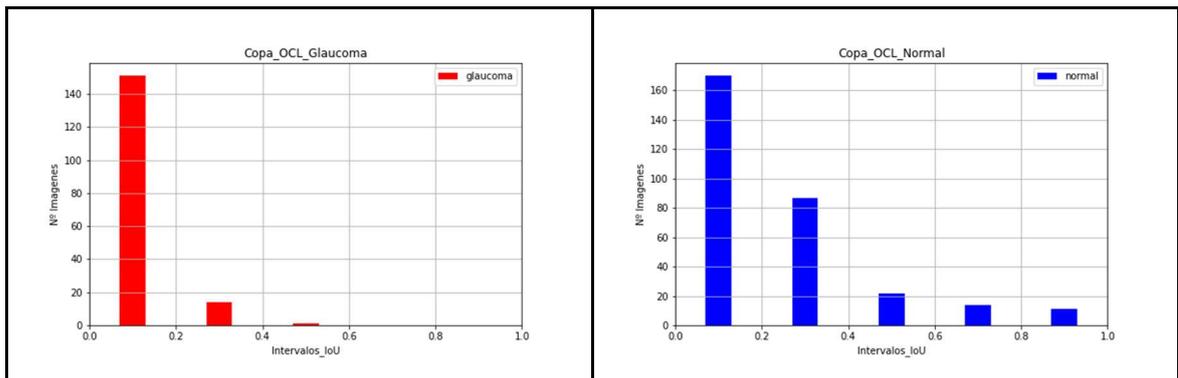


Figura 172. IoU entre mapas OCL y copa para ojos con glaucoma (izquierda) y ojos normales (derecha).

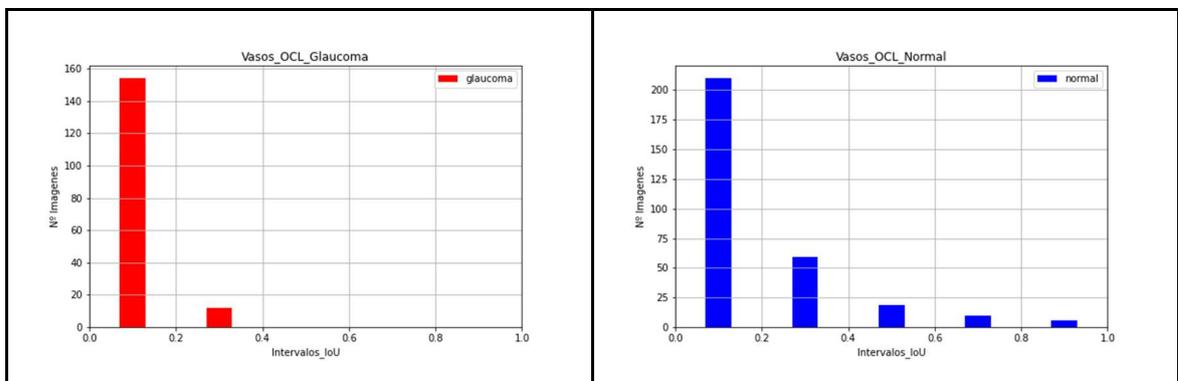


Figura 173. IoU entre mapas OCL y vasos para ojos con glaucoma (izquierda) y ojos normales (derecha).

Discusión de resultados de OCL

Se muestran diferencias entre sectores superiores e inferiores. Las zonas superiores muestran mayor representatividad sobre todo para los casos de glaucoma. El disco parece importante en el caso de imágenes normales.

5.2.5.4 Comparativa de resultados de los tres métodos con la DenseNet

Con esta red no hay un acuerdo claro entre los métodos. Mientras el Grad-Cam genera mapas donde todos los sectores (la totalidad del disco óptico) son muy relevantes, en los mapas Score-CAM no hay ningún sector relevante, ni para imágenes normales ni para imágenes de glaucoma. Esto quiere decir que los mapas no se centran en el disco óptico, sino fuera del mismo. El OCL muestra cierta importancia de los sectores superiores y del disco.

5.3. Segundo estudio: análisis de resultados de todas las redes con una misma técnica de visualización

Se estudiarán los mapas de calor obtenidos con la misma técnica de visualización aplicada sobre distintas redes, sobre el total del conjunto de imágenes de RIM-ONE



DL que está bien clasificado por todas las redes. Se intenta averiguar si existen en general zonas en la imagen que sean de interés común para todas las redes.

Los datos serán expuestos en forma de histogramas. En el eje x se representa el índice IoU entre los mapas generados y una región concreta del ojo, dividido en 5 intervalos. En el eje y, el número de mapas que tienen el IoU comprendido en ese intervalo.

5.3.1 GradCam

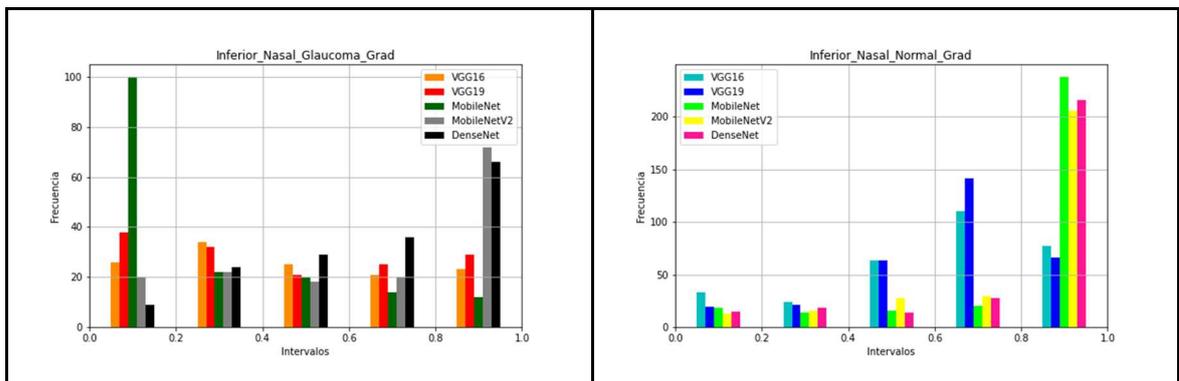


Figura 174. IoU entre mapas Grad-CAM y sector nasal inferior para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

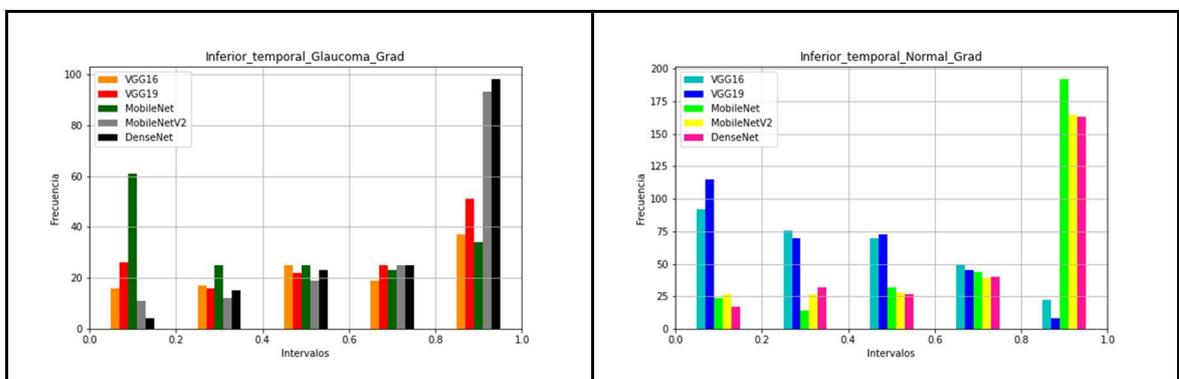


Figura 175 IoU entre mapas Grad-CAM y sector temporal inferior para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

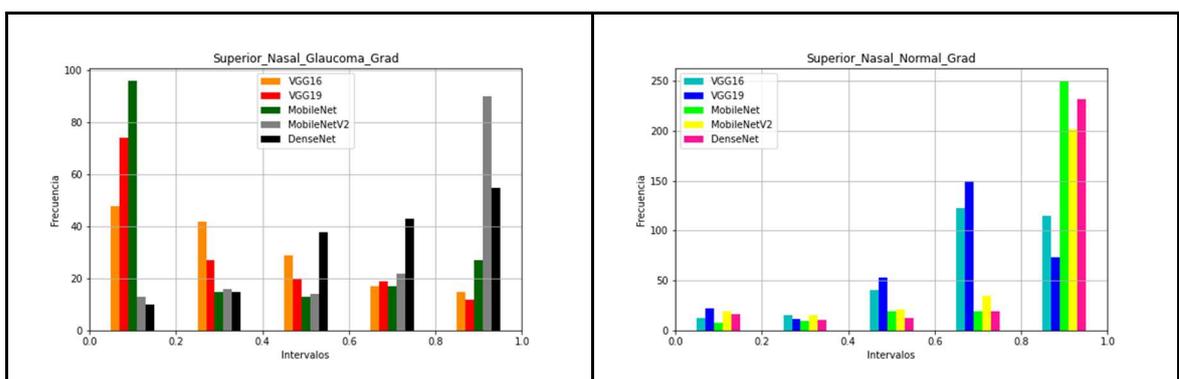




Figura 176. IoU entre mapas Grad-CAM y sector nasal superior para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

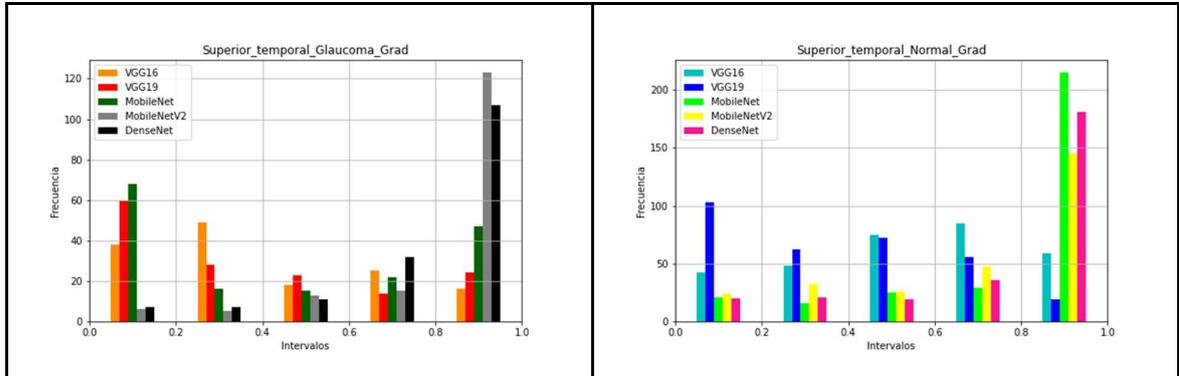


Figura 177. IoU entre mapas Grad-CAM y sector temporal superior para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

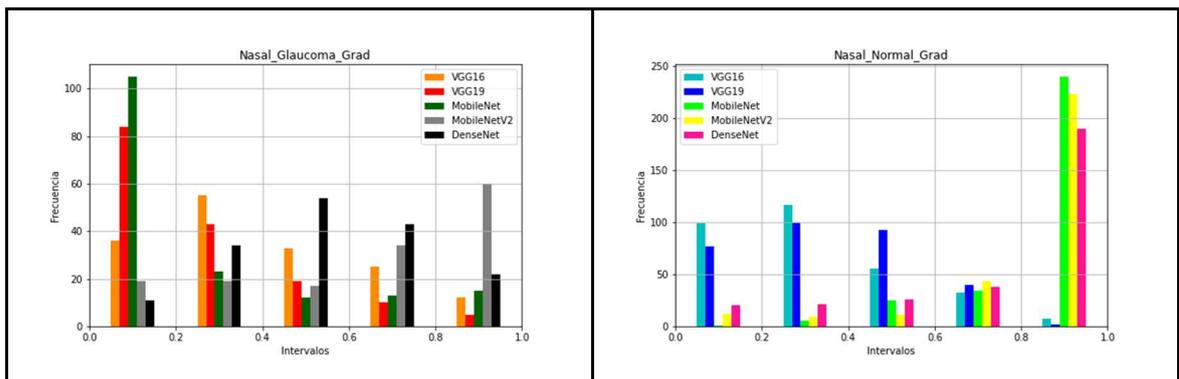


Figura 178. IoU entre mapas Grad-CAM y sector nasal para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

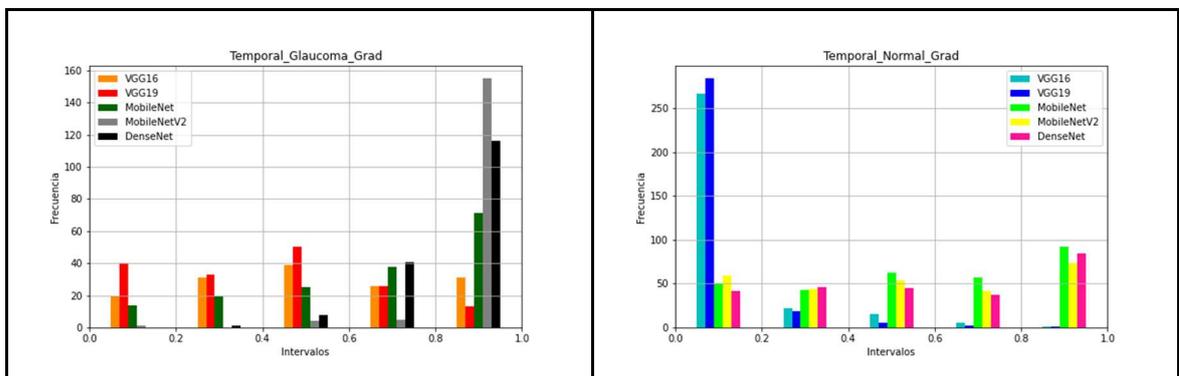


Figura 179. IoU entre mapas Grad-CAM y sector temporal para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

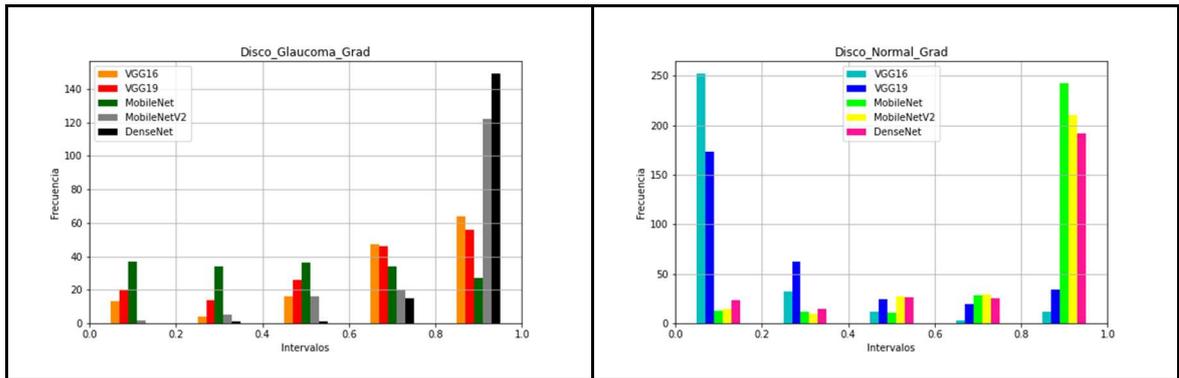


Figura 180. IoU entre mapas Grad-CAM y disco para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

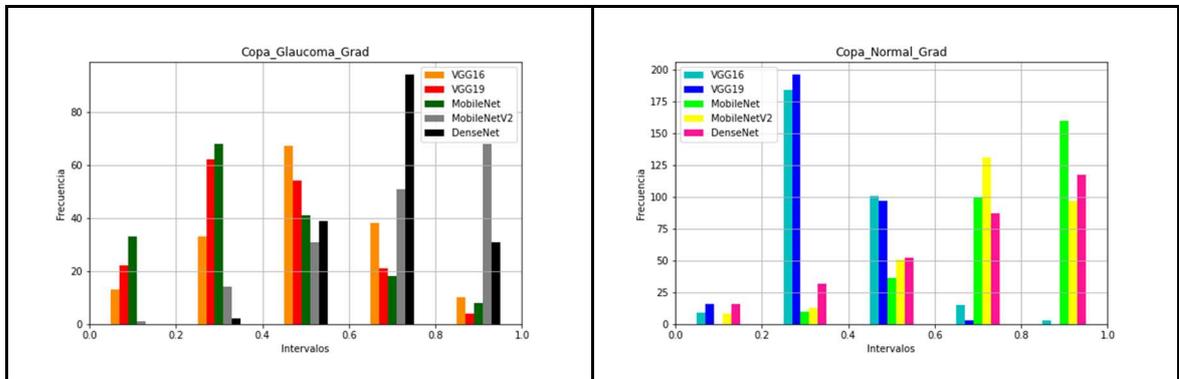


Figura 181. IoU entre mapas Grad-CAM y copa para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

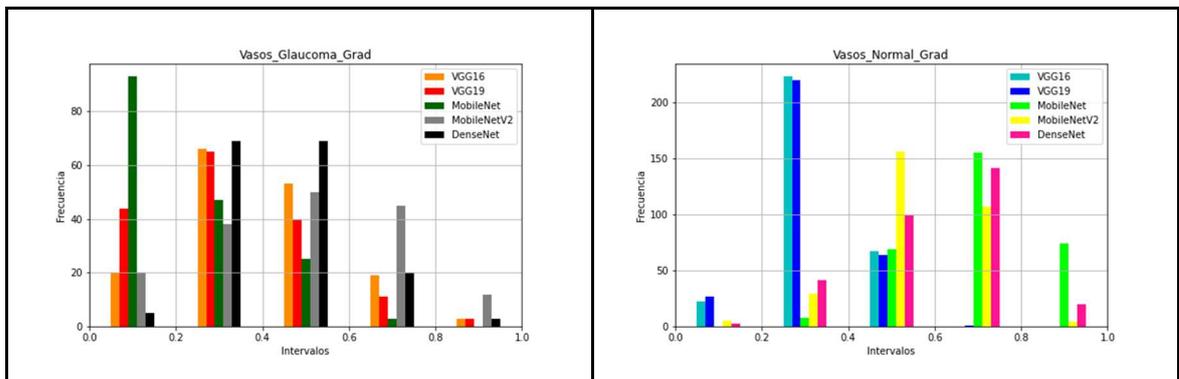


Figura 182. IoU entre mapas Grad-CAM y copa para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

Discusión de resultados de Grad-CAM

A la vista de los resultados, según el método Grad-CAM, las distintas redes no se fijan siempre en las mismas regiones. Las coincidencias son las siguientes:

- Sectores inferiores y sectores superiores: familia MobileNet y DenseNet, para imágenes normales y con glaucoma.
- Sector temporal: familia MobileNet y DenseNet, solo para glaucoma.



- Sector nasal: familia MobileNet y DenseNet, solo para normales.
- Disco: todas (menos MobileNet) en el caso de glaucoma, familia MobileNet y DenseNet, solo para normales.
- Copa: MobileNetV2 y DenseNet, en glaucomas, familia MobileNet y DenseNet, en normales.
- Vasos: ninguna red se fija de manera significativa en los vasos.

En general, parece que la mayoría se fija en el disco, y que los vasos no son relevantes, según Grad-CAM.

5.3.2 Score-CAM

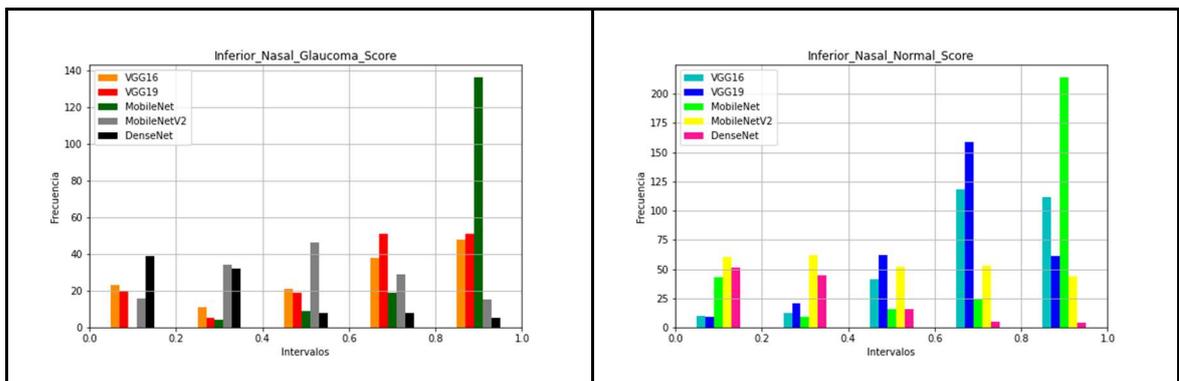


Figura 183. IoU entre mapas Score-CAM y sector nasal inferior para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

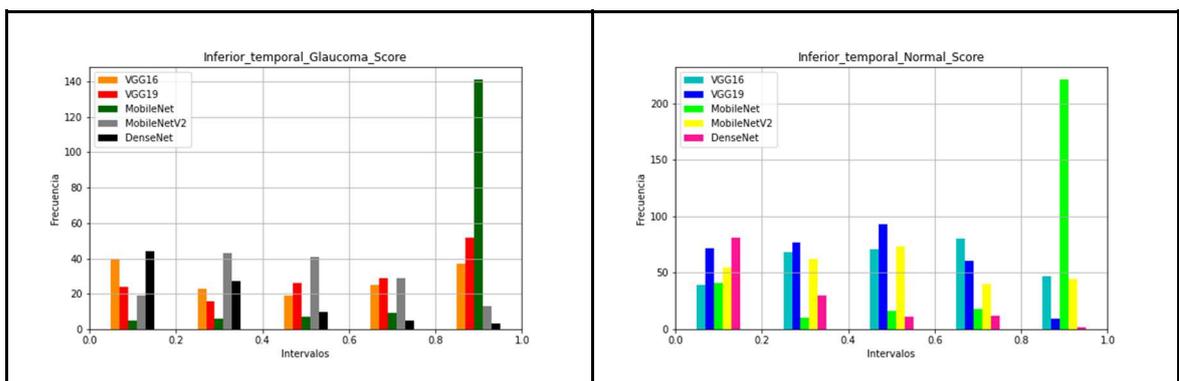


Figura 184. IoU entre mapas Score-CAM y sector temporal inferior para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

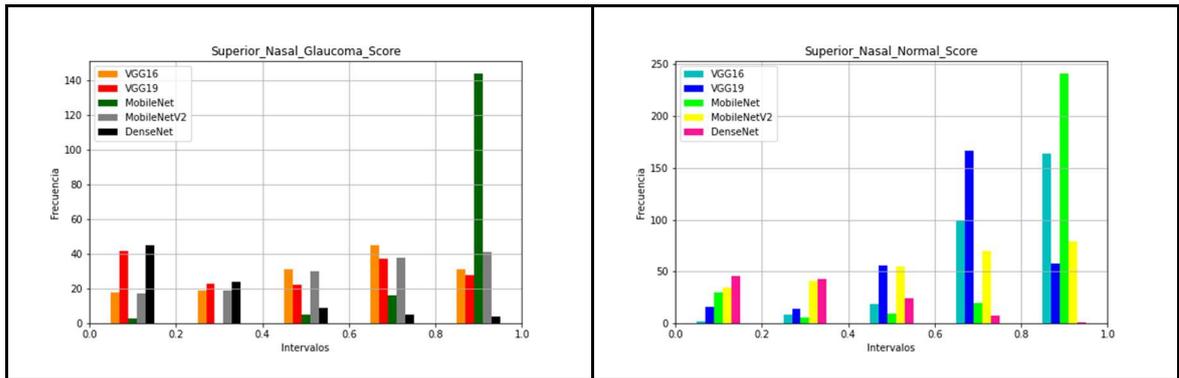


Figura 185. IoU entre mapas Score-CAM y sector nasal superior para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

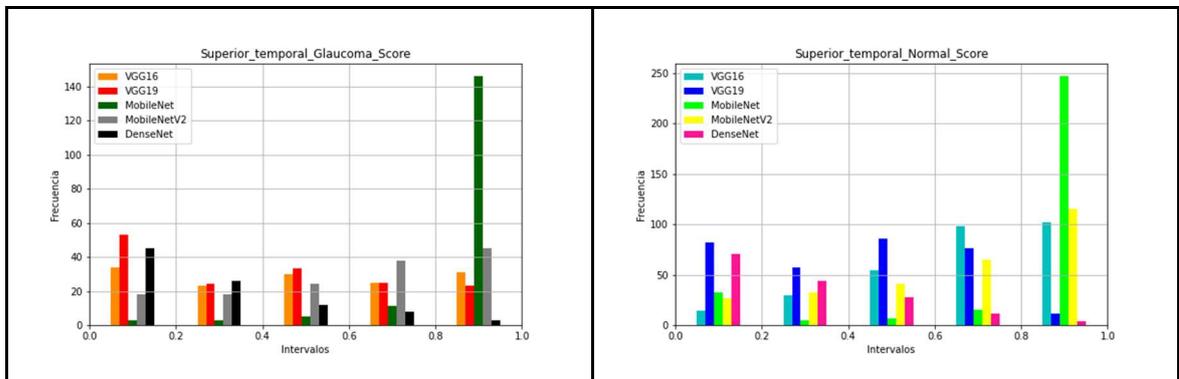


Figura 186. IoU entre mapas Score-CAM y sector temporal superior para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

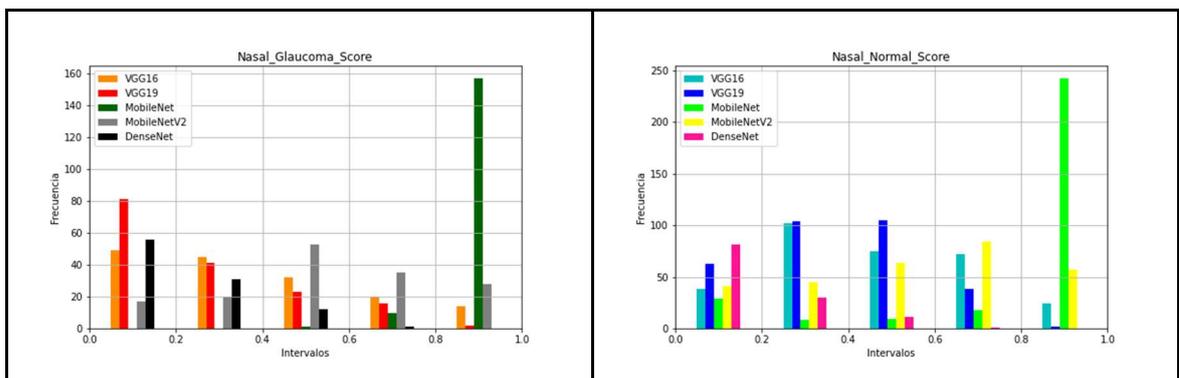


Figura 187. IoU entre mapas Score-CAM y sector nasal para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

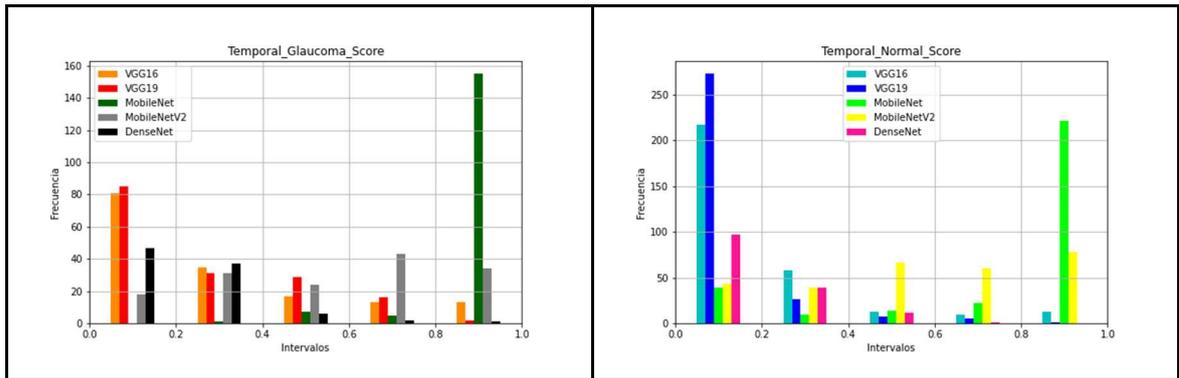


Figura 188. IoU entre mapas Score-CAM y sector temporal para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

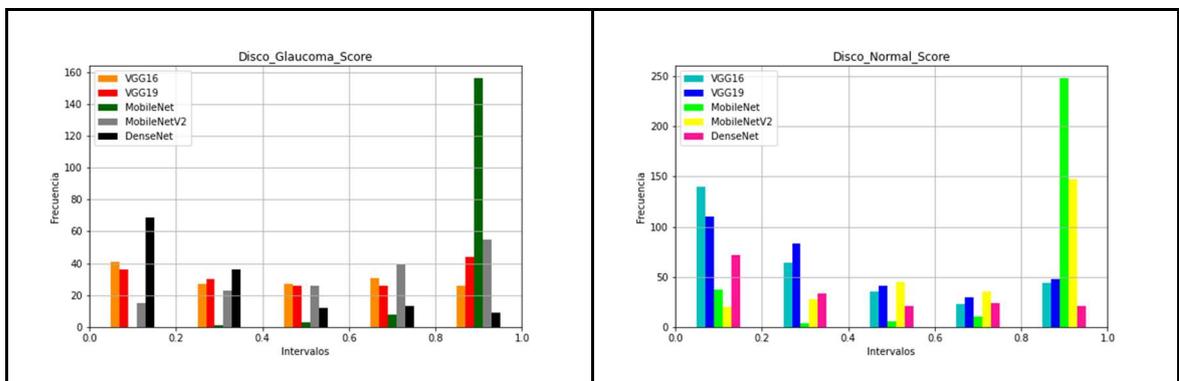


Figura 189. IoU entre mapas Score-CAM y disco para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

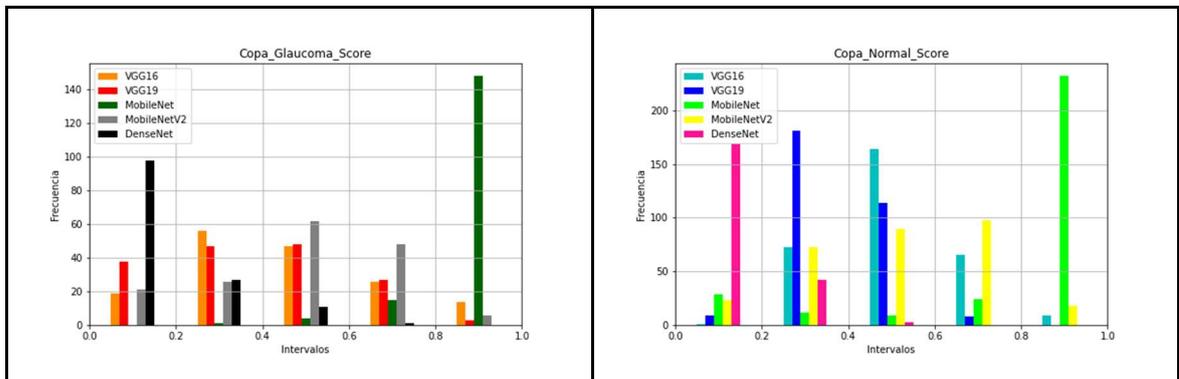


Figura 190. IoU entre mapas Score-CAM y copa para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

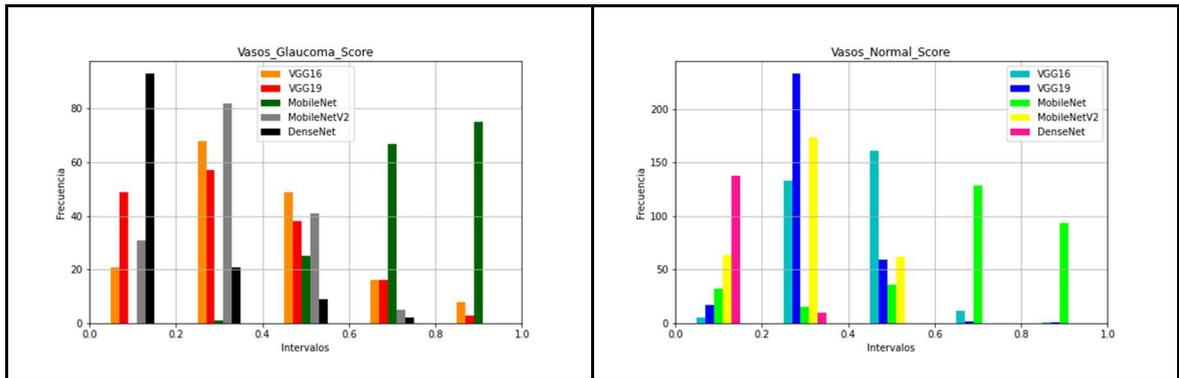


Figura 191. IoU entre mapas Score-CAM y vasos para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

Discusión de resultados de Score-CAM

A la vista de los resultados, según el método Score-CAM, las distintas redes no se fijan siempre en las mismas regiones. Se puede observar que:

- Sectores inferiores y sectores superiores: muy importantes para MobileNet en imágenes normales y con glaucoma. La zona inferior de la imagen parece relevante para la familia VGG en imágenes de glaucoma.
- Sectores temporal y nasal: relevantes para familia MobileNet, en normales y glaucoma. Esto implica que esta red se fija en todos los sectores, es decir, en todo el disco.
- Disco: todas (menos DenseNet) para normales y glaucomas.
- Copa: MobileNetV2 en glaucomas y normales.
- Vasos: MobileNet en glaucomas y normales..

En general, parece que la mayoría se fija en el disco. Las demás regiones son muy dependientes del tipo de red.

5.3.3 OCL

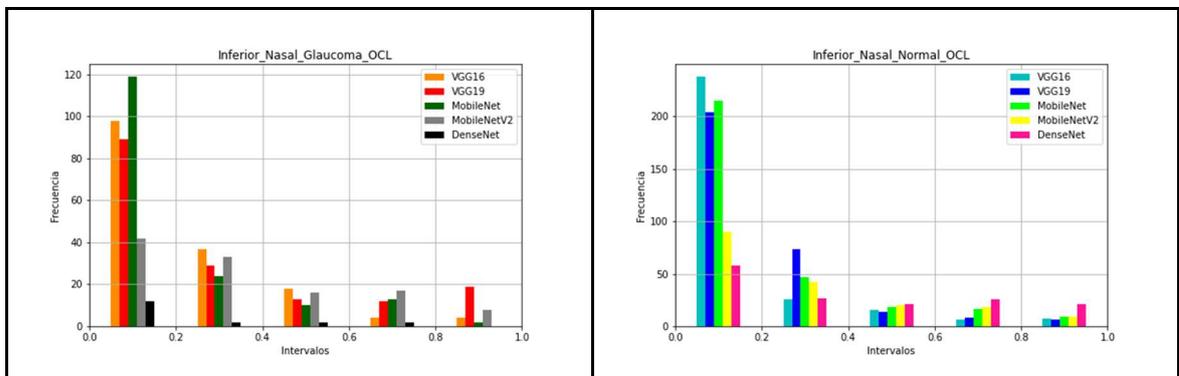


Figura 192 IoU entre mapas OCL y sector nasal inferior para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

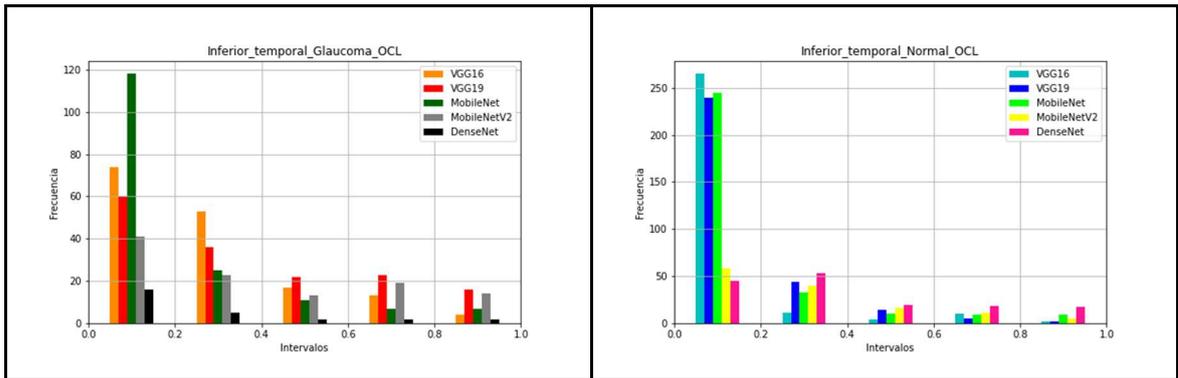


Figura 193. IoU entre mapas OCL y sector temporal inferior para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

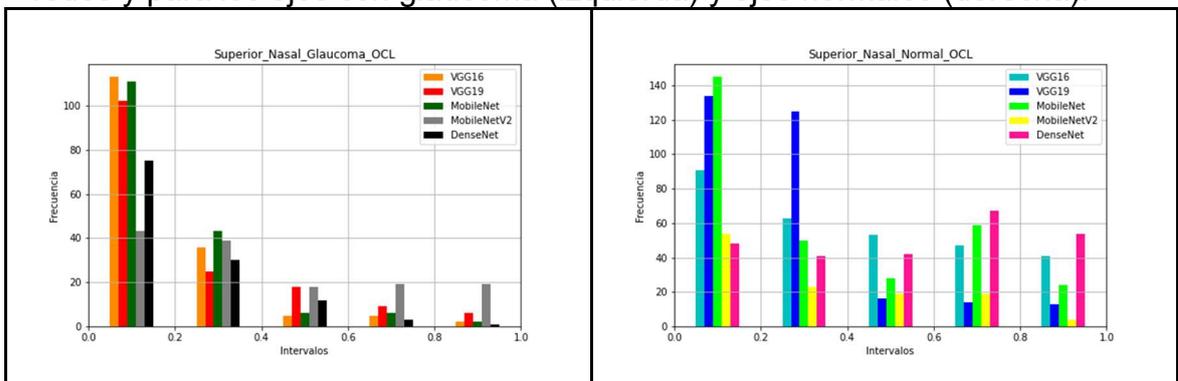


Figura 194. IoU entre mapas OCL y sector nasal superior para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

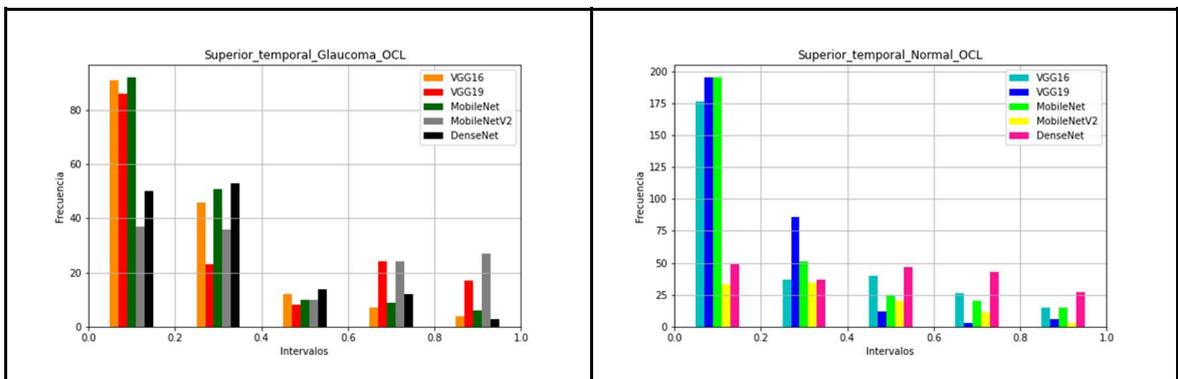


Figura 195. IoU entre mapas OCL y sector temporal superior para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

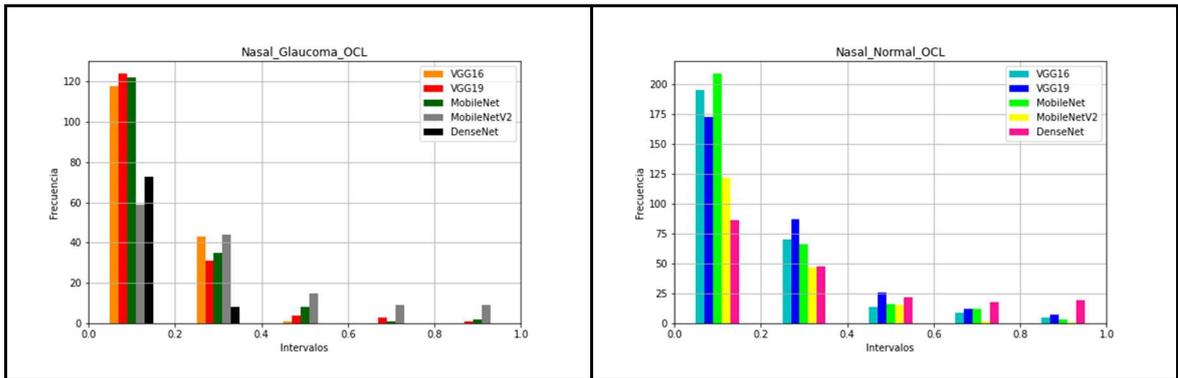


Figura 196. IoU entre mapas OCL y sector nasal para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

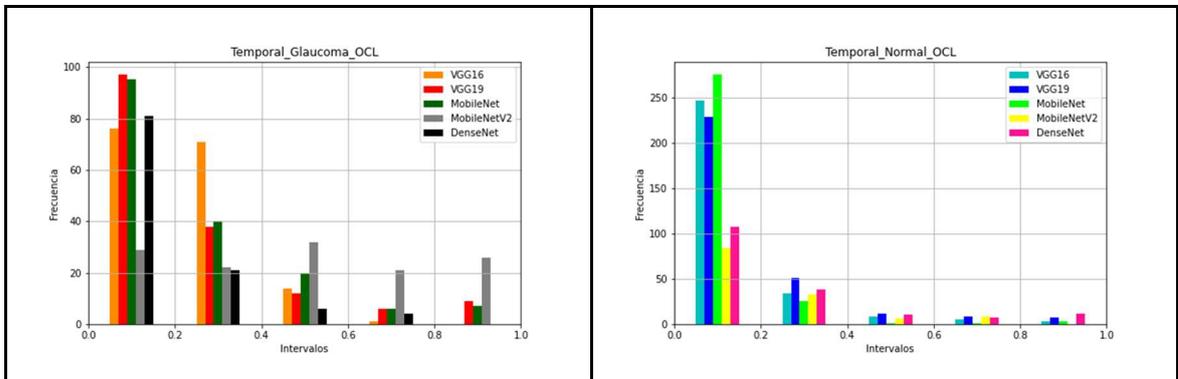


Figura 197. IoU entre mapas OCL y sector temporal para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

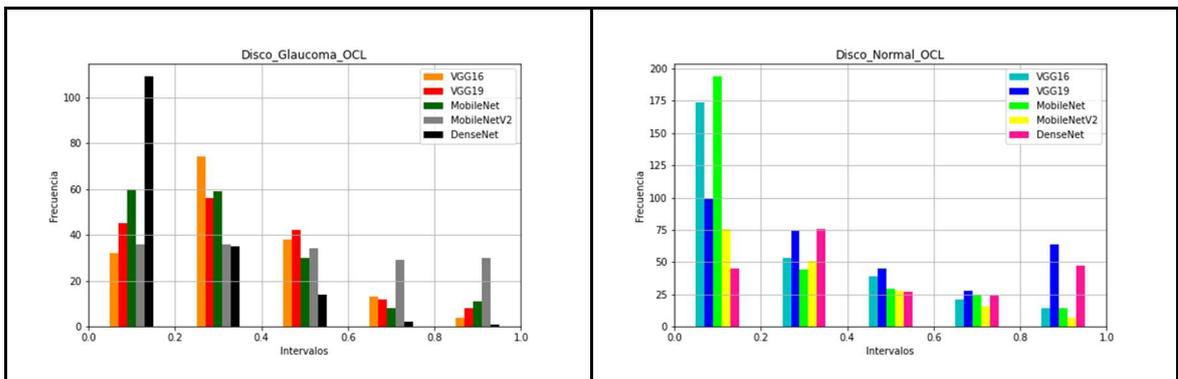


Figura 198. IoU entre mapas OCL y disco para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

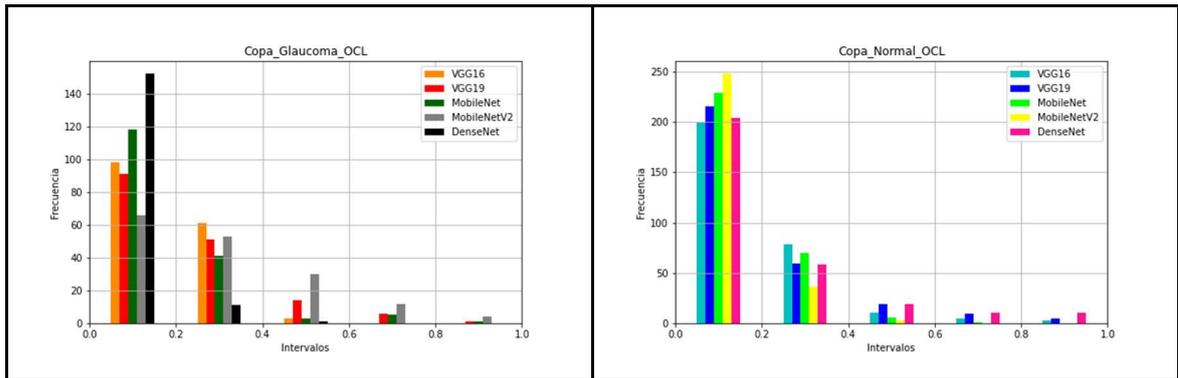


Figura 199. IoU entre mapas OCL y copa para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

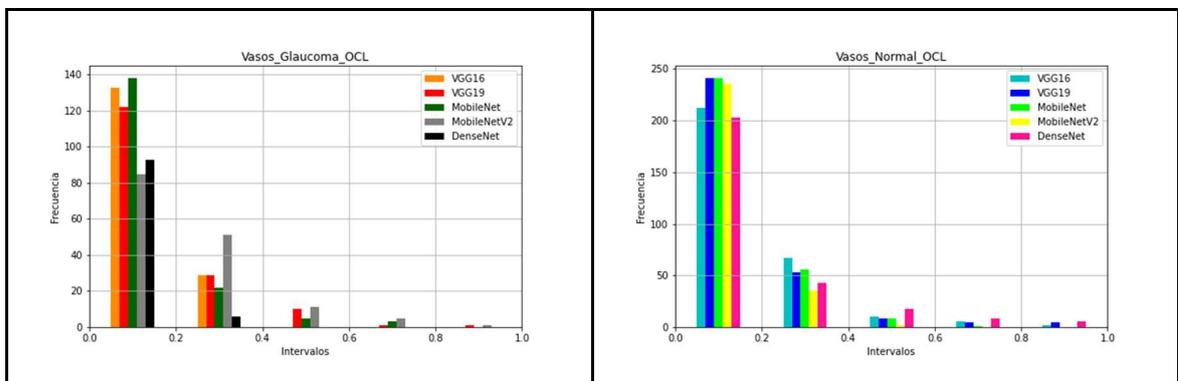


Figura 200. IoU entre mapas OCL y vasos para todas las redes y para los ojos con glaucoma (izquierda) y ojos normales (derecha).

Discusión de resultados de OCL

A la vista de los resultados, según el método OCL, las distintas redes no se fijan siempre en las mismas regiones. Se puede observar que:

- Los sectores superiores son un poco más relevantes que los inferiores para todas las redes, para imágenes normales y con glaucoma.
- Sector nasal poco importante en general para todas las redes.
- Sector temporal más importante para la MobileNetV2 con glaucomas.
- Entre disco, copa y vasos, la región más influyente es el disco, para todas las redes.



6. Conclusiones y líneas futuras

En este Trabajo Fin de Grado (TFG) se han analizado tres métodos de visualización ya existentes en la literatura científica para interpretar el comportamiento de cinco redes DL entrenadas para clasificar ojos sanos y con glaucoma a partir de retinografías. El interés del estudio es intentar identificar qué partes de la imagen son relevantes en la decisión de cada red según cada método de interpretación, y ver si los resultados son coherentes entre métodos, entre redes, y con los criterios seguidos por el médico para diagnosticar.

Analizando los resultados experimentales obtenidos se pueden sacar varias conclusiones. En primer lugar, se ha observado que las redes pertenecientes a la misma familia parece que funcionan de manera parecida. Por ejemplo, la familia VGG muestra resultados muy similares, lo que es esperable ya que la diferencia entre ellas radica en la mayor profundidad de capas de una sobre la otra.

Decidir cuál de las técnicas de visualización es la más idónea no es una tarea fácil. El Grad-CAM ofrece mapas de calor claros, pero es cierto que en las redes con más profundidad (más complejas) los mapas que genera son muy grandes y ocupan todos los sectores del ojo, por lo que no queda claro en qué región se está fijando la red exactamente. Son mapas demasiado gruesos. Esto puede ser debido a que a mayor complejidad (más capas en la red), mayor es la disminución del tamaño de las imágenes que llegan a la capa convolucional final, y por tanto, se pierde mucha resolución espacial. Por este motivo los sectores temporal y nasal aparecen como relevantes en estas redes complejas, mientras que no son importantes para las redes más sencillas.

El Score-CAM, que en la literatura científica está considerado como un método que da resultados parecidos a los del Grad-CAM, e incluso ligeramente mejores, por ofrecer mapas no tan gruesos, presenta dos comportamientos diferentes en este TFG. Por una parte, con las redes sencillas como las de la familia VGG, es cierto que ofrece resultados similares a los del Grad-CAM. En cambio, con el resto de las redes, que son bastante más complejas, los mapas que ofrece son diversos.

El método de oclusión probado en este TFG, el más básico, es muy sencillo de programar y probar. Pero sus resultados son altamente dependientes del tamaño de la ventana, y seguramente, serían diferentes si en vez de usar ventanas “negras” hubiéramos probado otros tipos de oclusión. Con este método es difícil asegurar la identificación de todos los patrones importantes para la red neuronal.

Analizando ahora los resultados desde el punto de vista médico, resulta relevante ver que en muchas técnicas y redes los sectores superiores e inferiores de la imagen parecen zonas importantes para la red. Esto es coherente con el criterio médico. Normalmente, el glaucoma es una enfermedad que se empieza a desarrollar por estos sectores.



En cambio, sorprende que la copa no parece ser relevante en la gran mayoría de los experimentos realizados. La relación copa-disco es un parámetro estudiado por el especialista con cuidado, porque en caso de glaucoma la copa suele crecer mucho. En las pruebas realizadas se observa que el disco es una región muy señalada, pero la copa no.

Se pueden plantear varias líneas futuras que continúen la investigación iniciada en este TFG:

- Con los datos recogidos en este TFG, se podrían diseñar otros estudios comparativos, por ejemplo, para una misma imagen, analizar la respuesta de la misma red con diferentes técnicas de visualización.
- Todos los experimentos realizados han sido sobre imágenes bien clasificadas. Igual podría ser interesante realizar un estudio de las imágenes mal clasificadas por las redes. El objetivo es ver en qué regiones se fija la red cuando se equivoca, para tratar de comprender su error.
- Se podría explotar el conocimiento de que los sectores más importantes son los superiores e inferiores, para diseñar un nuevo método de oclusión.
- Por supuesto, añadir nuevas técnicas de visualización y nuevas redes al estudio y realizar más experimentos.

7. Conclusions and Future lines

In this Final Degree, three visualization methods have been analyzed to interpret the behavior of five DL networks trained to classify healthy eyes and those with glaucoma from retinographics. The interest of the study is to try to identify which parts of the image are relevant in the decision of each network according to each interpretation method, and to see if the results are consistent between methods, between networks, and with the criteria followed by the doctor to diagnose.

The VGG family shows very similar results, which is normal since the difference lies in the greater depth of layers of one over the other.

Deciding which technique is the most suitable for this type of experimentation is not an easy task. Grad Cam has shown good results, although the deeper networks show high representations of all sectors. This may be due to the fact that the higher the complexity of the network, the higher the noise in the heat map obtained by the algorithm. This may be an explanation why the temporal and nasal sectors are more represented in these networks.



The Score-CAM, which in the scientific literature is considered a method that gives results similar to those of the Grad-CAM, and even slightly better, because it offers not so thick maps, presents two different behaviors in this TFG. On the one hand, with simple networks like those of the VGG family, it is true that it offers results similar to those of Grad-CAM. On the other hand, with the rest of the networks, which are much more complex, the maps it offers are diverse.

The simplest method of occlusion tested, is very easy to program and test. But their results are highly dependent on the size of the window, and surely, they would be different if instead of using “black” windows we had tried other types of occlusion. With this method it is difficult to ensure the identification of all the important patterns for the neural network.

Now analyzing the results from the medical point of view, it is relevant to see that in many techniques and networks the upper and lower sectors of the image appear to be important areas for the network. Normally, glaucoma is a disease that begins to develop in these sectors.

On the contrary, it is surprising that the cup does not seem to be relevant in the vast majority of the experiments carried out. The cup-disc relationship is a parameter carefully studied by the specialist, because in the case of glaucoma the cup tends to grow a lot. In the tests carried out, it is observed that the disc is a very marked region, but the cup is not.

There are many lines of future research, for example:

- With the data collected, other comparative studies could be designed, for example, for the same image, to analyze the response of the same network with different visualization techniques.
- All the experiments carried out have been on well classified images. Maybe, it could be interesting to carry out a study of images misclassified by networks. The objective is to see in which regions the network looks when it makes a mistake, to try to understand its error.
- A new method of occlusion could be designed taking into account the most important sectors are the upper and lower.
- Of course, add new visualization techniques and new networks to the study and do more experiments.

8. Bibliografía

[1] “Pandas.” [Online]. Available: <https://pandas.pydata.org/>.



- [2] “DB Imagenet.” [Online]. Available: <https://image-net.org/>.
- [3] “Algoritmo_Score-CAM.” <https://github.com/tabayashi0117/Score-CAM>.
- [4] “OpenCV.” [Online]. Available: <https://opencv.org/>.
- [5] Q. Abbas, “Glaucoma-Deep: Detection of Glaucoma Eye Disease on Retinal Fundus Images using Deep Learning,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 6, 2017, doi: 10.14569/ijacsa.2017.080606.
- [6] S. Bach, A. Binder, G. Montavon, F. Klauschen, K. R. Müller, and W. Samek, “On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation,” *PLoS One*, vol. 10, no. 7, pp. 1–46, 2015, doi: 10.1371/journal.pone.0130140.
- [7] M. Cesareo, A. Anastasio, and C. Nucci, “Tecniche di imaging nel glaucoma,” [Online]. Available: http://www.fgeditore.it/relazione_soi/pdf/capitolo_1/7Tecniche.pdf.
- [8] “VGG16_Photo” <https://ichi.pro/es/arte-con-ia-transferencia-de-estilo-neuronal-109749021858154>
- [9] P. Dabkowski and Y. Gal, “Real time image saliency for black box classifiers,” *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, pp. 6968–6977, 2017.
- [10] A. C. T. Felipe A. Medeiros, Alessandro A. Jammal, “From Machine to Machine: An OCT-Trained Deep Learning Algorithm for Objective Quantification of Glaucomatous Damage in Fundus Photographs,” *Ophthalmology*, vol. 126, no. 4, pp. 513–521, 2019.
- [11] -Matthew D. Zeiler and Rob Fergus, “Algoritmo-Visualizing-and-Understanding-Convolutional-neural-networks.” <https://github.com/saketd403/Visualizing-and-Understanding-Convolutional-neural-networks>.
- [12] F. Fumero, S. Alayon, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, “RIM-ONE: An open retinal image database for optic nerve evaluation,” *Proc. - IEEE Symp. Comput. Med. Syst.*, no. July, 2011, doi: 10.1109/CBMS.2011.5999143.
- [13] Y. M. G. B. J. A. H. I. G. W. J. S. S. R. Garnavi, “Understanding Deep Learning Decision for Glaucoma Detection using 3D Volumes,” *Investigative Ophthalmology & Visual Science*, vol. 61, p. 2022, 2020.
- [14] Google LLC, “Google Colab.” https://colab.research.google.com/notebooks/intro.ipynb#scrollTo=5fCEDCU_qrC0.
- [15] Google LLC, “TensorFlow.” <https://www.tensorflow.org/>.



- [16] M. ul Hassan, "VGG16 – Convolutional Network for Classification and Detection." <https://neurohive.io/en/popular-networks/vgg16/>.
- [17] T. Hirota, M., Mizota, A., Mimura, "Effect of color information on the diagnostic performance of glaucoma in deep learning using few fundus images," *Int Ophthalmol*, vol. 40, pp. 3013–3022, 2020.
- [18] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv*, 2017.
- [19] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 2261–2269, 2017, doi: 10.1109/CVPR.2017.243.
- [20] H. M. Keel S, Wu J, Lee PY, Scheetz J, "Visualizing Deep Learning Models for the Detection of Referable Diabetic Retinopathy and Glaucoma," *JAMA Ophthalmol*, vol. 137, no. 3, pp. 288–292, 2019.
- [21] Keras, "Keras." <https://keras.io/>.
- [22] R. and contributors Kotikalapudi, "Keras-vis," *GitHub*, 2017. <https://github.com/raghakot/keras-vis>.
- [23] R. Medica, D. E. C. Rica, and C. Lxxi, "El glaucoma y su tratamieto farmacológico," *Rev. médica Costa Rica y Centroamérica LXXI*, no. 610, pp. 297–303, 2014.
- [24] P. A. Miceli, W. D. Blair, and M. M. Brown, *Isolating Random and Bias Covariances in Tracks*. 2018.
- [25] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4510–4520, 2018, doi: 10.1109/CVPR.2018.00474.
- [26] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, 2020, doi: 10.1007/s11263-019-01228-7.
- [27] O. Sheremet, "Intersection over union (IoU) calculation for evaluating an image segmentation model," 2020. <https://towardsdatascience.com/intersection-over-union-iou-calculation-for-evaluating-an-image-segmentation-model-8b22e2e84686>.
- [28] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," *34th Int. Conf. Mach. Learn. ICML 2017*, vol. 7, pp. 4844–4866, 2017.



- [29] K. Simonyan, A. Vedaldi, and A. Zisserman, “Deep inside convolutional networks: Visualising image classification models and saliency maps,” *2nd Int. Conf. Learn. Represent. ICLR 2014 - Work. Track Proc.*, pp. 1–8, 2014.
- [30] J. Sivaswamy, S. R. Krishnadas, and A. Chakravarty, “A Comprehensive Retinal Image Dataset for the Assessment of Glaucoma from the Optic Nerve Head Analysis,” *JSM Biomed Imaging Data Pap 2(1) 1004*, vol. 2, pp. 1–7, 2015.
- [31] M. Sundararajan, A. Taly, and Q. Yan, “Axiomatic attribution for deep networks,” *34th Int. Conf. Mach. Learn. ICML 2017*, vol. 7, pp. 5109–5118, 2017.
- [32] D. S. W. Ting *et al.*, “Artificial intelligence and deep learning in ophthalmology,” *Br. J. Ophthalmol.*, vol. 103, no. 2, pp. 167–175, 2019, doi: 10.1136/bjophthalmol-2018-313173.
- [33] J. Torres, “Deep Learning Introducción practica con keras.” <https://torres.ai/escritor/>.
- [34] R. C. Fong and A. Vedaldi, “Interpretable Explanations of Black Boxes by Meaningful Perturbation. (arXiv:1704.03296v3 [cs.CV] UPDATED),” pp. 3429–3437, [Online]. Available: <http://arxiv.org/abs/1704.03296>.
- [35] D. Vaghjiani, S. Saha, Y. Connan, S. Frost, and Y. Kanagasingam, “Visualizing and Understanding Inherent Image Features in CNN-based Glaucoma Detection,” *2020 Digit. Image Comput. Tech. Appl. DICTA 2020*, pp. 1–3, 2020, doi: 10.1109/DICTA51227.2020.9363369.
- [36] H. Wang *et al.*, “Score-CAM: Score-weighted visual explanations for convolutional neural networks,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2020-June, pp. 111–119, 2020, doi: 10.1109/CVPRW50498.2020.00020.
- [37] M. H. Zhixi Li, Yifan He, Stuart Keel, Wei Meng, Robert T. Chang, “Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs,” *Ophthalmology*, vol. 125, no. 8, pp. 1199–1206.
- [39] “RIM-ONE databes.” <https://github.com/seva100/optic-nerve-cnn>.