



**Escuela Superior
de Ingeniería y Tecnología**
Universidad de La Laguna

Trabajo de Fin de Grado

Grado en Ingeniería Informática

Deep Learning para el diagnóstico de glaucoma en imágenes de fondo de ojo

Deep Learning to diagnose glaucoma in eye fundus images

Ángel Luis Morales Hernández

La Laguna, 5 de julio de 2019

D. **José Francisco Sigut Saavedra**, con N.I.F. 43.786.043-T profesor Titular de Universidad adscrito al Departamento de Ingeniería Informática y de Sistemas de la Universidad de La Laguna, como tutor

D. **Valentín Tinguaro Díaz Alemán**, con N.I.F. 54.043.074-C Médico Adjunto de Servicio de Oftalmología del Hospital Universitario de Canarias , como cotutor

C E R T I F I C A (N)

Que la presente memoria titulada:

“Deep Learning para el diagnóstico de glaucoma en imágenes de fondo de ojo”

ha sido realizada bajo su dirección por D. **Ángel Luis Morales Hernández**, con N.I.F. 54.109.201-E.

Y para que así conste, en cumplimiento de la legislación vigente y a los efectos oportunos firman la presente en La Laguna a 5 de julio de 2019

Agradecimientos

En primer lugar, agradecer a mi tutor D. José Francisco Sigut Saavedra por saber orientarme en este proyecto, por todo el apoyo y la ayuda dada.

También a mi co-tutor D. Valentín Tinguaro Díaz Alemán, por sus conocimientos médicos y su positivismo.

En segundo lugar, agradecer a mi pareja, el cual me ha animado siempre a continuar, apoyándome en los momentos más difíciles.

Por último, pero no menos importante a mi familia y amigos, ya que siempre estaban presentes cuando se les necesitaba y me ayudaban a desconectar y pasar un buen momento.

Licencia



© Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional.

Resumen

Las redes neuronales artificiales se aplican cada vez más en multitud de ámbitos sobre todo a partir del gran avance producido en los últimos años con el aprendizaje profundo (Deep learning) y las redes neuronales convolucionales para el procesamiento de imágenes. Uno de estos ámbitos es la medicina, mostrándose como herramientas de diagnóstico muy potentes aprendiendo a partir de conjuntos de imágenes de diferentes patologías. Sin embargo, un requisito imprescindible para la utilización generalizada y rutinaria de este tipo de sistemas en este ámbito es su capacidad para explicar las decisiones tomadas. Es por ello que este trabajo se ha planteado como una primera aproximación a este punto crucial a través del estudio de algunas de las técnicas de visualización e interpretación más estándar aplicadas al diagnóstico del glaucoma con el objetivo de ver hasta qué punto son capaces de arrojar luz sobre el comportamiento de la red neuronal.

Palabras clave: Red Neuronal Convolucional, Glaucoma, Grad-CAM, OCT, Retinografías, Imágenes de fondo de ojo

Abstract

Artificial neural networks are applied more and more in a multitude of areas, especially since the great advance produced in recent years with deep learning and convolutional neural networks for image processing. One of these areas is medicine, showing itself as very powerful diagnostic tools learning from sets of images of different pathologies. However, an essential requirement for the widespread and routine use of this type of systems in this area is its ability to explain the decisions taken. That is why this work has been proposed as a first approach to this crucial point through the study of some of the most standard visualization and interpretation techniques applied to the diagnosis of glaucoma in order to see to what extent they are capable of shedding light about the behavior of the neural network.

Keywords: CNN, Glaucoma, OCT, Grad-CAM, Retinography, Eye Fundus Image.

Índice general

Capítulo 1: Introducción.....	1
1.1 Objetivos propuestos.....	1
1.2 Antecedentes.....	1
1.3 Fases.....	1
Capítulo 2: Estado del arte.....	2
Capítulo 3: Desarrollo del proyecto.....	7
3.1 Materiales utilizados.....	7
3.1.1 Bases de datos de imágenes.....	7
3.1.2 Tecnologías utilizadas.....	8
3.2 Descripción de los métodos de visualización.....	9
3.2.1 Visualización de las activaciones de las neuronas	9
3.2.2 Visualización de los filtros de convolución.....	9
3.2.3 Visualización basada en Gradient Backpropagation	10
3.2.4 Visualización basada en Mapas de Activación de Clases	11
3.2.5 Visualización basada en oclusión	12
3.3 Resultados y discusión.....	12
3.3.1 Visualización por activaciones intermedias	12
3.3.2 Visualización por filtros.....	17
3.3.3 Visualización por Guided Backpropagation	21
3.3.4 Visualización por Grad-CAM	22
3.3.5 Visualización por oclusión	26
Capítulo 4: Conclusiones y líneas futuras.....	29
Capítulo 5: Summary and Conclusions.....	30
Capítulo 6: Presupuesto	31
6.1 Presupuesto personal	31

Índice de figuras

Figura 2.1: CAM de un ojo sano y otro glaucomatoso en la fila de arriba se representa un ojo glaucomatoso y en la de debajo a uno sano (T: Temporal, N:Nasal, S: Superior, I: Inferior)	2
Figura 2.2: Red Neuronal localizando las zonas sospechosas correctamente	3
Figura 2.3: Representación de los heatmaps creado por Grad-CAM.....	4
Figura 2.4: Mapas de calor de dos casos glaucomatosos y otro sano.....	5
Figura 2.5: Mapa de calor de la visualización por oclusión.....	6
Figura 3.1: Imagen en infrarrojo y color del nervio óptico de un ojo sano	8
Figura 3.2: Imagen en infrarrojo y color del nervio óptico de un ojo glaucomatoso	8
Figura 3.3: Arquitectura de una VGG16.....	9
Figura 3.4: Esquema de visualización de activaciones de las neuronas de las capas más altas.....	11
Figura 3.5: Activación intermedia de imagen con glaucoma a color block1_conv1.....	12
Figura 3.6: Activación intermedia de imagen con glaucoma a color block3_conv2.....	13
Figura 3.7: Activación intermedia de imagen con glaucoma a color block5_conv3.....	14
Figura 3.8: Activación intermedia de imagen sin glaucoma a color block1_conv1.....	15
Figura 3.9: Activación intermedia de imagen sin glaucoma a color block3_conv2.....	15
Figura 3.10: Activación intermedia de imagen sin glaucoma a color block5_conv3.....	16
Figura 3.11: Filtros imagen en color block1_conv1.....	17
Figura 3.12: Filtros imagen en color block3_conv2.....	18
Figura 3.13: Filtros imagen en color block5_conv3.....	18
Figura 3.14: Filtros imagen OCT block1_conv1.....	19

Figura 3.15: Filtros imagen OCT block3_conv2.....	20
Figura 3.16: Filtros imagen OCT block5_conv3.....	20
Figura 3.17: Guided Backpropagation de una imagen a color con glaucoma.....	21
Figura 3.18: Guided Backpropagation de una imagen a color sin glaucoma.....	21
Figura 3.19: Guided Backpropagation de una imagen en OCT con glaucoma.....	22
Figura 3.20: Guided Backpropagation de una imagen en OCT sin glaucoma.....	22
Figura 3.21: Grad-CAM de una OCT con glaucoma.....	23
Figura 3.22: Grad-CAM de una OCT sin glaucoma.....	23
Figura 3.23: Grad-CAM de una retinografía a color con glaucoma.....	24
Figura 3.24: Grad-CAM de una retinografía a color sin glaucoma.....	24
Figura 3.25: Mapa de calor (mayor de 75%) e imagen binaria.....	25
Figura 3.26: Imagen de fondo de ojo con glaucoma y partes segmentadas.....	25
Figura 3.27: Visualización por Oclusión de imagen con glaucoma y ventana cuadrada de 12.....	27
Figura 3.28: Visualización por Oclusión de imagen con glaucoma y ventana cuadrada de 16.....	27
Figura 3.29: Visualización por Oclusión de imagen con glaucoma y ventana cuadrada de 20.....	28

Índice de tablas

Tabla 3.1: Promedio de solape en imágenes normales y glaucomatosas.....26

Tabla 6.1: Presupuesto.....31

Capítulo 1: Introducción

1.1 Antecedentes

Una de las patologías más graves que pueden afectar al ojo es el glaucoma, la cual puede provocar ceguera irreversible si no se consigue tratar a tiempo. Esta enfermedad tiene un diagnóstico bastante complicado, más aún en sus etapas más tempranas. Existen técnicas diagnósticas como la Tomografía de Coherencia Óptica (OCT) que resultan muy eficaces y precisas, pero son muy costosas por lo que se hace necesario disponer de otro tipo de técnicas más asequibles como las basadas en retinografías que son, simplemente, fotos en color de la retina. Existen diferentes características de la retina que pueden servir como indicadores de un posible glaucoma, sobre todo en la cabeza del nervio óptico o papila. Estas características pueden ser extraídas según el criterio de un oftalmólogo experto o a través de un sistema automático que se encargue de ello. Es aquí donde entran en juego las redes neuronales convolucionales y el Deep learning. Si se dispone del número de datos suficientes, en este caso imágenes de ojos normales y glaucomatosos, podemos entrenar este tipo de redes para obtener un diagnóstico. A pesar de los buenos resultados que se han logrado con este tipo de sistemas, una asignatura pendiente que queda por resolver es su carácter de caja negra en el sentido de que no sabemos cómo se ha llegado a dicho diagnóstico, y al ser una aplicación médica es de vital importancia este hecho si se pretende utilizarla en este ámbito.

1.2 Objetivos propuestos

Este trabajo se plantea, principalmente, como una primera aproximación al análisis de técnicas de visualización e interpretación del comportamiento de las redes neuronales artificiales aplicadas al diagnóstico del glaucoma. Para ello se han probado algunas de las técnicas estándar más conocidas y utilizadas sobre una de las redes neuronales convolucionales más populares, la VGG16. El objetivo que se persigue es comprobar hasta qué punto estas técnicas permiten explicar las decisiones tomadas por la red y si existe coincidencia con lo que el médico especialista haría en cada caso. Además, dado que se han utilizado imágenes de diferente naturaleza, en color e infrarrojo, parece interesante ver, también, si existen diferencias entre unas y otras en lo que se refiere al uso de estas técnicas.

1.3 Fases

Para cumplir con los objetivos propuestos hizo falta dividir el trabajo en diferentes tareas a realizar:

- 1) Revisión de técnicas de visualización estándar relacionadas con el problema a tratar.
- 2) Partiendo de redes neuronales entrenadas con retinografías e imágenes de OCT, aplicar las técnicas seleccionadas y generar resultados.
- 3) Análisis de los resultados obtenidos y comparación entre las imágenes de infrarrojo y color.
- 4) Discusión con el médico y conclusiones.

Capítulo 2: Estado del arte

Aunque las técnicas de visualización e interpretación del comportamiento de las redes neuronales convolucionales se empiezan a usar cada vez con más frecuencia en todo tipo de problemas, en el caso concreto del diagnóstico basado en imágenes de fondo de ojo, su utilización es todavía bastante limitada. A continuación, se muestran algunos ejemplos de trabajos relacionados en los que se hace referencia a este tipo de técnicas, pero en ningún caso como el objetivo principal de los mismos sino de forma más bien anecdótica, sobre unas pocas imágenes, para justificar la coincidencia con lo predicho por el médico especialista.

El primer caso de todos es el de Stefan Maetschke et al, en el artículo “A feature agnostic approach for glaucoma detection in OCT volumes”, publicado en 2019 [1]. En dicho artículo se habla de cómo, para determinar si una persona padece glaucoma, se utilizan siempre imágenes segmentadas o partes oculares específicas. Por ello, se propuso que la red neuronal trabaje directamente con los volúmenes de datos en OCT sin modificar, usando una CNN de tres dimensiones. Se comparó la precisión de este procedimiento con otros algoritmos de machine learning basados en características, demostrando así su superioridad sobre los mismos a pesar de que debido a limitaciones de hardware, se tuvo que utilizar una red con solo cinco capas y filtros pequeños.

Como técnica de visualización se utilizó CAM (Class Activation Maps), como se muestra en la *Figura 2.1*.

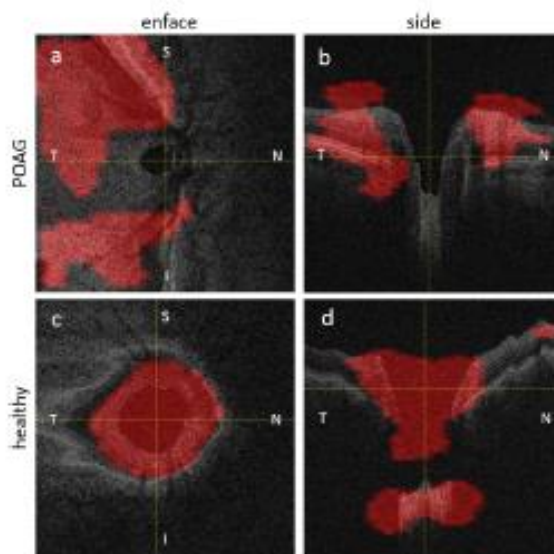


Figura 2.1: CAM de un ojo sano y otro glaucomatoso en la fila de arriba se representa un ojo glaucomatoso y en la de debajo a uno sano (T: Temporal, N:Nasal, S: Superior, I: Inferior).Extraído de [1]

El segundo caso es el de M. Kim et al, en el artículo “Computer-Aided Diagnosis and Localization of Glaucoma Using Deep Learning” [2], en este estudio el objetivo principal fue investigar si era viable crear una red neuronal convolucional de extremo a extremo, para el diagnóstico y la

localización del glaucoma, teniendo acceso solo a un número limitado de imágenes médicas sin segmentar. Se utilizaron imágenes de fondo de ojo en color (retinografía completa) y una red neuronal VGG16 y otra VGG19, modificando su estructura para aprovechar un modelo pre-entrenado de ImageNet.

Los resultados calculados sobre la base de datos de imágenes del Samsung Medical Center en Seul fueron prometedores. Para la visualización de la red se usó la técnica Grad-CAM como se muestra en la *Figura 2.2*.

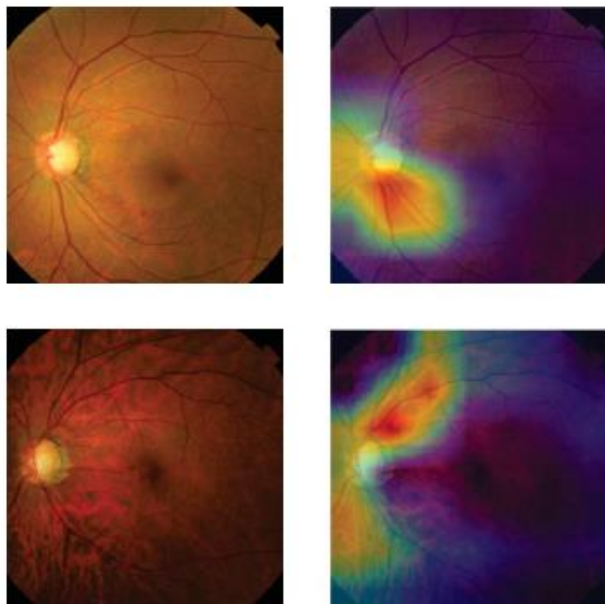


Figura 2.2: Red Neuronal localizando las zonas sospechosas correctamente. Extraído de [2].

El tercer caso es de S. Phan et al, en el estudio “Evaluation of Deep convolutional networks for glaucoma detection” [3], en este artículo, se utilizaron tres redes neuronales convolucionales y tres datasets de imágenes de fondo de ojo a color, para diferenciar si un paciente padece glaucoma o no. Los modelos de las redes utilizadas fueron ResNet, VGG y DenseNet.

El método de visualización que se usó fue CAM, como se muestra en la *Figura 2-3*. A partir de los resultados de este método, se demostró que el área del disco óptico sigue siendo una parte fundamental en el diagnóstico del glaucoma, aunque se comprobó que la calidad de las imágenes afectaba a la capacidad de discriminación.

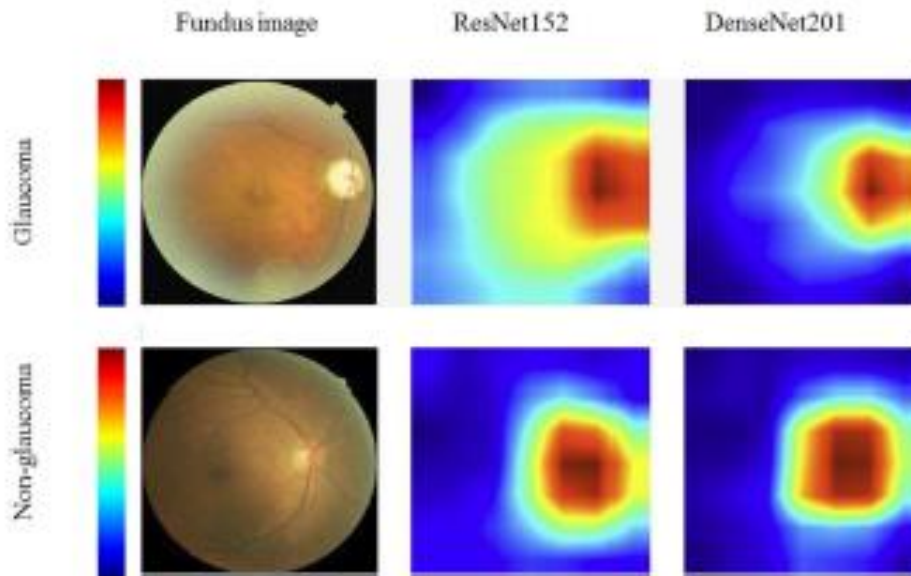


Figura 2.3: Representación de los heatmaps creado por Grad-CAM. Extraído de [3].

El cuarto caso es de G. An et al, en el artículo “Glaucoma Diagnosis with Machine Learning Based on Optical Coherence Tomography and Color Fundus Images” [4], en este estudio, se utilizaron imágenes de color y OCT para alimentar la entrada de la red neuronal, con el objetivo de diferenciar un ojo glaucomatoso de uno completamente sano. Para cumplir dicho cometido, se usó una red neuronal VGG19, utilizando un modelo de aprendizaje transferido (transfer learning) a partir de una red entrenada con ImageNet. El método de visualización utilizado fue Grad-CAM. Como se puede observar en los mapas de calor de la *Figura 2.4*, la primera fila es una representación a escala de grises de una imagen de fondo de ojo, la segunda es un mapa de grosor de la capa de la fibra nerviosa de la retina y la tercera es otro mapa de grosor de la célula ganglionar de la mácula. Como conclusión, el experimento dio resultados satisfactorios ya que podría usarse potencialmente para detectar un estado temprano de la enfermedad.

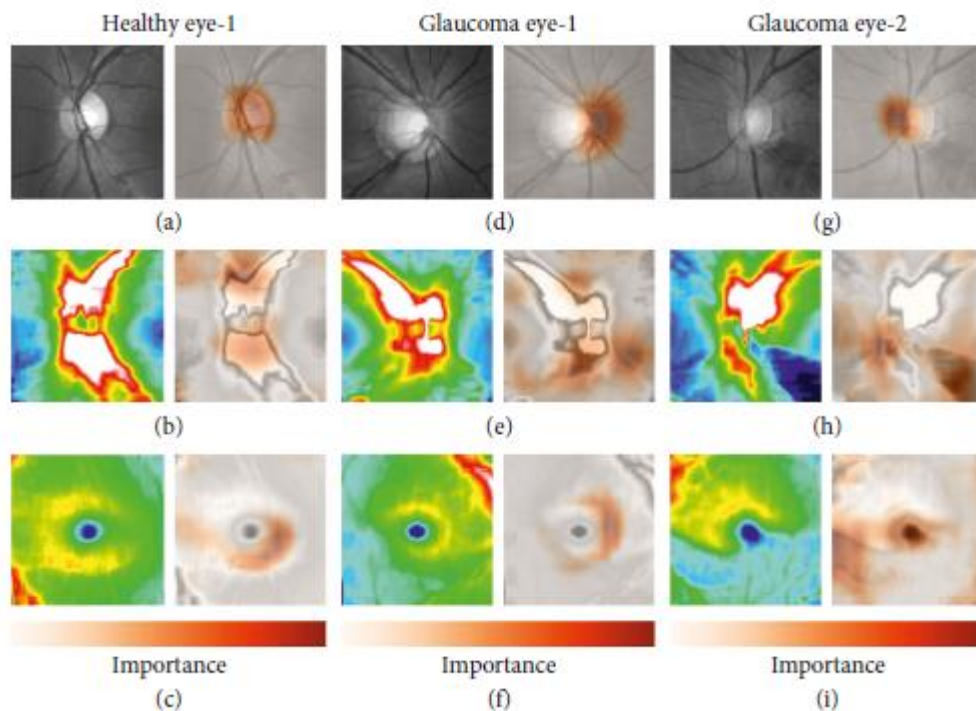


Figura 2.4: Mapas de calor de dos casos glaucomatosos y otro sano. Extraído de [4]

En este último caso, es el de M. Christopher et al, en el artículo “Performance of Deep Learning Architectures and Transfer Learning for Detecting Glaucomatous Optic Neuropathy in Fundus Photographs” [5], en el cual, el estudio se centró en detectar glaucoma utilizando imágenes de fondo de ojo en pacientes que son clínica y demográficamente diferentes, diferenciándolos por la raza, la edad y el sexo. Referente a las redes neuronales, se usaron tres modelos diferentes de estas: una VGG16, una Inception v3 y una ResNet50, y por cada una de ellas se evaluaron dos versiones: una nativa y la otra de aprendizaje transferido (Transfer Learning).

Para el desarrollo de la investigación se utilizaron imágenes recopiladas a partir de dos estudios diferentes: el primero de ellos fue el ADAGES (The African Descent and Glaucoma Evaluation Study) y el segundo fue el DIGS (Diagnostic Innovation in Glaucoma Study). Se consiguió detectar y localizar el glaucoma con una gran precisión en la mayoría de los casos, siendo el modelo de aprendizaje transferido con la arquitectura ResNet50 el que mejores resultados obtuvo.

En este caso, la técnica de visualización utilizada fue la de oclusión para cuantificar el impacto de cada región en la predicción del modelo, dando como resultado un mapa de calor con las zonas más importantes de la imagen, como se muestra en la *Figura 2.5*.

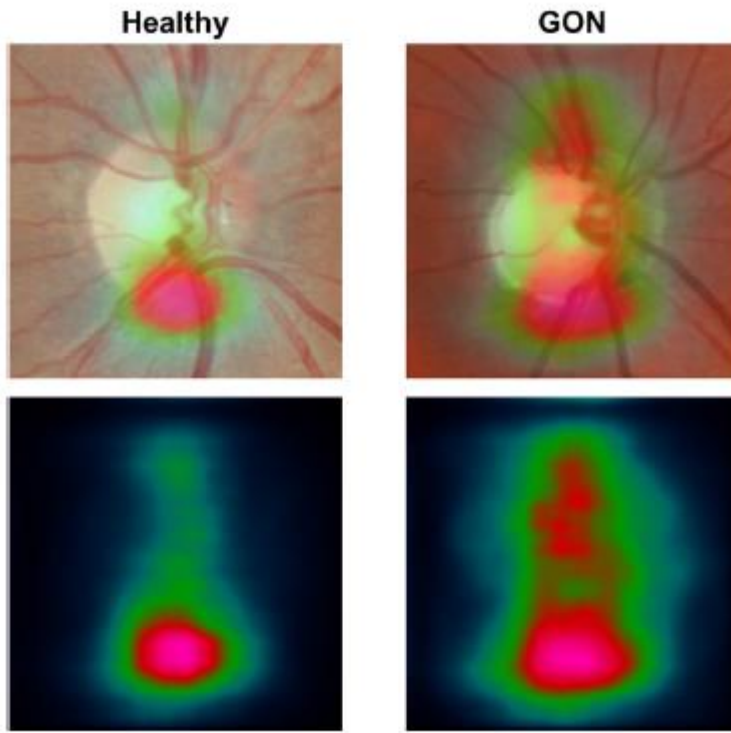


Figura 2.5: Mapa de calor de la visualización por oclusión. Extraído de [5]

Capítulo 3: Desarrollo del proyecto

En este apartado se expondrá los materiales y procedimientos utilizados en la realización de este TFG, así como los resultados obtenidos y los problemas encontrados en este trabajo.

3.1 Materiales utilizados

3.1.1 Bases de datos de imágenes

En el diagnóstico del glaucoma, la cabeza del nervio óptico (nos referiremos a ella también como disco óptico o papila) juega un papel fundamental y es por ello por lo que las imágenes utilizadas en este trabajo están centradas en esta parte de nuestra anatomía. Además, dichas imágenes son de dos tipos. Por un lado, se han utilizado imágenes de color del disco óptico en RGB, adquiridas con un aparato llamado retinógrafo, y por otro, imágenes equivalentes capturadas con un Tomógrafo de Coherencia Óptica.

Las imágenes en color están agrupadas en una base de datos llamada RIM-ONE enfocada principalmente en el diagnóstico de glaucoma. Actualmente hay tres versiones públicamente disponibles, cuyas imágenes se han recortado para descartar aquellas partes de la retina que no pertenecen a la zona de la cabeza del nervio óptico. Para este trabajo, hemos considerado una nueva versión que combina las tres anteriores, eliminando algunas repeticiones e imágenes con un diagnóstico no demasiado claro. En total, se han usado 495 imágenes, 314 correspondientes a casos diagnosticados como normales y 181 como glaucomas.

Para dos de las técnicas analizadas, concretamente la Grad-CAM y la de oclusión, se segmentaron manualmente 30 imágenes en color normales y 30 con glaucoma con el propósito de distinguir las tres partes principales: fondo de la imagen, disco óptico y vasos. Para ello se utilizó la aplicación MRSeg desarrollada a partir de un trabajo de fin de grado en el año 2013, por Sergio Jesús García Llarena [6].

Como se acaba de comentar, para la captura de imágenes de papila en infrarrojo se utilizó un Tomógrafo de Coherencia Óptica (OCT). El OCT es una técnica diagnóstica que utiliza luz de baja coherencia del infrarrojo cercano para registrar imágenes de dos y tres dimensiones. Concretamente, en nuestro trabajo hemos utilizado un OCT-Spectralis® (Heidelberg Engineering, Heidelberg, Alemania) de tipo spectral-domain que trabaja a una longitud de onda de 870 nm. En total, se han usado 316 imágenes de las que 131 fueron diagnosticadas como normales por el especialista y los 185 restantes como casos de glaucoma.

Hay que decir que, aunque las pruebas experimentales llevadas a cabo se han realizado sobre todas las imágenes, a efectos de poder mostrar algunos resultados en esta memoria, el oftalmólogo ha seleccionado cuatro ejemplos representativos, dos normales en color e infrarrojo (*Figura 3.1*) y dos con glaucoma en color e infrarrojo (*Figura 3.2*). Los resultados obtenidos para todas las imágenes con algunas de las técnicas utilizadas se pueden consultar en el siguiente enlace:

<https://drive.google.com/open?id=1mN-b1QwX0E1Y5GGW5USau-RGP6rZ5ECh>

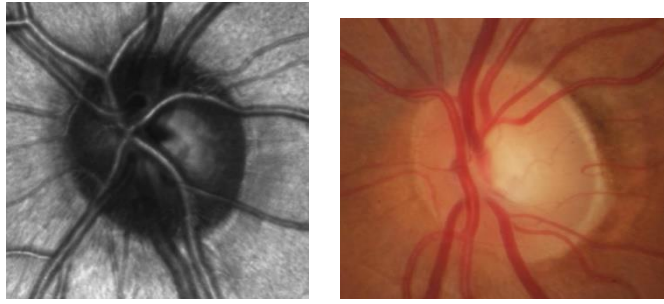


Figura 3.1: Imagen en infrarrojo y color del nervio óptico de un ojo sano

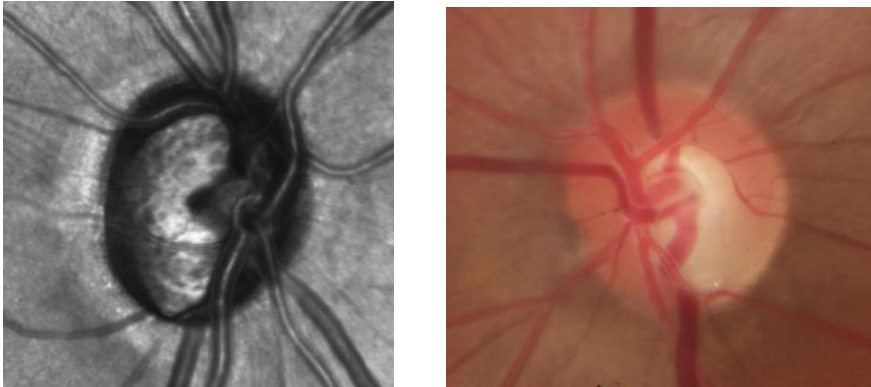


Figura 3.2: Imagen en infrarrojo y color del nervio óptico de un ojo glaucomatoso

3.1.2 Tecnologías utilizadas

Para este proyecto, se ha utilizado Keras que es una biblioteca desarrollada en Python y centrada en facilitar el uso de Redes Neuronales y Deep Learning, siendo su desarrollador principal y mantenedor el ingeniero de Google, François Chollet. También se usó una red neuronal convolucional VGG16 para trabajar con los datasets de OCT y RIM-ONE. Las VGG provienen del Grupo de Geometría Visual de Oxford (Visual Geometry Group) que siendo introducidas por Simonyan y Zisserman han conseguido ser una de las arquitecturas de redes neuronales convolucionales más ampliamente usadas. Hay dos tipos representativos de esta CNN, la VGG16 y la VGG19, teniendo cada una 16 y 19 capas de pesos respectivamente. Como ya se ha mencionado anteriormente, en este proyecto se usará la primera de estas arquitecturas, compuesta fundamentalmente por tres tipos de capas [7]:

- Capas convolucionales en las que, como su propio nombre indica, se llevan a cabo operaciones de convolución de forma jerárquica sobre la salida de capas anteriores. La convolución es una operación de sumas y productos que permite generar los mapas de características.
- Capas de pooling que consiste, básicamente, en hacer un downsampling de los mapas de características, usualmente utilizando el promedio o el máximo.
- Capas totalmente conectadas en las que se pierde la información espacial y son las que anteceden, habitualmente, a la capa de salida.

En la *Figura 3.3* se puede observar la estructura de una red neuronal VGG16 con los elementos mencionados.

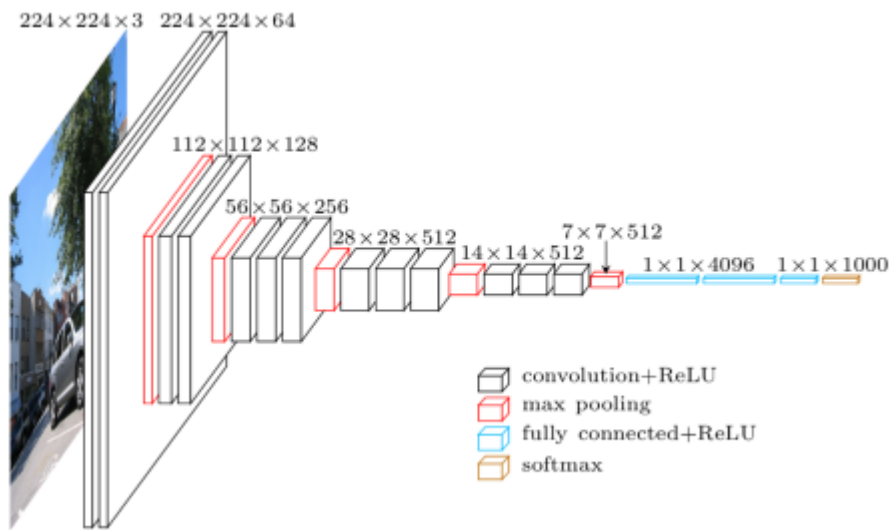


Figura 3.3: Arquitectura de una VGG16. Extraído de [7].

3.2 Descripción de los métodos de visualización

3.2.1 Visualización de las activaciones de las neuronas

Para esta sección[8], se entenderá por activación la salida de las diferentes capas de la red neuronal para una determinada imagen de entrada. Cada capa o mapa de características tiene tres dimensiones: anchura, altura y profundidad, correspondiendo esta última a los diferentes canales que codifican características relativamente independientes. La idea de esta técnica es visualizar en 2D las activaciones de las neuronas de dichos canales.

3.2.2 Visualización de los filtros de convolución

Esta forma de visualización [9], al contrario que la expuesta anteriormente, no toma como referencia una imagen de entrada concreta, sino que es un método más global que pretende mostrar el comportamiento de los filtros que la red ha aprendido a través de los parámetros o pesos de los que dependen. Para las primeras capas es posible interpretar los valores de estos pesos ya que suelen ser detectores de bordes u otras características de muy bajo nivel, pero para capas más profundas es muy difícil llegar a tener intuición sobre el problema. Es por eso por lo que en lugar de intentar visualizar los pesos directamente, se hace a través de un procedimiento más indirecto que busca generar imágenes que maximicen la salida de los diferentes filtros. El procedimiento en sí consiste en aplicar el gradiente estocástico descendente para ajustar los valores de una imagen de entrada que maximice las activaciones de las neuronas de un filtro o canal específico, partiendo de una imagen

con valores aleatorios. Gracias a este procedimiento, se puede saber a qué patrones o texturas son más sensibles los distintos filtros, aunque a veces resulte complicada la interpretación de los patrones visualizados.

3.2.3 Visualización basada en Gradient Backpropagation

Este tipo de visualización es un nombre genérico para diferentes técnicas que tienen como característica común que en todas ellas se invierte el flujo de información a través de la red, propagando una señal hacia atrás desde las neuronas de salida de un determinado canal a través de todas las capas hasta llegar a la imagen de entrada con el fin de saber qué píxeles de la misma son los que han contribuido en mayor medida a la activación de dicha neurona. Las tres técnicas más estándar en este contexto son: Vanilla Backpropagation, Deconvolutional Networks y Guided Backpropagation.

Vanilla Backpropagation es la técnica más básica y consiste en propagar la señal hacia atrás, después de propagarla hacia adelante, para calcular el gradiente de la activación con respecto a la imagen de entrada.

Las Deconvnet o redes deconvolucionales, son modelos de redes neuronales convolucionales conteniendo los mismos componentes de pooling y de filtrado, pero usándolos de manera inversa, es decir, en vez de mapear los píxeles a las características, hace lo contrario. Esto ayuda a tener siempre una forma de volver a los píxeles de la imagen y comprobar qué partes hacen que se active el mapa de características utilizado. La principal diferencia con la técnica anterior es la manera en la que se gestiona la propagación hacia atrás a través de las unidades lineales rectificadas (ReLU).

Con los dos métodos anteriores se obtienen, por lo general, imágenes bastante ruidosas. La técnica Guided Backpropagation[10] se puede considerar como una modificación de la técnica Deconv y se denomina así porque se añade una señal que sirve de guía con respecto al backpropagation usual, dando como resultado imágenes con una mayor calidad de visualización. Por este motivo, en este trabajo, nos hemos centrado en esta última técnica. La *Figura 3.4* muestra de forma esquemática el funcionamiento de los tres métodos y sus diferencias.

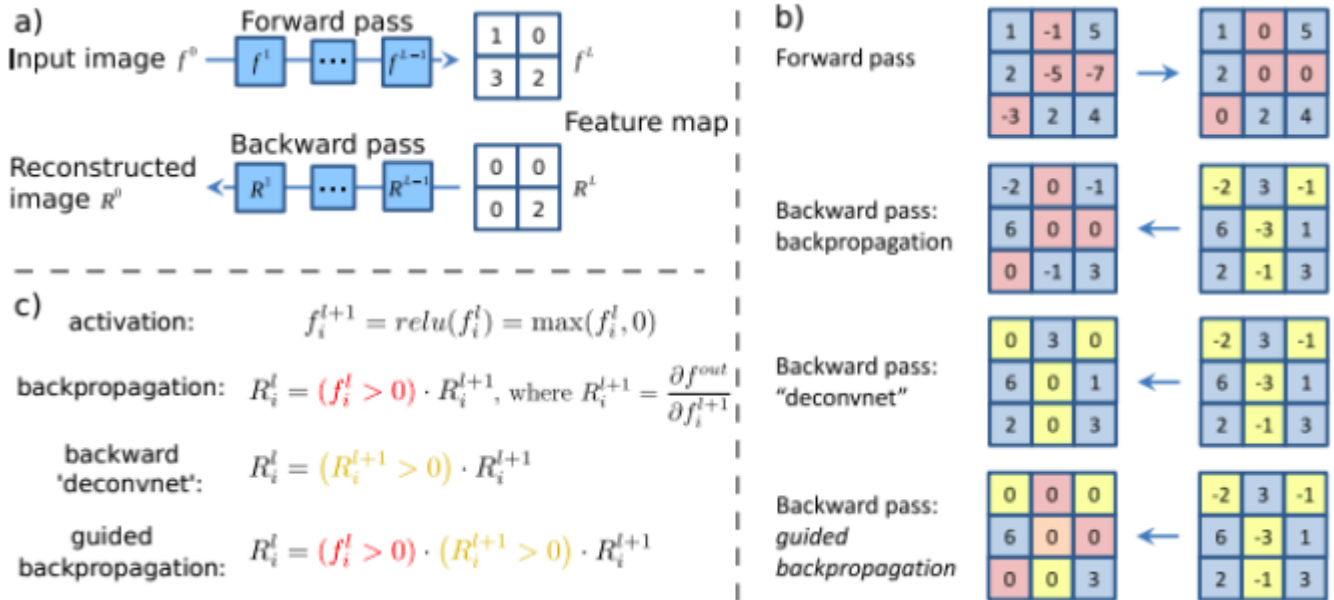


Figura 3.4: Esquema de visualización de activaciones de las neuronas de las capas más altas

3.2.4 Visualización basada en Mapas de Activación de Clases

En este apartado se abordan dos de las técnicas más utilizadas para la visualización del comportamiento de las redes: CAM (Class Activation Map) y Grad-CAM.

El método CAM permite, dada una imagen de entrada, saber qué partes de dicha imagen han sido más determinantes con respecto al valor de probabilidad asignado a una cierta clase utilizando los mapas de activación de las neuronas de la capa de convolución más profunda y que, por lo tanto, responde a características de más alto nivel. El principal inconveniente de esta técnica es que requiere la modificación de la arquitectura de la red para reemplazar las capas totalmente conectadas por capas de convolución y pooling, lo que además implica tener que reentrenar el sistema.

Grad-CAM[11] se puede considerar como una generalización de la técnica CAM que evita estos inconvenientes. Para ello, aprovecha la información del gradiente de las activaciones de la capa de convolución más profunda con respecto a una determinada clase de salida. A continuación, se muestra de forma esquemática lo que serían los pasos a seguir para aplicar esta técnica:

1. Tomamos el mapa de características de la capa de convolución final como F_1, F_2, \dots, F_n . La imagen de entrada la tomamos como I_0 , la clase de referencia c y la predicción para esa clase como $S_c(I_0)$.
2. Los pesos (w_1, w_2, \dots, w_n) para cada pixel en F_1, F_2, \dots, F_n son calculados a partir del gradiente de la predicción con respecto a cada mapa de características:

$$w_i = \frac{\partial S_c}{\partial F_i} \bigg|_{F_i} \quad \forall i = 1 \dots n$$

3. Los pesos y las activaciones correspondientes del mapa de características son multiplicadas

para calcular las activaciones ponderadas (A_1, A_2, \dots, A_n) para cada píxel en el mapa de características.

$$A_i = w_i * F_i \quad \forall i = 1 \dots n$$

- Las activaciones ponderadas de los distintos canales de la capa se suman a nivel de píxel para indicar la importancia de cada uno en el mapa definitivo de importancia de las características $H_{i,j}$ como:

$$H_{i,j} = \sum_{k=1}^n A_k(i, j) \quad \forall i = 1 \dots n$$

Dicho mapa debe ser re-escalado a la resolución de la imagen original y se representa a través de un mapa de calor.

3.2.5 Visualización basada en oclusión

Esta sencilla técnica[12] se encuadra en lo que se conoce como métodos de perturbación que consisten en alterar la imagen de entrada y ver el efecto que esto tiene en la probabilidad asignada por una determinada clase. En el caso concreto de la oclusión, lo más habitual es poner a cero una parte de la imagen (normalmente una ventana cuadrada) y ver en qué partes de esta disminuye el valor de probabilidad lo que resulta indicativo de la importancia de esa parte específica en la predicción de la red. Al igual que ocurre con las técnicas del apartado anterior, el resultado también se suele mostrar en forma de mapa de calor.

3.3 Resultados y discusión

3.3.1 Visualización por activaciones intermedias

A continuación, se presentan los resultados obtenidos para las imágenes de prueba seleccionadas. Por simplicidad y motivos de espacio, se muestran solo los resultados correspondientes a la primera capa (block1_conv1), una intermedia (block2_conv2) y la última capa de convolución (block5_conv3) para una imagen de color y otra de infrarrojo.

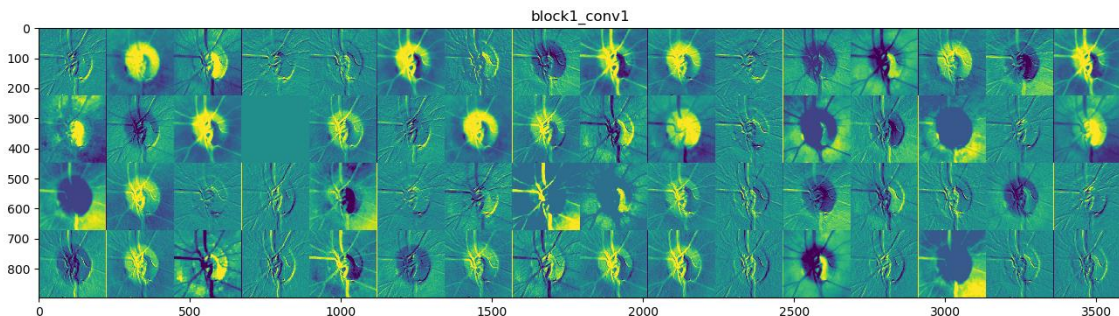


Figura 3.5: Activación intermedia de imagen con glaucoma a color block1_conv1

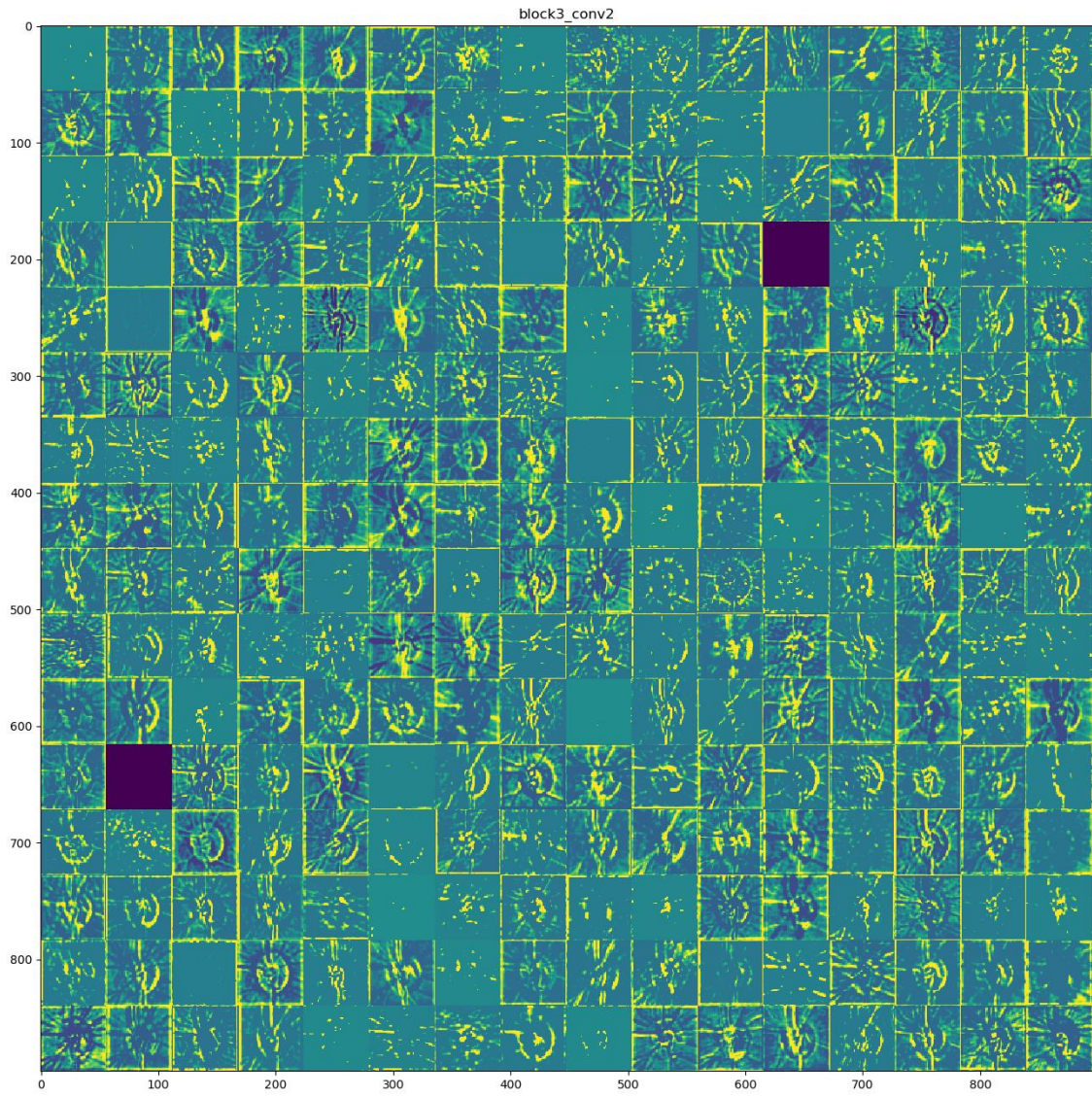


Figura 3.6: Activación intermedia de imagen con glaucoma a color block2_conv2

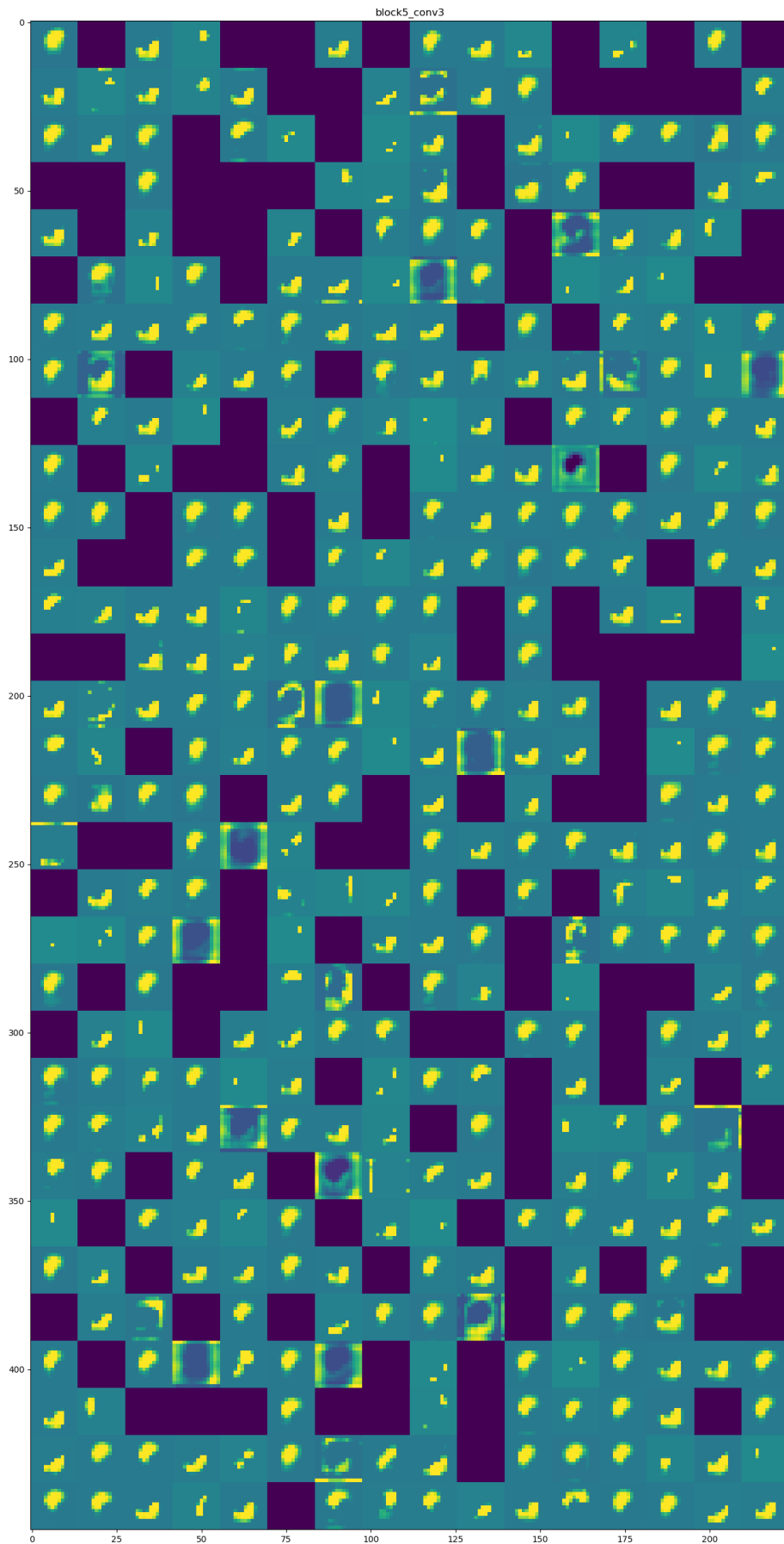


Figura 3.7: Activación intermedia de imagen con glaucoma a color `block5_conv3`

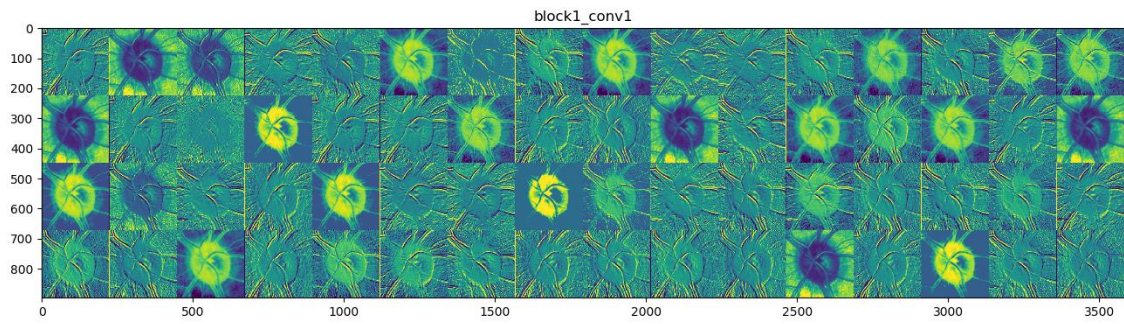


Figura 3.8: Activación intermedia de imagen sin glaucoma en OCT block1_conv1

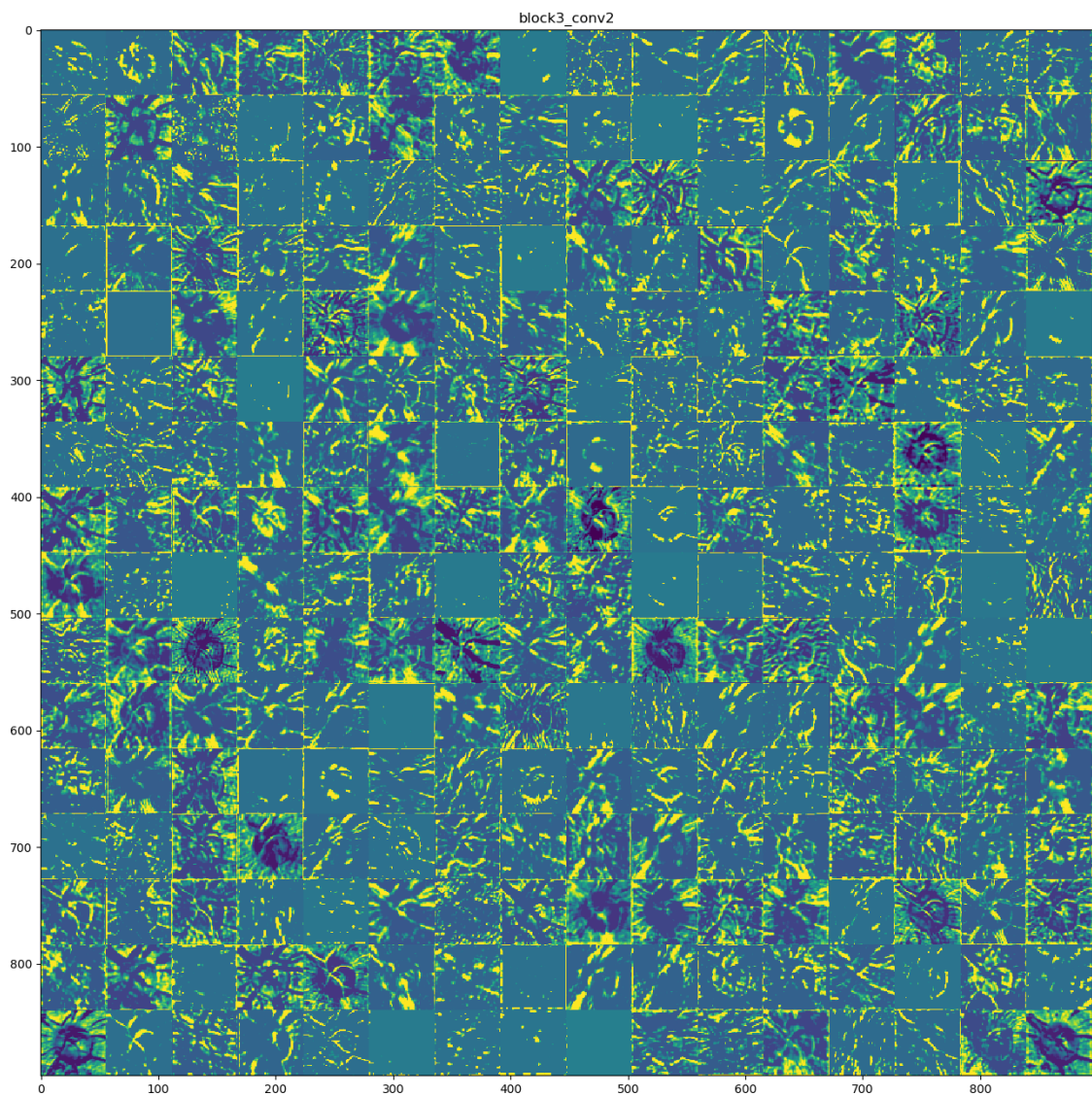


Figura 3.9: Activación intermedia de imagen sin glaucoma en OCT block3_conv2

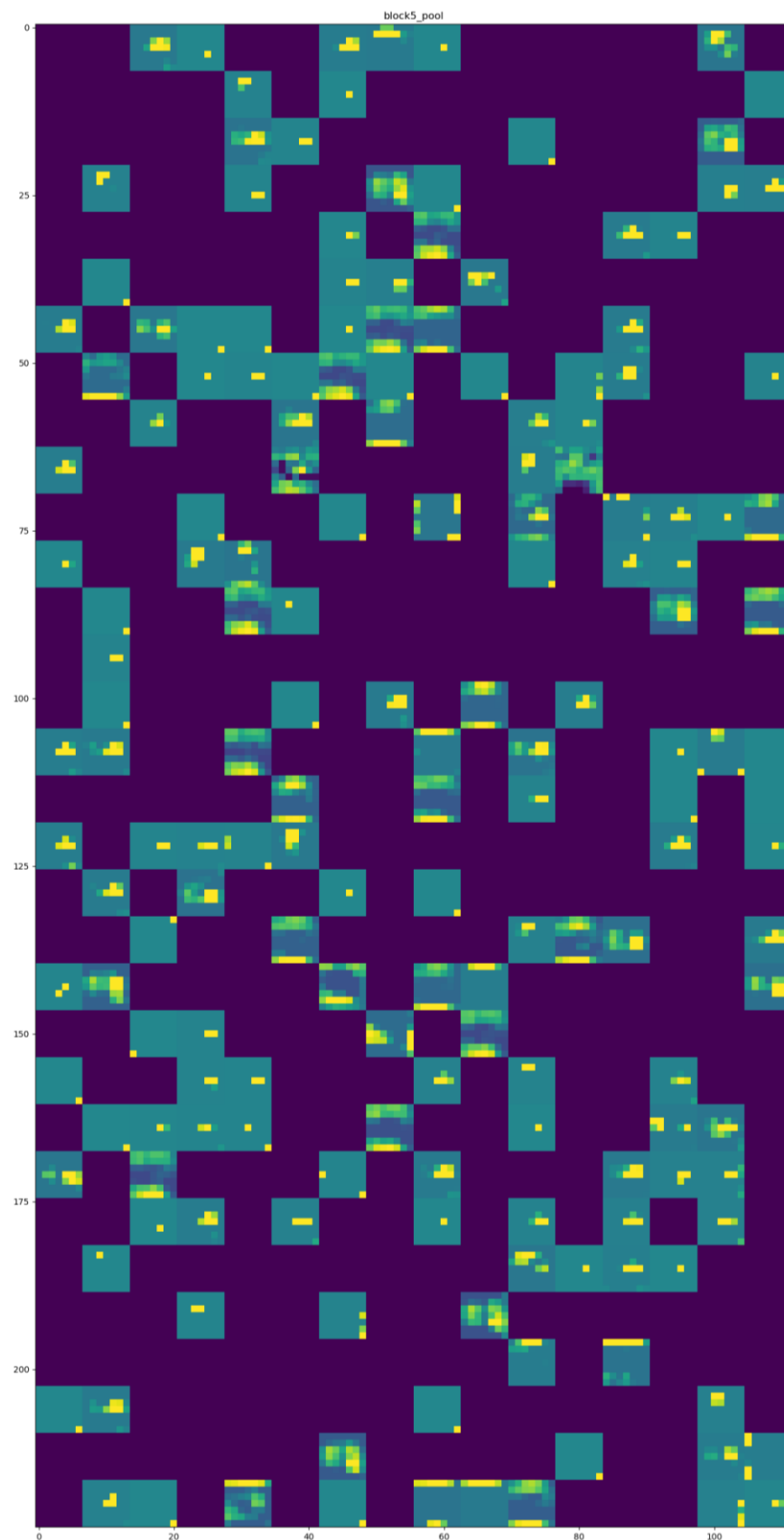


Figura 3.10: Activación intermedia de imagen sin glaucoma en OCT block5_conv3

Lo que se puede observar es más o menos lo esperado. Las neuronas de las primeras capas responden a características de bajo nivel, como bordes, aunque también se aprecia una activación conjunta en partes significativas como vasos y partes de la papila, y bastante redundancia. A medida que vamos hacia capas más profundas, las activaciones se hacen más difíciles de interpretar, y ya en la última capa de convolución hay bastantes filtros que no dan ninguna respuesta lo cual se debe a que el patrón codificado por dichos filtros no aparece en la imagen concreta que se ha utilizado como imagen de entrada.

3.3.2 Visualización por filtros

Por simplicidad, se muestran los resultados correspondientes a los 64 filtros con mayores activaciones en cada capa, y se incluyen las mismas capas que en el apartado anterior.

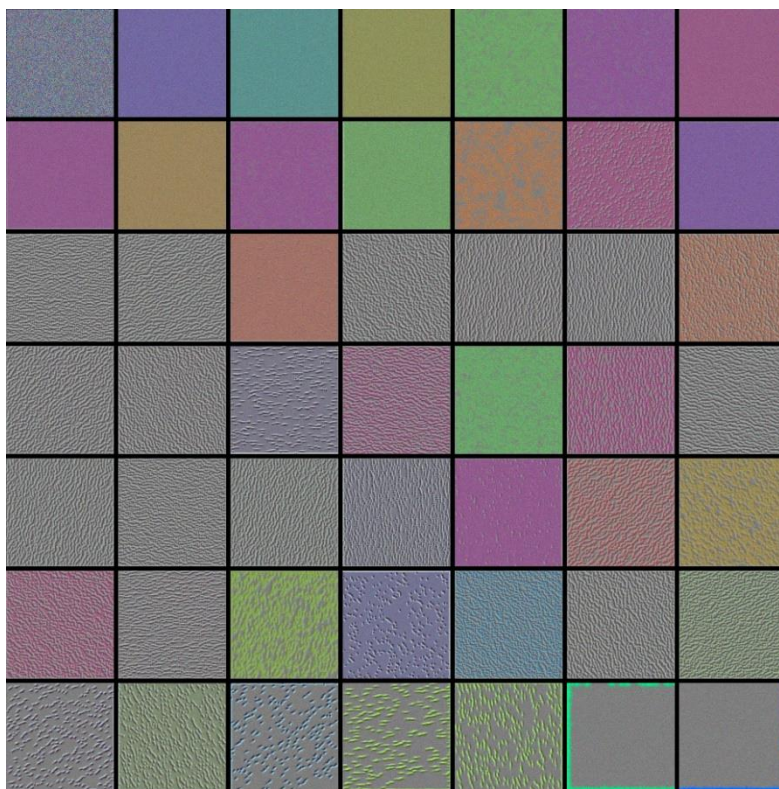


Figura 3.11 Filtros imagen en color block_conv1

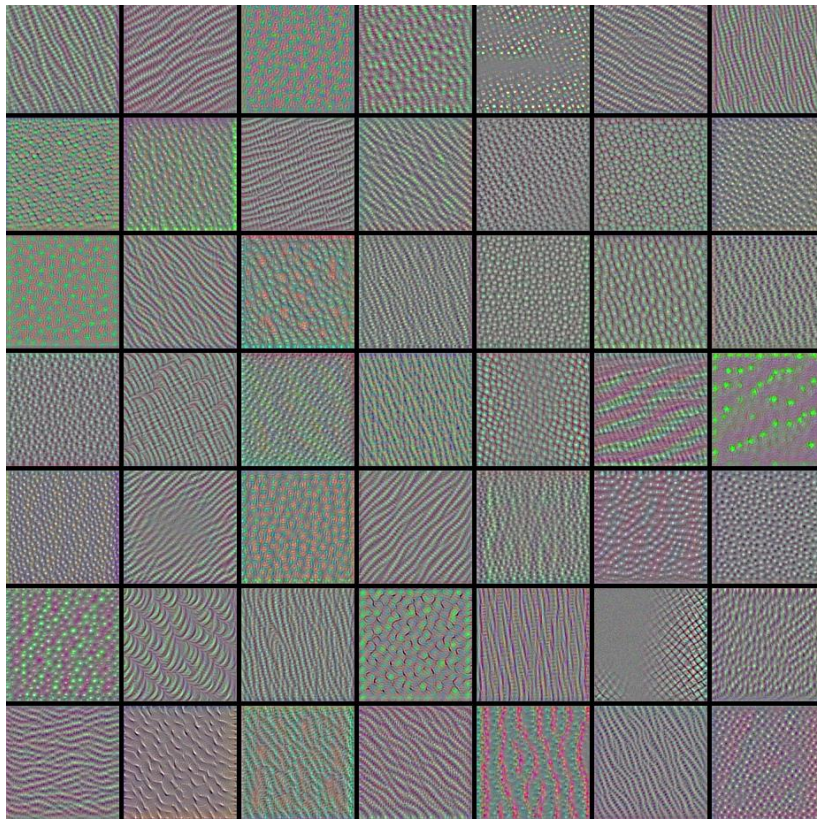


Figura 3.12 Filtros de imagen a color block3_conv2

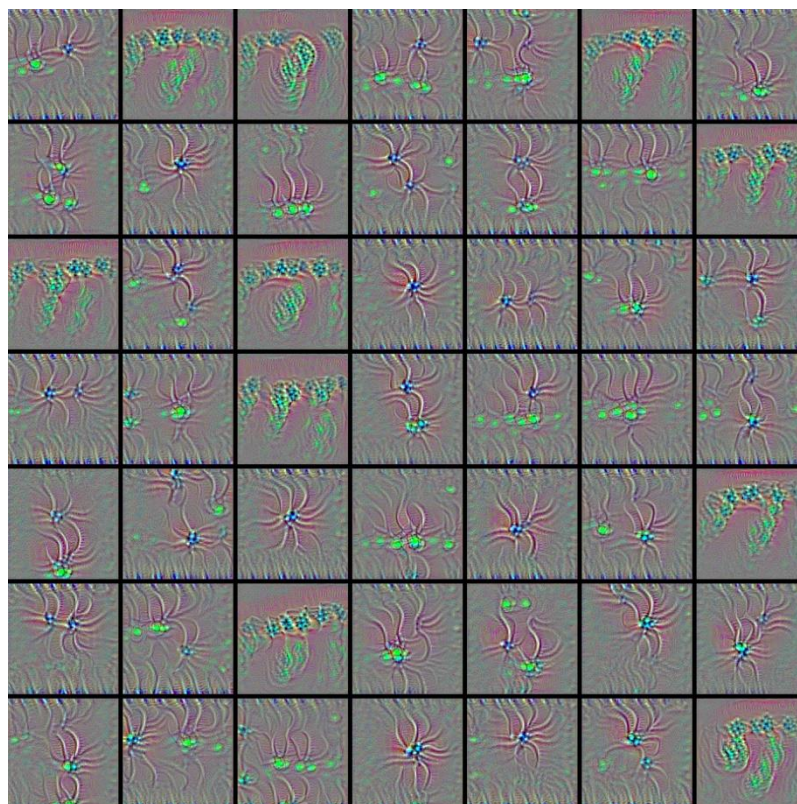


Figura 3.13 Filtros de imagen a color block5_conv3

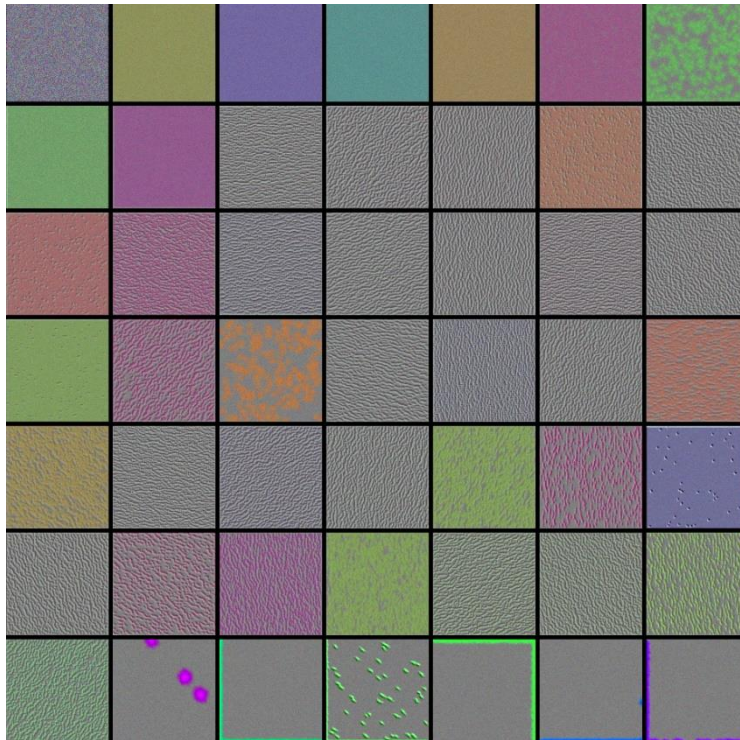


Figura 3.14: Filtros OCT block1_conv1

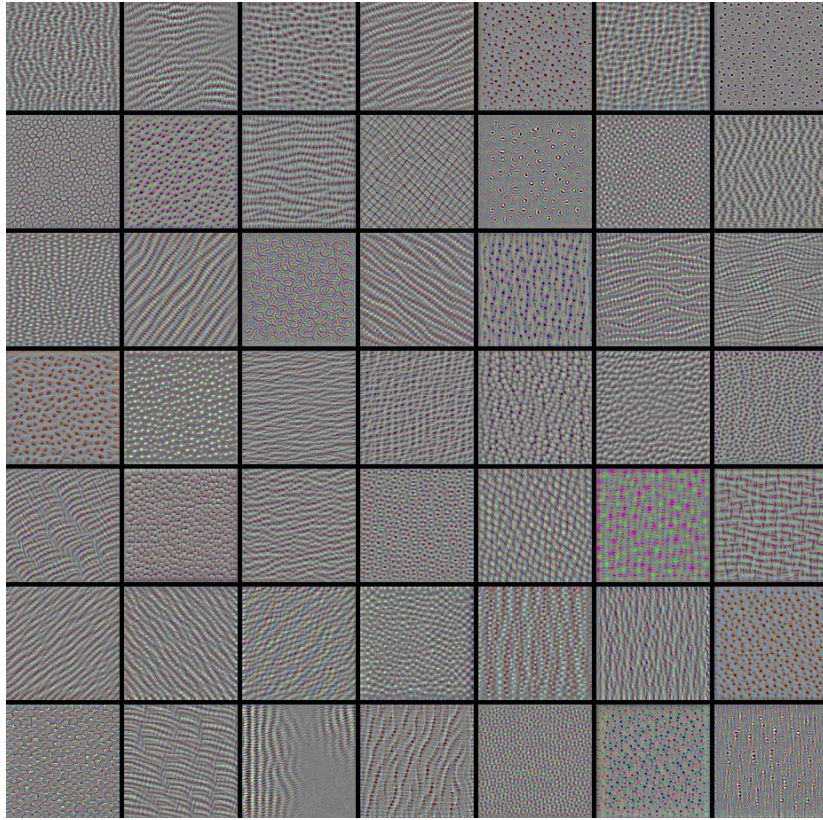


Figura 3.15: Filtros OCT block3_conv2

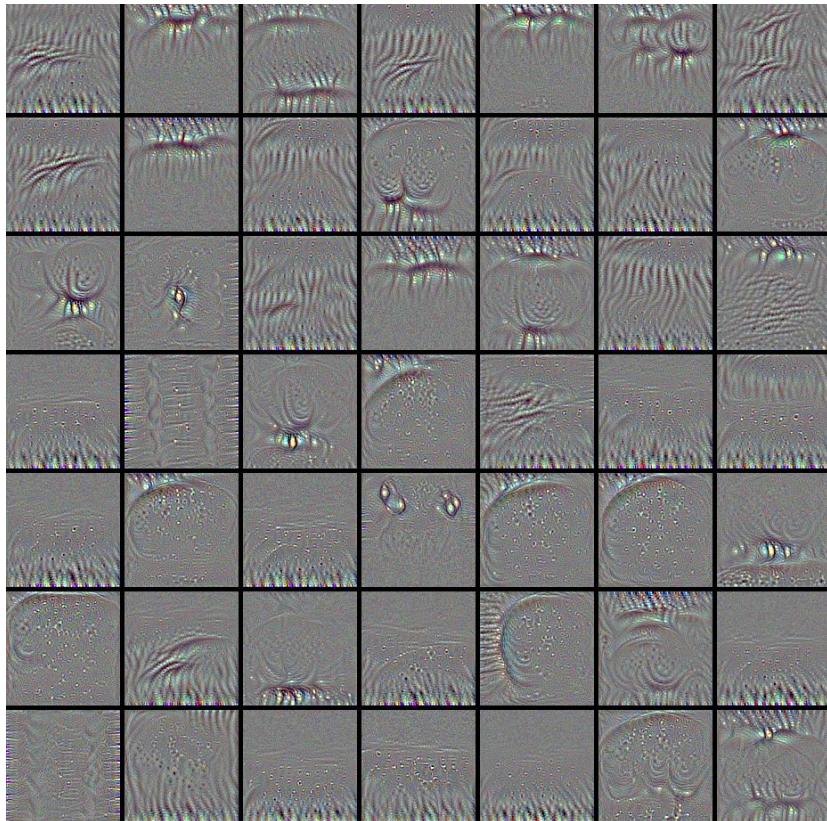


Figura 3.16: Filtros OCT block5_conv3

Los patrones que aparecen en la primera capa y en la intermedia son similares a los que se pueden encontrar en redes entrenadas para otro tipo de problemas que no tienen que ver con este trabajo. Sin embargo, en la última capa sí que se observan algunos patrones más característicos que pueden identificarse con las estructuras habituales en este tipo de imágenes, especialmente en el caso de las imágenes de infrarrojo donde destacan lo que parecen fibras nerviosas e incluso el perfil de lo que sería la papila. En las imágenes en color hay un patrón que se repite mucho que se podría identificar con el paquete vascular de la papila.

3.3.3 Visualización por Guided Backpropagation

A continuación, se muestran los resultados encontrados tras la aplicación de la técnica Guided Backpropagation.



Figura 3.17: Guided Backpropagation de una imagen a color con glaucoma

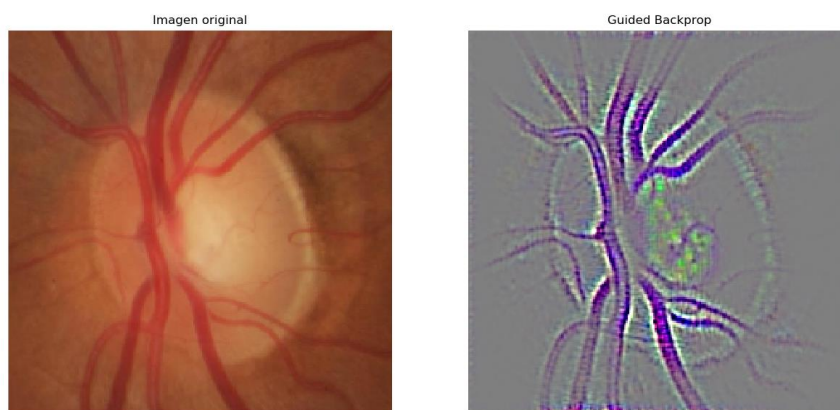


Figura 3.18: Guided Backpropagation de una imagen a color sin glaucoma

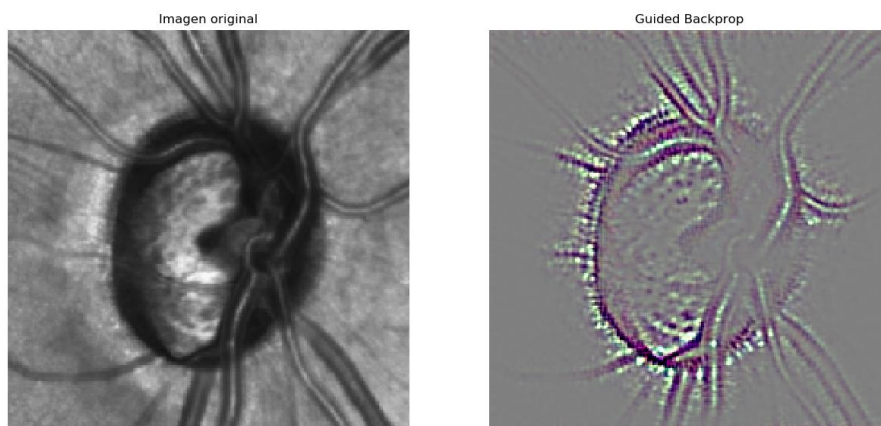


Figura 3.19: Guided Backpropagation de una imagen en OCT con glaucoma

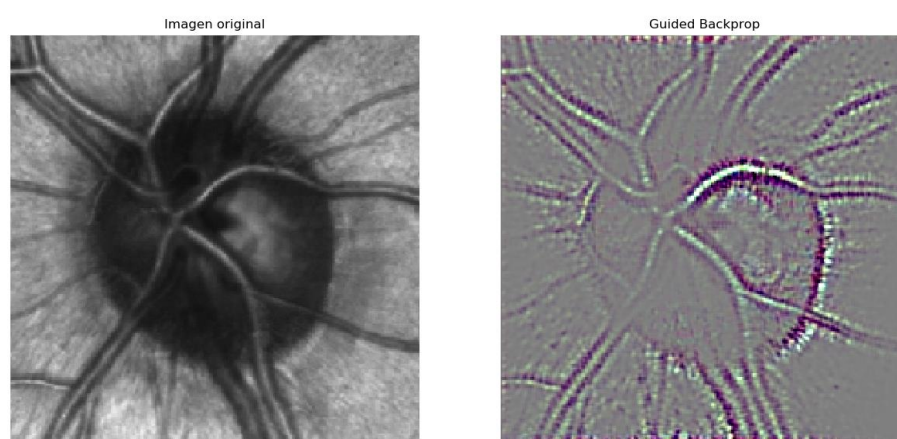


Figura 3.20: Guided Backpropagation de una imagen en OCT sin glaucoma

Como ya comentamos en la descripción de este método, se trata de ver a nivel de pixel la influencia de los mismos en la activación, en este caso, de las neuronas de la capa de convolución más profunda (block5-conv3). La principal ventaja de esta técnica es, precisamente, que da bastante detalle sobre aquello en lo que más se fija la red, pero, por contra, no discrimina bien las clases consideradas, lo que puede llevar a destacar partes de la imagen que no son realmente tan relevantes para una determinada clase, en este caso, glaucoma o no glaucoma. Además, como se puede observar, la salida es bastante ruidosa, lo que dificulta su interpretación.

3.3.4 Visualización por Grad-CAM

A continuación, se muestran los resultados obtenidos con la técnica Grad-CAM.

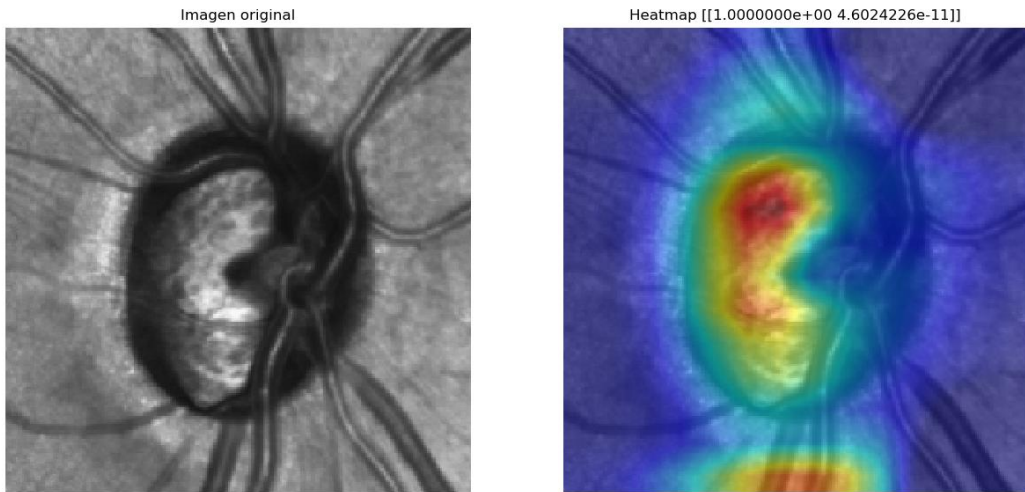


Figura 3.21: Grad-CAM de una OCT con glaucoma

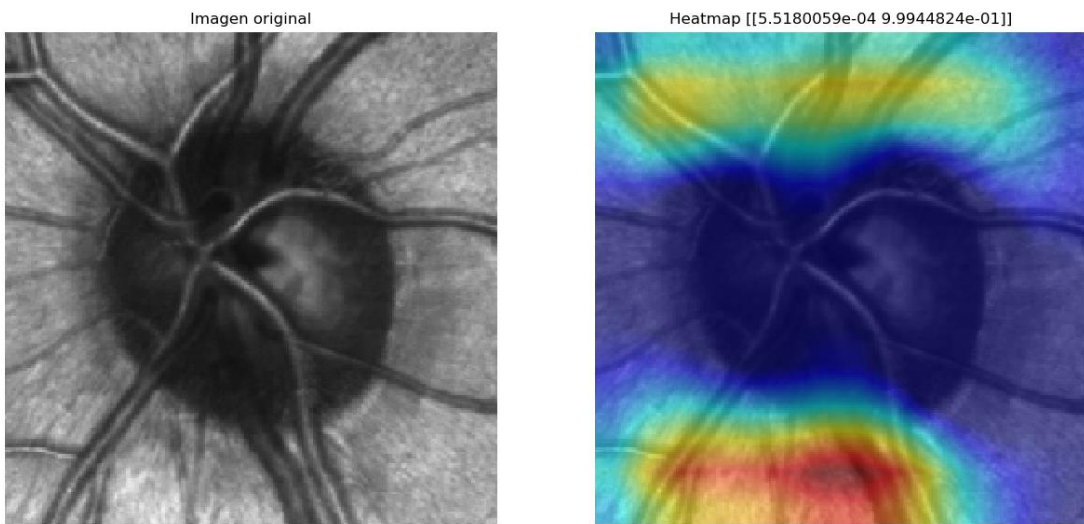


Figura 3.22: Grad-CAM de una OCT sin glaucoma

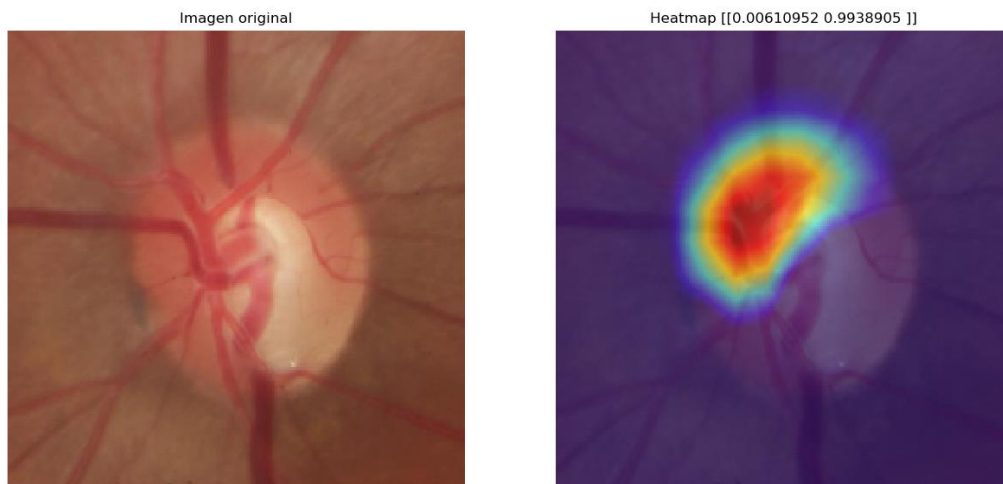


Figura 3.23: Grad-CAM de una retinografía a color con glaucoma



Figura 3.24: Grad-CAM de una retinografía a color sin glaucoma

Como ya se comentó anteriormente, se trata de la técnica que más se ha trabajado en este proyecto y la que ha proporcionado resultados más concluyentes. De la observación de las figuras se desprende, claramente, que, en el caso de las imágenes de color, la red es muy papilo-dependiente y focaliza su atención en el interior del disco tanto en zona de los vasos como fuera de ella. Sin embargo, en el caso de las imágenes de OCT se fija mucho más en la zona fuera del disco donde se pueden apreciar las fibras nerviosas que en color son muy difíciles de distinguir y que constituyen una pista clave para el diagnóstico de la enfermedad. De hecho, se ve que en el caso de la imagen de infrarrojo sin glaucoma (que es donde las fibras están intactas), apenas repara en lo que pasa dentro de la papila.

Aparte de esta valoración cualitativa, nos planteamos un estudio un poco más cuantitativo y para ello se utilizaron las imágenes segmentadas a las que se hacía referencia en la sección 3.1.1. En la *Figura 3.26* se muestra una de estas imágenes segmentadas. Lo que se hizo fue umbralizar los mapas

de calor para quedarnos con la región correspondiente al 25% más alto y obtener así una máscara binaria que pudiéramos solapar con las máscaras de fondo, vasos y disco obtenidas a través de la segmentación manual con la herramienta MRSeg. La *Figura 3.25* muestra un ejemplo de la umbralización a la que hacemos referencia.

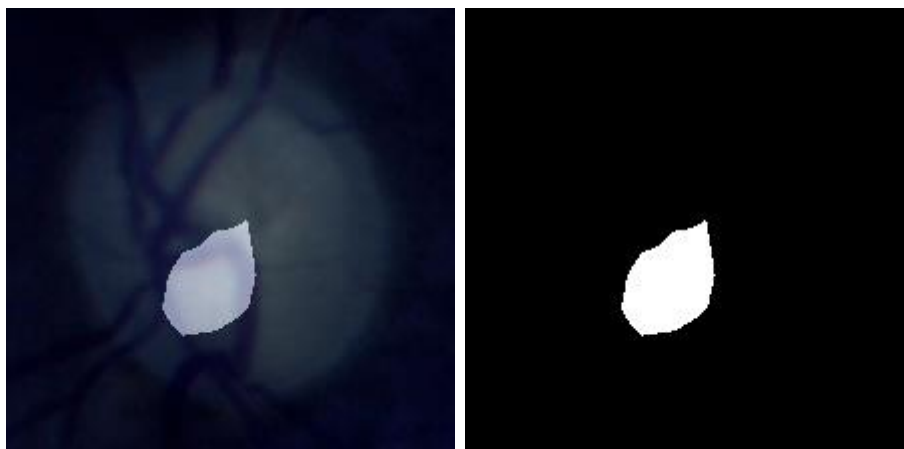


Figura 3.25: Mapa de calor (mayor de 75%) e imagen binaria



Figura 3.26: Imagen de fondo de ojo con glaucoma y partes segmentadas

Lo que hicimos fue calcular un porcentaje de solape entre la máscara del mapa de calor y las de las estructuras para determinar, en promedio, qué partes eran las que jugaban un papel más importante en cada caso: glaucoma y no glaucoma. Además, hay que aclarar que solo se calculó para las imágenes en color. Los resultados encontrados fueron los que se muestran en la siguiente tabla:

	Imágenes normales:	Imágenes con glaucoma:
Solape promedio con el disco (desviación típica)	0.52 (0.11)	0.88 (0.15)
Solape promedio con los vasos (desviación típica)	0.48 (0.11)	0.11 (0.15)
Solape promedio con el fondo (desviación típica)	0.00 (0.00)	0.01 (0.03)

Tabla 3.1: Promedio de solape en imágenes normales y glaucomatosas

Lo que se observa en la tabla se puede explicar desde el punto de vista médico teniendo en cuenta que el daño glaucomatoso comienza con la pérdida de fibras del anillo neuroretiniano y peripapilares, y esta pérdida es la que diagnostica el glaucoma. A medida que la enfermedad avanza posteriormente se afecta también la morfología del paquete vascular. Esto explica por qué el promedio en la zona del disco en glaucoma es tan alto. Si la muestra fuese solo de casos avanzados lo más probable es que no existiese esa diferencia.

Por último, hay que comentar que a diferencia de la técnica de Guided Backpropagation, aquí nos encontramos con una buena discriminación entre clases, pero muy poco detalle ya que, como se explicó en la descripción del método, el cálculo se hace en la capa convolucional más profunda que en VGG16 tiene una resolución espacial de 14x14 por lo que es necesario hacer un re-escalado para ponerla en la resolución original de 224x224 con lo que ello supone.

3.3.5 Visualización por oclusión

Los resultados encontrados con la aplicación de la técnica de oclusión sobre estas imágenes no han sido demasiado buenos, en particular con las imágenes normales, ya que en muchos casos no parecía tener ninguna incidencia el hecho de “tapar” ciertas partes de la imagen. A continuación, se muestra lo que se ha obtenido para la imagen de color con glaucoma con una oclusión de una ventana cuadrada de lado: 12, 16 y 20.



Figura 3.27: Visualización por Oclusión de imagen con glaucoma y ventana cuadrada de 12

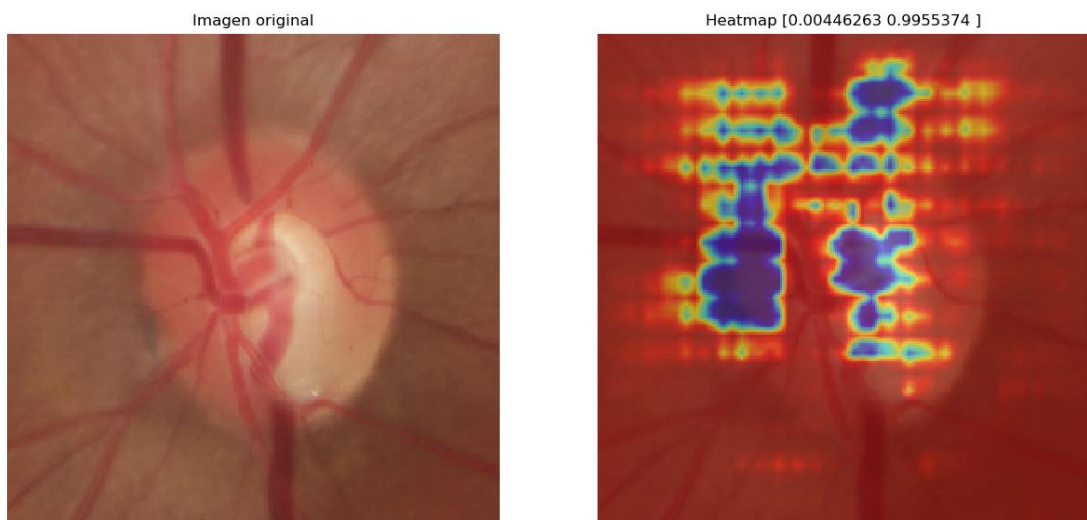


Figura 3.28: Visualización por Oclusión de imagen con glaucoma y ventana cuadrada de 16

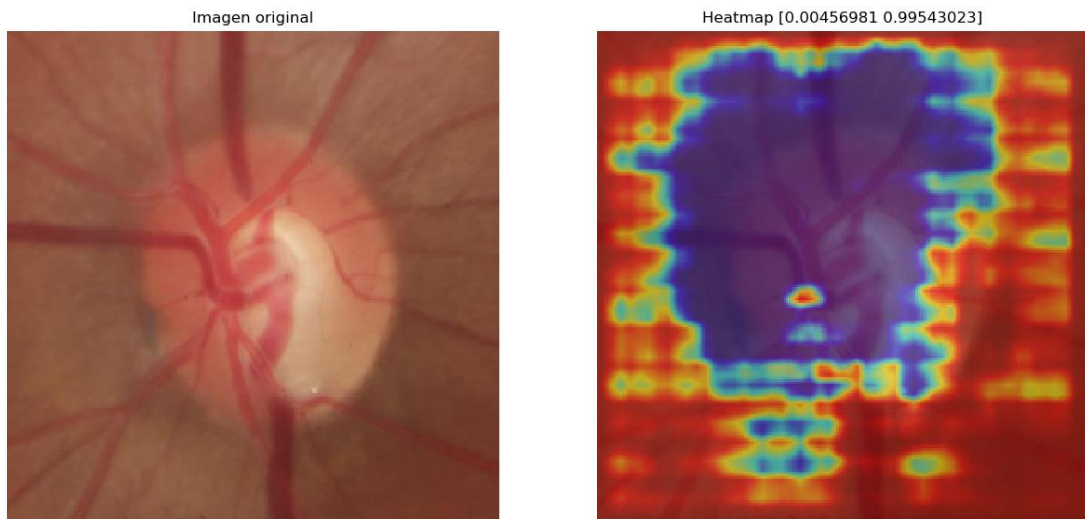


Figura 3.29: Visualización por Oclusión de imagen con glaucoma y ventana cuadrada de 20

Se aprecia claramente la influencia que tiene el tamaño elegido para la ventana. A pesar de tratarse de una técnica muy sencilla e intuitiva, ya que se trabaja a nivel de la imagen de entrada, el hecho de tener que ocluir una parte de esta introduce artefactos cuyos efectos sobre la probabilidad de predicción son impredecibles. De hecho, se realizó un experimento adicional que consistió en aprovechar la segmentación manual a la que se hizo referencia en el apartado anterior para ocluir partes de la imagen semánticamente significativas pero los resultados que se obtuvieron fueron disparatados ya que poner a cero partes tan grandes de la imagen confundía totalmente a la red que nunca había visto en el entrenamiento algo similar.

Capítulo 4: Conclusiones y líneas futuras

Dado que este trabajo se ha planteado como una primera aproximación al estudio de las técnicas de visualización e interpretación del comportamiento de las redes neuronales en el diagnóstico del glaucoma, tampoco cabe esperar que podamos obtener conclusiones muy definitivas al respecto. En primer lugar, de la aplicación de las diferentes técnicas estándar, podemos concluir que la información aportada por las mismas es limitada y difícil de interpretar en algunos casos. Coincidimos con otros investigadores en destacar la técnica Grad-CAM como la que aporta una información más clara e intuitiva, aunque también se han señalado sus limitaciones en el capítulo anterior. En los casos en los que sí ha sido posible interpretar los resultados obtenidos, hemos encontrado una buena coincidencia cualitativa y cuantitativa con el análisis médico, siendo mayor para las imágenes de infrarrojo, lo cual entra dentro de lo esperado dado que son imágenes de mejor calidad y aportan una información más clara y precisa sobre la patología a diagnosticar.

Son muchas las posibles líneas futuras de trabajo. Por supuesto, probar otro tipo de técnicas más sofisticadas y otros modelos de redes neuronales convolucionales. En general, hacer un estudio más profundo de este tema de gran importancia en un ámbito como la medicina.

Capítulo 5: Summary and Conclusions

Given that this work has been proposed as a first approach to the study of the visualization and interpretation of the behavior of neural networks in the diagnosis of glaucoma, we can not expect to obtain very definitive conclusions about it. In the first place, from the application of the different standard techniques, we can conclude that the information provided by them is limited and difficult to interpret in some cases. We agree with other researchers in highlighting the Grad-CAM technique as the one that provides the most clear and intuitive information, although its limitations have also been pointed out in the previous chapter. In cases where it has been possible to interpret the results obtained, we have found a good qualitative and quantitative coincidence with the medical analysis, being higher for infrared images, which is within the expected since they are images of better quality and provide clearer and more accurate information about the pathology to be diagnosed.

There are many possible future lines of work. Of course, to try other types of more sophisticated techniques and other models of convolutional neural networks. In general, to make a deeper study of this issue of great importance in a field such as medicine.

Capítulo 6: Presupuesto

En este capítulo se presentará el presupuesto por horas trabajadas según las actividades expuestas en la *Tabla 6.1*.

6.1 Presupuesto personal

Actividad	Horas trabajadas	Precio/Hora
Búsqueda de información	48 horas	10€/h
Investigación	110 horas	14€/h
Realización de la documentación	50 horas	10€/h
Programación	65 horas	15€/h
Estudio de resultados	37 horas	12€/h
TOTAL	310 horas	3939€

Tabla 1: Presupuesto

Bibliografía

- [1] S. Maetschke, B. Antony, H. Ishikawa, G. Wollstein, J. S. Schuman, R. Garvani, “A feature agnostic approach for glaucoma detection in OCT volumes”, pp. 13, May 2019 Disponible en: <https://arxiv.org/pdf/1807.04855.pdf> Accedido por última vez: 05/07/2019
- [2] M. Kim, H. Park, J. Zuallaert, O. Janssens, S. V. Hoecke, W. D. Neve, “Computer-Aided Diagnosis and Localization of Glaucoma Using Deep Learning” 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), December 2018 Disponible en: Disponible en: <https://ieeexplore.ieee.org/document/8621168> Accedido por última vez: 05/07/2019
- [3] S. Phan, S. Satoh, Y. Yoda, K. Kashiwagi, T. Oshika, “Evaluation of Deep convolutional networks for glaucoma detection”, Japanese Journal of Ophthalmology, vol. 64, pp 276-283, May 2019 Disponible en: <https://link.springer.com/article/10.1007/s10384-019-00659-6> Accedido por última vez: 05/07/2019
- [4] G. An, K. Omodaka, K. Hashimoto, S. Tsuda, Y. Shiga, N. Takada, T. Kikawa, H. Yokota, M. Akiba, T. Nakazawa, “Glaucoma Diagnosis with Machine Learning Base don Optical Coherence Tomography and Color Fundus Images”, Journal of Healthcare Engineering, vol. 2019, pp. 9, February 2019 Disponible en: <https://www.hindawi.com/journals/jhe/2019/4061313/> Accedido por última vez: 05/07/2019
- [5] M. Christopher, A. Belghith, C. Bowd, J. A. Proudfoot, M. H. Goldbaum, R.N. Weinreb, C. A. Girkin, J. M. Liebmann, L. M. Zangwill, “Performance of Deep Learning Architectures and Transfer Learning for Detecting Glaucomatous Optic Neuropathy in Fundus Photographs”, Scientific Reports, November 2018 Disponible en: <https://www.nature.com/articles/s41598-018-35044-9> Accedido por última vez: 05/07/2019
- [6] S. J. García Llarena, “Mejora de las prestaciones de una aplicación semiautomática de segmentación de imágenes”, 2013
- [7] F. Chollet, “Deep Learning for Computer Vision” in *Deep Learning with Python*, T. Arritola, Manning Publications Co.: Shelter Island, New York, 2018, pp. 119-177
- [8] F. Chollet, “Visualizing what convnets learn” in *Deep Learning with Python*, T. Arritola, Manning Publications Co.: Shelter Island, New York, 2018, pp. 160-167
- [9] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, H. Lipson, “Understanding Neural Networks Through Deep Visualization”, June 2015 Disponible en: http://yosinski.com/media/papers/Yosinski_2015_ICML_DL_Understanding_Neural_Networks_Through_Deep_Visualization_.pdf Accedido por última vez: 05/07/2019
- [10] J. T. Springenberg, A. Dosovitskiy, T. Brox, M. Riedmiller, “Striving for Simplicity: The All Convolutional Net”, December 2014 Disponible en: <https://arxiv.org/abs/1412.6806?context=cs> Accedido por última vez: 05/07/2019
- [11] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization”, October 2016

Disponible en: <https://arxiv.org/abs/1610.02391> Accedido por última vez: 05/07/2019

[12] M.T. Zeiler, R. Fergus, “Visualizing and Understanding Convolutional Networks”, November 2013, Disponible en: <https://arxiv.org/abs/1311.2901> Accedido por última vez: 05/07/2019

[13] R. Ghosh, S. Jain, “Visualizing Deep Learning Networks – Part I”, September 2017

Disponible en: http://blog.qure.ai/notes/visualizing_deep_learning Accedido por última vez: 05/07/2019

[14] R. Ghosh, S. Jain, “Visualizing Deep Learning Networks – Part II”, September 2017

Disponible en: <http://blog.qure.ai/notes/deep-learning-visualization-gradient-based-methods>
Accedido por última vez: 05/07/2019