



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

## Trabajo de Fin de Grado

---

**Redes Generativas Antagónicas para  
predecir la evolución del Glaucoma**

*Generative Adversarial Networks for predicting the evolution  
of Glaucoma*

Alejandro Peraza González

---

La Laguna, 7 de julio de 2022

D. **Rafael Arnay del Arco**, con N.I.F. 78.569.591-G profesor Contratado Doctor de Universidad adscrito al Departamento de Ingeniería Informática y de Sistemas de la Universidad de La Laguna, como tutor

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 cotutor

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

Que la presente memoria titulada:

*"Redes Generativas Antagónicas para predecir la evolución del Glaucoma"*

ha sido realizada bajo su dirección por D. **Alejandro Peraza González**, con N.I.F. 79.092.686-X.

Y para que así conste, en cumplimiento de la legislación vigente y a los efectos oportunos firman la presente en La Laguna a 7 de julio de 2022

# Agradecimientos

Quiero agradecer a mis tutores Rafael Arnay del Arco y José Francisco Sigut Saavedra por el apoyo recibido así como la guía ofrecida para llevar a cabo este trabajo. Asimismo, agradezco a todos los profesores que me han instruido a lo largo de mi transcurso en la Universidad de La Laguna que han hecho posible que me forme como un futuro Ingeniero Informático.

Me gustaría mostrar especial gratitud a todos mis seres queridos, de los que he recibido un constante apoyo no solo durante la elaboración de este trabajo. Este apoyo me ha ayudado a superar múltiples obstáculos con los que me he enfrentado.

# Licencia

\* Si quiere permitir que se compartan las adaptaciones de tu obra mientras se comparta de la misma manera y NO quieres permitir usos comerciales de tu obra indica:



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

## Resumen

*El glaucoma es un conjunto de enfermedades que causa daños tanto en el nervio óptico como en la retina, habitualmente generando pérdidas en el campo de visión del paciente. El tipo más común de glaucoma es muy difícil de detectar para la persona que lo padece, ya que no aprecia la pérdida en su campo de visión hasta que la enfermedad esté ya bastante avanzada. Esto origina una gran preocupación por el diagnóstico precoz de la enfermedad. Una vez un paciente es diagnosticado de glaucoma y medicado correctamente, cabe la posibilidad de que la enfermedad continúe progresando. Esta progresión es el foco de investigación de este trabajo. El objetivo principal es proporcionar al oftalmólogo una imagen de la progresión de la enfermedad, transcurrido un cierto periodo de tiempo, a partir de una imagen actual que represente el estado en el que se encuentra el sujeto en ese momento.*

*Las redes generativas antagónicas permiten la creación de imágenes fotorrealistas a partir de un conjunto de imágenes. Existe un tipo específico de red generativa antagónica que es utilizado para llevar a cabo tareas de traducción de imagen a imagen, las redes Pix2Pix. Estas redes son planteadas como un framework versátil que permita utilizar la misma arquitectura y función de pérdida para problemas distintos. En este trabajo se hace uso de estas redes para, partiendo de una imagen de un ojo glaucomatoso, generar una imagen falsa que comparta características con ella y que refleje la evolución de la enfermedad.*

**Palabras clave:** Glaucoma, Retinografía, Mapa de células ganglionares, Tomografía de coherencia óptica (OCT), Red generativa antagónica (GAN), Software de traducción de imagen a imagen basado en GANs (Pix2Pix)

## **Abstract**

*Glaucoma is a group of diseases that cause damage to both the optic nerve and the retina, usually resulting in a loss of the patient's field of vision. The most common type of glaucoma is very difficult to detect for the person who suffers from it, since he/she does not notice the loss in his/her field of vision until the disease is already quite advanced. This causes great concern for early diagnosis of the disease. Once a patient is diagnosed with glaucoma and properly medicated, the disease is likely to continue to progress. This progression is the focus of research in this paper. The main objective is to provide the ophthalmologist with an image of the progression of the disease, after a certain period of time, from a current image that represents the state in which the subject is at that moment.*

*Generative adversarial networks allow the creation of photorealistic images from a set of images. There is a specific type of generative adversarial networks that is used to perform image-to-image translation tasks, the Pix2Pix networks. These networks are proposed as a versatile framework that allows using the same architecture and loss function for different problems. In this work we make use of these networks to generate a false image that shares characteristics with an image of a glaucomatous eye and that reflects the evolution of the disease.*

**Keywords:** Glaucoma, Retinography, ganglion cell map, Optical coherence tomography (OCT), Generative adversarial network (GAN), Image-to-image translation software based on GANs (Pix2Pix)

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Objetivos y fases . . . . .	3
1.3. Antecedentes y estado del arte . . . . .	3
1.3.1. Antecedentes . . . . .	3
1.3.2. Estado del arte . . . . .	4
<b>2. Redes Generativas Antagónicas</b>	<b>6</b>
<b>3. Desarrollo del proyecto</b>	<b>9</b>
3.1. Tecnologías utilizadas . . . . .	9
3.2. Entrenamiento de GAN simple con imágenes de ranas de CIFAR-10 . . . . .	9
3.3. Entrenamiento de GAN y DCGAN con retinografías de RIM-ONE-DL . . . . .	10
3.4. Modelo empleado para predecir la evolución de una imagen de un ojo glaucomatoso . . . . .	13
3.4.1. Función objetivo de las redes Pix2Pix . . . . .	13
3.4.2. Generador utilizando red en U . . . . .	14
3.4.3. Discriminador utilizando <i>PatchGAN</i> . . . . .	15
3.5. Mapas de células ganglionares . . . . .	16
3.6. Preprocesamiento de las imágenes de mapas de células ganglionares . . . . .	17
3.6.1. Recogida de datos de las imágenes para su posterior alineación . . . . .	19
3.6.2. Alineación de las imágenes . . . . .	20
3.6.3. Escalado y emparejamiento de las imágenes . . . . .	20
3.7. Diagrama del proceso completo . . . . .	21
<b>4. Resultados</b>	<b>23</b>
4.1. Análisis y comparación cuantitativa de las imágenes generadas . . . . .	24
4.2. Análisis cuantitativo de los resultados . . . . .	25
4.3. Análisis cualitativo de imágenes generadas . . . . .	28
4.3.1. Análisis cualitativo experto . . . . .	28
4.3.2. Tripletas de imágenes con mapas de calor de la similitud estructural . . . . .	32
4.4. Experimento de clasificación entre imágenes falsas y reales . . . . .	36
<b>5. Conclusiones y líneas futuras</b>	<b>38</b>
5.1. Conclusiones . . . . .	38
5.2. Líneas futuras . . . . .	38
<b>6. Conclusion and future lines of work</b>	<b>40</b>
6.1. Conclusions . . . . .	40

6.2. Future lines of work . . . . .	40
<b>7. Presupuesto</b>	<b>42</b>
<b>A. Tablas de similitud estructural entre imagen falsa e imagen final</b>	<b>43</b>
A.1. Primera y última imagen . . . . .	43
A.2. Seis meses de diferencia entre las imágenes de un par . . . . .	44
A.3. Un año de diferencia entre las imágenes de un par . . . . .	45

# Índice de Figuras

1.1. Ejemplo de retinografía proveniente de RIM-ONE-DL . . . . .	1
1.2. Ejemplo de mapa de células ganglionares . . . . .	1
1.3. Tomografía de coherencia óptica de la mácula en el dominio espectral. . . . .	5
2.1. Arquitectura de una GAN. Imagen obtenida en Gen (2018) . . . . .	6
3.1. Imágenes reales y generadas de ranas en 32x32 . . . . .	10
3.2. Imagen falsa generada en las primeras iteraciones . . . . .	10
3.3. Imágenes de RIM-ONE-DL (Batista et al., 2020) de infrarrojo en escala de grises . . . . .	11
3.4. Intento de crear imágenes de fondo de ojo con resolución 32x32 . . . . .	11
3.5. Conjunto de imágenes generadas por una DCGAN entrenada con retinografías	12
3.6. Boceto de red en U obtenido en uNe (2019) . . . . .	14
3.7. Capas y partes de la mácula en una OCT en dominio espectral. Imagen obtenida de Reyes (2014) . . . . .	16
3.8. Imagen resaltando las partes de interés . . . . .	17
3.9. Imágenes de mapas de células ganglionares iniciales . . . . .	18
3.10 Figura 3.11 tras alinear y recortar . . . . .	19
3.11 Alineación de la mácula con el nervio óptico . . . . .	19
3.12 Resultados de entrenar la red Pix2Pix empleando imágenes de tamaño 192x192 píxeles . . . . .	21
3.13 Diagrama del proceso desarrollado . . . . .	22
4.1. Histograma de las medias de SSIM entre sectores. Pares de primera y última muestra de cada paciente . . . . .	27
4.2. Histograma de las medias de SSIM entre sectores. Pares de imágenes con hasta seis meses de diferencia . . . . .	27
4.3. Histograma de las medias de SSIM entre sectores. Pares de imágenes con hasta un año de diferencia . . . . .	28
4.4. “Aumento en extensión y profundidad de la lesión temporal inferior” . . . . .	29
4.5. “Cambios leves a nivel superior de la GCL” . . . . .	29
4.6. “Pérdida de la GCL a nivel temporal inferior y a nivel superior” . . . . .	30
4.7. Imagen generada por la red basada en la imagen 4.6a . . . . .	30
4.8. “Disminución difusa de la GCL” . . . . .	31
4.9. “Disminución generalizada de la GCL.” . . . . .	31
4.10 Imagen final del paciente, siendo la imagen inicial: 4.9a . . . . .	32
4.11. “Reducción de la GCL a nivel temporal y temporal inferior.” . . . . .	32
4.12 Par de imágenes reales junto al mapa de calor de SSIM (1) . . . . .	33
4.13 Par de imagen inicial con la imagen generada junto al mapa de calor de SSIM (1) . . . . .	33

4.14	Par de imagen generada y la imagen final del paciente junto al mapa de calor de SSIM (1)	33
4.15	Par de imágenes reales junto al mapa de calor de SSIM (2)	34
4.16	Par de imagen inicial con la imagen generada junto al mapa de calor de SSIM (2)	34
4.17	Par de imagen generada y la imagen final del paciente junto al mapa de calor de SSIM (2)	34
4.18	Par de imágenes reales junto al mapa de calor de SSIM (3)	35
4.19	Par de imagen inicial con la imagen generada junto al mapa de calor de SSIM (3)	35
4.20	Par de imagen generada y la imagen final del paciente junto al mapa de calor de SSIM (3)	35

# Índice de Tablas

4.1. Matriz de confusión de los resultados obtenidos en el experimento de clasificación . . . . .	36
7.1. Resumen de tipos . . . . .	42
A.1. Tabla de SSIM entre la imagen real tras la evolución y la imagen falsa. Pares de primera y última imagen . . . . .	43
A.2. Tabla de SSIM entre la imagen real tras la evolución y la imagen falsa. Pares de imágenes con seis meses de diferencia . . . . .	44
A.3. Tabla de SSIM entre la imagen real tras la evolución y la imagen falsa. Pares de imágenes con un año de diferencia (1) . . . . .	45
A.4. Tabla de SSIM entre la imagen real tras la evolución y la imagen falsa. Pares de imágenes con un año de diferencia (2) . . . . .	46
A.5. Tabla de SSIM entre la imagen real tras la evolución y la imagen falsa. Pares de imágenes con un año de diferencia (3) . . . . .	47

# Capítulo 1

## Introducción

### 1.1. Motivación

Este trabajo de fin de grado tiene como propósito predecir la evolución de pacientes de glaucoma medicados empleando técnicas de aprendizaje profundo (*Deep Learning*) para generar imágenes sintéticas de retinografía (Figura 1.1) o de células ganglionares (Figura 1.2) que representen una evolución de la enfermedad.

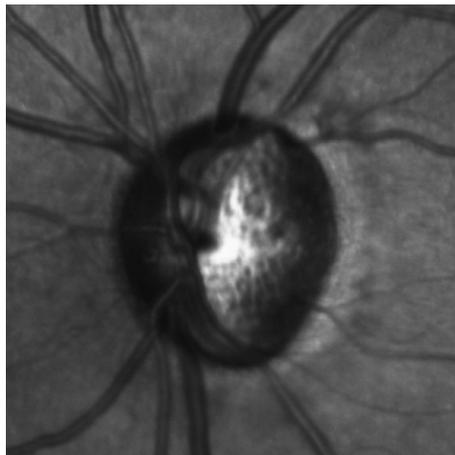


Figura 1.1: Ejemplo de retinografía proveniente de RIM-ONE-DL

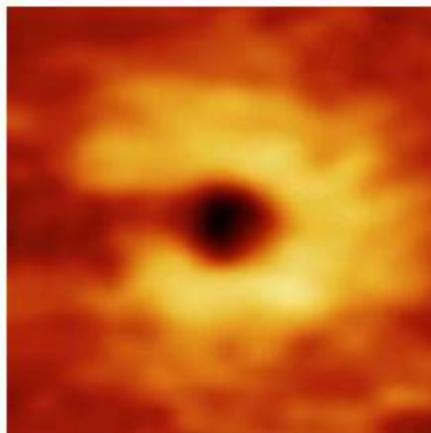


Figura 1.2: Ejemplo de mapa de células ganglionares

El glaucoma consiste en un conjunto de enfermedades que puede causar daños en el nervio óptico así como en la retina pudiendo generar pérdidas en el campo de visión.

Antes de describir la enfermedad, es necesario tener en cuenta ciertas partes de la anatomía ocular que serán de interés para comprender el glaucoma y el trabajo realizado.

- Humor acuoso. Se trata de una sustancia fluida dividida en dos cavidades, una de ellas situada frente al iris (sección coloreada del ojo) y la otra justo detrás.
- Mácula lútea. Zona ovalada amarilla localizada en el fondo del ojo cuyo centro es conocido como la fovea. Se especializa en la visión de detalles.
- Nervio óptico. Nervio encargado de transmitir la información visual desde la retina al cerebro.
- Retina. Fina capa que cubre la superficie del fondo del ojo.
- Papila óptica. Cavidad circular en la retina a través de la cual salen los axones de las células ganglionares de la retina que conforman el nervio óptico.

Existen dos tipos de glaucoma:

- Glaucoma de ángulo abierto: Es el tipo de glaucoma más común. El ángulo iridocorneal se mantiene normal, este contiene la vía trabecular a través de la cual se drena alrededor del 90 % del humor acuoso. Este tipo de glaucoma es menos abrupto y la pérdida de visión suele ocurrir sin que el paciente se percate.
- Glaucoma de ángulo cerrado: Se produce un decrecimiento del ángulo iridocorneal. Conlleva una elevación brusca de la presión intraocular, dolor intenso y disminución de la agudeza visual entre otros.

Debido a que el glaucoma de ángulo abierto es el más frecuente y habitualmente los pacientes no tienen constancia de que padecen la enfermedad, surge la necesidad de realizar diagnósticos a las personas que más posibilidades tengan de padecer la enfermedad con el fin de diagnosticarla lo antes posible. El diagnóstico del glaucoma es un campo de investigación bastante desarrollado por su alta relevancia. Aunque este trabajo no se centre en el diagnóstico, la línea de trabajo planteada permitiría a especialistas médicos estudiar, por ejemplo, cómo continuará progresando un paciente que se ha observado que su enfermedad progresa negativamente.

Este trabajo de fin de grado y la línea de trabajo que conlleva, nace con la motivación de facilitar a doctores, investigadores o cualquier interesado, el estudio de la evolución de un paciente de glaucoma. Serviría tanto con pacientes cuya enfermedad progresa negativamente, de tal manera que puedan analizar un posible estadio futuro del paciente, así como con pacientes cuya enfermedad ha parado de evolucionar gracias a la medicación. Sirviendo esto último de manera didáctica para aprender cómo hubiera avanzado la enfermedad si no se hubiese llegado a medicar correctamente.

## 1.2. Objetivos y fases

Los principales objetivos de este trabajo de fin de grado son:

- Realizar una investigación acerca de las redes generativas antagónicas.
- Lograr generar imágenes falsas que estén cercanas a la realidad por medio de redes generativas antagónicas.
- Generar imágenes en las que se demuestre una evolución temporal de la enfermedad.
- Conseguir imágenes de una resolución en la que se puedan diferenciar características a grandes rasgos.
- Recoger los resultados y realizar análisis de ellos.

Las principales fases que se han llevado a cabo han sido:

- Llevar a cabo una investigación inicial de los antecedentes de la línea de trabajo planteada y aprender acerca de las redes generativas antagónicas.
- Realizar pruebas con datos no relacionados con el glaucoma con el fin de encontrar arquitecturas de redes generativas antagónicas que puedan servir para pasar a imágenes de fondo de ojo.
- Estudiar redes Pix2Pix y realizar pruebas con datos de muestra.
- Establecer los pasos necesarios para preprocesar los datos tal y como se han empleado en la arquitectura estudiada.
- Probar la red Pix2Pix con los datos preprocesados y analizar los resultados iniciales.
- Realizar estudios con mayor profundidad de los resultados de manera cualitativa por expertos así como emplear métricas que expliquen los resultados obtenidos de forma cuantitativa.

## 1.3. Antecedentes y estado del arte

### 1.3.1. Antecedentes

Las Redes Generativas Antagónicas surgieron en 2014 (Goodfellow et al., 2014) y desde entonces, se han ido generalizando una serie de técnicas o pasos a seguir para lograr mejores resultados. En 2015 se presentan las DCGAN (Deep Convolutional Generative Adversarial Networks) Radford et al. (2015) obteniéndose una mejora en la similitud de las imágenes generadas respecto a las imágenes de la entrada de la red.

Debido al insurgente crecimiento de las GAN junto al desarrollo de nuevas técnicas relacionadas, han sido un gran foco para la investigación de los últimos años. Las redes generativas antagónicas han sido utilizadas en múltiples áreas de investigación relacionadas con el glaucoma:

- Diagnóstico de la enfermedad. Por la naturaleza del glaucoma, al ser difícilmente reconocible por el paciente de la constante pérdida de campo de visión, el diagnóstico es de enorme importancia así como lograr un diagnóstico lo más pronto posible. Han sido publicados en los últimos años múltiples artículos donde se busca un acercamiento a este problema utilizando redes generativas antagónicas de maneras muy diversas (Bisneto et al. (2020), Jain et al. (2022), Wang et al. (2019)).
- Síntesis de imágenes de ojos glaucomatosos. Las redes generativas antagónicas permiten la creación de imágenes sintéticas, lo cual ha sido probado para generar imágenes de fondo de ojo (Yoo et al. (2020), Diaz-Pinto et al. (2019), Yu (2019)) o mapas de vasos sanguíneos o ambos (Chen et al., 2021). Asimismo, las redes generativas antagónicas han sido probadas para sintetizar angiografías basadas en imágenes de fondo de ojo (Tavakkoli et al., 2020).
- Segmentación del disco óptico y la copa óptica. Una de las posibles maneras de diagnosticar glaucoma es comparando el tamaño del disco óptico con el de la copa óptica. La segmentación de estas partes supone despejar todos los detalles innecesarios (para esta tarea) que se encuentran en una imagen de fondo de ojo. Esto ha sido probado con redes generativas antagónicas, en específico, con cGANs (Liu et al., 2019) y mediante DCGANs (Jiang et al., 2019).
- Predicción de la evolución de un paciente de glaucoma. Se trata del mayor foco de investigación de este trabajo y en el que menos desarrollo se ha realizado. A pesar de que no haya ningún trabajo que utilice como entrada imágenes del fondo de un ojo glaucomatoso o mapas de células ganglionares para predecir su progresión en un intervalo de tiempo, en el artículo Hassan et al. (2020) se ha empleado una Red Generativa Adversativa Condicional con el mismo objetivo pero utilizando tomografías de coherencia óptica macular.

### 1.3.2. Estado del arte

El estado del arte acerca del uso de las redes generativas antagónicas para tareas relacionadas con el glaucoma, tanto de tareas de diagnóstico o síntesis como de segmentación, viene recogido en Qasim et al. (2021). En este artículo se realiza un estudio en el que tratan de buscar trabajos donde se mencione el glaucoma o términos relacionados así como técnicas de aprendizaje profundo (siendo las GAN la alternativa de mayor interés para este trabajo). Se logra obtener un diagrama con los trabajos encontrados en el cual tanto las cGAN como las DCGAN (explicadas posteriormente en el capítulo 2) destacan frente a otras variaciones de redes generativas antagónicas. En el estudio mencionado se plantean una serie de desafíos de las redes generativas antagónicas para detectar enfermedades así como los beneficios de usarlas y algunas recomendaciones para investigaciones futuras.

Las redes generativas antagónicas surgieron en 2014 siendo una arquitectura de aprendizaje profundo relativamente joven a la vez que un gran foco de investigación. Debido a la capacidad de crear datos sin diseñar específicamente una función de densidad de probabilidad, las GAN han captado la atención de investigadores de la imagen médica. En un estudio del estado del arte más reciente (Yi et al., 2019), se recogen las aplicaciones de las redes generativas antagónicas en cuanto a imágenes médicas: reconstrucción, síntesis de imágenes, segmentación, clasificación, registro y otros casos de uso.

Respecto a la predicción de la evolución de un problema médico ocular, el estado del arte está principalmente constituido por un *framework* denominado *Disease Forecast via Progression Learning* (DFPL) descrito en Wu et al. (2020). Este *framework* se basa en la irreversibilidad de la artrofia peripapilar (se trata de un síntoma común entre múltiples enfermedades oculares). Para ello, utilizan una red generativa antagónica que trata de predecir el grado de progresión de la enfermedad utilizando un acercamiento que, por su naturaleza, permite ser aplicado a varios tipos de enfermedades oculares.

En cuanto a la predicción de la evolución de ojos glaucomatosos, el principal trabajo encontrado presente en el estado del arte, se trata del artículo Hassan et al. (2020). Los resultados obtenidos sugieren que con solo dos visitas previas, podría ser suficiente para predecir la siguiente tomografía de coherencia óptica (TCO) del paciente seis meses después. Aunque este trabajo haya obtenido buenos resultados, la predicción de la evolución del glaucoma no está suficientemente explorada y se podrían encontrar resultados de gran ayuda para médicos. En la Figura 1.3 podemos observar una tomografía de coherencia óptica como las utilizadas en el artículo mencionado.

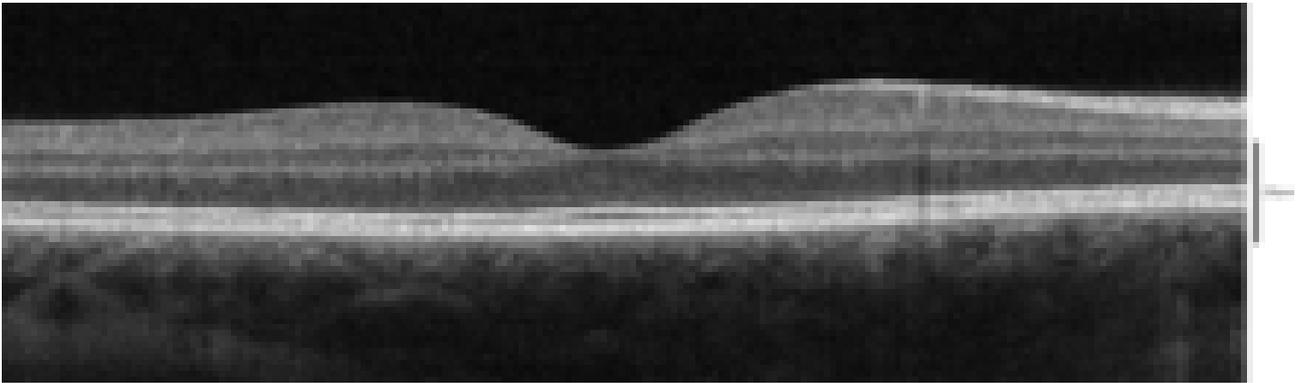


Figura 1.3: Tomografía de coherencia óptica de la mácula en el dominio espectral.

# Capítulo 2

## Redes Generativas Antagónicas

Este capítulo está destinado a explicar en detalle las redes generativas antagónicas (GAN de ahora en adelante) ya que para este trabajo de fin de grado se planteaba la exploración de estas redes con el fin de conocerlas, utilizarlas y aprender acerca de sus fortalezas y limitaciones.

Como se ha mencionado previamente, las GANs surgieron en 2014 planteándose como un marco para estimar modelos generativos a través de un proceso adversario. Estas redes consisten en entrenar simultáneamente dos modelos: un modelo generativo G que captura la distribución de los datos, y un modelo discriminatorio D que estima la probabilidad de que una muestra provenga de los datos de entrenamiento y no del modelo generativo G. El procedimiento de entrenamiento de G consiste en maximizar la probabilidad de que D cometa un error (Goodfellow et al., 2014). A grandes rasgos, el generador intenta engañar al discriminador mientras que el discriminador trata de diferenciar cada vez mejor entre muestras reales y generadas.

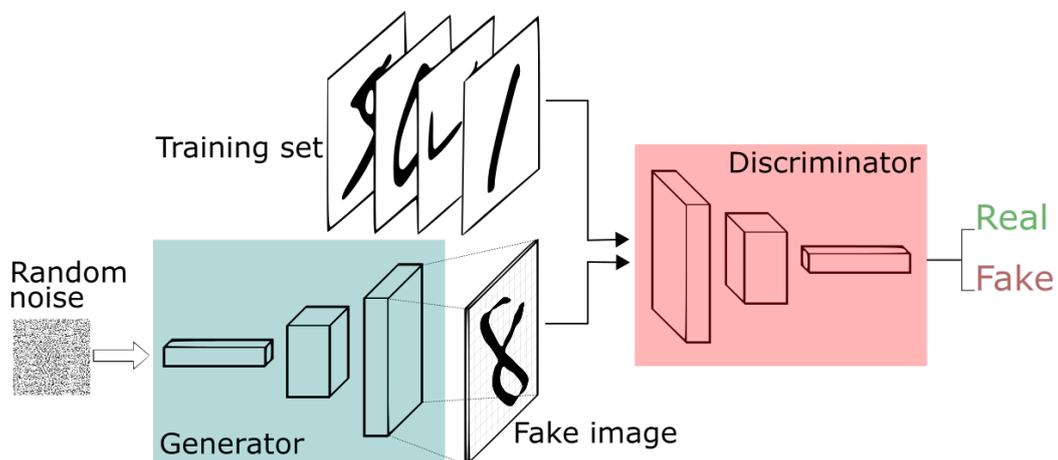


Figura 2.1: Arquitectura de una GAN. Imagen obtenida en Gen (2018)

En la Figura 2.1 se muestra como el modelo generativo G recibe ruido aleatorio con el que crea un resultado falso que aleatoriamente será entregado al modelo discriminatorio D. A pesar de que no se vaya a profundizar en esto, es importante saber que las GANs aprenden mediante una función de pérdida que se adapte a los datos, pueden por tanto aplicarse a una multitud de tareas que tradicionalmente requerirían tipos de funciones de pérdida muy diferentes.

El marco de modelización adversario es sencillo de aplicar cuando ambos modelos son perceptrones multicapa. Para aprender la distribución del generador  $p_g$  sobre los datos  $x$ , se define una prioridad sobre variables de ruido de entrada  $p_z(z)$ , representando un mapeo al espacio de datos como  $G(z; \theta_g)$ , donde  $G$  es una función diferenciable representada por un perceptrón multicapa con parámetros  $\theta_g$ . Se define también un segundo perceptrón multicapa  $D(x; \theta_d)$  cuya salida es un valor escalar.  $D(x)$  representa la probabilidad de que  $x$  haya venido de los datos en lugar de  $p_g$ . Se entrena  $D$  para que maximice la probabilidad de asignar correctamente la etiqueta tanto a los datos de entrenamiento como a las muestras creadas por  $G$ . Asimismo, se entrena  $G$  para minimizar  $\log(1 - D(G(z)))$ :

Es decir,  $D$  y  $G$  juegan el siguiente minimax de dos jugadores con función de valor  $V(G, D)$ :

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (2.1)$$

Uno de los posibles problemas que se podrían plantear al comprender el funcionamiento de las redes generativas antagónicas es acerca de cómo superan las primeras iteraciones en las que el generador  $G$  solo crea ruido aleatorio y por lo tanto el discriminador puede acertar con una enorme facilidad qué resultados provienen del conjunto de datos de entrenamiento y cuáles de  $p_g$ .

Analizando la Ecuación (2.1), se puede apreciar que el gradiente no sea suficiente para que  $G$  aprenda correctamente a crear resultados que logren convencer a  $D$  de que son reales. En los inicios del entrenamiento, cuando  $G$  es baja,  $D$  puede rechazar las muestras creadas con gran confianza debido a la enorme diferencia entre los datos de entrenamiento y los que provienen de  $p_g$ . En este caso,  $\log(1 - D(G(z)))$  se satura.

La solución planteada es entrenar  $G$  para que maximice  $\log D(G(z))$  en vez de minimizar  $\log(1 - D(G(z)))$ . Este cambio en la función objetivo permite la misma interacción entre  $G$  y  $D$ , pero se logran gradientes mucho más fuertes al principio del aprendizaje.

A pesar de que las GANs se propusieron como un marco innovador acompañado por resultados muy prometedores, con el paso del tiempo se han ido proponiendo modificaciones de la arquitectura mejorando la calidad o resolución de las imágenes generadas y creando nuevas posibilidades de uso. Estos son algunos tipos de GANs:

- GANs condicionales (Mirza and Osindero, 2014). Este tipo de variación, consiste en que el modelo generativo recibe datos que condicionan aquello que genere. Sirven por ejemplo para entrenar una GAN con imágenes de animales etiquetadas y luego inferir una imagen de una etiqueta en específico.
- DCGAN (Radford et al., 2015). Se proponen como una clase de redes convolucionales que funcionan correctamente en aprendizaje no supervisado. Se plantean una serie de requisitos y modificaciones respecto a las GAN iniciales para asegurar buenos resultados. Estos modelos han sido probados mientras se estudiaban las GAN.
- Traducción de imagen a imagen (Zhu et al., 2017). La traducción de imagen a imagen se puede definir como la tarea en la que se busca traducir una posible representación de una escena en otra, dada la información suficiente. Esta tarea puede se ha

planteado ser resuelta mediante redes generativas antagónicas. Este tipo de redes son las de mayor interés para este trabajo de fin de grado y se profundizará más en ellas posteriormente. Se trata de emplear GANs condicionadas a imágenes de entrada. Un ejemplo que se trabajó para comprender el software Pix2Pix propuesto en el artículo en el que se presentan estas redes fue el de traducción de imágenes de satélite a imágenes de Google Maps. En el desarrollo del proyecto se han empleado para generar imágenes futuras de un ojo glaucomatoso condicionadas a una imagen de un estadio previo.

- Síntesis de texto a imagen (Zhang et al., 2016). Estos modelos de GANs consisten en recibir texto como entrada y trata de generar una imagen que cumpla con la descripción dada. Específicamente se emplean *StackGAN (Stacked Generative Adversarial Networks)* con las que se generan imágenes de resolución 256x256 basadas en el texto dado. Para ello se usa primero una GAN en la primera fase que crea imágenes de baja resolución y con poco detalle a partir del texto, pero luego se usa otra GAN que recibe esta imagen de baja resolución y el texto de nuevo para aumentar su resolución y añadir detalles fotorrealistas.
- Superresolución (Ledig et al., 2016). Esta variación tiene como propósito aumentar la resolución de una imagen de entrada añadiendo detalles fotorrealistas y logrando crear imágenes casi indistinguibles de la real aumentando cuatro veces la resolución.
- CycleGAN (Zhu et al., 2017). Estos modelos consisten en tener dos conjuntos independientes de entrenamiento, por ejemplo, uno de caballos y otro de cebras. Lo que permite este tipo de modelo de GANs es convertir una imagen de un conjunto a otro en ambos sentidos. Un caso de uso muy interesante es pasar de fotografías a imágenes con el estilo de cuadros de Van Gogh.
- GANs progresivas (Karras et al., 2017). La primera capa del modelo generativo comienza creando imágenes de baja resolución y el resto de capas van añadiendo detalles y aumentando la resolución, logrando un entrenamiento más rápido y unas imágenes resultantes de mayor resolución. Uno de los problemas de las GANs convencionales es la baja resolución de las imágenes creadas.

# Capítulo 3

## Desarrollo del proyecto

### 3.1. Tecnologías utilizadas

Este trabajo requiere de la implementación de métodos basados en redes neuronales, esto principalmente se realiza en Python utilizando el conjunto de librerías TensorFlow y Keras o PyTorch y PyTorch Lightning. Debido a la experiencia previa con este lenguaje así como el soporte prestado a herramientas de aprendizaje profundo, el código desarrollado y utilizado es en el lenguaje Python. Los tres programas implementados para el preprocesamiento de las imágenes utilizadas fueron implementados en Python.

En cuanto a las librerías de implementación de redes neuronales, se ha utilizado *Tensorflow* junto a *Keras* ya que han sido utilizadas en proyectos previos. A pesar de que el entorno de librerías de *PyTorch* son una muy buena alternativa que permite implementaciones con mayor detalle y más bajo nivel, a principios del desarrollo aún no se habían utilizado para ningún proyecto.

En lo que se refiere a entrenamiento del modelo utilizado, se hizo uso de Google Colab utilizando ficheros IPYNB (*Interactive Python notebook*) y ejecutándolos en GPU en lugar de CPU. Este planteamiento tiene tanto ventajas como desventajas. Por un lado, ofrece las ventajas de un entorno de trabajo independiente del ordenador de trabajo utilizado, control de versiones y acceso sencillo al contenido de la cuenta de Drive de manera gratuita. Por otro lado, la principal desventaja es un mayor tiempo de cómputo (respecto a ejecución local con aceleración por GPU mediante tarjetas gráficas de alta gama) así como un tiempo de uso de GPU mensual limitado.

### 3.2. Entrenamiento de GAN simple con imágenes de ranas de CIFAR-10

Con motivo de estudiar el funcionamiento de las redes generativas antagónicas, se han leído las secciones en las que se hace una introducción al *Deep Learning* y aquellas en las que se habla de las GAN en Chollet (2017) y Goodfellow et al. (2016). Estos libros han servido como una introducción para entender el propósito de estas redes y profundizar teóricamente en su funcionamiento.

Para comenzar con una pequeña introducción en Tensorflow a las GAN, se han utilizado imágenes de ranas provistas por CIFAR-10 (Krizhevsky, 2009) junto a una GAN simple

que genere imágenes de resolución 32x32. En la Figura 3.1, se muestran imágenes de resolución 32x32 de las cuales las dos de la primera fila son reales y se han utilizado para entrenar la GAN. En la segunda fila, se muestran dos imágenes falsas creadas por el generador previamente entrenado. Ambas fueron creadas tras más de 9000 iteraciones de entrenamiento.

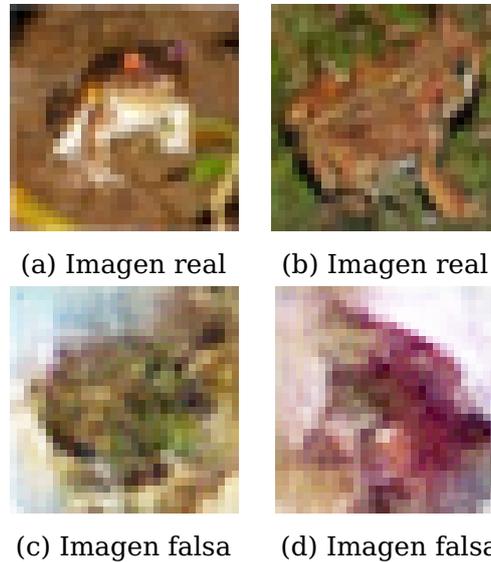


Figura 3.1: Imágenes reales y generadas de ranas en 32x32

Por otro lado, en la Figura 3.2 tenemos una imagen creada por el generador tras solo 100 iteraciones. Inicialmente el generador empieza creando imágenes que no se parecen a las que se quieren crear.

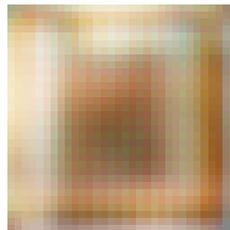


Figura 3.2: Imagen falsa generada en las primeras iteraciones

### 3.3. Entrenamiento de GAN y DCGAN con retinografías de RIM-ONE-DL

Se ha estudiado la posibilidad de trabajar con retinografías. Esta técnica consiste en tomar fotografías de fondo de ojo. Las imágenes están centradas en el disco óptico, lo cual consideramos que ayudaría ya que no son extremadamente diferentes entre ellas. En especial se han usado las imágenes del conjunto de datos RIM-ONE-DL (Batista et al., 2020) centradas en el nervio óptico. Este *dataset* está compuesto por 313 retinografías de sujetos normales y 172 retinografías de pacientes con glaucoma. De estas imágenes solo

serían utilizadas las de pacientes de glaucoma, ya que este trabajo consiste en predecir la evolución de un ojo glaucomatoso.

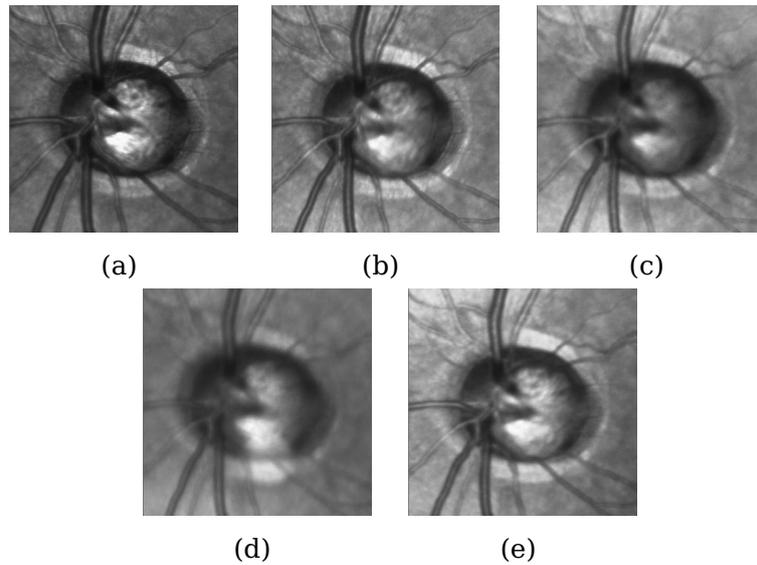


Figura 3.3: Imágenes de RIM-ONE-DL (Batista et al., 2020) de infrarrojo en escala de grises

Se han realizado intentos de generar imágenes sintéticas usando redes Generativas Antagónicas, pero sin lograr que la red convergiera y se obtengan imágenes que se asemejen a retinografías (Figura 3.4). Estos modelos requieren de muchas más imágenes para converger, esto empeora debido a la cantidad de detalle presente en estas imágenes. Por estas razones, solo las 172 imágenes disponibles de pacientes de glaucoma en RIM-ONE-DL no serían suficientes.

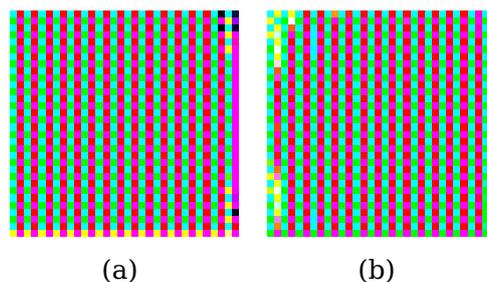


Figura 3.4: Intento de crear imágenes de fondo de ojo con resolución 32x32

Con el fin de lograr crear imágenes de fondo de ojo que se asemejen a la realidad, se han probado las DCGAN, las cuales consisten en unas pequeñas modificaciones a las GAN ya usadas con el fin de asegurar mejores resultados, estos cambios son:

- Reemplazar las capas de agrupamiento (*pooling layer*) por *strided convolutions* en el discriminador y *fractional-strided convolutions* en el caso del generador.
- Usar normalización en lotes tanto en el generador como el discriminador.

- Eliminar las capas ocultas completamente conectadas para arquitecturas más complejas.
- Usar activación *ReLU* en el generador para todas las capas excepto la de salida, para la cual se debe usar *Tanh*
- Usar activación *LeakyReLU* en todas las capas del discriminador.

Para comprender las DCGAN, se ha seguido un ejemplo que trataba los datos del *dataset* FASHION-MNIST (Maynard-Reid, 2021) con el que se ha aprendido cómo implementarlas y su funcionamiento. Una vez probado con los datos iniciales, se ha probado con las retinografías previamente mencionadas pero con resolución 28x28 ya que era la usada inicialmente por el ejemplo que se ha seguido. Los resultados se muestran en la Figura 3.5 y se asemejan a retinografías reales pero carecen de la resolución suficiente para apreciar detalles en la imagen.

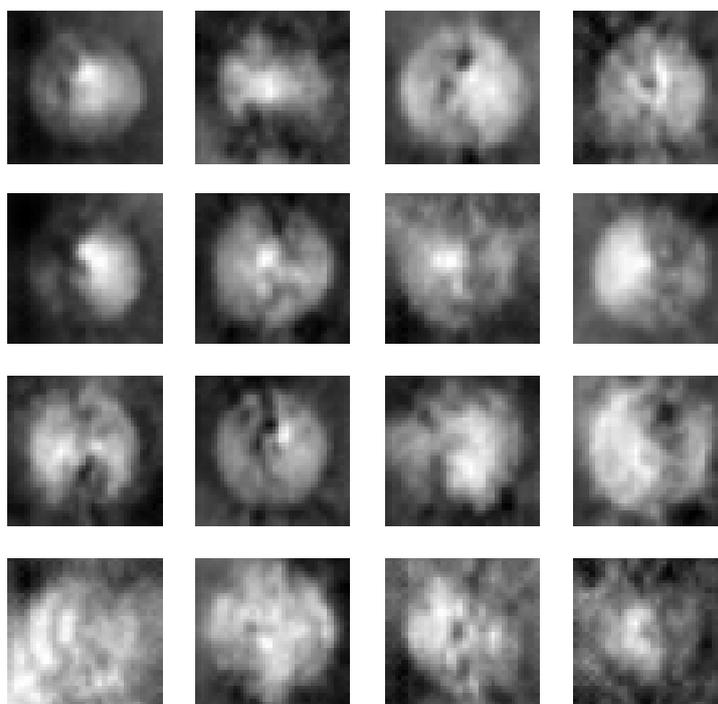


Figura 3.5: Conjunto de imágenes generadas por una DCGAN entrenada con retinografías

A pesar de haber generado imágenes que se asemejen con la realidad, es importante saber que los cambios en el nervio óptico no son fácilmente apreciables, causando que la evolución entre de un paciente sea muy difícilmente apreciable. A esto se le suma el hecho de que las DCGAN trabajan con imágenes de baja resolución, es una limitación común de las GAN.

Por estas razones, se ha optado por descartar las retinografías y emplear los mapas de células ganglionares con el fin de que se planteen diferencias notorias en la imagen producto de la evolución de un ojo glaucomatoso.

### 3.4. Modelo empleado para predecir la evolución de una imagen de un ojo glaucomatoso

Se ha empleado una red Pix2Pix (red traducción de imagen a imagen), la cual se ha descrito previamente como una CGAN (red generativa antagónica condicionada) en la que la condición es una imagen de entrada. Para entender el razonamiento detrás de usar este tipo de redes, se debe mantener en mente el objetivo de este trabajo. El objetivo es, dada una imagen de cualquier momento de un paciente de glaucoma medicado, tratar de predecir una imagen futura tras cierto intervalo de tiempo.

Las redes Pix2Pix (Isola et al., 2016), nos permiten lograr esto mismo, comenzando por entrenar la red con pares de imágenes donde una de ellas es la imagen de un estadio inicial y la otra del par corresponde a un momento posterior. Una vez entrenada la red, se consigue poder pasarle una imagen del mapa de células ganglionares de un paciente y obtener una predicción de cómo sería la imagen correspondiente a un estadio futuro de dicho paciente.

Estas redes proponen solución a grandes problemas de la traducción de imagen a imagen así como problemas de las redes generativas antagónicas:

- Resolver problemas que tradicionalmente requieren funciones de pérdida muy diferentes. Esto es logrado ya que las redes Pix2Pix no solo aprenden a realizar un mapeado de la imagen de entrada a la imagen resultante, sino que también aprenden una función de pérdida que sirve para entrenar este mapeado.
- Un problema de las redes generativas antagónicas es que requieren utilizar imágenes de muy baja resolución. En las redes Pix2Pix se utilizan redes PatchGAN (serán explicadas en detalle posteriormente), las cuales permiten que las imágenes de entrada y salida sean de resoluciones variables, funcionando correctamente con resoluciones tanto bajas como altas.
- Uno de los mayores problemas de las tareas de traducción de imagen a imagen es que mapean una entrada de alta resolución a una matriz resultante también de alta resolución. En la mayor parte de estas tareas, la entrada tiene una estructura similar a la imagen de salida. En base a esto, se utiliza una red en  $U$  ( $U$ -Net) como se puede ver en la Figura 3.6. Esta arquitectura contiene saltos entre las capas del decodificador y las del codificador, lo que permite pasar información entre la entrada y salida con el fin de que se pueda analizar la estructura o ciertas características encontradas en la imagen de entrada y utilizarlas para crear la imagen resultante.

#### 3.4.1. Función objetivo de las redes Pix2Pix

Las redes Pix2Pix, son redes generativas antagónicas condicionadas a una imagen inicial, en la Ecuación (3.1) podemos apreciar la función objetivo de una cGAN,  $G$  trata de minimizar la función objetivo frente a  $D$ , que trata de maximizarlo:  $G^* = \arg \min_G \max_D L_{cGAN}(G, D)$ .

$$L_{cGAN}(G, D) = E_{x,y}[\log D(x, y)] + E_{x,z}[\log(1 - D(x, G(x, z)))] \quad (3.1)$$

La función objetivo utilizada para las redes Pix2Pix se encuentra en la Ecuación (3.3) donde la distancia L1 viene descrita en la Ecuación (3.2).

$$L_{L1}(G) = E_{x,y,z}[||ky - G(x,z)k||_1] \quad (3.2)$$

$$G^* = arg \min_G \max_D L_{cGAN}(G, D) + \lambda L_{L1}(G) \quad (3.3)$$

El ruido aleatorio  $z$  desaparece en la Ecuación (3.3), aún así, la red es capaz de aprender un mapeado entre  $x$  e  $y$ , el problema es que los resultados serían deterministas (para una imagen de entrada siempre se generaría la misma imagen). A pesar de la ausencia de este ruido aleatorio, se utilizan capas *dropout* para aportar cierto nivel de aleatoriedad a las imágenes generadas.

### 3.4.2. Generador utilizando red en U

Las redes Pix2Pix utilizan modificaciones del generador y discriminador de las DCGAN Radford et al. (2015).

Lo más destacable del generador es el uso de una red en U (Ronneberger et al., 2015). A diferencia de una red codificador-decodificador, en las redes en U se forman conexiones entre las capas del codificador y el decodificador tal y como se puede apreciar en la Figura 3.6. Estas conexiones permiten que, una vez pasada la capa que forma el cuello de botella, se utilice información de bajo nivel al codificar la imagen creada.

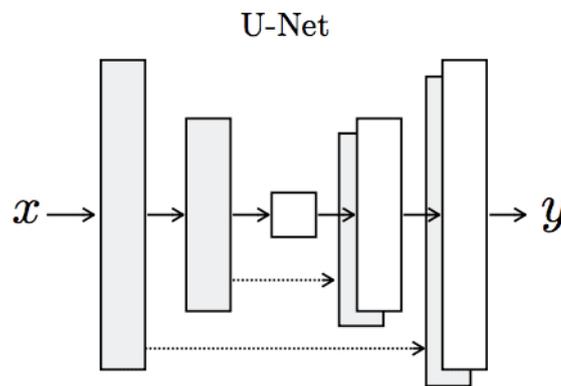


Figura 3.6: Boceto de red en U obtenido en uNe (2019)

Estas conexiones son entre las capas  $i$  y las  $n - 1$ , siendo  $n$  el número total de capas. Dichas conexiones concatenan todos los canales de la capa  $i$  con los de la capa  $n - 1$ .

Esto es implementado mediante las siguientes capas en la red neuronal utilizada para diseñar el generador:

- **Convolución bidimensional.** Esta capa aplica una operación llamada convolución que aplica un filtro a la entrada y produce un mapa de características. Este filtro es de menor tamaño que la entrada y para aplicarlo se utiliza el producto escalar. Esta

capa se utiliza con unos parámetros específicos de tal manera que el resultado tiene la mitad del tamaño de la entrada. Por ejemplo, siendo las imágenes de resolución  $256 \times 256$  y en color, el tamaño de la entrada es  $256 \times 256 \times 3$  (3 bandas de color en RGB) y por lo tanto, tras pasar por una convolución bidimensional, el mapa de características resultante tendría tamaño  $128 \times 128 \times 3$ .

- Leaky ReLU (Lea) y ReLU (ReL). Las funciones de activación ReLU son funciones lineales en la dimensión positiva y cero en la dimensión negativa. Las funciones de activación Leaky ReLU, en lugar de ser cero en la dimensión negativa, en la dimensión negativa tiene una pequeña pendiente.
- Normalización en lotes. Frente a una entrada de una capa intermedia de la red, normaliza los datos de entrada convirtiéndolos en un formato estándar. Ayuda a que las siguientes capas puedan procesar de manera más eficiente los datos.
- *Dropout*. Esta capa sirve para evitar sobreajuste a los datos de entrenamiento, causa que algunas neuronas sean quitadas u omitidas aleatoriamente.
- Concatenación. El proceso de concatenación de dos capas simplemente añade a una capa el contenido de otra. Esto es usado en el generador para añadir las conexiones de la red en U mencionadas previamente.
- Convolución bidimensional traspuesta. Sin entrar en detalle, su propósito es el contrario que la convolución bidimensional, es utilizada para aumentar el tamaño de la entrada. En este caso, se parametriza de tal manera que para una entrada de tamaño  $64 \times 64 \times 3$ , el resultado sería de tamaño  $128 \times 128 \times 3$ . La combinación de estas dos capas nos permite extraer las características de la imagen de entrada al implementarlas en forma de red en U, reduciendo el tamaño con convoluciones y aumentándolo para producir una imagen nueva mediante convoluciones traspuestas.

Con el fin de reproducir la arquitectura de una red en U, se crean bloques codificadores y decodificadores. Los bloques codificadores son utilizados en la primera parte de la red en U, permitiendo extraer las características de la imagen. Los decodificadores se implementan en la segunda parte para decodificar las características extraídas y crear una nueva imagen de igual tamaño; Se añaden conexiones desde los codificadores mediante concatenación.

Un bloque codificador está compuesto por una convolución bidimensional, una normalización en lotes (a excepción del primer codificador) y una función de activación Leaky ReLU. Por otro lado, un decodificador contiene una convolución bidimensional traspuesta seguida de una capa de normalización en lotes, *dropout*, la concatenación para añadir la conexión y por último, una activación ReLU.

### 3.4.3. Discriminador utilizando *PatchGAN*

La idea principal es que al usar la función de pérdida L1 (Ecuación (3.2)) se obtendrían resultados borrosos pero se capturarían con alta precisión las frecuencias bajas. Para modelar estructuras de altas frecuencias, se plantean las *PatchGAN*, las cuales solo penalizan la estructura en cada parche. El discriminador usando estas *PatchGAN* discrimina si cada parche de tamaño  $N \times N$  corresponde a una imagen real o a una falsa. Posteriormente

se obtiene el promedio de estas clasificaciones y será el resultado que produzca  $D$ , el discriminador.

El tamaño de estos mencionados parches, se ha probado variando desde  $1 \times 1$ , usándose PixelGANs, hasta  $286 \times 286$  lo cual corresponde al uso de ImageGANs. Las redes Pix2Pix utilizan PatchGANs de tamaño  $70 \times 70$ , siendo el tamaño con el que se obtuvieron mejor resultados. La mayor ventaja de dividir la imagen en parches y utilizar PatchGANs es la capacidad de poder utilizar imágenes de entrada de tamaños muy variables y con alta resolución. En específico, las imágenes utilizadas en este trabajo son de resolución  $256 \times 256$ . Se asume independencia entre los píxeles de parches distintos, clasificando cada uno por separado.

Para implementar las mencionadas PatchGAN del discriminador, se utiliza una sucesión de convoluciones bidimensionales, normalización en lotes (a excepción de la primera convolución) y LeakyReLU como función de activación. Estas capas han sido descritas brevemente en la sección 3.4.2.

### 3.5. Mapas de células ganglionares

Con el propósito de emplear imágenes en las que el cambio a lo largo del tiempo sea más notorio, aunque sea leve igualmente, se prueban los mapas de células ganglionares (GCL). Tanto los mapas de células ganglionares como los escaneos de sección transversal de la mácula, son tomados mediante una técnica llamada tomografía de coherencia óptica (OCT). Una tomografía de coherencia óptica ocular nos permite obtener mapas de células ganglionares (ejemplo mostrado en la Figura 1.2), mientras que una tomografía de coherencia óptica en el dominio espectral sirve para obtener escaneos de sección transversal de la mácula (ejemplo mostrado en la Figura 1.3).

Ambos tipos de imágenes sirven para analizar el estado de la mácula y la pérdida de fibras nerviosas. La mácula está compuesta por múltiples capas que se pueden diferenciar al analizar una OCT en el dominio espectral tal y como se muestra en la Figura 3.7. En el centro de la mácula se localiza la fovea.

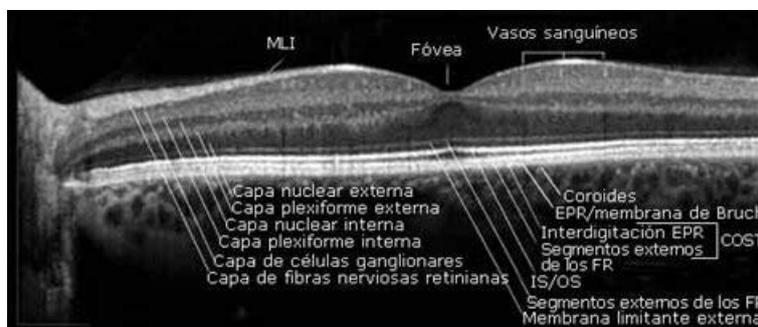


Figura 3.7: Capas y partes de la mácula en una OCT en dominio espectral. Imagen obtenida de Reyes (2014)

En la Figura 3.8 podemos apreciar un mapa de células ganglionares resaltando ciertas partes de interés. En el borde derecho, se encuentra la leyenda en la que se dispone de

colores más claros las partes de la GCL de mayor grosor y de colores más oscuros las de menor grosor. La fovea es de menor grosor que el resto y forma parte de la mácula. La mácula es de mayor tamaño, aproximadamente el área mostrada. Tras perder parte de las fibras nerviosas, el grosor de la mácula reduce y esto se ve reflejado en los mapas de células ganglionares con menos zonas de color amarillo. Asimismo, se muestra también el nervio óptico que será de interés para poder alinear horizontalmente el mapa de células ganglionares.

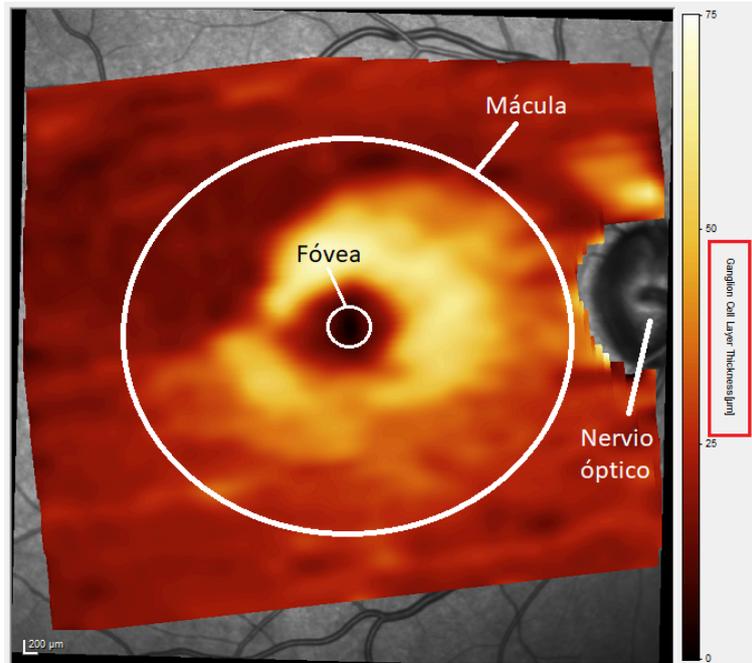
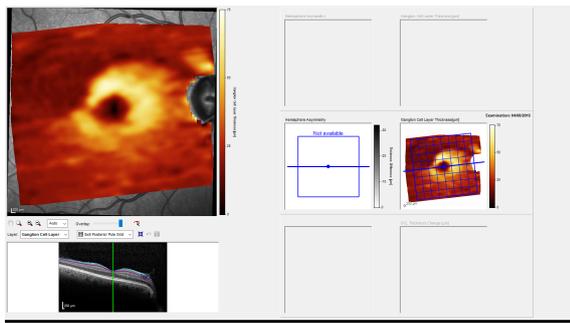


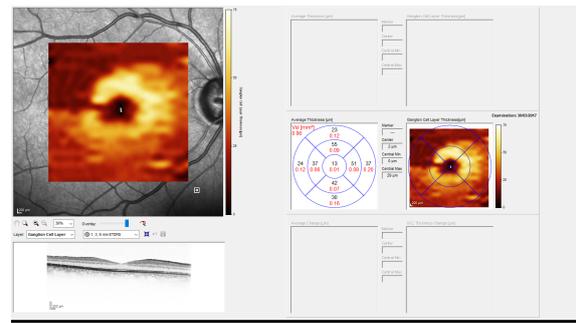
Figura 3.8: Imagen resaltando las partes de interés

### 3.6. Preprocesamiento de las imágenes de mapas de células ganglionares

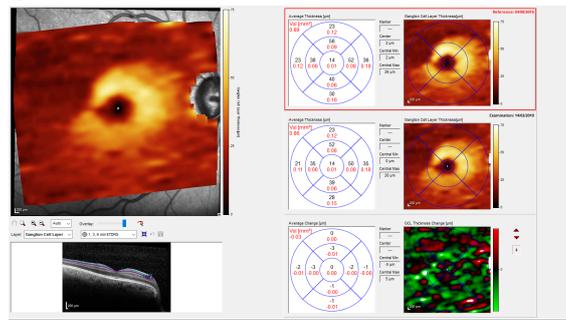
En la Figura 3.9, se pueden apreciar las imágenes iniciales que se quieren utilizar. Fueron proporcionadas por el hospital universitario y provienen del resultado de emplear un programa específico. En estas imágenes, tenemos en la zona superior izquierda un mapa de células ganglionares y en la parte inferior izquierda, se dispone un escaneo de sección transversal de la mácula. En la zona derecha de estas imágenes, se encuentra el proceso de alineación del mapa de células ganglionares con el nervio óptico así como un mapa calor.



(a) Mapa de células ganglionares rotado



(b) Mapa de células ganglionares alineado



(c) Mapa de células ganglionares rotado

Figura 3.9: Imágenes de mapas de células ganglionares iniciales

A pesar de que los escaneos de sección transversal de la mácula- sean utilizados en el estado del arte, se plantearon ciertos problemas al crear imágenes biológicamente plausibles. Hemos intentado probar los mapas de células ganglionares para comprobar su viabilidad. En la zona derecha de estas imágenes, se encuentra el proceso de alineación del mapa de células ganglionares con el nervio óptico así como un mapa de calor.

Para utilizar los mapas de células ganglionares con la red Pix2Pix, es preprocesar las imágenes dadas para que sean como la Figura 3.10, una imagen donde únicamente se encuentra el mapa de células ganglionares. Para ello, las imágenes deben estar horizontalmente rectas y recortadas adecuadamente. Una vez estén preprocesadas, podrán ser emparejadas entre ellas de distintas maneras, por ejemplo, pares de imágenes de un paciente entre las cuales ha transcurrido menos de un año.

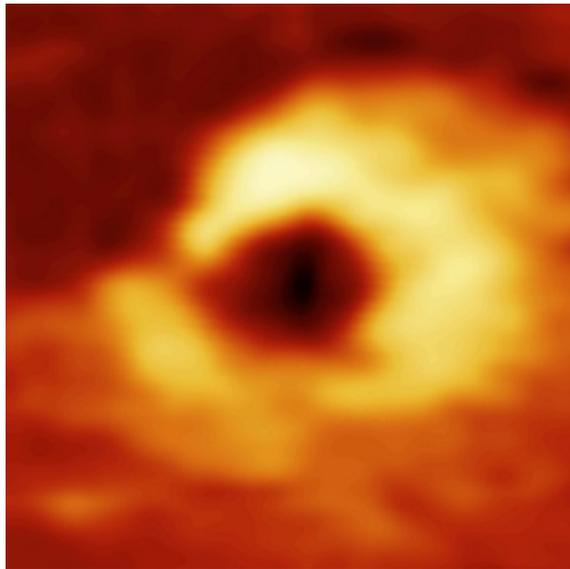


Figura 3.10: Figura 3.11 tras alinear y recortar

### 3.6.1. Recogida de datos de las imágenes para su posterior alineación

Hasta el momento se ha mencionado que se quieren alinear las imágenes pero no cómo. Para ello, podemos tomar como referencia el centro de la mácula, la fovea (la zona oscura rodeada por un disco amarillo) y el centro del nervio óptico, este se encuentra a la izquierda o derecha de la imagen dependiendo del ojo y está en escala de grises (en la figura 3.9a es el óvalo oscuro situado a la derecha de la mácula).

Teniendo el centro de la mácula y el del nervio óptico como referencias, se puede trazar una recta y luego alinear la imagen con el ángulo que forma esta recta respecto al eje horizontal. Esta sería la idea mediante la cual las imágenes son alineadas horizontalmente manteniendo la fovea como centro de la imagen.

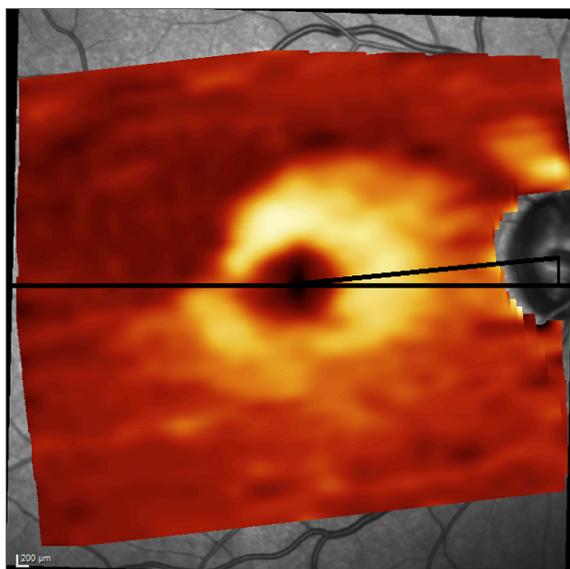


Figura 3.11: Alineación de la mácula con el nervio óptico

Para obtener estos datos se ha elaborado un programa simple con interfaz gráfica en el que se puede especificar un directorio e irá mostrando las imágenes (incluso de subdirectorios). El usuario tendrá que realizar tres clicks por cada imagen, el primero en el centro de la mácula, el segundo en el centro del nervio óptico y el último en el borde del nervio óptico. Los dos primeros clicks ya contienen la información suficiente para alinear posteriormente la imagen de manera horizontal y mantener la mácula centrada. Aún así, se solicita un tercer click que podría ser usado para disponer de un tamaño máximo de la imagen recortada. Ya que es un proceso manual y tedioso, es más conveniente disponer de todo lo necesario y no tener que repetir el proceso posteriormente.

### **3.6.2. Alineación de las imágenes**

Para alinear las imágenes se ha desarrollado un programa que recibe un JSON (la salida del programa anterior) y genera un árbol de directorios, como el original indicado en las rutas provistas en el fichero de entrada, con las imágenes alineadas y recortadas quedando solamente un cuadrado alrededor del centro de la mácula que contiene toda la información necesaria de la mácula.

El alineado se realiza como se ilustra en 3.11 y la rotación de la imagen se realiza respecto al centro de la mácula para no perder información en caso de que se hiciera respecto al centro de la imagen original. Las imágenes resultantes tienen tamaños distintos entre ellas aunque posteriormente sea necesario que todas tengan las mismas dimensiones. Esto se ha hecho con el fin de que la imagen final que contenga la mácula dependa de la resolución que tiene y no tenga un tamaño predefinido, ya que las imágenes de entrada tienen dimensiones muy dispares.

### **3.6.3. Escalado y emparejamiento de las imágenes**

Con el fin de resolver la diferencia en resoluciones en las imágenes de entrada, tras haberlas rotado y recortado proporcionalmente al tamaño de cada una, se redimensiona cada una de ellas. Inicialmente se buscaba el tamaño de imagen más pequeño y se establecía ese tamaño a todas las imágenes restantes.

Como ya se ha mencionado previamente, la red Pix2Pix tiene una red de tipo red en U. El número de capas de esta red implica el tamaño de las imágenes utilizadas, si la red en U tiene 8 capas en el decodificador y el codificador, las imágenes de entrada y salida serán de resolución  $2^8 \times 2^8$ , es decir,  $256 \times 256$ . Ya que inicialmente se probó el tamaño mínimo entre las imágenes disponibles después de ser recortadas, había que modificar el número de filtros en las capas de la red en U. El propósito de esta red es comenzar con un tamaño de entrada ( $256 \times 256$  en este caso), reducirlo hasta un punto mínimo dividiendo entre dos en cada capa y luego volver a aumentar, multiplicando por dos en cada capa, de nuevo este tamaño hasta volver al inicial; haciendo esto, se trata de generar una salida con las mismas características que la entrada inicial.

La modificación mencionada previamente reside en que, si tamaño inicial de las imágenes hubiera sido  $212 \times 212$  en lugar de  $256 \times 256$ , se empleaban ocho capas para un

tamaño de  $256 \times 256$  ( $2^8 = 256$ ). Ya que 212 es menor que 256, se necesitará un menor número de capas. El problema consiste en que, debido a que 212 no es potencia de dos, con ningún número de capas se consigue restaurar el tamaño inicial. Esto puede verse ilustrado en la Representación (3.4) donde se reduce a la mitad el tamaño de las imágenes en la parte del decodificador y se duplica en la parte del codificador.

$$212 \rightarrow 106 \rightarrow 53 \rightarrow 26 \rightarrow 13 \rightarrow 6 \rightarrow 3 \rightarrow 2 \rightarrow //1// \rightarrow 2 \rightarrow [4]!! \rightarrow 8 \rightarrow 16 \dots \quad (3.4)$$

Tras haberse encontrado que era imposible hacer uso de la resolución  $212 \times 212$ , se optó por reducir el tamaño a  $192 \times 192$  ya que  $2^7 = 192$  reduciendo por lo tanto el número de capas en ambas partes de la red en U en uno. Haciendo esto se logró que el modelo funcionase con los datos de entrada, pero al visualizar los resultados (3.12), simplemente se produce una pérdida de nitidez entre la imagen de origen y la generada, no se ve reflejado ningún indicio de evolución.

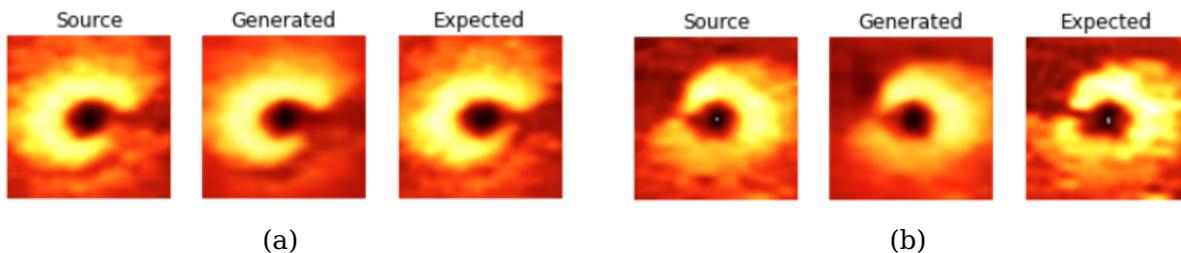


Figura 3.12: Resultados de entrenar la red Pix2Pix empleando imágenes de tamaño  $192 \times 192$  píxeles

Frente a este problema, se abandonó la idea de intentar siempre bajar la resolución de las imágenes y se establecieron todas a la resolución que se planteaba en la red original de  $256 \times 256$  obteniendo resultados notoriamente mejores y en los que se aprecia cierto grado de evolución entre la imagen fuente y la generada. Esto posiblemente se deba a que la arquitectura esté optimizada para esta resolución, pero los cambios aplicados teóricamente deberían ser válidos.

### 3.7. Diagrama del proceso completo

A continuación, en la Figura 3.13 se muestra un diagrama en el que se puede apreciar el proceso partiendo del conjunto de imágenes iniciales, sin preprocesar, hasta conseguir un conjunto de imágenes generadas por el modelo utilizado. Se encuentra la fase de preprocesamiento a partir de la cual se obtienen los mapas de células ganglionares emparejados entre sí. Posteriormente se separan las imágenes entre un conjunto de entrenamiento (70 % de los pares de imágenes) y un conjunto de testeo (30 % de los pares de imágenes). Utilizando el conjunto de entrenamiento, se entrena el modelo y, una vez entrenado, se infieren imágenes a partir de las del conjunto de testeo.

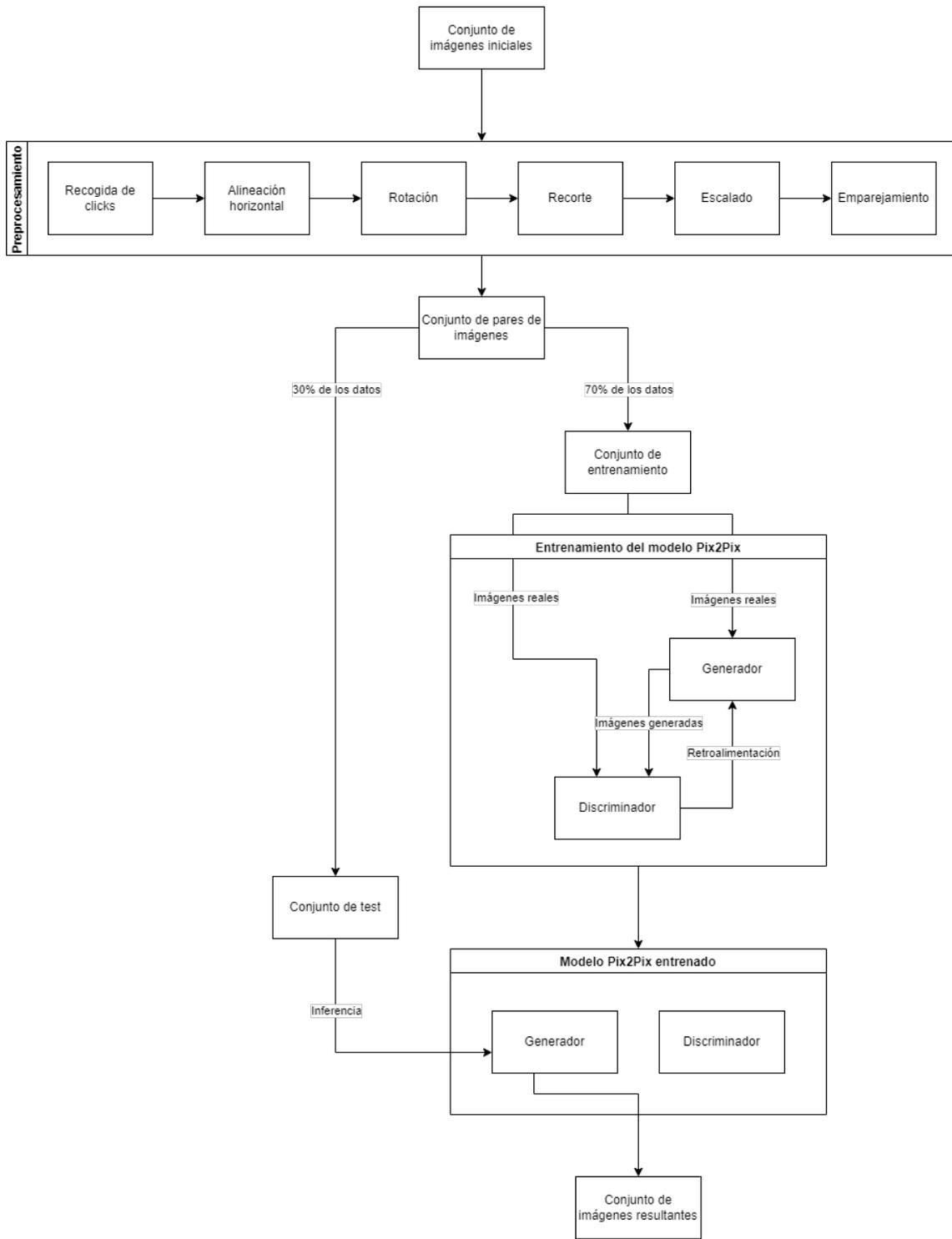


Figura 3.13: Diagrama del proceso desarrollado

# Capítulo 4

## Resultados

Los resultados en general han sido bastante prometedores. Se ha conseguido crear imágenes de mapas de células ganglionares que se asemejen en gran medida a imágenes reales. Esto se planteaba inicialmente como una posibilidad. Este pensamiento venía causado por la escasez de pacientes cuyos mapas de células ganglionares puedan ser usables. Las imágenes resultantes han sido tan parecidas a la realidad debido a la gran similitud en las imágenes del conjunto utilizado. Aunque haya diferentes patrones, en todas ellas se encuentra presente la misma gama de colores y la estructura característica de la mácula.

El modelo se ha entrenado con pares de imágenes del paciente de tres formas distintas: la primera y última imagen, pares de imágenes en los que hayan pasado menos de seis meses entre una muestra y la siguiente y por último, pares de imágenes donde haya pasado un año a lo sumo. Han sido utilizadas 406 imágenes de mapas de células ganglionares de 76 pacientes distintos. Una vez seleccionado el método de emparejamiento, los pares resultantes se dividen en un conjunto de entrenamiento (70%) y un conjunto de *test* (30%). Para el caso de la primera y última imagen, ya que solo se dispondría de un par de imágenes por cada paciente, el conjunto de *test* contendrá imágenes de pacientes que nunca han sido utilizados para el entrenamiento. En cuanto a los pares de como mucho 6 meses y como mucho un año, no se da el caso. Con alta probabilidad, pacientes utilizados para el entrenamiento serán utilizados para el *test*, aún así, ninguno de los pares utilizados en el entrenamiento será utilizado para generar el conjunto de imágenes falsas finales.

Se plantean una serie de experimentos distintos con el fin de tratar de determinar cuál es la mejor forma de emparejar las muestras de un paciente. El experimento donde se utiliza la primera y última imagen se ha planteado con fines didácticos. Nos permite estudiar los pares de imágenes donde más tiempo ha tenido la enfermedad de evolucionar, habiendo la mayor diferencia posible entre la imagen inicial y la final. Respecto a los otros dos experimentos, ya que desconocíamos cuál podría ser un intervalo adecuado para emparejar las imágenes, por esto han sido realizadas pruebas tanto con 6 meses de diferencia como un año de diferencia. Debido a que las muestras son principalmente tomadas espaciadas alrededor de estos intervalos de tiempo, escoger otros intervalos implica un menor número de pares por paciente así que han sido descartados.

Antes de mostrar los resultados obtenidos y analizarlos cuantitativamente, cabe destacar que las imágenes inferidas por la red han sido creadas solamente basadas en imágenes

previas sin utilizar datos clínicos del paciente, esto se plantea como una línea futura de este trabajo.

## 4.1. Análisis y comparación cuantitativa de las imágenes generadas

Una vez la red Pix2Pix utilizada fue entrenada, al inferir un mapa de células ganglionares basado en uno real del conjunto de prueba, fue necesaria la capacidad de poder evaluar el resultado obtenido comparándolo principalmente con el mapa de células ganglionares correspondiente.

Para realizar una comparación entre imágenes, una solución sencilla para medir la calidad de la imagen es utilizar el error cuadrático medio (ECM) entre la imagen real y la imagen creada por el generador. El error cuadrático medio de un estimador mide el promedio de los errores al cuadrado, es decir, la diferencia entre el estimador y lo que se estima (wik, 2020). Esta métrica nos permitiría calcular la diferencia entre una imagen y otra. En la Ecuación (4.1), tenemos el vector de predicciones  $\hat{Y}$  y el vector de valores  $Y$ . Es decir, el ECM es la media  $\frac{1}{n} \sum_{i=1}^n$  de los errores al cuadrado  $(\hat{Y}_i - Y_i)^2$ .

$$ECM = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \quad (4.1)$$

Ya que nos interesaría calcular el ECM entre imágenes, tendríamos que calcularlo para matrices. En la Ecuación (4.2) tenemos la ecuación en notación matricial, donde  $e_i$  corresponde a  $(\hat{Y}_i - Y_i)$  y  $e$  la matriz de tamaño  $nx1$ .

$$ECM = \frac{1}{n} \sum_{i=1}^n (e_i)^2 = \frac{1}{n} e^T e \quad (4.2)$$

El problema del error cuadrático medio es que compara imágenes calculando el error entre píxeles y, por la naturaleza de las imágenes de mapas de células ganglionares, la sección periférica de la imagen de color rojizo no es de gran interés y afectarían cambios sutiles en los píxeles al ECM. Frente a este problema, la métrica de la similitud estructural (SSIM) Wang et al. (2004) surge para tratar de resolver estos problemas y acercarse más a la forma en la que los humanos percibimos los colores. Esta métrica compara los patrones locales de la intensidad de los píxeles de los cuales la luminosidad y el contraste han sido previamente normalizados.

Dadas dos imágenes A y B, la formula de la similitud estructural píxel a píxel es:

$$SSIM(x, y) = (l(x, y))^\alpha (c(x, y))^\beta (s(x, y))^\gamma \quad (4.3)$$

En la Ecuación 4.3, tenemos tres componentes: la luminosidad  $l$ , el contraste  $c$  y un componente estructural  $s$ . Todos estos componentes reciben las imágenes A y B, asimismo, en todas las medias, varianzas y covarianzas se omite  $(x, y)$ . Cabe mencionar que los parámetros  $\alpha$   $\beta$  y  $\gamma$  serán mayor que 0.

$$l(x, y) = \frac{2\mu_A\mu_B + C_1}{\mu_A^2 + \mu_B^2 + C_1} \quad (4.4)$$

$$c(x, y) = \frac{2\delta_A\delta_B + C_2}{\delta_A^2 + \delta_B^2 + C_2} \quad (4.5)$$

$$s(x, y) = \frac{2\delta_A\delta_B + C_2}{\delta_A^2 + \delta_B^2 + C_2} \quad (4.6)$$

En las Ecuaciones 4.4, 4.5 y 4.6 se describen el componente de la luminosidad, el del contraste y el componente estructural respectivamente. Por otro lado,  $C_1 = (K_1L)^2$ ,  $C_2 = (K_2L)^2$  y  $C_3 = \frac{C_2}{2}$  siendo  $L = 2^8 - 1$  para imágenes de componentes de 8 bits y  $K_1 = 0,01$  mientras que  $K_2 = 0,03$ . Conociendo la similitud estructural píxel a píxel, la similitud estructural media entre dos imágenes A y B sería:

$$MSSIM(A, B) = \frac{1}{wh} \sum_x \sum_y SSIM(x, y) \quad (4.7)$$

La Ecuación (4.7) es la forma utilizada en este trabajo para calcular la similitud entre imágenes. Era importante seleccionar una métrica adecuada para comprobar cuantitativamente la diferencia entre las imágenes falsas y reales (principalmente). También nos sirve para comprobar la diferencia entre la imagen inicial y aquella tras pasar un intervalo de tiempo (ambas imágenes reales de un paciente), así como comparar la imagen falsa generada por la red con la que correspondía a la evolución del paciente.

## 4.2. Análisis cuantitativo de los resultados

Con el fin de separar las secciones de una imagen, la similitud estructural ha sido calculada dividiendo cada imagen en una matriz de tamaño 3x3 y luego se comparan cada uno de los nueve sectores con los nueve sectores de otra imagen (uno a uno). Vamos a centrarnos en la similitud entre las imágenes finales, la imagen real final y la que ha sido generada por la red. Los otros dos pares posibles de imágenes (las dos imágenes reales o la imagen real inicial con la imagen generada) nos servirían para estudiarlos en conjunto. De todos modos, la información que contienen es redundante y podemos observar directamente los resultados. Con analizar si la imagen generada se parece a la imagen final, los otros dos pares posibles no aportarían nada más. El caso sería distinto si hubiera pasos intermedios en la evolución de un paciente o en la generación de la imagen falsa.

En el anexo A podemos encontrar las tablas de similitud estructural de cada experimento. Dichas tablas contienen los cuantiles 0, 0.25, 0.5, 0.75 y 1 de los resultados de calcular la similitud estructural entre los nueve sectores (de la subdivisión 3x3) entre dos imágenes. Estos cuantiles nos permiten analizar el mínimo (cuantil 0), la mediana de los sectores (cuantil 0.5) y el máximo (cuantil 1).

Podemos apreciar como la mediana de gran parte de las imágenes se encuentra entre 0.7 y 0.85, habiendo pocos casos en los que se salga de este intervalo. Existen múltiples

imágenes donde el mínimo se encuentra por debajo de 0.6, por sí mismo esto no plantea un problema ya que corresponde a solo un sector de la imagen, es posible que el resto de sectores se parezcan mayoritariamente a la imagen final del paciente. Es interesante destacar que el valor máximo no parece bajar de 0.7, siendo un valor relativamente alto de similitud estructural. Esto se puede deber al parecido entre las imágenes utilizadas, donde en todas ellas la fovea es la parte central y representa una región más oscura.

Por otro lado, hemos calculado la media de los sectores para obtener un único valor de similitud estructural por cada par de imágenes. Podemos observar el comportamiento de estos datos en forma de histograma.

Cabe destacar que, por la naturaleza de los experimentos, no todos tienen el mismo número de pares de imágenes. El experimento de la primera y última imagen de cada paciente tiene tantos pares como pacientes (el conjunto de testeo tiene 21 imágenes, se puede ver en la Tabla A.1). Entre el caso de hasta seis meses de diferencia, el conjunto de testeo contiene 25 imágenes (como se puede apreciar en la Tabla A.1). Por último, el experimento de hasta un año de diferencia es el que más pares de imágenes dispone y un 30 % de ellas pertenecen al conjunto de *test*, siendo 91 imágenes (ver Tablas A.3, A.4 y A.5).

A pesar de la diferencia de tamaño del conjunto de *test* de cada experimento, podemos observar un comportamiento bastante similar entre los histogramas 4.1, 4.2 y 4.3. En los tres de ellos, la mayor parte de los datos se encuentra entre valores de 0.75 y 0.85 de similitud estructural promedio. Habiendo casos específicos con similitud estructural alrededor de 0.65 y muy pocos con similitud estructural promedio menor que esto.

Podemos concluir que en los tres experimentos se obtuvieron buenos resultados en cuanto a similitud estructural.

Por otro lado, para obtener valores únicos que representen a cada experimento, se ha obtenido una media y una desviación estándar de los valores de similitud estructural promedio de cada imagen:

- Primera y última imagen de cada paciente. Media: 0.7717 Desviación estándar: 0.0564
- Hasta seis meses de diferencia entre las imágenes del par. Media: 0.7656 Desviación estándar: 0.0518
- Hasta un año de diferencia entre las imágenes del par. Media: 0.7808 Desviación estándar: 0.0747

Los tres experimentos llevados a cabo llegaron a resultados muy similares tanto en cuanto a la media como la desviación estándar. La media más alta corresponde al caso de un año de diferencia mientras que la desviación estándar más baja pertenece al experimento de seis meses de diferencia. Cabe recordar que nos hemos centrado en las imágenes finales: la imagen final del par utilizado y la imagen que ha sido generada por la red.

Al igual que previamente, los resultados son muy similares entre sí. No hay un caso que sea claramente mejor que el resto. Aún así, es seguro asumir que los tres han logrado

buenos resultados en cuanto a similitud estructural. Parece que los valores de SSIM mayor a 0.9 son muy difíciles de lograr incluso si la imagen parece lo suficiente parecida. Asimismo, se ha observado que casos de SSIM menor a 0.5 han sido prácticamente nulos.

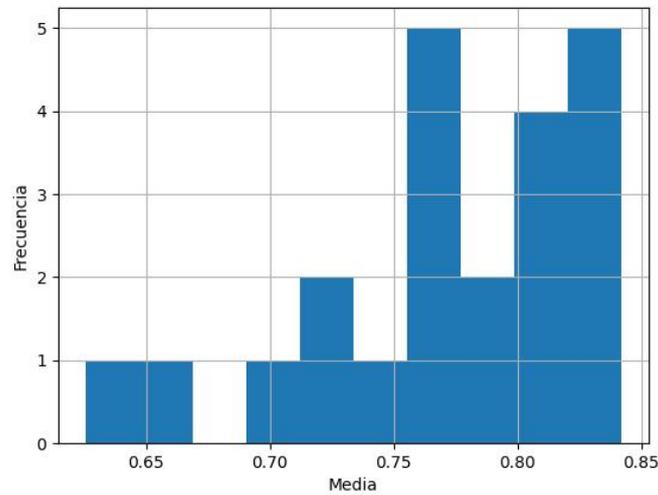


Figura 4.1: Histograma de las medias de SSIM entre sectores. Pares de primera y última muestra de cada paciente

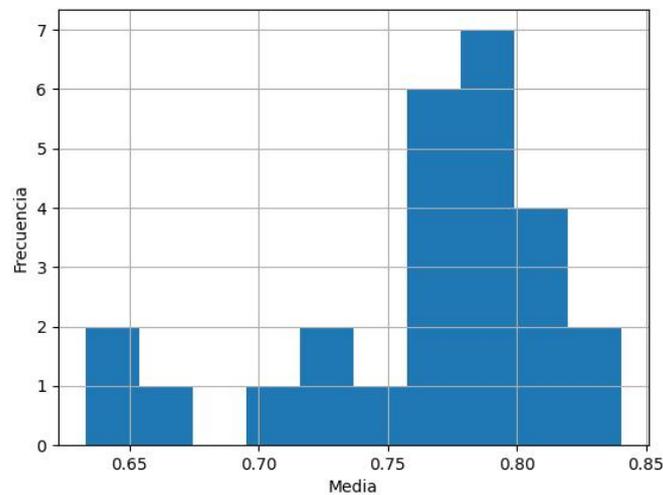


Figura 4.2: Histograma de las medias de SSIM entre sectores. Pares de imágenes con hasta seis meses de diferencia

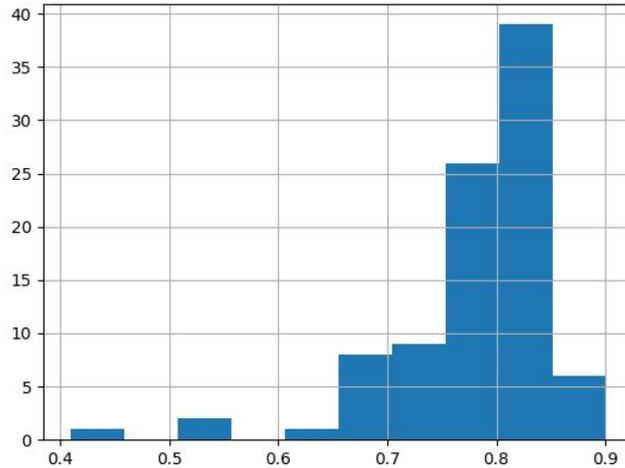


Figura 4.3: Histograma de las medias de SSIM entre sectores. Pares de imágenes con hasta un año de diferencia

### 4.3. Análisis cualitativo de imágenes generadas

Esta parte de la memoria sirve para comprender mejor las imágenes utilizadas, los mapas de células ganglionares, así como apreciar y entender los resultados de similitud estructural obtenidos mediante algunos casos concretos en la sección 4.3.2. Asimismo, se ha realizado un análisis experto por un colaborador externo en el apartado 4.3.1 cuyo principal propósito es estudiar cuáles serían los puntos de interés de un mapa de células ganglionares así como la viabilidad de utilizar estas imágenes que han sido generadas mediante el modelo utilizado.

En los apartados 4.3.1 y 4.3.2, todas las imágenes utilizadas corresponden al entrenamiento donde los pares de imágenes estaban formados por la primera y última muestra de cada paciente. Por esto, todas las tripletas de imágenes (dos imágenes reales y una generada) estarán constituidas por: la primera toma de mapa de células ganglionares del paciente, la última y la imagen generada por el generador G condicionado a la mencionada imagen inicial del paciente en concreto.

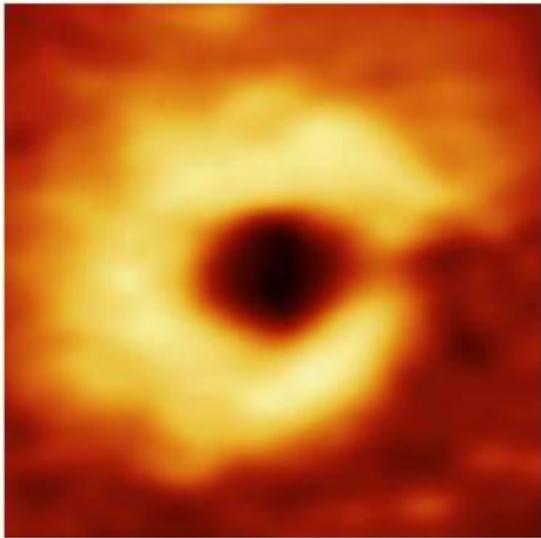
#### 4.3.1. Análisis cualitativo experto

Con el fin de analizar cualitativamente la generación de imágenes, han sido estudiadas y descritas por un colaborador externo, un doctor experto en las técnicas utilizadas en este trabajo (imágenes de fondo de ojo y mapas de células ganglionares). Este pequeño experimento ha sido especialmente útil para entender mejor los mapas de células ganglionares y cuáles serían las partes en las que un experto interesado se fijaría. Para este análisis, se utilizaron las imágenes generadas tras entrenar al modelo con los pares de imágenes de la primera y última muestra de cada paciente.

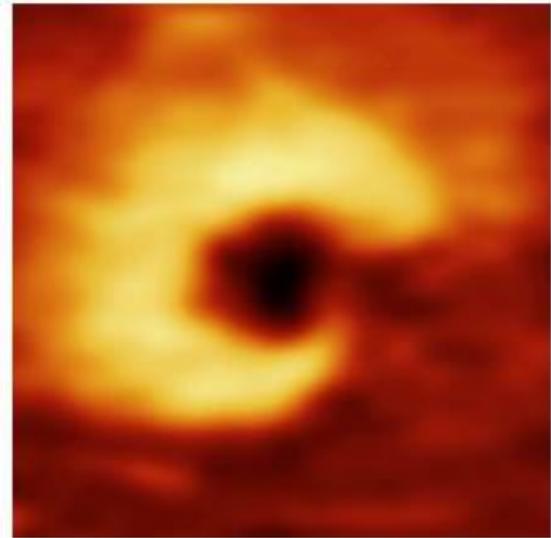
En las descripciones recibidas, se menciona una parte temporal de la imagen, esto se refiere a una separación vertical donde el lado del ojo más cercano a la nariz sería la parte

nasal y el lado más cercano a los huesos temporales del cráneo es referido como temporal. Asimismo, una pérdida de GCL se ve reflejada en el mapa de células ganglionares al pasar una zona de color amarillo a tonos más rojizos, perdiendo grosor.

### Pares de imágenes reales

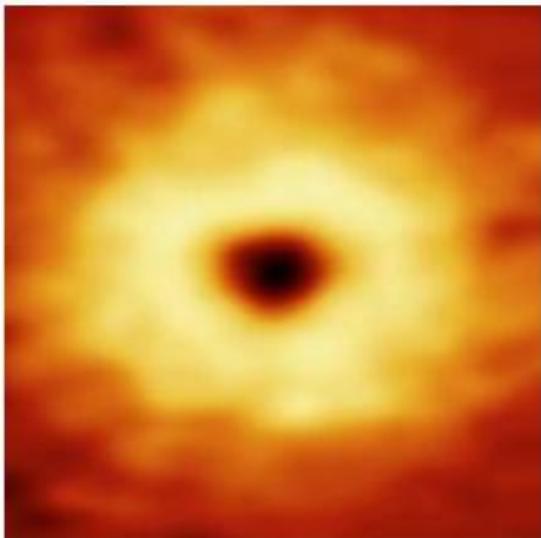


(a) imagen real inicial

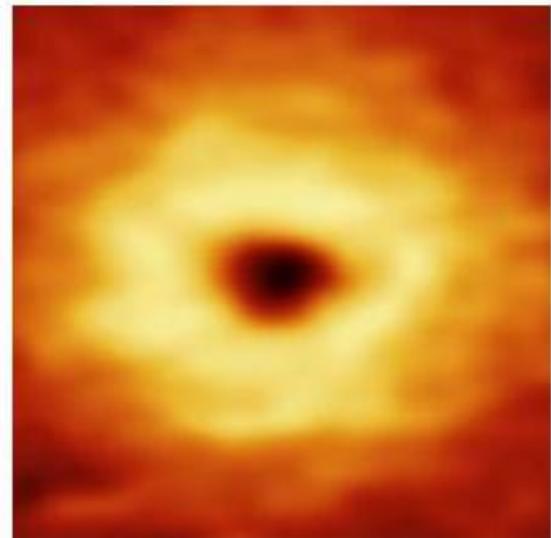


(b) imagen real final

Figura 4.4: "Aumento en extensión y profundidad de la lesión temporal inferior"

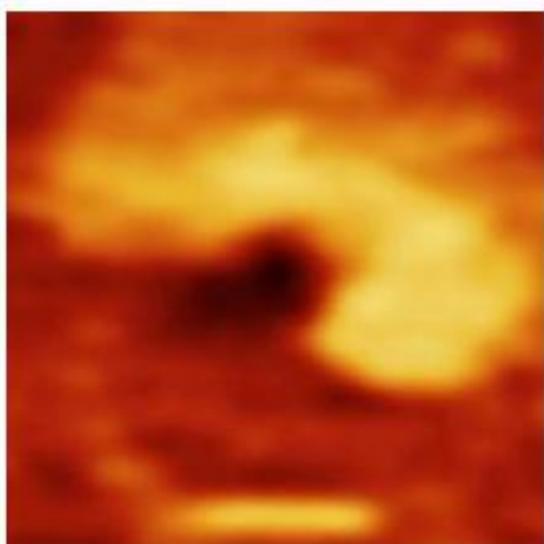


(a) imagen real inicial

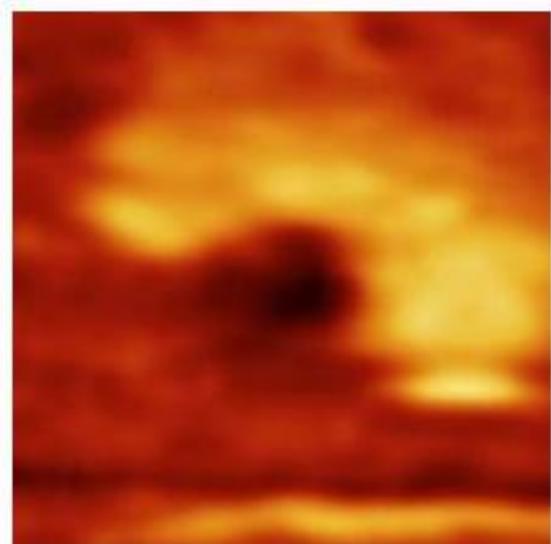


(b) imagen real final

Figura 4.5: "Cambios leves a nivel superior de la GCL"



(a) imagen real inicial



(b) imagen real final

Figura 4.6: “Pérdida de la GCL a nivel temporal inferior y a nivel superior”

En el caso de la Figura 4.6, es interesante mencionar el artefacto producido por el aparato de medición en la parte inferior de la imagen final 4.6b, el cual no es presente al inferir una imagen del modelo condicionada a la imagen inicial del paciente, tal y como se puede ver en la Figura 4.7.

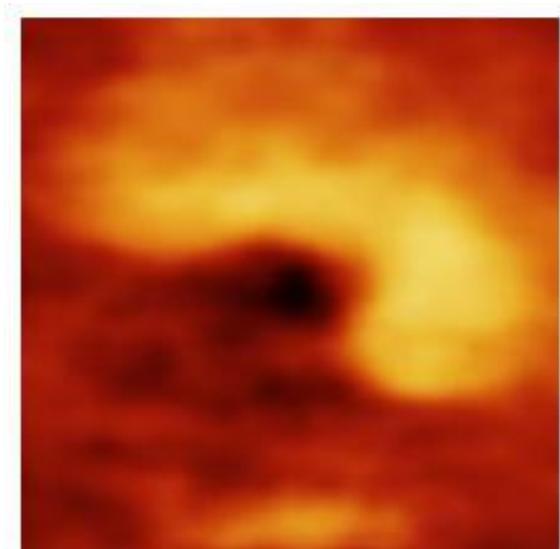
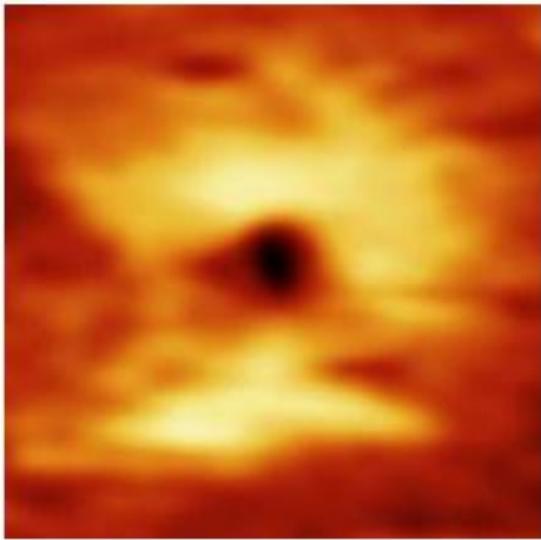
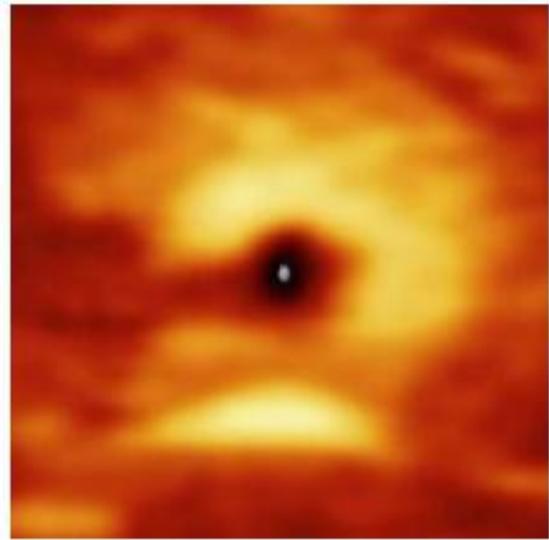


Figura 4.7: Imagen generada por la red basada en la imagen 4.6a



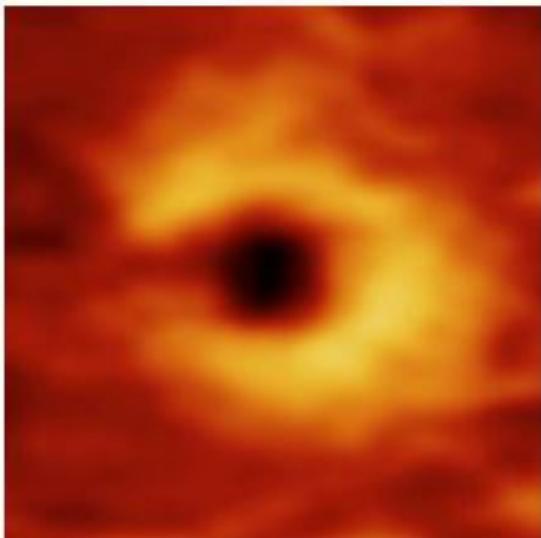
(a) imagen real inicial



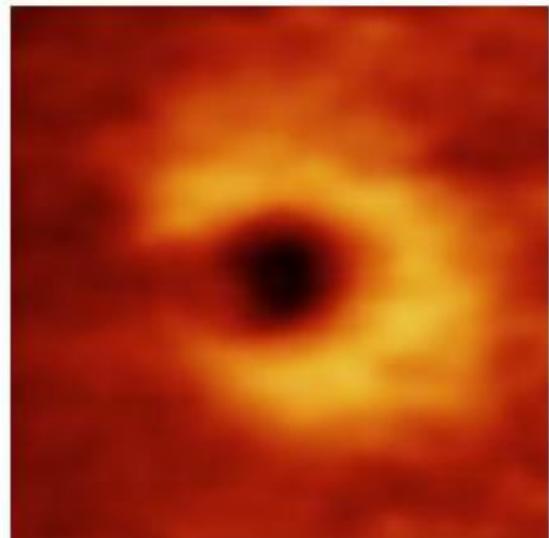
(b) imagen real final

Figura 4.8: “Disminución difusa de la GCL”

#### **Pares de imágenes de la imagen inicial real y la imagen generada**



(a) imagen real inicial



(b) imagen generada

Figura 4.9: “Disminución generalizada de la GCL.”

Cabe destacar que en la imagen original correspondiente a la imagen final del par del paciente de la Figura 4.9, el aparato de medición produjo un artefacto que empeora mucho la calidad general de la imagen (4.10). Aparentemente, estos artefactos, a pesar de que ocurran en la realidad, debido a su infrecuencia, la red no ha aprendido a generarlos. Esto provoca que el doctor opine que, para estos casos, la imagen generada por la red es de mayor utilidad que la real, ya que no presenta dicho artefacto.

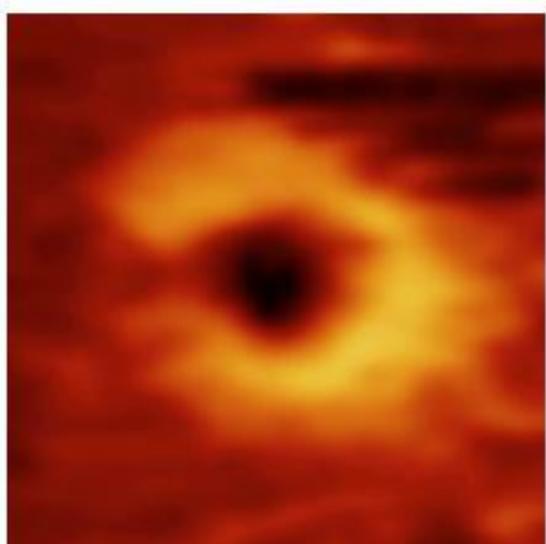
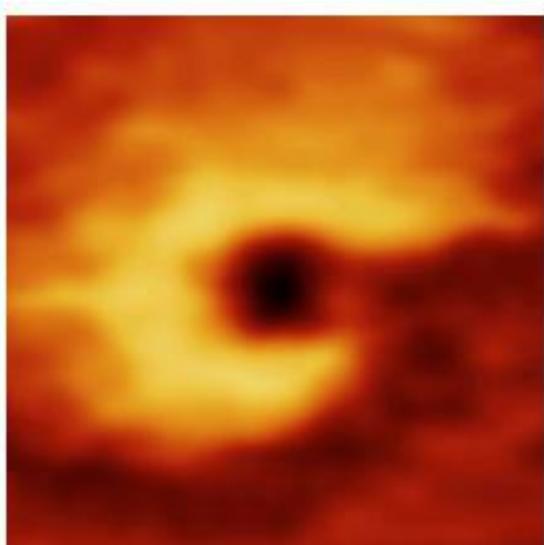
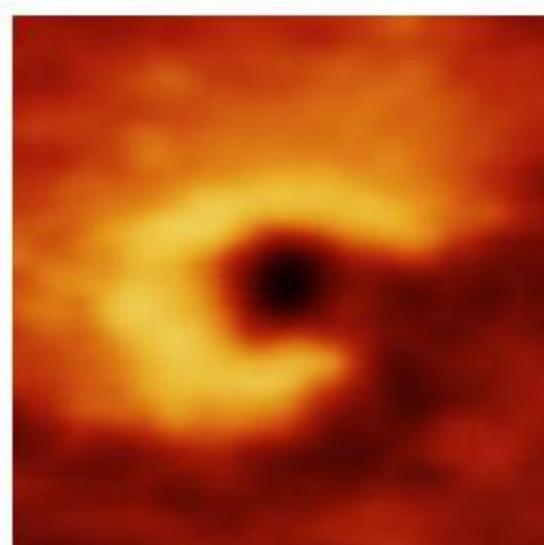


Figura 4.10: Imagen final del paciente, siendo la imagen inicial: 4.9a



(a) imagen real inicial



(b) imagen generada

Figura 4.11: “Reducción de la GCL a nivel temporal y temporal inferior.”

### 4.3.2. Tripletas de imágenes con mapas de calor de la similitud estructural

A continuación se presentan ejemplos de pares de imágenes de las tres siguientes posibles combinaciones:

- Imágenes reales. La imagen inicial junto a la imagen final del paciente.
- Imagen inicial con la imagen generada.
- Imagen generada y la imagen final.

Junto a estos pares de imágenes, con el objetivo de analizar los resultados de similitud estructural obtenidos, se han realizado mapas de calor con el valor de SSIM correspondiente al sector de las subdivisiones de tamaño 3x3. Haber separado las imágenes en sectores nos permite fijarnos en partes específicas de la imagen y estudiar el comportamiento de la similitud estructural.

En las Figuras 4.12, 4.14 y 4.13 tenemos un caso en el que la progresión de la enfermedad es muy lenta, las tres imágenes son muy similares entre sí. Centrando la atención en las imágenes reales, podemos apreciar que todos los valores de similitud estructural son mayores que 0.8, excepto uno con valor de 0.74, que sería el sector en el que más diferencia hay entre las imágenes. Esto último generalmente está relacionado con aquellas partes en las que es más visible el efecto degenerativo del progreso de la enfermedad.

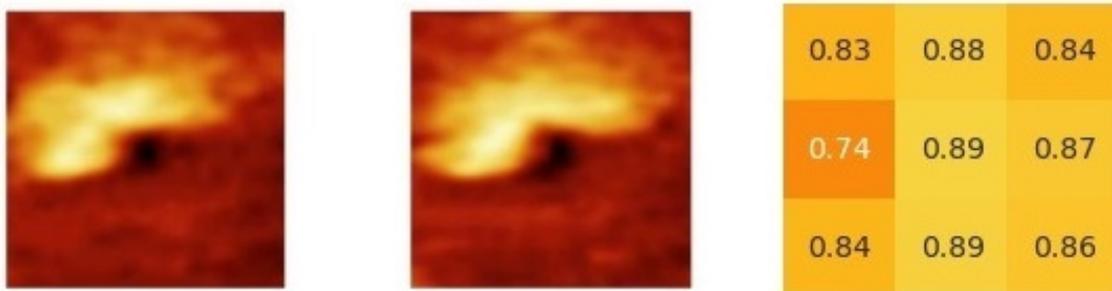


Figura 4.12: Par de imágenes reales junto al mapa de calor de SSIM (1)

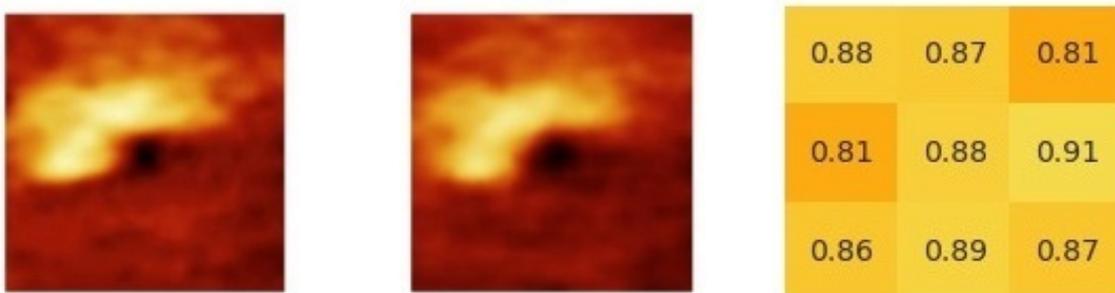


Figura 4.13: Par de imagen inicial con la imagen generada junto al mapa de calor de SSIM (1)

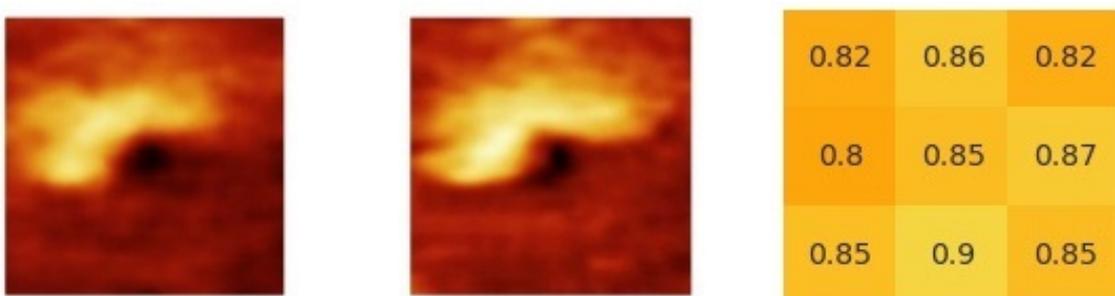


Figura 4.14: Par de imagen generada y la imagen final del paciente junto al mapa de calor de SSIM (1)

En contra parte al grupo de imágenes precedente, las Figuras 4.15, 4.17 y 4.16 corresponden a un paciente en el que la enfermedad ha progresado de manera más acelerada. Esto se puede apreciar en una pérdida difusa de la GCL, ya que la zona de color amarillento se ve reducida en tamaño desde muchas partes.

Lo más destacable de este paciente, además de presentar un caso de mayor progresión, podemos localizar un artefacto producido por el aparato de medición en la imagen final (segunda imagen de la Figura 4.15) localizado en la parte superior derecha, que no ha sido generado por la red (segunda imagen de la Figura 4.16). Esto produce que la imagen generada sea mucho mejor para analizar el caso del paciente, como bien nos fue informado por el colaborador externo mencionado en la sección 4.3.1.

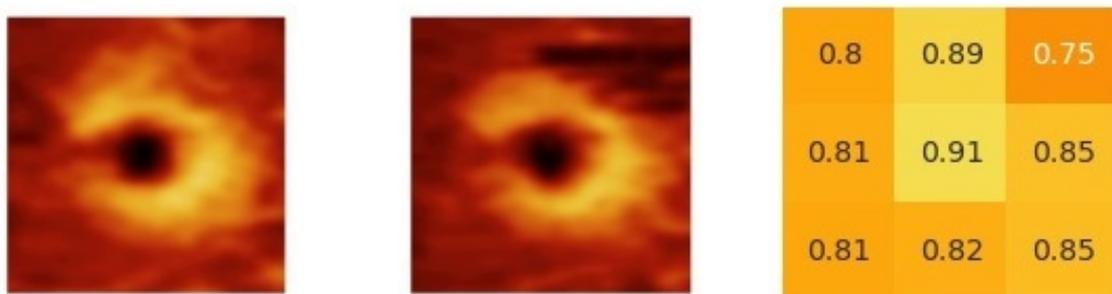


Figura 4.15: Par de imágenes reales junto al mapa de calor de SSIM (2)

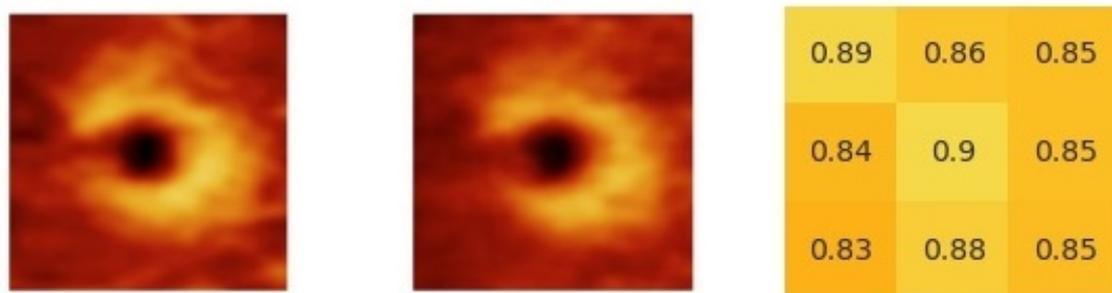


Figura 4.16: Par de imagen inicial con la imagen generada junto al mapa de calor de SSIM (2)

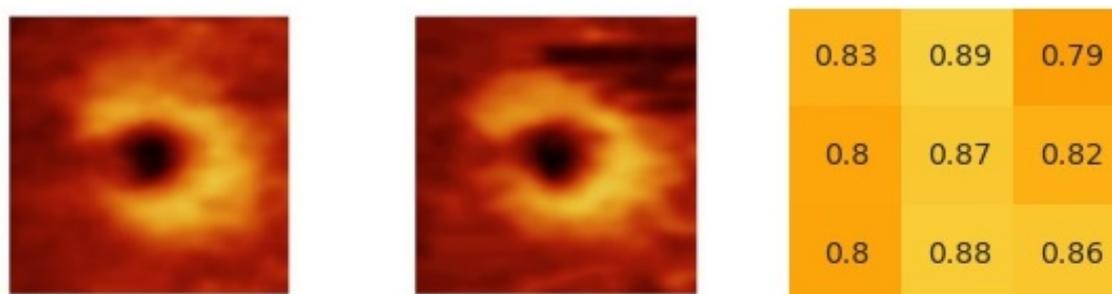


Figura 4.17: Par de imagen generada y la imagen final del paciente junto al mapa de calor de SSIM (2)

El siguiente conjunto de imágenes corresponde a un paciente de progresión rápida, al igual que el anterior. A lo largo de las Figuras 4.18, 4.20 y 4.19 podemos ver sectores de la subdivisión 3x3 con similitud estructural menor a 0.8, especialmente en la segunda fila de la primera columna y la tercera de la tercera columna. En estas zonas parece haber mayor diferencia, aún así, a simple vista, la parte inferior derecha de las imágenes es muy parecida. Estos casos han generado dudas acerca del uso de la similitud estructural para este trabajo y se plantea como línea futura la posibilidad de utilizar métricas alternativas.

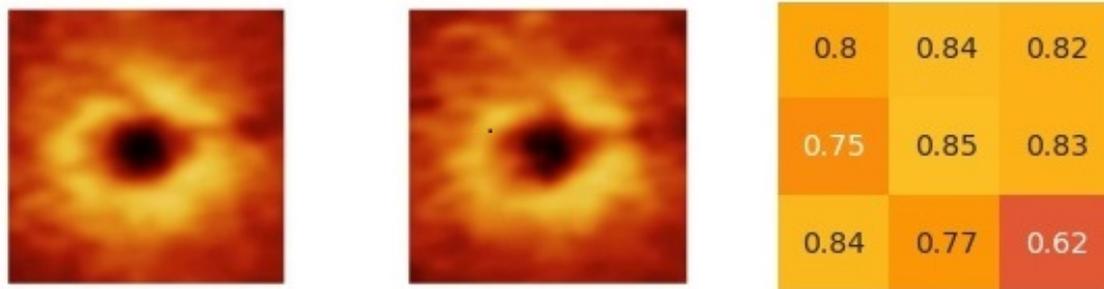


Figura 4.18: Par de imágenes reales junto al mapa de calor de SSIM (3)

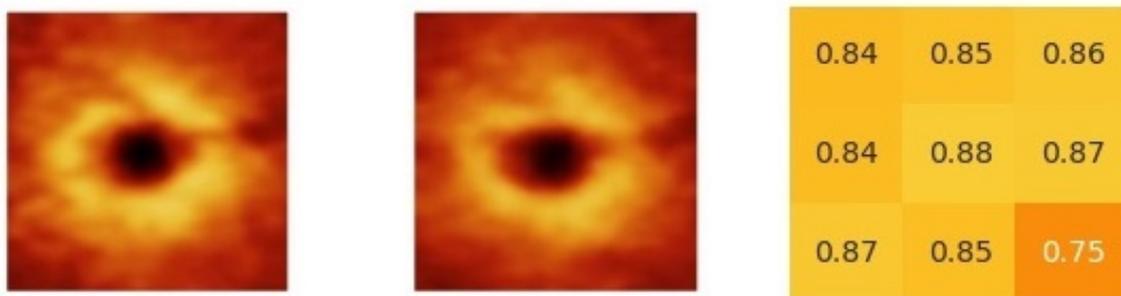


Figura 4.19: Par de imagen inicial con la imagen generada junto al mapa de calor de SSIM (3)

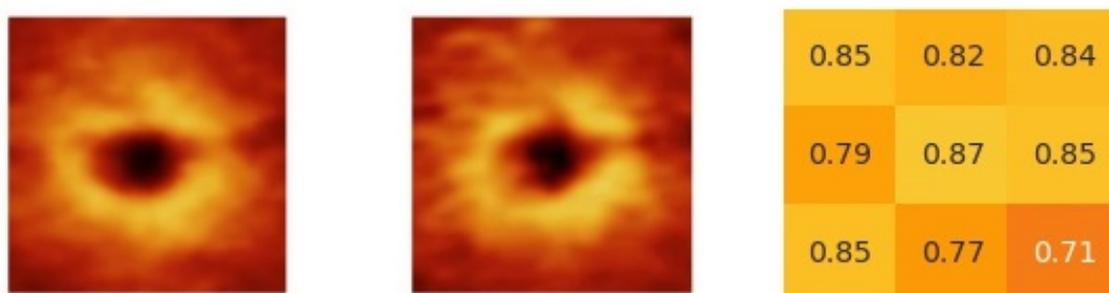


Figura 4.20: Par de imagen generada y la imagen final del paciente junto al mapa de calor de SSIM (3)

## 4.4. Experimento de clasificación entre imágenes falsas y reales

Los mapas de células ganglionares creados por la red pix2pix implementada no presentan claros signos de ser imágenes falsas. Para tratar de corroborar esto, se ha llevado a cabo un experimento con el colaborador experto mencionado previamente. Dicho experimento consiste en una colección de imágenes reales y falsas mezcladas y sin etiquetar para que el doctor las clasifique y estudiar los resultados.

Un conjunto de 44 imágenes fueron entregadas (donde 22 de ellas eran reales y 22 falsas). Las imágenes reales pertenecen al conjunto de entrenamiento utilizado para los pares de imágenes de la primera y última muestra de cada paciente mientras que las falsas, pertenecen al conjunto de test de dicho entrenamiento. Todas las imágenes entregadas pertenecen a pacientes diferentes.

		Etiquetas		
		Real	Falsa	
Predicciones	Real	17	14	31
	Falsa	5	8	13
		22	22	44

Tabla 4.1: Matriz de confusión de los resultados obtenidos en el experimento de clasificación

Analizando la matriz de confusión dispuesta en la Tabla 4.1, podemos extraer fácilmente los siguientes datos:

- Han sido clasificadas 31 imágenes como reales. 70 %, 31/44
- Han sido clasificadas 13 imágenes como falsas. 30 %, 13/44
- Falsos positivos: 14 imágenes han sido clasificadas como falsas y eran reales. 45 %, 14/31
- Falsos negativos: 5 imágenes han sido clasificadas como reales y eran falsas. 38 %, 5/13
- Aciertos positivos: 17 imágenes han sido clasificadas como reales y eran reales. 55 %, 17/31
- Aciertos negativos: 8 imágenes han sido clasificadas como falsas y eran falsas. 62 %, 8/13
- Aciertos totales: 25 imágenes han sido clasificadas correctamente. 57 %, 25/44
- Fallos totales: 19 imágenes han sido clasificadas erróneamente. 43 %, 19/44

Estos datos son bastante favorables para la realidad de las imágenes generadas, ha habido un 57% de aciertos, lo cual está muy cerca de una clasificación aleatoria. Asimismo se aprecia que muchas más imágenes han sido clasificadas como reales, un 70%, esto implica que muchas imágenes falsas han sido interpretadas como reales (Han ocurrido 14 falsos positivos, un 31% de las imágenes).

# Capítulo 5

## Conclusiones y líneas futuras

### 5.1. Conclusiones

A lo largo de esta memoria, se ha hablado del Glaucoma junto a la posibilidad de predecir cómo evoluciona un ojo glaucomatoso utilizando redes generativas antagónicas. Se ha hablado de los pasos recorridos para la selección de un método y un conjunto de imágenes adecuadas para esta tarea: la utilización de redes Pix2Pix, redes generativas antagónicas condicionadas a imágenes que sirven para realizar tareas de traducción de imagen a imagen y, los mapas de células ganglionares, imágenes con tonos amarillentos y rojizos que representan el grosor de la mácula.

Tras introducir el problema y hablar del desarrollo llevado a cabo, incluyendo las explicaciones necesarias para comprender el contexto, se describen los resultados obtenidos mediante tablas de similitud estructural de los pares de imágenes. A parte de estos datos, se analizan casos individuales de pacientes de glaucoma junto al mapa de calor de SSIM correspondiente. Por último se realizan dos experimentos con un colaborador externo experto: el primero consiste en describir las imágenes observadas con el fin de comprender en qué partes se fijaría mientras que, el otro experimento, se trata de una clasificación entre imágenes reales y falsas.

Las redes generativas antagónicas han sido sorprendentes, desde los primeros momentos en los que se probaron GAN sencillas o DCGANs para generar imágenes artificiales de ranas o retinografías hasta las redes Pix2Pix por la capacidad que tienen de elaborar imágenes que sean fotorrealistas. No solo estas imágenes tienen el fotorrealismo necesario para engañar a un usuario corriente, también se logra con expertos como se ha observado en el experimento llevado a cabo en la sección 4.3.1.

Asimismo, los resultados han sido muy prometedores y llaman a la creación de múltiples posibles líneas de futuro. Este trabajo se planteaba como un estudio de la viabilidad de generar imágenes de ojos glaucomatosos artificialmente. Teniendo en cuenta la reducida muestra que ha estado disponible, 77 pacientes con un número variable de muestras cada uno, obtener imágenes que se asemejen a la realidad ya es un gran logro.

### 5.2. Líneas futuras

Se plantean líneas futuras muy diversas, tanto para probar nuevos métodos como para mejorar el trabajo actual y expandirlo:

- Buscar una alternativa a la similitud estructural para comparar la calidad de las imágenes generadas. Se encontraron algunos casos en los que la similitud estructural no reflejaba la similitud a simple vista entre mapas de células ganglionares, por esto se plantea la posibilidad de diseñar o utilizar técnicas alternativas para comparar imágenes.
- Probar a utilizar escaneos de sección transversal de la mácula al igual que en Hassan et al. (2020) y comparar los resultados obtenidos.
- Probar con imágenes de campos de visión en lugar de mapas de células ganglionares. Un campo de visión ayuda a identificar el estado de un paciente de glaucoma, al igual que un mapa de células ganglionares, por lo tanto, se plantea seguir el mismo proceso pero utilizando como entrada campos de visión.
- Generar mapas de células ganglionares o campos de visión a partir de retinografías. Los aparatos requeridos para obtener mapas de células ganglionares o campos de visión son de un coste mucho mayor que los empleados para obtener retinografías. Lograr esto sería posible si la información necesaria para recrearlos estuviera reflejada en una retinografía, aún así, esto se desconoce por el momento.
- Repetir los entrenamientos cuando haya un mayor número de imágenes disponibles. Sabiendo que los resultados han sido positivos dado el reducido conjunto de imágenes, el método debería ser probado cuando el número de individuos y muestras aumente.
- Clasificar el progreso de evolución de un ojo glaucomatoso en grupos que correspondan a posibles patrones de evolución. Esta posibilidad de trabajo parece estar ya estudiada, pero es interesante la combinación de estos patrones de evolución con cGANs.
- Condicionar la creación de imágenes a datos clínicos del paciente. Actualmente las imágenes son generadas basadas en la información existente en la imagen dada, los resultados podrían ser más fiables si se condiciona la generación de imágenes a datos clínicos que influyan en la enfermedad (por ejemplo, la edad).

# Capítulo 6

## Conclusion and future lines of work

### 6.1. Conclusions

Throughout this report, Glaucoma has been discussed along with the possibility of predicting how a glaucomatous eye evolves using antagonistic generative networks. The steps taken to select a method and a set of images suitable for this task have been discussed: the use of Pix2Pix networks, antagonistic generative networks conditioned on images that serve to perform image-to-image translation tasks, and ganglion cell maps, images with yellowish and reddish tones that represent the thickness of the macula.

After introducing the problem and discussing the development carried out, including the explanations necessary to understand the context, the results obtained are described by means of structural similarity tables of the image pairs. Apart from these data, individual cases of glaucoma patients are analyzed together with the corresponding SSIM heat map. Finally, two experiments are performed with an external expert collaborator: the first one consists in describing the observed images in order to understand in which parts it would fixate while the other experiment is a classification between real and fake images.

Antagonistic generative networks have been amazing, from the early times when simple GANs or DCGANs were tested to generate artificial images of frogs or retinographs to Pix2Pix networks for their ability to produce images that are photorealistic. Not only do these images have the photorealism necessary to fool an ordinary user, it is also achieved with experts as observed in the experiment conducted in section 4.3.1.

Also, the results have been very promising and call for the creation of multiple possible lines for the future. This work was intended as a study of the feasibility of imaging glaucomatous eyes artificially. Considering the small sample that has been available, 77 patients with a variable number of samples each, obtaining images that resemble reality is already a great achievement.

### 6.2. Future lines of work

A wide range of future directions are planned, both to test new methods and to improve as well as expand on current work:

- Search for an alternative to structural similarity to compare the quality of the generated images. Some cases were found where the structural similarity did not

reflect the naked eye similarity between ganglion cell maps, hence the possibility of designing or using alternative techniques to compare images.

- Try using cross-sectional scans of the macula as in Hassan et al. (2020) and compare the results obtained.
- Test with images of fields of view instead of ganglion cell maps. A field of vision helps to identify the status of a glaucoma patient, as does a ganglion cell map, therefore, we propose to follow the same process but using fields of vision as input.
- To generate ganglion cell maps or fields of vision from retinographs. The devices required to obtain ganglion cell maps or fields of vision are much more expensive than those used to obtain retinographs. Achieving this would be possible if the information needed to recreate them were reflected in a retinogram, yet this is unknown at this time.
- Repeat the training when a greater number of images are available. Knowing that the results have been positive given the small set of images, the method should be tested when the number of individuals and samples increases.
- Classify the progression of a glaucomatous eye into groups corresponding to possible progression patterns. This possibility of work seems to be already studied, but combining these evolution patterns with cGANs is interesting.
- Condition the creation of images to clinical data of the patient. Currently, images are generated based on existing information in the given image, results could be more reliable by conditioning image generation to clinical data influencing the disease (e.g., age).

# Capítulo 7

## Presupuesto

En este capítulo se hablará brevemente del presupuesto que conlleva este trabajo.

Para empezar, deben ser contadas las horas de trabajo. Con el fin de saber cuál sería un coste adecuado por hora, se ha consultado el salario promedio por hora de un ingeniero de aprendizaje profundo en España, que se encuentra alrededor de 20 euros/hora.

El entrenamiento del modelo implementado fue realizado en Google Colab con una cuenta institucional con un plan gratuito para Google Colab. Esto supone una limitación en el tiempo de ejecución mensual utilizando GPU, lo cual es indispensable. Por estas razones, el coste del equipo utilizado no se toman en cuenta en el presupuesto mostrado en la Tabla 7.1.

<b>Tipos</b>	<b>Cantidad</b>	<b>Coste</b>
Horas de trabajo (20 €/hora)	300	6000€

Tabla 7.1: Resumen de tipos

# Apéndice A

## Tablas de similitud estructural entre imagen falsa e imagen final

### A.1. Primera y última imagen

	Cuantiles				
	0.0	0.25	0.5	0.75	1.0
0	0.741	0.767	0.821	0.852	0.863
1	0.528	0.584	0.601	0.693	0.738
2	0.681	0.737	0.753	0.798	0.818
3	0.58	0.722	0.795	0.827	0.886
4	0.614	0.702	0.794	0.837	0.877
5	0.793	0.804	0.833	0.848	0.862
6	0.71	0.779	0.81	0.844	0.871
7	0.801	0.816	0.845	0.861	0.895
8	0.54	0.682	0.711	0.739	0.867
9	0.601	0.72	0.782	0.869	0.883
10	0.503	0.741	0.83	0.851	0.875
11	0.697	0.826	0.852	0.855	0.872
12	0.744	0.765	0.81	0.822	0.856
13	0.413	0.568	0.708	0.788	0.835
14	0.608	0.735	0.764	0.798	0.85
15	0.679	0.696	0.721	0.752	0.781
16	0.694	0.736	0.753	0.849	0.873
17	0.756	0.801	0.838	0.857	0.876
18	0.671	0.683	0.708	0.742	0.791
19	0.756	0.786	0.827	0.853	0.861
20	0.787	0.819	0.844	0.866	0.88
21	0.647	0.713	0.745	0.789	0.812

Tabla A.1: Tabla de SSIM entre la imagen real tras la evolución y la imagen falsa. Pares de primera y última imagen

## A.2. Seis meses de diferencia entre las imágenes de un par

	Cuantiles				
	0.0	0.25	0.5	0.75	1.0
0	0.732	0.764	0.805	0.811	0.83
1	0.671	0.727	0.781	0.806	0.824
2	0.598	0.623	0.682	0.72	0.757
3	0.716	0.789	0.82	0.827	0.869
4	0.738	0.784	0.812	0.82	0.838
5	0.662	0.707	0.735	0.761	0.818
6	0.684	0.752	0.782	0.803	0.836
7	0.724	0.764	0.775	0.798	0.843
8	0.715	0.744	0.773	0.782	0.798
9	0.667	0.718	0.732	0.765	0.837
10	0.612	0.704	0.726	0.756	0.817
11	0.702	0.726	0.763	0.787	0.824
12	0.691	0.801	0.812	0.817	0.839
13	0.751	0.77	0.797	0.823	0.838
14	0.769	0.783	0.819	0.838	0.858
15	0.729	0.745	0.782	0.824	0.855
16	0.688	0.732	0.781	0.784	0.832
17	0.56	0.592	0.63	0.67	0.731
18	0.492	0.623	0.654	0.682	0.756
19	0.745	0.754	0.764	0.787	0.833
20	0.717	0.756	0.785	0.818	0.855
21	0.735	0.768	0.787	0.806	0.821
22	0.636	0.696	0.726	0.733	0.784
23	0.759	0.787	0.797	0.821	0.847
24	0.797	0.813	0.828	0.868	0.909
25	0.785	0.824	0.84	0.855	0.881

Tabla A.2: Tabla de SSIM entre la imagen real tras la evolución y la imagen falsa. Pares de imágenes con seis meses de diferencia

### A.3. Un año de diferencia entre las imágenes de un par

	Cuantiles				
	0.0	0.25	0.5	0.75	1.0
0	0.682	0.704	0.773	0.815	0.833
1	0.778	0.831	0.84	0.857	0.881
2	0.798	0.818	0.862	0.879	0.922
3	0.767	0.791	0.83	0.842	0.863
4	0.571	0.672	0.697	0.745	0.767
5	0.424	0.461	0.557	0.616	0.752
6	0.604	0.698	0.725	0.744	0.798
7	0.688	0.787	0.812	0.822	0.867
8	0.74	0.804	0.847	0.864	0.89
9	0.681	0.765	0.809	0.831	0.854
10	0.716	0.746	0.772	0.825	0.877
11	0.744	0.757	0.784	0.816	0.842
12	0.753	0.783	0.82	0.849	0.862
13	0.736	0.761	0.787	0.806	0.839
14	0.379	0.693	0.773	0.811	0.847
15	0.754	0.832	0.85	0.858	0.876
16	0.777	0.837	0.855	0.858	0.888
17	0.761	0.776	0.792	0.81	0.896
18	0.771	0.792	0.823	0.845	0.859
19	0.735	0.81	0.84	0.872	0.892
20	0.77	0.803	0.804	0.821	0.859
21	0.696	0.791	0.826	0.839	0.865
22	0.758	0.824	0.847	0.87	0.915
23	0.684	0.733	0.794	0.816	0.838
24	0.81	0.825	0.849	0.874	0.902
25	0.608	0.659	0.727	0.798	0.841
26	0.549	0.642	0.66	0.696	0.758
27	0.653	0.712	0.767	0.805	0.837
28	0.809	0.827	0.841	0.885	0.905
29	0.655	0.716	0.73	0.756	0.788
30	0.626	0.69	0.728	0.77	0.797
31	0.727	0.773	0.814	0.844	0.865
32	0.682	0.763	0.772	0.816	0.851
33	0.692	0.703	0.725	0.784	0.836
34	0.852	0.861	0.886	0.894	0.9
35	0.827	0.849	0.856	0.898	0.92
36	0.774	0.799	0.817	0.824	0.838
37	0.794	0.804	0.844	0.859	0.873
38	0.501	0.657	0.731	0.773	0.833

Tabla A.3: Tabla de SSIM entre la imagen real tras la evolución y la imagen falsa. Pares de imágenes con un año de diferencia (1)

	Cuantiles				
39	0.724	0.794	0.799	0.843	0.874
40	0.459	0.58	0.613	0.636	0.779
41	0.222	0.296	0.415	0.506	0.628
42	0.576	0.764	0.823	0.843	0.873
43	0.787	0.814	0.837	0.842	0.86
44	0.678	0.732	0.749	0.803	0.868
45	0.752	0.786	0.833	0.855	0.869
46	0.766	0.791	0.818	0.837	0.877
47	0.675	0.73	0.791	0.823	0.858
48	0.793	0.858	0.87	0.892	0.899
49	0.765	0.806	0.829	0.832	0.896
50	0.712	0.737	0.761	0.816	0.823
51	0.742	0.781	0.79	0.846	0.871
52	0.716	0.733	0.782	0.823	0.843
53	0.715	0.761	0.801	0.812	0.866
54	0.804	0.813	0.854	0.869	0.895
55	0.626	0.705	0.73	0.757	0.809
56	0.697	0.746	0.787	0.8	0.817
57	0.77	0.808	0.848	0.856	0.881
58	0.718	0.775	0.815	0.824	0.874
59	0.74	0.783	0.824	0.842	0.882
60	0.694	0.721	0.772	0.811	0.855
61	0.563	0.666	0.736	0.756	0.814
62	0.571	0.686	0.709	0.738	0.763
63	0.792	0.82	0.834	0.845	0.883
64	0.667	0.738	0.776	0.83	0.855
65	0.712	0.762	0.818	0.843	0.85
66	0.656	0.733	0.775	0.839	0.872
67	0.563	0.639	0.663	0.702	0.761
68	0.761	0.779	0.84	0.847	0.852
69	0.722	0.768	0.777	0.805	0.82
70	0.71	0.779	0.795	0.818	0.86
71	0.762	0.779	0.813	0.835	0.88
72	0.45	0.654	0.748	0.821	0.878
73	0.717	0.789	0.802	0.826	0.884
74	0.857	0.892	0.903	0.915	0.941
75	0.487	0.669	0.716	0.77	0.817
76	0.592	0.675	0.699	0.718	0.75
77	0.807	0.825	0.872	0.882	0.918
78	0.744	0.775	0.829	0.838	0.885
79	0.74	0.79	0.802	0.832	0.864
80	0.755	0.817	0.83	0.85	0.863

Tabla A.4: Tabla de SSIM entre la imagen real tras la evolución y la imagen falsa. Pares de imágenes con un año de diferencia (2)

	Cuantiles				
81	0.802	0.826	0.837	0.87	0.875
82	0.711	0.776	0.806	0.832	0.853
83	0.639	0.796	0.827	0.844	0.873
84	0.756	0.805	0.815	0.823	0.84
85	0.576	0.614	0.675	0.797	0.866
86	0.727	0.793	0.806	0.822	0.853
87	0.619	0.779	0.809	0.836	0.861
88	0.765	0.795	0.834	0.855	0.889
89	0.343	0.508	0.55	0.596	0.713
90	0.772	0.815	0.822	0.858	0.879
91	0.738	0.759	0.808	0.817	0.834

Tabla A.5: Tabla de SSIM entre la imagen real tras la evolución y la imagen falsa. Pares de imágenes con un año de diferencia (3)

# Bibliografía

- jun 2018. [Online]. Available: <https://thedata scientist.com/tips-tricks-neural-networks/generative-adversarial-network/>
- F. J. F. Batista, T. Diaz-Aleman, J. Sigut, S. Alayon, R. Arnay, and D. Angel-Pereira, "Rim-one dl: A unified retinal image database for assessing glaucoma using deep learning," *Image Analysis & Stereology*, vol. 39, no. 3, pp. 161–167, 2020. [Online]. Available: <https://www.ias-iss.org/ojs/IAS/article/view/2346>
- "Architecture of the u net generator model," May 2019. [Online]. Available: <https://machinelearningmastery.com/wp-content/uploads/2019/05/Architecture-of-the-U-Net-Generator-Model-1.png>
- P. G. Reyes, "Cátedra n° 03 - conceptos básicos de oct," ago 2014. [Online]. Available: <https://tecnologiamedicaoftalmo.blogspot.com/2019/03/catedra-n-03-conceptos-basicos-de-oct.html>
- I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014. [Online]. Available: <https://arxiv.org/abs/1406.2661>
- A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015. [Online]. Available: <https://arxiv.org/abs/1511.06434>
- T. R. V. Bisneto, A. O. de Carvalho Filho, and D. M. V. Magalhães, "Generative adversarial network and texture features applied to automatic glaucoma detection," *Applied Soft Computing*, vol. 90, p. 106165, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1568494620301058>
- S. Jain, S. Indora, and D. K. Atal, "Rider manta ray foraging optimization-based generative adversarial network and cnn feature for detecting glaucoma," *Biomedical Signal Processing and Control*, vol. 73, p. 103425, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1746809421010223>
- Z. Wang, Z. Wang, G. Qu, F. Li, Y. Yuan, D. S. Lam, X. Zhang, Y. Zhang, and Y. Qiao, "Intelligent glaucoma diagnosis via active learning and adversarial data augmentation," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 1234–1237.
- T. K. Yoo, I. H. Ryu, J. K. Kim, I. S. Lee, J. S. Kim, H. K. Kim, and J. Y. Choi, "Deep learning can generate traditional retinal fundus photographs using ultra-widefield images via generative adversarial networks," *Computer Methods*

- and Programs in Biomedicine, vol. 197, p. 105761, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169260720315947>
- A. Diaz-Pinto, A. Colomer, V.Ñaranjo, S. Morales, Y. Xu, and A. F. Frangi, "Retinal image synthesis and semi-supervised learning for glaucoma assessment," *IEEE Transactions on Medical Imaging*, vol. 38, no. 9, pp. 2211–2218, 2019.
- Z. Yu, "Retinal image synthesis from multiple-landmarks input with generative adversarial networks," *BioMedical Engineering OnLine*, vol. 18, 05 2019.
- J. S. Chen, A. S. Coyner, R. P. Chan, M. E. Hartnett, D. M. Moshfeghi, L. A. Owen, J. Kalpathy-Cramer, M. F. Chiang, and J. P. Campbell, "Deepfakes in ophthalmology: Applications and realism of synthetic retinal images from generative adversarial networks," *Ophthalmology Science*, vol. 1, no. 4, p. 100079, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666914521000816>
- A. Tavakkoli, S. Amit, K. F. Hossain, and S. Zuckerbrod, "A novel deep learning conditional generative adversarial network for producing angiography images from retinal fundus photographs," *Scientific Reports*, vol. 10, 12 2020.
- S. Liu, J. Hong, X. Lu, X. Jia, Z. Lin, Y. Zhou, Y. Liu, and H. Zhang, "Joint optic disc and cup segmentation using semi-supervised conditional gans," *Computers in Biology and Medicine*, vol. 115, p. 103485, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482519303543>
- Y. Jiang, N. Tan, and T. Peng, "Optic disc and cup segmentation based on deep convolutional generative adversarial networks," *IEEE Access*, vol. 7, pp. 64 483–64 493, 2019.
- O.Ñ. Hassan, S. Sahin, V. Mohammadzadeh, X. Yang, N. Amini, A. Mylavarapu, J. Martinyan, T. Hong, G. Mahmoudinezhad, D. Rueckert, K.Ñouri-Mahdavi, and F. Scalzo, "Conditional gan for prediction of glaucoma progression with macular optical coherence tomography," 2020. [Online]. Available: <https://arxiv.org/abs/2010.04552>
- A. Qasim, S. Abdullah, J. Che-Hamzah, and A. Abdul Ghani, "Accuracy of using generative adversarial networks for glaucoma detection during the covid-19 pandemic: A systematic review and bibliometric analysis (preprint)," *Journal of Medical Internet Research*, vol. 23, 01 2021.
- X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Medical Image Analysis*, vol. 58, p. 101552, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1361841518308430>
- B. Wu, S. Ren, J. Li, X. Sun, S. Li, and Y. Wang, "Forecasting irreversible disease via progression learning," 2020. [Online]. Available: <https://arxiv.org/abs/2012.11107>
- M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014. [Online]. Available: <https://arxiv.org/abs/1411.1784>
- J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," 2017. [Online]. Available: <https://arxiv.org/abs/1703.10593>

- H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. Metaxas, "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks," 2016. [Online]. Available: <https://arxiv.org/abs/1612.03242>
- C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," 2016. [Online]. Available: <https://arxiv.org/abs/1609.04802>
- T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," 2017. [Online]. Available: <https://arxiv.org/abs/1710.10196>
- F. Chollet, *Deep Learning with Python*. Manning Publications Company, 2017. [Online]. Available: <https://books.google.es/books?id=Yo3CAQAACAAJ>
- I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- A. Krizhevsky, "Learning multiple layers of features from tiny images," 2009.
- M. Maynard-Reid, "Get started: Dcgan for fashion-mnist," nov 2021. [Online]. Available: <https://pyimagesearch.com/2021/11/11/get-started-dcgan-for-fashion-mnist/>
- P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2016. [Online]. Available: <https://arxiv.org/abs/1611.07004>
- O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015. [Online]. Available: <https://arxiv.org/abs/1505.04597>
- [Online]. Available: <https://paperswithcode.com/method/leaky-relu>
- [Online]. Available: <https://paperswithcode.com/method/relu>
- "Error cuadrático medio," Oct 2020. [Online]. Available: [https://es.wikipedia.org/wiki/Error\\_cuadrático\\_medio](https://es.wikipedia.org/wiki/Error_cuadrático_medio)
- Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, pp. 600 – 612, 05 2004.