

TESIS DE MAESTRÍA

MÉTODO BASADO EN APRENDIZAJE PROFUNDO PARA LA MITIGACIÓN  
DE ARTEFACTOS ESPEJO Y EL *ALIASING* PRODUCIDO POR EL  
SUBMUESTREO LATERAL EN EL DOMINIO DE FOURIER EN LA  
TOMOGRAFÍA DE COHERENCIA ÓPTICA

DIEGO A. PULGARÍN  
dapulgaris@eafit.edu.co

UNIVERSIDAD EAFIT  
ESCUELA DE CIENCIAS APLICADAS E INGENIERÍA  
MAESTRÍA EN FÍSICA APLICADA  
LABORATORIO DE ÓPTICA Y FOTÓNICA  
MEDELLÍN  
2025



MÉTODO BASADO EN APRENDIZAJE PROFUNDO PARA LA MITIGACIÓN  
DE ARTEFACTOS ESPEJO Y EL *ALIASING* PRODUCIDO POR EL  
SUBMUESTREO LATERAL EN EL DOMINIO DE FOURIER EN LA  
TOMOGRAFÍA DE COHERENCIA ÓPTICA

DIEGO A. PULGARÍN

Tesis presentada como requisito parcial para optar al título de Magíster en Física  
Aplicada.

Directores

Ph.D RENÉ RESTREPO GÓMEZ  
Director

PH.D CARLOS TRUJILLO ANAYA  
Co-director

PH.D(c) JUAN JOSÉ CADAVID  
Co-director

UNIVERSIDAD EAFIT  
ESCUELA DE CIENCIAS APLICADAS E INGENIERÍA  
MAESTRÍA EN FÍSICA APLICADA  
LABORATORIO DE ÓPTICA Y FOTÓNICA  
MEDELLÍN  
2025



*“La energía no se crea ni se destruye, solo se transforma... en cafeína y noches sin dormir.”*

**Ley de conservación de la energía del estudiante**



# Agradecimientos

*La presente tesis se dedica a mi madre, en reconocimiento a su inquebrantable apoyo y a la confianza depositada en mi capacidad para alcanzar esta meta.*

Quisiera agradecer a Mis asesores, René, Carlos y Juan José, por su invaluable apoyo y contribuciones a lo largo de esta investigación. Su guía experta y las enriquecedoras discusiones en el laboratorio fueron fundamentales para la culminación de este trabajo.

A Santi, cuyo compañerismo y apoyo fueron invaluable durante nuestra estancia en Europa. Las reflexiones compartidas sobre la existencia humana fueron momentos de profunda conexión que dejaron una huella imborrable en mi. Simplemente sin su presencia la pasantía no hubiese sido igual!

A mis compañeros del laboratorio de óptica, por acogerme y hacer de la maestría no solo una experiencia académica, sino por brindarme momentos y recuerdos que llevaré por siempre en mi ser. A todos esos compañeros que ahora tengo el placer de llamar colegas y amigos.

A mis profesores que, durante mi maestría, compartieron su vasto conocimiento y contribuyeron significativamente a mi formación profesional.

A los miembros del grupo de investigación del Wellman Center for Photomedicine, por su valiosa contribución y por el material suministrado, que fueron fundamentales para el desarrollo de los productos de investigación de esta tesis y sus artículos derivados.

A todo el equipo de Vicomtech, especialmente a Arantza, Aitor y Ander, por abrirme las puertas de sus instalaciones en Donosti. Su hospitalidad y calidez hicieron de mi estadía una experiencia gratificante, aportandome conocimientos invaluable para mi crecimiento personal y profesional, Eskerrik asko!

Finalmente, extendiendo mi más sincero agradecimiento a la Vicerrectoría de Ciencia, Tecnología e Innovación, por su apoyo fundamental al permitirme cursar esta maestría a través de su programa interno de becas.

---

# RESUMEN

La tomografía de coherencia óptica (OCT) ha revolucionado el diagnóstico médico al ofrecer imágenes tridimensionales no invasivas de tejidos biológicos con alta resolución. La implementación inicial, conocida como OCT en el dominio del tiempo (TD-OCT), dependía de un barrido mecánico en su brazo de referencia para adquirir cada perfil de profundidad (*A-line*). Este diseño limitaba intrínsecamente la velocidad de adquisición, lo que comprometía la calidad de las imágenes *in-vivo* al hacerlas susceptibles a artefactos causados por el movimiento del paciente. Para superar esta restricción, se desarrolló OCT en el dominio de Fourier (FD-OCT), una técnica que captura simultáneamente toda la información de un *A-line*. Aunque este avance aumentó significativamente la velocidad y la sensibilidad, también introdujo nuevos desafíos técnicos. Uno de ellos es el artefacto espejo, que surge debido a que la señal interferométrica adquirida es real, generando una copia conjugada que se superpone a la imagen tras aplicar la Transformada de Fourier. Otro desafío fundamental es el aliasing. En la práctica clínica, existe un compromiso entre reducir el tiempo de adquisición para minimizar los artefactos de movimiento del paciente y cumplir con el teorema de muestreo de Nyquist. Para evitar el desenfoque o la distorsión causados por movimientos involuntarios, se propone reducir la densidad del muestreo lateral. Aunque esta estrategia disminuye el tiempo de la adquisición, un muestreo insuficiente provoca aliasing, donde las frecuencias espaciales altas de la estructura del tejido se registran incorrectamente como frecuencias más bajas, degradando la fidelidad estructural de la imagen.

En el problema del submuestreo lateral, relacionado con el aliasing, se planteó una solución para reducir el tiempo de adquisición, asumiendo la máxima velocidad física de los espejos galvanométricos, minimizando el impacto de movimientos involuntarios del paciente en la calidad de los tomogramas. Por otro lado, en el problema del artefacto espejo se buscó eliminar las copias especulares generadas en el tomograma, lo que permitió una mejor utilización del rango de profundidad que se puede registrar en el tomograma. Para abordar los artefactos de aliasing y espejo, la literatura reporta diversas técnicas que a menudo requieren modificaciones en el hardware del sistema OCT, lo cual puede implicar mayores costos, complejidad y una aplicabilidad limitada a instrumentos ya existentes. En contraste, este trabajo propone una solución basada en aprendizaje profundo que opera exclusivamente en la etapa de postprocesamiento digital. La principal ventaja de este enfoque es que no requiere intervenciones físicas

---

en el montaje óptico, permitiendo su aplicación a datos adquiridos con sistemas OCT convencionales. Se exploraron arquitecturas generativas adversarias (GANs) para esta tarea, identificando el modelo Pix2Pix como el más eficaz para la corrección de ambas problemáticas. La efectividad de la metodología fue validada tanto cualitativa como cuantitativamente, y los hallazgos se detallan en la sección de resultados. Este trabajo no solo demuestra la capacidad de las redes generativas adversarias para abordar estos desafíos, sino que también sienta las bases para su aplicación en entornos clínicos, donde la eficiencia y precisión en la adquisición de los datos de OCT son fundamentales.

---

# ÍNDICE GENERAL

AGRADECIMIENTOS	V
RESUMEN	VI
ÍNDICE DE FIGURAS	XI
LISTA DE ACRÓNIMOS	XVI
<b>1 Introducción</b>	<b>18</b>
1.1 Señales Digitales . . . . .	21
1.1.1 Muestreo y Aliasing . . . . .	22
1.1.2 Artefactos Espejo . . . . .	25
1.1.3 Mitigación de Artefactos en el Dominio de Fourier . . . . .	27
1.2 Planteamiento del Problema . . . . .	29
1.3 Objetivos . . . . .	30
1.3.1 Objetivo General . . . . .	30
1.3.2 Objetivos Específicos . . . . .	30
<b>2 Tomografía de coherencia óptica</b>	<b>32</b>
2.1 Coherencia de la Luz . . . . .	33
2.1.1 Coherencia Temporal . . . . .	33
2.2 Interferometría de Baja Coherencia . . . . .	35
2.2.1 Modelo Matemático del Campo Eléctrico . . . . .	35
2.2.2 Interferómetro de Michelson y Propagación de los Haces . . . . .	36
2.2.3 Formación de la Señal Interferométrica . . . . .	36
2.2.4 Componentes de la Señal Interferométrica . . . . .	37
2.2.5 Influencia del Espectro de la Fuente de Luz . . . . .	37

2.3	OCT en el dominio de Fourier . . . . .	38
2.3.1	<i>Spectral Domain OCT</i> . . . . .	38
2.3.2	<i>Swept Source OCT</i> . . . . .	39
2.4	<i>Forward Model</i> . . . . .	41
2.5	Muestreo Axial y Lateral en OCT . . . . .	43
2.5.1	Adquisición Axial (A-lines) . . . . .	43
2.5.2	Adquisición Lateral (Escaneo) . . . . .	43
2.5.3	Condiciones de Muestreo y Aliasing en OCT . . . . .	44
2.5.4	Solapamiento de Haces y Continuidad de la Imagen . . . . .	44
2.6	Formación de Artefactos de Espejos en OCT . . . . .	46
2.7	Supresión de artefactos espejo durante la adquisición . . . . .	47
2.7.1	<i>Phase shifting</i> . . . . .	47
2.7.2	<i>Phase shifting</i> con doble brazo de referencia . . . . .	48
2.7.3	<i>Heterodyne SS-OCT</i> con dispositivos de desplazamiento de frecuencia . . . . .	49
2.8	Supresión de artefactos espejo en post-procesamiento . . . . .	51
2.8.1	Algoritmo de DEFR . . . . .	51
<b>3</b>	<b>Fundamentos de Aprendizaje profundo</b>	<b>54</b>
3.1	Fundamentos del Aprendizaje Profundo . . . . .	55
3.2	Arquitecturas para la Reconstrucción de tomogramas submuestreados	57
3.2.1	Pix2Pix . . . . .	57
3.2.2	CCAR Net . . . . .	60
<b>4</b>	<b>Implementación y Evaluación del Aprendizaje Profundo en OCT</b>	<b>66</b>
4.1	Metodología para la Selección de Hiperparámetros . . . . .	66
4.2	Preparación del Conjunto de Datos . . . . .	68
4.2.1	Preparación de tomogramas Submuestreados Lateralmente . . . . .	70
4.2.2	Preparación de tomogramas con artefactos espejo . . . . .	72
4.3	Métricas de Evaluación . . . . .	74

4.4	Resultados y evaluación de los modelos . . . . .	77
4.4.1	Reconstrucción de tomogramas submuestreados mediante aprendizaje profundo . . . . .	77
4.4.2	Mitigación de artefactos espejo usando aprendizaje profundo .	92
<b>5</b>	<b>Conclusiones y trabajo futuro</b>	<b>103</b>
5.1	Conclusiones basadas en los objetivos . . . . .	103
5.2	Conclusiones basadas en los resultados . . . . .	104
5.2.1	Reducción de tiempo en la adquisición de tomogramas . . . .	104
5.2.2	Mitigación del artefacto espejo . . . . .	105
5.3	Trabajo futuro . . . . .	106
5.4	Productos Derivados de la Maestría . . . . .	107
	<b>REFERENCIAS</b>	<b>109</b>

---

# ÍNDICE DE FIGURAS

1.1.	Representación de un volumen OCT que muestra la estructura tridimensional de la mácula y las capas de la retina, utilizando los planos y ejes de referencia estándar en la técnica. . . . .	18
1.2.	Comparación de dos configuraciones de OCT: LF-SDOCT y FF-SSOCT, junto con la reconstrucción de una A-line en OCT a partir de su interferograma espectral . . . . .	20
1.3.	Ejemplo de muestreo a Nyquist y Sub-Nyquist en OCT. A la izquierda, se observan dos haces que se solapan adecuadamente, asegurando una condición de Nyquist. A la derecha, la falta de solapamiento de estos mismos haces genera una condición Sub-Nyquist, propiciando aliasing. . . . .	22
1.4.	Comparación entre una señal de 100 Hz y su versión submuestreada, junto con sus transformadas rápidas de Fourier. La señal submuestreada ha sido interpolada en el dominio de la frecuencia para permitir una representación con el mismo número de puntos que la señal original, facilitando la comparación. En la FFT de la señal submuestreada, el lóbulo principal se centra en la frecuencia aparente del aliasing, lo que ilustra el efecto del submuestreo en el dominio de la frecuencia. . . . .	23
1.5.	Comparación de un tomograma muestreado a la tasa de Nyquist y uno submuestreado. La flecha amarilla indica la dirección en que se aplicó el submuestreo. . . . .	25
1.6.	Representación comparativa de la señal real y las componentes real e imaginaria de la señal compleja en el dominio del tiempo (fila superior), junto con las transformadas de Fourier correspondientes para la señal real y la señal compleja (filas intermedias e inferior, respectivamente). . . . .	27
1.7.	Comparación de una b-scan de una fovea humana sin y con artefactos espejo. . . . .	28
2.1.	Representación del primer grado de coherencia $ g^{(1)} $ en función del retraso normalizado $\tau/\tau_c$ . . . . .	34

2.2. Esquema de un interferómetro de Michelson, donde se observa la fuente de luz de baja coherencia y los caminos del haz de referencia y el haz objeto. . . . .	35
2.3. Esquema de un sistema SD-OCT basado en un interferómetro de Michelson, donde la adquisición axial se obtiene al calcular la transformada de Fourier del interferograma espectral capturado con un espectrómetro digital. . . . .	40
2.4. Simulación de una <i>B-scan</i> usando el modelo directo de OCT. . . . .	43
2.5. Patrón de escaneo en OCT bajo condiciones de muestreo de Nyquist y sub-Nyquist. . . . .	45
2.6. OCT con <i>phase shifting</i> utilizando un acoplador de fibra 3x3. . . . .	48
2.7. Segmento anterior del ojo. En (a) se observa la presencia de la estructura con su respectiva imagen gemela, mientras que en (b) se visualiza únicamente la estructura. La comparación entre ambas imágenes muestra que al eliminar la imagen gemela se duplica el rango de profundidad: en (a) se observa un alcance de 4 mm, mientras que al eliminar el complejo conjugado, se obtiene un rango de 8 mm en el eje de profundidad. . . . .	49
2.8. Diagrama del sistema Heterodyne SS-OCT que utiliza un modulador optoacústico (AO) para generar un desplazamiento de frecuencia en la señal. Se indican los componentes clave: AO (modulador optoacústico), FFPI (interferómetro de Fabry-Pérot de fibra), LPF (filtro de paso bajo) y CASS (sistema de adquisición de señal coherente). . . . .	50
2.9. Comparación entre una señal sin y con <i>Heterodyne</i> OCT. . . . .	51
2.10. <i>B-scans</i> que ilustran el efecto del solapamiento en la visualización de la fovea y la capacidad del método DEFR para mitigar la imagen especular. La fila superior (a-d) muestra la misma estructura con diferentes grados de solapamiento, mientras que la fila inferior (e-h) presenta la misma estructura con el método DEFR aplicado, mostrando una reducción significativa de la imagen especular y mejorando la visualización de la fovea. . . . .	53
3.1. Esquema del generador en arquitecturas adversaria es basadas en U-net. Las conexiones en azul cian representan <i>skip-connections</i> en la arquitectura Pix2Pix y <i>gate-connections</i> en CCAR Net. . . . .	58

4.1. Conjunto de datos utilizados durante el entrenamiento. En (a) se muestra un plano *en-face* de un tomograma experimental. De (b) a (f) se presentan los tomogramas sintéticos donde las B-scans (c) a (f) conservan la misma resolución lateral y axial mostrada en (b). . . . . 70

4.2. Ilustración del proceso de submuestreo lateral aplicado al conjunto de datos. La imagen de la izquierda representa un plano *en-face* del volumen original (“Ground truth”), mientras que la imagen de la derecha muestra el resultado del submuestreo lateral (“Entrada de la red”), donde los B-scans en índices pares han sido reemplazados por valores de cero, creando un patrón intercalado. . . . . 72

4.3. Muestra de tomogramas utilizados para entrenar los modelos que buscan mitigar los artefactos espejo. los grupos de imágenes del recuadro azul muestrean las estructuras usadas como objetivo. Los grupos rojos muestran las mismas estructuras con artefactos espejo y con distintos grados de solapamiento. De (a) a (d) se presentan tomogramas sintéticos, y en (e) se muestra un tomograma experimental de la fóvea, adquirido en el Wellman Center for Photomedicine. Este es usado para la validación de los modelos. . . . . 73

4.4. Gráficos del MSE por época para distintas configuraciones de hiperparámetros. Las líneas continuas representan los resultados de cada configuración, mientras que las líneas punteadas indican el valor promedio del MSE para cada caso. . . . . 78

4.5. Comparación de planos *en-face* de la fóvea. De izquierda a derecha: tomograma original, tomograma reconstruido por la arquitectura cGAN propuesta, tomograma subsampleado interpolado linealmente. Finalmente, la flecha azul indica el eje de subsampleo. . . . . 81

4.6. Comparación de planos ZY de la fóvea. De izquierda a derecha: tomograma original, tomograma reconstruido por la arquitectura cGAN propuesta, tomograma submuestreado interpolado linealmente. La flecha azul indica el eje de submuestreo. . . . . 82

4.7. Diferencias de fase en dirección del eje de las B-scans Y en dirección de las A-lines para el plano *en-face* del tomograma experimental de la fovea. . . . . 83

- 4.8. Comparación entre muestreo de Nyquist y sub-Nyquist en el eje Y con distintos tiempos de adquisición. Fila (a) muestra un plano *en-face* del nervio óptico humano, fila (b) muestra el plano lateral. La columna izquierda corresponde al muestreo de Nyquist con un tiempo de adquisición T, la columna central al muestreo sub-Nyquist reconstruido por cGAN con un tiempo de adquisición T/2, y la columna derecha al muestreo sub-Nyquist en el eje Y reconstruido por cGAN con un tiempo de adquisición T. . . . . 84
- 4.9. En a) se muestra el Comportamiento del índice de similitud estructural (SSIM) y el error cuadrático medio (MSE) a través de los planos *En-face*. En b) se presenta la Variación en la relación señal-ruido pico (PSNR) y la diferencia de histograma en los planos *En-face*. . . . . 86
- 4.10. Evaluación del SSIM en función del número de planos en-face diferentes conjuntos de datos: experimentales y sintéticos combinados (rojo), únicamente experimentales (azul) y únicamente sintéticos (verde). Las líneas punteadas horizontales representan el valor medio del SSIM para cada conjunto de datos. . . . . 88
- 4.11. Comparación de tomogramas de fovea humana después de aplicar TNode: La columna izquierda muestra el tomograma original muestreado a la tasa de Nyquist. El cuadrante superior derecho muestra el tomograma muestreado por debajo de Nyquist reconstruido utilizando el método de aprendizaje profundo propuesto con una cGAN. Cada región resaltada permite una evaluación detallada de la calidad de reconstrucción por diferentes métodos. . . . . 90
- 4.12. Comparación de SSIM entre los tomogramas muestreado a Nyquist y reconstruido mediante Pix2Pix antes y después de aplicar TNode. La línea azul representa el SSIM después de aplicar TNode, y la línea roja representa el SSIM antes de aplicar TNode. Las líneas discontinuas indican los valores promedio de SSIM para cada caso. . . . . 91
- 4.13. Gráficos del MSE por época para distintas configuraciones de hiperparámetros. Las líneas continuas representan los resultados de cada configuración, mientras que las líneas punteadas indican el valor promedio del MSE para cada caso. el recuadro negro con líneas punteadas representan los resultados de los hiperparámetros exclusivos de la red CCARNet . . . . . 93

4.14. Comparación de B-scans con artefactos espejo (primera columna), reconstruidos por la red Pix2Pix (segunda columna), reconstruidos por la red CCARNet (tercera columna) y sin artefactos espejo (cuarta columna). Las regiones marcadas con círculos destacan áreas donde la estructura original se superpone con su artefacto espejo. Las filas muestran distintos grados de solapamiento entre la estructura y su artefacto, lo que permite evaluar el desempeño de la red en diversos escenarios. . . . . 94

4.15. Correlación de la fase a lo largo de los ejes axial y rápido en imágenes B-scan de OCT. La columna izquierda muestra la correlación de fase de un B-scan con artefactos de conjugado complejo, mientras que las siguientes columnas presentan reconstrucciones utilizando Pix2Pix, CCARNet y una imagen libre de artefactos espejo como referencia. . . . . 97

4.16. Comparación cualitativa y cuantitativa de la supresión del artefacto espejo. La columna izquierda muestra las imágenes reconstruidas en cuatro condiciones: a) imagen con artefacto de espejo, b) reconstrucción del campo complejo con Pix2Pix, c) reconstrucción empleando CCARNet y d) imagen libre de artefacto. La columna derecha presenta los perfiles de intensidad correspondientes a las regiones de interés marcadas en las imágenes. . . . . 99

4.17. Variación del SSIM en rojo y del MSE en verde a lo largo de las B-scans del volumen. Se muestra la media y desviación estándar para la solución híbrida con Pix2Pix. . . . . 101

4.18. Métricas de la evolución del PSNR en rojo y de la diferencia de histograma en verde a lo largo de las B-scans del volumen para la solución híbrida con Pix2Pix. Se incluye la media y desviación estándar de ambas métricas. . . . . 102

---

# LISTA DE ACRÓNIMOS

<b>CCAR net</b> ,	<i>Complex conjugate artifact removal network.</i>
<b>CNN</b> ,	Redes neuronales convolucionales.
<b>DEFR</b> ,	Rango completo codificado por dispersión.
<b>FD-OCT</b> ,	Tomografía de coherencia óptica en el dominio de Fourier.
<b>FID</b> ,	<i>Frechet Inception Distance.</i>
<b>FOV</b> ,	Campo de visión.
<b>GPU</b> ,	Unidades de procesamiento gráfico.
<b>MAE</b> ,	Error absoluto medio.
<b>MSE</b> ,	Error cuadrático medio.
<b>NA</b> ,	Apertura numérica.
<b>OCT</b> ,	Tomografía de coherencia óptica.
<b>OFDI</b> ,	Imagen en el dominio de la frecuencia óptica.
<b>PSF</b> ,	Función de dispersión de punto.
<b>PSNR</b> ,	Relación Señal-Ruido Pico.
<b>PZT</b> ,	Transductor piezoeléctrico.
<b>ROI</b> ,	región de interés.
<b>SD-OCT</b> ,	Tomografía de coherencia óptica en el dominio espectral.
<b>SGD</b> ,	<i>Stochastic gradient descent.</i>
<b>SNR</b> ,	Relación señal ruido.
<b>SS-OCT</b> ,	Tomografía de coherencia óptica con fuente de barrido.
<b>SSIM</b> ,	Índice de Similitud Estructural.
<b>TD-OCT</b> ,	Tomografía de coherencia óptica en el dominio del tiempo.
<b>cGAN</b> ,	Red generativa adversaria condicional.



## INTRODUCCIÓN

La tomografía de coherencia óptica, conocida como OCT, es una técnica interferométrica para la adquisición de imágenes volumétricas de muestras biológicas, que tiene la ventaja de ser no invasiva. Esto la convierte en un área de gran interés en campos de la medicina, como la oftalmología, la gastroenterología, la dermatología, entre otras [1,2]. La adquisición de información se logra enfocando un haz en un punto de la muestra y capturando la luz retrodispersada en el tejido [3]. Este procedimiento se lleva a cabo a lo largo de toda la extensión del tejido, obteniendo en cada punto líneas de interferencia en profundidad, conocidas como *A-lines* (eje axial). El proceso se realiza en la dirección del eje x, también conocido como eje rápido, lo que da como resultado la adquisición de una *B-scan*. En última instancia, la recolección de todas las *B-scans* constituye el volumen total [4].

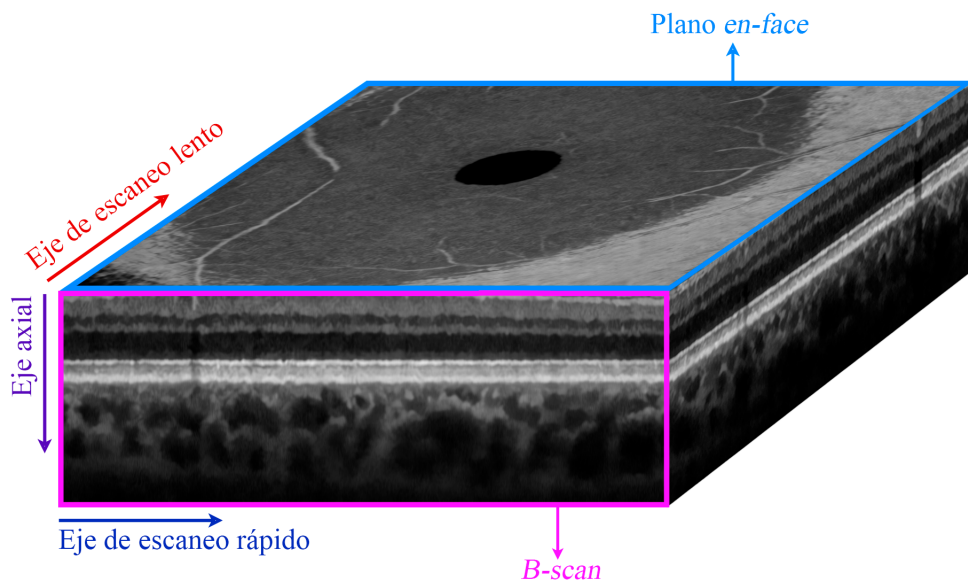


Figura 1.1: Representación de un volumen OCT que muestra la estructura tridimensional de la mácula y las capas de la retina, utilizando los planos y ejes de referencia estándar en la técnica.

El diagrama presentado en la Figura 1.1 muestra los ejes fundamentales en un

---

volumen de OCT. El eje axial, representado en color morado, corresponde a la dirección de profundidad de las imágenes tomográficas. El escaneo rápido (eje azul) se realiza horizontalmente durante la adquisición de un B-scan, mientras que el escaneo lento (eje rojo) describe el movimiento del escáner a través del tejido para generar un volumen completo. Además, el plano *en-face* muestra la proyección del volumen en una vista de planta, ofreciendo una perspectiva adicional utilizada en el diagnóstico.

En OCT, existen varias técnicas que difieren según el equipo utilizado para la adquisición de señales o la muestra que se está analizando. En el caso de OCT en dominio de Fourier (FD-OCT), la adquisición se puede realizar mediante una fuente de barrido que emite a diferentes longitudes de onda en el tiempo, técnica conocida como OCT de fuente barrido (SS-OCT), o empleando una fuente de amplio espectro con un espectrómetro, configuración denominada OCT de dominio espectral (SD-OCT) [5]. La Figura 1.2 compara dos implementaciones de estas técnicas: (a) un sistema SDOCT, basado en una fuente de banda ancha y un espectrómetro digital que captura las franjas de interferencia, y (b) un sistema SSOCT, en el que una fuente de barrido ilumina simultáneamente la muestra y el espejo de referencia, formando la imagen en un detector 2D.

El primer sistema OCT en el dominio de Fourier se informó en 1995 [6], donde el brazo de referencia se mantiene estacionario y la información de profundidad se obtiene mediante una transformada de Fourier de las franjas de interferencia espectral. Este nuevo enfoque revolucionó los sistemas OCT y aumentó drásticamente la velocidad con la que se obtienen las imágenes, y mejoró la relación señal-ruido (SNR) [7, 8].

La velocidad de adquisición es un aspecto crítico para la obtención de imágenes de calidad en OCT, ya que la presencia de movimientos involuntarios del paciente durante el escaneo puede generar artefactos y desenfocos que afectan tanto la visualización como la precisión del diagnóstico [10]. En OCT en dominio de Fourier, la velocidad de adquisición está determinada por dos factores principales: la tasa de *A-lines*, que define la resolución axial y depende del tiempo de exposición, la sensibilidad y el ancho de banda espectral de la fuente [11]; y la velocidad de escaneo lateral, que está limitada por la respuesta de los espejos galvanométricos o el sistema de adquisición utilizado. El balance entre ambos parámetros es crucial para optimizar la calidad de imagen sin comprometer la estabilidad temporal del escaneo. Aunque se han desarrollado sistemas de OCT ultrarrápidos en el rango de MHz [12], los mecanismos de escaneo y los digitalizadores pueden convertirse en factores limitantes [13]. Para abordar estas limitaciones, una estrategia de hardware común es el uso de espejos galvanométricos, que permiten escanear la muestra empleando diferentes tasas de muestreo en los ejes rápido y lento [14–16].

Sin embargo, para reducir el tiempo de adquisición y preservar la integridad

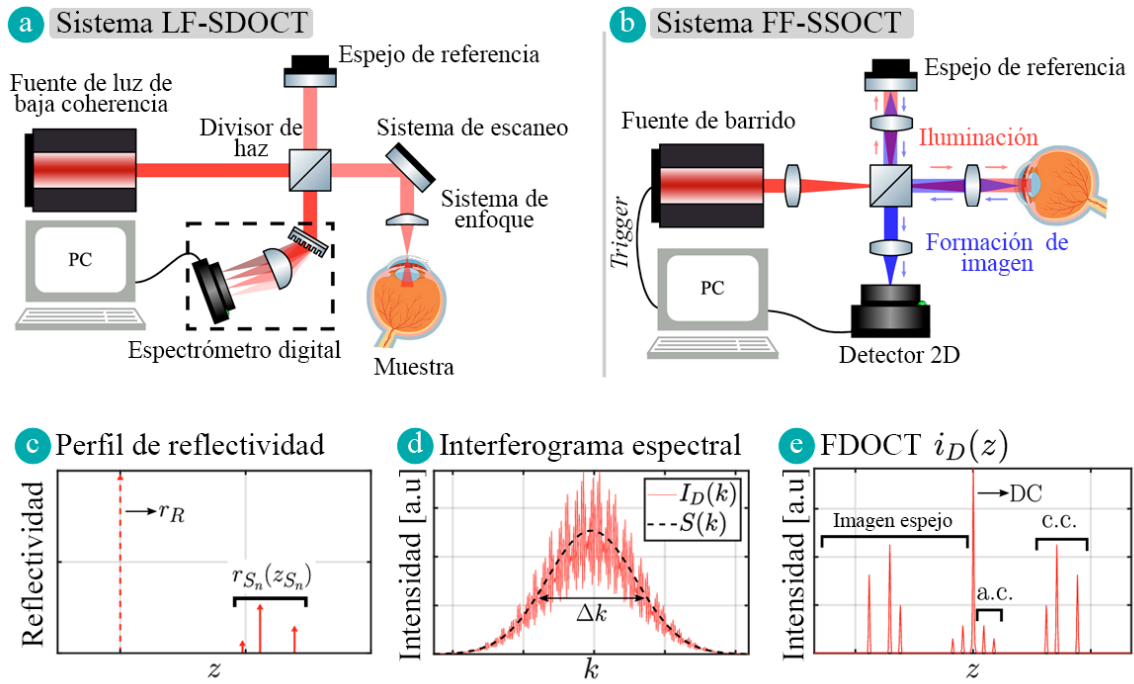


Figura 1.2: Comparación de dos configuraciones de OCT: LF-SDOCT y FF-SSOCT, junto con la reconstrucción de una A-line en OCT a partir de su interferograma espectral. Imagen adaptada de [9].

de la señal, es fundamental mantener un muestreo lateral eficiente. Esto significa ajustar el muestreo en ambos ejes a la tasa de Nyquist. Cuando el muestreo lateral se encuentra por debajo del límite de Nyquist, denominado como submuestreo, se disminuye la demanda en el tiempo de adquisición al reducir el solapamiento del haz de iluminación entre una adquisición y la siguiente a lo largo del tejido. No obstante, esta estrategia conlleva a la pérdida de información espacial y a la aparición de aliasing. Un solapamiento de frecuencias en el dominio recíproco que afecta la capacidad de reconstrucción de los detalles finos de las estructuras presentes en la imagen OCT.

A pesar de los beneficios de los sistemas FD-OCT, la transición de OCT en el dominio temporal a OCT en el dominio de Fourier introdujo varios artefactos no triviales. Como se ilustra en la Figura 1.2, la reconstrucción de una A-line a partir del interferograma espectral conlleva a la presencia de componentes indeseados. En (d) se observa el interferograma registrado en el dominio espectral, cuya transformada de Fourier genera la señal en el dominio espacial mostrada en (e). En esta representación, se pueden identificar artefactos inherentes a FD-OCT, tales como los componentes de autocorrelación (A.C.), que surgen por la interferencia entre distintos puntos reflectores dentro del rango de coherencia de la fuente [8]. Además, uno de los problemas centrales de este trabajo es el artefacto del conjugado complejo, el cual

se debe a que la señal adquirida solo representa la parte real de la componente compleja del campo electromagnético de la luz interferida por la muestra y el haz de referencia [17]. Como resultado, la transformada de Fourier de la señal presenta copias reflejadas a ambos lados de la frecuencia cero [18], dando lugar a la imagen espejo mostrada en (e).

Ambos problemas, el aliasing causado por el submuestreo lateral y los artefactos espejo generados por la simetría de la transformada de Fourier, se manifiestan en el dominio de la frecuencia, y afectan la calidad del volumen OCT. Debido a esto, se requieren estrategias específicas para su corrección y reconstrucción. A continuación, se abordarán de forma general los fundamentos del procesamiento de señales digitales, con énfasis en cómo el aliasing y los artefactos espejo se manifiestan y afectan la representación de la señal en el dominio de la frecuencia.

## 1.1. Señales Digitales

Como se mencionó previamente, la adquisición de imágenes en OCT se realiza de manera secuencial, capturando en profundidad línea a línea la información de interferencia a lo largo de la muestra [8]. Este proceso produce un conjunto de señales que codifican la estructura interna del tejido y que luego deben ser procesadas para reconstruir la imagen tomográfica final [19]. Es esencial entender el tratamiento matemático de estas señales digitales para comprender fenómenos como el aliasing y los artefactos espejo.

De forma general, una **señal digital** es una representación discreta en el tiempo o en el espacio de una señal continua, obtenida mediante un proceso de muestreo y de cuantificación [20]. Matemáticamente, una señal continua en el tiempo  $x(t)$  puede ser muestreada a intervalos uniformes  $T_s$  para obtener una señal discreta  $x[n]$  [21], donde  $n$  es un número entero que representa el índice de la muestra:

$$x[n] = x(nT_s), \quad n \in \mathbb{Z}. \quad (1.1)$$

El intervalo de muestreo  $T_s$  está relacionado con la frecuencia de muestreo  $f_s$  mediante  $T_s = \frac{1}{f_s}$  [22]. La señal digital  $x[n]$  es una secuencia de valores que puede ser procesada y almacenada en sistemas digitales.

En el dominio de la frecuencia, la Transformada de Fourier de Tiempo Discreto (DTFT) de la señal  $x[n]$  se define como:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-i\omega n}, \quad (1.2)$$

donde  $\omega$  es la frecuencia angular digital, con unidades de radianes por muestra de la señal [20]. Debido a la naturaleza discreta de  $x[n]$ , su transformada  $X(e^{j\omega})$  es periódica con periodo  $2\pi$ .

En OCT, las señales digitales resultan de la interferencia entre el haz de referencia y el haz que se retroesparce en la muestra, después de ser muestreadas por los sistemas de adquisición [19]. Estas señales contienen información tanto de amplitud como de fase, y su correcta adquisición y procesamiento son fundamentales para una reconstrucción fiel de la imagen del tejido.

### 1.1.1. Muestreo y Aliasing

El muestreo puede modelarse matemáticamente mediante la multiplicación de  $x(t)$  por un tren de impulsos  $\delta_T(t)$ :

$$x_s(t) = x(t) \cdot \delta_T(t) = x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_s). \quad (1.3)$$

Donde La Transformada de Fourier de la señal muestreada  $x_s(t)$  es:

$$X_s(f) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} X\left(f - \frac{k}{T_s}\right), \quad (1.4)$$

donde  $X_s(f)$  es la Transformada discreta de Fourier de la señal muestreada  $x_s(t)$ . Esta expresión muestra que el espectro de la señal muestreada es una suma de infinitas copias desplazadas del espectro original, espaciadas a intervalos de  $f_s = \frac{1}{T_s}$ .

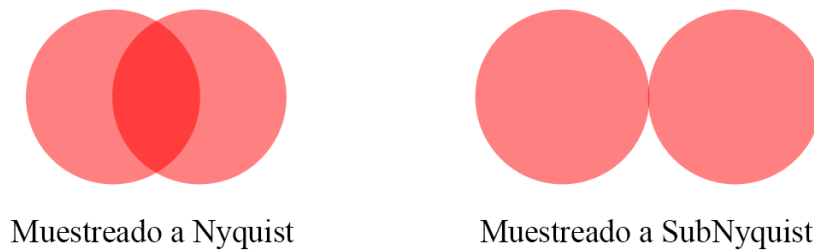


Figura 1.3: Ejemplo de muestreo a Nyquist y Sub-Nyquist en OCT. A la izquierda, se observan dos haces que se solapan adecuadamente, asegurando una condición de Nyquist. A la derecha, la falta de solapamiento de estos mismos haces genera una condición Sub-Nyquist, propiciando aliasing.

El **aliasing** ocurre cuando la frecuencia de muestreo  $f_s$  no es suficientemente alta para representar la totalidad del contenido frecuencial de la señal original en el ancho

frecuencial del muestreo. Según el **Teorema de Muestreo de Nyquist-Shannon**, para evitar el aliasing, la frecuencia de muestreo debe ser al menos el doble de la máxima frecuencia presente en la señal [21], es decir:

$$f_s \geq 2f_{\max}. \quad (1.5)$$

Si esta condición no se cumple, los espectros desplazados se solapan, y las componentes de alta frecuencia de la señal original se pliegan sobre las de baja frecuencia, distorsionando la señal reconstruida [20]. Este fenómeno, conocido como *aliasing*, resulta en pérdida de información y la aparición de artefactos en la señal muestreada.

En el caso de OCT, el cumplimiento de la condición de Nyquist depende del solapamiento de los haces durante el escaneo lateral o entre A-lines adyacentes. La Figura 1.3 ilustra este concepto: en la imagen de la izquierda, los haces se superponen lo suficiente, asegurando un muestreo adecuado según Nyquist. En contraste, la imagen de la derecha muestra una condición Sub-Nyquist, donde la separación entre haces impide la correcta reconstrucción de la señal, dando lugar a aliasing.

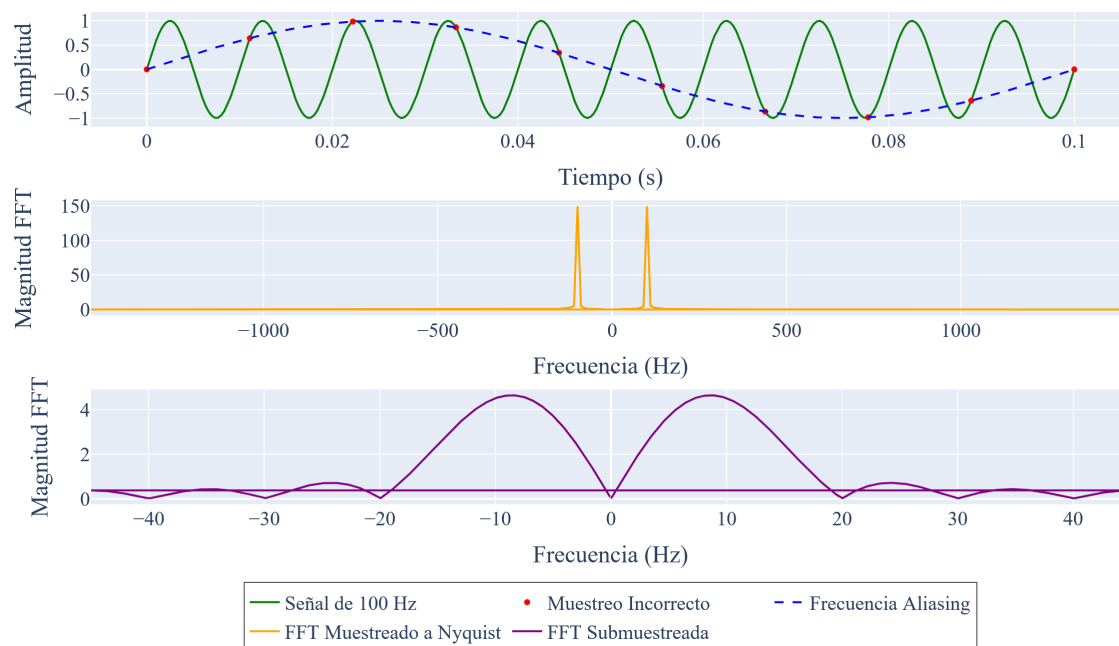


Figura 1.4: Comparación entre una señal de 100 Hz y su versión submuestreada, junto con sus transformadas rápidas de Fourier. La señal submuestreada ha sido interpolada en el dominio de la frecuencia para permitir una representación con el mismo número de puntos que la señal original, facilitando la comparación. En la FFT de la señal submuestreada, el lóbulo principal se centra en la frecuencia aparente del aliasing, lo que ilustra el efecto del submuestreo en el dominio de la frecuencia.

En la Figura 1.4 se compara una señal de 100 Hz con su versión submuestreada, así como las respectivas transformadas rápidas de Fourier. En la primera fila, se muestra la señal original de 100 Hz (verde), la señal submuestreada (puntos rojos) y la señal reconstruida por aliasing (línea azul discontinua). La segunda fila presenta la FFT de la señal muestreada a la tasa de Nyquist, destacando la posición de los picos en la frecuencia de 100 Hz. Finalmente, en la tercera fila, se observa la FFT de la señal submuestreada, donde el aliasing introduce una distorsión en el espectro de frecuencia, desplazando la energía de la señal hacia frecuencias más bajas.

En el contexto de OCT, el muestreo no solo se realiza en el dominio espectral, sino también en el dominio espacial durante el escaneo lateral de la muestra. Si consideramos una señal espacial  $s(x)$  que representa la reflectividad del tejido a lo largo de una dimensión lateral, y la muestreamos a intervalos  $\Delta x$ , la condición para evitar el aliasing espacial es análoga:

$$\Delta x \leq \frac{1}{2f_{x,\max}}, \quad (1.6)$$

donde  $f_{x,\max}$  es la máxima frecuencia espacial presente en la muestra. Si el muestreo lateral no cumple esta condición, las estructuras finas del tejido no pueden ser capturadas correctamente, y se produce aliasing espacial.

Para ilustrar el aliasing, consideremos una señal senoidal continua de frecuencia  $f_0$ :

$$x(t) = \sin(2\pi f_0 t). \quad (1.7)$$

Al muestrear esta señal con una frecuencia  $f_s$ , la señal discreta es:

$$x[n] = \sin(2\pi f_0 n T_s) = \sin\left(2\pi \frac{f_0}{f_s} n\right), \quad (1.8)$$

Si  $f_s < 2f_0$ , la relación  $\frac{f_0}{f_s}$  excede 0,5, y la frecuencia aparente en la señal muestreada será distinta, manifestándose el aliasing.

En OCT, el submuestreo lateral se podría utilizar para reducir el tiempo de adquisición. Sin embargo, al disminuir la frecuencia de muestreo espacial  $f_{s,x}$  por debajo de  $2f_{x,\max}$ , se introduce aliasing, lo que resulta en imágenes con artefactos y pérdida de resolución lateral.

La Figura 1.5 muestra una comparación entre un tomograma muestreado a la tasa de Nyquist (izquierda) y otro submuestreado (derecha). En esta imagen, se observa el impacto del submuestreo en la reconstrucción de las estructuras internas del tejido. Las áreas señaladas por los recuadros rojo y azul destacan regiones donde se ha

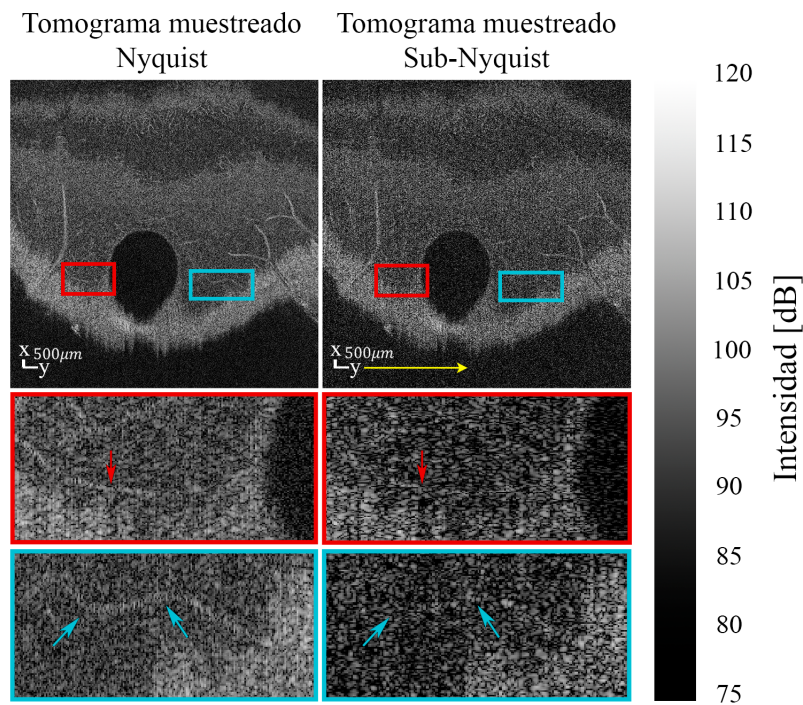


Figura 1.5: Comparación de un tomograma muestreado a la tasa de Nyquist y uno submuestreado. La flecha amarilla indica la dirección en que se aplicó el submuestreo.

producido una pérdida de detalle debido al submuestreo. Las flechas en las regiones de interés señalan estructuras específicas cuya reconstrucción ha sido afectada por la menor tasa de muestreo.

Es importante en OCT diseñar adecuadamente los parámetros de muestreo para cumplir con las condiciones de Nyquist, tanto en el dominio espectral como espacial, o implementar técnicas de procesamiento de señales avanzadas para mitigar los efectos del aliasing cuando el submuestreo es inevitable debido a limitaciones prácticas.

### 1.1.2. Artefactos Espejo

Los artefactos espejo en OCT son el resultado de la naturaleza de la señal adquirida y de cómo esta es transformada al dominio de la frecuencia mediante la Transformada de Fourier [18, 23]. Para entender este fenómeno, es fundamental revisar primero el comportamiento de la Transformada de Fourier en señales reales y complejas.

#### Transformada de Fourier de Señales Reales

La Transformada de Fourier de una señal  $s(t)$ , en su forma continua, está definida como:

$$S(f) = \int_{-\infty}^{\infty} s(t)e^{-i2\pi ft} dt. \quad (1.9)$$

Cuando la señal  $s(t)$  es real, la propiedad de simetría Hermítica [20] se cumple, lo cual implica que la transformada en el dominio de la frecuencia tiene la siguiente propiedad:

$$S(f) = S^*(-f). \quad (1.10)$$

Esto significa que el espectro  $S(f)$  es simétrico respecto al eje de la frecuencia  $f = 0$ , es decir, el valor de la Transformada de Fourier en una frecuencia positiva  $f$  es el conjugado complejo de su valor en la frecuencia negativa  $-f$ . Como consecuencia, la información de la señal está duplicada en ambas direcciones (positiva y negativa) del espectro, es decir si se tiene una señal coseno a una frecuencia  $\omega_0$  se generará en el espacio de frecuencias dos deltas de Dirac, uno en  $\omega_0$  y otro en  $-\omega_0$  [18]. Esta propiedad es el origen de los *artefactos espejo* en OCT, cuando se adquiere una señal real.

### Transformada de Fourier de Señales Complejas

A diferencia de las señales reales, las señales complejas no generan copias espectrales especulares en el dominio de la frecuencia. Esto se debe a que las señales complejas contienen tanto la componente en fase como la componente en cuadratura, lo que permite representar completamente la señal sin necesidad de duplicar la información en frecuencias positivas y negativas [18].

Matemáticamente, una señal compleja  $s(t)$  se puede descomponer mediante la fórmula de Euler:

$$s(t) = A(t)e^{i\phi(t)} = A(t)[\cos(\phi(t)) + i \sin(\phi(t))], \quad (1.11)$$

donde  $A(t)$  es la amplitud de la señal y  $\phi(t)$  es su fase. Esta expresión muestra que una señal compleja está formada por dos componentes: una parte real (en fase) proporcional al coseno, y una parte imaginaria (en cuadratura) proporcional al seno.

En la Figura 1.6 se muestra una señal en el dominio del tiempo que consta de dos componentes frecuenciales: una componente de frecuencia positiva y otra de frecuencia negativa. Esta configuración se ha utilizado para simular de manera ilustrativa cómo puede producirse el solapamiento de frecuencias en señales reales y complejas.

En la primera fila de la gráfica se observa la señal real  $A \cos(\omega t + \theta)$ , la cual

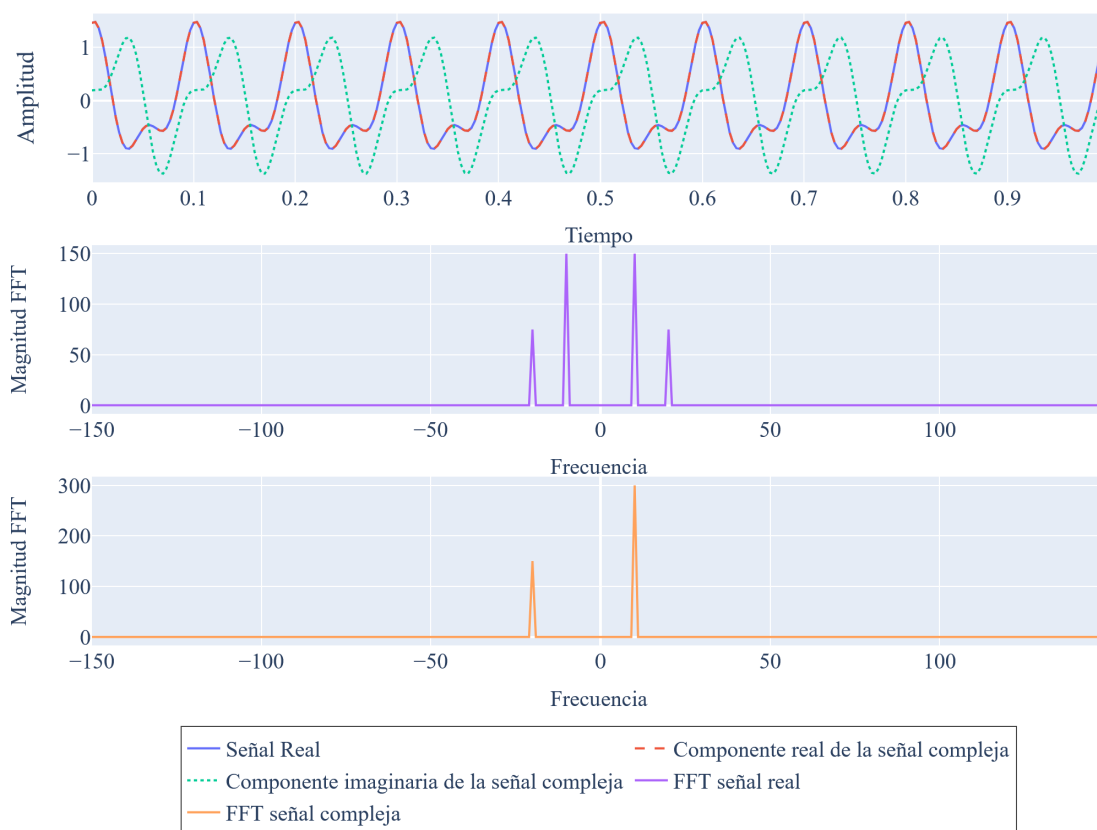


Figura 1.6: Representación comparativa de la señal real y las componentes real e imaginaria de la señal compleja en el dominio del tiempo (fila superior), junto con las transformadas de Fourier correspondientes para la señal real y la señal compleja (filas intermedias e inferior, respectivamente).

está formada por la suma de las componentes de frecuencia positiva y negativa. Como se evidencia en la Transformada de Fourier de esta señal (segunda fila), ambas frecuencias aparecen reflejadas simétricamente con respecto a la frecuencia cero. En contraste, la señal compleja  $Ae^{-i\omega t + \theta}$ , mostrada en la misma figura, exhibe una Transformada de Fourier distinta (tercera fila). A diferencia de la señal real, la transformada de la señal compleja no muestra una imagen conjugada, ya que al utilizar la representación exponencial compleja se elimina la simetría inherente de las señales reales. Esto evita la aparición del artefacto de imagen gemela, lo cual es importante para mejorar la calidad de las imágenes reconstruidas en OCT.

### 1.1.3. Mitigación de Artefactos en el Dominio de Fourier

Como se mencionó previamente en las secciones 1.1.1 y 1.1.2, en OCT existen diversos factores que afectan la calidad de la imagen, generando artefactos en el dominio de Fourier que distorsionan la representación del tejido. Dos de los problemas más relevantes en este contexto son el aliasing, que ocurre debido al submuestreo

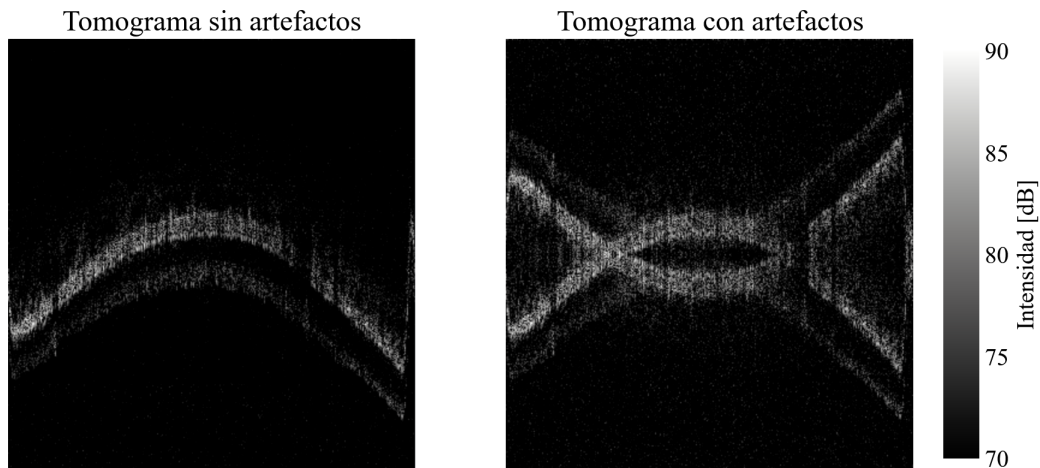


Figura 1.7: Comparación de una b-scan de una fovea humana sin y con artefactos espejo.

lateral, y el artefacto del complejo conjugado, que resulta de la simetría inherente de la Transformada de Fourier aplicada a señales reales.

Para abordar estos problemas, se han propuesto soluciones tanto en el ámbito del hardware como en el del procesamiento de señales. En el caso del aliasing, una estrategia consiste en mejorar la velocidad de adquisición mediante sistemas de OCT ultrarrápidos, como los basados en escáneres resonantes, que permiten aumentar la tasa de muestreo sin comprometer la calidad de la imagen [24]. Alternativamente, el uso de redes neuronales convolucionales ha demostrado ser efectivo en la reconstrucción de señales submuestreadas, permitiendo recuperar las frecuencias faltantes y mitigar los efectos del aliasing [25, 26].

Para eliminar el artefacto espejo, una de las estrategias más utilizadas consiste en descartar la mitad de la información espectral siempre y cuando esta no esté solapada con la estructura real, es decir que tanto la imagen real como la imagen gemela estén situadas cada una a un lado del *zero path delay*, esto elimina la imagen reflejada pero también reduce el rango de profundidad utilizable. Como alternativa a esta pérdida de información, se han desarrollado métodos basados en hardware y técnicas instrumentales que buscan determinar la señal compleja o cuadratura de la señal, permitiendo suprimir el artefacto espejo en la etapa de adquisición. Estas estrategias serán detalladas en la sección 2.7. En el contexto del procesamiento de señales, el aprendizaje profundo ha emergido como una herramienta clave para eliminar el artefacto directamente en la etapa de posprocesamiento. Modelos basados en redes generativas adversarias han sido empleados para distinguir y suprimir la información duplicada, obteniendo reconstrucciones libres de artefactos sin necesidad de modificar el proceso de adquisición [27, 28].

Si bien cada una de estas estrategias ha mostrado resultados prometedores, persisten limitaciones en su aplicabilidad y desempeño. La siguiente sección plantea en detalle el problema y las dificultades asociadas con la eliminación de estos artefactos en imágenes OCT.

## 1.2. Planteamiento del Problema

Como se evidenció en las secciones anteriores, las imágenes OCT están sujetas a artefactos causados por distorsiones en el dominio de Fourier. Ambos problemas conducen a la aparición de estructuras no deseadas en las imágenes OCT, afectando la interpretación y precisión diagnóstica de estas imágenes. La eliminación efectiva de ambos artefactos es importante para mejorar la calidad de las imágenes y permitir un rango de profundidad utilizable mayor en OCT.

Los artefactos en OCT, derivados tanto de distorsiones en el dominio de Fourier como del fenómeno del complejo conjugado, representan un obstáculo significativo para el diagnóstico preciso en oftalmología y otras aplicaciones biomédicas [29]. Específicamente, estos artefactos pueden enmascarar detalles críticos de la microestructura tisular, lo que dificulta la detección temprana de enfermedades como el glaucoma y la degeneración macular [30, 31]. OCT se ha convertido en un estándar en aplicaciones como en la oftalmología, proporcionando información en tiempo real sobre la estructura y función del ojo para diagnosticar enfermedades, evaluar su progresión y la respuesta a terapias [30]. Sin embargo, la presencia de artefactos como el aliasing debido al submuestreo lateral y los artefactos espejo producidos por la ambigüedad del complejo conjugado limitan la calidad y la interpretabilidad de las imágenes obtenidas.

El aliasing o submuestreo lateral ocurre cuando la frecuencia de muestreo espacial es insuficiente para representar con precisión las estructuras finas dentro del tejido, lo que lleva a la aparición de artefactos que pueden simular o enmascarar patologías reales. Por otro lado, los artefactos espejo son una consecuencia inherente del procesamiento de la señal OCT en el dominio de Fourier. Dado que la transformada de Fourier de una señal interferométrica espectral de valor real es simétrica hermitiana, se produce una superposición de reflectores a diferencias de trayectoria óptica positivas y negativas, generando imágenes especulares que dificultan la interpretación de la estructura real del tejido. Esta ambigüedad del complejo conjugado puede llevar a la confusión entre estructuras reales y artefactos, especialmente en sistemas de alta resolución donde los términos espejo pueden volverse más visibles.

Las técnicas tradicionales de corrección, como el *phase shifting* y la modulación de frecuencia [32], aunque efectivas en ciertos escenarios, poseen limitaciones en resolución espacial, velocidad de procesamiento y elevado costo de instrumentación

[33]. Además, algunas de estas técnicas requieren la adquisición de múltiples imágenes que pueden no ser instantáneas. Por tal motivo, el uso de técnicas basadas en Deep Learning, proveen una solución más versátil, robusta, y generalizable, que permite afrontar tanto los efectos de *aliasing* como los artefactos espejo producidos por el complejo conjugado [27,28]. El uso de estas redes permite generalizar contra diferentes tipos de ruidos, y diferentes equipos de adquisición. Además, se han demostrado éxitos del aprendizaje profundo en la eliminación de artefactos de imágenes médicas, ya que las redes neuronales pueden aprender a reconocer y corregir patrones complejos de artefactos [34,35]. Arquitecturas como U-Net han demostrado ser especialmente efectivas para la segmentación rápida y precisa de imágenes biomédicas, y las Redes Generativas Adversarias, incluyendo sus variantes condicionales, han mostrado potencial en la traducción de imagen a imagen y la eliminación de ruido. Sin embargo, a diferencia de las metodologías existentes, esta tesis se enfoca en la reconstrucción de la señal compleja, preservando la información tanto de amplitud como de fase, en lugar de limitarse al problema de imágenes de intensidad con un rango típico de 0 a 255. Adicionalmente, mientras que los trabajos previos suelen abordar el *aliasing* o los artefactos espejo de forma individual, esta investigación busca ofrecer una solución para ambos problemas.

El objetivo principal de este trabajo es desarrollar una estrategia de aprendizaje profundo capaz de mitigar estos artefactos, mejorando la precisión y confiabilidad de las imágenes OCT. Para ello, se entrenará un modelo que reconstruya tomogramas libres de artefactos utilizando tanto datos sintéticos como experimentales, optimizando la visualización y el rango de profundidad para aplicaciones biomédicas y clínicas.

### 1.3. Objetivos

#### 1.3.1. Objetivo General

Desarrollar una metodología basada en aprendizaje profundo para abordar dos limitaciones en OCT, la presencia de artefactos espejo, inherentes a la transformada de Fourier, y el *aliasing* resultante del submuestreo lateral, ambos fenómenos que se manifiestan en el dominio de Fourier. El objetivo es mitigar ambos problemas para aumentar la profundidad de imagen, preservar la integridad estructural y dimensional de los tejidos, y potencialmente reducir el tiempo de adquisición sin modificaciones ópticas.

#### 1.3.2. Objetivos Específicos

- Construir una base de datos de imágenes OCT simuladas utilizando un modelo directo que incluya la generación controlada de artefactos espejo y simule los

efectos del submuestreo lateral en el dominio de Fourier, para entrenar y evaluar modelos de aprendizaje profundo.

- Implementar modelos de aprendizaje profundo para la mitigación de artefactos espejo y para la reducción del *aliasing* producido por el submuestreo lateral en imágenes OCT, utilizando como datos de entrada tomogramas simulados y tomogramas experimentales.
- Analizar la generalización de los modelos de aprendizaje profundo en la detección y corrección de artefactos espejo y de aliasing en imágenes OCT utilizando diferentes tipos de muestras biológicas experimentales.
- Evaluar cuantitativamente el rendimiento de los modelos de aprendizaje profundo implementados para la mitigación de artefactos espejo y la reducción del *aliasing*, utilizando métricas de evaluación objetivas para comparar los resultados obtenidos con respecto a los datos de referencia

# TOMOGRAFÍA DE COHERENCIA ÓPTICA

La Tomografía de Coherencia Óptica es una técnica de imagen no invasiva basada en la interferometría que permite obtener imágenes tomográficas de alta resolución de la estructura interna de muestras biológicas [36]. OCT mide el “eco” de la luz retrodispersada por la muestra, de forma similar a cómo la ecografía mide el eco del sonido. Esta técnica ha sido ampliamente utilizada en áreas como la oftalmología, la dermatología y la cardiología [1, 2, 37], debido a su capacidad para proporcionar imágenes en tiempo real de los tejidos con resolución micrométrica. Existen dos tipos principales de OCT: en el dominio del tiempo y en el dominio de Fourier [8]. En TD-OCT, la luz de una fuente de baja coherencia es dividida en un interferómetro de Michelson en dos brazos: uno dirigido hacia la muestra y el otro hacia un espejo de referencia. Este se mueve linealmente hacia adelante y hacia atrás a una velocidad constante. Este movimiento controlado permite modular la longitud de la trayectoria óptica y, por lo tanto, el tiempo de vuelo de la luz reflejada. La interferencia resultante entre la luz retrodispersada por la muestra y la luz reflejada por el espejo genera una señal que está directamente relacionada con la profundidad y estructura de la muestra [9]. La intensidad de la señal permite reconstruir un perfil de profundidad o *A-line* de la muestra, con una resolución axial determinada por la longitud de coherencia de la fuente de luz utilizada [38, 39].

Con el desarrollo de OCT en el dominio de Fourier, se logró una mejora significativa en la velocidad y sensibilidad de adquisición de imágenes [39, 40]. En FD-OCT, la interferencia se captura sin necesidad de mover el espejo de referencia, ya que la información de profundidad se codifica en las frecuencias de la señal interferométrica [8]. Esto se hace ya sea utilizando un espectrómetro para medir el espectro completo de la señal, más conocido como Spectral-Domain OCT [41], o mediante el uso de una fuente de barrido que cambia su longitud de onda con el tiempo, denominado como Swept-Source OCT [41]. El análisis posterior mediante la Transformada de Fourier permite obtener la distribución de reflectividad en profundidad de la muestra.

## 2.1. Coherencia de la Luz

La coherencia es una propiedad fundamental de la luz que describe la correlación entre las fases de las ondas electromagnéticas en diferentes puntos en el espacio y el tiempo [42]. Esta propiedad es crucial para técnicas interferométricas como OCT, ya que determina la capacidad de la luz para interferir de manera específica y diferenciar estructuras aledañas. [43]. La coherencia de la luz se puede clasificar en dos tipos: coherencia temporal y coherencia espacial.

### 2.1.1. Coherencia Temporal

La coherencia temporal se refiere a la correlación de la fase de la onda de luz a lo largo del tiempo en un mismo punto en el espacio [42, 44, 45]. La coherencia temporal está relacionada con la longitud de coherencia, la cual es la distancia a lo largo de la dirección de propagación de la luz en la que un tren de ondas electromagnéticas con un ancho de banda espectral permanecen en fase [42]. La longitud de coherencia  $l_c$  está inversamente relacionada con el ancho de banda espectral  $\Delta\lambda$  de la fuente de luz. En OCT la longitud de coherencia se expresa mediante la siguiente relación,

$$l_c = \frac{2 \ln 2}{\pi} \frac{\lambda_0^2}{\Delta\lambda}, \quad (2.1)$$

donde  $\lambda_0$  es la longitud de onda central de la fuente de luz. Las fuentes de luz con un amplio ancho de banda espectral, como las fuentes de luz de baja coherencia utilizadas en OCT, tienen una longitud de coherencia corta, usualmente entre los  $5\mu m$  y los  $100\mu m$ . Esto permite que la interferencia ocurra solo cuando las diferencias de camino óptico entre los haces de referencia y muestra están dentro de este rango de coherencia [43].

### Función de Coherencia de Primer Grado

La coherencia de la luz puede cuantificarse usando la función de coherencia de primer grado, definida como,

$$g^{(1)}(\tau) = \frac{\langle E(t)E^*(t + \tau) \rangle}{\langle |E(t)|^2 \rangle}, \quad (2.2)$$

donde  $E(t)$  es el campo eléctrico de la onda,  $E^*(t + \tau)$  es su conjugado complejo en un tiempo retardado  $\tau$ , y los corchetes  $\langle \cdot \rangle$  representan el promedio temporal. La función  $g^{(1)}(\tau)$  mide el grado de correlación de las fases de la onda a diferentes tiempos [42]. La magnitud de  $g^{(1)}(\tau)$  alcanza su valor máximo de 1 cuando  $\tau = 0$  y disminuye a medida que aumenta  $\tau$ . La escala en la cual  $|g^{(1)}(\tau)|$  se reduce

significativamente se denomina longitud de coherencia.

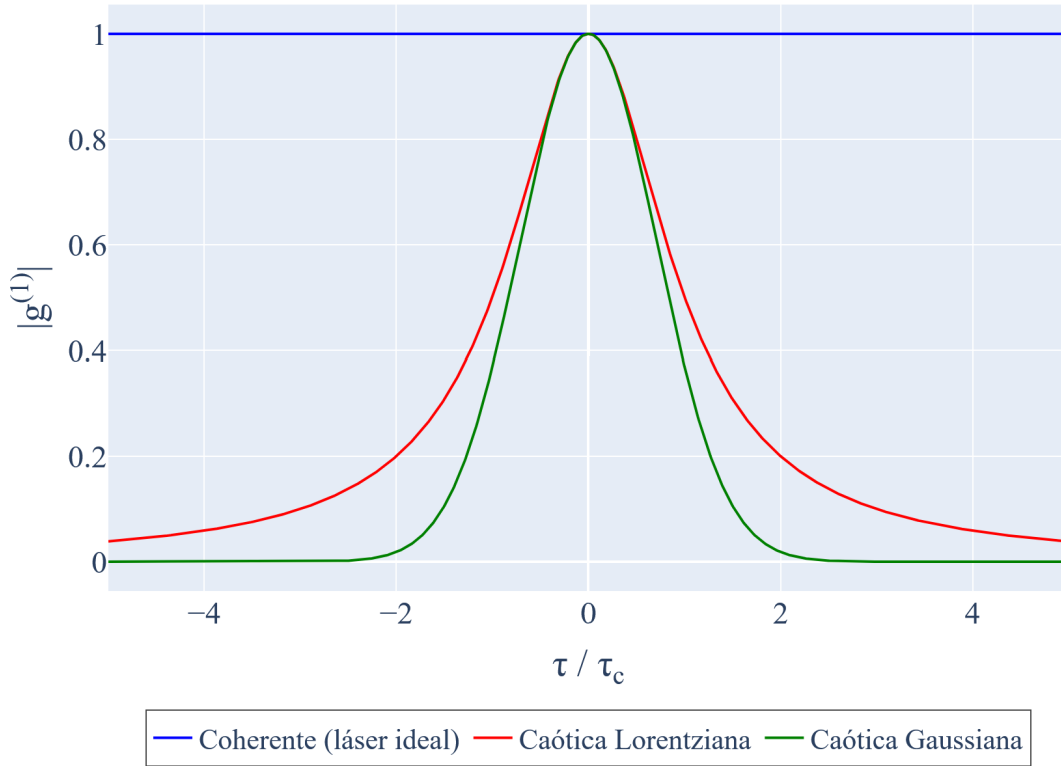


Figura 2.1: Representación del primer grado de coherencia  $|g^{(1)}|$  en función del retraso normalizado  $\tau/\tau_c$ .

En el contexto de OCT, la función de coherencia de primer grado  $g^{(1)}$  está directamente relacionada con el *zero path* o frecuencia cero. Este corresponde a la posición donde las diferencias de camino óptico entre los brazos de referencia y muestra son iguales, lo que genera una autocorrelación o un contraste interferométrico máximo. Este máximo ocurre cuando las diferencias de tiempo  $\tau$  entre los frentes de onda reflejados están dentro de la longitud de coherencia de la fuente. Matemáticamente, esto se traduce en que  $g^{(1)}(\tau)$  alcanza su valor máximo cuando  $\tau = 0$  [19].

La Figura 2.1 ilustra el comportamiento del primer grado de coherencia  $|g^{(1)}|$  en función del retraso normalizado  $\tau/\tau_c$ , comparando tres tipos de fuentes de luz: una fuente coherente ideal (curva azul), una fuente caótica o parcialmente coherente con un perfil Lorentziano (curva roja) y otra con un perfil Gaussiano (curva verde). La fuente coherente ideal mantiene su coherencia a lo largo de un mayor rango de retraso, mientras que las fuentes caóticas presentan una caída más rápida, dependiendo de sus perfiles espectrales.

En el dominio de Fourier, la frecuencia cero representa este punto de máxima interferencia. A medida que las diferencias de camino óptico se desvían del *zero path*,

la correlación de las ondas reflejadas disminuye, lo que provoca una caída en  $|g^{(1)}(\tau)|$  y, en consecuencia, una reducción del contraste interferométrico [19]. Por lo tanto, la función de coherencia de primer grado no solo caracteriza la coherencia de la luz, sino que también define la posición y la intensidad relativa de la señal en el dominio de Fourier.

## 2.2. Interferometría de Baja Coherencia

La interferometría de baja coherencia es el principio central sobre el cual se basa OCT. Este proceso se realiza mediante un interferómetro, generalmente de tipo Michelson, que utiliza una fuente de luz con un amplio espectro o de baja coherencia. La interferencia resultante solo se produce cuando las diferencias de camino óptico entre los dos haces se encuentran dentro del rango de la longitud de coherencia de la luz utilizada [23, 46]. Esto es fundamental para la resolución axial del sistema OCT.

### 2.2.1. Modelo Matemático del Campo Eléctrico

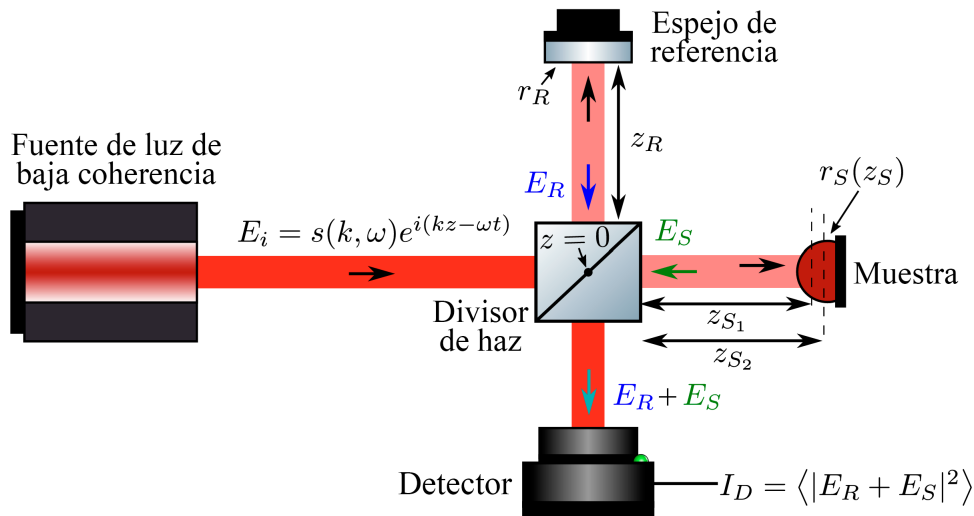


Figura 2.2: Esquema de un interferómetro de Michelson, donde se observa la fuente de luz de baja coherencia y los caminos del haz de referencia y el haz objeto. Adaptado de [9].

Consideremos la luz emitida por la fuente de baja coherencia como la que se observa en la figura 2.2. El campo eléctrico de esta luz puede representarse en forma compleja como:

$$E_i(k, \omega) = s(k, \omega)e^{i(kz - \omega t)}, \quad (2.3)$$

donde  $s(k, \omega)$  es la amplitud espectral del campo eléctrico, dependiente del número de onda  $k = 2\pi/\lambda$  y de la frecuencia angular  $\omega = 2\pi\nu$ . Aquí,  $k$  y  $\omega$  describen la relación entre la frecuencia espacial y temporal en la propagación, mientras que  $e^{i(kz-\omega t)}$  representa la dependencia tanto espacial ( $z$ ) como temporal ( $t$ ) del campo eléctrico [47].

### 2.2.2. Interferómetro de Michelson y Propagación de los Haces

En un interferómetro de Michelson, representado en la Figura 2.2, la luz emitida por la fuente se divide en dos haces: uno que se dirige hacia la muestra llamado haz objeto y otro hacia un espejo de referencia denominado haz de referencia [46]. El haz de referencia se propaga una distancia  $z_R$  antes de reflejarse en el espejo, con una reflectividad  $r_R$ . La intensidad reflejada de este haz es  $R_R = |r_R|^2$ , y el campo eléctrico reflejado cuando regresa al divisor se describe por

$$E_R = \frac{E_i}{2} r_R e^{i2kz_R}. \quad (2.4)$$

El factor  $e^{i2kz_R}$  surge debido al corrimiento de fase que experimenta el haz durante su propagación de ida y vuelta al espejo.

Por otro lado, el haz objeto se propaga hasta llegar a la muestra, la cual posee una reflectividad dependiente de la profundidad  $r_S(z_S)$ . Para simplificar, consideramos una serie de  $N$  reflectores a través de la muestra, de modo que la reflectividad total se puede expresar como,

$$r_S(z_S) = \sum_{n=1}^N r_{S_n} \delta(z_S - z_{S_n}), \quad (2.5)$$

donde  $r_{S_n}$  es el coeficiente de reflexión del  $n$ -ésimo reflector, y  $z_{S_n}$  es su posición en profundidad. El campo eléctrico reflejado por la muestra al regresar al divisor se representa como,

$$E_S = \frac{E_i}{4} \sum_{n=1}^N r_{S_n} e^{i2kz_{S_n}}. \quad (2.6)$$

### 2.2.3. Formación de la Señal Interferométrica

La señal de interferencia que se detecta es la intensidad total, dada por,

$$I(k, \omega) = |E_R + E_S|^2 = (E_R + E_S)(E_R + E_S)^*, \quad (2.7)$$

donde  $*$  indica el conjugado complejo. Esta intensidad se detecta mediante un fotodetector en el caso de SS-OCT o en un arreglo de pixeles en SD-OCT, generando una fotocorriente proporcional a  $I(k, \omega)$ . Esta fotocorriente puede describirse como,

$$I_D(k, \omega) = \rho \langle I(k, \omega) \rangle, \quad (2.8)$$

donde  $\rho$  es el factor de respuesta del detector y  $\langle \cdot \rangle$  denota la integración media sobre el tiempo de respuesta. Al sustituir  $E_R$  y  $E_S$  y expandir la expresión, se obtiene,

$$I_D(k, \omega) = \rho \left\langle \left| \frac{s(k, \omega)}{2} r_R e^{i2kz_R} + \frac{s(k, \omega)}{2} \sum_{n=1}^N r_{S_n} e^{i2kz_{S_n}} \right|^2 \right\rangle. \quad (2.9)$$

#### 2.2.4. Componentes de la Señal Interferométrica

Tras realizar el promedio temporal y asumir que la frecuencia temporal de la onda es mucho mayor que el tiempo de respuesta del detector, se obtiene una expresión final independiente del tiempo descrito en [47, 48] como,

$$\begin{aligned} I_D(k) = & \frac{\rho}{4} [S(k)(r_R + r_{S1} + r_{S2} + \dots + r_{SN})] \\ & + \frac{\rho}{2} \left[ S(k) \sum_{n=1}^N \sqrt{r_R r_{S_n}} \cos[2k(z_R - z_{S_n})] \right] \\ & + \frac{\rho}{4} \left[ S(k) \sum_{m \neq n=1}^N \sqrt{r_{S_n} r_{S_m}} \cos[2k(z_{S_n} - z_{S_m})] \right]. \quad (2.10) \end{aligned}$$

La expresión resultante consta de tres términos. El primer término representa la componente DC de la señal, que es independiente de la diferencia de camino óptico y actúa como un escalamiento constante en el detector. El segundo término contiene información de la posición en profundidad del reflector  $n$ -ésimo, y constituye la señal de interés en OCT, ya que es la que permite obtener la estructura interna de la muestra. Finalmente, el tercer término describe la interferencia que ocurre entre los distintos reflectores en la muestra, lo cual se manifiesta como artefactos en la señal y generalmente corresponde al speckle.

#### 2.2.5. Influencia del Espectro de la Fuente de Luz

Para sistemas OCT, se suele utilizar una fuente con un espectro de distribución gaussiana, lo que resulta en una señal interferométrica que también posee un perfil

de intensidad gaussiano. Matemáticamente, si la distribución espacial en el dominio  $z$  de la fuente es  $\gamma(z) = e^{-z^2\Delta k^2}$ , su transformada de Fourier se expresa como,

$$S(k) = \frac{1}{\Delta k \sqrt{\pi}} e^{-\left[\frac{(k-k_0)}{\Delta k}\right]^2}, \quad (2.11)$$

donde  $k_0$  es el número de onda central y  $\Delta k$  el ancho de banda espectral. Esta distribución afecta la resolución axial de OCT. El espectro gaussiano de la fuente de luz en OCT presenta ventajas significativas. Su forma gaussiana minimiza los lóbulos laterales en la PSF, lo que se traduce en una resolución axial óptima y una mejor distinción de estructuras cercanas. Además, concentra la energía de la fuente, mejorando la relación señal-ruido resultando en imágenes más nítidas. La distribución espacial de la fuente en el dominio  $z$ , dada por  $\gamma(z) = e^{-z^2\Delta k^2}$ , muestra una dependencia cuadrática de la profundidad debido a la atenuación de la luz en el tejido por dispersión y absorción. Esta atenuación es más pronunciada a medida que aumenta la profundidad. Si bien un espectro gaussiano estrecho (ancho de banda pequeño) puede emborronar detalles espaciales debido a una mayor longitud de coherencia, en la práctica se busca un equilibrio entre resolución axial y penetración mediante la selección de un ancho de banda adecuado para la aplicación específica [36].

### 2.3. OCT en el dominio de Fourier

La Tomografía de Coherencia Óptica en el dominio de Fourier representa una mejora significativa respecto a TD-OCT, ya que permite una mayor sensibilidad y velocidad de adquisición [39,41]. A diferencia de TD-OCT, FD-OCT no requiere mover el espejo de referencia, lo que facilita la adquisición simultánea de la información de profundidad a lo largo de la muestra [36, 37, 41]. La señal interferométrica se adquiere directamente en el espacio  $k$ , y una posterior transformada de Fourier permite obtener el perfil de profundidad  $i_D(z)$  de la muestra [43]. En esta sección, se describen las dos aproximaciones principales de FD-OCT: OCT en el dominio espectral y la OCT con fuente de barrido.

#### 2.3.1. *Spectral Domain OCT*

En SD-OCT, la señal interferométrica  $I_D(k)$  se adquiere utilizando un espectrómetro digital [41, 43], como se muestra en la Figura 2.3. El espectrómetro separa las diferentes componentes espectrales de la luz reflejada por la muestra, capturando la intensidad de todo el ancho de banda en función del número de onda  $k$  en una sola medición. La adquisición en el dominio espectral permite utilizar la transformada de Fourier para reconstruir el perfil de profundidad  $i_D(z)$  de la muestra,

$$i_D(z) = \text{FT}_k\{I_D(k)\}, \quad (2.12)$$

donde  $\text{FT}_k$  denota la transformada de Fourier con respecto al número de onda  $k$ . La relación entre el espacio axial y el dominio del número de onda es fundamental en SD-OCT, ya que el perfil de profundidad se extrae de la interferencia capturada en el espectro [43].

La señal interferométrica en FD-OCT contiene múltiples componentes que se pueden expresar como,

$$\begin{aligned} i_D(z) = \frac{\rho}{8}\gamma(z) & \left[ r_R + \sum_{n=1}^N r_{S_n} \right] \\ & + \frac{\rho}{4} \sum_{n=1}^N \sqrt{r_R r_{S_n}} [\gamma(2[z_R - z_{S_n}]) + \gamma(-2[z_R - z_{S_n}])] \\ & + \frac{\rho}{8} \sum_{n \neq m=1}^N \sqrt{r_{S_n} r_{S_m}} [\gamma(2[z_{S_n} - z_{S_m}]) + \gamma(-2[z_{S_n} - z_{S_m}])]. \end{aligned} \quad (2.13)$$

donde  $\gamma(z) = \text{FT}_k\{S(k)\}$  es la función de coherencia de la fuente de luz. El primer término representa el componente de fondo, el segundo término contiene la información del perfil reflectivo y la posición de los reflectores, que es de interés en OCT, y el tercer término describe los artefactos de interferencia entre los reflectores de la muestra.

En SD-OCT, la resolución axial depende de la longitud de coherencia de la luz, relacionada con el ancho de banda espectral. La función  $S(k)$  típicamente sigue una distribución gaussiana la cual está definida por la ecuación 2.11. La resolución axial está directamente influenciada por el ancho de la función de coherencia  $\gamma(z)$ , lo cual afecta la capacidad para distinguir detalles en la estructura de la muestra.

### 2.3.2. Swept Source OCT

En SS-OCT, en lugar de un espectrómetro, se utiliza una fuente de luz sintonizable que barre en longitudes de onda a lo largo de un amplio espectro [47]. La señal interferométrica  $I_D(k)$  se adquiere mediante un fotodetector de un solo pixel mientras la longitud de onda de emisión central  $k$  de la fuente de luz se barre de manera controlada. Este proceso también se conoce como imagen en el dominio de frecuencia óptica (OFDI).

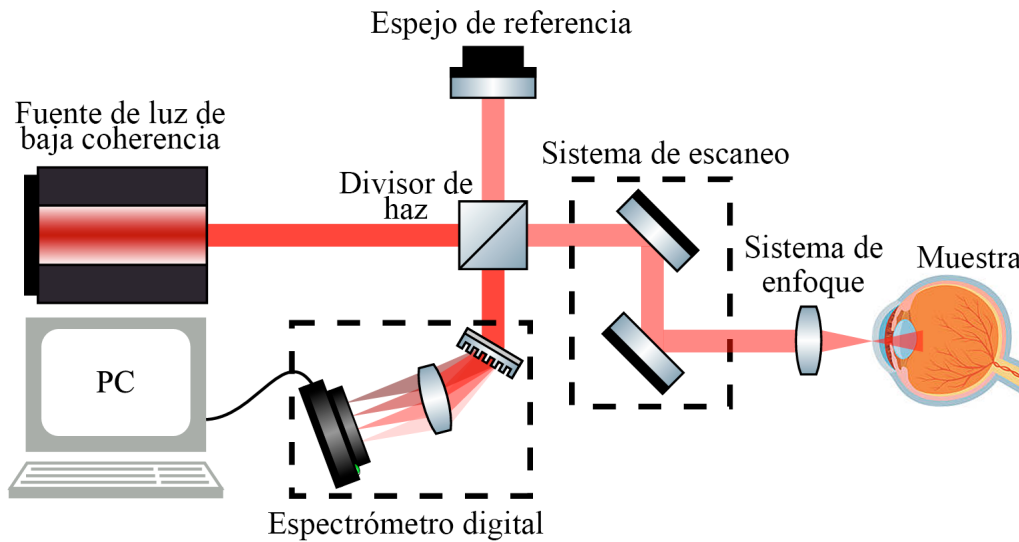


Figura 2.3: Esquema de un sistema SD-OCT basado en un interferómetro de Michelson, donde la adquisición axial se obtiene al calcular la transformada de Fourier del interferograma espectral capturado con un espectrómetro digital. Adaptado de [9]

El método de barrido espectral permite alcanzar tasas de adquisición altas, alcanzando frecuencias de hasta 200 kHz o incluso en el orden de los MHz con instrumentación especializada [47, 49]. Esto mejora significativamente la velocidad de adquisición en comparación con TD-OCT [24], ya que la señal interferométrica se obtiene de forma continua en el tiempo mientras se barre la longitud de onda. En contraste, SD-OCT adquiere múltiples espectros interferométricos de forma discreta, lo que limita su velocidad de adquisición en comparación con SS-OCT. La señal adquirida se transforma posteriormente mediante la transformada de Fourier para obtener el perfil de profundidad de la muestra [43].

Una ventaja clave de SS-OCT es que permite la utilización de fuentes de luz altamente sintonizables con un espectro estrecho, lo cual puede mejorar la calidad de la señal y la sensibilidad [49]. Sin embargo, la velocidad del barrido está limitada por la tasa de adquisición de la fuente de luz y del sistema de detección.

Ambos enfoques, SD-OCT y SS-OCT, presentan la característica de que la transformada de la señal interferométrica obtenida es simétrica respecto a la frecuencia cero debido a las propiedades hermíticas de la transformada de Fourier de señales reales [47]. Esto lleva a la aparición de artefactos espejo, los cuales pueden ser filtrados o eliminados mediante diversos métodos, los cuales se discutirán en los siguientes capítulos. Es importante señalar que, aunque la validación experimental de la metodología propuesta en esta tesis se realice con datos adquiridos mediante un sistema SD-OCT, el problema de los artefactos espejo que se busca resolver es

una característica fundamental presente tanto en ambos sistemas.

## 2.4. Forward Model

El modelo directo es una representación numérica o simulación computacional del proceso de formación de imágenes en OCT, que describe cómo se genera la señal de OCT medida en función de la estructura subyacente del tejido, las propiedades ópticas y los parámetros del sistema OCT, como el escaneo confocal [50] [51]. En otras palabras, el modelo directo puede predecir la señal OCT dada una estructura de tejido conocida y una configuración del sistema. Este modelo típicamente tiene en cuenta la propagación de la luz a través del tejido, considerando factores como la dispersión, absorción y el índice de refracción, además de simular el proceso de interferencia y el proceso de detección y digitalización del sistema [9]. Asimismo, incorpora pasos de procesamiento de señal, como la transformación de Fourier, para generar las imágenes tomográficas finales.

De acuerdo con los principios de óptica de Fourier, la señal de OCT obtenida en el plano del detector con coordenadas transversales  $(x, y)$  para un número de onda específico  $k$  se puede describir mediante la convolución de la función de punto esparcido (PSF) del sistema, denotada como  $h(x, y, z; k)$ , con el potencial de esparcimiento de la muestra, representado como  $\eta(x, y, z)$ . Por tanto la señal de OCT se puede expresar matemáticamente como  $S(x, y; k) = h(x, y, z; k) \otimes \eta(x, y, z)$ , donde el símbolo  $\otimes$  denota la convolución [52]. Por lo tanto,  $S(x, y; k)$  se puede describir como:

$$S(x, y; k) = \iiint h(x - x', y - y', z'; k) \eta(x', y', z') dx' dy' dz', \quad (2.14)$$

OCT se basa en la geometría de doble paso de reflexión, por lo que los haces incidentes y detectados son los mismos, y por ende, podemos decir que  $g_i(x, y, z; k)$  y  $g_d(x, y, z; k)$  son iguales a  $g(x, y, z; k)$  [50]. La integración sobre  $z'$  indica que la luz es capturada simultáneamente en todas las profundidades, lo cual es una característica de la detección en el dominio de Fourier.

La función  $h(x, y, z; k)$  se puede expresar como el producto de los campos complejos incidente y detectado. Esto se puede expresar como  $h(x, y, z; k) = k^2 |A(k)|^2 g^2(-x, -y, z; k)$ , donde  $|A(k)|^2$  denota la densidad del espectro de potencia de la fuente. Debido a la geometría de reflexión, las coordenadas laterales  $(x, y)$  pueden invertirse en  $(x, y) = (-x, -y)$ . Para derivar un modelo para el haz  $g(x, y, z; k)$ , consideramos que el haz es un haz gaussiano normalizado en el plano focal  $z = z_0$  y las coordenadas transversales son  $r_0 = (x, y, z_0)$ . Por lo tanto, podemos expresar el haz como la ecuación 2.15:

$$g_0(\mathbf{r}_0; k) = \frac{1}{2\pi w_0^2(k)} e^{-r_0^2/[2w_0^2(k)]}, \quad (2.15)$$

donde  $w_0(k) = \alpha/k$  y representa el radio de la cintura del haz para el número de onda  $k$ , siendo  $\alpha = \pi/NA$ , donde  $NA$  es la apertura numérica del sistema. Usando la descomposición en ondas planas con la coordenada de frecuencia transversal  $q = (q_x, q_y, 0)$ , el haz en  $r = (x, y, z)$  en cualquier plano  $z$  se describe como [50],

$$g(\mathbf{r}; k) = \frac{1}{(2\pi)^2} \iint e^{i(z-z_0)\sqrt{k^2-\mathbf{q}^2}} e^{-\mathbf{q}^2\alpha^2/2k^2} e^{i\mathbf{q}\cdot\mathbf{r}} d^2\mathbf{q}. \quad (2.16)$$

Para modelar la señal de interferencia, podemos reemplazar la expresión de  $h(x, y, z; k)$  en la ecuación 2.14, obteniendo la expresión de la ecuación 2.17, descrita en detalle en [53],

$$S(\mathbf{r}'; k) = \frac{A(k)}{(2\pi)^2 k} \iiint f^2(\mathbf{r} - \mathbf{r}'; k) \eta(\mathbf{r}) d^2\mathbf{r} dz, \quad (2.17)$$

El término  $f^2(r; k)$  está dado por:

$$f^2(\mathbf{r}; k) = \frac{1}{8\pi^2} \left( \frac{\alpha^2}{k^2} + \frac{iz}{k} \right)^{-1} \iint e^{-\mathbf{q}^2\alpha^2/4k^2} e^{iz\sqrt{4k^2-\mathbf{q}^2}} e^{-i\mathbf{q}\cdot\mathbf{r}} d^2\mathbf{q}. \quad (2.18)$$

Finalmente, la contribución de todos los dispersores puntuales en la muestra, ponderada por la función  $f^2(r - r'; k)$  y escalada por un valor proporcional a la intensidad de la fuente de luz  $A(k)$ , está representada por la señal  $S(\mathbf{r}'; k)$ , la cual se mide para el número de onda instantáneo  $k$  cuando el haz está ubicado en la posición instantánea  $\mathbf{r}'$ . Esta señal luego se utiliza para reconstruir una imagen de la muestra, la cual puede proporcionar información sobre la estructura interna y las propiedades de la muestra.

La Figura 2.4 muestra una simulación basada en el modelo directo de OCT. En la imagen de la izquierda, se representan puntos reflectores distribuidos en capas y esferas que simulan estructuras tisulares y vasos sanguíneos. La imagen de la derecha corresponde a la reconstrucción OCT generada a partir de esta muestra simulada. El algoritmo utilizado para producir estas imágenes fue implementado inicialmente en [9] y posteriormente generalizado a tres dimensiones en [26].

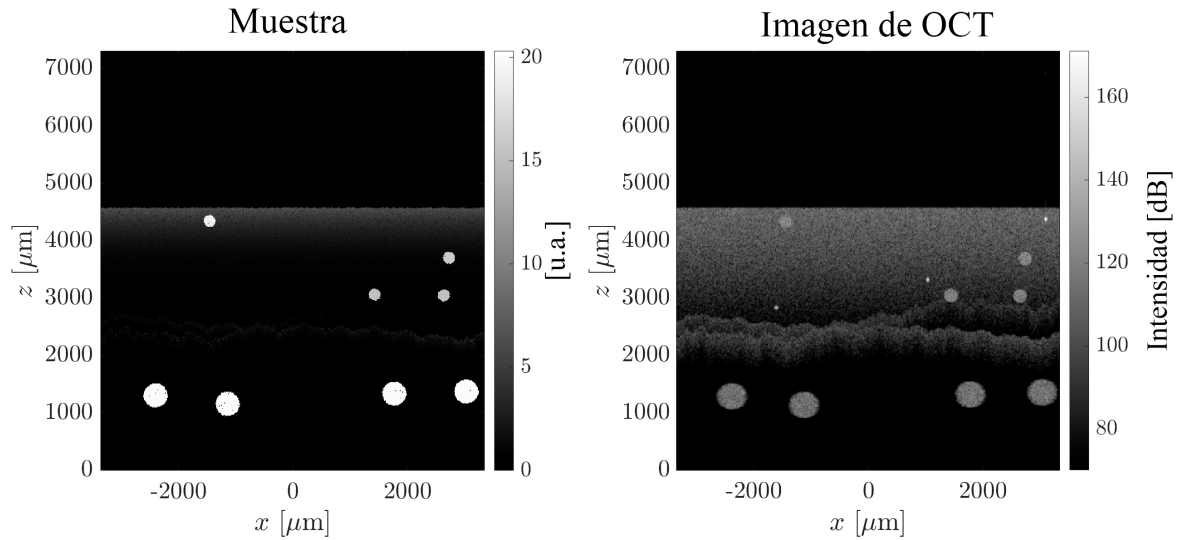


Figura 2.4: Simulación de una *B-scan* usando el modelo directo de OCT.

## 2.5. Muestreo Axial y Lateral en OCT

La adquisición rápida de imágenes en OCT es fundamental para mitigar artefactos y desenfoques inducidos por el movimiento del sujeto [10]. Este proceso involucra dos dimensiones principales de muestreo: axial (profundidad) y lateral (horizontal).

### 2.5.1. Adquisición Axial (A-lines)

La tasa de adquisición de líneas de interferencia (*A-lines*) varía significativamente entre las arquitecturas OCT. En los sistemas SS-OCT, esta tasa está directamente ligada a la velocidad de barrido de la fuente de luz, representando un equilibrio entre exposición, sensibilidad y ancho de banda espectral [11]. Los sistemas de OCT ultrarrápidos han alcanzado frecuencias en el rango de MHz [12], pero las limitaciones en los mecanismos de escaneo y digitalización aún influyen en la calidad de la imagen y la velocidad de adquisición [13]. En contraste, la adquisición axial en sistemas SD-OCT se basa en la dispersión espectral de la luz reflejada y su detección simultánea, lo que también impone consideraciones sobre la sensibilidad y la resolución axial.

### 2.5.2. Adquisición Lateral (Escaneo)

La adquisición en las dimensiones laterales para la creación de imágenes *en-face* y de B-scans se realiza comúnmente mediante escáneres, como los espejos galvanométricos. Estos sistemas emplean distintas tasas de muestreo para los ejes rápido y lento, lo cual es crítico para obtener imágenes de alta calidad [14–16]. La

implementación de escáneres resonantes ha demostrado mejorar la velocidad de adquisición lateral sin problemas de submuestreo, manteniendo la fidelidad espacial [24]; sin embargo, su costo y complejidad pueden limitar su viabilidad. En sistemas con espejos galvanométricos, la baja velocidad de escaneo lateral puede provocar pérdida de correlación temporal, generando discontinuidades de fase y degradación de la señal [54, 55].

Reducir el tiempo total de adquisición, manteniendo la fidelidad en la reconstrucción tanto axial como lateral, es crucial, especialmente en campos de visión amplios [56]. Para minimizar el tiempo sin comprometer la calidad, la densidad de *A-lines* y la velocidad de escaneo lateral deben ajustarse adecuadamente para capturar la información necesaria que represente la continuidad del tejido y evite el aliasing.

### 2.5.3. Condiciones de Muestreo y Aliasing en OCT

Como se menciona en la sección 1.1.1, el aliasing surge cuando la tasa de muestreo es insuficiente, resultando en un solapamiento de frecuencias que distorsiona la reconstrucción espacial de la imagen, lo cual afecta especialmente la resolución lateral. El submuestreo lateral, por debajo de la tasa de Nyquist, implica que el haz no se solapa entre escaneos en el tejido, lo que reduce el factor de superposición a 0%. Esta estrategia de muestreo, aunque reduce el tiempo de adquisición, genera discontinuidades estructurales en la imagen y afecta la precisión diagnóstica.

Para abordar esta discontinuidad espacial entre estructuras finas generada por el submuestreo en la dirección de las B-scans o eje lento, se propone utilizar la información adyacente en las *A-lines* para inferir las estructuras faltantes y lograr la continuidad de la imagen. Este enfoque se puede considerar un problema heurístico, donde se conocen los límites entre las regiones pero no la información intermedia. El submuestreo lateral aplicado, cumple con la condición de Nyquist solo en dirección de las *A-lines* o eje rápido, lo que permite inferir la información espacial faltante entre las B-scans y reconstruir de manera precisa la estructura del tejido [26, 57].

### 2.5.4. Solapamiento de Haces y Continuidad de la Imagen

El solapamiento de los haces en OCT es esencial para mantener la continuidad estructural de la imagen. Cuando el sistema cumple con la condición de Nyquist, el factor de superposición del haz focalizado sobre el tejido alcanza un 50%, lo que permite que las *A-lines* adyacentes contengan información redundante para garantizar la reconstrucción precisa de la señal. El submuestreo lateral, sin embargo, elimina esta redundancia en uno de sus ejes y reduce el factor de superposición al 0%, dificultando la obtención de detalles finos y la continuidad de la estructura en

la imagen. La Figura 2.5 compara dos patrones de escaneo en OCT bajo diferentes condiciones de muestreo. A la izquierda de la imagen, se observa el escaneo bajo la condición de Nyquist, donde la distancia entre haces ( $d$ ) es igual a  $\Delta x/2$ , lo que asegura un solapamiento adecuado entre los haces y evita el aliasing en ambos ejes. A la derecha, se presenta el escaneo bajo la condición sub-Nyquist, en la cual la distancia entre haces es  $\Delta x$  en el eje  $y$ , produciendo aliasing y pérdida de resolución espacial en esta dirección.

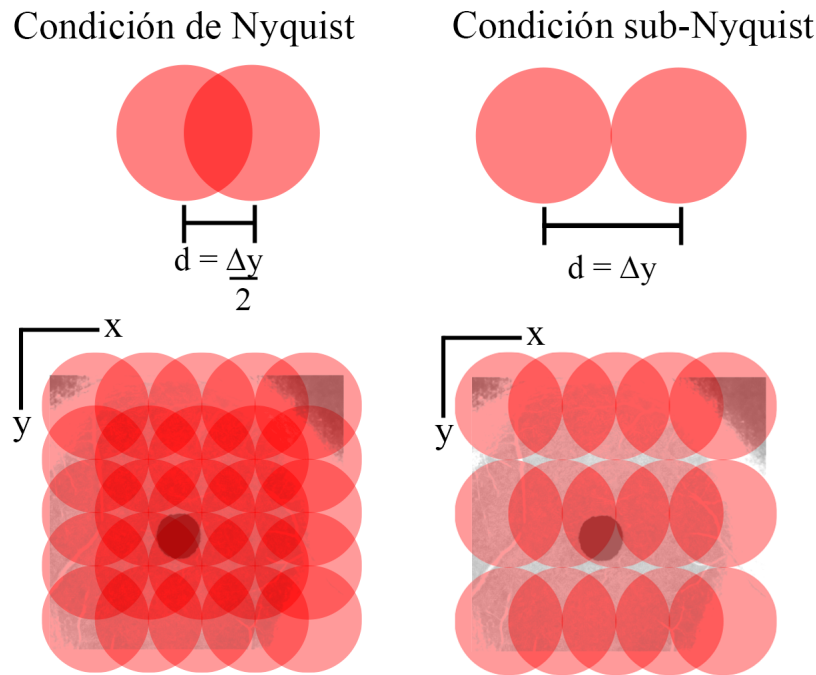


Figura 2.5: Patrón de escaneo en OCT bajo condiciones de muestreo de Nyquist y sub-Nyquist.

El submuestreo lateral se presenta como una estrategia eficiente para reducir el tiempo de adquisición en OCT, ya que permite adquirir menos B-scans en el eje lateral, el cual suele ser el componente más lento del proceso de escaneo. Sin embargo, aunque esta técnica conserva la velocidad de adquisición, puede generar una pérdida significativa de información espacial, afectando tanto la calidad de las imágenes como la precisión diagnóstica debido a la distorsión en la información real.

Para compensar los efectos de la reducción del factor de superposición, se han desarrollado técnicas que aprovechan la información adyacente de las *A-lines* para inferir las estructuras ausentes, permitiendo así que, incluso bajo condiciones de submuestreo, se recupere la señal original y se preserve la resolución lateral necesaria para un diagnóstico adecuado. Históricamente, la investigación en OCT se ha enfocado en disminuir el tiempo de adquisición mediante el aumento de la tasa de escaneo axial, pasando de los sistemas TD-OCT a los sistemas FD-OCT, e incluso alcanzando

el rango de MHz en sistemas de investigación [58]. Sin embargo, estas mejoras a menudo implican un encarecimiento de los sistemas, dificultando su implementación comercial [26]. En contraposición, un enfoque computacional busca disminuir el tiempo de adquisición reduciendo el número de líneas necesarias para el tomograma completo, es decir, adquiriendo datos submuestreados y utilizando métodos para su reconstrucción.

Desde la perspectiva del aprendizaje profundo, el problema de reconstruir información faltante debido al submuestreo lateral en OCT presenta similitudes con el problema de superresolución, donde se busca estimar una imagen de alta resolución a partir de una de baja resolución [26]. En este contexto, se han explorado arquitecturas como las Redes Generativas Adversarias (GANs) [59], que han demostrado un desempeño notable en tareas de superresolución, aunque su complejidad y costo computacional pueden ser considerables [26].

En el ámbito específico de la aplicación de aprendizaje profundo a OCT, se han investigado métodos para la reconstrucción de tomogramas submuestreados en profundidad (eje axial) [60], lo cual difiere del submuestreo lateral que se aborda en este trabajo. Un enfoque más cercano se encuentra en la tomografía fotoacústica, donde se ha aplicado una Fully Dense U-Net para la reconstrucción de imágenes submuestreadas espacialmente [61]. Sin embargo, esta última aplicación se limita al procesamiento de la intensidad de la señal, mientras que el presente trabajo se enfoca en el procesamiento del campo complejo completo de la señal OCT, incluyendo tanto la amplitud como la fase retro-esparcida por la muestra [57].

De esta manera, el uso de métodos basados en aprendizaje profundo se presenta como una línea de investigación activa y prometedora para abordar las limitaciones impuestas por el submuestreo en diversas modalidades de imagenología, incluyendo la OCT, ofreciendo el potencial de optimizar la calidad de la imagen y su valor diagnóstico [26, 57].

## 2.6. Formación de Artefactos de Espejos en OCT

En un sistema OCT, se adquiere la intensidad de la luz retrodispersada por la muestra, lo que implica que la señal medida en el detector es real. Como se menciona en la sección 1.1.2, la naturaleza real de la señal lleva a la aparición de artefactos conocidos como artefactos espejo, los cuales son generados por la simetría inherente en la Transformada de Fourier de señales reales.

Desde un punto de vista físico, esto significa que: por cada punto en la muestra, se generan dos señales, una correspondiente a la verdadera posición de profundidad y otra a una imagen especular de esta, conocida como conjugado complejo. Este fenómeno puede representarse mediante la siguiente expresión, que muestra la contribución de

ambas partes del espectro,

$$S(f) = \frac{1}{2} [S_{\text{pos}}(f) + S_{\text{neg}}(-f)], \quad (2.19)$$

donde  $S_{\text{pos}}(f)$  representa el espectro en frecuencias positivas y  $S_{\text{neg}}(-f)$  el espectro reflejado en frecuencias negativas. En el dominio espacial, este comportamiento da lugar a la aparición de artefactos espejo, en los cuales las estructuras de la muestra son visibles tanto en su posición real como en una posición especular.

Como consecuencia, solo la mitad del rango de profundidad disponible es utilizada de manera efectiva para representar la información real de la muestra, mientras que la otra mitad queda ocupada por la imagen especular del conjugado complejo. Este comportamiento limita el uso completo del sistema OCT, ya que reduce el rango de profundidad utilizable para obtener imágenes claras y sin duplicados [32, 62, 63].

## 2.7. Supresión de artefactos espejo durante la adquisición

Como se ha mostrado, los artefactos se originan debido a que en OCT solo se adquiere la parte de intensidad de la señal compleja. Dado que los instrumentos utilizados para la captura de datos necesariamente adquieren datos con valores reales, la reconstrucción del campo complejo en OCT se vuelve difícil de determinar. Es importante señalar que la corrección del artefacto espejo en OCT no se limita a una simple operación bidimensional sobre B-scans individuales. Dado que el artefacto espejo se genera a lo largo de cada A-line, su manifestación en un volumen tridimensional por ende es difícil de determinar. La estructura del tejido biológico raramente presenta una morfología plana y uniforme con una orientación espacial definida. Por lo tanto, la posición y la forma del artefacto espejo varían a través de los diferentes B-scans que componen el volumen. Una corrección simplificada que asuma una relación directa entre el artefacto en un B-scan y su correspondiente en otros B-scan puede resultar inadecuada y potencialmente introducir errores, especialmente en tejidos con estructuras complejas y no uniformes. Esta consideración subraya la necesidad de métodos de corrección robustos que tengan en cuenta la naturaleza volumétrica del artefacto espejo y la variabilidad de las estructuras tisulares en las tres dimensiones.

### 2.7.1. *Phase shifting*

Un método para estimar la componente imaginaria a partir del valor real registrado mediante la adquisición de múltiples interferogramas en el mismo punto, pero con un desplazamiento de fase relativo [64, 65]. Esto se puede realizar de varias maneras. En [66] se ha reportado el uso de un transductor piezoeléctrico (PZT) para desplazar

el espejo de referencia una distancia conocida, así como un enfoque descrito en [67], que cambia la fase utilizando un modulador optoeléctrico.

El uso de un PZT para inducir el desplazamiento de fase tiende a limitar la velocidad de adquisición, ya que estará restringido por el tiempo de respuesta del PZT [4, 68], por lo tanto, el desplazamiento de fase con un PZT es más sensible al movimiento del paciente que cuando se utiliza, por ejemplo, un modulador electro-óptico [64].

La técnica conocida como *phase-shifting OCT* tiene una desventaja en comparación con otras técnicas, ya que se requieren varios interferogramas en el mismo punto de la muestra, lo que incrementa el tiempo de adquisición debido al aumento en el número de interferogramas. Algunos algoritmos de *phase-shifting* utilizados en OCT requieren cinco interferogramas para recuperar la señal compleja [69], lo que aumenta el tiempo necesario para capturar un tomograma en un factor de 5. Otros algoritmos que han sido desarrollados, solo requieren dos valores de fase relativos, esto con el fin reducir el tiempo de adquisición de las imágenes. Sin embargo, estos algoritmos solo son aplicables para anchos de banda limitados [70].

### 2.7.2. *Phase shifting* con doble brazo de referencia

En [66] se describe una metodología que reduce el tiempo de adquisición al capturar simultáneamente dos interferogramas para reconstruir la señal compleja, mitigando así los problemas asociados con los movimientos involuntarios de los pacientes. La Figura 2.6 muestra una configuración donde se implementan dos brazos de referencia a diferentes distancias de la frecuencia cero para capturar dos imágenes con la fase relativa del "mismo" punto de la muestra. Posteriormente, se emplea un algoritmo de dos pasos, como se muestra en [71], para recuperar la señal compleja.

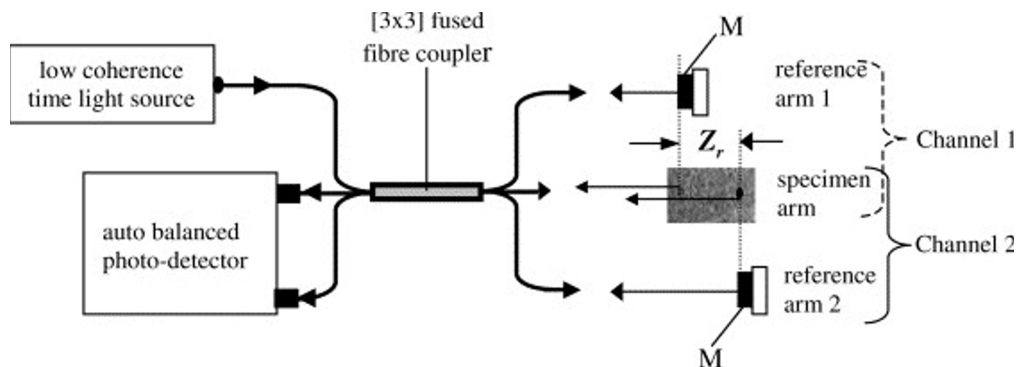


Figura 2.6: OCT con *phase shifting* utilizando un acoplador de fibra 3x3. Tomado de [66]

A continuación, se presentan los resultados reportados en [66] al utilizar un acoplador de fibra 3x3 en un sistema SS-OCT. En la Figura 2.7 se observa una

B-scan de un tomograma del segmento anterior del ojo. En este se puede ver que la profundidad de la imagen se duplicó, pasando de una profundidad de 4 mm a 8 mm, obteniendo así una vista completa de la estructura ocular.

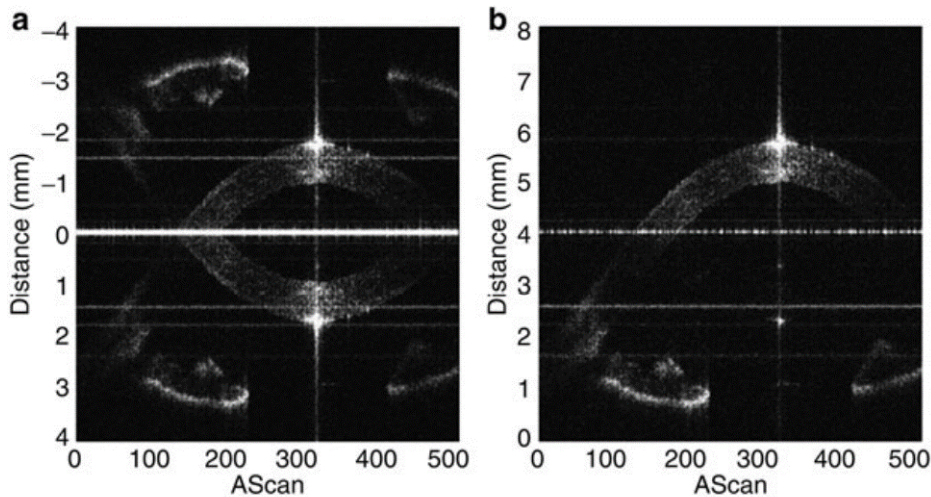


Figura 2.7: Segmento anterior del ojo. En (a) se observa la presencia de la estructura con su respectiva imagen gemela, mientras que en (b) se visualiza únicamente la estructura. La comparación entre ambas imágenes muestra que al eliminar la imagen gemela se duplica el rango de profundidad: en (a) se observa un alcance de 4 mm, mientras que al eliminar el complejo conjugado, se obtiene un rango de 8 mm en el eje de profundidad. Tomado de [66].

### 2.7.3. *Heterodyne* SS-OCT con dispositivos de desplazamiento de frecuencia

Como se mencionó en la sección 2.3.2, los sistemas SS-OCT están compuestos por una fuente de barrido que varía su frecuencia en función del tiempo y un detector, que captura la señal. En [32] se propone añadir una frecuencia de batimiento  $\omega_d$  al brazo de referencia. De esta manera, los términos de auto-correlación se mantendrán centrados cerca del componente DC, mientras que los términos de correlación cruzada se desplazarán lejos del centro alrededor de la frecuencia portadora  $\omega_d$ . Aunque la transformada de Fourier de la señal desplazada en frecuencia sigue siendo simétrica y hermítica, las franjas desplazadas en frecuencia están lo suficientemente alejadas de la banda base o frecuencia cero, para poder distinguir entre desplazamientos negativos y positivos. Los desplazamientos positivos están por encima de  $\omega_d$ , mientras que los desplazamientos negativos están por debajo de  $\omega_d$ , pero por encima de la banda base, siempre y cuando  $\omega_d$  sea mayor que la frecuencia más alta de la señal. [66].

La Figura 2.8 muestra el esquema de un sistema SS-OCT, con una fuente de barrido centrada en 1300nm y un ancho de banda de 100nm, utilizando también un

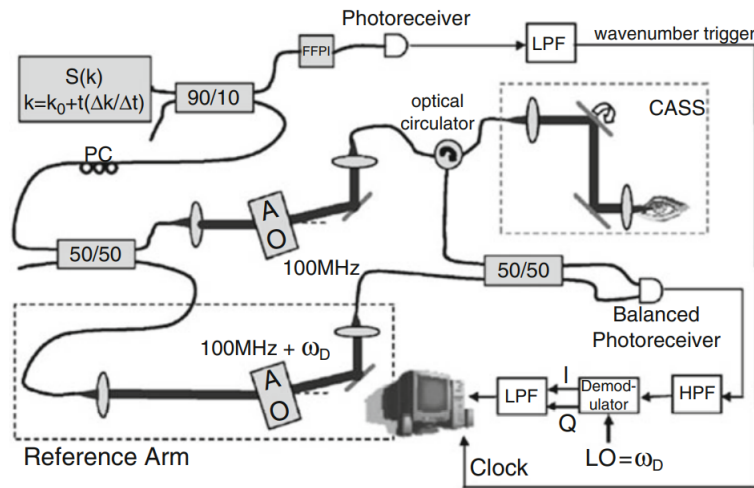


Figura 2.8: Diagrama del sistema Heterodyne SS-OCT que utiliza un modulador optoacústico (AO) para generar un desplazamiento de frecuencia en la señal. Se indican los componentes clave: AO (modulador optoacústico), FFPI (interferómetro de Fabry-Pérot de fibra), LPF (filtro de paso bajo) y CASS (sistema de adquisición de señal coherente). Tomado de [66].

fotodetector balanceado para reducir el ruido. El montaje muestra cómo se utiliza un filtro pasa-bandas alrededor de la frecuencia portadora y su posterior demodulación para recuperar las componentes en fase y en cuadratura de la señal OCT compleja. También se utilizan dos moduladores optoacústicos centrados en 100 MHz, con una frecuencia de desplazamiento ajustable por el usuario en uno de los moduladores para proporcionar la frecuencia portadora.

Después de adquirir la señal, se utiliza un filtro pasa-altos para eliminar las componentes DC y de auto-correlación. La señal filtrada se demodula con un oscilador sintonizado a la misma frecuencia que la portadora. Las componentes en cuadratura se pueden recuperar para reconstruir la señal OCT compleja.

Las Figuras 2.9(a) y 2.9(b) muestran la señal OCT sin desplazamiento de frecuencia, donde se puede ver que no es posible diferenciar entre dos señales ubicadas a  $\pm 50\mu\text{m}$ . En 2.9c y 2.9d se observa una señal OCT modulada por una frecuencia de 20kHz. En esta situación, ya es posible diferenciar entre las dos señales, lo que permite filtrar una de las copias para obtener información en todo el rango.

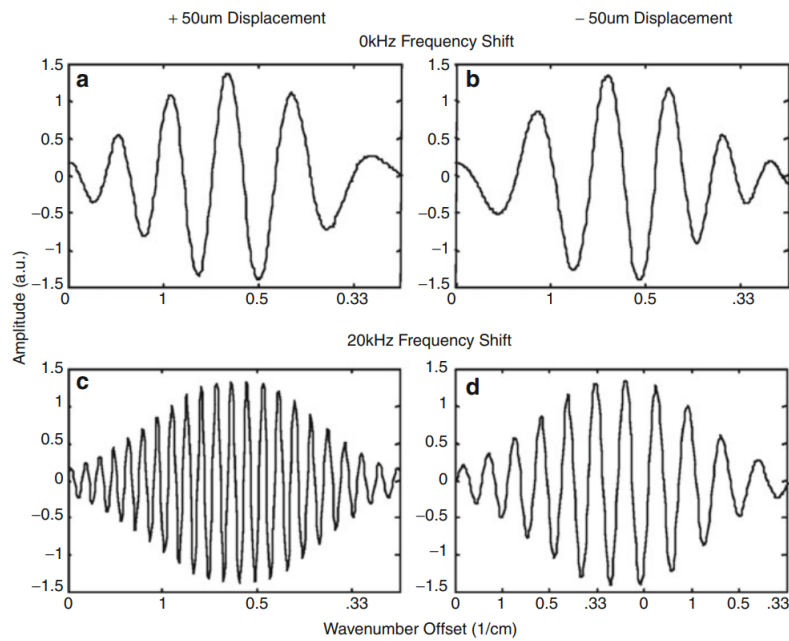


Figura 2.9: Comparación entre una señal sin y con *Heterodyne* OCT. Tomado de [66].

## 2.8. Supresión de artefactos espejo en post-procesamiento

Como se evidencia en la sección anterior, los métodos basados en instrumentación proporcionan una solución al problema del artefacto espejo. Sin embargo, requieren elementos ópticos adicionales que aumentan el costo de los sistemas OCT. Algunos otros métodos, como los basados en el desplazamiento de fase, incluso pueden aumentar el tiempo de adquisición, lo que los hace sensibles a los movimientos involuntarios del paciente, produciendo sobreexposición y errores de enfoque. Una alternativa a esto se encuentra en los métodos de post-procesamiento, donde no se compromete el tiempo de adquisición.

### 2.8.1. Algoritmo de DEFR

El algoritmo de *Dispersion encoded full range* (DEFR)-OCT o en español como: Rango completo codificado por dispersión. Es un algoritmo iterativo que utiliza la dispersión entre los brazos de referencia y de la muestra en FD-OCT para cancelar efectivamente los complejos conjugados en cada una de las *A-lines* y generar tomogramas de rango completo. El esquema de DEFR-OCT, se presenta en [63]. Este algoritmo permite distinguir las estructuras reales de los artefactos espejo producidos por los complejos conjugados. Aunque DEFR-OCT tiene una mayor complejidad de post-procesamiento, no se compromete la velocidad de adquisición, ya que no se requieren mediciones adicionales de *A-lines*. Esto hace que esta técnica sea robusta a las fluctuaciones de fase. El algoritmo utiliza compensación numérica

de la dispersión y muestra una resolución similar al conseguido con métodos basados en instrumentación [63].

Los elementos residuales de los términos de la imagen especular se reducen aún más al incorporar información adicional, como el espectro de potencia de la fuente utilizada en la adquisición del tomograma. La relación de supresión de los artefactos espejo supera los 50 dB y, por lo tanto, es comparable a las técnicas que utilizan instrumentación óptica para adquirir múltiples *A-lines* [72]. El desajuste de dispersión entre los brazos se utiliza para codificar la fase espectral de la señal. Esta codificación hace que la posición de los componentes individuales de la señal compleja correspondientes a las estructuras reales y a los artefactos espejo se desplacen en direcciones opuestas dentro de un rango de  $\pm\Delta x$  en el dominio de profundidad. El algoritmo DEFR resultante explota este desplazamiento diferencial a través de un procedimiento iterativo. Inicialmente, asume que una parte de la señal corresponde a la imagen real y la otra al complejo conjugado. Utilizando el conocimiento previo de la fase de la dispersión inducida en el brazo de referencia, el algoritmo aplica una compensación de dispersión específica. En cada iteración, esta compensación favorece la concentración de la energía de la señal en la posición real de la estructura, mientras que la energía del complejo conjugado se dispersa. Al iterar este proceso, el algoritmo refina la estimación de la posición y la amplitud de las estructuras reales, discriminándolas de los artefactos espejo que no se enfocan correctamente con la misma compensación de dispersión.

Finalmente, en la Figura 2.10 se observa una comparación entre un grupo de *B-scans* con diferentes grados de solapamiento y su respectiva mitigación de las copias especulares empleando el algoritmo DEFR. La prueba se realizó colocando la muestra a diferentes distancias de la frecuencia cero, esto con el fin de evaluar el método en diferentes condiciones de solapamiento entre los artefactos y la estructura real. Se puede ver cómo en los dos primeros casos la estructura y la copia están separadas entre sí. Esto hace posible eliminarlas simplemente realizando un recorte a la altura de la frecuencia cero. En las dos últimas pruebas, la estructura y los artefactos espejo están uno sobre el otro, lo que hace imposible realizar este procedimiento. En el caso de las imágenes de la segunda fila en la figura 2.10 se observa una reducción significativa de los artefactos espejo producidos por los complejos conjugados, obteniendo una imagen adecuada de la fovea.

Si bien el método DEFR representa una estrategia efectiva para la reducción de artefactos espejo, presenta ciertas limitaciones inherentes en comparación con las aproximaciones basadas en Deep Learning. En primer lugar, la implementación del método DEFR requiere la adquisición previa de información adicional para caracterizar el perfil de dispersión del brazo de referencia, un paso necesario para corregir el ensanchamiento axial de las estructuras reales. Esto depende de una medición de

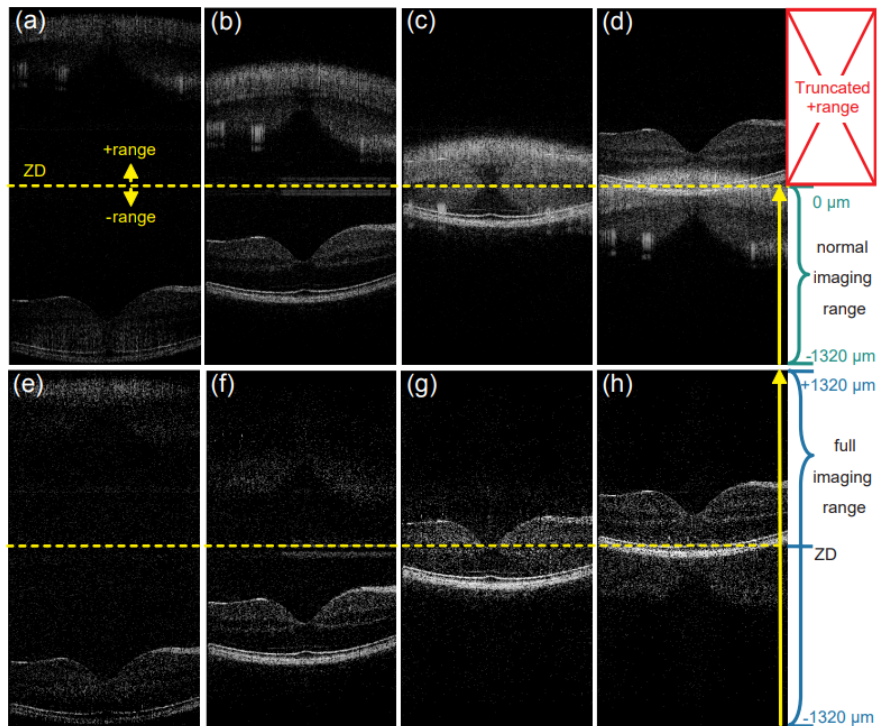


Figura 2.10: *B-scans* que ilustran el efecto del solapamiento en la visualización de la fovea y la capacidad del método DEFR para mitigar la imagen especular. La fila superior (a-d) muestra la misma estructura con diferentes grados de solapamiento, mientras que la fila inferior (e-h) presenta la misma estructura con el método DEFR aplicado, mostrando una reducción significativa de la imagen especular y mejorando la visualización de la fovea. Tomado de [72].

calibración específica para cada sistema de OCT lo cual introduce una complejidad adicional y limita su aplicabilidad directa a diferentes configuraciones. En segundo lugar, la naturaleza iterativa del proceso de corrección y eliminación de la imagen gemela, aplicada A-line por A-line, si bien puede beneficiarse del procesamiento en paralelo mediante GPUs, aún impone restricciones en el procesamiento en tiempo real. Además, el criterio de detención del algoritmo iterativo se basa en la evaluación de la atenuación de la imagen gemela en relación con la estructura real, lo que introduce una no deterministicidad en el número de iteraciones necesarias para una corrección adecuada. En contraste, las técnicas de Deep Learning ofrecen la ventaja de eliminar el complejo conjugado sin la necesidad de información adicional específica del sistema, aprendiendo directamente de los datos existentes. Sumado a las capacidades de procesamiento paralelo inherentes a los *frameworks* de aprendizaje profundo, esta característica permite una eliminación de artefactos potencialmente más eficiente y en principio generalizable a diferentes sistemas de adquisición.

# FUNDAMENTOS DE APRENDIZAJE PROFUNDO

El aprendizaje profundo se ha convertido en una herramienta esencial en el campo de las aplicaciones médicas debido a su capacidad para procesar grandes cantidades de datos y extraer características complejas de forma automática. Desde la detección de enfermedades a través de imágenes médicas hasta el análisis de señales fisiológicas, los algoritmos de aprendizaje profundo han demostrado superar las metodologías tradicionales en tareas como la clasificación, segmentación y reconstrucción de imágenes [73, 74]. En el contexto de OCT, el aprendizaje profundo ha abierto nuevas posibilidades para mejorar la calidad de las imágenes y la precisión de los diagnósticos, enfrentando desafíos específicos como el ruido, la resolución limitada y la presencia de artefactos.

En el ámbito de OCT, varias aplicaciones han aprovechado las ventajas del aprendizaje profundo para resolver problemas particulares. Un ejemplo de ello es la eliminación de ruido de correlación que genera *speckle* o *despeckling* en las imágenes de OCT [75, 76], que mejora significativamente la claridad de las estructuras analizadas. Otra aplicación destacada es la supresión del artefacto del complejo conjugado reportadas en [28, 77], en las cuales para abordar este problema, se han desarrollado múltiples métodos que emplean aprendizaje profundo, los cuales serán discutidos en detalle más adelante en este capítulo. Además de las metodologías para eliminar el artefacto espejo, existen otras aplicaciones relevantes en las que se emplea aprendizaje profundo en OCT, tal como se reporta en [78] que busca mejorar la resolución axial en un sistema SD-OCT, y las publicadas en [79], en donde no solo buscan reducir el ruido sino que también busca reducir la sombra en la retina, las cuales pueden obstruir la vista de estructuras importantes. Este rango de aplicaciones demuestran la amplia adaptabilidad y eficacia del aprendizaje automático para mejorar los tiempos de adquisición y procesamiento de datos en OCT.

### 3.1. Fundamentos del Aprendizaje Profundo

Para comprender la aplicación de las redes neuronales en OCT, es fundamental familiarizarse con algunos conceptos técnicos esenciales que guían el proceso de entrenamiento y optimización de estos modelos.

#### Funciones de pérdida

Uno de los elementos clave en el entrenamiento de redes neuronales es la función de pérdida. Esta función mide la diferencia entre las predicciones del modelo y los valores reales, proporcionando una métrica que guía el proceso de aprendizaje [80]. En el contexto de OCT, se suelen emplear varias funciones de pérdida que permiten al modelo aprender a reconstruir imágenes de alta fidelidad.

Las funciones de pérdida más comunes incluyen el Error Cuadrático Medio (*Mean Squared Error*, MSE) y el Error Absoluto Medio (*Mean Absolute Error*, MAE). El MSE se calcula promediando el cuadrado de las diferencias entre los valores predichos y los reales. Esto da más peso a los errores grandes, lo que puede ser útil para penalizar discrepancias significativas en la reconstrucción de estructuras finas en los tomogramas. Por otro lado, el MAE calcula el promedio de las diferencias absolutas, siendo menos sensible a los valores atípicos, lo cual es útil cuando se busca un modelo robusto ante pequeñas variaciones en la señal [81].

En el caso específico de OCT, también se utilizan funciones de pérdida más avanzadas, como la pérdida adversaria en arquitecturas generativas, que fomenta la generación de imágenes más fidedignas a las registradas por el instrumento. Esta pérdida se emplea en redes generativas adversarias condicionales, como la arquitectura *Pix2Pix*, donde el modelo generador busca mejorar la calidad de las imágenes para “engañar” a un discriminador que evalúa si las imágenes generadas son realistas [82]. Además, la pérdida de reconstrucción evalúa qué tan bien el modelo reproduce las estructuras del tejido, jugando un papel crucial en la preservación de detalles importantes [59].

La selección de la función de pérdida es crítica para el rendimiento del modelo, ya que determina la forma en que éste se adapta a los datos de entrada y salida. Una adecuada combinación de funciones de pérdida puede mejorar la calidad de las imágenes reconstruidas y la capacidad del modelo para generalizar a nuevos datos.

#### Optimizador

Los optimizadores son algoritmos que ajustan los pesos de la red neuronal para minimizar la función de pérdida durante el proceso de entrenamiento, guiando al modelo hacia una mejor solución [83]. La forma en que los optimizadores actualizan

los pesos afecta directamente la rapidez y eficacia del aprendizaje.

Uno de los algoritmos más básicos es el *Stochastic Gradient Descent* (SGD), que actualiza los pesos del modelo utilizando el gradiente de la función de pérdida calculado a partir de un subconjunto aleatorio de los datos [84]. Sin embargo, para problemas más complejos, se requieren optimizadores más avanzados que puedan adaptarse mejor a los cambios en la función de pérdida.

Entre los optimizadores más comunes en las aplicaciones de aprendizaje profundo se encuentran Adam y RMSprop. Adam combina las ventajas del SGD con el uso de información de los gradientes pasados, específicamente la media móvil exponencial del gradiente (que estima la dirección promedio del descenso) y la media móvil exponencial del cuadrado del gradiente (que estima la magnitud promedio de los gradientes recientes). Esto permite un ajuste más eficiente y estable de los pesos [85–87] y lo hace especialmente útil en problemas donde los datos presentan ruido o donde la función de pérdida tiene una superficie compleja, con múltiples mínimos locales, mesetas o valles estrechos, lo que dificulta la convergencia para optimizadores más simples como el SGD estándar. Por otro lado, RMSprop adapta la tasa de aprendizaje de cada parámetro, específicamente los pesos y los sesgos de la red neuronal, manteniendo un promedio móvil de los gradientes recientes. Los sesgos son términos adicionales en cada capa de la red que permiten aprender patrones incluso cuando las entradas son nulas, actuando como un valor constante que se suma a la combinación lineal de las entradas ponderadas. Esto permite que el algoritmo ajuste la tasa de aprendizaje de cada conexión de forma individual. Esto resulta en un entrenamiento más rápido y estable en problemas donde las derivadas de la función de pérdida varían significativamente [88].

La elección del optimizador depende de las características específicas del problema y del modelo. En el caso de las imágenes OCT, donde se requiere preservar detalles finos y mantener una estabilidad en la reconstrucción, la selección del optimizador adecuado es crucial para garantizar una convergencia rápida y un resultado óptimo. Para seleccionar el optimizador más adecuado, se llevó a cabo una evaluación comparativa de diferentes optimizadores sobre un conjunto de datos de validación, midiendo su desempeño a través del error cuadrático medio (MSE). Los detalles de esta metodología se presentan en la sección 4.1.

### Tasa de aprendizaje

La tasa de aprendizaje o *learning rate* es un parámetro fundamental en el entrenamiento de redes neuronales, ya que controla el tamaño de los pasos que da el optimizador al actualizar los pesos de la red [89]. Un valor de tasa de aprendizaje demasiado alto puede hacer que el modelo no converja, ya que los cambios en los pesos serán demasiado amplios y pueden saltar por encima del mínimo de la función

de pérdida. Por el contrario, una tasa de aprendizaje demasiado baja puede resultar en un proceso de entrenamiento extremadamente lento, ya que los pasos serán muy pequeños y el modelo podría quedar atrapado en los mínimos locales [90].

Existen varias estrategias para ajustar la tasa de aprendizaje de manera eficiente. Una de las técnicas más comunes es el uso de tasas de aprendizaje adaptativas, donde la tasa se ajusta automáticamente durante el proceso de entrenamiento en función de los gradientes calculados [90]. Optimizadores como Adam y RMSprop incorporan esta característica, lo que les permite adaptarse a diferentes dinámicas de aprendizaje en distintas etapas del proceso.

Otra técnica útil es el decaimiento de la tasa de aprendizaje, donde el valor de la tasa se reduce gradualmente a medida que avanza el entrenamiento. Esto permite que el modelo explore la función de pérdida con pasos más grandes al inicio y realice ajustes más finos al final del proceso [89,91]. En algunos casos, se utilizan estrategias cíclicas, como las tasas de aprendizaje cíclicas, que varían periódicamente la tasa para escapar de mínimos locales y mejorar la convergencia [92–94].

## 3.2. Arquitecturas para la Reconstrucción de tomogramas submuestreados

En el contexto de la reconstrucción de imágenes OCT, las arquitecturas de redes neuronales juegan un papel fundamental. A continuación, se describen las arquitecturas utilizadas en este trabajo y su relevancia para los problemas específicos abordados. En trabajos previos, como se describe en [26], se exploró el uso de arquitecturas U-Net y autoencoders para la reconstrucción de tomogramas submuestreados lateralmente, obteniendo resultados prometedores. En ese estudio, el entrenamiento se llevó a cabo utilizando datos sintéticos y se validó con tomogramas experimentales de la córnea de un cerdo, así como con otros tomogramas sintéticos no presentes en la fase de entrenamiento. Inspirándose en dicha metodología, en este trabajo se propone la incorporación de redes generativas, específicamente el modelo Pix2Pix, como una alternativa a los modelos usados en ese trabajo.

### 3.2.1. Pix2Pix

La arquitectura Pix2Pix es un tipo de Red Generativa Adversaria Condicional utilizada para la traducción de imagen a imagen. Esta arquitectura se compone de dos redes principales: un generador y un discriminador. El generador es el encargado de producir imágenes reconstruidas a partir de las imágenes de entrada, ya sea para **completar los tomogramas submuestreados lateralmente o para eliminar los artefactos espejo presentes en las imágenes OCT**. El discriminador se encarga

de evaluar la autenticidad de las imágenes generadas por el generador, comparándolas con imágenes reales. A través de este proceso de aprendizaje adversario, el generador mejora continuamente la calidad de sus salidas para engañar al discriminador, mientras este último se vuelve más efectivo en distinguir las imágenes falsas de las reales [27, 95].

En la figura 3.1 se muestra el esquemático del generador en la arquitectura Pix2Pix empleada, el cual tiene una estructura basada en un modelo *encoder-decoder*, que se utiliza para capturar y aprender las características de las imágenes de entrada. La parte del *encoder* utiliza una serie de capas convolucionales para extraer características a múltiples escalas, reduciendo progresivamente la dimensionalidad de la imagen de entrada [96]. Durante este proceso, se aplican convoluciones con diferentes números de filtros para captar patrones de distintos niveles de complejidad. El *stride* se utiliza para reducir la resolución de la imagen a medida que avanza el proceso de codificación, permitiendo que el modelo capture características globales de la imagen.

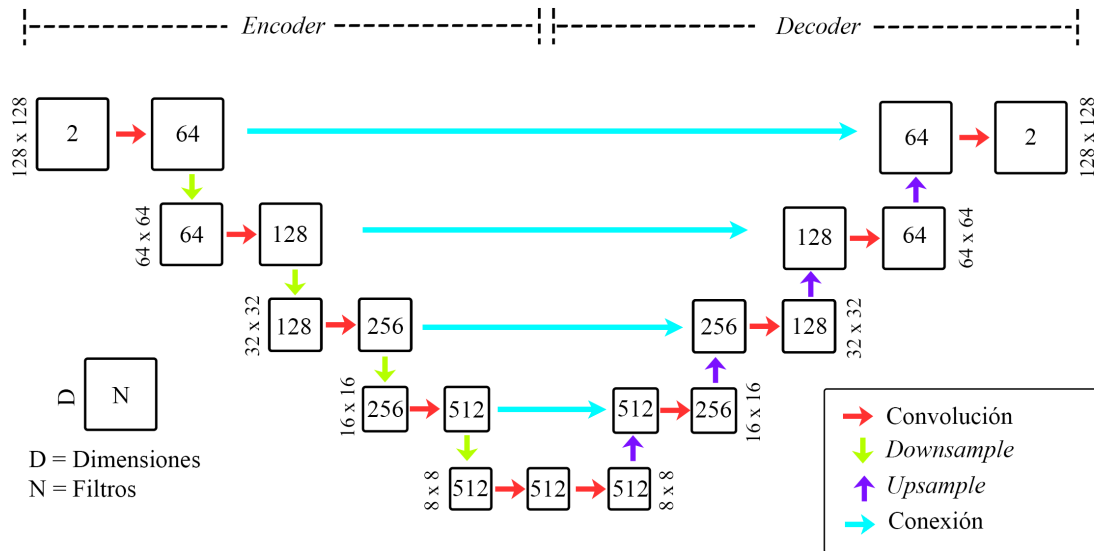


Figura 3.1: Esquema del generador en arquitecturas adversaria es basadas en U-net. Las conexiones en azul cian representan *skip-connections* en la arquitectura Pix2Pix y *gate-connections* en CCAR Net.

En la fase de *decoder*, el generador utiliza capas de convolución transpuesta, también conocidas como deconvoluciones para reconstruir la imagen de salida a partir de las características aprendidas [96]. Aquí es donde se restaura la resolución de la imagen original. Uno de los elementos clave de esta arquitectura es el uso de *skip connections* o conexiones de salto, que conectan directamente las capas correspondientes del *encoder* con las del *decoder*. Estas conexiones permiten transferir información de alta frecuencia del *encoder* al *decoder*, mejorando la precisión y

el detalle en la imagen reconstruida [97]. Esto podría ser especialmente útil en problemas como la reconstrucción de tomogramas submuestreados y en la mitigación de artefactos espejo, donde es crucial preservar tanto los detalles finos como la estructura general.

El discriminador, por otro lado, opera como una red convolucional similar a la fase del *encoder* del generador, la cual evalúa segmentos locales o “parches” de la imagen para decidir si son reales o generados [27]. Específicamente, una PatchGAN compara las predicciones del discriminador para cada parche de forma *element-wise* o elemento a elemento en el espacio latente. Esto permite que la arquitectura se enfoque en la calidad de las características locales, lo cual es fundamental en imágenes OCT, donde la fidelidad en las texturas y las estructuras es crítica para el diagnóstico.

La arquitectura Pix2Pix utiliza una combinación de dos funciones de pérdida para guiar el entrenamiento: la pérdida adversaria y la pérdida L1. La pérdida adversaria se basa en la interacción entre dos redes neuronales: el generador ( $G$ ) y el discriminador ( $D$ ). El objetivo del generador es crear imágenes  $G(x)$  que sean lo suficientemente realistas como para engañar al discriminador, haciéndole creer que son imágenes reales  $y$ . El discriminador, por otro lado, intenta distinguir entre las imágenes reales  $y$  y las imágenes generadas  $G(x)$ . La función de pérdida adversaria se puede expresar como:

$$\mathcal{L}_{adv}(G, D) = \mathbb{E}_{x, y \sim p_{data}(x, y)}[\log D(x, y)] + \mathbb{E}_{x \sim p_{data}(x)}[\log(1 - D(x, G(x)))]. \quad (3.1)$$

El primer término de esta ecuación busca maximizar la probabilidad de que el discriminador clasifique correctamente las imágenes reales. El segundo término busca minimizar la probabilidad de que el discriminador clasifique correctamente las imágenes generadas (es decir, que las clasifique como reales). Esta “competición” entre las dos redes impulsa al generador a aprender a producir imágenes cada vez más convincentes y similares a las reales en términos de su distribución general y características visuales. En esencia, la pérdida adversaria asegura que las imágenes generadas se vean plausibles y coherentes con el conjunto de datos real.

La pérdida L1, también conocida como la norma L1 o la distancia Manhattan, mide la diferencia absoluta entre el valor de cada píxel en la imagen generada  $G(x)$  y el valor del píxel correspondiente en la imagen real  $y$ . Para una imagen de  $N$  píxeles, la función de pérdida L1 se define como:

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y \sim p_{data}(x,y)} [\|y - G(x)\|_1] = \mathbb{E}_{x,y \sim p_{data}(x,y)} \left[ \frac{1}{N} \sum_{i=1}^N |y_i - G(x)_i| \right]. \quad (3.2)$$

Al minimizar esta pérdida, se penalizan las grandes diferencias entre las imágenes a nivel de píxel. Esto tiene el efecto de forzar al generador a producir imágenes que sean muy similares a las imágenes reales en términos de su contenido detallado, incluyendo la estructura y los valores de intensidad de cada punto. A diferencia de otras pérdidas como la L2 (error cuadrático medio), la pérdida L1 tiende a preservar mejor los bordes y los detalles finos en las imágenes reconstruidas, ya que penaliza linealmente las diferencias [98, 99].

La función de pérdida total para el entrenamiento de Pix2Pix es una combinación ponderada de estas dos pérdidas:

$$\mathcal{L}_{total}(G, D) = \mathcal{L}_{adv}(G, D) + \lambda \mathcal{L}_{L1}(G), \quad (3.3)$$

donde  $\lambda$  es un factor de ponderación que controla la importancia relativa de la pérdida L1 en comparación con la pérdida adversaria. Para este trabajo, se adoptó el valor de  $\lambda = 100$ , siguiendo la implementación establecida en la publicación original de Pix2Pix por Isola et al en [100]. y la práctica estándar en implementaciones subsecuentes.

Esta combinación de la pérdida adversaria, que fomenta la generación de imágenes realistas a nivel general, y la pérdida de L1, que promueve la precisión en la reconstrucción de los detalles a nivel de píxel, permite que Pix2Pix sea efectivo tanto para rellenar las áreas submuestreadas en los tomogramas como para eliminar los artefactos espejo, proporcionando una solución integral para mejorar la calidad de las imágenes OCT [95].

### 3.2.2. CCAR Net

La arquitectura CCAR net (Complex Conjugate Artifact Removal net) es una arquitectura generativa adversaria diseñada específicamente para abordar el problema de los artefactos de conjugado complejo en las imágenes OCT, la cual se reporta en [101]. La CCAR net emplea imágenes B-scan en intensidad normalizada entre 0 y 1 como entrada, lo que limita su capacidad para preservar información de fase crucial en ciertas aplicaciones de OCT. Sin embargo, ofrece un enfoque valioso para eliminar artefactos de las imágenes OCT utilizando técnicas avanzadas de aprendizaje profundo.

La CCAR net sigue un enfoque de *encoder-decoder*, similar a la arquitectura Pix2Pix descrita anteriormente la cual se ilustra en la figura 3.1. El *encoder* extrae características de las imágenes de entrada utilizando bloques de convoluciones, reduciendo la resolución a través de capas de *down-sampling* para capturar características globales. Posteriormente, el *decoder* reconstruye la imagen a su resolución original mediante capas de *up-sampling*, utilizando convoluciones transpuestas para recuperar los detalles finos. Este proceso se ve reforzado por las conexiones tipo compuerta o *gate connections*.

Un elemento distintivo de la CCAR net es la modificación de las conexiones tipo *skip* que se encuentran en los modelos U-Net. En lugar de transferir directamente todas las características del *encoder* al *decoder*, la CCAR net emplea *gate connections* para filtrar y controlar la cantidad de información que se transfiere. Estas conexiones se diseñan para realzar los detalles de la estructura en áreas libres de artefactos, evitando la propagación de los artefactos espejo [101].

Para entender cómo funciona este mecanismo, es importante describir las características de bajo nivel ( $x_1$ ) y las características de alto nivel ( $x_2$ ). Las características de bajo nivel ( $x_1$ ) se refieren a los detalles finos en la imagen, como bordes y texturas, que son extraídos por las primeras capas del *encoder*. A diferencia de los métodos de procesamiento de imágenes tradicionales que emplean filtros diseñados manualmente, como los filtros de paso alto para detectar bordes, en una red neuronal convolucional estos filtros son “aprendidos”. Cada capa aplica un conjunto de kernels cuyos coeficientes se optimizan durante el entrenamiento para detectar patrones que minimicen la función de pérdida global. En las capas iniciales, estos filtros aprendidos se especializan en identificar formas visuales simples. Funcionalmente, muchos de ellos se comportan de manera análoga a detectores de bordes, esquinas o gradientes, ya que responden a las altas frecuencias espaciales de la imagen. Por lo tanto, en lugar de un único filtro de paso alto predefinido, la red aprende un banco de filtros diversos que, en conjunto, extraen un rico repertorio de características de bajo nivel. Por otro lado, las características de alto nivel ( $x_2$ ) representan patrones más abstractos, profundos y globales, capturados por las capas más profundas del *encoder* [101, 102]. Estas características codifican información sobre la forma general y la organización estructural del tejido, proporcionando una visión más abstracta de la imagen.

Las *gate connections* en la CCAR Net juegan un papel crucial en la combinación de características de bajo nivel ( $x_1$ ) y alto nivel ( $x_2$ ). Específicamente, toman las características de alto nivel ( $x_2$ ) y, mediante una capa lineal, las transforman para luego dividir las de forma adaptativa en tres variables intermedias:  $g_1$ ,  $g_2$ , y  $g_3$ , tal como se describe en la Ecuación 3.4. Estas variables intermedias permiten a la red controlar el flujo de información y seleccionar qué aspectos de las características de alto nivel son relevantes para modular las características de bajo nivel [101]. La

operación de *gate* utiliza estas variables intermedias para modular la combinación de  $x_1$  y  $x_2$  según la Ecuación 3.5. Este proceso de filtrado implica aplicar funciones de activación tangente hiperbólica ( $\tanh$ ) y sigmoide ( $\sigma$ ), así como una convolución a la variable intermedia  $g_1$ , para generar la salida ( $y$ ) que se concatena con  $x_2$  y sirve como entrada al siguiente bloque de extracción de características. La elección de las funciones de activación es importante porque introducen no linealidad y permiten controlar el rango de los valores de las variables intermedias, actuando como “compuertas” suaves. La función sigmoide ( $\sigma$ ), con su rango de salida entre 0 y 1, es comúnmente utilizada para implementar mecanismos de atención o gating, ya que puede interpretarse como una medida de cuánto “dejar pasar” de la información. La función tangente hiperbólica ( $\tanh$ ), con su rango entre -1 y 1, puede ayudar a modular la información con valores tanto positivos como negativos, lo que puede ser útil para refinar las características. Existen otras funciones de activación que podrían utilizarse en mecanismos de gating, como la ReLU o sus variantes (LeakyReLU, PReLU), aunque la combinación de sigmoide para el gating propiamente dicho y otras no linealidades como  $\tanh$  es frecuente en arquitecturas de atención y fusión de características. Este enfoque permite que la red seleccione de manera precisa qué detalles conservar (mejorando las características de bajo nivel  $x_1$ ) y cuáles descartar, lo que resulta fundamental para eliminar los artefactos espejo sin perder información crítica.

$$g_1, g_2, g_3 = \text{split}(\text{linear}(x_2)), \quad (3.4)$$

$$y = \text{concat}(x_1 \times \tanh(\text{conv}(g_1)), \sigma(g_2), \tanh(g_3)). \quad (3.5)$$

Además, la arquitectura incorpora convoluciones separables en profundidad en el módulo de extracción de características. Esta técnica divide la convolución en dos pasos: una convolución grupal para capturar las relaciones espaciales y una convolución puntual para mezclar características entre canales [103]. Este enfoque permite mejorar la extracción de características relevantes en las imágenes OCT mientras reduce los costos computacionales, algo fundamental cuando se trabaja con datos volumétricos como los obtenidos en OCT.

El discriminador de la CCAR net sigue la arquitectura *PatchGAN*, que se especializa en evaluar la imagen a nivel de parches en lugar de considerar la imagen completa [104]. Esta metodología podría ser particularmente útil en OCT, ya que permitiría que el modelo se enfoque en las texturas y detalles locales esenciales para una reconstrucción fiel. El discriminador se compone de tres capas de convolución progresiva con un *stride* de 2, que extrae características de alto nivel tanto de las imágenes generadas como de las imágenes reales, reescaladas a un tamaño de  $32 \times 32$

píxeles.

La tarea del discriminador es distinguir entre imágenes reales y generadas a nivel de parche, evaluando las diferencias locales que exponen discrepancias espaciales. Este enfoque ayuda a mejorar la calidad de la imagen generada, ya que obliga al generador a reproducir detalles estructurales precisos en las áreas afectadas por los artefactos [27, 95, 101].

Para guiar el proceso de entrenamiento, la CCAR net utiliza una función de pérdida compuesta que incluye:

- **Pérdida de contenido ( $L_{content}$ ):** Cuantifica las diferencias a nivel de píxeles entre la imagen generada ( $G(x)_{i,j}$ ) y la imagen real ( $y_{i,j}$ ), asignando un mayor peso a los píxeles que se desvían significativamente de la referencia. Matemáticamente, se define como:

$$L_{content}(G) = \frac{1}{w \times h} \sum_{0 \leq i \leq w, 0 \leq j \leq h} ((y_{i,j} - G(x)_{i,j} + 1)^2 \times |y_{i,j} - G(x)_{i,j}|), \quad (3.6)$$

donde  $w$  y  $h$  denotan la resolución del mapa de salida,  $x$  representa la imagen de entrada,  $y$  es la imagen del *ground truth*, y  $G$  es la función del generador. El término adicional  $|y_{i,j} - G(x)_{i,j}|$  asigna un peso mayor a las diferencias de píxeles más grandes.

- **Pérdida de características ( $L_{feature}$ ):** Incorporada para abordar el desenfoque que puede surgir del uso de la pérdida de contenido. Incluye componentes de pérdidas perceptuales ( $L_{perceptual}$ ) y de estilo ( $L_{style}$ ), que se enfocan en las diferencias en las características a nivel de parche entre las imágenes sintéticas y las reales.
  - **Pérdida perceptual ( $L_{perceptual}$ ):** Se calcula como la diferencia  $L_1$  entre los mapas de características de las imágenes generada y real extraídos por la capa  $l$ -ésima de una red VGG19 pre-entrenada ( $\phi_l$ ).

$$L_{perceptual}(G) = \frac{1}{C_l H_l W_l} \sum_{i,j,k} |\phi_l(y)_{i,j,k} - \phi_l(G(x))_{i,j,k}|, \quad (3.7)$$

donde  $C_l$ ,  $H_l$  y  $W_l$  representan el número de canales, la altura y el ancho del mapa de características de la  $l$ -ésima capa de VGG19, respectivamente.

- **Pérdida de estilo ( $L_{style}$ ):** Se define como la diferencia  $L_1$  entre las matrices de Gram de los mapas de características de las imágenes generada y real extraídos por las capas  $j$ -ésimas de una red VGG19 pre-entrenada.

La matriz de Gram de un mapa de características  $F^l \in \mathbb{R}^{C_l \times H_l \times W_l}$  se calcula como  $G_{ij}^l = \sum_{k=1}^{H_l W_l} F_{ik}^l F_{jk}^l$ .

$$L_{style}(G) = \frac{1}{C_j^2 H_j^2 W_j^2} \sum_l |Gram(\phi_j(y))_l - Gram(\phi_j(G(x)))_l|_1, \quad (3.8)$$

donde  $\phi_j$  representa las características extraídas por la  $j$ -ésima capa de VGG19. En el contexto de imágenes de OCT, la pérdida de estilo se refiere a la diferencia en el patrón de *speckle* entre las imágenes sintéticas y reales.

La pérdida de características total se define como una combinación ponderada de las pérdidas perceptual y de estilo:

$$L_{feature}(G) = \lambda_{perceptual} + L_{perceptual}(G) + \lambda_{style} L_{style}(G), \quad (3.9)$$

donde  $\lambda_{perceptual}$  y  $\lambda_{style}$  son pesos para equilibrar ambas pérdidas.

- **Pérdida adversaria** ( $L_{adversarial}$ ): Similar a la utilizada en Pix2Pix [27], entrena al generador ( $G$ ) para engañar al discriminador ( $D$ ), mejorando la calidad global de las imágenes generadas. Se define como:

$$L_{adversarial}(G, D) = \mathbb{E}_{y \sim p_{data}(y)} [\log D(y)] + \mathbb{E}_{x \sim p_{data}(x)} [\log(1 - D(G(x)))]. \quad (3.10)$$

El objetivo del generador es minimizar esta pérdida, mientras que el objetivo del discriminador es maximizarla.

La función de pérdida total para el entrenamiento de la CCAR-Net se resume como una combinación ponderada de estas tres pérdidas:

$$L_{total}(G) = lL_{content}(G) + mL_{feature}(G) + nL_{adversarial}(G, D), \quad (3.11)$$

donde  $l$ ,  $m$  y  $n$  son hiperparámetros que controlan la importancia relativa de cada término de pérdida.

El uso de esta función de pérdida compuesta permite que la CCAR net aprenda a eliminar los artefactos espejo, preservando al mismo tiempo las características estructurales esenciales en las imágenes OCT [101]. Aunque este enfoque ha demostrado ser efectivo, se encuentra limitado al uso de imágenes de intensidad. **En contraste, la presente propuesta busca trabajar directamente en el campo complejo, lo que proporciona una ventaja significativa al abordar la naturaleza intrínseca de los datos en OCT.** Para lograr esto, se ha modificado la

arquitectura de la CCAR para que reciba y procese dos canales: uno con la amplitud normalizada y otro con la fase. De esta forma, la CCAR aprende a generar los componentes reales e imaginarios que conforman el fasor que representa la señal compleja, permitiéndole modelar la relación entre la amplitud y la fase de las señales OCT y trabajar directamente con la información intrínseca de los datos.

# IMPLEMENTACIÓN Y EVALUACIÓN DEL APRENDIZAJE PROFUNDO EN OCT

En este capítulo se detallan los procedimientos llevados a cabo para implementar y evaluar modelos de aprendizaje profundo aplicados a OCT. La metodología empleada para la selección de hiperparámetros, que incluye la descripción de los parámetros clave y el proceso de optimización, se detalla en la subsección 4.1. Se busca optimizar el rendimiento de los modelos en la solución de dos problemas específicos presentes en el dominio de Fourier: la reconstrucción de tomogramas submuestreados lateralmente y la mitigación de artefactos espejo. Además, se describen las métricas de evaluación utilizadas para validar los resultados obtenidos y se presentan los análisis detallados de los modelos entrenados para cada problema.

### 4.1. Metodología para la Selección de Hiperparámetros

La selección de hiperparámetros es un paso crítico en el desarrollo de modelos de aprendizaje profundo, ya que impacta directamente en el rendimiento y la capacidad de generalización de los modelos. En el contexto de OCT, donde la calidad de la reconstrucción y la preservación de los detalles estructurales son esenciales, es importante realizar una selección cuidadosa de los hiperparámetros que influyen en el proceso de entrenamiento.

#### Hiperparámetros Clave

En los experimentos realizados, se variaron los siguientes hiperparámetros para optimizar el rendimiento de los modelos:

- **Profundidad de la red:** La profundidad de la red se modificó variando el

número de capas convolucionales en el generador, probando configuraciones con 4, 5 y 6 capas. A medida que se incrementa la profundidad, el número de filtros se mantiene fijo de forma ascendente, comenzando con 64 filtros y aumentando hasta 512. Se eligió usar múltiplos de 32 para la cantidad de filtros debido a consideraciones de eficiencia de memoria, paralelización en GPUs y convenciones en el diseño de redes convolucionales. Esta estructura permite que la red profundice su capacidad de aprendizaje sin tener que explorar combinaciones complejas de diferentes números de filtros por capa.

- **Optimizador:** Se probaron diferentes optimizadores, incluyendo Adam, RMS-prop y SGD, para evaluar su efecto en la convergencia del modelo y en la calidad de las imágenes generadas. La elección del optimizador influye en cómo el modelo minimiza la función de pérdida durante el entrenamiento.
- **Porcentaje de *dropout*:** El *dropout* es una técnica de regularización utilizada para prevenir el sobreajuste. Con el objetivo de evaluar su impacto en la generalización del modelo, se varió el porcentaje de *dropout* en las capas del generador entre tres valores típicos encontrados en la literatura: 20 %, 30 % y 50 % [105–107].
- **Coefficientes de las funciones de pérdida personalizadas (solo para CCAR net):** Para la CCAR net, se ajustaron los coeficientes que ponderan las diferentes componentes de la función de pérdida compuesta (pérdida de contenido, pérdida de características y pérdida adversaria) para encontrar la mejor combinación que preserve las estructuras anatómicas mientras elimina los artefactos de conjugado complejo.

### Estrategia de Selección de Hiperparámetros y Criterios de Evaluación

La estrategia de selección de hiperparámetros consistió en la evaluación iterativa de diferentes configuraciones, explorando un rango de valores para cada hiperparámetro. Los rangos de valores se definieron basados en valores típicos encontrados en la literatura [108–110], acotando el espacio de búsqueda debido a limitaciones de tiempo de entrenamiento. En lugar de una búsqueda exhaustiva en cuadrícula, se optimizó cada hiperparámetro individualmente, seleccionando el valor que mejor desempeño presentó según la métrica de Error Cuadrático Medio (MSE). Esta decisión se fundamentó en **limitaciones computacionales inherentes al entorno de desarrollo, que incluyen un tiempo de ejecución restringido para la experimentación y una capacidad de memoria limitada en la infraestructura disponible**. El entrenamiento de los modelos se llevó a cabo utilizando principalmente el computador de desarrollo e investigación APOLO, el cual cuenta con 125 GB de RAM HBM2, 3 tarjetas de video NVIDIA V100 con 32 GB de VRAM, y un procesador Intel Xeon

E5-2683v4 a 2.10 GHz. Es importante mencionar que los recursos de APOLO aunque son muy buenos para realizar la búsqueda en cuadrícula de los hiperparámetros, son de uso compartido para la universidad y centros aliados, y opera bajo un sistema de procesamiento por lotes o *SLURM* con colas de trabajo y con tiempos máximos de ejecución restringidos lo cual lo hace de acceso limitado. Para contrarrestar estas limitantes, Adicionalmente, se utilizó un computador personal de carácter comercial con un procesador Ryzen 7 5800H, 16 GB de RAM DDR4 y una tarjeta de video NVIDIA RTX 3050 de 4 GB GDDR6 para algunas etapas de experimentación. Aunque esta estrategia permite una optimización eficiente dentro de estas restricciones, **existe el riesgo de no identificar la combinación de hiperparámetros que maximizaría el rendimiento del modelo. Esto podría traducirse en una menor capacidad de generalización a nuevos datos o en una precisión ligeramente inferior a la que se podría alcanzar con una exploración más completa del espacio de búsqueda.** Para el **learning rate**, se implementó una estrategia de *decay rate* exponencial con un factor de decaimiento de 0,96. El valor inicial del *learning rate* se fijó en 0,0002 y disminuye gradualmente a lo largo de las épocas de entrenamiento. Este enfoque permite que, en las etapas iniciales del entrenamiento, el modelo realice ajustes significativos en sus pesos para escapar rápidamente de posibles regiones subóptimas en el espacio de búsqueda. A medida que el entrenamiento avanza y el modelo se acerca a un mínimo (convergencia), la tasa de aprendizaje disminuye progresivamente, permitiendo ajustes más pequeños y precisos para afinar los pesos y evitar oscilaciones alrededor del mínimo óptimo.

### Optimización Específica

En el caso de la CCAR net, además de los hiperparámetros mencionados, se realizó un ajuste específico de los coeficientes que ponderan las distintas componentes de la función de pérdida compuesta. Estos coeficientes desempeñan un papel crucial en el equilibrio entre la eliminación de artefactos y la preservación de los detalles estructurales. Los coeficientes de la pérdida de contenido, la pérdida de características y la pérdida adversaria se ajustaron para optimizar la reconstrucción.

Por otra parte, para las arquitecturas de tipo Pix2Pix, se experimentó con diferentes cantidades de filtros y porcentajes de *dropout* con el objetivo de mejorar la capacidad del modelo para generalizar y aprender de los datos de entrenamiento, manteniendo siempre el enfoque en la evaluación mediante el MSE.

## 4.2. Preparación del Conjunto de Datos

La preparación de los datos es un paso crucial para garantizar que los modelos sean capaces de reconstruir B-scans con alta fidelidad en distintos tomogramas.

Un tomograma está compuesto por diversas distribuciones de *speckle*, intensidades de *multiple scattering* y distribuciones de intensidad en el tejido [111, 112]. Por lo tanto, un conjunto de entrenamiento robusto debe incluir muestras significativas de estas variaciones, lo que permite al modelo generalizar adecuadamente diferentes estructuras espaciales. Sin embargo, este enfoque presenta un reto importante, la posible dependencia de los datos, que podría llevar al sobreajuste de la red [113, 114]. Además, las configuraciones instrumentales intrínsecas de los sistemas OCT, como la dispersión en el medio biológico y las propiedades de coherencia, pueden sesgar el modelo al alterar la imagen adquirida. Para mitigar estos efectos, se entrena la red utilizando una mezcla de tomogramas experimentales obtenidos con diferentes sistemas OCT y tomogramas sintéticos generados mediante el modelo directo.

Para generar los datos simulados de entrenamiento, se producen tomogramas sintéticos que presentan patrones aleatorios de esparsos tridimensionales que emulan estructuras biológicas, como capas y estructuras vasculares, utilizando el modelo directo descrito en la sección 2.4, el cual genera un tomograma OCT basado en la morfología de la muestra [115]. Se emplean tomogramas complejos con canales reales e imaginarios con el fin de realizar un análisis de estabilidad de fase, así como representar la amplitud y la intensidad de los tejidos. El análisis de fase, el cual se explica a detalle en la sección 4.3 permite comprender la continuidad y correlación estructural cuantificando los cambios de fase y los gradientes locales entre A-lines adyacentes o repetidas en un mismo punto [9].

Los datos sintéticos generados incluyen diversas estructuras y capas para simular dispersores puntuales en los tejidos. Sin embargo, el modelo directo solo describe las distribuciones de *speckle* orientadas al tejido, sin incluir el ruido fotónico ni el *multiple scattering*. Para incorporar estos efectos, los datos experimentales incluyen tanto ruido no relacionado con el tejido como señales de dispersión del tejido, así como otras características instrumentales como la dispersión. Utilizar tomogramas experimentales de diferentes instrumentos permite que el modelo aprenda estas características, evitando el sobreajuste a un único conjunto de datos. En la tabla 4.1 se presentan los parámetros de simulación empleados para generar los tomogramas sintéticos, estos parámetros son los que se encuentra típicamente en un sistema OCT comercial.

Tabla 4.1: Parámetros de simulación utilizados en el estudio

Parámetro	Valor
Longitud de onda central	1310 nm
FWHM (Ancho de banda a mitad de altura)	120 nm
Apertura numérica	> 0,1
Número de puntos de dispersión (scatters)	< 170000
Resolución axial	> 7,5 $\mu$ m

## 4.2.1. Preparación de tomogramas Submuestreados Lateralmente

La Figura 4.1 presenta algunos de los tomogramas utilizados durante el entrenamiento de la red, etiquetados de (a) a (f). En (a), se observa un tomograma experimental del nervio óptico en vista *en-face*. De (b) a (f), se muestran B-scans de tomogramas sintéticos generados mediante el modelo directo descrito en la Sección 2.4.

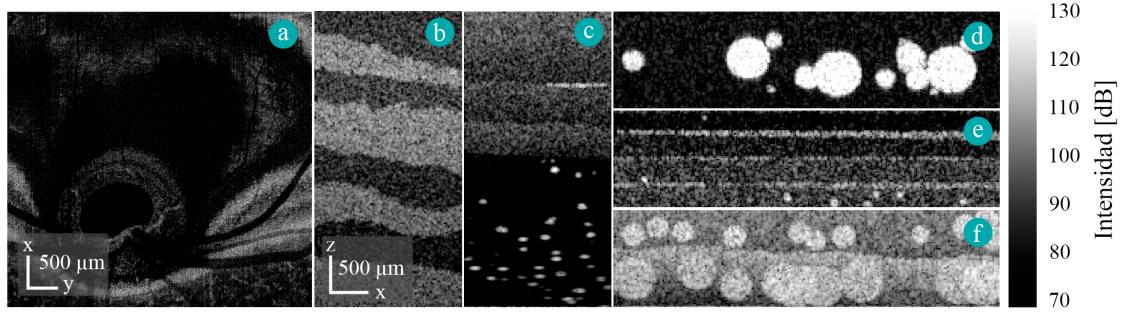


Figura 4.1: Conjunto de datos utilizados durante el entrenamiento. En (a) se muestra un plano *en-face* de un tomograma experimental. De (b) a (f) se presentan los tomogramas sintéticos donde las B-scans (c) a (f) conservan la misma resolución lateral y axial mostrada en (b).

El volumen sintético utilizado para el entrenamiento está compuesto por elementos con dimensiones  $(Z, X, Y, Ch) = (256, 256, 256, 2)$ , mientras que el tomograma experimental empleado en esta misma fase presenta dimensiones de  $(100, 928, 960, 2)$ . Los tomogramas destinados a la validación tienen dimensiones de  $(586, 896, 960, 2)$  y  $(800, 1024, 992, 2)$ , donde  $Ch$  representa los canales correspondientes a las partes real e imaginaria (Re/Im). Los datos experimentales utilizados en la validación fueron adquiridos por el Wellmann Center for Photomedicine mediante un sistema Spectralis OCT de Heidelberg Engineering, equipado con una interfaz oftálmica y un barrido de longitud de onda personalizado. Este sistema, basado en OCT de dominio espectral, emplea un diodo superluminescente con una longitud de onda central de 870 nm y un ancho de banda de 50 nm, operando a una tasa de A-lines de 85 kHz. Estos volúmenes complejos se utilizan para representar de manera adecuada tanto la amplitud como la fase de los tejidos en la entrada del generador, y se someten a un proceso de normalización para garantizar la consistencia en la distribución estadística de los datos.

El proceso de normalización aplicado a los tomogramas se realiza en dos etapas principales: el tratamiento de la amplitud y la conservación de la fase. Inicialmente, se separa la información de amplitud ( $A$ ) y fase de los tomogramas. La fase ( $\Phi$ ) se obtiene a partir de la relación entre la parte imaginaria ( $iB$ ) y la parte real ( $A$ ) de los datos complejos, tal como se expresa en la Ecuación 4.1:

$$(\Phi) = \text{atan2} \left( \frac{B}{A} \right). \quad (4.1)$$

Posteriormente, se aplica una normalización a la amplitud ( $A$ ) en escala logarítmica utilizando el método de mínimos y máximos. Este proceso transforma la amplitud original a un rango  $[0, 1]$  mediante la siguiente relación:

$$A_{\text{norm}} = \frac{A_{\log} - \text{mín}(A_{\log})}{\text{máx}(A_{\log}) - \text{mín}(A_{\log})}. \quad (4.2)$$

Es importante destacar que, durante este proceso de normalización de la amplitud, la fase ( $\Phi$ ) calculada en el primer paso se conserva sin modificaciones. Esto se realiza con el objetivo de mantener la representación de los datos en su forma fasorial original. Finalmente, se recombina la amplitud normalizada  $A_{\text{norm}}$  con la fase intacta  $\Phi$  para formar un nuevo tomograma en su representación fasorial  $R_{\text{norm}}$ , tal como se describe en la siguiente ecuación:

$$R_{\text{norm}} = |A_{\text{norm}}| e^{(i\Phi)}. \quad (4.3)$$

La obtención de esta nueva forma fasorial  $R_{\text{norm}}$  asegura que el conjunto de datos normalizado mantenga una distribución estadística coherente, un aspecto relevante señalado por Sun et al. [116]. En su trabajo, Sun et al. concluyen que la mayoría de los métodos de normalización como Batch Normalization, Layer Normalization, etc. operan bajo un principio unificado: **proyectan las activaciones internas de la red o sus pesos sobre un espacio acotado, conceptualmente similar a una esfera**. La importancia de esto radica en que, al eliminar la ambigüedad de escala, el proceso de optimización se vuelve más estable y predecible. En la práctica, esto ayuda a mitigar problemas comunes como el desvanecimiento del gradiente, permitiendo una convergencia más rápida y robusta del modelo. Por lo tanto, al normalizar nuestros datos fasoriales, estamos aplicando este principio para facilitar que la red aprenda las características relevantes de manera efectiva.

Tras la normalización de la amplitud de este fasor, tanto para los datos sintéticos como para los experimentales, se procede a la creación de dos conjuntos de datos diferenciados. El primer conjunto, denominado conjunto de condicionamiento el cual es usado como entrada de la red. Incluye tanto los canales real como imaginario. Este volumen es submuestreado lateralmente en dirección del eje  $y$ , asignando un valor de cero a determinados B-scans ubicados en índices pares tal como se ilustra en la Figura 4.2. Esta asignación de valores crea una disposición intercalada de estructuras y B-scans vacíos, donde los valores iniciales funcionan como semillas para el proceso de reconstrucción, afectando únicamente el estado inicial del optimizador durante

las primeras épocas de entrenamiento. Este procedimiento se aplica a ambos tipos de volúmenes —sintéticos y experimentales— que, originalmente, no habían sido submuestreados lateralmente con el fin de emular las discontinuidades observadas en los B-scans experimentales. El conjunto de datos generado a partir de este proceso se etiqueta como falso. Se define como falso el tomograma submuestreado digitalmente, en virtud de que contiene información que no es real desde el experimento, y se alterará en el proceso generativo.

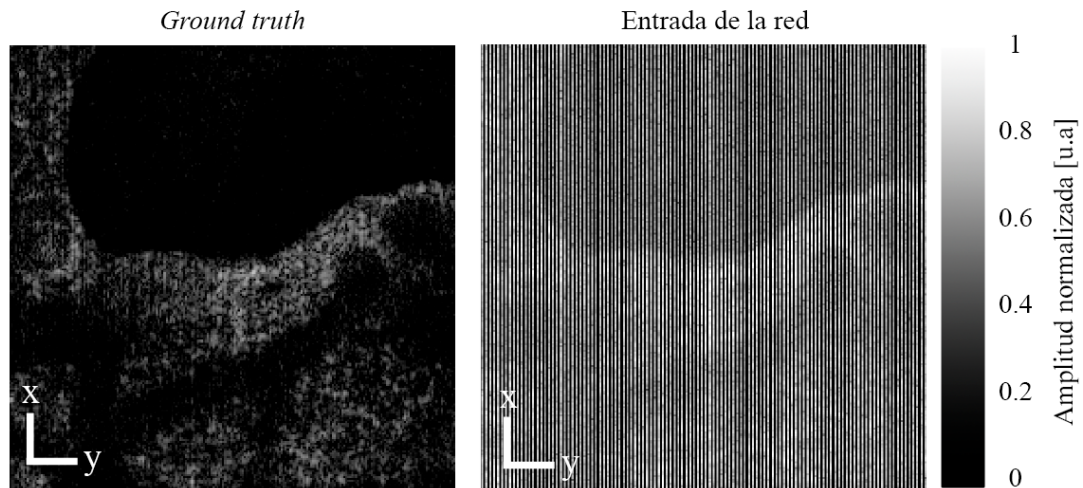


Figura 4.2: Ilustración del proceso de submuestreo lateral aplicado al conjunto de datos. La imagen de la izquierda representa un plano *en-face* del volumen original (“Ground truth”), mientras que la imagen de la derecha muestra el resultado del submuestreo lateral (“Entrada de la red”), donde los B-scans en índices pares han sido reemplazados por valores de cero, creando un patrón intercalado.

El segundo conjunto de datos consiste en los B-scans originales los cuales no están submuestreados digitalmente, se utilizan como referencia y se etiquetan como *ground truth*. Durante el entrenamiento, el modelo debe inferir las estructuras espaciales de los B-scans ausentes, minimizando las diferencias con las estructuras originales. La reconstrucción de los B-scans sigue un proceso que avanza por niveles de profundidad, generando de manera progresiva los B-scans correspondientes a cada nivel hasta alcanzar el plano *en-face* más profundo. Esto asegura la reconstrucción completa del volumen. Utilizando la técnica de *sliding window* sobre los datos sintéticos y experimentales, se obtienen un conjunto total de 10,496 planos *en-face* submuestreados los cuales se usarán para entrenar la red.

#### 4.2.2. Preparación de tomogramas con artefactos espejo

Los tomogramas experimentales utilizados como referencia para el entrenamiento de la red fueron adquiridos mediante el sistema de OCT descrito en la Sección 4.2.1.

Durante la adquisición, la muestra se posicionó de tal manera que su espectro en el espacio de frecuencias ocupara una región desplazada del origen. Esta configuración espacial en el dominio de Fourier evitó la superposición de las imágenes reales y su artefactos espejo en el dominio de profundidad. Al realizar un corte en el DC, se pudo extraer una única copia del tomograma sin la interferencia del artefacto gemelo. Es importante notar que este método implica una reducción en la profundidad de escaneo, ya que la mitad del rango de frecuencias se descartó en el proceso. Los artefactos espejo necesarios para el entrenamiento de la red en los datos experimentales se generaron posteriormente mediante un procedimiento específico, el cual se detallará más adelante.

Posteriormente, la Figura 4.3 muestra ejemplos de tomogramas utilizados para entrenar la red en la tarea de eliminación de artefactos espejo. Estos tomogramas, tanto sintéticos como experimentales (con artefactos generados), ilustran los datos de entrada para el proceso de aprendizaje.

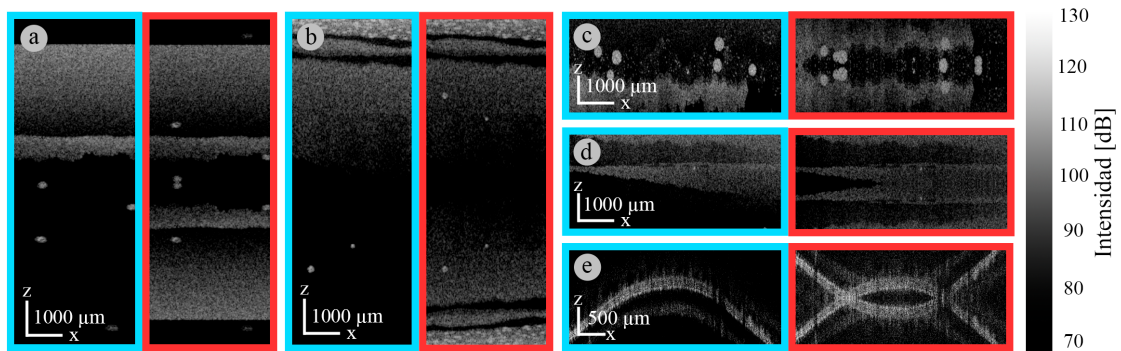


Figura 4.3: Muestra de tomogramas utilizados para entrenar los modelos que buscan mitigar los artefactos espejo. los grupos de imágenes del recuadro azul muestrean las estructuras usadas como objetivo. Los grupos rojos muestran las mismas estructuras con artefactos espejo y con distintos grados de solapamiento. De (a) a (d) se presentan tomogramas sintéticos, y en (e) se muestra un tomograma experimental de la fóvea, adquirido en el Wellman Center for Photomedicine. Este es usado para la validación de los modelos.

Los volúmenes sintéticos empleados para el entrenamiento tienen dimensiones  $(Z, X, Y, Ch) = (512, 512, 20, 2)$ , siendo  $Ch$  los canales correspondientes a las partes real e imaginaria (Re/Im). En total, se dispone de 198 tomogramas sintéticos. Por otro lado, los volúmenes experimentales comprenden tres tomogramas humanos, correspondientes a la fóvea, y el nervio óptico. Cada uno con dimensiones  $(1024, 1024, 512, 2)$ , donde la disposición de los ejes sigue la misma estructura  $(Z, X, Y, Ch)$ .

El proceso de generación de los artefactos espejo comienza con la aplicación de la transformada inversa de Fourier en el eje  $Z$ , lo que permite recuperar las franjas de interferencia en el dominio de frecuencia espacial ( $k$ ). A partir de estas franjas,

representadas por la señal compleja  $S(k)$ , se compone una nueva señal  $S'(k)$  de la siguiente forma:

$$S'(k) = |S(k)| \cos(\phi), \quad (4.4)$$

donde  $|S(k)|$  representa la magnitud de la señal compleja en el dominio  $k$ , y  $\phi = \text{atan2}(\text{Im}(S(k)), \text{Re}(S(k)))$  es su fase. Esta operación asegura que las componentes reales de las franjas de interferencia se mantengan. Al aplicar nuevamente la Transformada de Fourier en el eje  $Z$  a  $S'(k)$ , los artefactos espejo aparecen simétricamente con respecto al componente de frecuencia cero (DC), generando la imagen gemela característica de las estructuras.

Una vez obtenidos los tomogramas con y sin artefactos espejo, se procede a la normalización siguiendo el mismo procedimiento descrito en la subsección 4.2.1, donde se transforma la amplitud a una escala logarítmica normalizada y se conserva la fase  $\Phi$  para mantener la representación fasorial de la señal.

Para los tomogramas experimentales, cuyas dimensiones son mayores a las de los tomogramas sintéticos, se seleccionan regiones de interés (*regions of interest*, *ROIs*) con dimensiones (512, 512, 2) que cubren el área central de los volúmenes experimentales después de la transformación de las franjas de interferencia. La selección de estas regiones se realiza tomando como referencia la frecuencia central en el eje  $Z$ , de la siguiente manera:

$$Z/2 - 256 \leq Z < Z/2 + 256, \quad (4.5)$$

mientras que para el eje  $X$  se toman dos regiones de interés: ( $0 \leq X \leq 512$ ) y ( $512 \leq X \leq 1024$ ), obteniendo así dos volúmenes de entrada a partir de cada tomograma experimental.

Este proceso asegura que las regiones seleccionadas sean simétricas alrededor de la frecuencia central en el espacio de Fourier, lo que es crucial para la correcta eliminación de los artefactos de espejo durante el entrenamiento de la red.

### 4.3. Métricas de Evaluación

Dada la naturaleza de la información contenida en las imágenes OCT, resulta esencial no solo recuperar una apariencia visualmente similar a las imágenes originales, sino también preservar las estructuras y detalles cruciales para el diagnóstico y análisis clínico. Para este propósito, se emplea un conjunto de métricas de evaluación que permiten cuantificar la eficacia de los métodos utilizados en términos de similitud, calidad y preservación de características.

### Índice de Similitud Estructural

El Índice de Similitud Estructural (SSIM) es una métrica utilizada para evaluar la similitud perceptual entre dos imágenes. A diferencia de las métricas basadas en la intensidad de los píxeles, SSIM se enfoca en medir cambios en las estructuras de la imagen, teniendo en cuenta aspectos como el contraste, la luminancia y las características locales [117]. En OCT, el uso de SSIM es especialmente relevante, ya que permite evaluar si las estructuras anatómicas y los detalles finos del tejido han sido preservados tras la reconstrucción.

La expresión matemática del SSIM se basa en la comparación de ventanas locales de las imágenes, y se define como,

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (4.6)$$

donde  $x$  e  $y$  representan las ventanas locales de las imágenes a comparar,  $\mu_x$  y  $\mu_y$  son las medias locales,  $\sigma_x^2$  y  $\sigma_y^2$  las varianzas locales,  $\sigma_{xy}$  la covarianza local, y  $C_1$  y  $C_2$  son constantes para estabilizar la división. SSIM toma valores entre 0 y 1, siendo 1 la similitud estructural perfecta.

### Relación Señal-Ruido Pico (PSNR)

La Relación Señal-Ruido Pico (PSNR) mide la calidad de la reconstrucción comparando la imagen generada con la imagen original. Se define en términos de la proporción entre la potencia máxima de una señal y la potencia del ruido, y se expresa en decibelios (dB), donde valor de PSNR más alto indica una mayor calidad de reconstrucción.

La fórmula para calcular la PSNR es:

$$\text{PSNR} = 10 \cdot \log_{10} \left( \frac{\text{MAX}^2}{\text{MSE}} \right), \quad (4.7)$$

donde MAX es el valor máximo posible de la intensidad de la imagen y MSE es el error cuadrático medio entre la imagen original y la imagen generada.

Es importante destacar que la PSNR es sensible a valores atípicos. Si un solo píxel tiene un valor de intensidad extremadamente alto, esto puede disparar la métrica, incluso si la reconstrucción general es deficiente. Además, la PSNR se centra en la diferencia de intensidad píxel a píxel y no considera la estructura de la imagen.

### Diferencia de histogramas mediante similitud de coseno

La diferencia de histogramas es una métrica empleada para evaluar el grado de similitud entre las distribuciones de intensidad de dos imágenes. Para este propósito, se utiliza la similitud de coseno, una medida que cuantifica el ángulo entre dos vectores en un espacio multidimensional, en este caso, los histogramas normalizados de las imágenes en comparación. Debido a su independencia de las escalas de la información comparada, la similitud de coseno es especialmente útil para comparar histogramas que pueden tener diferentes rangos de valores. Esta métrica toma valores entre -1 y 1, donde 1 indica que los histogramas son idénticos, 0 implica que no existe correlación entre ellos, y -1 representa una anti-correlación completa, lo que significa que las distribuciones de intensidad son opuestas. Por ejemplo, si un histograma tiene una alta frecuencia en ciertos bins, el otro histograma tendrá una baja frecuencia (o ninguna) en esos mismos bins, y viceversa.

La expresión matemática de la similitud de coseno entre dos histogramas  $h_1$  y  $h_2$  se define como,

$$\text{Similitud de Coseno} = \frac{h_1 \cdot h_2}{\|h_1\| \|h_2\|}, \quad (4.8)$$

donde  $h_1$  y  $h_2$  son los vectores de los histogramas,  $h_1 \cdot h_2$  representa el producto punto entre ambos vectores, y  $\|h_1\|$  y  $\|h_2\|$  son las normas de los respectivos vectores.

En el contexto de OCT, esta métrica permite evaluar cuán similares son las distribuciones de intensidad entre un tomograma original y uno procesado o reconstruido. Dado que en las imágenes OCT las intensidades están relacionadas con la reflectividad de los tejidos, una similitud alta entre histogramas implica que las estructuras tisulares y sus características se han mantenido tras el procesamiento, lo que sugiere una reconstrucción exitosa o la correcta mitigación de artefactos.

### Diferencia de Fase

En el enfoque orientado a la reconstrucción del campo complejo en OCT, la diferencia de fase se establece como una métrica fundamental. A diferencia de las métricas tradicionales que se centran en la intensidad o la similitud estructural, la diferencia de fase evalúa directamente la precisión con la que se reconstruye la fase del campo complejo. Esto resulta relevante para aplicaciones avanzadas de OCT, donde la fase contiene información clave sobre las propiedades de los tejidos y su dinámica [118].

Para calcular la diferencia de fase, se evalúa la variación en el ángulo entre píxeles adyacentes, lo cual es equivalente a una correlación de retardo unitario, en las direcciones  $x$  e  $y$  de la imagen reconstruida. Formalmente, esta correlación de

fase en las dos direcciones se define de la siguiente manera:

Sea  $\phi_{x,y}$  la fase del píxel en la posición  $(x, y)$  de la imagen reconstruida. La diferencia de fase en la dirección  $x$  se expresa como,

$$\Delta\phi_x(x, y) = \arg [\phi_{x+1,y} \cdot \phi_{x,y}^*], \quad (4.9)$$

donde  $\phi_{x,y}^*$  representa el complejo conjugado de  $\phi_{x,y}$ , y  $\arg$  denota la operación que calcula el ángulo de un número complejo. De manera similar, la diferencia de fase en la dirección  $y$  se calcula como,

$$\Delta\phi_y(x, y) = \arg [\phi_{x,y+1} \cdot \phi_{x,y}^*]. \quad (4.10)$$

Este proceso permite detectar inconsistencias locales en la estimación de la fase. Posteriormente, se calculan la desviación estándar y el promedio de estas diferencias de fase para obtener una medida cuantitativa de la estabilidad y consistencia de la fase para un B-scan específico.

Un valor bajo de desviación estándar indica una mayor correspondencia en la estimación de la fase, lo que sugiere una reconstrucción más fiel del campo complejo. Sin embargo, la definición de "bajo" es relativa y depende del contexto específico de la aplicación y la calidad esperada de la reconstrucción. En general, se busca obtener valores de desviación estándar lo más cercanos posible a cero, lo que indicaría una fase perfectamente consistente.

Por lo tanto, esta métrica proporciona una evaluación directa de la calidad de la reconstrucción de fase, superando las limitaciones de las métricas basadas únicamente en la intensidad o la estructura.

## 4.4. Resultados y evaluación de los modelos

En esta sección se presentan los resultados obtenidos al aplicar los modelos de aprendizaje profundo a los dos problemas principales: la reconstrucción de tomogramas submuestreados lateralmente y la eliminación de artefactos espejo. A continuación, se detallan los resultados específicos para cada uno de estos desafíos, evaluando la efectividad de los modelos en ambos contextos.

### 4.4.1. Reconstrucción de tomogramas submuestreados mediante aprendizaje profundo

Para el ajuste y validación de los modelos se realizaron cuatro series de pruebas, cada una diseñada para optimizar diferentes aspectos del entrenamiento. Los

entrenamientos se ejecutaron durante 100 épocas usando el modelo Pix2Pix. Para las validaciones de dichos entrenamientos se emplearon 3 B-scan de uno de los tomogramas de validación, los cuales no han sido vistos previamente por la red.

La primera serie de pruebas consistió en la variación del conjunto de datos de la sección 4.2.1, manteniendo todos los hiperparámetros constantes mientras se cambiaba la fuente de datos entre tres opciones: tomogramas simulados, experimentales y una combinación de ambos. Se procuró mantener la misma cantidad de *slices* para cada configuración, evitando introducir sesgos debidos a diferencias en el tamaño del conjunto de datos. En la Figura 4.4, específicamente en el gráfico superior izquierdo, se muestran los resultados de esta prueba, donde se observa que la combinación de datos simulados y experimentales resultó en un mejor desempeño. Esto sugiere que la red aprovecha la diversidad en los datos para mejorar la precisión del modelo, logrando un menor valor de MSE respecto a las opciones que utilizan únicamente datos simulados o experimentales.

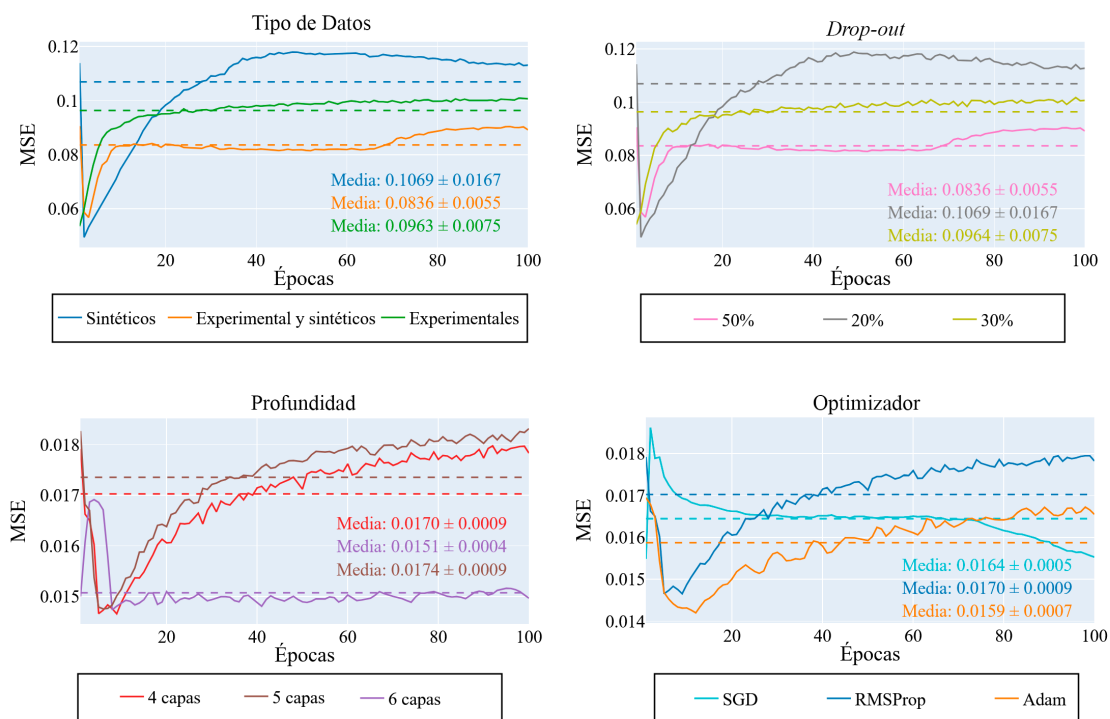


Figura 4.4: Gráficos del MSE por época para distintas configuraciones de hiperparámetros. Las líneas continuas representan los resultados de cada configuración, mientras que las líneas punteadas indican el valor promedio del MSE para cada caso.

La segunda serie de pruebas se enfocó en la regularización a través del uso de *dropout*. Manteniendo constante la configuración óptima del conjunto de datos encontrada en la prueba anterior, se variaron los valores de *dropout* entre 20 %, 30 % y 50 %. Los resultados, representados en el gráfico superior derecho de la Figura 4.4, indican que el mejor valor de regularización fue 50 %. Esto sugiere que un mayor

grado de regularización ayuda a la red a evitar el sobreajuste, permitiendo una mejor generalización.

La tercera serie de pruebas se centró en la profundidad de la red. Inicialmente, los modelos se entrenaron con solo dos capas de convolución para reducir el tiempo de cómputo. Sin embargo, con el fin de capturar mayores detalles y mejorar el rendimiento, se realizaron nuevas pruebas incrementando la profundidad a 4, 5 y 6 capas. En el gráfico inferior izquierdo de la Figura 4.4 se observa que la red con 6 capas obtuvo los valores más bajos de MSE, indicando que una mayor profundidad contribuye a una mejor representación y aprendizaje de las características complejas del tomograma.

Por último, se llevó a cabo una serie de pruebas para determinar el optimizador más adecuado. Para estas pruebas, se utilizaron las mejores configuraciones encontradas en las pruebas previas: uso combinado de datos, regularización al 50 % y una red de 6 capas. Cabe destacar que, en las pruebas previas, se utilizó Adam como optimizador fijo. Se compararon varios optimizadores, y los resultados, visibles en el gráfico inferior derecho de la Figura 4.4, muestran que Adam fue el mejor optimizador. Este desempeño indica que Adam maneja de manera efectiva la convergencia del modelo, aprovechando la configuración previa óptima para lograr menores valores de MSE.

Si bien se reconoce que la elección de Adam como optimizador fijo en las pruebas previas pudo haber influido en la selección de los demás hiperparámetros, esta decisión se tomó debido a limitaciones de tiempo y complejidad computacional. La selección inicial de Adam se justifica por su capacidad para ofrecer una convergencia rápida, eficiencia computacional y bajos requerimientos de memoria [119, 120], características ventajosas en las etapas exploratorias de la investigación. Además, Adam ha demostrado un rendimiento superior en la clasificación de imágenes OCT en comparación con otros optimizadores como SGD y RMSProp logrando alta precisión y baja pérdida [121–123]. La estrategia de optimización acumulativa, que incorpora progresivamente las mejoras encontradas en cada etapa, permitió asegurar que cada experimento se beneficiara de las configuraciones optimizadas en pruebas previas, logrando así un modelo final eficiente.

Los hiperparámetros del modelo se resumen en la Tabla 4.2. La arquitectura U-Net y PatchGAN se complementan mediante la configuración de los pesos iniciales, una tasa de *dropout* del 50 % para prevenir el sobreajuste, y el optimizador Adam para facilitar la convergencia durante el entrenamiento.

Arquitectura	Peso $\sigma$	Dropout	Optimizador	Funciones de Pérdida
U-net — PatchGAN	0.02	0.5	Adam	MSE, MAE ( $G$ ) — Binary Cross-Entropy ( $D$ )

Tabla 4.2: Hiperparámetros del Modelo Pix2Pix

### Reconstrucción de la Correlación Espacial del Tejido con Predicción del Modelo

Las capacidades del modelo Pix2Pix para reproducir la continuidad estructural en un tejido pueden evaluarse desde las perspectivas de intensidad y fase. En la Figura 4.5 se muestran las vistas *en-face* de datos originales, predicciones del modelo, datos interpolados a partir de tomogramas submuestreados de un mismo tomograma y región. La interpolación, una técnica clásica de procesamiento de imágenes, se utiliza para estimar los valores de píxeles faltantes en imágenes submuestreadas, permitiendo comparar el rendimiento de nuestro modelo con métodos tradicionales de recuperación de información. Además de establecer el modelo con mejor rendimiento, se busca comparar las capacidades de Pix2Pix con técnicas clásicas de procesamiento como la interpolación, para así determinar el beneficio del uso de tecnologías de aprendizaje profundo sobre el procesamiento tradicional en la recuperación de información perdida en OCT.

Al analizar la versión submuestreada, la cual corresponde a la última columna de la Figura 4.5, se observa cómo la degradación de la resolución lateral impide representar correctamente las estructuras vasculares más pequeñas. Estas estructuras discontinuas comienzan a comportarse como patrones de ruido espacialmente no correlacionados, lo que compromete la integridad estructural. En la perspectiva lateral, figura 4.6, algunos puntos clave están ausentes en comparación con el tomograma original, lo que resalta la pérdida de información entre píxeles causada por el submuestreo. Esta pérdida de datos distorsiona la representación de las propiedades espaciales del tejido, tal y como se observa en el recuadro amarillo de la Figura 4.5.

La interpolación se presenta como una primera estrategia para completar la información faltante, considerando únicamente las condiciones adyacentes. Sin embargo, esta metodología no logra recuperar correctamente la información y no proporciona continuidad estructural para las características más pequeñas. Utilizando un interpolador lineal a lo largo de todos los planos *en-face*, se realizó una reconstrucción rápida. No obstante, como un problema de regresión, las intensidades obtenidas no tienen en cuenta los cambios de estructura o formas a través de los B-scans, proporcionando únicamente valores de intensidad logarítmica que minimizan los residuos, pero carecen de continuidad estructural.

En la segunda columna de las figuras 4.5 y 4.6, se observan las estructuras

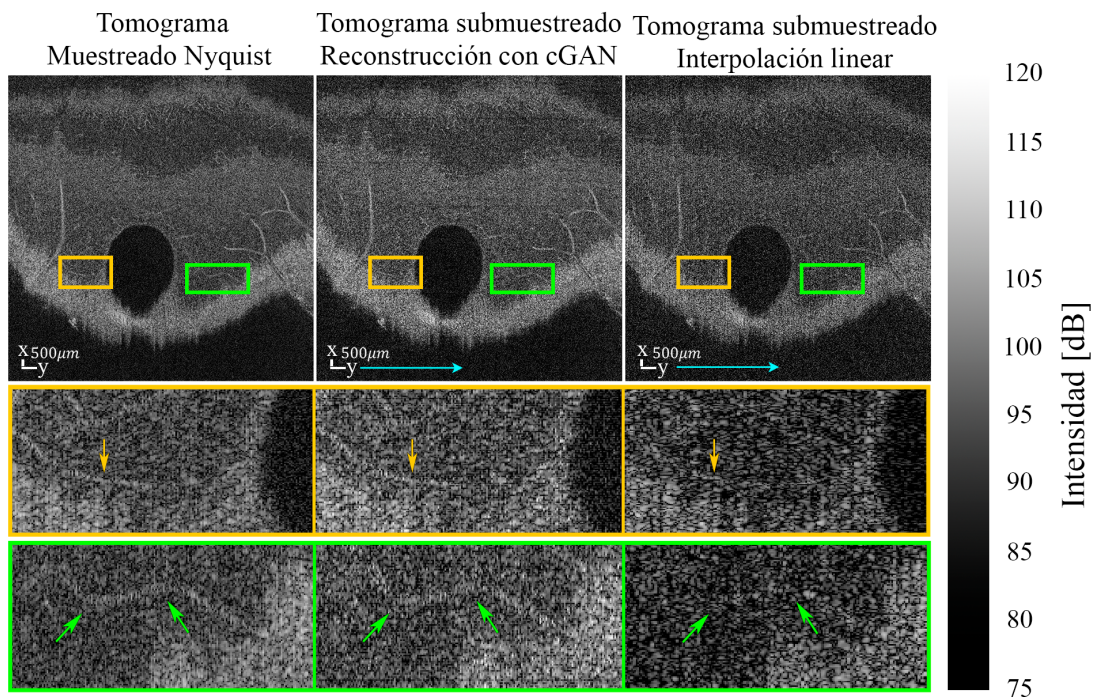


Figura 4.5: Comparación de planos *en-face* de la fóvea. De izquierda a derecha: tomograma original, tomograma reconstruido por la arquitectura cGAN propuesta, tomograma subsampleado interpolado linealmente. Finalmente, la flecha azul indica el eje de subsampleo.

predichas por el modelo Pix2Pix entrenado, donde la orientación preferencial de las estructuras del tejido que se revelan como patrones de intensidad coherentes a lo largo de los B-scans, sin embargo, se nota un cambio en la distribución de *speckle*. Este comportamiento se debe a que el modelo se enfoca en recrear las estructuras del tejido en lugar de reproducir el ruido o el *multiple scattering*. Aunque se presentan cambios en algunas propiedades del *speckle*, la continuidad de los patrones estructurales se recupera a lo largo de los B-scans.

Si observamos los puntos clave, recuadros verde y amarillo en la Figura 4.5, tanto en el tomograma original en la primera columna como en el predicho en la segunda columna, se aprecia una correspondencia estructural entre tejidos vasculares. En la perspectiva lateral mostrado en la figura 4.6, en el caso de la predicción, se observa que los puntos clave se han recuperado, y todos los B-scans siguen un patrón estructural que describe la forma y geometría del tejido de manera congruente y cercana al original, desde una perspectiva cualitativa. Además, se observa que las propiedades de profundidad, como las sombras de los tejidos superiores, también se recuperan en la predicción, mientras que el ruido aumenta en las regiones donde no hay tejido presente.

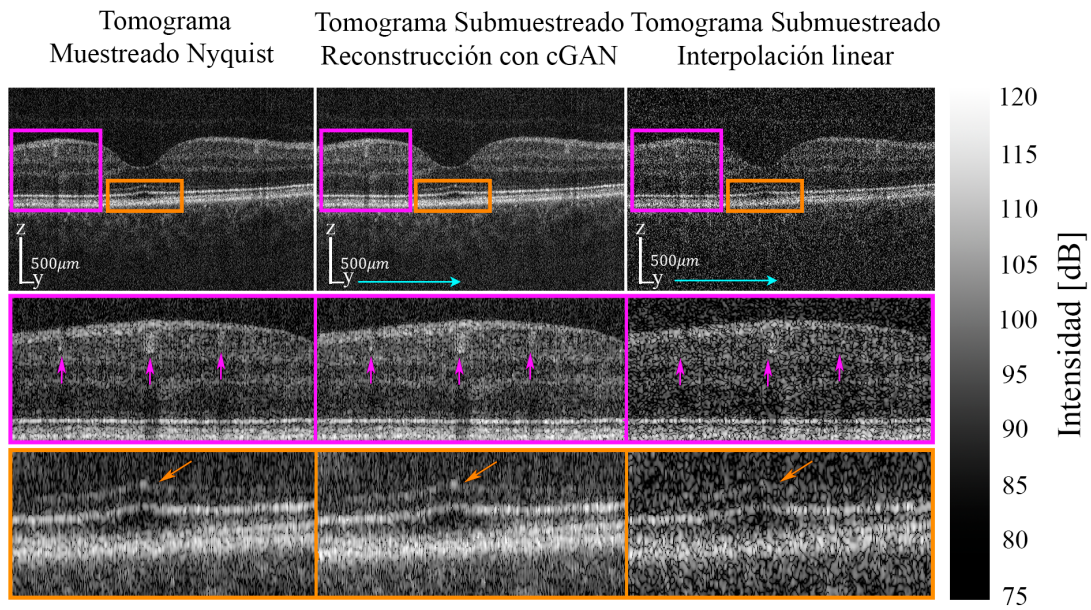


Figura 4.6: Comparación de planos ZY de la fovea. De izquierda a derecha: tomograma original, tomograma reconstruido por la arquitectura cGAN propuesta, tomograma submuestreado interpolado linealmente. La flecha azul indica el eje de submuestreo.

### Análisis de la Reconstrucción de la Estabilidad y Correlación de Fase en el Tejido

Para obtener un análisis cuantitativo más detallado sobre la continuidad estructural del tejido recuperado, se evalúan los cambios de fase locales y la estabilidad de fase. En este análisis, se submuestran lateralmente los canales real e imaginario, y se aplican predicciones a ambas componentes usando Pix2pix. Posteriormente, se calculan las fases a partir de las predicciones y se observan las perspectivas *en-face* de los mapas de fase obtenidos. Para este análisis de fase, se utiliza la misma región *en-face* mostrada en la Figura 4.5.

Las diferencias de fase locales se calculan desplazando un pixel las fases adyacentes en dirección del eje de las B-scans, y en dirección de las A-lines, luego, restándolas fasorialmente de forma independiente con respecto al mapa no desplazado, esto es equivalente al análisis de correlación de retardo unitario del campo complejo. La Figura 4.7 muestra las diferencias de fase obtenidas. Cabe destacar que el tomograma original está estabilizado en fase a lo largo del eje de las A-lines, pero no en el eje de las B-scan [124–126] por la forma en que se adquirió, lo cual se refleja en un mayor nivel de ruido en las diferencias de fase en el eje de las B-scan.

En la última columna de la Figura 4.7, se observa que la estabilidad de fase se pierde parcialmente cuando el tomograma es subsampleado lateralmente. Las

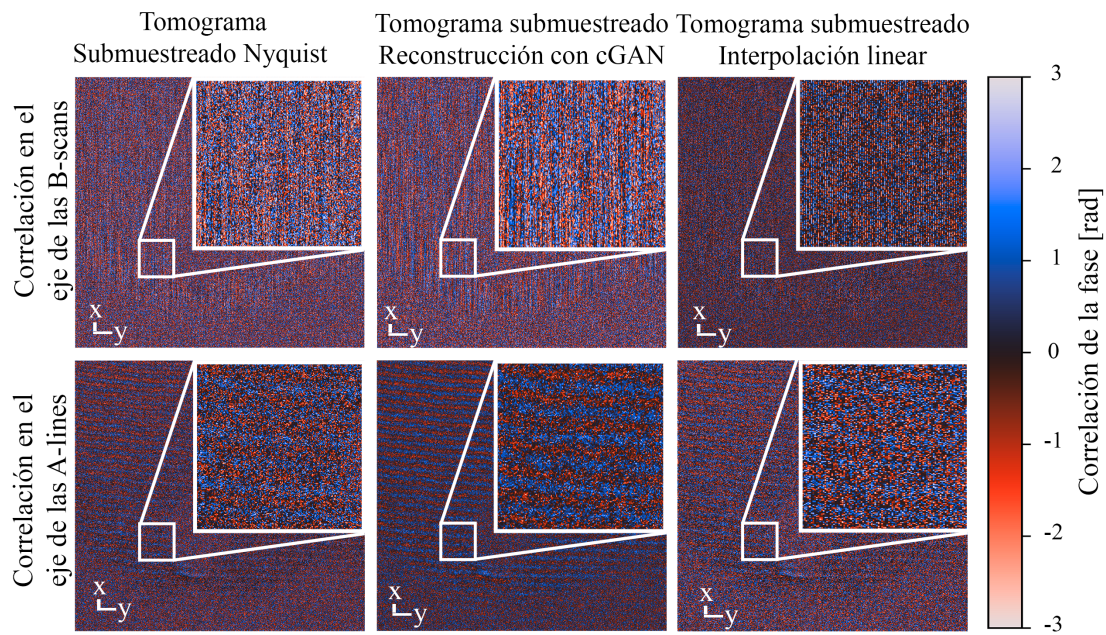


Figura 4.7: Diferencias de fase en dirección del eje de las B-scans Y en dirección de las A-líneas para el plano *en-face* del tomograma experimental de la fovea.

diferencias de fase muestran gradientes de fase locales aleatorios, con dislocaciones de fase ( $|\Delta\phi| \geq \pi$ ) que se extienden a lo largo de la región *en-face*, lo que implica la presencia de estructuras adyacentes no correlacionadas.

Tomando como referencia las versiones submuestreadas, se puede observar cómo las técnicas de reconstrucción afectan la estabilidad espacial en el tejido. En el caso de la reconstrucción con interpolación lineal en la Figura 4.7 tercera columna, la estabilidad de fase no se recupera, y se muestran mayores desplazamientos de fase aleatorios. Este comportamiento corrobora que el proceso de regresión no considera el comportamiento estructural, aplicando solo valores numéricos que satisfacen la función de pérdida, pero perdiendo completamente el contexto espacial de las estructuras. Este resultado concuerda con las observaciones hechas en la comparación de intensidades en la Figura 4.5, donde se demuestra que la interpolación daña aún más la correlación estructural.

En la segunda columna de la Figura 4.7, se observan diferencias de fase con un predominante tono oscuro, lo que indica una tendencia hacia cero. Esta tendencia implica una transición más suave entre estructuras adyacentes, lo que sugiere que los cambios espaciales son menos abruptos y están más correlacionados espacialmente. En comparación con el tomograma original, no solo se recupera la estabilidad de fase inicial, sino que los patrones espaciales también muestran correlación con límites de fase similares entre sí. Las diferencias de fase tienden a variar principalmente en las regiones con mayor absorción, donde se presentan fases aleatorias. Sin embargo, el

modelo intenta predecir fases correlacionadas incluso en estas áreas. Este comportamiento sugiere que el modelo se enfoca en características del tejido más que en características de ruido, como se discutió previamente con las intensidades.

De este modo, el modelo demuestra ser capaz de interpretar el contexto espacial de las estructuras entre píxeles en señales dominadas por tejido, recreando transiciones estructurales congruentes, a pesar de alterar algunas propiedades de distribución de intensidad, lo cual se discutirá más adelante en el análisis SSIM. A partir de la estabilidad observada en las diferencias de fase, se sugiere que la generación de datos correlacionados es independiente del valor complejo inicial de los B-scans adyacentes, lo que indica que el modelo tiende a estabilizar los canales del campo complejo para recrear la fase.

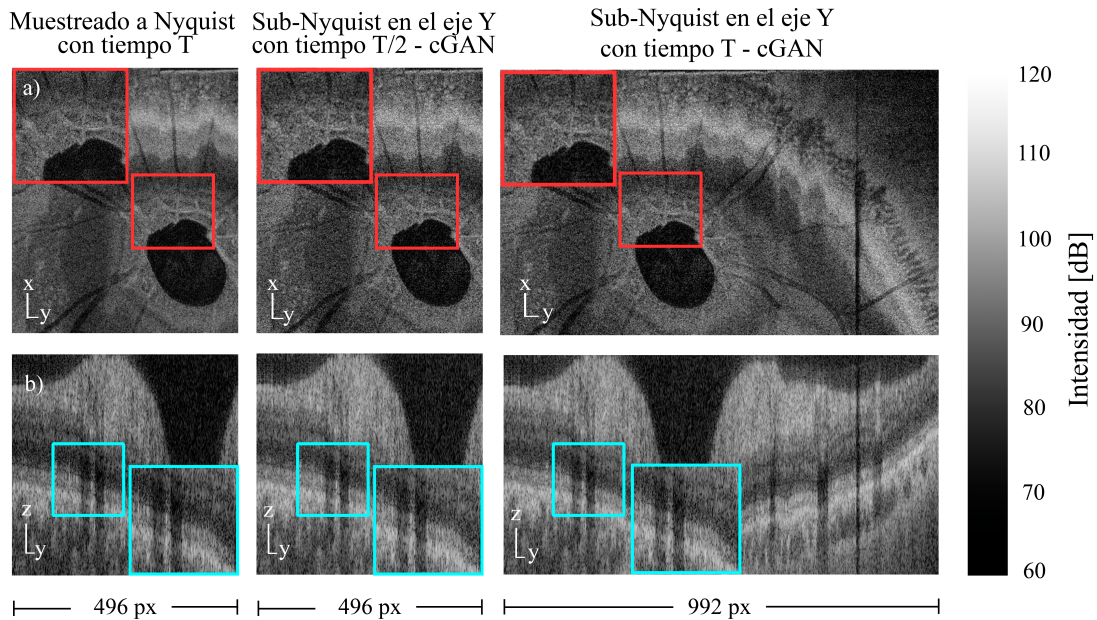


Figura 4.8: Comparación entre muestreo de Nyquist y sub-Nyquist en el eje Y con distintos tiempos de adquisición. Fila (a) muestra un plano *en-face* del nervio óptico humano, fila (b) muestra el plano lateral. La columna izquierda corresponde al muestreo de Nyquist con un tiempo de adquisición T, la columna central al muestreo sub-Nyquist reconstruido por cGAN con un tiempo de adquisición T/2, y la columna derecha al muestreo sub-Nyquist en el eje Y reconstruido por cGAN con un tiempo de adquisición T.

La Figura 4.8 muestra tres estructuras predichas por el modelo a partir de tomogramas experimentales subsampleados lateralmente, con el mismo campo de visión (FOV) y la mitad del tiempo de adquisición, y el doble del FOV con el mismo tiempo de adquisición.

Esta figura también ilustra una posible aplicación derivada de la metodología de submuestreo y predicción del modelo. Por ejemplo, considere que un volumen

muestreado según la tasa de Nyquist con dimensiones laterales de 496 píxeles requiere un tiempo  $T$  para su adquisición, obteniendo un FOV definido del tejido (ver primera columna de la Figura 4.8). Si se quisiera aumentar la dimensión lateral en un factor de 2, se requeriría el doble de tiempo  $T$  para duplicar el FOV. No obstante, si aplicamos la metodología de subsamplio y predicción, podríamos reducir el tiempo de adquisición a la mitad para escanear el mismo volumen con el mismo FOV (ver segunda columna de la Figura 4.8), o podríamos subsampliar lateralmente y reconstruir un tomograma de 992 píxeles de tamaño lateral en el mismo tiempo  $T$  para duplicar el FOV (ver tercera columna de la Figura 4.8).

Con esta estrategia, es posible optimizar la cantidad de datos adquiridos a una misma tasa de A-lines, lo que reduce el tiempo requerido o aumenta el campo de visión en un menor tiempo. Esto ayuda a reducir la decorrelación de fase durante periodos prolongados de adquisición. En la Figura 4.8, también se puede observar el rendimiento del modelo propuesto en distintos tejidos. Se aprecia que el modelo logra una mayor fidelidad en las distribuciones de intensidad en las regiones más densas en tejido.

### Desempeño del Modelo en Volúmenes Completos: Métricas

Aunque el modelo ha sido verificado en diferentes regiones y tejidos, es importante analizar su comportamiento en todas las posibles distribuciones dentro del volumen completo del tomograma de la fovea. Para ello, se utilizan cuatro métricas que permiten evaluar la similitud estructural y las diferencias escaladas, desde regiones dominadas por ruido fotónico hasta zonas con presencia parcial de tejido en un tomograma ocular. En cada plano *en-face* del conjunto de datos predicho y original, se calculan las métricas, cuyos valores se promedian y se asignan a cada índice de plano correspondiente. Los resultados se presentan en la Figura 4.9.

Cada métrica evalúa el desempeño del modelo para reconstruir regiones con predominio de ruido fotónico, donde las estructuras de tejido son mínimas o inexistentes, hasta llegar a zonas densas en tejido ocular y vascular. En el gráfico superior de la Figura 4.9, la curva de SSIM permite analizar los cambios estructurales a diferentes profundidades. Para una visualización completa de la evolución estructural a lo largo de todos los planos *en-face*, se ha incluido un video detallado en el siguiente repositorio: [Material suplementario](#). Este video complementa el análisis estático de la Figura 4.9, permitiendo observar la continuidad y las variaciones estructurales de manera dinámica a través del volumen reconstruido. En la región inicial, donde predominan el ruido y pequeñas regiones de tejido, se observa que el modelo predice distribuciones de ruido que no corresponden completamente con las originales. A medida que las estructuras de tejido se vuelven dominantes en la proyección *en-face*, la similitud estructural aumenta, reflejando la mayor densidad de

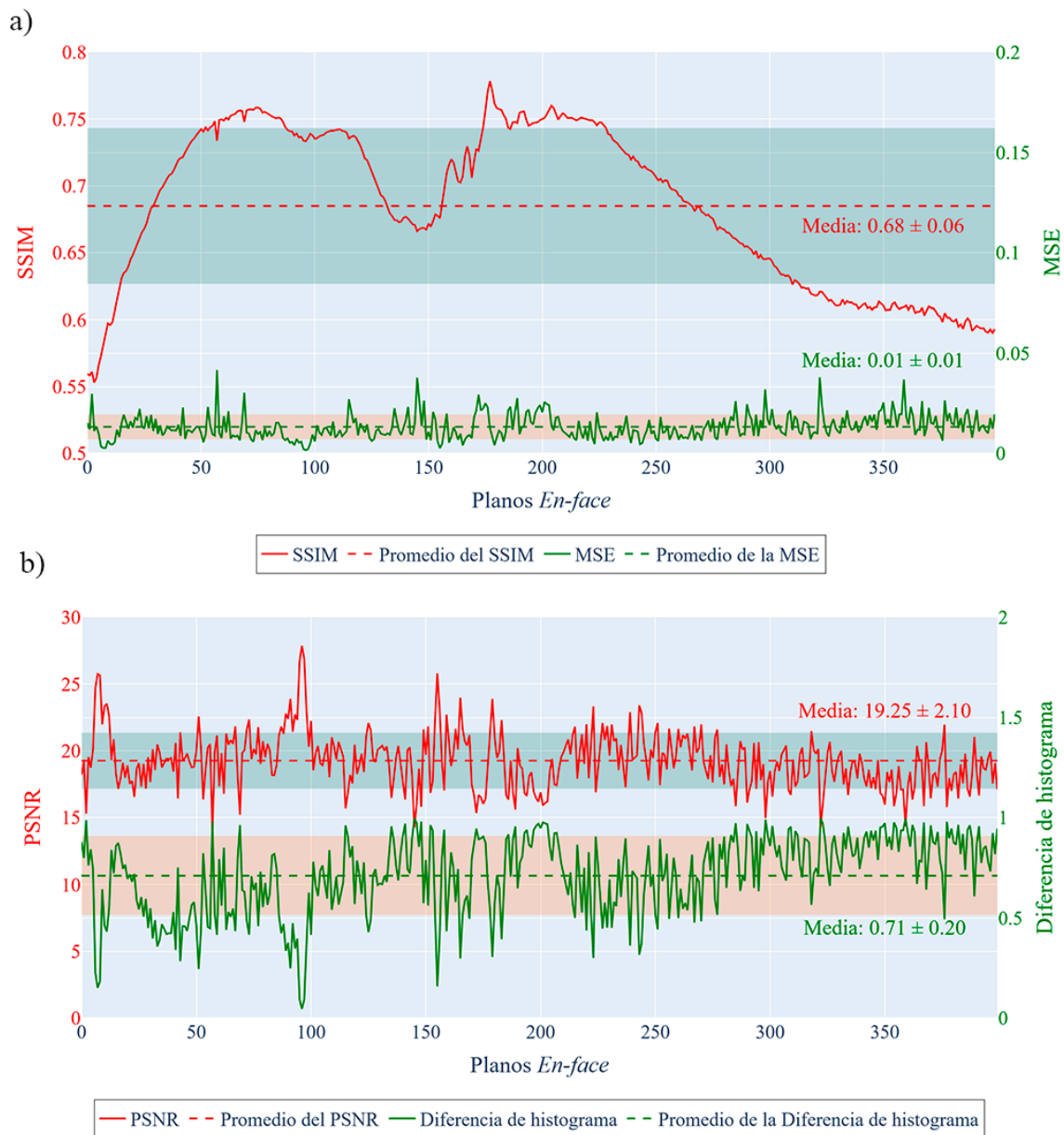


Figura 4.9: En a) se muestra el Comportamiento del índice de similitud estructural (SSIM) y el error cuadrático medio (MSE) a través de los planos *En-face*. En b) se presenta la Variación en la relación señal-ruido pico (PSNR) y la diferencia de histograma en los planos *En-face*.

los elementos estructurales. Al avanzar a través del tomograma ocular, se alcanza la región macular, donde debido a su densidad y absorción, el ruido fotónico aumenta y la SSIM disminuye, hasta que el tejido vuelve a ser dominante, lo que incrementa de nuevo la similitud estructural. Finalmente, al salir de las regiones densas en tejido, el ruido y otros fenómenos de dispersión se hacen más dominantes, lo que reduce la SSIM, marcando cambios en la distribución de intensidades hasta llegar al final del tomograma, en regiones vacías de tejido.

La métrica de MSE describe una comparación promedio píxel a píxel de las escalas absolutas de las intensidades logarítmicas. El hecho de que el error oscile de manera constante en torno a su media muestra que el error es consistente, con poca varianza, indicando que la intensidad promedio es cercana a la escala original. El valor medio de MSE es de aproximadamente  $\overline{MSE} \approx 0,012\text{dB}^2$ . En el gráfico inferior, se utiliza la similitud coseno para estimar la diferencia de histograma normalizada, proyectando los diferentes vectores de intensidades logarítmicas del original y el predicho. Para la construcción del histograma, se utilizaron 256 bins, los cuales son intervalos que agrupan rangos de intensidad, permitiendo una resolución adecuada para capturar las variaciones de intensidad en las imágenes normalizadas entre 0 y 1. La similitud de coseno muestra una correspondencia general entre las proyecciones de barras del histograma, donde las regiones con mayor densidad de tejido son menos similares que las regiones dominadas por ruido.

Este resultado indica que las propiedades de *speckle* en estas regiones densas de tejido no son completamente iguales. Por ejemplo, en el plano *en-face* 100, pueden volverse antagónicas. Combinando SSIM y similitud coseno, se puede cuantificar que, aunque la similitud estructural se recupera en las predicciones y las transiciones de fase son suaves, la descripción de intensidades del modelo puede distanciarse de la distribución original en algunos casos particulares. En consecuencia, para mejorar la representación de las propiedades de *speckle* y la similitud de los histogramas, se considera que el modelo podría aumentar su rendimiento al mejorar su capacidad para describir características de tejido más pequeñas, una posible alternativa a explorar es la incorporación de más datos experimentales con el fin de que el modelo pueda aprender mejor de las distribuciones de esparcimiento multicapas y de esparcimiento en vasos sanguíneos.

Finalmente, en las regiones de tejido comprendidas entre los planos *en-face* 50 – 120, los valores de MSE se mantienen consistentemente cercanos a la media, lo que destaca la habilidad de la red para reconstruir con precisión las regiones densas en información de tejido. A medida que el tomograma se desplaza hacia la zona de desvanecimiento, a partir del plano 240, se observa un aumento en los valores de MSE, lo que indica ciertos desafíos que enfrenta la red a medida que las estructuras del tejido se vuelven menos definidas.

El cálculo de la Relación Señal-Ruido Pico (PSNR) utiliza la definición en la Ecuación (4.7). Generalmente, un PSNR  $> 25\text{dB}$  se considera de buena calidad en aplicaciones de compresión de imágenes [127], mientras que valores PSNR  $< 15\text{dB}$  indican baja calidad. Esta métrica permite cuantificar la similitud entre las imágenes en cuanto a la escala de diferencia entre intensidades. El PSNR presenta fluctuaciones que indican baja calidad de señal en regiones de ruido. En las zonas de tejido, los valores de PSNR se estabilizan por encima de la media. En la zona de desvanecimiento,

muestra una tendencia negativa lineal, sugiriendo una transición gradual de regiones con tejido definido hacia áreas con menor cantidad de tejido y, potencialmente, mayor ruido. Un MSE bajo entre los tomogramas de referencia y los predichos se traduce en un PSNR alto, lo que indica que el cGAN ha logrado una reconstrucción cercana en la escala de señal. Los picos pronunciados, especialmente alrededor de los planos 60 y 380, sugieren regiones con mayores diferencias en las predicciones, coincidiendo con las regiones periféricas donde está presente el ruido fotónico. Los resultados indican la necesidad de incluir información sobre el ruido fotónico y ajustar el modelo para mejorar el aprendizaje en estas zonas, especialmente en las regiones de inicio y finalización del tomograma.

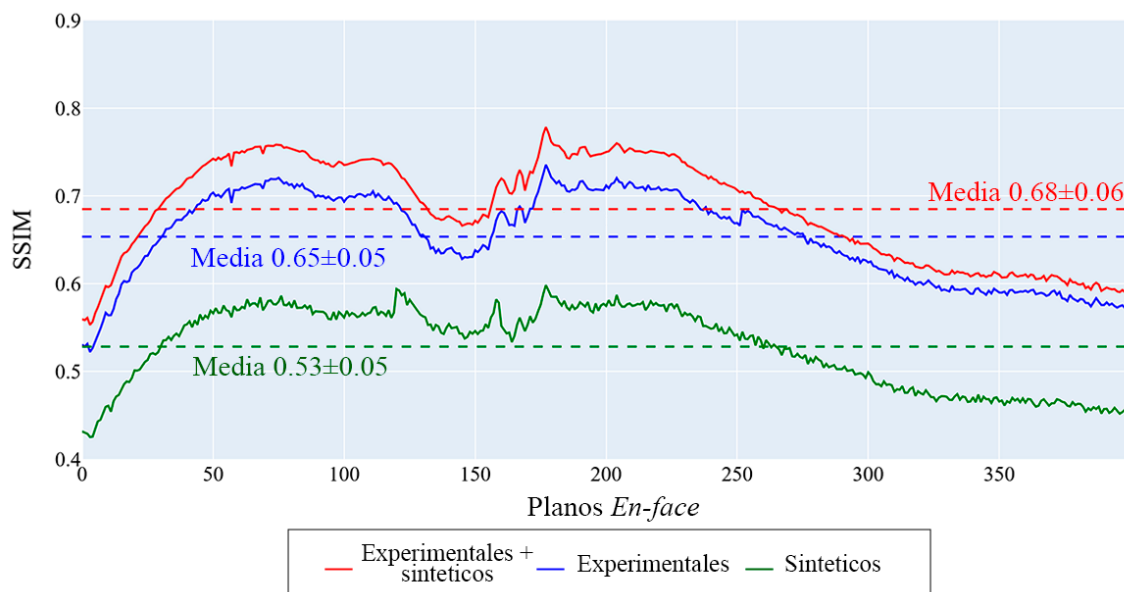


Figura 4.10: Evaluación del SSIM en función del número de planos en-face diferentes conjuntos de datos: experimentales y sintéticos combinados (rojo), únicamente experimentales (azul) y únicamente sintéticos (verde). Las líneas punteadas horizontales representan el valor medio del SSIM para cada conjunto de datos.

Los resultados de la evaluación presentados en la figura 4.10 confirman la tendencia observada en las pruebas iniciales (Figura 4.4, gráfico superior izquierdo), en esta prueba se mantienen constantes los hiperparámetros de la red Pix2Pix en los valores óptimos determinados tras la etapa de optimización. La única diferencia entre los tres modelos evaluados radica en la naturaleza del conjunto de datos de entrenamiento: una combinación de tomogramas experimentales y sintéticos, únicamente tomogramas experimentales o únicamente tomogramas sintéticos. A pesar de utilizar la misma configuración de hiperparámetros optimizada para todos los casos, la combinación de datos experimentales y sintéticos continúa demostrando un rendimiento superior en términos de SSIM en comparación con el entrenamiento utilizando conjuntos de datos de una sola naturaleza. Este hallazgo refuerza la hipótesis de que la diversidad

inherente a la combinación de datos reales y simulados enriquece el proceso de aprendizaje del modelo, independientemente de la configuración específica de los hiperparámetros. La inclusión de tomogramas sintéticos, que pueden ofrecer una cobertura más amplia del espacio de características o resaltar aspectos específicos de la estructura que podrían ser menos frecuentes en los datos experimentales, complementa la información proporcionada por las imágenes reales, las cuales aportan la complejidad y variabilidad propias de las mediciones. Esta sinergia permite al modelo generalizar mejor y producir reconstrucciones de mayor fidelidad, incluso al variar la profundidad dentro del volumen analizado, como se evidencia en la evaluación a lo largo de los planos en-face. La consistencia de este resultado, incluso con hiperparámetros optimizados, subraya la robustez del beneficio de combinar datos de ambas fuentes para la tarea de generación de tomogramas.

### Rendimiento con TNode

TNode es un algoritmo de reducción de speckle basado en el principio de promedios no locales, diseñado para disminuir las distribuciones de intensidad no tisular y resaltar las estructuras del tejido [38]. Es importante destacar que también existe una implementación de TNode basada en aprendizaje profundo, reportada en [75], lo cual concuerda con el enfoque de este trabajo y demuestra las fortalezas del aprendizaje profundo en el procesamiento de imágenes OCT. Dado que TNode opera sobre intensidades, el análisis se centra en comparaciones de planos en intensidades logarítmicas. Se compara el tomograma original con el tomograma reconstruido mediante el modelo Pix2Pix, ambos procesados con TNode, para evaluar la capacidad del modelo en la recuperación de la continuidad y similitud estructural.

La Figura 4.11 muestra los resultados de aplicar TNode a ambos tomogramas. El tomograma muestreado a Nyquist en la columna izquierda y el tomograma reconstruido por el modelo en la columna derecha presentan zonas de estructuras densas. Al comparar la región resaltada en magenta, se observa que el tomograma reconstruido por el modelo presenta bordes más definidos en estructuras pequeñas, indicando una mejor recuperación de detalles finos en comparación con el tomograma original. Además, la aplicación de TNode al tomograma reconstruido mitiga el efecto secundario del modelo de añadir varianza de speckle, resultando en una mejor representación de la relación señal-ruido del tejido y una recuperación de la continuidad espacial, conservando únicamente las estructuras espaciales altamente correlacionadas.

Para evaluar la capacidad del modelo en la reconstrucción de la similaridad estructural, se aplicó TNode para reducir el ruido fotónico y otras distribuciones no tisulares. La Figura 4.12 compara el índice de similitud estructural entre el tomograma original y el reconstruido por el modelo, antes y después de aplicar

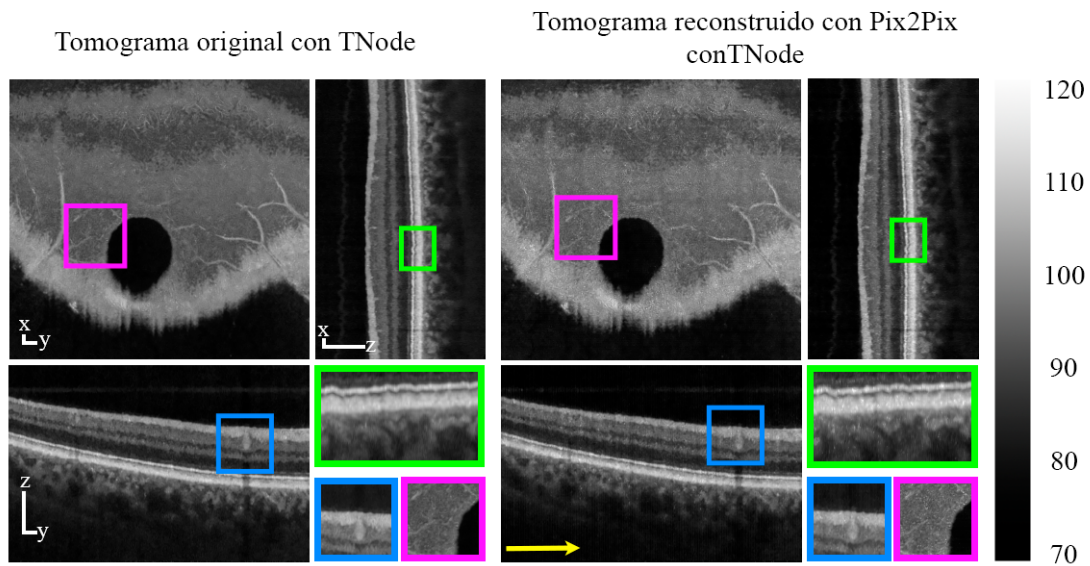


Figura 4.11: Comparación de tomogramas de fóvea humana después de aplicar TNode: La columna izquierda muestra el tomograma original muestreado a la tasa de Nyquist. El cuadrante superior derecho muestra el tomograma muestreado por debajo de Nyquist reconstruido utilizando el método de aprendizaje profundo propuesto con una cGAN. Cada región resaltada permite una evaluación detallada de la calidad de reconstrucción por diferentes métodos.

TNode.

La Figura 4.12 muestra que las diferencias en SSIM se deben principalmente a la interpretación del ruido por parte del modelo, lo que afecta la similitud entre planos *en-face*, pero no la continuidad estructural. El desplazamiento entre las gráficas indica que aproximadamente el 30% de la diferencia se debe a la dispersión múltiple y otros fenómenos de dispersión fuera del dominio tisular, que el modelo no representa completamente.

La Tabla 4.3 presenta una comparación de diferentes métricas de evaluación antes y después de aplicar TNode.

Condición	SSIM	MSE	PSNR	Dif. Histograma
pre-TNode	$0.685 \pm 0.0583$	$0.013 \pm 0.0062$	$19.25 \pm 2.10$	$0.711 \pm 0.1959$
TNode	$0.887 \pm 0.0756$	$0.0007 \pm 0.0012$	$34.04 \pm 4.37$	$0.266 \pm 0.2663$

Tabla 4.3: Comparación de las métricas promedio para planos *en-face* antes y después de aplicar TNode. Las métricas incluyen el Índice de SSIM, MSE, PSNR y la Diferencia de Histograma.

Los resultados de la Tabla 4.3 indican que el modelo Pix2Pix logra métricas adecuadas para aplicaciones médicas, con mejoras significativas en SSIM, MSE y PSNR. Aunque TNode mejora aún más estas métricas, es un postprocesamiento general

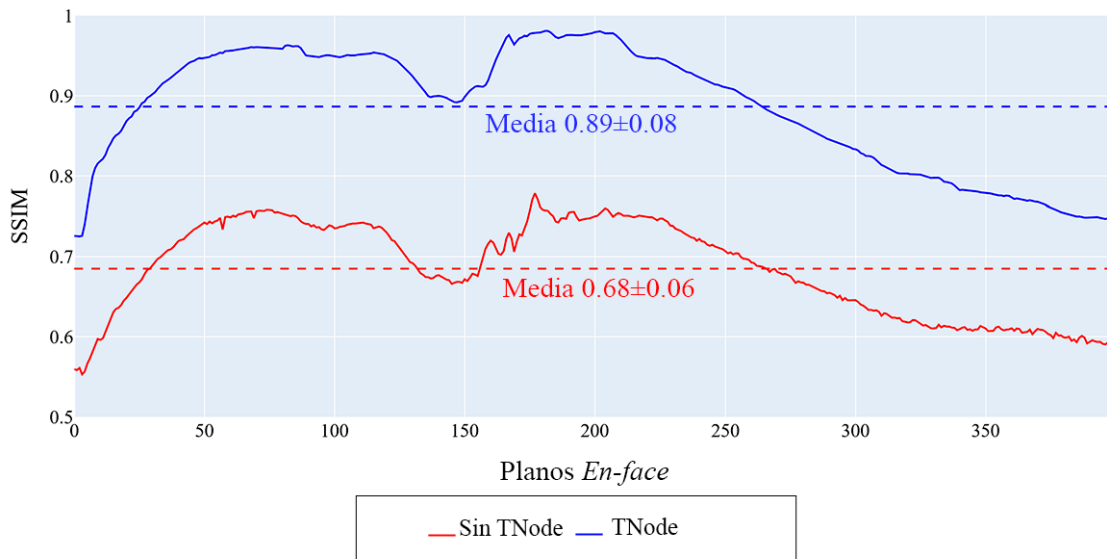


Figura 4.12: Comparación de SSIM entre los tomogramas muestreado a Nyquist y reconstruido mediante Pix2Pix antes y después de aplicar TNode. La línea azul representa el SSIM después de aplicar TNode, y la línea roja representa el SSIM antes de aplicar TNode. Las líneas discontinuas indican los valores promedio de SSIM para cada caso.

aplicado a todos los tomogramas. El enfoque principal de este trabajo es demostrar la capacidad del modelo Pix2Pix para reconstruir tomogramas submuestreados, logrando resultados cercanos a los obtenidos con tomogramas bien muestreados, que requieren más tiempo de adquisición.

En resumen, Pix2Pix logra reconstrucciones de buena calidad, y TNode puede refinar aún más estos resultados. Esto demuestra el potencial del modelo de aprendizaje profundo para recuperar detalles perdidos por el submuestreo, ofreciendo una solución viable para reducir los tiempos de adquisición sin comprometer significativamente la calidad de imagen. TNode permite una representación más cercana a los datos originales, indicando que las diferencias entre las intensidades reconstruidas por el modelo y los datos originales se deben principalmente a distribuciones de ruido de speckle presentes en los tomogramas experimentales. Si bien TNode podría aplicarse directamente a tomogramas submuestreados, es importante señalar que, al ser un algoritmo de despeckling, no sería capaz de recuperar la información estructural deformada o faltante debido al submuestreo inherente en la adquisición. Por lo tanto, su principal utilidad en este contexto radica en el refinamiento de las reconstrucciones generadas por el modelo de aprendizaje profundo, actuando sobre los artefactos de ruido más que sobre las deficiencias estructurales causadas por la falta de información.

#### 4.4.2. Mitigación de artefactos espejo usando aprendizaje profundo

La Figura 4.13 muestra los resultados obtenidos al realizar pruebas de variación de hiperparámetros para las arquitecturas evaluadas. Es importante mencionar que, dado el éxito demostrado previamente al combinar tomogramas sintéticos y experimentales en el problema del submuestreo lateral, todas las pruebas realizadas en esta etapa consideran exclusivamente esta combinación de datos. Este enfoque busca mantener la consistencia en los datos de entrada y evaluar de manera más precisa el impacto de cada hiperparámetro.

En primer lugar, se observa la variación de la profundidad de la red en términos del número de capas. Los resultados indican que la configuración con siete capas proporciona el menor MSE, con un valor promedio de  $0,0473 \pm 0,0029\text{dB}^2$ . Este comportamiento podría explicarse por la capacidad de las redes más profundas de capturar características de mayor nivel, especialmente en datos con estructuras complejas, como los presentes en OCT. Sin embargo, al aumentar la profundidad a más de siete capas, el MSE incrementa ligeramente, lo que sugiere un posible sobreajuste o dificultades en la optimización debido al aumento en la complejidad del modelo.

Posteriormente, con la arquitectura óptima en profundidad seleccionada, se evaluaron distintas configuraciones de regularización. La regularización del 20% arrojó el mejor desempeño, con un MSE promedio de  $0,0539 \pm 0,0038$ . Este resultado sugiere que un nivel moderado de regularización permite reducir la sobreadaptación al conjunto de datos de entrenamiento sin comprometer la capacidad de la red para generalizar. Por el contrario, niveles más altos de regularización, como el 50%, aumentan el MSE, probablemente debido a la restricción excesiva que impone al modelo.

Finalmente, se analizaron distintos optimizadores, utilizando las configuraciones óptimas de profundidad y regularización identificadas previamente. El optimizador Adam presentó el mejor desempeño, con un MSE promedio de  $0,0391 \pm 0,0005$ . Este valor supera significativamente los resultados obtenidos con RMSprop, que registró un MSE promedio de  $0,1002 \pm 0,1713$ , lo que representa una diferencia de 0,0611 unidades de MSE. Además, Adam también superó a SGD, que obtuvo un MSE promedio de  $0,0426 \pm 0,0025$ , lo que implica una diferencia de 0,0035 unidades de MSE, y una diferencia en la desviación estándar de 0,0020. Estos resultados confirman que Adam maneja de manera más efectiva la convergencia del modelo en este contexto. Este resultado puede explicarse por la capacidad de Adam para combinar las ventajas del descenso por gradiente estocástico con un ajuste adaptativo de la tasa de aprendizaje, lo que resulta particularmente útil en problemas donde los gradientes pueden variar significativamente entre las iteraciones. Además, su uso generalizado en problemas de aprendizaje profundo, incluidos los relacionados con

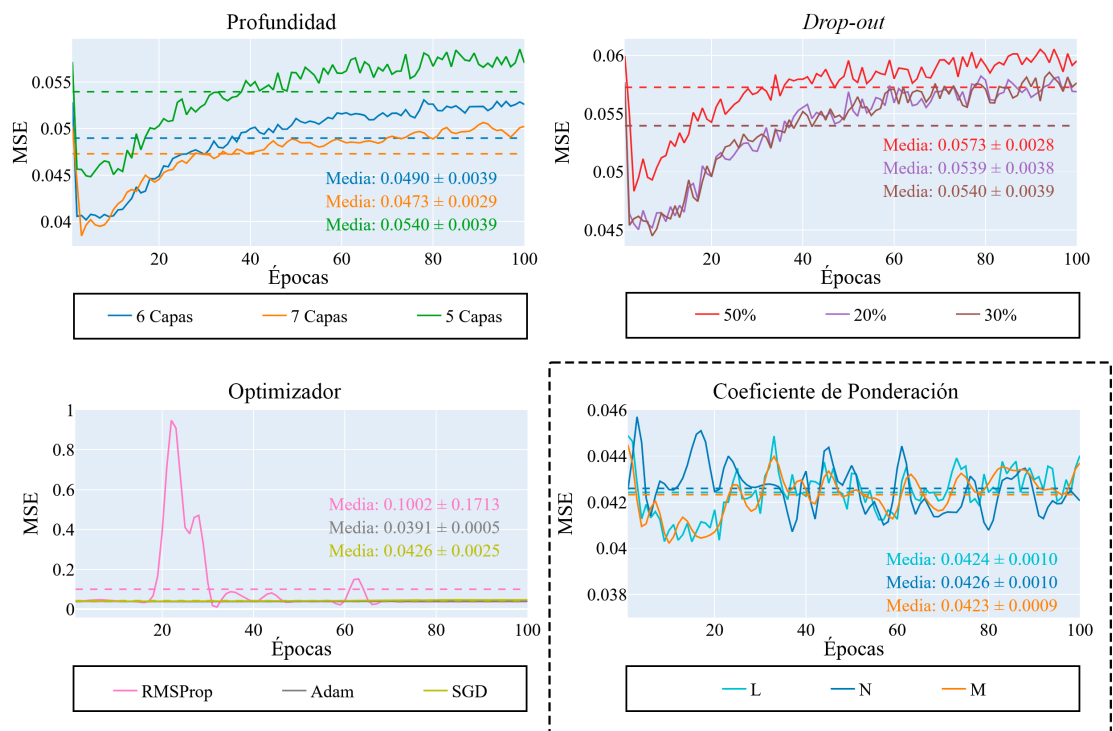


Figura 4.13: Gráficos del MSE por época para distintas configuraciones de hiperparámetros. Las líneas continuas representan los resultados de cada configuración, mientras que las líneas punteadas indican el valor promedio del MSE para cada caso. el recuadro negro con líneas punteadas representan los resultados de los hiperparámetros exclusivos de la red CCARNet

datos biológicos, refuerza su idoneidad para la reconstrucción de tomogramas en este contexto.

De forma específica, en la arquitectura CCARNet, los coeficientes de ponderación analizados (L, N y M) mostraron un desempeño consistente, con valores promedio de MSE en un rango estrecho entre 0,0423 y 0,0426  $dB^2$ . Esto sugiere que esta arquitectura es robusta frente a pequeñas variaciones en los coeficientes de ponderación, lo que refuerza su adecuación para problemas que requieren un mse mínimo y resultados estables a través de diferentes configuraciones de ponderación.

### Evaluación cualitativa de las redes en la reconstrucción de B-scans con artefactos espejo

La Figura 4.14 ilustra los resultados de la reconstrucción de un B-scan correspondiente a la fóvea, evaluando el impacto de diferentes porcentajes de solapamiento entre las estructuras reales de la retina y sus respectivas copias conjugadas. En los círculos de colores se señalan regiones críticas donde las capas de la retina intersectan con el artefacto espejo, permitiendo analizar la capacidad de las arquitecturas Pix2Pix

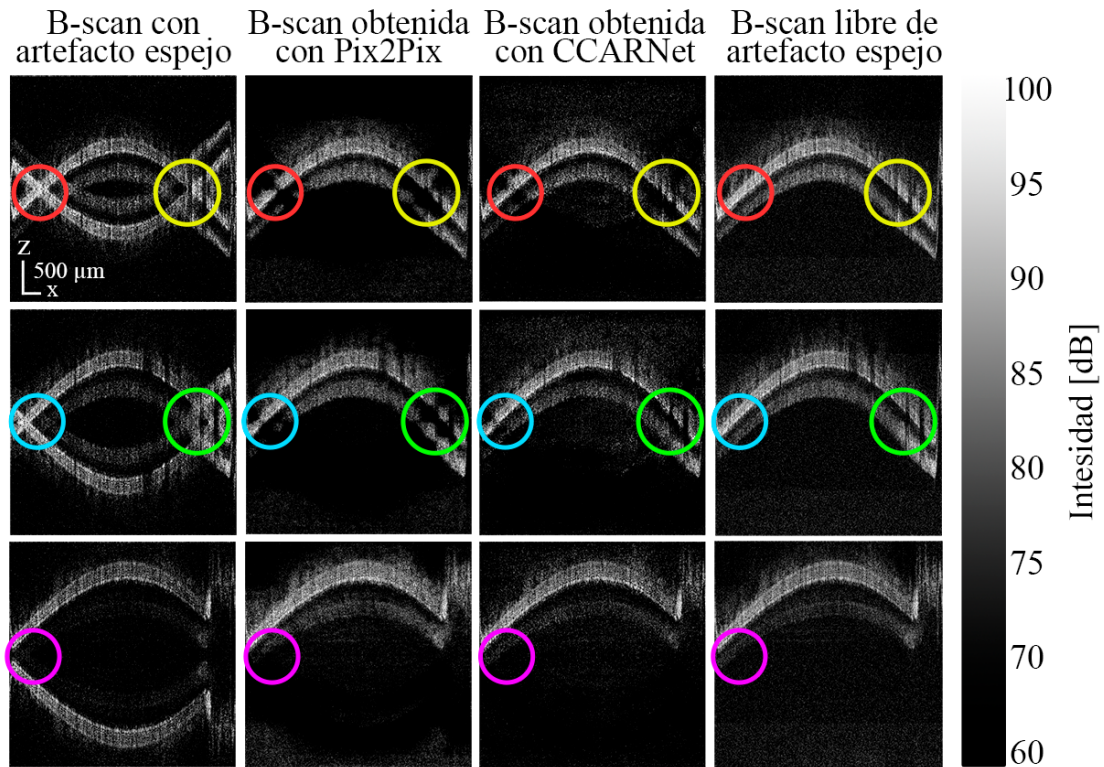


Figura 4.14: Comparación de B-scans con artefactos espejo (primera columna), reconstruidos por la red Pix2Pix (segunda columna), reconstruidos por la red CCARNet (tercera columna) y sin artefactos espejo (cuarta columna). Las regiones marcadas con círculos destacan áreas donde la estructura original se superpone con su artefacto espejo. Las filas muestran distintos grados de solapamiento entre la estructura y su artefacto, lo que permite evaluar el desempeño de la red en diversos escenarios.

y CCARNet para mitigar este efecto y preservar la información estructural.

En las regiones marcadas con los círculos amarillo, verde, rojo y azul, ambas arquitecturas presentan pérdida de información estructural, evidenciando la dificultad inherente de reconstruir áreas donde el tejido real está significativamente solapado con su copia conjugada. Sin embargo, se observa que la red Pix2Pix sufre una degradación más severa en estas áreas, perdiendo detalles importantes en las interfaces entre capas. Esto podría atribuirse a una menor capacidad de Pix2Pix para modelar la interacción compleja entre el artefacto y las estructuras originales, especialmente en zonas donde el solapamiento cubre la mayor parte de tejido. Este comportamiento sugiere que, a medida que aumenta el grado de solapamiento entre las estructuras reales y su copia conjugada, el riesgo de pérdida de información se incrementa significativamente.

Por otro lado, los círculos fucsia señalan áreas donde la copia conjugada apenas se toca con las estructuras reales. En estas regiones, se observa una reconstrucción de información esperada, aunque con menor impacto en las características visibles de las

capas. Este escenario, además de ser predecible, no ofrece resultados particularmente valiosos desde una perspectiva de procesamiento, ya que en la línea típica de procesamiento de OCT este caso simplemente se manejaría como un corte directamente en la frecuencia cero para separar las copias conjugadas.

El análisis cualitativo del ruido de fondo también destaca diferencias entre las arquitecturas. Pix2Pix genera un fondo más uniforme, probablemente debido a una tendencia a suavizar las intensidades en regiones de bajo contraste. Por el contrario, CCARNet conserva un patrón de ruido más cercano al target libre de artefactos, reflejando su capacidad para modelar mejor las características estadísticas del ruido intrínseco en los datos experimentales.

Finalmente podemos observar como este primer análisis indica que posiblemente CCARNet logra una mejor preservación de las estructuras originales en comparación con Pix2Pix, especialmente en áreas con solapamiento significativo.

#### **Análisis de la Correlación de Fase para la eliminación del artefacto espejo**

La Figura 4.15 muestra la correlación de la fase en dos direcciones: el eje axial ( $z$ ) y el eje rápido ( $x$ ), calculada para un B-scan con artefacto espejo, así como para las reconstrucciones obtenidas mediante Pix2Pix, CCARNet, y la B-scan objetivo libre de artefacto. Este análisis permite explorar cómo las redes reconstruyen la relación de fase en el campo complejo, proporcionando una perspectiva complementaria al análisis basado en intensidad.

Desde el punto de vista físico y óptico, la correlación de fase en el eje axial ( $z$ ) refleja la continuidad en las oscilaciones del campo óptico en profundidad, lo cual está directamente relacionado con el nivel de coherencia local entre los frentes de onda. Mientras que en el eje rápido ( $x$ ) evalúa las relaciones laterales entre A-lines. En el B-scan original con artefacto espejo, las regiones solapadas entre las estructuras reales y sus copias presentan una fuerte correlación de fase, ya que el artefacto es una réplica del campo complejo invertido. Este fenómeno explica la simetría observada en la correlación de fase alrededor del centro de la imagen, donde el artefacto y las estructuras reales se intersectan.

Al analizar las reconstrucciones generadas por Pix2Pix y CCARNet, se observa que ambas redes fallan en reconstruir adecuadamente el campo complejo. En lugar de eliminar completamente la información asociada al artefacto espejo, las redes parecen asignar valores similares de fase en las regiones previamente ocupadas por la copia conjugada. Esto se manifiesta en la correlación como patrones persistentes en las mismas ubicaciones donde se encontraba el artefacto. Este fenómeno sugiere que las redes no logran modelar correctamente las discontinuidades o diferencias de fase introducidas por el artefacto, limitándose a redistribuir la información de manera

uniforme.

Desde el punto de vista del procesamiento de señales, este comportamiento puede interpretarse como una deficiencia en la capacidad de las redes para aprender la relación causal entre las componentes real e imaginaria del campo complejo. En un sistema de OCT, la fase del campo está intrínsecamente relacionada con el número de onda ( $k$ ) y la profundidad ( $z$ ). La interferencia producida por las reflexiones en las capas del tejido actúa como una “portadora” que modula la fase, desplazando el tomograma en el plano  $z$  y permitiendo codificar la fase en el coseno de la señal interferométrica. Si la red no reconstruye adecuadamente estas relaciones, los valores de fase en las regiones reconstruidas carecen de consistencia física, lo que resulta en la persistencia de fase falsas en las áreas donde antes se encontraba el artefacto.

En la correlación del eje rápido ( $x$ ), se observa un comportamiento similar, incluso más marcado. Las redes generan patrones de fase que, aunque eliminan la simetría más evidente del artefacto espejo, no reflejan las transiciones naturales esperadas en la correlación de fase lateral. Esto se debe, en parte, a que la información mezclada en los campos complejos de reflejo y original dificulta la reconstrucción, generando discontinuidades en la fase. Esto sugiere que las redes tienden a suavizar la fase en estas regiones, asignando valores que si bien eliminan el artefacto, no logran una reconstrucción perfecta de los detalles locales del campo complejo.

Si bien ambas redes logran atenuar visualmente el artefacto espejo en términos de intensidad, no reconstruyen de manera fiel las propiedades del campo complejo subyacente. Esto tiene implicaciones importantes en aplicaciones donde la información de fase es crítica, como en la estimación de propiedades ópticas del tejido o en el análisis de dispersión múltiple. Estos resultados sugieren la necesidad de enfoques complementarios que permitan mejorar la reconstrucción de la fase, asegurando una representación más física y consistente del campo complejo en OCT.

La Figura 4.16 muestra un análisis complementario diseñado para mejorar la reconstrucción del campo complejo en las regiones evaluadas. El procedimiento se basa en reconstruir una señal compleja a partir de la magnitud de la entrada y la fase de la salida de las redes, utilizando la expresión:

$$S_{\text{rec}} = |S_{\text{in}}| \cdot e^{i \cdot \phi_{\text{out}}}, \quad (4.11)$$

donde  $S_{\text{rec}}$  representa la señal reconstruida,  $|S_{\text{in}}|$  es la magnitud del campo complejo de entrada, y  $\phi_{\text{out}}$  es la fase de la salida de la red. Este análisis se centra en dos capas biológicas de la retina, identificadas como la esclera ubicada en la capa más externa y la coroides que se encuentra en la capa más interna. Dichas regiones de interés están marcadas con líneas y recuadros verdes y fucsias respectivamente,

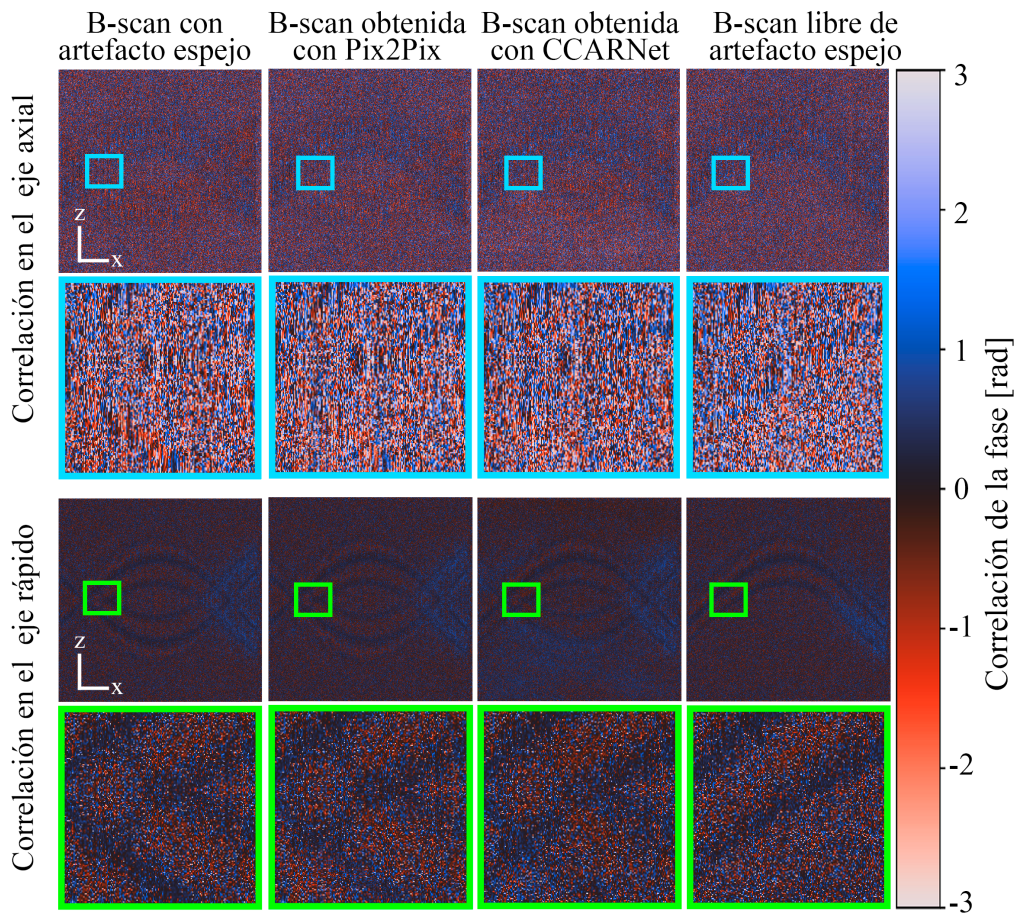


Figura 4.15: Correlación de la fase a lo largo de los ejes axial y rápido en imágenes B-scan de OCT. La columna izquierda muestra la correlación de fase de un B-scan con artefactos de conjugado complejo, mientras que las siguientes columnas presentan reconstrucciones utilizando Pix2Pix, CCARNet y una imagen libre de artefactos espejo como referencia.

las cuales se encuentran tanto en el semi-eje negativo como en el semi-eje positivo del eje de frecuencias. En esta *A-line* específica, las frecuencias no se mezclan, por lo que las capas reales están claramente definidas en el semi-eje negativo, mientras que el semi-eje positivo corresponde a la copia generada por el artefacto espejo. En el semi-eje negativo, las comparaciones reflejan degradación en las estructuras reales. Para la capa externa, el target presenta un valor medio en la magnitud normalizada de  $0,46 \pm 0,20$ , mientras que Pix2Pix y CCARNet alcanzan valores de  $0,43 \pm 0,21$  y  $0,29 \pm 0,28$ , respectivamente. Estos resultados indican que CCARNet presenta una mayor degradación en la esclera, ya que su valor medio es significativamente menor que el del target y el de Pix2Pix. En la capa interna o coroides, el target tiene un valor medio de  $0,24 \pm 0,12$ . Pix2Pix alcanza un valor medio de  $0,22 \pm 0,12$ , mientras que CCARNet logra  $0,12 \pm 0,08$ . Similar al caso de la esclera, CCARNet degrada más la intensidad de la estructura real en comparación con Pix2Pix. Esta degradación

observada en CCARNet podría estar relacionada con el funcionamiento de sus gate connections. Es posible que estas conexiones estén atenuando excesivamente las características de bajo nivel, que contienen los detalles finos de las estructuras reales, o que los pesos asociados a las gate connections no estén convergiendo a valores óptimos durante el entrenamiento, resultando en una combinación inadecuada de características. Esto podría llevar a que la red priorice la eliminación del artefacto espejo sobre la preservación de los detalles finos de las estructuras reales.

En el semi-eje positivo, las comparaciones evalúan la atenuación de la copia generada por el artefacto espejo. Para la esclera, el target presenta un valor medio en la magnitud normalizada de  $0,46 \pm 0,20$ . En este caso, Pix2Pix y CCARNet logran valores medios de  $0,16 \pm 0,07$  y  $0,14 \pm 0,06$ , respectivamente. Esto corresponde a relaciones de atenuación del  $0,35 \pm 0,21$  % para Pix2Pix y del  $0,30 \pm 0,19$  % para CCARNet. Aquí, un menor valor indica una mayor atenuación de la copia, lo que posiciona a CCARNet como la red más efectiva para esta capa. En la coroides, el target tiene un valor medio de  $0,24 \pm 0,12$ . Pix2Pix alcanza  $0,09 \pm 0,04$ , mientras que CCARNet logra  $0,11 \pm 0,09$ . Esto corresponde a relaciones de atenuación del 37,5 % para Pix2Pix y del 45,8 % para CCARNet. En este caso, Pix2Pix atenúa mejor la copia en comparación con CCARNet.

Este análisis destaca dos comportamientos clave. En el semi-eje negativo, CCARNet degrada más las intensidades de las estructuras originales en comparación con la Pix2Pix, lo que es visualmente evidente en la distribución de frecuencias y los picos de magnitud. Sin embargo, en el semi-eje positivo, CCARNet logra una mejor atenuación de la copia en la esclera, mientras que Pix2Pix muestra un desempeño superior en la atenuación de la copia en la coroides. Estos resultados subrayan la necesidad de un balance entre la preservación de las estructuras reales y la atenuación efectiva del artefacto espejo.

La Tabla 4.4 presenta los valores promedio de las métricas calculadas para las diferentes soluciones evaluadas. A continuación, se discuten los resultados y sus posibles implicaciones.

En términos de la métrica SSIM, la solución híbrida con CCARNet ( $0,36 \pm 0,01$ ) supera a la solución híbrida con Pix2Pix ( $0,34 \pm 0,01$ ) y a las salidas directas de Pix2Pix ( $0,33 \pm 0,03$ ) y CCARNet ( $0,27 \pm 0,02$ ). Este resultado es interesante, ya que en el análisis previo se observó que la reconstrucción compuesta con CCARNet tiende a degradar la señal. Una posible explicación radica en que CCARNet reconstruye mejor la esclera, una estructura más amplia y homogénea, que típicamente tiene un mayor peso en la métrica SSIM. Según la literatura, la esclera tiene un grosor promedio de  $[400 - 500]\mu m$  en regiones centrales, mientras que la coroides es significativamente más delgada, con un grosor de  $[150 - 250]\mu m$ . Estas diferencias de tamaño podrían justificar el mejor desempeño de CCARNet en esta métrica estructural.

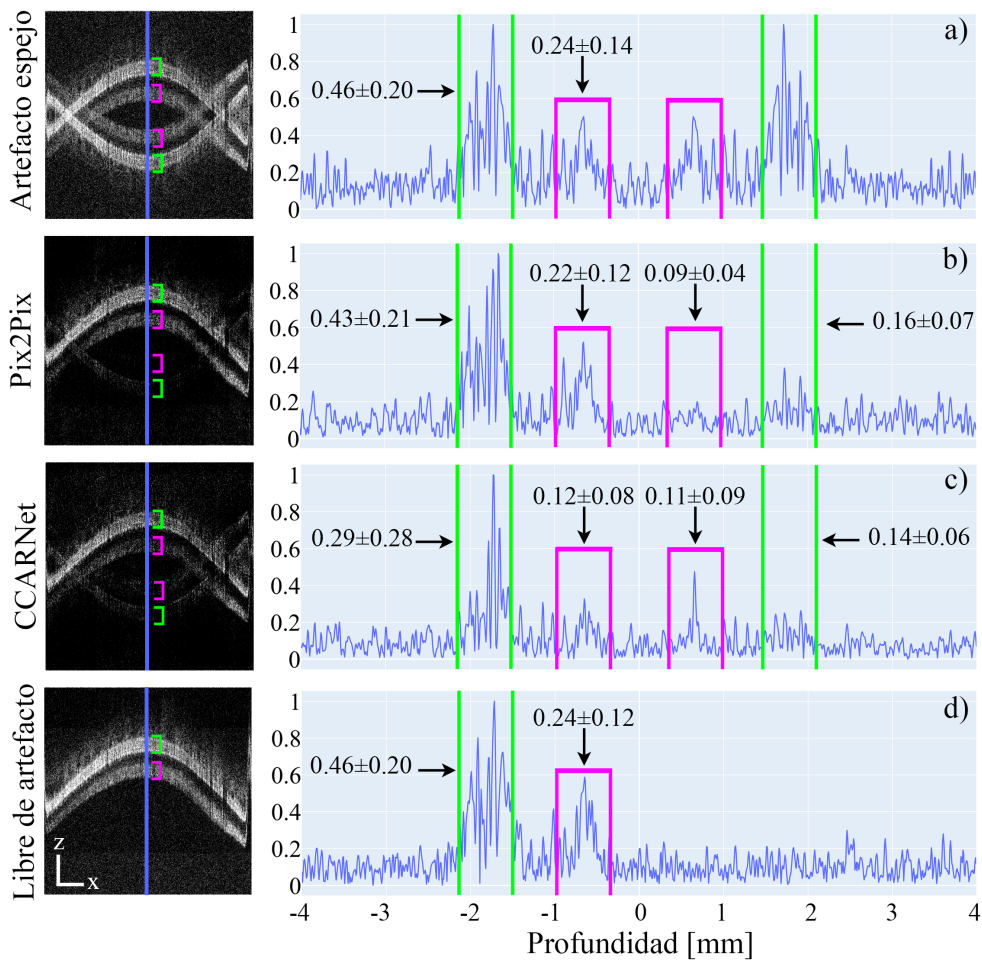


Figura 4.16: Comparación cualitativa y cuantitativa de la supresión del artefacto espejo. La columna izquierda muestra las imágenes reconstruidas en cuatro condiciones: a) imagen con artefacto de espejo, b) reconstrucción del campo complejo con Pix2Pix, c) reconstrucción empleando CCARNet y d) imagen libre de artefacto. La columna derecha presenta los perfiles de intensidad correspondientes a las regiones de interés marcadas en las imágenes.

Por otro lado, la métrica PSNR muestra que las soluciones híbridas con Pix2Pix y CCARNet logran los mejores resultados, con valores de  $21,21 \pm 0,50\text{dB}$  y  $21,01 \pm 0,55\text{dB}$ , respectivamente. Estos valores superan ampliamente los obtenidos por las salidas directas de las redes Pix2Pix ( $11,60 \pm 1,16\text{dB}$ ) y CCARNet ( $11,82 \pm 1,01\text{dB}$ ). Además, los valores de PSNR alcanzados por las soluciones híbridas son consistentes con los reportados en la literatura de OCT, donde se sitúan típicamente entre  $20 - 25\text{dB}$  para reconstrucciones de buena calidad. Este comportamiento puede explicarse por el uso de la magnitud de la entrada en las soluciones híbridas, que preserva mejor la calidad de la señal de intensidad, compensando las limitaciones de las redes en la reconstrucción directa del campo complejo.

En cuanto a la métrica de diferencia de coseno del histograma, las salidas directas de las redes presentan los mejores resultados,  $0,76 \pm 0,06$  para Pix2Pix y  $0,58 \pm 0,07$  para CCARNet en comparación con las soluciones híbridas,  $0,12 \pm 0,05$  para ambas recordemos que las soluciones híbridas son los tomogramas reconstruidos a partir de la amplitud original con artefactos espejo y la información de fase obtenida a través de los modelos entrenados. Este resultado se debe a que las salidas directas mantienen una distribución estadística más cercana a los datos originales. En contraste, las soluciones híbridas combinan información de diferentes fuentes, magnitud de entrada y fase de salida, lo que altera la distribución y penaliza esta métrica.

En términos generales, Pix2Pix y su solución híbrida destacan por sus mejores resultados en comparación con los de CCARNet. La solución híbrida con Pix2Pix obtiene los mejores resultados en dos de las cuatro métricas principales, mostrando un equilibrio entre la preservación estructural y la calidad de la señal. Esto sugiere que Pix2Pix es una opción más consistente en contextos donde se requiere una combinación de fidelidad visual y robustez estadística.

Tabla 4.4: Valores promedio de las métricas para los modelos evaluados

Métrica	Híbrido CCARNet	Híbrido Pix2Pix	Pix2Pix	CCARNet
SSIM	$0,36 \pm 0,01$	$0,34 \pm 0,01$	$0,33 \pm 0,03$	$0,27 \pm 0,02$
MSE	$0,01 \pm 0,01$	$0,01 \pm 0,01$	$0,07 \pm 0,02$	$0,07 \pm 0,02$
PSNR	$21,01 \pm 0,55$	$21,21 \pm 0,50$	$11,60 \pm 1,16$	$11,82 \pm 1,01$
Hist. Diff.	$0,12 \pm 0,05$	$0,12 \pm 0,05$	$0,76 \pm 0,06$	$0,58 \pm 0,07$

La Figura 4.17 presenta la evolución de las métricas SSIM y MSE a lo largo de las B-scans del volumen utilizando el método híbrido con Pix2Pix. Ambas métricas muestran valores promedio estables, con un SSIM de  $0,344 \pm 0,013$  y un MSE de  $0,008 \pm 0,001\text{dB}^2$ , lo que refleja la consistencia del método frente a variaciones en el grado de solapamiento. Sin embargo, se observan pequeñas fluctuaciones en estas métricas que parecen estar relacionadas con la distribución del solapamiento en el volumen.

En la primera región del volumen, B-scans 0 al 200, donde el solapamiento se encuentra en la zona central de las estructuras, el SSIM y el MSE permanecen cercanos a sus valores promedio, mostrando una reconstrucción más estable. Esto sugiere que el método es consistente en el solapamiento central, donde las características estructurales están mejor definidas. En contraste, en la región intermedia del volumen desde la B-scans 201 a la 400, donde el solapamiento se desplaza hacia los bordes, se observan mayores fluctuaciones, especialmente en el MSE, con picos que alcanzan valores más altos. Este comportamiento puede atribuirse a que en esta región las estructuras que se reconstruyen tienen un mayor nivel de detalle y complejidad en comparación con las regiones con menor solapamiento. La reconstrucción de detalles finos es más susceptible a errores, lo que se traduce en un aumento en el MSE.

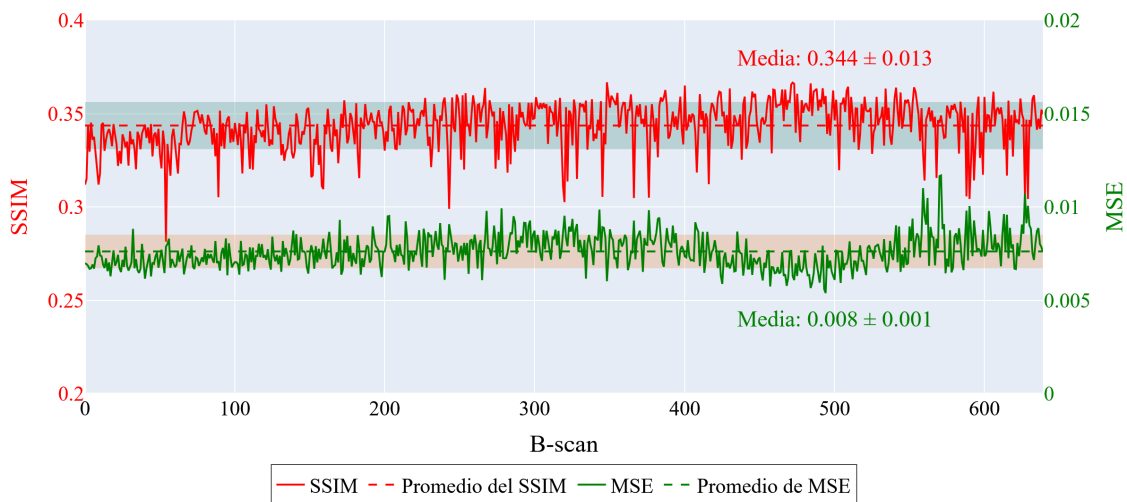


Figura 4.17: Variación del SSIM en rojo y del MSE en verde a lo largo de las B-scans del volumen. Se muestra la media y desviación estándar para la solución híbrida con Pix2Pix.

Finalmente, en la última región del volumen desde la B-scans 401 a la 600, donde el solapamiento desaparece por completo, ambas métricas se estabilizan nuevamente, indicando que la ausencia de solapamiento facilita una reconstrucción más consistente.

Además, se aprecia visualmente una correlación inversa entre las métricas SSIM y MSE a lo largo del volumen. En general, las caídas en el SSIM coinciden con aumentos en el MSE, lo cual es consistente con su naturaleza complementaria: una mayor similitud estructural está asociada con un menor error cuadrático medio, y viceversa. Este comportamiento refuerza la importancia de analizar ambas métricas de manera conjunta para evaluar la calidad de la reconstrucción.

La Figura 4.18 muestra la evolución de las métricas PSNR y diferencia de histograma a lo largo de las B-scans del volumen. Aunque el PSNR tiene una media global de  $21,211 \pm 0,498\text{dB}$ , se observa un patrón oscilante que se aleja de la media, especialmente entre las B-scans 400 y 550, donde la métrica presenta un aumento significativo. Este comportamiento sugiere una mejora en la reconstrucción de la intensidad en estas regiones, posiblemente asociada a una menor complejidad estructural o la ausencia de solapamiento significativo. Sin embargo, esta oscilación indica que la calidad de reconstrucción en términos de intensidad no es completamente consistente en estas áreas.

En contraste, la métrica de diferencia de histograma muestra una mayor estabilidad a lo largo de todo el volumen, con una media de  $0,120 \pm 0,051$  y pequeñas fluctuaciones alrededor de este valor. Esta estabilidad sugiere que, a pesar de las variaciones en la intensidad reconstruida, como lo indica el PSNR, los patrones estadísticos de las intensidades se mantienen consistentes en comparación con el

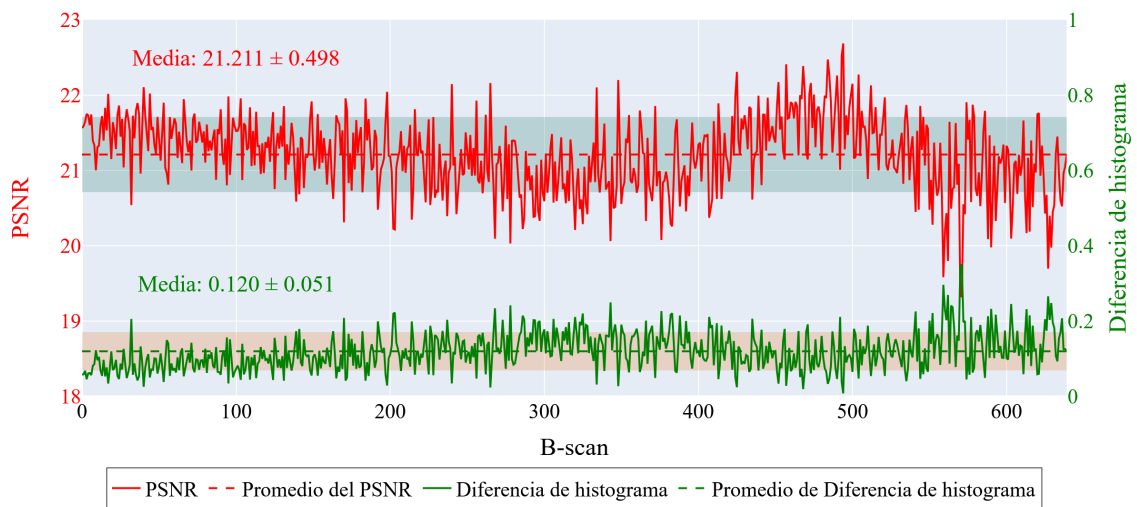


Figura 4.18: Métricas de la evolución del PSNR en rojo y de la diferencia de histograma en verde a lo largo de las B-scans del volumen para la solución híbrida con Pix2Pix. Se incluye la media y desviación estándar de ambas métricas.

objetivo.

La discrepancia entre el comportamiento del PSNR y la diferencia de histograma resalta que estas métricas miden aspectos complementarios de la reconstrucción. Mientras que el PSNR captura la calidad de la señal reconstruida en términos de intensidad, la diferencia de histograma evalúa la similitud en la distribución estadística de la intensidad.

# CONCLUSIONES Y TRABAJO FUTURO

### 5.1. Conclusiones basadas en los objetivos

A partir de los objetivos planteados al inicio de este trabajo, se pueden establecer las siguientes conclusiones:

- En cuanto al primer objetivo, se logró construir una base de datos de tomogramas simulados que integran tanto artefactos espejo como los efectos del submuestreo lateral en el dominio de Fourier, mediante la implementación del modelo directo. Esta base de datos sirvió como referencia para el entrenamiento de los modelos de aprendizaje profundo. La calidad y diversidad de los datos generados permitieron analizar el rendimiento de los modelos en condiciones controladas y reproducibles, sentando una base sólida para la investigación.
- Respecto al segundo objetivo, se implementaron y entrenaron modelos de aprendizaje profundo independientes para la mitigación de artefactos espejo y la reducción del *aliasing* producido por el submuestreo lateral en imágenes OCT. Aunque el objetivo inicial fue entrenar las redes usando imágenes de intensidad de B-scans para el problema de los artefactos espejo y planos *en-face* para el problema de submuestreo lateral, todos los entrenamientos se realizaron utilizando como entrada los tomogramas complejos. En el caso de los artefactos espejo, las redes fueron entrenadas para reconstruir un tomograma complejo sin la copia, mientras que en el caso del submuestreo lateral, se buscó reconstruir un tomograma con muestreo Nyquist a partir de un tomograma sub-Nyquist. Este enfoque permitió preservar información de fase esencial para procesos de post-procesamiento, como la corrección de movimiento de cuerpo rígido y la corrección de *phase jitter*.
- En relación al tercer objetivo, se analizó la generalización de los modelos de aprendizaje profundo al aplicarlos a muestras biológicas experimentales,

incluyendo aquellas con submuestreo lateral aplicado. En el problema de los tomogramas submuestreados lateralmente, se observó que los modelos representan una alternativa efectiva para reducir el tiempo de adquisición en OCT, evidenciándose la recuperación en la continuidad de estructuras que se perdían debido al submuestreo. Estos resultados destacan el potencial de estas técnicas para mejorar la calidad de las imágenes mientras se optimiza el proceso de adquisición. Por otro lado, en el problema de los artefactos espejo, aunque los modelos lograron eliminar el artefacto, se observó una degradación significativa de la señal, particularmente en la fase. La red tendió a asignar valores constantes en los píxeles correspondientes a la copia, en lugar de generalizar el ruido de fondo circundante. Esto sugiere que el problema de los artefactos espejo podría no ser generalizable para cualquier estructura o sistema OCT, y podría requerir estrategias específicas para determinadas técnicas y áreas médicas.

- Finalmente, en cuanto al cuarto objetivo, se evaluó cuantitativamente el rendimiento de los modelos de aprendizaje profundo implementados para la mitigación de artefactos espejo y la reducción del *aliasing*, utilizando métricas de evaluación objetivas para comparar los resultados obtenidos con respecto a los datos de referencia (*ground truth*). El análisis cuantitativo basado en estas métricas evidenció que, en el caso de los tomogramas submuestreados lateralmente, los modelos lograron recuperar la continuidad estructural de las imágenes con resultados positivos y consistentes, lo que valida su potencial para mejorar el proceso de adquisición en OCT. Por otro lado, en el problema de los artefactos espejo, aunque cualitativamente se confirmó la eliminación del artefacto, las métricas revelaron una degradación en la calidad de la señal reconstruida, especialmente en la fase, con valores uniformes que afectaron la similitud estructural en regiones específicas. Este contraste resalta la necesidad de enfoques más específicos para abordar cada problema de manera independiente y adecuada.

## 5.2. Conclusiones basadas en los resultados

### 5.2.1. Reducción de tiempo en la adquisición de tomogramas

Los resultados obtenidos a lo largo de este capítulo confirman que el método propuesto, basado en aprendizaje profundo, es una herramienta eficaz para reconstruir tomogramas subsampleados lateralmente. El modelo cGAN, al utilizar información incompleta del volumen original, logra generar estructuras congruentes y coherentes tanto en términos de intensidad como de fase. Las métricas evaluadas, como SSIM, MSE y PSNR, demuestran que el modelo es capaz de preservar la continuidad estructural y minimizar las diferencias en las distribuciones de intensidad, incluso en

regiones densas en tejido y zonas con ruido fotónico predominante. Esto valida la capacidad del modelo para abordar escenarios donde la adquisición de datos completa no es viable debido a limitaciones de tiempo o recursos.

Además, este enfoque permite reducir significativamente el tiempo de adquisición, lo que resulta en una mejora importante para aplicaciones clínicas, donde la velocidad de escaneo es crucial para minimizar los efectos de movimientos involuntarios del paciente durante el procedimiento. Al disminuir el tiempo necesario para adquirir los datos y mantener la fidelidad estructural del tomograma, se reduce el efecto de distorsiones y artefactos generados por dichos movimientos. Por tanto, el uso de redes neuronales profundas, como el modelo cGAN, no solo optimiza la eficiencia de adquisición de imágenes, sino que también mejora la calidad y precisión de los resultados, lo que constituye una contribución valiosa para la tomografía de coherencia óptica en entornos clínicos.

### 5.2.2. Mitigación del artefacto espejo

Durante este trabajo se evaluaron diferentes métodos para la supresión de artefactos espejo en OCT, con un enfoque en redes generativas adversarias y la implementación de un método híbrido. A partir de los análisis realizados, se extraen las siguientes conclusiones clave:

El mejor desempeño en las métricas de calidad, como SSIM y PSNR, fue obtenido con el método híbrido basado en Pix2Pix. Sin embargo, aunque los resultados muestran una mejora relativa frente a las salidas directas de las redes, en términos absolutos, la similitud estructural sigue siendo limitada. Este desempeño subraya la necesidad de enfoques complementarios que permitan mejorar la reconstrucción estructural y garantizar una representación más fiel del tejido biológico en los tomogramas de OCT.

El análisis a lo largo de las B-scans del volumen demuestra que el grado de solapamiento de las estructuras tiene un impacto significativo en el desempeño de las reconstrucciones. En el caso específico del tejido de la fovea, se observó que el método híbrido responde de manera más estable en regiones donde el solapamiento es central, mientras que en zonas con solapamiento en los bordes, las métricas tienden a fluctuar más, particularmente el PSNR. Esto resalta que el desempeño del método depende en gran medida de las características del tejido y de la distribución del solapamiento, limitando su generalización a otros escenarios. Esta conclusión difiere del problema de los tomogramas submuestreados lateralmente, donde el uso de datos sintéticos permitió obtener redes que respondían de manera positiva a estructuras no vistas durante el entrenamiento.

Una contribución importante de este trabajo es la introducción del método híbrido,

que combina la magnitud del tomograma de entrada con la fase generada por la red. Este enfoque fue implementado para mitigar la degradación estructural causada por las redes generativas. Sin embargo, esta solución no pertenece completamente a la metodología de las redes neuronales, ya que requiere como insumo adicional, la magnitud de la señal de entrada para reconstruir el tomograma final. Los resultados mostraron que este método logró una diferencia del 65 % entre la copia y la estructura real en términos de intensidad, preservando la estructura original y mitigando de manera visible el artefacto espejo. Este comportamiento puede ser relevante en la interpretación médica, ya que facilita la identificación del tejido biológico al reducir el remanente del artefacto de forma perceptible.

En conclusión, los resultados obtenidos sugieren que el método híbrido que usa como parámetros la amplitud del tomograma original y la fase generada por el modelo, podría beneficiarse de estrategias adicionales para optimizar la reconstrucción, como la integración de algoritmos inspirados en técnicas como DEFR. No obstante, estas mejoras potenciales deben equilibrarse cuidadosamente con los tiempos de inferencia en postprocesamiento, especialmente en aplicaciones médicas donde la rapidez en el análisis de los datos es fundamental. Finalmente, los resultados alcanzados en este trabajo destacan la necesidad de acotar el problema a sistemas de OCT específicos y aplicaciones médicas concretas, dado que la generalización de las soluciones propuestas no fue alcanzada, como sí ocurrió en el problema de los tomogramas submuestreados lateralmente.

### 5.3. Trabajo futuro

Para abordar el problema del artefacto espejo en OCT, se propone acotar el problema a un tejido biológico en particular, con el objetivo de “especializar” el modelo Pix2Pix hacia el reconocimiento de estructuras específicas. Una posible aplicación médica podría enfocarse en tejidos donde el tamaño de la estructura sea lo suficientemente grande y se requiera una mayor profundidad. Aunque en este trabajo se utilizó la fovea como fuente de estudio, con un tamaño aproximado de 3.5 mm que permite separar la estructura de la frecuencia cero y conservar su totalidad, este enfoque podría no ser aplicable a otras estructuras. Por ejemplo, en el caso de la córnea, cuya profundidad alcanza entre 10 mm y 11 mm, sería más difícil separar la estructura del centro y conservarla en su totalidad, ya que esta podría solaparse con su copia conjugada. Este escenario representa un desafío significativo para la técnica propuesta y ofrece una oportunidad para ampliar su aplicabilidad.

En futuras investigaciones, se podrían explorar otras arquitecturas generativas que han mostrado resultados prometedores en tareas de generación y reconstrucción de imágenes. Los modelos de difusión, por ejemplo, generan imágenes de alta calidad

a través de un proceso iterativo de eliminación de ruido [128], lo que podría ser beneficioso para reconstruir la señal OCT libre de artefactos espejo al aprender la distribución subyacente de los datos complejos. Asimismo, los modelos basados en mecanismos de atención, como los Transformers, han demostrado ser efectivos para capturar dependencias a largo alcance en los datos [129], lo cual podría ser útil para modelar la relación entre la estructura real y su complejo conjugado, permitiendo una supresión más inteligente y contextualizada del artefacto.

Adicionalmente, se propone un refinamiento del método híbrido desarrollado en este trabajo. Una posibilidad sería integrar técnicas de filtrado espacial o selectivo, similares a las utilizadas en algoritmos como DEFR, para mitigar aún más la copia remanente, en caso de que esta persista incluso después de acotar el problema a un tejido específico. Este refinamiento podría mejorar la calidad de la reconstrucción y proporcionar una solución más robusta y adaptable para aplicaciones médicas que requieran un análisis preciso de estructuras complejas. Sin embargo, es importante considerar el impacto de estas modificaciones en los tiempos de inferencia, especialmente en escenarios donde el procesamiento en tiempo real es crítico.

En cuanto al trabajo futuro relacionado con los tomogramas submuestreados lateralmente, se propone enfocar la metodología en el área de la cardiología, donde de forma convencional se da el submuestreo en el eje axial. Esto se debe a que las tomografías de las arterias coronarias dependen de una sonda o catéter que realiza un barrido circular a través de ellas, ocasionando submuestreo debido a la necesidad de un menor tiempo de adquisición, dado la naturaleza invasiva del procedimiento. Este enfoque no solo le otorgaría un contexto médico específico a la metodología planteada en este trabajo, sino que también permitiría validar su utilidad en aplicaciones prácticas donde las limitaciones de adquisición son críticas. Además, el uso de redes generativas adversarias podría explorarse en combinación con técnicas de regularización e interpolación que permitan reforzar la continuidad estructural y preservar detalles relevantes en las imágenes reconstruidas. Otro aspecto relevante sería investigar la incorporación de modelos de dominio específico que aprovechen las características únicas de las imágenes de arterias coronarias, lo cual podría mejorar aún más la generalización y el desempeño de las redes en este tipo de datos.

## 5.4. Productos Derivados de la Maestría

El desarrollo de esta maestría ha resultado en diversas producciones académicas. Entre ellas, se presentó una ponencia en el congreso internacional *“10th International Symposium “Optics & its applications”*”, donde se compartieron avances de la investigación. Asimismo, se realizó una pasantía en el centro de investigación Vicomtech en España - País Vasco. Durante esta estancia, se profundizó en la generación de

señales de audio digitales mediante la implementación de modelos avanzados de inteligencia artificial. Esta experiencia permitió no solo consolidar los fundamentos teóricos del procesamiento digital de señales, sino también adquirir habilidades prácticas en el diseño y entrenamiento de redes neuronales profundas para aplicaciones de audio. Se exploraron diversas arquitecturas de modelos generativos, como redes adversarias generativas y autoencoders variacionales, aplicándolas a la síntesis de audio y la manipulación de características espectrales. Este trabajo contribuyó a afianzar la capacidad para abordar problemas complejos de procesamiento de señales utilizando técnicas de aprendizaje profundo, y a comprender la importancia de la colaboración interdisciplinaria en la investigación. Además, se encuentran en proceso de elaboración dos artículos científicos que profundizan en los hallazgos de la investigación: *“Generative model-based method for the reconstruction of optical coherence tomography images with lateral subsampling”* y *“Suppression of Mirror Artifacts in Optical Coherence Tomography Using Generative Adversarial Networks”*. Estos artículos, actualmente en construcción, tienen como objetivo contribuir al avance del conocimiento en el campo de la tomografía de coherencia óptica y su aplicación en diversas áreas.

---

# REFERENCIAS

- [1] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, and C. A. Puliafito, “Optical coherence tomography,” *Science (New York, N.Y.)*, vol. 254, pp. 1178–1181, Nov. 1991. Place: United States.
- [2] E. A. Swanson, J. A. Izatt, M. R. Hee, D. Huang, C. P. Lin, J. S. Schuman, C. A. Puliafito, and J. G. Fujimoto, “In vivo retinal imaging by optical coherence tomography,” *Opt. Lett.*, vol. 18, pp. 1864–1866, Nov 1993.
- [3] W. Drexler, Y. Chen, A. D. Aguirre, B. Považay, A. Unterhuber, and J. G. Fujimoto, *Ultrahigh Resolution Optical Coherence Tomography*. Society of Photo-Optical Instrumentation Engineers, 2015.
- [4] S. Aumann, S. Donner, J. Fischer, and F. Müller, *Optical Coherence Tomography (OCT): Principle and Technical Realization*, pp. 59–85. Cham: Springer International Publishing, 2019.
- [5] W. Wieser, B. R. Biedermann, T. Klein, C. M. Eigenwillig, and R. Huber, “Multi-megahertz oct: High quality 3d imaging at 20 million a-scans and 4.5 voxels per second,” *Opt. Express*, vol. 18, pp. 14685–14704, Jul 2010.
- [6] A. Fercher, C. Hitzenberger, G. Kamp, and S. El-Zaiat, “Measurement of intraocular distances by backscattering spectral interferometry,” *Optics Communications*, vol. 117, no. 1, pp. 43–48, 1995.
- [7] E. J. McDowell, A. K. Ellerbee, M. A. Choma, B. E. Applegate, and J. A. Izatt, “Spectral domain phase microscopy for local measurements of cytoskeletal rheology in single cells,” *Journal of biomedical optics*, vol. 12, p. 044008, Aug. 2007. Place: United States.
- [8] W. Drexler and J. Fujimoto, *Optical Coherence Tomography: Technology and Applications*. Biological and Medical Physics, Biomedical Engineering, Springer Berlin Heidelberg, 2008.
- [9] S. Lopera-Ruiz, “Computational aberration correction in optical coherence tomography with phase-unstable systems,” *EAFIT*, 2020.

- [10] R. de Kinkelder, J. Kalkman, D. J. Faber, O. Schraa, P. H. Kok, F. D. Verbraak, and T. G. van Leeuwen, “Heartbeat-induced axial motion artifacts in optical coherence tomography measurements of the retina,” *Investigative ophthalmology & visual science*, vol. 52, no. 6, pp. 3908–3913, 2011.
- [11] T. Klein and R. Huber, “High-speed oct light sources and systems [invited],” *Biomed. Opt. Express*, vol. 8, pp. 828–859, Feb 2017.
- [12] T. S. Kim, J. Joo, I. Shin, P. Shin, W. J. Kang, B. J. Vakoc, and W.-Y. Oh, “9.4 mhz a-line rate optical coherence tomography at 1300 nm using a wavelength-swept laser based on stretched-pulse active mode-locking,” *Scientific Reports*, vol. 10, no. 1, p. 9328, 2020.
- [13] M. Draelos, C. Viehland, R. P. McNabb, A. N. Kuo, and J. A. Izatt, “Adaptive point-scan imaging beyond the frame rate–resolution limit with scene-reactive scan trajectories,” *Optica*, vol. 9, pp. 1276–1288, Nov 2022.
- [14] V.-F. Duma, K. sung Lee, P. Meemon, and J. P. Rolland, “Experimental investigations of the scanning functions of galvanometer-based scanners with applications in oct,” *Appl. Opt.*, vol. 50, pp. 5735–5749, Oct 2011.
- [15] N. D. Shemonski, F. A. South, Y.-Z. Liu, S. G. Adie, P. Scott Carney, and S. A. Boppart, “Computational high-resolution optical imaging of the living human retina,” *Nature Photonics*, vol. 9, no. 7, pp. 440–443, 2015.
- [16] C. Mnerie, S. Preitl, and V.-F. Duma, “Control architectures of galvanometer-based scanners for an increased precision and a faster response,” in *Fifth International Conference on Lasers in Medicine: Biotechnologies Integrated in Daily Medicine* (C. Todea, A. G. Podoleanu, and V.-F. Duma, eds.), vol. 8925, p. 892500, International Society for Optics and Photonics, SPIE, 2014.
- [17] J. Fujimoto, “Fujimoto, j.g. optical coherence tomography for ultrahigh resolution in vivo imaging. nat. biotechnol. 21, 1361-1367,” *Nature biotechnology*, vol. 21, pp. 1361–7, 12 2003.
- [18] J. W. Goodman, “Introduction to fourier optics,” *Introduction to Fourier optics, 3rd ed., by JW Goodman. Englewood, CO: Roberts & Co. Publishers, 2005*, vol. 1, 2005.
- [19] K. Irsch, “Optical principles of oct,” in *Albert and Jakobiec’s Principles and Practice of Ophthalmology*, pp. 1083–1096, Springer, 2022.
- [20] A. V. Oppenheim, *Discrete-time signal processing*. Pearson Education India, 1999.

- [21] J. H. McClellan, R. Schafer, and M. Yoder, *Dsp first*. Prentice Hall Press, 2015.
- [22] C. Luo and J. H. McClellan, “Discrete random sampling theory,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5430–5434, IEEE, 2013.
- [23] E. Hecht, *Optics*. Pearson Education India, 2012.
- [24] W. Wieser, B. R. Biedermann, T. Klein, C. M. Eigenwillig, and R. Huber, “Multi-megahertz oct: High quality 3d imaging at 20 million a-scans and 4.5 gvoxels per second,” *Opt. Express*, vol. 18, pp. 14685–14704, Jul 2010.
- [25] T. Yu, “Applications of deep learning for ophthalmic oct,” Master’s thesis, Simon Fraser University, 2022.
- [26] J. H. O. Ocampo, “Método de aprendizaje profundo para la reconstrucción de imágenes de tomografía de coherencia Óptica con submuestreo lateral,” trabajo de grado, Universidad EAFIT, Medellín, Colombia, Julio 2021.
- [27] C.-M. Huang, E. Wijanto, and H.-C. Cheng, “Applying a pix2pix generative adversarial network to a fourier-domain optical coherence tomography system for artifact elimination,” *IEEE Access*, vol. 9, pp. 103311–103324, 2021.
- [28] V. Bellemo, L. Schmetterer, and X. Liu, “Complex conjugate artifact removal in FD-OCT using generative adversarial network,” in *Optical Coherence Tomography and Coherence Domain Optical Methods in Biomedicine XXVII* (J. A. Izatt and J. G. Fujimoto, eds.), vol. 12367, p. 123670U, International Society for Optics and Photonics, SPIE, 2023.
- [29] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. F. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, “Optical coherence tomography,” *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.
- [30] J. G. Fujimoto, “Optical coherence tomography in ophthalmology,” *Progress in retinal and eye research*, vol. 23, no. 1, pp. 7–31, 2004.
- [31] A. F. Fercher, “Optical coherence tomography,” *Journal of biomedical optics*, vol. 1, no. 2, pp. 157–173, 1996.
- [32] A. Maheshwari, M. A. Choma, and J. A. Izatt, “Heterodyne swept-source optical coherence tomography for complete complex conjugate ambiguity removal,” *Journal of Biomedical Optics*, vol. 10, no. 6, p. 064005, 2005.
- [33] A. F. Fercher, “Optical coherence tomography,” *Journal of biomedical optics*, vol. 1, no. 2, pp. 157–173, 1996.

- [34] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
- [35] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [36] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, "Optical coherence tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.
- [37] T. Yonetsu, B. E. Bouma, K. Kato, J. G. Fujimoto, and I.-K. Jang, "Optical coherence tomography," *Circulation Journal*, vol. 77, no. 8, pp. 1933–1940, 2013.
- [38] C. Cuartas-Vélez, R. Restrepo, B. E. Bouma, and N. Uribe-Patarroyo, "Volumetric non-local-means based speckle reduction for optical coherence tomography," *Biomedical optics express*, vol. 9, pp. 3354–3372, July 2018. Place: United States.
- [39] R. Leitgeb, C. K. Hitzenberger, and A. F. Fercher, "Performance of fourier domain vs. time domain optical coherence tomography," *Opt. Express*, vol. 11, pp. 889–894, Apr 2003.
- [40] M. A. Choma, M. V. Sarunic, C. Yang, and J. A. Izatt, "Sensitivity advantage of swept source and fourier domain optical coherence tomography," *Opt. Express*, vol. 11, pp. 2183–2189, Sep 2003.
- [41] J. W. Zahid Yaqoob and C. Yang, "Spectral domain optical coherence tomography: a better oct imaging strategy," *BioTechniques*, vol. 39, no. sup6, pp. S6–S13, 2005. PMID: 20158503.
- [42] J. W. Goodman, *Statistical optics*. John Wiley & Sons, 2015.
- [43] H. H. Hopkins, "Applications of coherence theory in microscopy and interferometry\*," *J. Opt. Soc. Am.*, vol. 47, pp. 508–526, Jun 1957.
- [44] C. Wheeler, R. Ramsier, and P. Henriksen, "An investigation of the temporal coherence length of light," *European journal of physics*, vol. 24, no. 4, p. 443, 2003.
- [45] W. Gao, "Effects of temporal and spatial coherence on resolution in full-field optical coherence tomography," *Journal of Modern Optics*, vol. 62, no. 21, pp. 1764–1774, 2015.
- [46] D. Malacara, "Optical shop testing," *Wiley Interscience*, 2007.

- [47] J. A. Izatt and M. A. Choma, “Theory of optical coherence tomography,” in *Optical Coherence Tomography: Technology and Applications*, pp. 47–72, Springer, 2008.
- [48] C. C. Vélez, S. R. Lopera, N. R. U. Patarroyo, and R. Restrepo, “Lab-made accessible full-field optical coherence tomography imaging system,” *Óptica pura y aplicada*, vol. 52, no. 3, p. 3, 2019.
- [49] F. Lavinsky and D. Lavinsky, “Novel perspectives on swept-source optical coherence tomography,” *International journal of retina and vitreous*, vol. 2, pp. 1–11, 2016.
- [50] T. S. Ralston, D. L. Marks, P. S. Carney, and S. A. Boppart, “Inverse scattering for optical coherence tomography,” *J. Opt. Soc. Am. A*, vol. 23, pp. 1027–1037, May 2006.
- [51] T. S. Ralston, D. L. Marks, S. A. Boppart, and P. S. Carney, “Non-paraxial solution to inverse scattering in optical coherence tomography,” in *Frontiers in Optics*, p. FMI3, Optica Publishing Group, 2006.
- [52] B. J. Davis, S. C. Schlachter, D. L. Marks, T. S. Ralston, S. A. Boppart, and P. S. Carney, “Nonparaxial vector-field modeling of optical coherence tomography and interferometric synthetic aperture microscopy,” *J. Opt. Soc. Am. A*, vol. 24, pp. 2527–2542, Sep 2007.
- [53] D. L. Marks, T. S. Ralston, P. S. Carney, and S. A. Boppart, “Inverse scattering for rotationally scanned optical coherence tomography,” *J. Opt. Soc. Am. A*, vol. 23, pp. 2433–2439, Oct 2006.
- [54] J. Fujimoto, “Fujimoto, j.g. optical coherence tomography for ultrahigh resolution in vivo imaging. nat. biotechnol. 21, 1361-1367,” *Nature biotechnology*, vol. 21, pp. 1361–7, 12 2003.
- [55] N. Nassif, B. Cense, B. H. Park, S. H. Yun, T. C. Chen, B. E. Bouma, G. J. Tearney, and J. F. de Boer, “In vivo human retinal imaging by ultrahigh-speed spectral domain optical coherence tomography,” *Opt. Lett.*, vol. 29, pp. 480–482, Mar 2004.
- [56] D. M. Sampson, A. M. Dubis, F. K. Chen, R. J. Zawadzki, and D. D. Sampson, “Towards standardizing retinal optical coherence tomography angiography: a review,” *Light: science & applications*, vol. 11, no. 1, p. 63, 2022.
- [57] Y. Wang, A. Warter, M. Cavichini-Cordeiro, W. R. Freeman, D.-U. G. Bartsch, T. Q. Nguyen, and C. An, “Learning to correct axial motion in oct for 3d

- retinal imaging,” in *2021 IEEE International Conference on Image Processing (ICIP)*, pp. 126–130, IEEE, 2021.
- [58] W. Wieser, B. R. Biedermann, T. Klein, C. M. Eigenwillig, and R. Huber, “Multi-megahertz oct: High quality 3d imaging at 20 million a-scans and 4.5 voxels per second,” *Optics express*, vol. 18, no. 14, pp. 14685–14704, 2010.
- [59] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, 2014.
- [60] Y. Zhang, T. Liu, M. Singh, E. Çetintaş, Y. Luo, Y. Rivenson, K. V. Larin, and A. Ozcan, “Neural network-based image reconstruction in swept-source optical coherence tomography using undersampled spectral data,” *Light: Science & Applications*, vol. 10, no. 1, p. 155, 2021.
- [61] A. DiSpirito, D. Li, T. Vu, M. Chen, D. Zhang, J. Luo, R. Horstmeyer, and J. Yao, “Reconstructing undersampled photoacoustic microscopy images using deep learning,” *IEEE transactions on medical imaging*, vol. 40, no. 2, pp. 562–570, 2020.
- [62] D. G. Fischer and B. Ovrzyn, “Phase-shifting optical coherence tomography,” in *Coherence Domain Optical Methods in Biomedical Science and Clinical Applications V* (V. V. Tuchin, J. A. Izatt, and J. G. Fujimoto, eds.), vol. 4251, pp. 97 – 101, International Society for Optics and Photonics, SPIE, 2001.
- [63] B. Hofer, B. Považay, B. Hermann, A. Unterhuber, G. Matz, and W. Drexler, “Dispersion encoded full range frequency domain optical coherence tomography,” *Opt. Express*, vol. 17, pp. 7–24, Jan 2009.
- [64] D. Y. Kim, J. S. Werner, and R. J. Zawadzki, “Comparison of phase-shifting techniques for in vivo full-range, high-speed Fourier-domain optical coherence tomography,” *Journal of biomedical optics*, vol. 15, p. 056011, Oct. 2010. Place: United States.
- [65] Q. Liu, L. Li, H. Zhang, W. Huang, and X. Yue, “Simultaneous dual-wavelength phase-shifting interferometry for surface topography measurement,” *Optics and Lasers in Engineering*, vol. 124, p. 105813, 2020.
- [66] O. Carrasco-Zevallos and J. A. Izatt, “Complex Conjugate Removal in SS Optical Coherence Tomography,” in *Optical Coherence Tomography*, p. 255, springer, 2015.
- [67] Z. Ren, K. Cui, L. Kong, Y. Huang, and R. Zhu, “A high-performance multistep phase-shifting interferometric fiber optic sensing method with modulation error

- correction capability,” *IEEE Sensors Journal*, vol. 22, no. 11, pp. 10576–10582, 2022.
- [68] D. Y. Kim, J. S. Werner, and R. J. Zawadzki, “Evaluation of complex conjugate artifact removal methods used in spectrometer-based Fourier-domain optical coherence tomography systems: a comparative study,” in *Optical Coherence Tomography and Coherence Domain Optical Methods in Biomedicine XIV* (J. A. Izatt, J. G. Fujimoto, and V. V. Tuchin, eds.), vol. 7554, p. 75542X, International Society for Optics and Photonics, SPIE, 2010.
- [69] Q. Zhang, S. Zhong, J. Lin, W. Chen, M. Luo, J. Zhong, Y. Yu, Z. Peng, and S. Cheng, “Full-range fourier-domain optical coherence tomography based on mach–zehnder interferometer,” *Optics and Lasers in Engineering*, vol. 124, p. 105794, 2020.
- [70] E. Götzinger, M. Pircher, R. A. Leitgeb, and C. K. Hitzenberger, “High speed full range complex spectral domain optical coherence tomography,” *Opt. Express*, vol. 13, pp. 583–594, Jan 2005.
- [71] L. García-Lechuga, B. L. Ortiz, V. H. F. M. noz, N.-I. Toto-Arellano, J. C. Seck-Tuoh-Mora, J. C. González-Islas, and G. Rodríguez-Zurita, “Two-step phase shifting interferometry based in two coupled sagnac interferometers,” in *Imaging and Applied Optics 2014*, p. JTh1C.3, Optica Publishing Group, 2014.
- [72] B. Hofer, B. Považay, B. Hermann, A. Unterhuber, G. Matz, and W. Drexler, “Dispersion encoded full range frequency domain optical coherence tomography,” *Opt. Express*, vol. 17, pp. 7–24, Jan 2009.
- [73] I. A. Viedma, D. Alonso-Caneiro, S. A. Read, and M. J. Collins, “Deep learning in retinal optical coherence tomography (oct): A comprehensive survey,” *Neurocomputing*, vol. 507, pp. 247–264, 2022.
- [74] M. Pekala, N. Joshi, T. A. Liu, N. M. Bressler, D. C. DeBuc, and P. Burlina, “Deep learning based retinal oct segmentation,” *Computers in biology and medicine*, vol. 114, p. 103445, 2019.
- [75] B. R. Chintada, S. Ruiz-Lopera, R. Restrepo, M. Villiger, B. E. Bouma, and N. Uribe-Patarroyo, “Practical volumetric speckle reduction in oct using deep learning,” in *Optical Coherence Imaging Techniques and Imaging in Scattering Media V*, p. 126321F, Optica Publishing Group, 2023.
- [76] B. R. Chintada, S. Ruiz-Lopera, R. Restrepo, B. E. Bouma, M. Villiger, and N. Uribe-Patarroyo, “Probabilistic volumetric speckle suppression in oct using deep learning,” *Biomedical Optics Express*, vol. 15, no. 8, pp. 4453–4469, 2024.

- [77] B. Zhang, R. Zheng, and F. Wang, “A method of complex conjugate artifact elimination for ss-oct systems based on adversarial generative model,” in *2023 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)*, pp. 1176–1180, 2023.
- [78] Z. Yuan, D. Yang, H. Pan, and Y. Liang, “Axial super-resolution study for optical coherence tomography images via deep learning,” *IEEE Access*, vol. 8, pp. 204941–204950, 2020.
- [79] H. Cheong, S. K. Devalla, T. Chuangsuwanich, T. A. Tun, X. Wang, T. Aung, L. Schmetterer, M. L. Buist, C. Boote, A. H. Thiéry, and M. J. A. Girard, “Octgan: single step shadow and noise removal from optical coherence tomography images of the human optic nerve head,” *Biomed. Opt. Express*, vol. 12, pp. 1482–1498, Mar 2021.
- [80] Y. Tian, D. Su, S. Lauria, and X. Liu, “Recent advances on loss functions in deep learning for computer vision,” *Neurocomputing*, vol. 497, pp. 129–158, 2022.
- [81] Z. Allen-Zhu, Y. Li, and Z. Song, “A convergence theory for deep learning via over-parameterization,” in *International conference on machine learning*, pp. 242–252, PMLR, 2019.
- [82] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [83] J. Pomerat, A. Segev, and R. Datta, “On neural network activation functions and optimizers in relation to polynomial regression,” in *2019 IEEE International Conference on Big Data (Big Data)*, pp. 6183–6185, IEEE, 2019.
- [84] L. Bottou, “Stochastic gradient descent tricks,” in *Neural Networks: Tricks of the Trade: Second Edition*, pp. 421–436, Springer, 2012.
- [85] P. Zhou, J. Feng, C. Ma, C. Xiong, S. C. H. Hoi, *et al.*, “Towards theoretically understanding why sgd generalizes better than adam in deep learning,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 21285–21296, 2020.
- [86] N. S. Keskar and R. Socher, “Improving generalization performance by switching from adam to sgd,” *arXiv preprint arXiv:1712.07628*, 2017.
- [87] A. Gupta, R. Ramanath, J. Shi, and S. S. Keerthi, “Adam vs. sgd: Closing the generalization gap on image classification,” in *OPT2021: 13th Annual Workshop on Optimization for Machine Learning*, 2021.

- [88] S. De, A. Mukherjee, and E. Ullah, “Convergence guarantees for rmsprop and adam in non-convex optimization and an empirical comparison to nesterov acceleration,” *arXiv preprint arXiv:1807.06766*, 2018.
- [89] A. Lewkowycz, “How to decay your learning rate,” *arXiv preprint arXiv:2103.12682*, 2021.
- [90] N. A. Hamid, N. M. Nawi, R. Ghazali, and M. N. M. Salleh, “Solving local minima problem in back propagation algorithm using adaptive gain, adaptive momentum and adaptive learning rate on classification problems,” in *International Journal of Modern Physics: Conference Series*, vol. 9, pp. 448–455, World Scientific, 2012.
- [91] K. You, M. Long, J. Wang, and M. I. Jordan, “How does learning rate decay help modern neural networks?,” *arXiv preprint arXiv:1908.01878*, 2019.
- [92] L. N. Smith, “Cyclical learning rates for training neural networks,” in *2017 IEEE winter conference on applications of computer vision (WACV)*, pp. 464–472, IEEE, 2017.
- [93] J. Li and X. Yang, “A cyclical learning rate method in deep learning training,” in *2020 International Conference on Computer, Information and Telecommunication Systems (CITS)*, pp. 1–5, IEEE, 2020.
- [94] R. Gulde, M. Tuscher, A. Csiszar, O. Riedel, and A. Verl, “Deep reinforcement learning using cyclical learning rates,” in *2020 Third International Conference on Artificial Intelligence for Industries (AI4I)*, pp. 32–35, IEEE, 2020.
- [95] J. Henry, T. Natalie, and D. Madsen, “Pix2pix gan for image-to-image translation,” *Research Gate Publication*, pp. 1–5, 2021.
- [96] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [97] M. Drozdal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, “The importance of skip connections in biomedical image segmentation,” in *International workshop on deep learning in medical image analysis, international workshop on large-scale annotation of biomedical data and expert label synthesis*, pp. 179–187, Springer, 2016.
- [98] V. Agarwal, A. V. Gribok, and M. A. A. and, “Image restoration using l1 norm penalty function,” *Inverse Problems in Science and Engineering*, vol. 15, no. 8, pp. 785–809, 2007.

- [99] Y. Yang, H. Zheng, L. Zeng, X. Shen, and Y. Zhan, “ $l_1$ -regularized reconstruction model for edge-preserving filtering,” *IEEE Transactions on Multimedia*, vol. 25, pp. 4148–4162, 2023.
- [100] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” *CoRR*, vol. abs/1611.07004, 2016.
- [101] B. Zhang, R. Zheng, and F. Wang, “A method of complex conjugate artifact elimination for ss-oct systems based on adversarial generative model,” in *2023 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)*, pp. 1176–1180, IEEE, 2023.
- [102] G. Bertasius, J. Shi, and L. Torresani, “High-for-low and low-for-high: Efficient boundary detection from deep object features and its applications to high-level vision,” in *Proceedings of the IEEE international conference on computer vision*, pp. 504–512, 2015.
- [103] R. Zhang, F. Zhu, J. Liu, and G. Liu, “Depth-wise separable convolutions and multi-level pooling for an efficient spatial cnn-based steganalysis,” *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1138–1150, 2019.
- [104] C. Fan, H. Lin, and Y. Qiu, “U-patch gan: A medical image fusion method based on gan,” *Journal of Digital Imaging*, vol. 36, no. 1, pp. 339–355, 2023.
- [105] N. Huynh, D. Yan, Y. Ma, S. Wu, C. Long, M. T. Sami, A. Almudaifer, Z. Jiang, H. Chen, M. N. Dretsch, T. S. Denney, R. Deshpande, and G. Deshpande, “The use of generative adversarial network and graph convolution network for neuroimaging-based diagnostic classification,” *Brain Sciences*, vol. 14, no. 5, 2024.
- [106] Y. Khalil, S. Amirrajab, C. Lorenz, J. Weese, J. Pluim, and M. Breeuwer, “Reducing segmentation failures in cardiac mri via late feature fusion and gan-based augmentation,” *Computers in Biology and Medicine*, vol. 161, p. 106973, 04 2023.
- [107] H. Cheong, S. K. Devalla, T. H. Pham, L. Zhang, T. Tun, X. Wang, S. Perera, L. Schmetterer, T. Aung, C. Boote, A. Thiery, and M. Girard, “Deshadowgan: A deep learning approach to remove shadows from optical coherence tomography images,” *Translational Vision Science and Technology*, vol. 9, p. 23, 04 2020.
- [108] “Hyperparameter Tuning Strategies for GANs.”
- [109] O. Mounjid and X. Guo, “Convergence of gans training: A game and stochastic control methodology,” 2021.

- [110] rendyk, “Tuning the Hyperparameters and Layers of Neural Network Deep Learning,” May 2021.
- [111] G. R. Ge, J. P. Rolland, and K. J. Parker, “Speckle statistics of biological tissues in optical coherence tomography,” *Biomed. Opt. Express*, vol. 12, pp. 4179–4191, Jul 2021.
- [112] X. Li, Z. Dong, H. Liu, J. J. Kang-Mieler, Y. Ling, and Y. Gan, “Frequency-aware optical coherence tomography image super-resolution via conditional generative adversarial neural network,” *Biomed. Opt. Express*, vol. 14, pp. 5148–5161, Oct 2023.
- [113] M. Wang, W. Zhu, K. Yu, Z. Chen, F. Shi, Y. Zhou, Y. Ma, Y. Peng, D. Bao, S. Feng, L. Ye, D. Xiang, and X. Chen, “Semi-supervised capsule cgan for speckle noise reduction in retinal oct images,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 4, pp. 1168–1183, 2021.
- [114] Y. Ma, X. Chen, W. Zhu, X. Cheng, D. Xiang, and F. Shi, “Speckle noise reduction in optical coherence tomography images based on edge-sensitive cgan,” *Biomed. Opt. Express*, vol. 9, pp. 5129–5146, Nov 2018.
- [115] L. Veselka, L. Krainz, L. Mindrinos, W. Drexler, and P. Elbau, “A quantitative model for optical coherence tomography,” *Sensors*, vol. 21, no. 20, p. 6864, 2021.
- [116] J. Sun, X. Cao, H. Liang, W. Huang, Z. Chen, and Z. Li, “New interpretations of normalization methods in deep learning,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 5875–5882, Apr. 2020.
- [117] D. M. Rouse and S. S. Hemami, “Understanding and simplifying the structural similarity metric,” in *2008 15th IEEE international conference on image processing*, pp. 1188–1191, IEEE, 2008.
- [118] G. Schmidt, B. E. Bouma, and N. Uribe-Patarroyo, “Asynchronous, semi-reverberant elastography,” *Optica*, vol. 11, pp. 1285–1294, Sep 2024.
- [119] H. Ahmed, Q. Zhang, R. Donnan, and A. Alomainy, “Transformer enhanced autoencoder rendering cleaning of noisy optical coherence tomography images,” *Journal of Medical Imaging*, vol. 11, no. 3, p. 034008, 2024.
- [120] T. Marciniak and A. Stankiewicz, “Automatic diagnosis of selected retinal diseases based on oct b-scan,” *Klinika Oczna / Acta Ophthalmologica Polonica*, vol. 126, no. 1, pp. 8–14, 2024.

- [121] R. Loganathan and S. Latha, “The empirical comparison of deep neural network optimizers for binary classification of oct images,” *Library of Progress-Library Science, Information Technology & Computer*, vol. 44, no. 3, 2024.
- [122] R. K. Ara, A. Matiolański, A. Dziech, R. Baran, P. Domin, and A. Wiczorkiewicz, “Fast and Efficient Method for Optical Coherence Tomography Images Classification Using Deep Learning Approach,” *Sensors (Basel, Switzerland)*, vol. 22, June 2022. Place: Switzerland.
- [123] Y. Xu, R. Quan, W. Xu, Y. Huang, X. Chen, and F. Liu, “Advances in medical image segmentation: A comprehensive review of traditional, deep learning and hybrid approaches,” *Bioengineering*, vol. 11, no. 10, 2024.
- [124] S. Ruiz-Lopera, R. Restrepo, C. Cuartas-Vélez, B. E. Bouma, and N. Uribe-Patarroyo, “Computational adaptive optics in phase-unstable optical coherence tomography,” *Opt. Lett.*, vol. 45, pp. 5982–5985, Nov 2020.
- [125] S. Ruiz-Lopera, R. Restrepo, T. M. Cannon, M. Villiger, B. E. Bouma, and N. Uribe-Patarroyo, “Computational refocusing in polarization-sensitive optical coherence tomography with phase unstable systems,” in *Optical Coherence Tomography and Coherence Domain Optical Methods in Biomedicine XXV*, vol. 11630, p. 116300E, SPIE, 2021.
- [126] S. Ruiz-Lopera, R. Restrepo, T. M. Cannon, M. Villiger, B. E. Bouma, and N. Uribe-Patarroyo, “Computational refocusing in phase-unstable polarization-sensitive optical coherence tomography,” *Optics letters*, vol. 48, no. 18, pp. 4765–4768, 2023.
- [127] Y. Ling, Z. Dong, X. Li, Y. Gan, and Y. Su, “Deep learning empowered highly compressive ss-oct via learnable spectral&#x2013;spatial sub-sampling,” *Opt. Lett.*, vol. 48, pp. 1910–1913, Apr 2023.
- [128] V. Singh, “InDepth Guide to Denoising Diffusion Probabilistic Models DDPM,” Mar. 2023.
- [129] S. Li, R. Higashita, H. Fu, H. Li, J. Niu, and J. Liu, “Content-preserving diffusion model for unsupervised as-oct image despeckling,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023* (H. Greenspan, A. Madabhushi, P. Mousavi, S. Salcudean, J. Duncan, T. Syeda-Mahmood, and R. Taylor, eds.), (Cham), pp. 660–670, Springer Nature Switzerland, 2023.