



Universidad
Rey Juan Carlos

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA
DE TELECOMUNICACIÓN

GRADO EN INGENIERÍA BIOMÉDICA

TRABAJO FIN DE GRADO

**DISEÑO DE UN CONFORMADOR DE HAZ BASADO
EN DEEP LEARNING PARA LA MEJORA DE LA
IMAGEN ULTRASÓNICA**

Autor: Carolina Bertoncini Fernández

Tutor: Miguel Ángel Cámara Vázquez

Cotutor: Óscar Fernando Martínez Graullera

Curso académico: 2022/2023

*La motivación es lo que te pone en marcha, el
hábito lo que hace que sigas.*

Jim Ryun (1947)

Agradecimientos

Este proyecto pone fin a cuatro años de aprendizaje, no solo a nivel académico sino también a nivel personal. Me gustaría dar las gracias a todos aquellos que me han apoyado durante estos años y que han hecho posible la finalización de esta etapa.

En primer lugar, a mi tutor de la universidad Miguel Ángel Cámara, por haber confiado en mí y por haber podido contar con él en cualquier momento. Gracias por tu tiempo, tus consejos y por acompañarme en este proyecto.

Poder realizar este trabajo no hubiera sido posible sin la confianza que el CSIC y en concreto mi tutor de prácticas Óscar Martínez han depositado en mí, compartiendo su idea, sus recursos y principalmente su tiempo para que todo saliera lo mejor posible.

Muchas gracias.

Agradecer a los que empezaron siendo compañeros de clase y han acabado siendo amigos para toda la vida. Me llevo a personas increíbles de esta etapa. Y a mis amigos de siempre por sacar tiempo para vernos y apoyarme en todo.

Finalmente, dar gracias a mi familia y a Dani, por haberme comprendido en mis malas rachas y motivarme durante estos años a continuar. Gracias por todo el cariño que siempre he recibido y estar conmigo cuando lo he necesitado.

Resumen

Hoy en día los sistemas de imagen ultrasónica, concretamente las ecografías, se han convertido en una herramienta esencial de diagnóstico en medicina, permitiendo visualizar la anatomía interna de forma no invasiva mediante procesos menos costosos que otras técnicas similares, como son Rayos X o Resonancias Magnéticas. A pesar de ello, en este tipo de imágenes aparecen los llamados lóbulos secundarios, distorsionando la imagen y requiriendo un alto nivel de conocimientos y experiencia del operador para la interpretación de estas imágenes.

Por todo ello, el principal objetivo de este proyecto es el diseño de un conformador de haces de ultrasonido utilizando técnicas de Deep Learning con el fin de reducir estos artefactos y mejorar la calidad de imagen. En concreto, se plantea un estudio detallado de varios tipos de redes neuronales artificiales profundas a partir de dos tipos de datos extraídos de las imágenes, datos frecuenciales y datos temporales. Estas redes planteadas para resolver un problema de regresión se caracterizan por ser modelos sencillos con pocas neuronas dentro de cada capa oculta con el fin de disminuir la complejidad y no requerir de grandes recursos computacionales.

Se ha construido un generador de imágenes sintéticas en el entorno de programación *Python* que ha permitido realizar pruebas en condiciones controladas. A través de ello, se ha podido conocer el comportamiento de los modelos de redes neuronales frente a la resolución, el ruido y la intensidad de los blancos. Esto ha permitido llevar a cabo el diseño de los diferentes modelos de redes neuronales más eficientes de cara a la tarea a resolver.

El método propuesto proporciona un primer estudio positivo acerca la capacidad de las redes neuronales profundas en la eliminación de lóbulos secundarios, permitiendo además abordar la interpretabilidad de los sistemas aprendidos a partir de las observaciones. Los resultados obtenidos muestran la viabilidad de esta aproximación en el diseño de un conformador de haces ultrasónico a partir de técnicas de DL pudiendo visualizar de manera fiable aquellos blancos que presentan una intensidad mayor a -20dB.

Lista de Acrónimos y Abreviaturas

A continuación, se muestra la lista de acrónimos utilizados en el desarrollo de este proyecto:

ABLE Adaptative Beamforming

BD Big Data

BN Batch Normalization

CNN Red Neuronal Convolutacional del inglés *Convolutional Neural Network*

CT Tomografía Computarizada del inglés *Computed Tomography*

DAS Delay and Sum

DL Deep Learning

DNN Deep Neural Network

FPGAs Field Programmable Gate Array

FNN Feedforward Neural Network

FMC Full Matrix Capture

ML Machine Learning

MLP Perceptrón Multicapa del inglés *Multilayer Perceptron*.

MSE Mean Squared Error

MSLE Mean Squared Logaritmico Error

PA Lineal Phased Array

PLD Dispositivo Lógico Programable del inglés *Programmable Logic Device*

PSF Función de Dispersión de Punto del inglés *Point Spread Function*.

PSNR Peak Signal to Noise Ratio

RM Resonancia Magnética

RNA Red Neuronal Artificial

ROI Region of Interest

SAFT Synthetic Aperture Focusing Technique

SAU Synthetic Aperture Ultrasound

SNR Relación Señal a Ruido del inglés *Signal to Noise Ratio*.

TF Transformada de Fourier

TFG Trabajo Fin de Grado

TL Transfer Learning

US Ultrasonido

Índice general

Agradecimientos	5
Resumen	7
Lista de Acrónimos y Abreviaturas	9
Índice de Figuras	14
Índice de Tablas	17
1. Introducción	19
1.1. Motivación.....	19
1.2. Objetivos y Estructura de la memoria.....	21
2. Marco teórico	25
2.1. Propagación de ondas de ultrasonido.....	25
2.1.1. Ecuación de onda.....	27
2.1.2. Parámetros de onda.....	28
2.2. Física del ultrasonido	29
2.3. Pulso-Eco. Imágenes en Modo B	32
2.4. Factores de calidad de imagen.....	34
2.5. Transductores y arrays	35
2.5.1. Conformador de haces	36
2.5.2. Metodología del Conformador de haces ultrasónicos.....	41
3. Machine Learning y Deep Learning aplicado a imagen ultrasónica	45

3.1.	Inteligencia artificial en US: Motivación	45
3.2.	Introducción al ML	46
3.3.	Deep Learning	48
3.3.1.	DL aplicado a imagen ultrasónica	54
4.	La imagen de apertura sintética	57
4.1.	Apertura sintética.....	57
4.1.1.	Concepto de apertura sintética	58
4.1.2.	Técnica SAFT.....	58
4.1.3.	El Coarray	60
4.1.4.	Características de la imagen SAFT.....	62
5.	Materiales y Métodos	66
5.1.	Hipótesis de partida	66
5.1.1.	Estimación de la reflectividad	67
5.2.	Datos sintéticos.....	68
5.3.	Experimentos con Redes Neuronales.....	70
5.4.	Resolución de la imagen objetivo.....	72
5.4.1.	Estimación del tamaño de las redes	72
5.4.2.	Funciones de activación y función de coste.....	75
5.4.3.	Implementación y entrenamiento	76
5.5.	Evaluación.....	78
6.	Resultados basados en regresión	82
6.1.	Solución en tiempo	82
6.1.1.	Modelo red 1	82
6.1.2.	Modelo red 2	85
6.1.3.	Modelo red 3	87
6.2.	Solución en frecuencia.....	89
6.2.1.	Modelo red 1	89

6.2.2. Modelo red 2	91
6.2.3. Modelo red 3	93
7. Discusión y líneas de trabajo futuras.....	97
7.1. Discusión.....	97
7.2. Líneas Futuras	99
Bibliografía.....	102

Índice de Figuras

Figura 1.1 Esquema e imagen de lóbulos secundarios en la señal y en el haz de ultrasonidos	20
Figura 2.1. Espectro de vibraciones acústicas	26
Figura 2.2. Diagrama del sistema de imagen ultrasónica.....	26
Figura 2.4. Variación en la propagación de ondas para campo cercano y campo lejano	29
Figura 2.5. Reflexión: dispersión de la onda y vuelta al transductor. Refracción: cambio de dirección de la onda	30
Figura 2.6. Fenómeno de dispersión difusa, la energía se dispersa en todas las direcciones.....	31
Figura 2.7. Representación amplitud del eco. Imagen Modo A	33
Figura 2.8. Representación de la información como puntos brillantes en 2D. Imagen Modo B.....	33
Figura 2.9. Representación profundidad de penetración. Imagen Modo M	34
Figura 2.10. Ejemplo de Lineal Phased Array de 6 elementos	36
Figura 2.11. Deflexión del haz en emisión con un array lineal	37
Figura 2.12. Focalización del haz en emisión en el eje con un array lineal	38
Figura 2.13. Focalización con array lineal.....	40
Figura 2.14. Esquema simplificado del DAS Beamforming	41
Figura 2.15. Ejemplo de aplicación de diferentes tipos de ventanas para reducción de la amplitud.....	43
Figura 3.1. Overfitting: excesivo ajuste en los datos de entrenamiento. Underfitting: modelo excesivamente simple	48
Figura 3.2. Modelo de neurona estándar	49
Figura 3.3. Algoritmo descenso de gradiente	51

Figura 3.4. Ejemplo de convolución en 2D. Se dibujan cuadros con flechas para indicar cómo se forma el elemento superior izquierdo de la salida aplicando el kernel a la región superior izquierda correspondiente a los datos de entrada..... 53

Figura 3.5. Arquitectura CNN con dos convoluciones sucesivas y capa fully connected para resolver un problema de clasificación 54

Figura 4.1. Secuencia de activación de los elementos del array en un SAFT 60

Figura 4.2. Subaperturas, subcoarray y coarray al aplicar un SAFT sobre array de 8 elementos..... 62

Figura 4.3. Principio básico del SAFT..... 63

Figura 5.1. Imagen sintética en dB. Obtenida con parámetros de: $bw = 0.5$, $\lambda = 0.5$ y frecuencia central = 2. 69

Figura 5.2. Representación de 25a) Datos temporales b) Datos frecuencia 69

Figura 5.3. Imagen de referencia enfoque de regresión en dB 72

Figura 6.1. Esquema modelo red 1 74

Figura 6.2. Representación función de activación GeLU 75

Figura 6.3. Diagramas Coeficiente Dice y Distancia Hausdorff Red 1 Datos Temporales 83

Figura 6.5. Comparación imagen original - predicción modelo red 1 Datos Temporales 84

Figura 6.6. Diagramas Coeficiente Dice y Distancia Hausdorff Red 2 Datos Temporales 86

Figura 6.7. Comparación imagen original - predicción modelo red 2 Datos Temporales 86

Figura 6.8. Diagramas Coeficiente Dice y Distancia Hausdorff Red 3 Datos Temporales 88

Figura 6.9 Comparación imagen original - predicción modelo red 3 Datos Temporales 88

Figura 6.10. Diagramas Coeficiente Dice y Distancia Hausdorff Red 1 Datos Frecuenciales 90

Figura 6.11. Comparación imagen original - predicción modelo red 1 Datos Frecuenciales 90

Figura 6.12. Diagramas Coeficiente Dice y Distancia Hausdorff Red 2 Datos Frecuenciales 92

Figura 6.13. Comparación imagen original - predicción modelo red 2 Datos Frecuenciales	93
Figura 6.14. Diagramas Coeficiente Dice y Distancia Hausdorf Red 3 Datos Frecuenciales	94
Figura 6.15. Comparación imagen original - predicción modelo red 3 Datos Frecuenciales	95

Índice de Tablas

Tabla 2.1. Velocidad de propagación de onda en diferentes tejidos.....	28
Tabla 6.1. Resultados métricas modelo 1 Datos Temporales	83
Tabla 6.2. Resultados métricas modelo 2 Datos Temporales	85
Tabla 6.3. Resultados métricas modelo 1 Datos Temporales	87
Tabla 6.4. Resultados métricas modelo 1 Datos Frecuenciales	89
Tabla 6.5. Resultados métricas modelo 2 Datos Frecuenciales	91
Tabla 6.6. Resultados métricas modelo 3 Datos Frecuenciales	93

Capítulo 1

Introducción

Este primer capítulo del Trabajo de Fin de Grado (TFG) describe de forma general las líneas principales de este proyecto. En primer lugar, se detalla la motivación que ha impulsado el desarrollo de este, los objetivos que se han llevado a cabo, así como conceptos previos. Además, en la última parte se expone la estructura del proyecto.

1.1. Motivación

Los sistemas de imagen ultrasónica se han convertido en las últimas décadas en la modalidad de exploración preferida en una gran variedad de escenarios, siendo una herramienta esencial de diagnóstico en medicina, y en otros sectores como la industria, aplicándose en la realización de ensayos no destructivos.

La imagen médica mediante técnicas ultrasónicas (US) nos permite visualizar la anatomía interna de manera no invasiva, complementando la información obtenida con otros métodos como Rayos-X, resonancia magnética (RM) o tomografía computarizada (CT), sirviendo como un activo fundamental para el diagnóstico oportuno, la estadificación, el tratamiento y seguimiento de enfermedades. Las ventajas y diferencias de los US con respecto a estas radican en una mayor seguridad, ya que no producen radiaciones ionizantes, eliminando los efectos nocivos para el organismo humano. Además, es un proceso menos costoso que opera en tiempo real y no requiere de grandes

1. INTRODUCCIÓN

infraestructuras para su uso. Esta técnica se puede implementar con dispositivos portátiles, lo que resulta especialmente ventajoso en el caso de aplicaciones médicas, ya que no es necesario adaptar ninguna estancia, siendo suficiente un pequeño espacio y una camilla, permitiendo además su uso en entornos de emergencia, clínicas rurales y países en vías de desarrollo.

Gracias a lo anterior, la técnica de imagen por US ha pasado de ser una técnica sofisticada, costosa y solo accesible a unos pocos a convertirse en una exploración básica, incorporada en la cartera de servicios de múltiples especialidades creciendo de manera exponencial en el mercado a nivel mundial.

El tipo de imagen médica más común basada en US es la ecografía. Un sistema ultrasónico está compuesto por el transductor ultrasónico, encargado de generar los pulsos y recibir los ecos procedentes del medio, la electrónica de emisión y recepción de señal, que determina las características del haz ultrasónico generado y los sistemas de procesamiento de señal y visualización [1].

A pesar de todas las ventajas y mejoras que se obtienen de la aplicación de ultrasonido, se encuentran una serie de problemas asociados. Entre ellos se encuentra la aparición de lóbulos secundarios. Este problema consiste en que, un blanco o punto de interés se visualiza como un lóbulo principal, donde normalmente aparece junto con estos lóbulos mencionados considerados elementos no deseados en el haz de ultrasonidos. Surgen debido a ondas mecánicas propagadas a través de ángulos distintos de la trayectoria principal, causando defectos en la representación del punto de interés, conocido como difracción. Esto produce un gran número de artefactos que distorsionan la imagen y exigen del operador un alto nivel de experiencia para la interpretación de la misma.

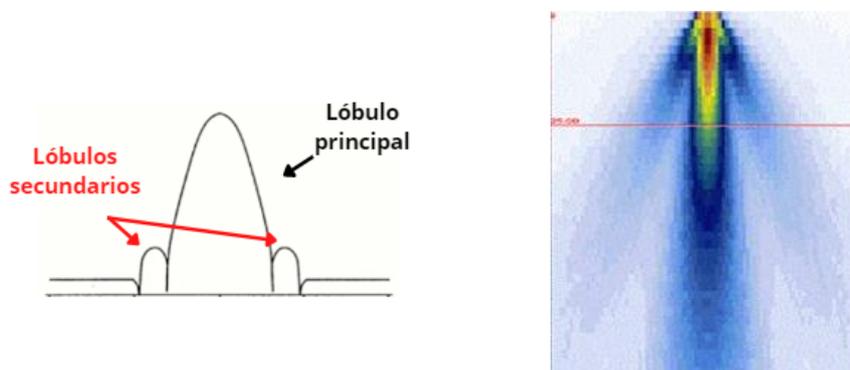


Figura 1.1 Esquema e imagen de lóbulos secundarios en la señal y en el haz de ultrasonidos [2].

Para lograr su eliminación y mejorar la resolución de la imagen se recurre a técnicas de inteligencia artificial. La inteligencia artificial se puede definir como el desarrollo de sistemas informáticos capaces de realizar tareas que requieren de la inteligencia humana. Algunas de sus aplicaciones en medicina son la ayuda del diagnóstico médico de forma rápida y la reducción de errores humanos entre otras [3].

Una de las ramas más utilizadas es el aprendizaje automático, conocido como *Machine Learning* (ML). Estas técnicas han demostrado estar entre los algoritmos más precisos y confiables de nuestros días, basándose principalmente en métodos de aprendizaje supervisado con redes neuronales mediante la utilización de un enfoque basado en datos. Para este tipo de aprendizaje se introducen una serie de datos a la entrada, los cuales sufren una serie de transformaciones para obtener la salida deseada. Los parámetros de estas transformaciones se aprenden a través de un algoritmo de descenso de gradiente diseñado para minimizar el error a la salida de la red [4].

Además de las técnicas de ML, dos de los términos más comunes que se escuchan relacionados con este son *Big Data* (BD) y *Deep Learning* (DL). La potencia del BD se encuentra en la disponibilidad de manejar grandes volúmenes de datos, añadido a la capacidad de las técnicas de DL, las cuales permiten explorar patrones lineales más complejos y aprender relaciones significativas dentro de los datos. Todo ello hace que en la actualidad se pueda extraer conocimiento valioso de los datos de una forma notablemente superior a la disponible en tiempos anteriores. Como ejemplo relevante para este TFG en los últimos años se han publicado una serie de trabajos sobre análisis de imágenes de ultrasonidos, reducción de ruido y segmentación con el fin de introducir estas tecnologías y algoritmos en el sector de la salud [5].

Teniendo en cuenta lo anterior, sería deseable poder reducir estos defectos de las imágenes mediante el uso de técnicas de ML o DL, mejorando su resolución y reduciendo la incertidumbre presente, pudiendo de esta manera facilitar el diagnóstico.

1.2. Objetivos y Estructura de la memoria

Resulta interesante conocer de antemano el funcionamiento de los dispositivos de ecografía, así como los diferentes tipos de datos que se pueden extraer de las ondas de

ultrasonido para, en la medida de lo posible hacer uso de los distintos algoritmos de inteligencia artificial. De esta forma, en un futuro se podría facilitar al personal sanitario su labor en el diagnóstico clínico de distintas patologías.

Por todo ello, el objetivo de este TFG es la creación de un filtro conformador para *arrays* lineales de transductores basado en técnicas de inteligencia artificial, para reducir aberraciones producidas por la difracción en imagen ultrasónica. (Una vez diseñado este conformador, el objetivo es poder llegar a probarlo sobre imágenes ecográficas reales).

Para conseguir el objetivo general anterior, los objetivos específicos del presente TFG se desglosan del siguiente modo.

- Diseñar un simulador de ondas y señales sintéticas de ecografía. Gracias a este, es posible construir imágenes sintéticas que permitan realizar distintas pruebas. Es necesario tener en cuenta que este generador permite controlar el nivel de ruido, la resolución, y el número de puntos de interés y su distribución, de manera que se puedan llevar a cabo distintos experimentos bajo condiciones controladas.
- Estudio y valoración de la información a introducir en el algoritmo de inteligencia artificial y su forma de entrada.
- Analizar el comportamiento de distintas redes neuronales frente a niveles de ruido, números de blancos en las imágenes y el efecto del número de capas ocultas.
- Diseño y estudio de un regresor que permita la reconstrucción de imágenes ecográficas en ausencia de defectos que impidan la correcta visualización de los blancos de interés.

Para conseguir los objetivos propuestos, se ha realizado un trabajo de investigación del ámbito que nos ocupa, así como un repaso de los temas fundamentales para el correcto desarrollo del TFG. Por todo ello, la estructura de la memoria se organiza de la siguiente manera:

- **Capítulo 1 – Introducción:** se detallan los motivos que llevaron al desarrollo del proyecto, los objetivos propuestos y la estructura de la memoria.

- **Capítulo 2 – Marco Teórico y Estado del Arte:** descripción de los conceptos básicos de este TFG. Centrándose en modelos de propagación de las ondas y funcionamiento y principios básicos del ultrasonido.
- **Capítulo 3 – Machine Learning y Deep Learning aplicado a imagen ultrasónica:** revisión del estado del arte de las diferentes técnicas de inteligencia artificial utilizadas en el procesamiento de señales ultrasónicas y en la obtención de una mejora de la calidad de imagen.
- **Capítulo 4 – La imagen de apertura sintética:** se realiza una descripción acerca del proceso de conformación de haces y una explicación de la imagen de apertura sintética y las técnicas existentes.
- **Capítulo 5 – Materiales y Métodos:** detalla la hipótesis de partida, las diferentes propuestas planteadas para la resolución del problema en función del conjunto de datos del que se dispone y las herramientas utilizadas para ello.
- **Capítulo 6 – Soluciones basadas en regresión:** se detallan las soluciones a las diversas pruebas realizadas enfocándose en la resolución de un problema de regresión tratando con datos temporales y frecuenciales, aportando y analizando los resultados correspondientes.
- **Capítulo 7 – Discusión y Líneas Futuras:** se lleva a cabo la discusión acerca de los resultados obtenidos, así como las conclusiones y propuestas de nuevas líneas de investigación.

Capítulo 2

Marco teórico

Este segundo capítulo proporciona los conocimientos necesarios para hacer posible el seguimiento del proyecto. En él se detallan conceptos relacionados con las ondas, su propagación y funcionamiento y principios básicos del ultrasonido, así como una descripción de las características básicas del transductor.

2.1. Propagación de ondas de ultrasonido

El sonido es cualquier fenómeno que lleve a cabo la propagación de ondas mecánicas a través de un medio físico elástico como consecuencia de la vibración de un cuerpo. Normalmente el medio por el que se propagan es el aire, aunque también puede ser en medios líquidos. Además, el sonido también es una transferencia de potencia, donde las vibraciones mecánicas se convierten en ondas de presión que transfieren energía al medio y a los objetos con los que entra en contacto la onda [6].

Por lo tanto, los ultrasonidos son una serie de ondas mecánicas de alta frecuencia (>20 kHz), que se encuentran por encima de la capacidad de audición del oído humano pero que presentan las mismas propiedades que las ondas audibles.

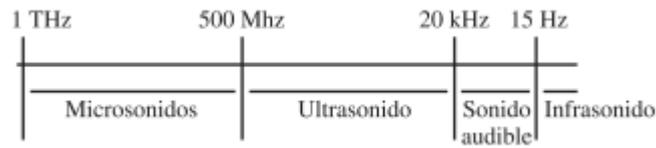


Figura 2.1. *Espectro de vibraciones acústicas donde se encuentran las ondas de ultrasonido entre 20kHz y 500Mhz [6].*

El proceso de generación de imagen se lleva a cabo mediante el envío de un pulso eléctrico al transductor el cual, gracias al efecto piezoeléctrico, los transforma en ondas mecánicas de ultrasonido. Este pulso ultrasónico se emite en una determinada dirección del espacio y al propagarse por el medio genera ondas reflejadas debido a las diferencias de impedancia acústica que encuentra en su camino. Estas partes de la onda reflejadas, denominadas ecos vuelven al origen y son recogidas por el transductor, transformándose en un pulso eléctrico, donde el tiempo de retorno de las ondas indica la profundidad de la superficie reflectante. Posteriormente las señales recogidas se componen en forma de imágenes [7].

Las ondas ultrasónicas que se transmiten a través de un medio se determinan en base al movimiento específico que cada tipo de onda provoca en los elementos del medio o la capacidad de imprimir un volumen al material.

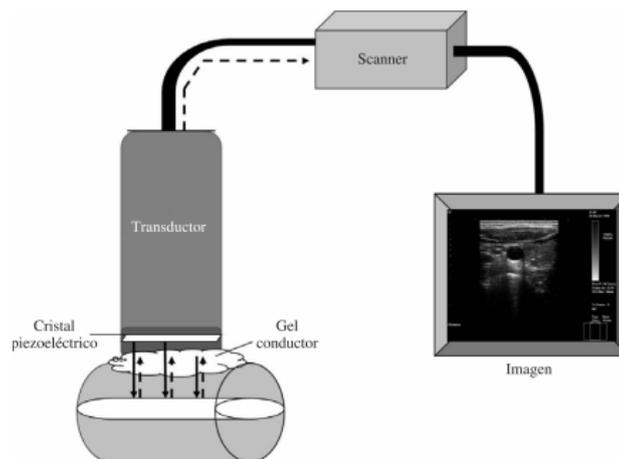


Figura 2.2. *Diagrama sistema de imagen ultrasónica. Se visualiza el transductor en contacto con el medio propagando y recogiendo las ondas mecánicas para posteriormente transformarlas en un pulso eléctrico y componer las imágenes [8].*

La ecografía médica se basa en la propagación de ondas longitudinales por el cuerpo, por lo que el movimiento de las partículas en el material se dirige en la dirección de propagación.

2.1.1. Ecuación de onda

Las ondas ultrasónicas generadas mediante transductores son señales gobernadas por las leyes de la física de propagación, en particular por la ecuación de onda (Ecuación (2.1)). Esta es una ecuación tridimensional dependiente del tiempo:

$$\frac{\partial^2 s}{\partial x^2} + \frac{\partial^2 s}{\partial y^2} + \frac{\partial^2 s}{\partial z^2} = \frac{1}{c^2} \frac{\partial^2 s}{\partial t^2} \quad (2.1)$$

donde $s(x, y, z, t) = s(\vec{x}, t)$ representa la presión del sonido en el espacio y en el tiempo, y c la velocidad de propagación.

La Ecuación (2.1) tiene muchas soluciones, y su derivación y obtención de la solución de manera detallada se encuentra en [9].

Es importante valorar que esta ecuación es válida únicamente en medios homogéneos, lineales y sin pérdidas, esto hace que en el tejido humano existan grandes complicaciones en la interacción con el sonido, debido a que no es un medio ideal.

Si se resuelve en coordenadas cartesianas, la solución presenta exponente complejo, representado en la expresión (2.2).

$$s(\vec{x}, t) = A e^{j(\omega t - \vec{k}\vec{x})} \quad (2.2)$$

Vemos que se obtiene una onda que viaja en la dirección x , donde A es una constante compleja, k es el número de onda ($k = \frac{2\pi}{\lambda}$) y w la frecuencia en radianes. (Por supuesto se pueden obtener soluciones correspondientes al movimiento de la onda en las direcciones de y y de z).

Sustituyendo la Ecuación (2.2) en la Ecuación (2.1) se obtiene la Ecuación (2.3). Indicando que, señales con la forma de la Ecuación (2.2) satisfacen la ecuación de onda.

$$|k|^2 = \frac{w^2}{c^2} \quad (2.3)$$

A esta solución de la ecuación de onda se denomina de ondas planas, debido a que, a cualquier instante de tiempo dado, las ondas se visualizan iguales a lo largo de cualquier plano. La linealidad de la ecuación de onda implica que multitud de ondas planas propagándose en direcciones diferentes pueden existir simultáneamente. Cualquier señal puede ser expresada como una superposición ponderada de exponenciales complejas [10].

La velocidad de propagación de la onda de ultrasonido varía en los diferentes tejidos biológicos. Valores típicos los podemos ver en la Tabla 2.1.

Tejido biológico	Velocidad (m/s)
Músculo	1585
Sangre	1570
Riñón	1561
Grasa	1549

Tabla 2.1. *Velocidad de propagación de onda en diferentes tejidos* [10].

2.1.2. Parámetros de onda

El campo ultrasónico está compuesto por dos regiones de interés, la región de campo cercano (región de Fresnel) y la región de campo lejano del *array* (región de Fraunhofer).

Para campo cercano las ondas emitidas por los diferentes elementos del *array* interfieren entre sí, generando una serie de mínimos y máximos a lo largo de la dirección de propagación hasta llegar a una profundidad determinada. La condición para campo cercano viene expresada en (2.4), donde Z_L es la profundidad, a el radio del transductor y λ la longitud de onda.

$$Z_L > \frac{a^2}{\lambda} \quad (2.4)$$

Mientras que para campo lejano, la diferencia que existe entre los trayectos recorridos por la onda emitida desde el elemento central y la del elemento más extremo del *array* es menor a media longitud de onda. De manera que el campo permanece invariante con la

distancia. Se utiliza la expresión (2.5) para determinar el valor de profundidad en función de las dimensiones del *array*.

$$Z_L = \frac{D^2}{4\lambda} \quad (2.5)$$

Como se visualiza en la Figura (2.4) las ondas se propagan como ondas planas o esféricas dependiendo de la distancia de propagación desde el punto de origen. Para campo cercano, el frente de onda se encuentra curvado mientras que se consideran ondas planas para campo lejano [11].

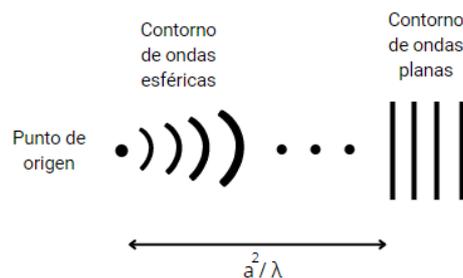


Figura 2.4. Variación en la propagación de ondas para campo cercano y campo lejano.

Es importante aclarar que la longitud de onda se define como la distancia propagada durante un periodo temporal. Su valor viene dado por la Ecuación (2.6) donde c es la velocidad de propagación, t el periodo y f su frecuencia temporal.

$$\lambda = c \cdot t = \frac{c}{f} \quad (2.6)$$

2.2. Física del ultrasonido

El factor más esencial de la física de los US es que la energía acústica es transportada por el tejido como una onda de sonido. Esta onda acústica se propaga mediante compresión longitudinal. Por lo tanto, al ser una onda, el US se encuentra sujeto a todos los comportamientos ondulatorios.

Reflexión

Cuando el transductor de US emite energía sonora en el cuerpo del paciente viaja a través del tejido sin impedimentos, hasta que se dispersa debido a cambios abruptos de impedancia acústica en las interfaces entre diferentes tipos de tejido. Parte del sonido vuelve al transductor, el cual lo recibe y procesa para extraer información, mientras que el resto de la onda continúa viajando hasta el siguiente reflector, y así sucesivamente.

Por lo tanto, la propiedad de reflexión del sonido hace posible la obtención de la imagen ultrasónica.

Es importante tener en cuenta que cada tipo de tejido presenta una impedancia acústica característica, definida como $z = \rho c$, donde ρ es la densidad del tejido y c la velocidad del sonido en el tejido.

Refracción

La refracción hace referencia al cambio de dirección de la onda a medida que viaja a través del tejido, causado por un cambio en la velocidad del sonido. Este fenómeno puede causar distorsiones en la imagen, debido a que las ondas de ultrasonido reflejadas utilizadas para formar la imagen pueden no viajar en línea recta desde el transductor al objetivo, y desde el objetivo al transductor.

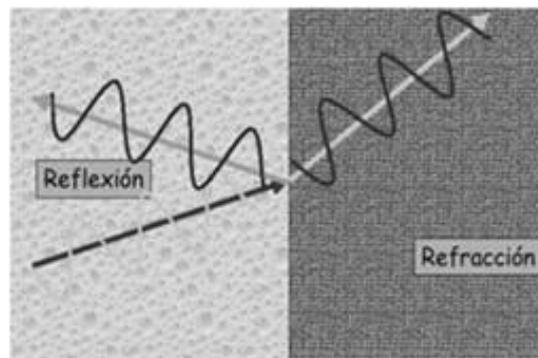


Figura 2.5. *Reflexión: dispersión de la onda y vuelta al transductor. Refracción: cambio de dirección de la onda [12].*

Dispersión

La dispersión es el resultado de la difracción. Cada volumen de tejido que contiene dos o más valores diferentes de impedancia es considerado un centro de dispersión.

Si el tamaño de la dispersión es pequeño comparado con la longitud de onda del US, la energía acústica se dispersa uniformemente en todas las direcciones. A esto se le llama dispersión difusa y la energía reflejada de vuelta al transductor es pequeña.

Por otra parte, si la dispersión es comparable con la longitud de onda, la energía se refleja en un ángulo igual al de incidencia. A esto se le denomina dispersión especular y la energía reflejada es mayor o menor dependiendo del ángulo incidente.

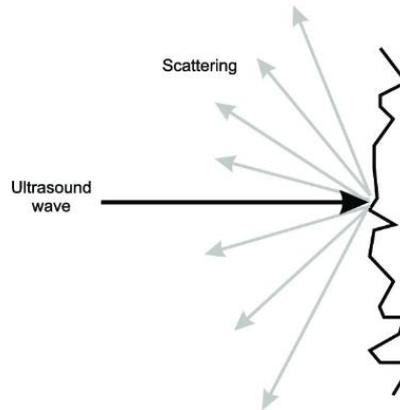


Figura 2.6. *Fenómeno de dispersión difusa, la energía se dispersa en todas las direcciones [13].*

Dentro de la dispersión podemos encontrar lo que se conoce como *speckle*. El *speckle* es el resultado de una propiedad importante de los US, la coherencia de onda. Este fenómeno se produce debido a que el frente de onda que avanza se refleja simultáneamente desde muchas dispersiones. El frente de onda que se recibe es la suma coherente de la amplitud del sonido reflejada por cada uno.

El *speckle* es a lo que se conoce como un “ruido multiplicativo” y su tamaño y características se encuentran determinadas por la frecuencia del transductor, la forma del pulso transmitido y la anchura del haz. Añade brillo o textura a la imagen de US, limitando la resolución espacial y de contraste

Atenuación

A medida que la onda de US se propaga a través del medio, su intensidad va disminuyendo, se va atenuando. Esto se produce por los fenómenos de absorción, reflexión y difracción.

La absorción se define como la conversión local de la energía de la onda de US a otras formas de energía, mayormente calor. La cantidad de energía que se absorbe depende de la distancia recorrida y la frecuencia de la onda.

La atenuación se puede modelar como se expresa en la Ecuación (2.7), donde α y β son parámetros acústicos del medio, y f es la frecuencia central del transductor [14].

$$\mu(f) = \alpha f^\beta \quad (2.7)$$

En la Ecuación (2.7) vemos como la atenuación es proporcional a la frecuencia central cuando $\beta = 1$. Consecuentemente, altas frecuencias de los transductores resultan en un aumento de la atenuación.

Para tejido humano, los valores típicos de α y β son $\beta = 1$ y $\alpha = 1\text{dB/cm/MHz}$ [14].

Difracción

Debido a la naturaleza de la onda ocurre la difracción, donde el haz emitido desde el transductor aumenta su diámetro con la distancia [11].

Dentro de la región de campo cercano el diámetro exterior del haz permanece uniforme; posteriormente el haz se propaga más allá de esta región. Por lo tanto, la difracción es el fenómeno determinante primario de la forma del haz de US.

2.3. Pulso-Eco. Imágenes en Modo B

La técnica de ultrasonidos de pulso – eco es un método de formación de imagen muy común. Se basa en ondas reflejadas o dispersadas desde las interfaces de tejido. A pesar de que la complejidad de la interacción del sonido con los tejidos blandos complica considerablemente su interpretabilidad, la versatilidad clínica de este método lo convierte en una de las técnicas de ultrasonido más importantes que existen en la actualidad [15]. Se encuentran disponibles dos tipos de examen de ultrasonido, métodos de generación de imagen por pulso-eco y métodos Doppler [10].

Los métodos de generación de imagen por pulso eco son los que interesan para el presente trabajo. Para estos, la demodulación de la amplitud da como resultado una serie

de métodos de visualización que muestran la amplitud del eco frente a la profundidad a lo largo de la trayectoria de los pulsos de ultrasonido que se emiten desde el transductor. Dependiendo de cómo se represente la información, estos métodos pueden ser clasificados como:

- **Modo A (Amplitud):** es el tipo más antiguo de representación y el instrumento de ultrasonidos de pulso – eco más simple. La amplitud de eco se representa en función de la profundidad de penetración.

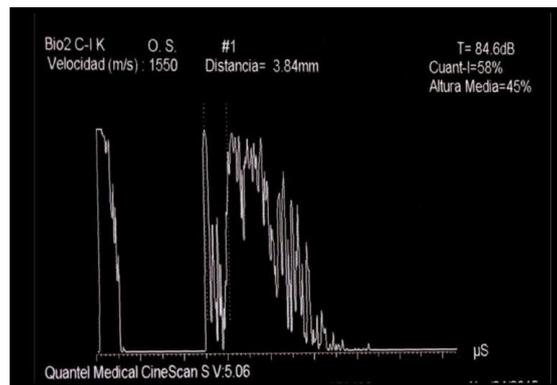


Figura 2.7. Representación amplitud del eco. Imagen Modo A [16].

- **Modo B (Brillo):** representa la información del modo A como puntos brillantes. Generalmente se utiliza escala de grises y una representación en 2 dimensiones de un escaneo sectorial.



Figura 2.8. Representación de la información como puntos brillantes en 2D. Imagen Modo B [16].

- **Modo C (Profundidad constante):** es muy similar al modo B. La diferencia es que el plano de exploración se fija a una profundidad constante desde la sonda.

- **Modo M (Movimiento):** representa la profundidad de penetración a lo largo de un eje, mientras que el resto de los ejes representan el tiempo. La amplitud se visualiza en escala de grises.

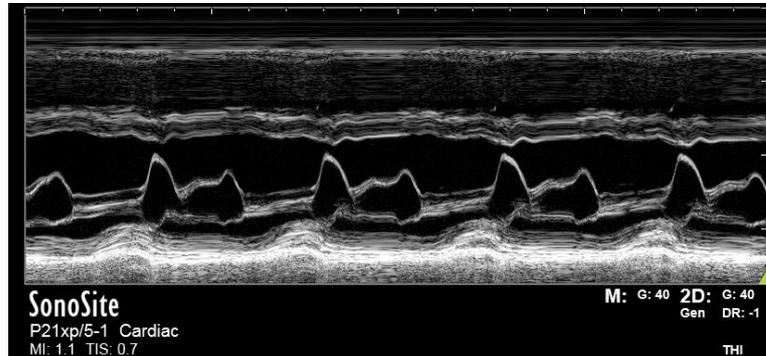


Figura 2.9. Representación profundidad de penetración. Imagen Modo M [16].

Las imágenes de Modo B permiten mostrar el interior de estructuras de tejido blando, por ello es el tipo de imagen ultrasónica más utilizada. Para el presente trabajo se utilizarán imágenes en este modo generadas a partir de métodos pulso-eco.

2.4. Factores de calidad de imagen

Para poder obtener un diagnóstico mediante imágenes de US es necesario que estas cuenten con la suficiente calidad para asegurar una correcta interpretación del clínico. Por lo que, es importante contar con criterios objetivos que permitan llevar a cabo una buena evaluación de la calidad de las imágenes ultrasónicas.

- **Resolución espacial:** mínima distancia a la que pueden estar situados dos reflectores para considerarlos separados. Su valor es clave para cuantificar la capacidad del sistema de resolver espacialmente pequeñas estructuras y de proporcionar una buena definición a los bordes en la imagen.

La resolución espacial para imagen en dos dimensiones está determinada por 2 componentes. La resolución axial depende de la longitud del pulso y hace referencia a la capacidad del sistema para separar objetos ubicados a lo largo del eje de propagación del haz. Mientras que la resolución lateral se mide perpendicular a la dirección de propagación del US, por lo que indica la

capacidad de separar objetos perpendicularmente al haz. Establece la relación entre la longitud de onda y el tamaño de la apertura.

- **Contraste:** mínima diferencia de impedancia acústica del medio que se puede mostrar con dos niveles de gris diferentes en la imagen. Un contraste adecuado es esencial en imagen médica donde las diferencias de impedancia pueden ser muy pequeñas, permitiendo una correcta discriminación del tejido para establecer un diagnóstico acertado.

Los principales factores que influyen en el contraste son la resolución espacial, el rango dinámico del sistema, el nivel de los lóbulos laterales y el ruido. La detección de objetos de alto contraste está limitada por la resolución espacial, mientras que los de bajo contraste por el ruido de la imagen.

Muchas veces al contraste se le hace referencia como rango dinámico, estableciendo una relación entre la señal más fuerte y las más débil que el sistema proporciona.

- **Relación temporal:** capacidad de detectar movimiento en la imagen dependiente del tiempo necesario para formarla. Generalmente se denomina cadencia de imágenes (*frame rate*), que es el número de imágenes adquiridas por unidad de tiempo.
- **Relación señal – ruido:** relación entre la información contenida en la señal frente al ruido que interfiere en la misma.

El ruido es fundamentalmente de dos tipos: ruido eléctrico y ruido de grano. El ruido eléctrico puede controlarse cuidando el diseño electrónico. Mientras que el ruido de grano hace referencia a los artefactos que aparecen en la imagen.

2.5. Transductores y arrays

Los transductores piezoeléctricos permiten convertir energía eléctrica en energía mecánica y viceversa emitiendo pulsos cortos y recibiendo los ecos reflectados. El patrón de radiación que presente el transductor es importante, debido a que de este dependen las prestaciones del sistema de imagen.

Los *arrays* son un conjunto de transductores que podemos modificar de forma activa para variar la forma del campo ultrasónico. Cada elemento puede transmitir y recibir

señales ultrasónicas. Existen diferentes tipos de transductores basados en *arrays*, pero este TFG se centra en lo que se denomina *Lineal Phased Arrays* (PA).

Un transductor de tipo *Lineal Phased Array* (Figura 2.10) consiste en una serie de elementos localizados a lo largo de una línea con la misma distancia entre ellos, donde cada uno de los elementos puede transmitir pulsos y recibir señales de eco individualmente. Esta implementación permite obtener un corte de la zona del espacio situada frente al transductor. La distancia a la que se encuentran los elementos es de $\lambda/2$ para evitar en la medida de lo posible lóbulos no deseados en el patrón de radiación. Se comenta con más detalle en apartados siguientes.

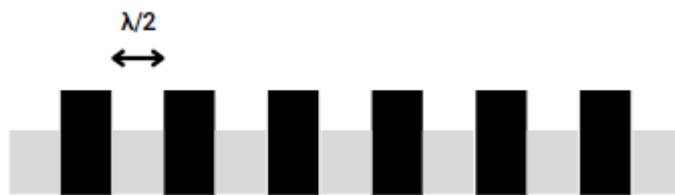


Figura 2.10. Ejemplo de *Lineal Phased Array* de 6 elementos.

Este tipo de transductores son los más utilizados en ecógrafos clínicos modernos.

Métodos de escaneo

Para formar una imagen, es necesario realizar un barrido transversal del haz sobre el objeto. Para imagen médica se utilizan dos métodos de barrido: lineal y sectorial.

- En un barrido sectorial, el haz es dirigido por señales que se aplican a los elementos del array con retrasos, que son una función lineal de la posición del elemento en la dirección de exploración.
- Para un barrido lineal, los elementos de un array de transductores son excitados uno por uno o bien por grupos de manera secuencial en una dirección determinada.

2.5.1. Conformador de haces

Previo al desarrollo de la imagen de apertura sintética y sus características es importante entender el proceso de formación de haces, conocido en inglés como *beamforming*.

Las técnicas de *beamforming* son unas de las partes más importantes en la imagen ultrasónica. Consisten en algoritmos de procesamiento de *arrays* que enfocan la capacidad de captura de señales del *array* en una dirección determinada [17]. Cada haz es controlado por métodos de control de haces, entre ellos se encuentran apertura dinámica, enfoque dinámico y apodización.

Deflexión ultrasónica

Una de las motivaciones de utilizar *arrays* de elementos es la posibilidad de deflectar electrónicamente el haz ultrasónico, lo que significa sin movimientos mecánicos. El principio se muestra para el haz emitido en la Figura 2.11.

En una excitación común cada elemento se retrasa un intervalo de tiempo determinado. La vibración de cada elemento se propaga en el medio; como cada uno ha sido excitado en un instante de tiempo diferente, el conjunto de las vibraciones crea un frente de ondas que se propaga en una dirección determinada, diferente del eje principal normal al transductor.

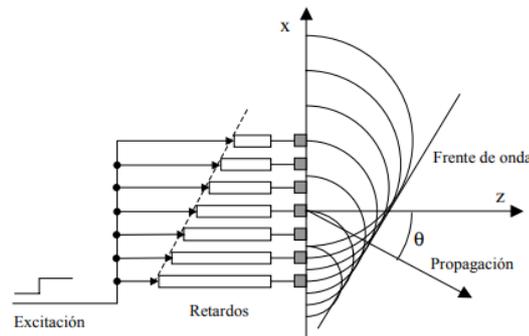


Figura 2.11. Deflexión del haz en emisión con un array lineal [10].

Para este proceso se considera cada elemento como un emisor puntual. Cuando estos tienen una dimensión finita d , las ondas individuales no pueden considerarse esféricas hasta una profundidad mucho mayor que la de su propio campo cercano, esto es $d^2/4\lambda$. Teniendo en cuenta que la distancia entre elementos es $d \approx \lambda/2$ las ondas pueden considerarse esféricas a partir de una distancia de $\lambda/16$ en casi todos los casos.

Si se invierte la dirección de propagación, puede realizarse una deflexión en recepción retrasando y sumando las señales recibidas. El encargado de realizar este proceso es el conformador de haz, como se verá posteriormente.

Focalización electrónica del haz ultrasónico

En los sistemas de imagen por ultrasonido no solo es importante la deflexión del haz, sino también la focalización, ya que las ondas esféricas se propagan en la región de campo cercano.

La profundidad a la que se encuentra el foco natural del array coincide con la distancia de campo lejano, sin embargo, operar en la región de campo cercano presenta la ventaja adicional de poder desplazar el foco natural del *array* a posiciones más cercanas a este, consiguiendo así concentrar toda la energía emitida en una zona, cuyas características vienen determinadas por la geometría del *array*.

A este proceso se le conoce como focalización del haz, de manera que, si se controlan los retardos aplicados en emisión o en recepción, el haz ultrasónico puede enfocarse en cualquier punto dentro del campo cercano de la apertura D , hasta una profundidad $D^2/4\lambda$. La Figura 2.12 muestra el proceso en emisión.

En este ejemplo, los retardos aplicados son simétricos alrededor del elemento central, para producir un foco en el eje del *array*. La idea es retrasar el disparo de cada elemento para que los frentes de onda individuales coincidan en el tiempo sobre el foco.

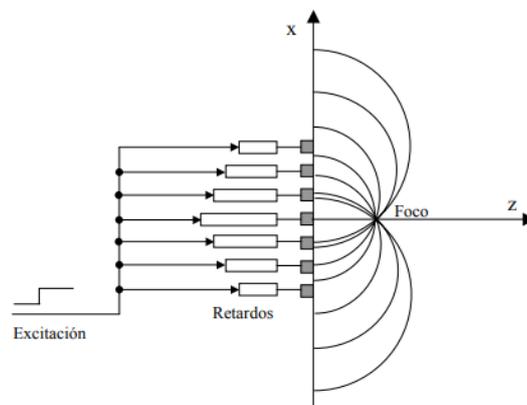


Figura 2.12. Focalización del haz en emisión en el eje con un array lineal [10].

En la Figura 2.12 se observa como un reflector puntual situado en cualquier punto de los frentes de onda fuera del foco produce una señal cuya amplitud es como mínimo $1/N$ de la que se produce cuando ese reflector se sitúa en el foco, donde confluyen las N ondas [18].

Existen dos diferencias notables entre focalizar en emisión y en recepción. En la práctica, en emisión, una vez producido el disparo de los elementos el foco no puede ser

modificado puesto que la onda ya se está propagando (foco fijo). En cambio, en recepción se abre la posibilidad de realizar una focalización dinámica, de forma que el foco vaya siguiendo al frente de ondas en su propagación.

Cálculo de los retardos de focalización y deflexión para una lente esférica

Se considera un array de N elementos con un espacio $d = \lambda/2$ entre estos y la focalización y deflexión en el punto (r, θ) , donde r es la distancia al centro del array y θ es el ángulo de deflexión. Colocando el origen de coordenadas en el centro del array, el retardo relativo al elemento n viene dado por la Ecuación (2.8) [1],[10].

$$\tau_n = (r - r_n)/c \quad (2.8)$$

donde r_n es la distancia entre el punto (r, θ) y el elemento n .

De acuerdo con el teorema del coseno, la distancia desde el punto al elemento n -ésimo es:

$$r_n = \sqrt{x_n^2 + r^2 - 2x_n r \sin\theta} \quad (2.9)$$

siendo x_n la posición del elemento n -ésimo.

El retardo unidireccional para el elemento n (Ecuación (2.10)) se obtiene sustituyendo la Ecuación (2.9) en la Ecuación (2.8):

$$\tau_n = \frac{1}{c} \left(R - \sqrt{x_n^2 + r^2 - 2x_n r \sin\theta} \right) = \tau_n^d + \tau_n^f \quad (2.10)$$

siendo τ_n^d el retardo debido a la deflexión y τ_n^f el de focalización. Sus expresiones respectivamente se encuentran representadas en (2.11) y (2.12).

$$\tau_n^d = \frac{x_n \sin\theta}{c} \quad (2.12)$$

$$\tau_n^f = \frac{r - \sqrt{x_n^2 + r^2 - 2x_n r \sin\theta} - x_n \sin\theta}{c} \quad (2.11)$$

La zona del foco es más estrecha cuanto más grande es el *array* y más corta cuanto más cerca de este nos encontremos. A mayor distancia, las ondas generadas comienzan a aplanarse, aumentando la profundidad del foco. El límite es la distancia de campo lejano, a partir de la cual no es posible focalizar el haz, tan solo deflectorlo [1].

Esto se visualiza en la Figura 2.13, en la cual se aprecia como la totalidad de la energía emitida por el *array* queda ahora concentrada en una pequeña región del espacio, consiguiendo así imágenes con mayor contraste a esta profundidad. Vemos, además como el haz tiende a estrecharse hasta alcanzar el foco, momento a partir del cual comienza a divergir.

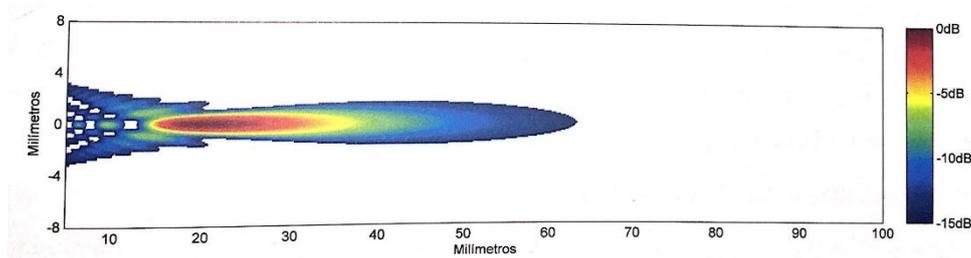


Figura 2.13. *Focalización con array lineal* [1].

En las ecuaciones anteriores se observa que el retardo de focalización para un ángulo de deflexión determinado varía con la distancia. La principal ventaja de la focalización es que se mejora la resolución lateral. De acuerdo con la forma en que se actualice dicho retardo se tratará de foco fijo o foco dinámico.

- **Foco fijo:** se establece un único retardo a una distancia concreta r . Un punto habitual para colocar el foco es la mitad de la región de interés. Alrededor del foco la resolución lateral es máxima, disminuyendo a medida que nos alejamos de él. Es la focalización utilizada en emisión.
- **Foco dinámico:** el retardo se ajusta en función de la distancia r . Es la focalización utilizada para recepción, de forma que los ecos procedentes de cualquier profundidad están continuamente en el foco.

2.5.2. Metodología del Conformador de haces ultrasónicos

La función del conformador de haces ultrasónico es alinear las señales correctamente y sumarlas en fase, de forma que se refuerce la señal respecto al ruido o las ondas que se propagan en otras direcciones diferentes a la de inspección [19].

En los sistemas de imagen ultrasónica al utilizar un rango corto para poder obtener imágenes de órganos humanos, la focalización es esencial. El conformador de haces de retardo y suma es el único método de conformación de haces adecuado para focalizar el haz a una distancia cercana. Es por ello por lo que en este capítulo tan solo se describe esta técnica.

Retardo y Suma

La formación de haces mediante retardo y suma (DAS, *Delay and Sum*) es el algoritmo de procesamiento de señales en *arrays* más antiguo y simple que se conoce.

La idea en la que se centra el DAS es simple, consiste en retrasar y combinar adecuadamente las señales recibidas con cada uno de los elementos del *array* en función a los retardos de deflexión que se desea aplicar, pudiendo de esta manera conformar y orientar el haz en recepción. Además, va a permitir reforzar la señal con respecto al ruido o las ondas que se propagan en otras direcciones. El encargado de realizar este proceso es el conformador de haz.

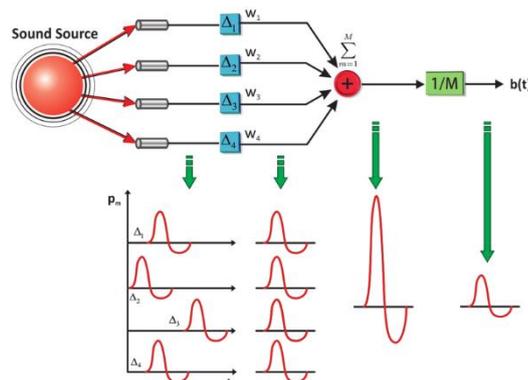


Figura 2.14. Esquema simplificado del DAS Beamforming [20].

Considerando un *array* de N elementos equiespaciados a una distancia d a lo largo del eje x , cuyo centro coincide con el eje de coordenadas, la localización del elemento n -ésimo viene determinada por la Ecuación (2.13) [10].

$$x_n = \left(n - \frac{N-1}{2}\right)d, \quad \text{para } 0 \leq n \leq N-1 \quad (2.13)$$

Si suponemos que la onda recibida por el elemento n -ésimo es $s_n(t) = f(x_n, t)$, la formación del haz consiste en aplicar un retardo τ_n a la salida de cada elemento y sumar las señales resultantes de forma que la salida del conformador de haz es:

$$S(t) = \sum_{n=0}^{N-1} w_n y_n(t - \tau_n) \quad (2.14)$$

Siendo w_n la ponderación de amplitud, también denominada apodización, pudiendo ser cualquier valor entre 0 y 1.

Métodos de control del haz

En general, un patrón de haz deseable es aquel que tiene un lóbulo principal estrecho y lóbulos laterales pequeños en relación con el lóbulo principal. Como se ha visto en el apartado anterior, la focalización dinámica es una forma de obtener un haz colimado. Sin embargo, hay métodos para mejorar la focalización dinámica en imagen médica ultrasónica como la apertura dinámica y la apodización [18].

Con la focalización dinámica, la relación entre el tamaño del *array* y la profundidad es mucho mayor a distancias cortas, lo que significa que la resolución lateral es menor en las proximidades del *array* que en las zonas más alejadas. Para obtener una resolución lateral uniforme a lo largo de la imagen, la cantidad de elementos utilizados como receptores varía en función de la profundidad haciendo que el tamaño del *array* se reduzca en profundidades cercanas en comparación con las profundidades lejanas. De manera que, esta variación del tamaño de la apertura con la profundidad es a lo que se le llama apertura dinámica.

El término apodización se emplea para mejorar la forma del haz y reducir los lóbulos laterales a costa de ensanchar ligeramente el lóbulo principal, obteniendo una mejora del rango dinámico, pero perdiendo resolución espacial. La reducción de la amplitud se realiza aplicando diferentes tipos de ventanas de apodización, como Hamming o Hanning. En la Figura 2.15 se muestra el patrón de un haz de un *array* de 128 elementos con y sin la ventana de apodización. Cuando se aplica la ventana, los lóbulos secundarios del patrón del haz están reducidos, pero el lóbulo principal se vuelve más ancho. Por lo tanto, existe una compensación entre la anchura del lóbulo principal y los lóbulos laterales.

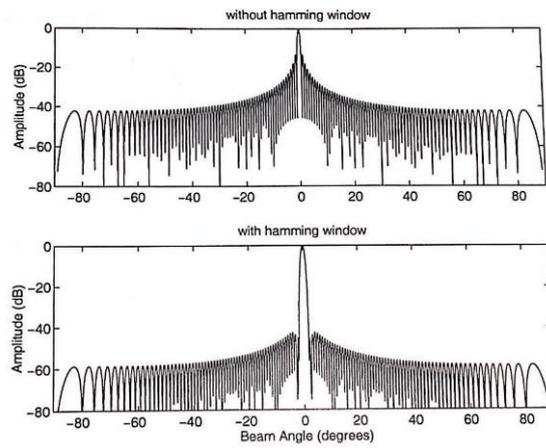


Figura 2.15. *Ejemplo de aplicación de diferentes tipos de ventanas para reducción de la amplitud [10].*

Capítulo 3

Machine Learning y Deep Learning aplicado a imagen ultrasónica

En este capítulo tres se lleva a cabo una explicación detallada acerca de las técnicas de *Machine Learning* y *Deep Learning*. Además, se reúne información relevante del estado del arte en la aplicación de los métodos descritos de inteligencia artificial en el ámbito de la imagen ecográfica.

3.1. Inteligencia artificial en US: Motivación

En el aspecto tecnológico, las sondas de ultrasonido se están volviendo cada vez más compactas y portátiles y la demanda de mercado de dispositivos pequeños de bajo coste se está expandiendo. Al mismo tiempo existe una gran tendencia hacia las imágenes en 3D y el uso de esquemas de imágenes de alta velocidad en fotogramas; ambos acompañados de un aumento drástico de las tasas de datos que suponen una pesada carga para la comunicación del sistema de sonda y los algoritmos de reconstrucción de imágenes posteriores [5].

Con la gran demanda actual de reconstrucción imágenes de alta calidad, las innovaciones en imágenes de ultrasonido modernas se basan en gran medida en el procesamiento de canales de recepción potentes. En este punto aparecen la inteligencia artificial y los métodos de aprendizaje profundo como estrategias que pueden desempeñar un papel convincente en este proceso.

Desafortunadamente como ya se ha comentado, las técnicas que ofrecen imágenes de alta calidad como la RM y CT por Rayos X presentan grandes inconvenientes en cuanto a la radiación ionizante que producen. Por el contrario, la ecografía es una gran herramienta de diagnóstico debido a ser mínimamente invasiva y de bajo coste; permitiendo además un alto grado de interactividad en tiempo real. Sin embargo, su principal desventaja es su reducida calidad de imagen en comparación con las modalidades anteriormente mencionadas.

Es importante tener en cuenta que la reconstrucción de imágenes de alta calidad a partir de los datos obtenidos resulta esencial para el diagnóstico por imágenes, facilitando un diagnóstico temprano, tratamientos adecuados y la posibilidad de alcanzar la personalización en la atención al paciente. Es por ello por lo que en los últimos años se ha producido un gran aumento de las herramientas utilizadas para analizar grandes cantidades de datos de manera eficiente, dando lugar a que el aprendizaje automático haya tomado un importante papel en innovación sanitaria [5].

Los sistemas desarrollados en el ámbito de la salud que utilizan aprendizaje automático deben ser fácilmente implementables, escalables, precisos, robustos y estables. Estos sistemas, aplicados en el ámbito que nos ocupa, han demostrado ser una herramienta poderosa para ayudar a los médicos en su ruta clínica diaria.

3.2. Introducción al ML

El ML es una técnica de aprendizaje automático que nos va a permitir abordar un problema aprendiendo de la experiencia, representada a través de observaciones. Requiere de la definición inicial de un vector de características que permitan describir la imagen ecográfica. A la extracción de características le sigue el entrenamiento del modelo sobre estas, como por ejemplo, la disposición espacial de los puntos de interés.

Dentro del campo del aprendizaje automático, se distinguen a menudo dos tipos principales de tareas: supervisadas y no supervisadas. La mayoría de los estudios realizados en el área del US dirigidas a mejorar los resultados de las imágenes obtenidas utilizan técnicas de aprendizaje supervisado.

El objetivo del aprendizaje supervisado es la deducción de una función a partir de los datos de entrenamiento etiquetados. Las dos técnicas supervisadas principales son la

clasificación y la regresión, que se diferencian por tener como resultado variables categóricas y continuas, respectivamente [22]. Mientras que el aprendizaje no supervisado tiene como objetivo descubrir estructuras ocultas en los conjuntos de datos que no se están etiquetados. El método de aprendizaje no supervisado más común es la agrupación o *clustering*, utilizada para agrupar datos con una estructura similar.

Previo al diseño del modelo tiene lugar la etapa de preparación de los datos, en la que el objetivo principal es manipular y transformar los datos originales para que la información contenida en ellos se encuentre accesible de forma más sencilla e interpretable.

Estos datos con los que nos quedamos finalmente se van a dividir en 3 subconjuntos: entrenamiento (*train*), *test* y validación. El conjunto de *train* nos permite definir el diseño y los parámetros del modelo, mientras que el de *test* evalúa las prestaciones finales de este. Por otro lado, el conjunto de validación se utiliza para evaluar su capacidad de generalización.

La generalización es un factor muy importante a la hora de definir el tipo de modelo que se requiere, ya que es de suma importancia obtener una buena generalización, evitando tanto el sobreajuste como el subajuste.

- **Sobreajuste (*overfitting*):** se ajusta en exceso a los datos de entrenamiento, lo que resulta en una mala capacidad para clasificar a pacientes nuevos (conjunto de *test*). Es un modelo muy complejo que genera pocos errores en entrenamiento.
- **Subajuste (*underfitting*):** modelo simple con muchos errores en entrenamiento. El algoritmo de aprendizaje prestará más atención a construir una frontera suave que a reducir el error entre la salida del modelo y el valor deseado.
- **Adecuada generalización:** se necesita un compromiso entre la complejidad del modelo y los errores que cometen, para ello se recurren a técnicas de validación cruzada. En [23] se puede obtener una explicación detallada para poder llevar a cabo este tipo de técnica.

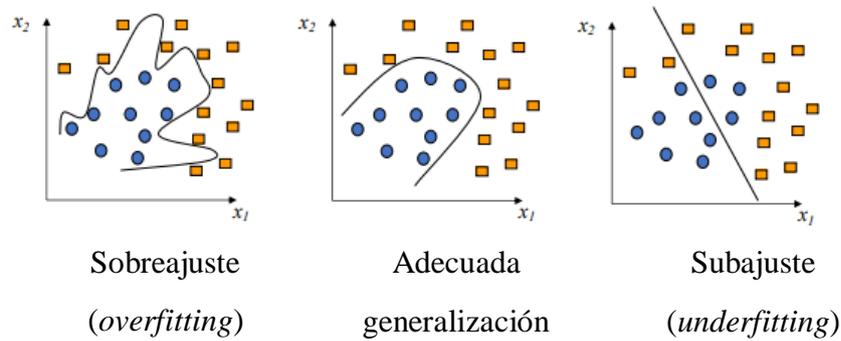


Figura 3.1. *Overfitting: excesivo ajuste en los datos de entrenamiento. Underfitting: modelo excesivamente simple* [24].

Además, por otro lado, podemos contar con métodos paramétricos, que proponen un modelo en forma de una función parametrizada y métodos no paramétricos. Estos no realizan ninguna suposición sobre la forma funcional del modelo, en su lugar buscan un modelo que se aproxime lo mejor posible a los datos (sin subajustar ni sobreajustar). Este último tipo de modelos son los que utilizaremos en el desarrollo del proyecto.

3.3. Deep Learning

Los modelos de DL, como las redes neuronales artificiales (RNA), están diseñados para extraer automáticamente las características óptimas para realizar una tarea.

Un modelo matemático sencillo de neurona consiste en una combinación lineal de las entradas, seguida de una función no lineal, como se visualiza en la Figura 3.2.

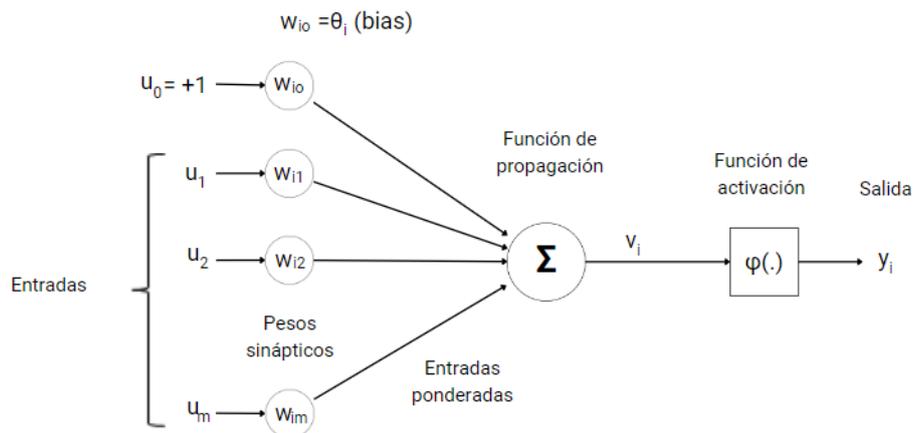


Figura 3.2. Modelo de neurona estándar con entradas ponderadas, función de propagación, función de activación y su salida correspondiente.

Sus componentes son:

- **Entradas ponderadas (u):** hacen referencia a las conexiones sinápticas de las neuronas cerebrales, donde el peso de la conexión (w) equivale a la fuerza o efectividad de la sinapsis. El valor de los pesos y el signo de estos definen el tipo y la intensidad de la influencia de una entrada sobre otra.
- **Función de propagación:** calcula el valor de entrada total, como una simple suma ponderada de todas las entradas recibidas multiplicadas por sus pesos.
- **Función de activación (g):** define el comportamiento de la neurona. Existen diferentes tipos de funciones, desde funciones de umbral a funciones no lineales. Permiten calcular el nivel o estado de activación de la neurona en función de la entrada total.
- **Salida (\hat{y}):** calcula la salida de la neurona en función de la activación de esta.

En la Ecuación (3.1) se define matemáticamente el modelo de neurona estándar, donde m es el número de entradas e i la capa oculta a la que pertenece cada uno de los pesos.

$$\hat{y} = \varphi\left(\sum_{i=1}^m u_i w_{im}\right) \quad (3.1)$$

Generalmente estos modelos incluyen un término conocido como *bias* (θ_i), interpretado como el valor que la neurona debe superar para activarse. Supone un

valor de sesgo el cual permite desplazar la función de activación para garantizar un aprendizaje exitoso. La Ecuación (3.1) extendida añadiendo el valor de *bias* se encuentra en la Ecuación (3.2).

$$\hat{y} = \varphi(\vartheta_i + \sum_{i=1}^m u_i w_{im}) \quad (3.2)$$

La neurona presenta la capacidad de aprendizaje modificando sus parámetros. El elemento esencial de las neuronas es su organización en capas: de entrada, salida y ocultas, de manera que el conocimiento se encuentra almacenado y distribuido en los pesos de la red.

Por lo tanto, el procedimiento de aprendizaje de la red (denominado entrenamiento), consiste en ajustar los parámetros libres para aproximar la función que modela la relación entre los vectores de características x y las salidas deseadas y , de forma que el conjunto S ($S = \{(x^{(i)}, y^{(i)})\}_{i=1}^N$) debe ser lo suficientemente representativo de la tarea a resolver.

Atendiendo al flujo de los datos, se distinguen redes unidireccionales (*feedforward*, *FNN*, *Feedforward Neural Networks*), donde la información fluye en un único sentido, desde las entradas ponderadas hasta obtener la salida y a través de cálculos intermedios f , a los que se denominan capas. Para estos modelos no hay conexiones de retroalimentación en las que las salidas del modelo se retroalimentan a sí mismo. Cuando esto ocurre se las denominan redes neuronales recurrentes (*feedback*).

El proceso de aprendizaje se concreta en la optimización de una función de coste o una función de error ($C(y, \hat{y})$) construida a partir del conjunto de aprendizaje. Esta función define la manera en la que la red neuronal va a aprender y por ello es de vital importancia elegirla bien. De manera simple, la función de coste define de una forma matemática la relación que existe entre el resultado que la red consigue (\hat{y}), y el resultado objetivo o *target* (y).

El fin último es encontrar una configuración de pesos que corresponda a un mínimo de la función de coste propuesta, esto se realiza por un algoritmo iterativo denominado descenso por gradiente. De manera que los valores que se guardan son aquellos que representan la combinación lineal existente entre las neuronas de una capa y cada una de las neuronas de la capa siguiente.

Como se expresa en la Ecuación (3.3) el tipo de método de optimización escogido y aplicado a la función de coste proporciona una regla de actualización de los pesos. En función de los casos usados en el entrenamiento, la regla modifica iterativamente los pesos hasta alcanzar una configuración adecuada.

Para ello, calcula la derivada parcial respecto a cada parámetro en el punto de evaluación. La derivada indica el valor y sentido en que se encuentra el mínimo más próximo (puede ser un mínimo local o global, el método no lo puede diferenciar). El resultado de la derivada se le resta a cada uno de los parámetros multiplicado por la tasa de aprendizaje (valor entre 0 y 1).

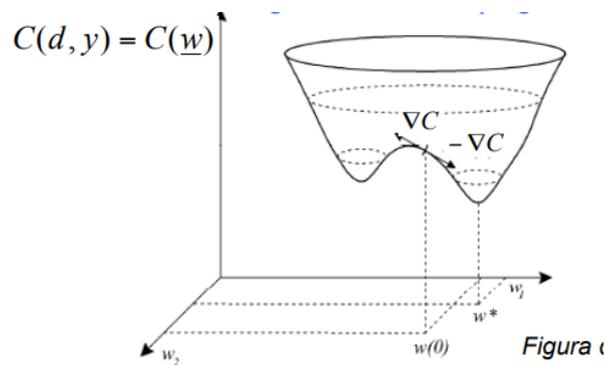


Figura 3.3. Representación gráfica del proceso que lleva a cabo el algoritmo de descenso de gradiente [24].

$$w_e^{(iter+1)} = w_e^{(iter)} - \eta \nabla_{w_e} C \quad (3.3)$$

Para la Ecuación (3.3) $w_e^{(iter)}$ son los pesos de la red en cada una de las iteraciones y η la tasa de aprendizaje.

La tasa de aprendizaje o *learning rate* constituye un importante parámetro dentro del entrenamiento de las redes neuronales. Permite variar la rapidez de actualización de los pesos. Es muy importante la selección de un valor adecuado ya que, un valor demasiado bajo puede provocar que nunca se alcance el mínimo, mientras que uno demasiado alto podría saltárselo.

Esta actualización de los pesos se realiza gracias a la función de coste mediante la técnica de *backpropagation*, encargada de realimentar la red neuronal desde el final al inicio con los datos obtenidos de la función de coste y mediante ellos actualizar los pesos en cada capa a fin de conseguir, por lo general, un menor coste en la siguiente iteración.

Esta técnica de *backpropagation* se realiza por la regla de la cadena de cálculo, donde se sigue un orden de operaciones específico para ser lo más eficiente posible. Suponemos x como un número real y definimos f y g como funciones que mapean de un número real a otro. Si $y = g(x)$ y $z = f(g(x)) = f(y)$ la regla de la cadena se visualiza en la Ecuación (3.4).

$$\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx} \quad (3.4)$$

La red neuronal más simple utilizada es el perceptrón (MLP). Funciona como un tipo de neurona artificial, ya que solo está formada por una capa, pero por su capacidad para construir aproximadores universales de funciones es el modelo de red más utilizado con un esquema FNN.

Además, existen arquitecturas con muchas capas, a las que se les consideran redes neuronales profundas (DNN, del inglés *Deep Neural Networks*), entre las cuales las más utilizadas son las RNA y las redes neuronales convolucionales (CNN, *Convolutional Neural Networks*).

Estos modelos generalmente se representan componiendo de forma conjunta muchas funciones diferentes. Siguiendo una estructura de cadenas, representada en la expresión (3.5).

$$f(x) = f^{(i)}(f^{(i-1)}(f^{(i-2)} \dots (x))) \quad (3.5)$$

Para este caso, f^1 recibiría el nombre de primera capa de la red, f^2 la segunda capa y así sucesivamente.

En cuanto a las CNN sabemos que una convolución consiste en la transformación de una región de la imagen mediante un filtro específico. Estas son capaces de extraer características e información teniendo en cuenta la relación de cada uno de los puntos de la imagen con su entorno.

Se realiza a partir de la combinación lineal de los píxeles de la imagen original multiplicados por los pesos de una máscara previamente definida. Cada peso de esta máscara define la importancia de cada píxel en la imagen. Esto se encuentra expresado en la Ecuación (3.6), siendo x los valores de la imagen de entrada, w la máscara o *kernel* y t el tiempo. Para el caso que se describe el tiempo se encuentra discretizado, ya que al

trabajar con datos en ordenadores solo puede tomar valores enteros. De manera que, si además se asume que x y w también están definidos solo por valores enteros, la convolución discreta se puede expresar como la Ecuación (3.7) siendo a la posición de cada punto en la imagen.

La representación gráfica del proceso de convolución en una imagen se visualiza en la Figura 3.4.

$$s(t) = (x * w)(t) \tag{3.6}$$

$$s(t) = \sum_{a=-\infty}^{\infty} x(a)w(t - a) \tag{3.7}$$

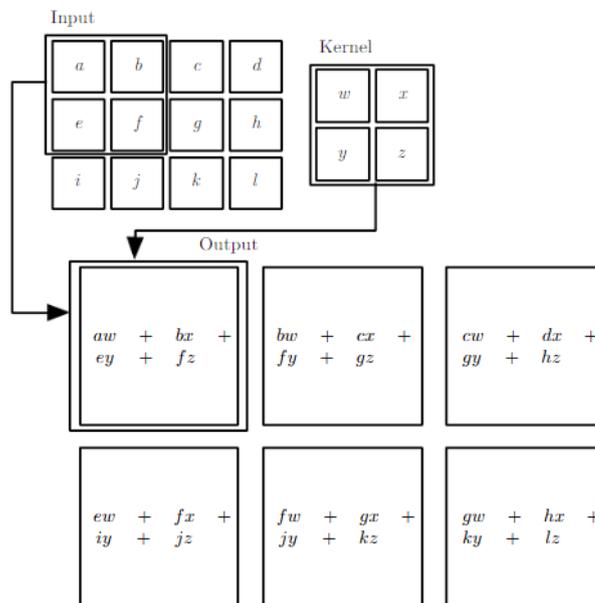


Figura 3.4. Ejemplo de convolución en 2D. Se dibujan cuadros con flechas para indicar cómo se forma el elemento superior izquierdo de la salida aplicando el kernel a la región superior izquierda correspondiente a los datos de entrada [26].

Generalmente en las aplicaciones de ML la entrada suele ser una matriz multidimensional de datos y el *kernel* una matriz multidimensional de parámetros que se adaptan mediante el algoritmo de aprendizaje.

El algoritmo de aprendizaje que se sigue para entrenar una CNN completa es un *forward pass* que calcula la salida a partir de una entrada y un *backbardware pass* para actualizar los pesos de la red. Se propaga la red por una o varias convoluciones hasta que

se ha extraído toda la información necesaria. Esta información se “aplana” (capa *flatten*) y se introduce en una red neuronal de las previamente definidas a modo de vector a partir de una capa *fully connected*, obteniendo una salida que se comparará con el valor real [21], [27].

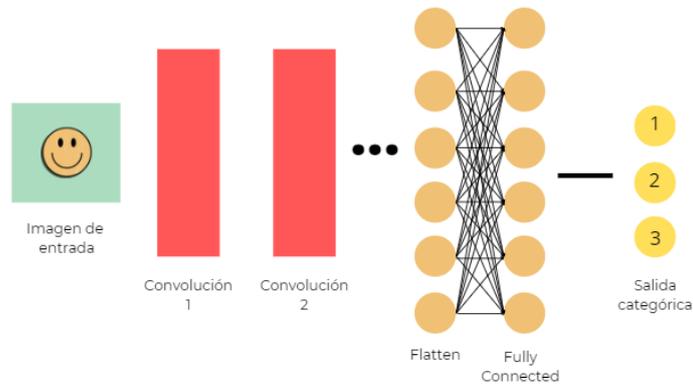


Figura 3.5. *Arquitectura CNN con dos convoluciones sucesivas y capa fully connected para resolver un problema de clasificación.*

Estas técnicas de aprendizaje profundo descritas se benefician de las grandes cantidades de datos que se utilizan para ser entrenadas y de metodologías optimizadas para tratar eficientemente los datos.

3.3.1. DL aplicado a imagen ultrasónica

En el ámbito que nos concierne diversos estudios utilizan modelos de DL. Concretamente RNA, debido a su capacidad de aprendizaje para actuar como poderosos agentes artificiales y procesadores de señales para mejorar la resolución y el contraste de las imágenes, en combinación con CNN obteniendo características relativas a la información espacial de las imágenes.

En [16] y [27] se puede ver el uso de DNN para realizar un aprendizaje en la formación de haces ultrasónicos rápidos y de alta calidad con muy pocos datos de entrenamiento, aplicándose a técnicas de onda plana y de apertura sintética, a lo que denominan técnica ABLE (del inglés *Adaptive Beamforming by Deep Learning*). Supone el comienzo hacia la implementación en tiempo real de un filtro adaptativo de formación de haces en sistemas de ultrasonido.

3. ML Y DL APLICADO A IMAGEN US

Debido a la importancia de obtener imágenes de alta calidad, en [29] se plantea el uso de dos CNN profundas para realizar una evaluación de la calidad de imágenes fetales mediante la implementación de un control de calidad en el examen clínico obstétrico. Para ello se generan una serie de puntuaciones a partir de anotaciones dadas por médicos expertos y se aplican a las redes mencionadas. Debido a la escasez de muestras etiquetadas se utiliza un método conocido como *transfer learning* (TL), el cual consiste en transferir el conocimiento adquirido por una arquitectura de DL entrenada en un conjunto de datos para su posterior utilización en otro conjunto diferente de datos y en una aplicación práctica relacionada pero diferente [30].

Por otro lado, una de las mejoras de la calidad de imagen ecográfica consiste en la eliminación del ruido y en la mejora de la resolución, muchas veces perdida debido al *speckle* [4], una textura granulosa estacionaria tratada como un ruido indeseable debido a la degradación que produce en las imágenes de diagnóstico clínico de modo B. En este trabajo se plantea el uso de simulaciones lineales simples de imágenes de ultrasonido junto con dos CNN para entrenar y aprender a generar un formador de haces que permita reducir el moteado.

Por todo ello, actualmente se encuentra en estudio el impacto del aprendizaje profundo en una gran variedad de aspectos de imágenes de ultrasonido, poniendo especial interés en los métodos que explotan tanto el poder de los datos como la estructura de señal para producir soluciones robustas y eficientes de los datos.

Capítulo 4

La imagen de apertura sintética

El siguiente capítulo describe en líneas generales el proceso de conformación de haces y la focalización de estos como antesala al desarrollo de las técnicas de apertura sintética, en concreto de la técnica SAFT, utilizada para la formación de imágenes ultrasónicas.

4.1. Apertura sintética

El desarrollo de sistemas ecográficos de alta calidad y con buenas prestaciones pasa por el empleo de transductores ultrasónicos basados en *arrays*. El uso de estos dispositivos permite componer aperturas de mayor tamaño, consiguiendo haces más estrechos que mejoran la resolución lateral del sistema, pero a su vez lleva asociado un elevado coste, tanto en tamaño y número de recursos hardware requeridos, como en el tiempo necesario para la adquisición y el procesamiento de los datos.

Por otro lado, otro requerimiento fundamental de este tipo de sistemas es la adquisición de imágenes en tiempo real. En este caso, la principal limitación viene impuesta por el tiempo que tarda el sonido en propagarse por el medio.

Todo ello dificulta la elaboración de sistemas compactos, de bajo consumo, que puedan operar en tiempo real. En este sentido, las técnicas de apertura sintética han demostrado ser un método eficaz para alcanzar dichos objetivos, minimizando el tamaño de los sistemas y acelerando la captura de imágenes [31].

4.1.1. Concepto de apertura sintética

Como ya se ha comentado, en la ecografía médica convencional se utiliza un *array* de elementos para generar el haz de transmisión. Los sistemas de imagen ultrasónica utilizan entre 64 y 128 canales de emisión/recepción. Para reducir el coste del sistema, el número de canales debería reducirse, y las técnicas de apertura sintética (SAU, *Synthetic Aperture Ultrasound*) son un buen método para ello.

Estas técnicas SAU surgen como un intento de aplicar en los sistemas ecográficos las técnicas de procesamiento desarrolladas en *Radar* y *Sonar*. Estos sistemas se basan en la emisión y recepción secuencial con una antena que se desplaza a velocidad constante sobre la región de interés [32].

Apertura Efectiva

El concepto de apertura efectiva se ha aplicado al análisis del patrón del haz. Las ecuaciones y desarrollo para determinar cuál es el número adecuado de elementos para este tipo de aperturas se encuentran en [33]. Obteniendo finalmente que el número de elementos utilizados en apertura efectiva, n_{TR} es:

$$n_{TR} = n_T + n_R - 1 \quad (4.1)$$

siendo n_T el número de elementos que transmiten y n_R el número de elementos que reciben.

4.1.2. Técnica SAFT

Actualmente existen diversos métodos propuestos de apertura sintética para imagen ultrasónica, pero para este trabajo nos centraremos específicamente en uno, denominada técnica de enfoque de apertura sintética (SAFT, *Synthetic Aperture Focusing Technique*) [34].

Fue la primera técnica de apertura sintética en aparecer. Se basa en la activación secuencial en emisión y recepción de cada elemento del *array* (como se visualiza en la Figura 4.1.). De forma que, cada elemento activo transmite un pulso corto, las señales

reflejadas por el objeto se reciben en este mismo elemento y se almacenan. Posteriormente se aplica una etapa de post-procesamiento para componer la imagen final. Este proceso se repite para los N elementos que conforman el *array*. Reduciendo la complejidad debido a que solo se necesita un único conjunto en el circuito para transmitir y recibir.

Para poder llevar a cabo dicha operación, se debe de realizar el proceso de *beamforming* comentado anteriormente, el cual focaliza dinámicamente cada punto de la imagen en emisión y recepción consiguiendo la máxima calidad a todas las profundidades.

Las técnicas SAFT además de reducir el coste y el *hardware* asociado al *array*, permiten lograr una mayor resolución espacial, debido a que el enfoque dinámico completo se realiza tanto en transmisión como en recepción, logrando de esta manera una mejora de la resolución lateral en una amplia zona focal y en el contraste de las imágenes ultrasónicas al reducir los efectos de la dispersión [35]. Además, este tipo de técnicas también permitirán acelerar el proceso de captura de imágenes, ya que al emplear tan solo un elemento para transmitir el tiempo total de la captura de la imagen será proporcional al tamaño del *array*.

Desafortunadamente, como principales desventajas de este tipo de técnicas podemos citar:

- Baja relación señal a ruido (SNR, *Signal to Noise Ratio*) debido a la pequeña apertura activa (elemento único). Al emitir con pocos elementos la energía radiada no es muy grande, reduciendo la capacidad de penetración en el material. De la misma manera, al recibir pocos elementos, el número de señales que se combinan para obtener una línea de imagen también es menor. Por lo tanto, cuanto mayor es la reducción de *hardware*, menor es la SNR.
- Baja resolución de contraste debido a los altos valores de los lóbulos laterales en el patrón de radiación, aun cuando los elementos del *array* se encuentran separados una distancia inferior a media longitud de onda.
- Grandes exigencias de cómputo, dado que para estas técnicas la composición de imagen se realiza una vez que todas las señales han sido ya adquiridas, por lo que es necesario disponer de un sistema de procesamiento rápido que pueda manejar todos los datos sin introducir retardos adicionales en el sistema.

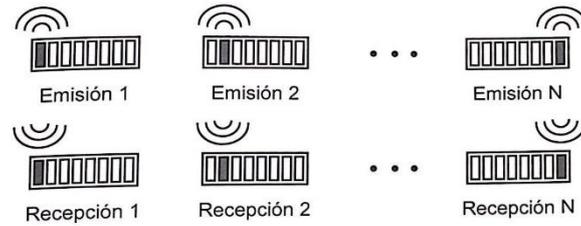


Figura 4.1. *Secuencia de activación de los elementos del array en un SAFT.*

4.1.3. El Coarray

El *coarray* es una herramienta matemática utilizada como solución para el diseño de sistemas que operen en tiempo real, entre los cuales presentan problemas como los descritos anteriormente para las técnicas SAFT [36]. Van a permitir planificar el modo de operación y evitar la aparición de los lóbulos secundarios trabajando con un número reducido de señales.

Como punto de partida es importante tener en cuenta que, el campo recibido en el plano del *array* y la distribución de amplitud de cada uno de los elementos de este se encuentran relacionados por la Transformada de Fourier (TF).

Para este tipo de técnicas relacionadas con la TF la característica básica es lo que se conoce como función de dispersión de punto (PSF), la cual consiste básicamente en la respuesta del sistema de formación de imágenes ante un punto fuente, jugando un papel básico en la obtención de características de la imagen.

Para cada tipo de imagen, la transformada inversa de la PSF actúa como una función de ponderación de las frecuencias espaciales de la distribución original [37]. Por lo tanto, cualquier influencia que se pueda tener sobre la calidad de la imagen estimada final, será únicamente a través del control que se pueda ejercer sobre las características de la PSF. En gran medida, las características de la transformada inversa de la PSF para cualquier tipo de imagen están determinadas por la geometría de la apertura.

Para el ámbito que nos ocupa el *coarray* supone una herramienta matemática utilizada para estudiar el patrón del haz generado en sistemas pulso-eco. Básicamente se trata de la apertura virtual de un sistema que produce, en un sentido, el mismo patrón del haz que el sistema original trabajando en emisión y recepción.

Cálculo del coarray

Suponiendo un *array* lineal de N elementos, siendo a_n los pesos de los transductores. En campo lejano y asumiendo señales de banda estrecha, el patrón de radiación se puede escribir como la expresión (4.2)[36].

$$f(u) = \sum_{n=0}^{N-1} a_n e^{jkx_n u} = \sum_{n=0}^{N-1} a_n (e^{jkd} u)^n \quad (4.2)$$

Donde $u = \sin(\theta)$, y θ es el ángulo medido desde la perpendicular del *array*. Si se realiza el cambio de variable $\Omega = kd \cdot \sin(\theta)$ la expresión anterior se corresponde con la Transformada de Fourier en Tiempo Discreto (DTFT) de la secuencia a_n .

En sistemas de apertura sintética, cada imagen escaneada se obtiene después de una secuencia de disparos de los elementos del *array*. De manera que el *coarray* que se obtiene al aplicar la técnica SAFT es el resultado de sumar un conjunto de subcoarrays (s_n^k), donde cada uno de ellos se obtiene al convolucionar un par de subaperturas en emisión (a_n^k) y recepción (b_n^k), las cuales determinan los elementos activos en cada adquisición. Por lo tanto, habrá por ellos tantas subaperturas en emisión y recepción como elementos tenga el *array*, teniendo cada una de ellas una única muestra distinta de cero, que representa la ganancia del único elemento activo en cada emisión-recepción.

La convolución de estas subaperturas da lugar a N *subcoarrays* de $2N - 1$ muestras también con un único valor distinto de cero. La Figura 4.2. muestra el proceso de cálculo completo para un *array* específico de 8 elementos.

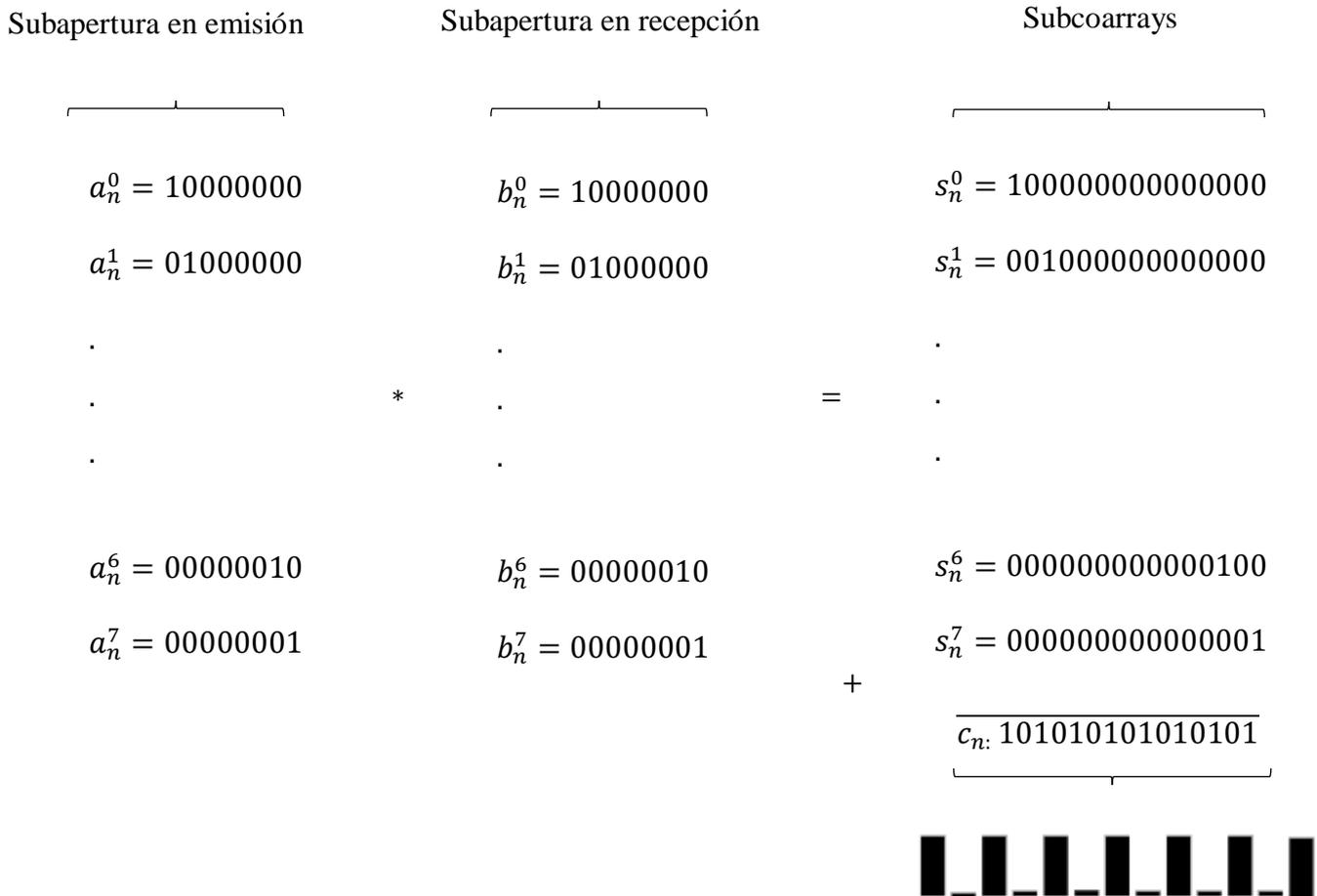


Figura 4.2. Subaperturas, subcoarray y coarray al aplicar un SAFT sobre array de 8 elementos.

Visualizando la Figura 4.2. se comprueba como los elementos activos del *coarray* se encuentran separados una distancia igual al doble de la que hay entre los elementos del *array* original. Es importante tener en cuenta que cuanto más pequeño es el elemento, más alta es la impedancia eléctrica y acústica que presenta, lo que empeora la SNR. Además, también es posible comprobar como el número de elementos activos en el *coarray* coincide con el número de señales recibidas por el sistema.

4.1.4. Características de la imagen SAFT

Para poder generar las imágenes de apertura sintética nos basaremos en un *array* de FMC (*Full Matrix Capture*). Consiste en una estrategia de adquisición de datos que

permite la captura de todas las combinaciones posibles de transmisión – recepción. De manera que, cada elemento del *array* se utiliza sucesivamente como transmisor, mientras que los demás elementos se utilizan como receptores para cada uno de los N disparos correspondientes al número de elementos del *array*. De esta manera se consigue aumentar el número de datos recogidos pudiendo lograr una mejora de la calidad de imagen.

El proceso de la mejora de la resolución espacial se visualiza en la Figura 4.3.

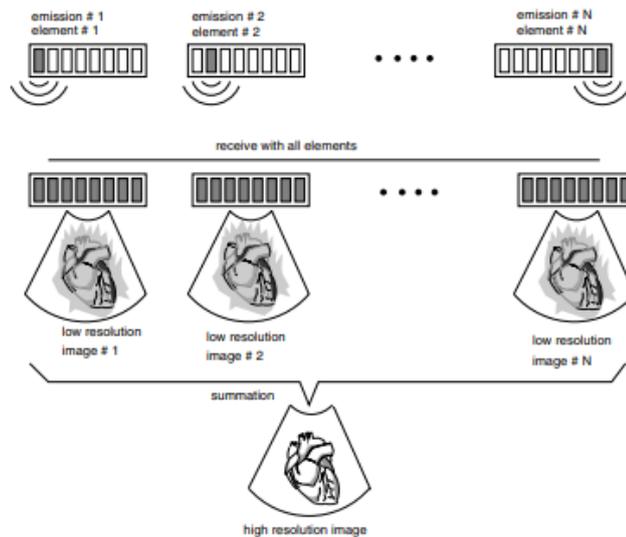


Figura 4.3. Principio básico del SAFT. Para cada iteración emite un elemento del array y todos reciben, sumando al final todos los ecos almacenados obteniendo una imagen de alta resolución [38].

Como sabemos cada elemento del *array* emite siendo todos los elementos los que reciben registrando los ecos del pulso transmitido, en esta etapa la imagen se sintetiza solo durante la recepción, correspondiente al proceso de enfoque dinámico y por lo tanto cada imagen que se va obteniendo es de baja resolución. Después de sucesivas transmisiones, hasta que se activa el último elemento del *array*, todos los ecos almacenados se suman de manera que se produce una imagen de alta resolución, enfocada dinámicamente durante la transmisión y la recepción.

Lo mismo ocurre para la dinámica de contraste, debido a que cada uno de los elementos del *array* emiten un pulso de forma individual, emitiendo pulsos de baja energía [39]. Además, se produce una disminución de la SNR.

En cuanto a la tasa de imagen es importante tener en cuenta que depende fuertemente de las características propias de cada sistema, como por ejemplo, del número de líneas

que es necesario representar, de la manera en la que estas se componen, la velocidad de propagación del sonido en el medio o el retardo introducido por la electrónica.

Uno de los principales parámetros que limita la tasa de imagen es el tiempo requerido para realizar la captura de las señales que conforman cada una de las líneas de la imagen. Este tiempo viene determinado por la velocidad de propagación del sonido en el medio, siendo un parámetro fijo que no se puede variar, y de la electrónica usada, pudiendo intervenir haciendo uso de un *hardware* más rápido, ya que el tiempo de captura de los datos también va a depender del número de disparos que es necesario realizar para obtener todas las líneas de imagen

Por tanto, las técnicas SAFT permiten acelerar este proceso reduciendo el número de disparos totales, ya que solo es necesario emitir y recibir tantas veces como elementos tiene el *array*.

Así pues, si el número de líneas que se desea adquirir es mucho más grande que el número de elementos que contiene el *array*, el SAFT conseguirá una reducción en el tiempo de adquisición y en consecuencia un incremento de la tasa de imagen.

Capítulo 5

Materiales y Métodos

Este quinto capítulo describe las distintas metodologías propuestas y utilizadas y desarrolladas a lo largo del TFG. En primer lugar, se centra en detallar la hipótesis de partida acerca del problema que queremos resolver y un análisis del conjunto de datos que componen la imagen. Se plantean las diferentes propuestas junto a una revisión de cada una de las técnicas de DL planteadas para la resolución del problema. Finalmente se realiza una explicación detallada acerca del desarrollo de las diferentes estrategias de RNA utilizadas basadas en regresión detallando cada uno de los pasos seguidos a la hora de resolver el problema sobre datos sintéticos.

5.1. Hipótesis de partida

Teniendo en cuenta lo descrito en capítulos anteriores, se ha revisado como, para la formación de imagen de US, la conformación de haces mediante la técnica DAS constituye una medida sólida de su reflectividad acústica. Aplicándose a imágenes de apertura sintética siguiendo la técnica SAFT, se pueden obtener imágenes de alta resolución y contraste.

En particular, es importante analizar como la distribución espacial de las muestras recibidas se organizan en el *coarray* y como se lleva a cabo la composición de los lóbulos secundarios. Con esta información es posible desarrollar técnicas de procesamiento de DL orientadas a evitar su formación o a reducir su influencia en la imagen. El trabajo que se presenta está dirigido a solucionar este problema a partir de un *array* lineal de 64 elementos a partir de una estrategia de adquisición en FMC.

5.1.1. Estimación de la reflectividad

Se define el conjunto FMC (\vec{x}) como una colección de muestras usadas para calcular la reflectividad en el punto \vec{x} de la Región de Interés (ROI, *Region of Interest*), a a partir de un *array* lineal de N elementos (64) [39].

Su expresión formal se corresponde a la Ecuación Figura 5.1. , donde \vec{x}_i y \vec{x}_j son las posiciones de los elementos emisores y receptores respectivamente. El valor utilizado para evaluar en \vec{x} la reflectividad correspondiente en la formación de imágenes ultrasónicas por la conformación de haces DAS, es la suma de todo el conjunto FMC (\vec{x}).

$$FMC(\vec{x}): \left\{ s_{ij}(\tau), \tau = \left| \frac{\vec{x} - \vec{x}_i}{c} \right| + \left| \frac{\vec{x} - \vec{x}_j}{c} \right|, \forall i, j = [1, N] \right\} \quad (5.1)$$

Se considera que cada muestra en el conjunto FMC (\vec{x}) se compone por la reflectividad $m_{ij}(\vec{x})$ en ese punto y un componente de ruido $n_{ij}(\tau)$, el cual incluye ruido térmico y toda la información acústica no relacionada con el punto \vec{x} .

$$s_{ij}(\tau) = m_{ij}(\vec{x}) + n_{ij}(\tau) \simeq m(\vec{x}) + n_{ij}(\tau) \quad (5.2)$$

Como la reflectividad en \vec{x} es la misma para los pares emisión-recepción, podemos escribirla como $m(\vec{x})$ (Ecuación (5.2)).

Mediante la técnica DAS comentada en capítulos anteriores, se puede obtener una estimación de la reflectividad como se encuentra representado en la Ecuación (5.3), donde si $n(\vec{x})$ es igual a cero, $\hat{m}(\vec{x}) = m(\vec{x})$.

$$\hat{m}(\vec{x}) = \left| m(\vec{x}) + \sum_{i=1}^N \sum_{j=i}^N \frac{n_{ij}(\tau)}{N^2} \right| = |m(\vec{x}) + n(\vec{x})| \quad (5.3)$$

Generación del espacio de muestras en el Coarray

Como se ha revisado en el capítulo anterior, el *coarray* es una herramienta matemática obtenida a partir de la convolución espacial en emisión $E[n]$ y la apertura de recepción $R[n]$ (Ecuación (5.4))

$$C[n] = E[n] * R[n] \quad (5.4)$$

Redefinimos FMC (\vec{x}) en el *coarray* agrupando las señales como se visualiza en la Ecuación (5.5), teniendo en cuenta que $n = i + j$.

$$FMC(\vec{x})[n]: \left\{ c[n] = \sum_{n=1}^{2N-1} s_{ij}(\tau), \quad i + j = n, \quad \forall s_{ij}(\tau) \in FMC(\vec{x}) \right. \quad (5.5)$$

Desde esta hipótesis planteada podemos empezar a describir los procedimientos para la obtención de las imágenes sintéticas, así como los tipos de datos y el tratamiento que se les realiza. Además, se plantean las diferentes propuestas de resolución del problema mediante redes neuronales.

5.2. Datos sintéticos

Se ha creado un conjunto de datos sintéticos en el entorno de programación *Python* a través de un generador de ondas sintéticas. Este se ha construido mediante pulsos gaussianos y operaciones matemáticas sencillas. Además, el generador cuenta con parámetros controlables de forma manual, pudiendo seleccionar la cantidad de ruido que se quiere introducir en las señales, el número de blancos que se quiere obtener, su distribución y la resolución. Esta posibilidad permite estudiar las propiedades que tienen los distintos métodos y compararlos entre ellos.

A estas imágenes son a las que denominaremos *imágenes originales*. Hacen referencia a las imágenes que se obtendrían directamente en el sistema de visualización de US. En la Figura 5.1. se puede observar la presencia de lóbulos secundarios y las dificultades que estos ofrecen para la visualización correcta de los blancos, principalmente los de menor intensidad.

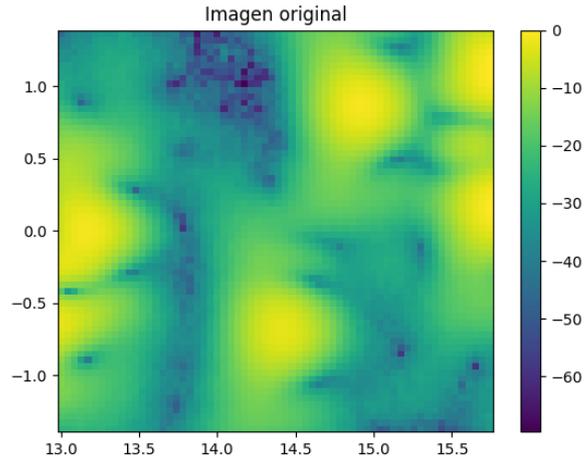


Figura 5.1. *Imagen sintética en dB. Obtenida con parámetros de: $bw = 0.5$, $\lambda = 0.5$ y frecuencia central = 2.*

Para cada imagen, se extraen dos tipos de datos: temporales y frecuenciales, organizados en matrices de 64×64 .

- Datos temporales: se calcula la señal en fase y en cuadratura para cada señal recibida, donde finalmente se obtiene la envolvente como el módulo de ambos valores.
- Datos frecuenciales: divididos en parte real e imaginaria. Aportan información espacial.

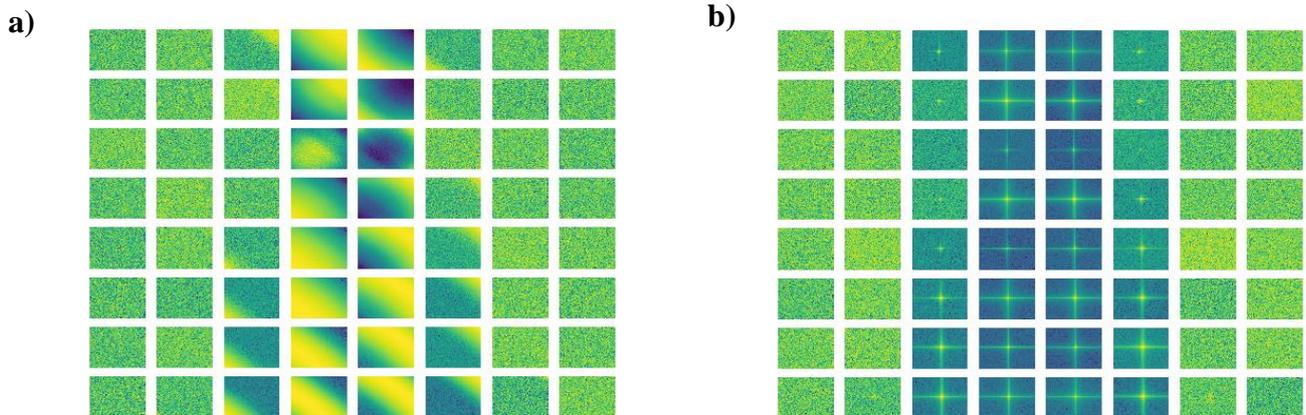


Figura 5.2. *Representación de a) Datos temporales b) Datos frecuencia.*

Como previamente se ha comentado, el interés de este trabajo reside en el uso de técnicas de *Deep Learning* con el objetivo de eliminar los lóbulos existentes en las imágenes ultrasónicas mejorando la calidad de imagen. Para ello se necesita un nuevo conjunto de imágenes a las que se ha denominado *imágenes de referencia*, debido a que son las imágenes ideales que se quieren obtener. En función del tipo de estrategia que se lleva a cabo (comentadas en el siguiente apartado) se obtendrán diferentes tipos de imágenes de referencia, debido a que la salida deseada será diferente.

5.3. Experimentos con Redes Neuronales

En esta sección se plantan distintos experimentos con diferentes algoritmos de RNA sobre los conjuntos de datos sintéticos creados por el generador vistos anteriormente. Las pruebas realizadas con cada uno de ellos consisten en estudiar el comportamiento de las RNA a la hora de eliminar los lóbulos secundarios y la capacidad de mostrar de manera clara los blancos presentes en las imágenes.

Se plantearon diferentes estrategias para abordar el problema: uso de CNN, planteamiento como un problema de clasificación y planteamiento como un problema de regresión.

Redes Neuronales Convolucionales (CNN)

Basándonos en una interpretación por similitud del *coarray* como una imagen; la primera aproximación que se llevó a cabo se basó en el uso de CNN por ser una herramienta de *Deep Learning* diseñada específicamente para el tratamiento de imágenes. Tomando estas como *inputs* y asignando importancias (pesos) a ciertos elementos en la imagen para así poder diferenciar unos de otros.

Sin embargo, la repetitividad estructural de las imágenes, con un patrón oscilante orientado siempre sobre la misma dirección, nos hizo considerar que la etapa convolucional introduce mucha información poco relevante. Con este tipo de algoritmos se obtenían estructuras repetitivas y unidireccionales, de forma que añaden más complejidad de la necesaria sin resolver de manera óptima, ya que no ofrece suficientes soluciones espaciales.

Esto hace que se descartara el uso de CNN y nos centrásemos en el desarrollo de RNA, ya que, además de obtener imágenes de US de la mayor calidad posible, el objetivo en el que nos enfocamos se basa en poder llegar a implementar la solución desarrollada en Dispositivos Lógicos Programables (PLD), como las FPGAs (*Field Programmable Gate Arrays*).

Planteamiento como un problema de clasificación

Para una tarea de clasificación su característica principal es que el número de etiquetas que se obtiene a la salida es numerable.

La ventaja de utilizar este tipo de planteamiento es su compatibilidad con la imagen clásica, permitiendo tratar los casos confusos. A pesar de ello debemos diseñar una función objetivo que responda como una función de probabilidad, aunque también es posible usar una función basada en el *randomwalk* a partir de la fase de las señales. Todo ello se pudo simplificar tomando una función de clasificación binaria, asignando clase “1” a los píxeles que contienen a un blanco y clase “0” a los píxeles que contienen valores de fondo. Por lo que para este caso la imagen de referencia consiste simplemente en una imagen que contiene valores de 0 y 1 en función de los valores de la imagen original.

Sin embargo, los resultados preliminares obtenidos resultaron poco concluyentes. Unido a la incapacidad de lograr una correcta correlación entre la amplitud obtenida y el valor asociado a los blancos hizo que se descartara esta línea de trabajo.

Planteamiento como un problema de regresión

Para un problema de regresión el conjunto de etiquetas que se obtienen a la salida no es numerable.

Para este planteamiento se trabaja sobre modelos de señales que nosotros introducimos (como gaussiana con deltas), donde la imagen de referencia consiste en campanas de gauss generadas sobre los blancos a partir de los datos obtenidos de la imagen original. Para su formación se ha generado una función que permite aumentar o disminuir el haz y la caída de las campanas en función de nuestro interés. Un ejemplo se puede ver en la Figura 5.3.25

Los resultados preliminares que fuimos obteniendo con este tipo de enfoque fueron mucho más concluyentes que con el resto de los planteamientos y se conseguía una

correlación entre la amplitud y el valor de los blancos. Por ello nos centramos en la realización de un algoritmo basado en un problema de este tipo.

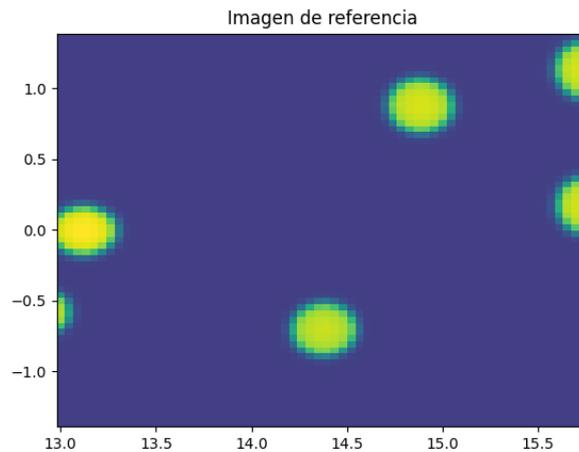


Figura 5.3. *Imagen de referencia enfoque de regresión en dB.*

5.4. Resolución de la imagen objetivo

Para el caso concreto de este TFG se han implementado 6 redes neuronales *fully connected*, de forma que todas las neuronas están totalmente conectadas entre capas, con el objetivo de resolver un problema de regresión.

Se han desarrollado de forma separada 3 redes neuronales para cada tipo de dato. Teniendo 3 redes neuronales para los datos temporales y otras 3 redes neuronales para los datos de frecuencia, de forma que podamos realizar una comparativa acerca de qué red y qué tipo de dato es el más eficaz a la hora de resolver la tarea objetivo.

5.4.1. Estimación del tamaño de las redes

Todas las redes neuronales contienen 1 capa de entrada, 11 capas ocultas y 1 capa de salida, variando el número de neuronas en las capas ocultas para cada uno de los 3 tipos de redes diseñadas.

El tamaño de las capas de entrada para todas las redes es el mismo, ya que depende del tamaño de la imagen objetivo que planteemos. Para el presente trabajo los datos se introducen en forma de tensor conteniendo 4096 neuronas de entrada (64×64). Lo mismo ocurre para la capa de salida, todas las redes contienen tan solo 1 neurona, ya que el objetivo es asociar uno a uno cada valor del píxel de la imagen original con los de la imagen de referencia.

El número de neuronas que contiene cada una de las 11 capas ocultas es lo que varía entre las redes, siendo su estructura la siguiente:

- Modelo red 1: [128, 32, 16, 16, 16, 16, 8, 8, 8, 4, 4]
- Modelo red 2: [128, 64, 64, 64, 64, 32, 32, 32, 16, 8, 8]
- Modelo red 3: [256, 256, 128, 128, 64, 64, 64, 32, 16, 8, 8]

Estas configuraciones son exploratorias y tienen como objetivo evaluar el desempeño del conformador con distintas cargas computacionales. Teniendo en cuenta el tipo de datos que se introduce se han tenido en cuenta diferentes consideraciones a la hora de diseñar los modelos:

- El número de neuronas por capa tiene que ser decreciente o igual a la capa anterior.
- El número de neuronas de la última capa solo se puede repetir una vez.
- El número de neuronas de la primera capa oculta debe ser del orden de $2N_e$ de forma que se encuentre en sintonía con los elementos del *coarray* ($2N_e - 1$). (N_e : el número de elementos).

Inicialmente se planteó un diseño de redes neuronales con capas iniciales que contenían un gran número de neuronas, obteniendo buena convergencia para los datos de entrenamiento, pero mala generalización para los datos de test, por lo que se dedujo que este tipo de modelos generaban sobreentrenamiento. Esto hizo que se plantease un diseño de redes simples, con diferentes capas ocultas y pocas neuronas en las capas iniciales, reduciendo la complejidad de los modelos y mejorando los resultados.

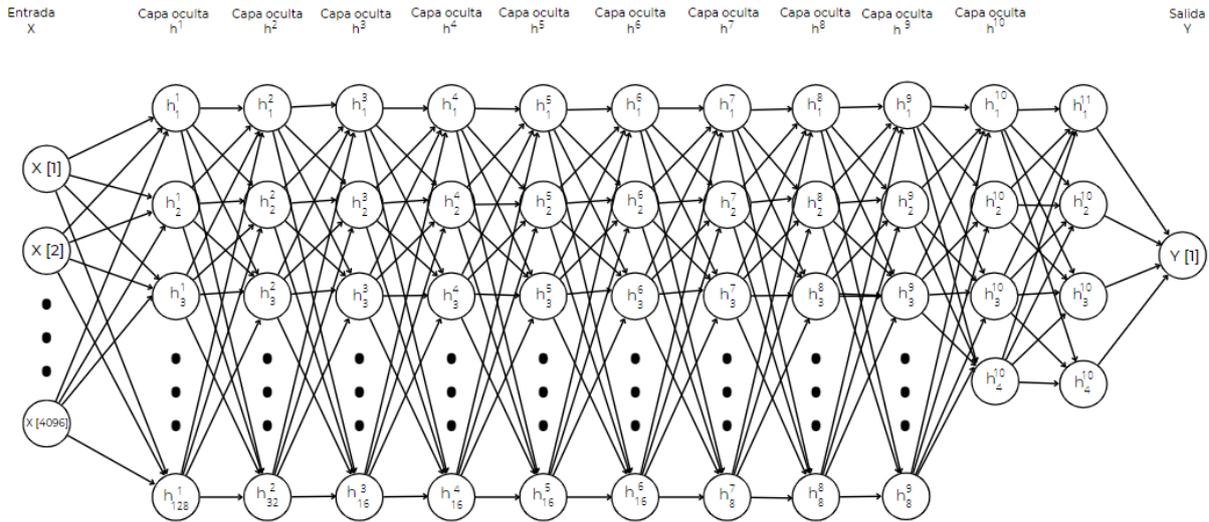


Figura 5.4. Esquema modelo red 1 con capa de entrada de 4096 neuronas, 11 capas ocultas con tamaño 128,32,16,16,16,16,8,8,4,4 respectivamente y una neurona en la capa de salida.

Las matrices de pesos ya mencionadas anteriormente w^n , son estructuras de vital importancia en lo que al entrenamiento de la red neuronal se refiere. Son estructuras de tantas filas como la capa de la que vengan y tantas columnas como la capa a la que vayan y son tan numerosas como intersecciones haya entre capas. Para este TFG corresponden doce matrices w^n . Por ejemplo, si estamos en la capa de entrada y hay que llegar a la primera capa oculta para el primer modelo de red (Figura 5.4.26), se tiene una matriz de pesos w^1 de dimensión 4096×128 .

Por otro lado es importante destacar los vectores de *bias*, b^n . Son estructuras constituidas por una serie de constantes las cuales interactúan con cada neurona. Actúan como regulador del umbral que rige la función de activación de la neurona en sí. Estos vectores se caracterizan por existir uno por intersección de capas y tener tantos puntos como número de elementos tenga la capa a la que se dirija la intersección. Por tanto, nuevamente para la Figura 5.4 entre la capa de entrada y la primera capa oculta, el vector b^1 tiene una dimensión de 128×1 .

Todos estos vectores y matrices se inicializan a cero y de manera aleatoria con distribución Gaussiana de media 0 y varianza 1, respectivamente. Con el paso de las iteraciones cambiarán sus valores (debido al método de *backpropagation*) y, una vez ha

finalizado el entrenamiento es con esos valores con los que nos quedamos para utilizar la red a modo de test con nuevas imágenes.

Es importante destacar que son estas estructuras las que iterativamente se van actualizando a fin de conseguir los valores óptimos para el problema que ocupa la red neuronal, por lo que, de estos valores depende el buen funcionamiento final de la red.

Finalmente obtenemos el vector de resultados, del mismo tamaño de los datos de entrada que introducimos, siendo 4096×1 . Para poder obtener una correcta visualización de los resultados se transforma a una matriz de 64×64 .

5.4.2. Funciones de activación y función de coste

En este punto se seleccionan los mecanismos de aprendizaje de la red neuronal, siendo clave las funciones de activación y la función de coste.

Las funciones de activación, como se ha comentado en el Apartado 3.3 son elementos no lineales, por lo general, que se operan al final de cada proceso en cada neurona para definir de manera adecuada la salida de la misma.

Para las neuronas de las capas intermedias, sin incluir la capa de salida se ha optado por una función de activación de tipo GeLU (*Gaussian Error Linear Units*), debido a que incorporan no linealidad al multiplicar de manera estocástica por 0 o 1 dependiendo de un muestreo a una distribución normal estándar. Su representación gráfica se visualiza en la Figura 5.5 .

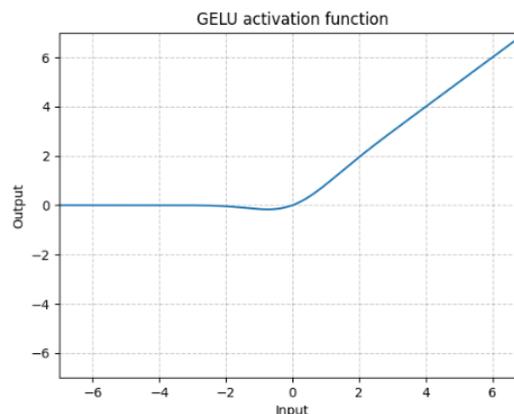


Figura 5.5 Representación gráfica función de activación GeLU [24].

Por otro lado, para la capa de salida se ha optado por una función de activación de tipo lineal. De manera que lo que se obtiene es lo que se calcula en cada elemento de salida, sin modificar, considerándose la salida de la red neuronal. Se llevó a cabo esta implementación debido a que, de seguir utilizando GeLU nunca se conseguirían valores negativos menores de -2.

En cuanto a la función de coste, para este TFG se ha escogido el MSLE (*Mean Squared Logarithmic Error*). Es una variación del clásico error cuadrático medio, el cual solo se preocupa por la diferencia relativa entre el valor real y el predicho, teniendo en cuenta la diferencia porcentual entre ellos. El MSLE penaliza más las subestimaciones que las sobreestimaciones, introduciendo de esta manera una asimetría en la curva del error. Lo cual nos interesa ya que una subestimación indica la posible pérdida de un blanco.

$$MSLE(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^N (\log(y + 1) - \log(\hat{y} + 1))^2 \quad (5.6)$$

Como se visualiza en la Ecuación (5.6) el error es la media sobre los datos vistos de las diferencias al cuadrado entre los valores reales y predichos transformados logarítmicamente.

Además, a la hora de formalizar el modelo se ha añadido como métrica el valor obtenido de MSE (*Mean Squared Error*), permitiendo evaluar cómo está siendo el entrenamiento del modelo. Esta función de métrica es muy similar a la función de pérdida, excepto que los resultados de la evaluación de esta no se utilizan para entrenar el modelo.

El MSE, expresado en la Ecuación (5.7) se define como el valor medio de los cuadrados de la diferencia entre los valores predichos y los valores reales.

$$MSE(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^N (y_n - \hat{y}_n)^2 \quad (5.7)$$

5.4.3. Implementación y entrenamiento

El entrenamiento que se lleva a cabo en este trabajo consiste en una base de datos de 1000 imágenes, donde de cada una se extraen dos matrices de datos (tiempo y frecuencia) de dimensiones (4096, 64, 64). Esto implica que la memoria utilizada es bastante grande

y resulte complicado poder introducir toda esta base de datos en la red neuronal. Por ello, para poder simplificar el entrenamiento se ha realizado lo que se denomina en términos de programación un bucle, de manera que, en cada iteración el generador forme una imagen y se extraigan los datos correspondientes (sean temporales o frecuencia dependiendo del examen que queramos realizar). Finalmente se introducen estos datos en la red neuronal con los parámetros que se van a describir a continuación, los cuales nos van a permitir determinar el tipo de entrenamiento óptimo para el conjunto de datos que introducimos. Este proceso se repite para las 1000 imágenes definidas como el conjunto de datos de entrenamiento.

Como hemos visto anteriormente, la principal razón de ejecutar una red neuronal es la de minimizar la función de coste con cada iteración. De manera natural esta función se calcula para cada uno de los vectores de entrada y, por ende, se ejecuta el método de *backpropagation* para cada vector, es decir, para cada iteración; esto da como resultado un entrenamiento muy preciso pero muy lento. Es por ello por lo que se implementa la idea de los *batches*. Estas estructuras son bloques de vectores de entrada de la gran cantidad de datos que se tiene. De manera que se calcula la función de coste por *batch* y no por entrada, acelerando el proceso de aprendizaje, aunque no resulta tan preciso como la manera inicial.

Para este trabajo el tamaño del *batch* escogido irá variando en función de la cantidad de imágenes que ya hayan sido entrenadas. Siendo: 256, 128, 64 y 32 cuando el número de imágenes sea menor de 200, 400, 600 y 1000 respectivamente. Como se puede ver son valores que guardan una relación con el número de elementos que contiene el *coarray* y se varían con el objetivo de lograr un mayor aprendizaje.

Es importante tener en cuenta que si se elige un tamaño de *batch* muy grande, la función de coste se calculará sobre una gran cantidad de datos y el entrenamiento no será el óptimo.

Una vez llegado a este punto se escoge el tiempo de entrenamiento que se quiere emplear a partir del número de iteraciones totales que quieren realizarse. A esto se le denomina *epoch*. Consiste en el hecho de haber ejecutado en la red todos los vectores de entrada de la base de datos, de forma que diremos que ha transcurrido una *epoch* cuando se hayan agotado todos los recursos de nuestros datos. Por lo que elegimos el número de veces que se entrena la red con la misma base datos.

En el trabajo se ha definido 20 como el número de *epochs* óptimo para lograr un aprendizaje óptimo.

Finalmente se elige el optimizador. Los optimizadores son clases o métodos que se utilizan para cambiar los atributos del modelo de aprendizaje con el objetivo de reducir las pérdidas en función del valor de *learning rate*. De manera que, los optimizadores ayudan a obtener resultados más rápidos.

El optimizador elegido es *Adamax*, una variante del optimizador clásico *Adam* basado en la norma del infinito. Aplicado con un valor de *learning rate* de 0.001, por ser aquel valor que logaba una buena convergencia sin estancarse en mínimos locales.

En cuanto a la implementación de las redes neuronales, como ya se ha comentado, todas van a seguir el mismo esquema, variando el número de capas ocultas para cada modelo, tanto para los datos temporales como los de frecuenciales.

La estructura de cada capa incluye el número de neuronas por capa, una *Batch Normalization* (BN) y la función de activación correspondiente.

Utilizamos BN por ser una técnica para entrenar redes neuronales profundas que estandariza las entradas a una capa para cada *batch*. Tiene el efecto de estabilizar el proceso de aprendizaje y reducir drásticamente la cantidad de épocas de entrenamiento necesarias. El proceso consiste en ir normalizando los vectores de las capas ocultas utilizando el primer y segundo momento estadístico (media y varianza) del *batch* actual. Este paso de normalización se aplica justo antes de la función no lineal de activación.

5.5. Evaluación

Una vez completado el entrenamiento se lleva a cabo el proceso de evaluación a partir del conjunto de *test*. Para ello contamos con una base de datos de 200 imágenes sintéticas (20% de la cantidad de datos de entrenamiento).

La evaluación cuantitativa que se va a seguir para poder comparar cada uno de los resultados va a ser la misma para todas las redes, independientemente del tipo de dato utilizado.

Para evaluar la calidad local de las imágenes se utilizará la SNR (Ecuación (5.8)). Mientras que para una evaluación global las métricas utilizadas serán PSNR (*peak signal to noise ratio*) (Ecuación (5.9)), pérdida de MSLE, pérdida de MSE y correlación cruzada (ρ) (Ecuación (5.10)) entre la imagen de referencia y la predicha por el modelo.

$$SNR = \frac{\mu_{\hat{y}}}{\sigma_{\hat{y}}} \quad (5.8)$$

$$PSNR = 20 * \log_{10}\left(\frac{\max(\hat{y})}{\sqrt{MSE}}\right) \quad (5.9)$$

$$\rho = \frac{\sum_n (y - \mu_y)(\hat{y} - \mu_{\hat{y}})}{\sqrt{(\sum_n |y - \mu_y|^2) (\sum_n |\hat{y} - \mu_{\hat{y}}|^2)}} \quad (5.10)$$

Todos estos valores son calculados a partir de valores naturales de las imágenes, no en decibelios (dB).

Además, se tiene en cuenta la complejidad de la red neuronal determinado por el número de parámetros de aprendizaje.

Por otro lado, también resulta interesante contar con unas métricas muy utilizadas en segmentación de imágenes, pero que en nuestro caso nos van a permitir valorar la similitud y la diferencia existente entre la imagen resultado y la imagen de referencia. El primero de ellos es un método basado en la similitud, denominado coeficiente de DICE, expresado en la Ecuación (5.11). Toma valores entre 0 y 1, siendo 0 ningún tipo de similitud y 1 similitud total entre las muestras.

$$DICE (Y, \hat{Y}) = \frac{Y \cap \hat{Y}}{Y \cup \hat{Y}} \quad (5.11)$$

Y la Distancia de Hausdorff (Ecuación (5.12)), es un método que se caracteriza por estar basado en distancias, ya que encuentra las distancia entre los elementos mal ubicados con respecto a la posición donde deberían estar. En este TFG se utiliza para mejorar el proceso de evaluación pudiendo obtener información sobre la forma y el contorno de los blancos.

$$Hausdorff (Y, \hat{Y}) = \max(h(Y, \hat{Y}), \max(\hat{Y}, Y)) \quad (5.12)$$

$$h(Y, \hat{Y}) = \max_{y \in Y} \min_{\hat{y} \in \hat{Y}} \|y - \hat{y}\| \quad (5.13)$$

La función $h(Y, \hat{Y})$ expresada en la Ecuación (5.13) (5.3)26 es la distancia Hausdorff directa y da como resultado el máximo de las distancias mínimas de cada punto del conjunto Y al conjunto \hat{Y} utilizando la distancia euclídea.

Finalmente para realizar una evaluación cualitativa se obtiene para todas los modelos de red la predicción para una misma imagen, representadas en el capítulo siguiente. Pudiendo juzgar de esta manera la calidad del proceso y de los resultados obtenidos con un elemento visual, así como la cantidad de blancos perdidos o añadidos por el modelo.

Capítulo 6

Resultados basados en regresión

En este Capítulo 6 se comentan los resultados obtenidos para cada tipo de estrategia planteada, tanto de manera cuantitativa, con las métricas descritas en el capítulo anterior, como cualitativamente, ofreciendo una comparativa entre estas.

6.1. Solución en tiempo

Descripción de los resultados obtenidos para los 3 modelos de redes utilizando datos temporales.

6.1.1. Modelo red 1

El primer modelo de red consiste en una estructura: [128, 32, 16, 16, 16, 16, 8, 8, 8, 4, 4] la cual cuenta con 530741 parámetros entrenables de los 531253 parámetros que contiene.

Modelo	SNR (dB)	PSNR (dB)	MSLE	MSE	ρ
Datos					
Temporales	18.744	21.809	0.0529	0.6571	0.5854
1					

Tabla 6.1. *Resultados métricas modelo 1 Datos Temporales.*

Comenzando con el valor de correlación cruzada observamos como no se obtiene un valor que indique una alta correlación. Lo debemos comparar con el coeficiente de DICE y la distancia de Hausdorff, los cuales se visualizan en la Figura 6.1.

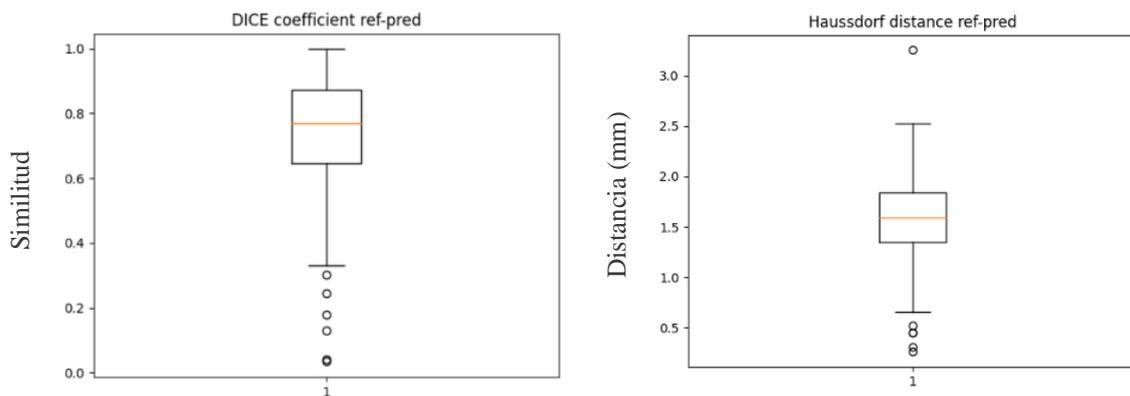


Figura 6.1. *Diagramas Coeficiente Dice y Distancia Hausdorff Red 1 Datos Temporales.*

El coeficiente de DICE presenta un valor alto, pero con una gran cantidad de *outliers* llegando algunos a alcanzar valores cercanos a cero. Esto hace que los resultados no sean óptimos del todo.

Mientras que para la distancia de Hausdorff podemos ver que la diferencia existente entre las dos imágenes no es muy notoria, obteniendo resultados bajos y con valores atípicos más bajos aún (a pesar de contar con un valor *outlier* mayor de 3.0).

De forma cualitativa se visualiza en la Figura 6.2. la predicción obtenida.

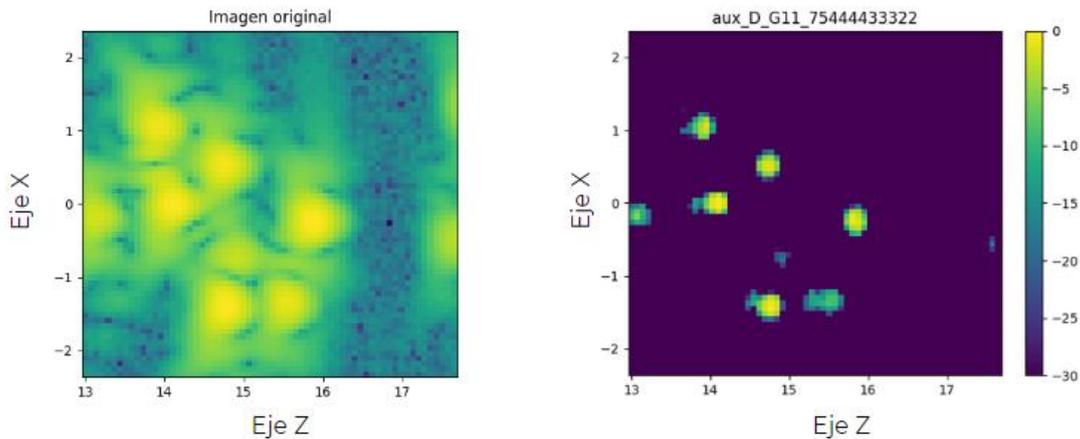


Figura 6.2. *Comparación imagen original - predicción modelo red 1 Datos Temporales.*

Se puede comprobar como los lóbulos secundarios han desaparecido. A pesar de ello se debe tener en cuenta que uno de los blancos que se intuye en la imagen original no aparece en la imagen resultado, aquel que se encuentra en las coordenadas [17.9,1.3]. Cabe destacar como en el modelo los blancos de intensidades menores a -20dB no los reproduce correctamente. Esto se comprueba con el blanco de coordenadas [14.8, -0.8]. Además, muchos de los blancos aparecen ligeramente duplicados, lo cual no nos interesa.

En cuanto a los valores de la Tabla 6.1. el valor de SNR es el más alto de los 3 modelos de datos temporales lo cual viene a indicar una mejor calidad de señal. En cuanto al PSNR indica la relación entre la máxima energía posible de una señal y el ruido que afecta a su buena representación. Los valores típicos suelen variar entre 30 y 50 dB, siendo mayor cuanto mejor es la codificación. Por lo que para este parámetro se obtiene un valor inferior al esperado debido a los resultados obtenidos.

Para los valores de error el MSLE es el que más nos interesa ya que es la función de coste para el entrenamiento del modelo. A menor valor mejores resultados, debido a que menor diferencia existirá entre la imagen resultado y la objetivo. Mientras que el MSE es superior a lo esperado, obteniendo un valor mucho más alto que el MSLE.

6.1.2. Modelo red 2

Este segundo modelo de red consiste en una estructura: [128, 64, 64, 64, 64, 32, 32, 32, 16, 8, 8] la cual cuenta con 551113 parámetros entrenables de los 552137 parámetros que contiene.

Modelo	SNR (dB)	PSNR (dB)	MSLE	MSE	ρ
Datos					
Temporales 3	15.348	22.315	0.0656	0.7196	0.5936

Tabla 6.2. *Resultados métricas modelo 2 Datos Temporales.*

Visualizando los resultados de la Tabla 6.2. se puede inferir la gran relación que guardan con los resultados obtenidos para el modelo 1. El valor de SNR es algo menor encontrándose por debajo del ratio habitual de valores (suele encontrarse entre 18 y 30dB) y el valor de PSNR es ligeramente algo mayor.

En cuanto a los valores de error comprobamos nuevamente como se obtienen valores muy similares a los del modelo anterior, siendo bastante más alto el MSE que el MSLE. Lo mismo ocurre con la correlación cruzada, siendo ligeramente superior este resultado.

Este modelo presenta una complejidad mayor y eso se puede comprobar en el aumento del número de parámetros entrenables que contiene.

Para los resultados de los coeficientes de DICE es importante remarcar la aparición de un valor atípico mayor que uno, lo cual indica un error obtenido en el cálculo. A pesar de ello este coeficiente es ligeramente inferior al obtenido en el anterior modelo, obteniendo nuevamente valores atípicos muy cercanos a cero. En cuanto a la distancia de Hausdorff la mediana se sitúa cercana a 1.5 mm, lo cual es un valor óptimo para este parámetro, indicando que no existe mucha diferencia entre las imágenes. (Figura 6.3. .

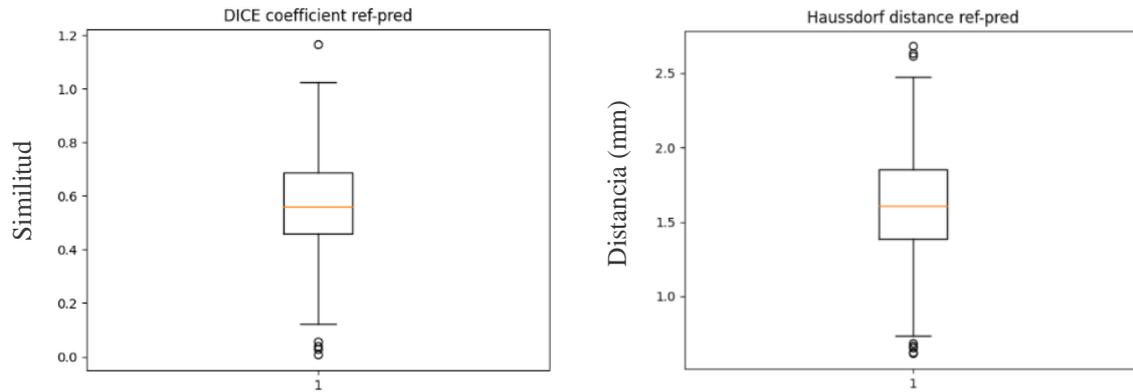


Figura 6.3. *Diagramas Coeficiente Dice y Distancia Hausdorff Red 2 Datos Temporales.*

Teniendo en cuenta los resultados cuantitativos obtenidos podemos deducir que los resultados cualitativos no van a diferir mucho del primer modelo. En cuanto a desventajas de este modelo se puede comprobar como nuevamente, el blanco mencionado en el modelo anterior vuelve a no aparecer. Por otro lado, como mejoras destaca una mejor visualización del blanco situado en las coordenadas [14.8, -0.8] y una ligera eliminación de los blancos duplicados.

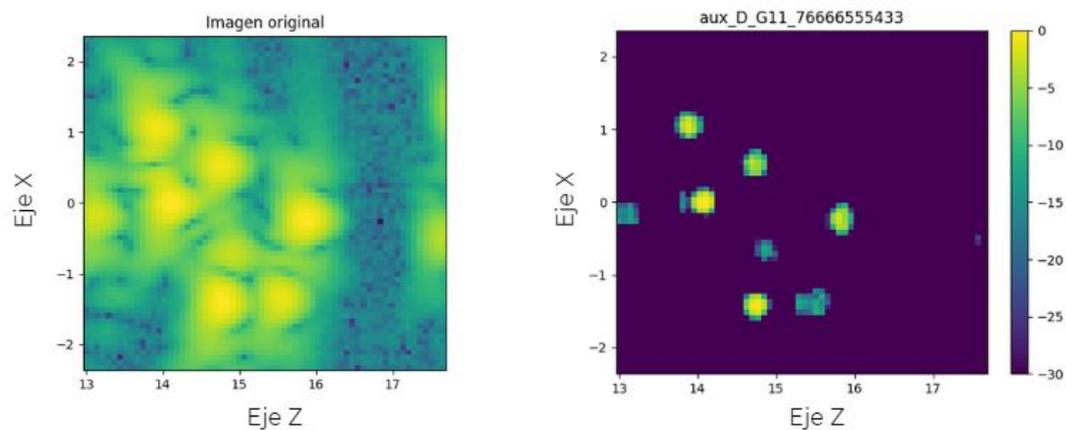


Figura 6.4. *Comparación imagen original - predicción modelo red 2 Datos Temporales.*

6.1.3. Modelo red 3

El tercer modelo de red consiste en una estructura: [256, 256, 128, 128, 64, 64, 64, 32, 16, 8, 8] la cual cuenta con 1185481 parámetros entrenables de los 1187529 parámetros que contiene.

Modelo	SNR (dB)	PSNR (dB)	MSLE	MSE	ρ
Datos					
Temporales	14.301	24.658	0.05085	1.0765	0.5966
2					

Tabla 6.3. *Resultados métricas modelo 1 Datos Temporales.*

Este tercer modelo es el que mayor complejidad presenta al ser aquel con mayor número de parámetros entrenables. Por el resultado de correlación podemos deducir que es el modelo que más relación guarda con la imagen objetivo (siendo ligeramente superior al modelo anterior). Comparándolo con los resultados de DICE y Hausdorff de la

Figura 6.5. se observa como para esta última métrica el valor de la mediana se sitúa ligeramente superior a los otros dos modelos, pero no aparecen *outliers* lo cual consideramos un resultado óptimo, ya que la mayor diferencia se encuentra levemente superior a 2.5 mm.

En cuanto al coeficiente de DICE nuevamente se obtienen *outliers* superiores a 1.0 debidos a errores y otros muy cercanos a cero.

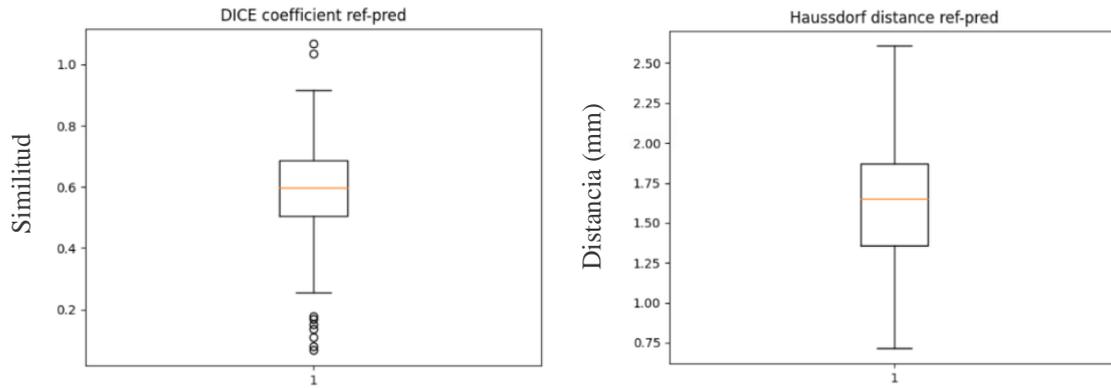


Figura 6.5. *Diagramas Coeficiente Dice y Distancia Hausdorff Red 3 Datos Temporales.*

Para los parámetros de SNR se comprueba como a medida que aumenta la complejidad del modelo, este disminuye indicando una peor calidad de señal. A la inversa ocurre con el PSNR aumentando a medida que la complejidad aumenta acercándose al valor umbral de 30 dB pero sin alcanzarlo en ninguna de las redes.

En cuanto a las métricas de error expresadas en la Tabla 6.3. se obtiene un valor menor para el MSLE, indicando que a medida que ha ido aumentando el número de neuronas por capa el entrenamiento ha ido mejorando. Mientras que por el contrario, el valor de MSE aumenta notablemente con respecto a los dos modelos anteriores.

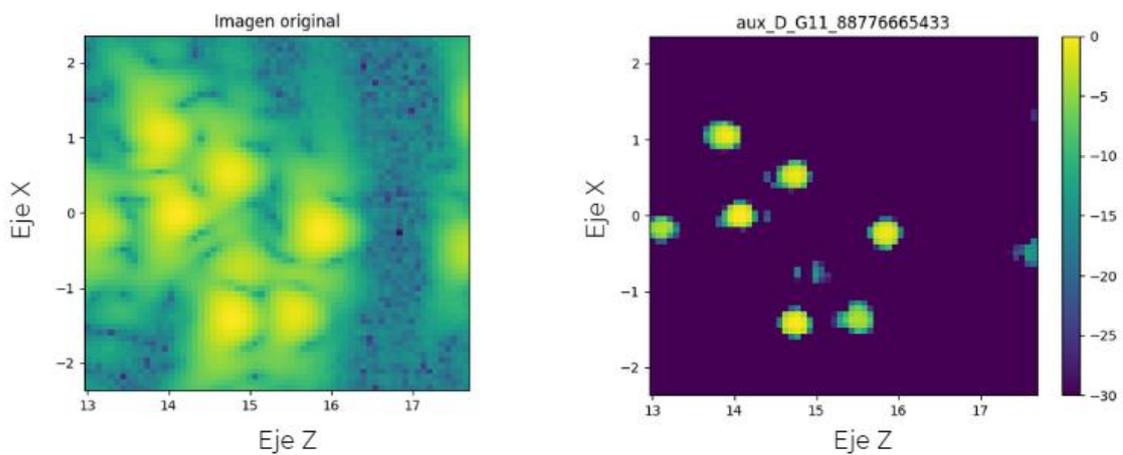


Figura 6.6. *Comparación imagen original - predicción modelo red 3 Datos Temporales.*

En la Figura 6.5. Figura 6.6. se obtiene como ventaja con respecto a los modelos anteriores la visualización completa del blanco en las coordenadas [17.9, -0.5], donde con las redes anteriores se visualizaba levemente. Se consigue intuir el blanco situado en [17.9,1.3], lo cual supone una gran mejora ya que en las soluciones de los modelos anteriores no ocurría, por lo que este modelo detecta con mayor calidad blancos menores a -20dB (sin superar los -25dB). Además, la amplitud de los blancos se asemeja en mayor medida a la imagen original, pudiendo valorar su intensidad.

Por otro lado, disminuye la visualización del blanco en [14.8, -0.8] y no se consigue corregir al completo la duplicación de algunos de los blancos, aunque mejora respecto a los modelos anteriores.

6.2. Solución en frecuencia

Descripción de los resultados obtenidos para los 3 modelos de redes para datos frecuenciales.

6.2.1. Modelo red 1

El primer modelo de red consiste en una estructura: [128, 32, 16, 16, 16, 16, 8, 8, 8, 4, 4] la cual cuenta con 530741 parámetros entrenables de los 531253 parámetros que contiene.

Modelo	SNR (dB)	PSNR (dB)	MSLE	MSE	ρ
Datos					
Frecuencia 1	30.241	28.921	0.0006	0.001	0.5848

Tabla 6.4. *Resultados métricas modelo 1 Datos Frecuenciales.*

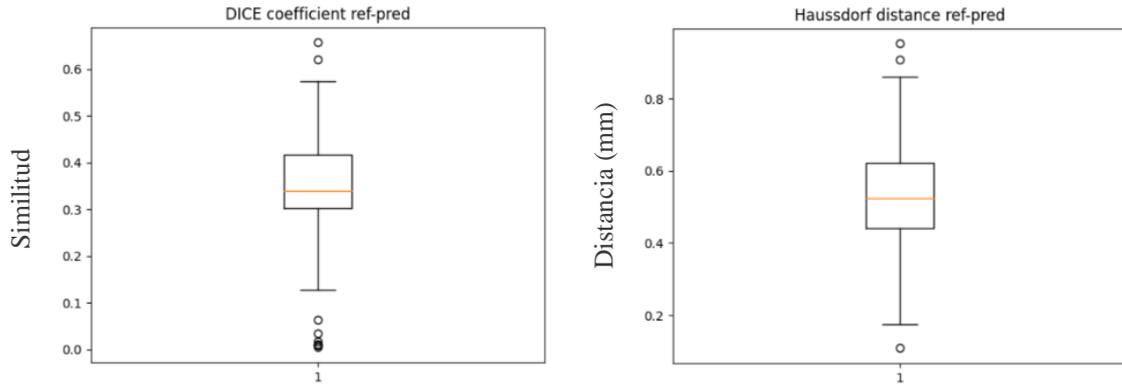


Figura 6.7. *Diagramas Coeficiente Dice y Distancia Hausdorff Red 1 Datos Frecuenciales*

Comparando los resultados de la Tabla 6.4. y la Figura 6.7. y tomando como referencia a su vez los resultados de los modelos de datos temporales, encontramos que la similitud dada por el coeficiente de DICE es menor de la que hemos ido obteniendo para los datos temporales, no llegando a 1. A diferencia de la distancia de Hausdorff, la cual es notablemente menor adquiriendo un valor de mediana cercano al 0.5. Una de las razones a las que puede deberse esta solución, y como se representa en la Figura 6.8. , es la forma de los blancos que se obtienen en la predicción, los cuales son circulares, asemejándose más a la forma de lo blancos originales, mientras que en los resultados obtenidos en los diseños anteriores adquirirían una forma levemente ovalada.

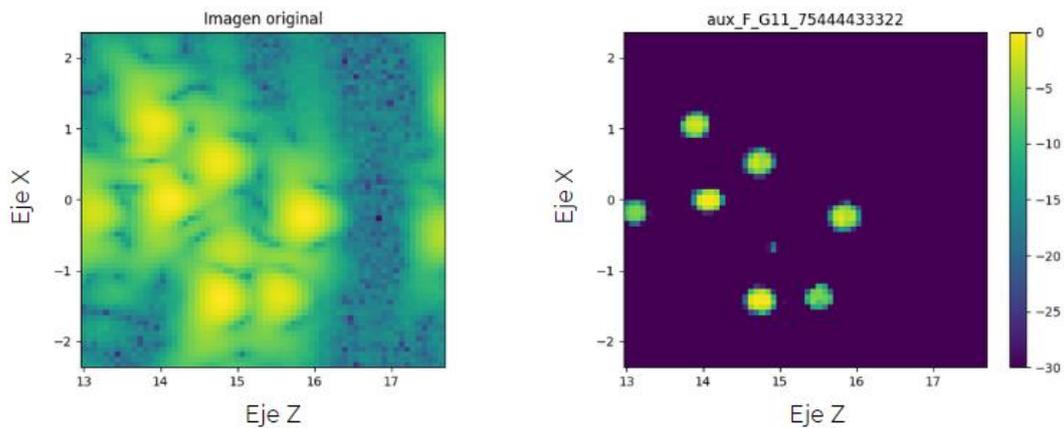


Figura 6.8. *Comparación imagen original - predicción modelo red 1 Datos Frecuenciales.*

En la Figura 6.8. se observa como los dos blancos que se encuentran a la derecha de la imagen (aquellos menores a -20dB) no se obtienen como resultado, y el blanco central en [14.8, -0.8] se visualiza con muy poca seguridad. Esto hace que prácticamente 3 puntos de interés desaparezcan. Por otro lado, es de importancia destacar la desaparición del duplicado que sufrían algunos blancos en los modelos de datos temporales, así como la forma de estos comentada anteriormente.

Estos resultados descritos se visualizan en la Tabla 6.4. en el valor de correlación cruzada, aunque comprobamos como en ninguno de los casos vistos es mayor de 0.6. En cuanto al resultado de SNR es mucho más alto que los obtenidos en los modelos de datos temporales superando el valor de 30dB donde se suele encontrar normalmente el máximo, mientras que el PSNR no llega aún ese valor umbral de 30dB. Para los parámetros de error, tanto el MSE como el MSLE presentan resultados notablemente más bajos que en los modelos temporales. Indicando por tanto mejores resultados en el entrenamiento que ha sido llevado a cabo por la red.

6.2.2. Modelo red 2

Este segundo modelo de red frecuencial, al igual que para los datos temporales consiste en una estructura: [128, 64, 64, 64, 64, 32, 32, 32, 16, 8, 8]. Contando con 551113 parámetros entrenables de los 552137 parámetros que contiene.

Modelo	SNR (dB)	PSNR (dB)	MSLE	MSE	ρ
Datos					
Frecuencia	18.7272	30.5936	0.0005	0.0007	0.5601
2					

Tabla 6.5. *Resultados métricas modelo 2 Datos Frecuenciales.*

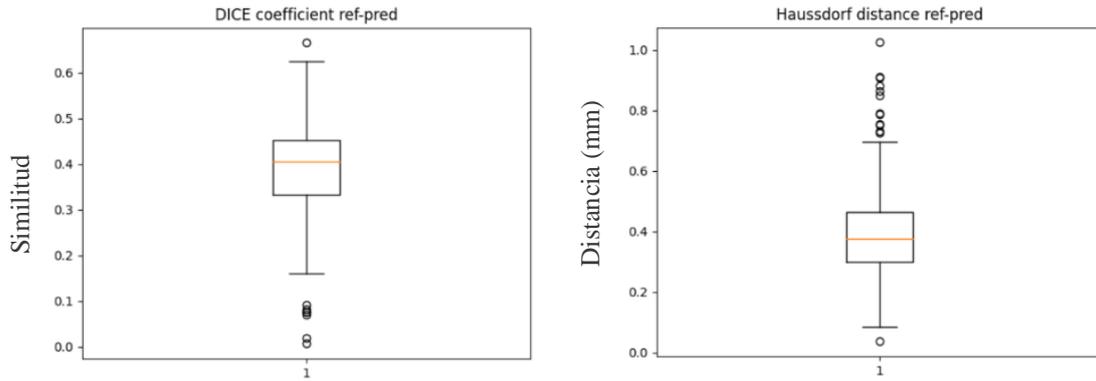


Figura 6.9. *Diagramas Coeficiente Dice y Distancia Hausdorff Red 2 Datos Frecuenciales.*

Se puede comprobar como el valor de correlación cruzada de la Tabla 6.5. se mantiene en unos valores similares a los ya obtenidos. Lo mismo ocurre para los coeficientes de Hausdorff y DICE de la Figura 6.9. Para el primero resalta como la mediana es ligeramente inferior a la del primer modelo de datos de frecuencia, pero aparecen gran cantidad de *outliers* por la zona superior. Mientras que el coeficiente de DICE muestra mayor grado de similitud que el modelo anterior entre nuestra imagen original y la predicha por el modelo, pero representando nuevamente valores *outliers* muy cercanos a 0 y sin lograr los resultados de este coeficiente para los datos temporales.

Nuevamente en los resultados cuantitativos de la Tabla 6.5. se detecta un descenso del valor de SNR, perdiendo calidad de señal, y un aumento de PSNR aproximándose al valor umbral determinado. De la misma manera, para los valores de error comprobamos como son notablemente inferiores a los obtenidos para los modelos temporales, y como el MSE disminuye con respecto al Modelo red 1 de datos frecuenciales.

En cuanto a los resultados cualitativos de la Figura 6.10. se concluye una falta de mejora sustancial con respecto al modelo anterior. Continúa sin mostrar los dos blancos menores de -20dB (que se encuentran a la derecha de la imagen) y el blanco [14.8, -0.8] sigue sin percibirse con calidad.

Además, se puede apreciar la aparición de una ligera reverberación en los puntos de interés.

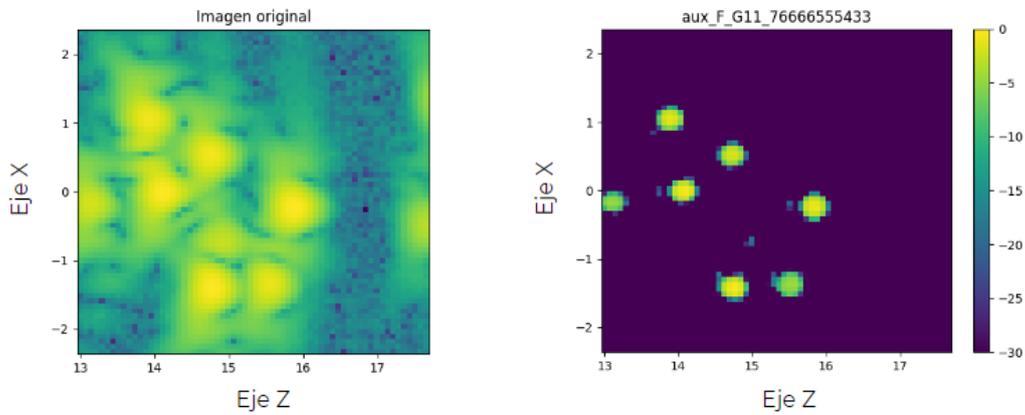


Figura 6.10. Comparación imagen original - predicción modelo red 2 Datos Frecuenciales.

6.2.3. Modelo red 3

Presenta una estructura: [256, 256, 128, 128, 64, 64, 64, 32, 16, 8, 8] contando con 1185481 parámetros entrenables de los 1187529 parámetros que contiene.

Modelo	SNR (dB)	PSNR (dB)	MSLE	MSE	ρ
Datos					
Frecuencia	17.5871	30.6133	0.00049	0.0007	0.5541
3					

Tabla 6.6. Resultados métricas modelo 3 Datos Frecuenciales

Este modelo presenta unos resultados muy similares al anterior. Como ya pasaba en los diseños de datos temporales, a medida que la complejidad del modelo ha ido aumentando, la calidad de la señal ha ido disminuyendo, comprobándolo en los valores de SNR. A diferencia de lo ocurrido con los valores de PSNR, aumentando con la complejidad.

Para las métricas de error de la Tabla 6.6. se comprueba como prácticamente son los mismos resultados que hemos obtenido anteriormente. Lo cual nos lleva a la conclusión

de que, los entrenamientos llevados a cabo con datos frecuenciales muestran mejores resultados que los obtenidos para datos temporales. (Pudiéndose deber a lo ya comentado de la forma de los blancos).

Como se representa en la Figura 6.11. el coeficiente de Hausdorf indica una mayor diferencia entre la imagen de referencia y la predicha con respecto a los dos modelos anteriores, debido a la presencia de un valor atípico situado en 1.6 mm. Mientras que para el coeficiente de DICE se obtiene prácticamente el mismo resultado que para el modelo de red anterior.

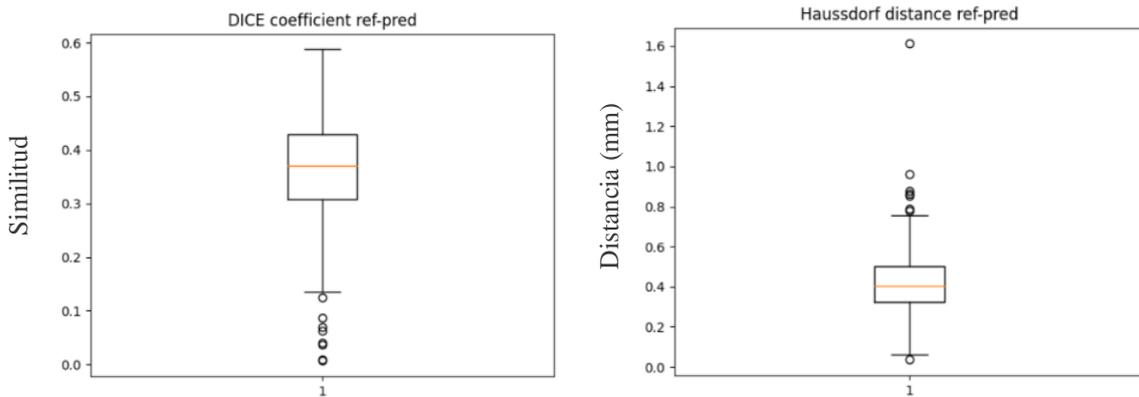


Figura 6.11. *Diagramas Coeficiente Dice y Distancia Hausdorf Red 3 Datos Frecuenciales.*

Finalmente, visualizando la Figura 6.12. se puede concluir con certeza la incapacidad que presentan los modelos de datos frecuenciales para mostrar blancos menores a -20dB. Mientras que, para este diseño de red, el blanco central (que en las dos redes anteriores se intuía con gran dificultad) se visualiza correctamente. Cabe destacar como ya se mencionó en el apartado anterior la ligera presencia de duplicados en el lado izquierdo de los blancos.

6. RESULTADOS BASADOS EN REGRESIÓN

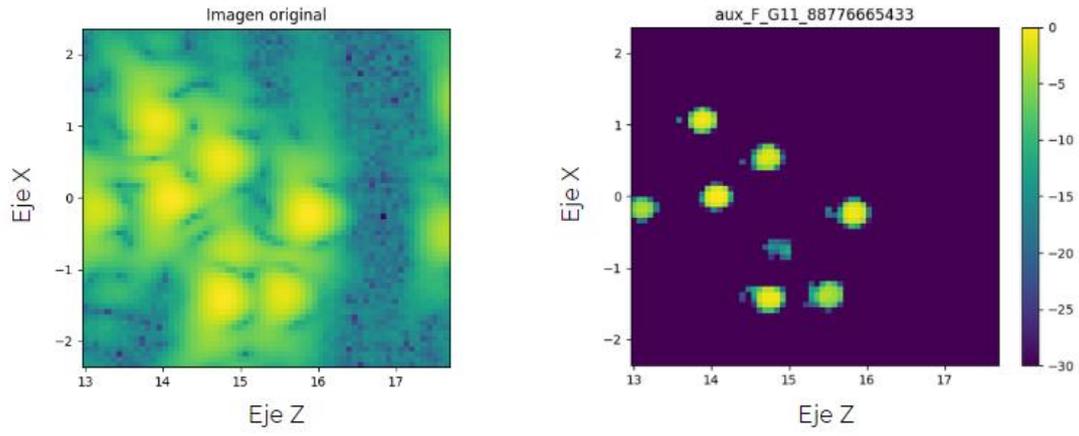


Figura 6.12. *Comparación imagen original - predicción modelo red 3 Datos Frecuenciales.*

Capítulo 7

Discusión y líneas de trabajo futuras

En este último capítulo se lleva a cabo la discusión acerca de los resultados obtenidos, así como conclusiones y posibles líneas de trabajo futuras a las que se ha llegado a lo largo de este TFG.

7.1. Discusión

Hasta el momento, se han hecho muchas contribuciones en DL y ML para mejora de la calidad de imagen ecográfica, pero la mayoría de las veces a partir de la implementación de CNN introduciendo como entrada estas imágenes. En el desarrollo de este proyecto, se ha buscado intencionadamente variar este tipo de modelos generando un conformador de haces a partir de técnicas de apertura sintética, extrayendo finalmente datos temporales y frecuenciales, utilizando para ello un conjunto de datos con variabilidad controlada permitiendo estudiar cómo mejora la calidad de las imágenes en función del tipo de dato y del modelo de red neuronal utilizado.

La gran cantidad de datos que se extraían de las imágenes hizo que se descartasen el uso de CNN y RNA muy profundas debido a la gran cantidad de parámetros que generaban y su dificultad para implementarlos en FPGAs. Es por ello por lo que el trabajo se basa en el uso de RNA con 11 capas ocultas variando el número de neuronas en cada capa.

A través de la realización del presente trabajo se puede extraer la capacidad que presentan las Redes Neuronales Artificiales profundas para llevar a cabo tareas de

reducción de ruido y mejora de la calidad de imagen US, a partir de datos frecuenciales y temporales extraídos directamente de imágenes sintéticas obtenidas a partir de un generador. Resulta esencial el diseño de este generador de datos sintéticos, permitiéndonos examinar los modelos en condiciones controladas y pudiendo comprobar el funcionamiento de los diseños redes neuronales antes de manejar imágenes reales de pacientes.

Las redes neuronales artificiales profundas que se proponen presentan un diseño sencillo, con no muchas neuronas en las capas ocultas. Esta decisión se tomó tras comprobar como aquellos diseños de modelos densos, con muchas capas ocultas y gran cantidad de neuronas en ellas generaban sobre-entrenamiento, obteniendo muy buenos resultados para los datos de entrenamiento, pero muy poca capacidad de generalización en los datos de *test*. Siendo el objetivo principal del presente trabajo encontrar un compromiso entre la complejidad de la red neuronal y la eliminación de los lóbulos secundarios.

En cuanto al uso de los diferentes tipos de datos extraídos de las imágenes se ha podido comprobar para el análisis cualitativo como los resultados obtenidos son muy parecidos, en cuanto a que, para ambos conjuntos de datos, a medida que la complejidad de la red aumentaba se obtenían mejoras en la visualización de los blancos. La predicción mediante el conjunto de datos temporales demuestra la capacidad que presentan los modelos de mostrar blancos menores a -20dB a medida que la complejidad aumenta (sin superar los -25dB). Como puntos negativos cabe destacar la aparición de artefactos de duplicado cercanos a los puntos de interés y la representación de los blancos con formas más ovaladas que aquellas que se visualizan en la imagen de referencia.

En los modelos de redes que usan los conjuntos de datos frecuenciales se puede constatar cómo, las imágenes predichas por los modelos representan los blancos con forma redondeada, siendo similar a la forma de estos representada en la imagen de referencia. Los puntos de interés aparecen con un menor duplicado a lo que observábamos en los resultados de los modelos de datos temporales, pero, a medida que la complejidad de los modelos aumenta, aparecen nuevamente de forma leve estos artefactos. Además, es importante destacar la incapacidad de estos modelos para representar blancos menores de -20 dB.

En cuanto a los resultados cuantitativos se ha detectado un patrón característico para los modelos de ambos conjuntos de datos. A medida que la complejidad de las redes aumentaba, los valores de SNR disminuían, al contrario de lo que ocurre para valores de PSNR, aumentando con la complejidad. Esto determina que, a medida que las redes presentan mayor cantidad de parámetros la calidad de la señal que obtenemos a la salida disminuye, mientras que la relación entre la máxima energía posible de la señal y el ruido aumenta.

A pesar de las mejoras obtenidas y la demostración de la capacidad de los modelos de redes neuronales para la reducción de los lóbulos secundarios, es importante destacar de forma general la pérdida de blancos con intensidades menores a -20dB y la aparición de blancos duplicados, impidiendo lograr una forma completamente redondeada de estos, acorde con la representada en la imagen objetivo.

7.2. Líneas Futuras

A continuación, se proponen las siguientes líneas para seguir avanzando en la investigación sobre generación de un conformador de haces para mejora de la calidad de imagen ultrasónica:

- Llevar a cabo un segundo estudio más profundo y sensible utilizando como base el primer estudio general realizado en este TFG.
- Diseñar un generador sintético de mayor complejidad añadiendo ruido aleatorio aproximándose más a una imagen real.
- Diseñar una función de error con el objetivo de que no reste tanta importancia a los valores pequeños.
- Intentar llevar a cabo el diseño de un modelo que combine los conjuntos de datos temporales y frecuenciales, con el objetivo de extraer la máxima cantidad de información a partir de ellos.

7. DISCUSIÓN Y LÍNEAS FUTURAS

- Llevar a cabo un estudio con imágenes médicas reales a través de la extracción de los diversos conjuntos de datos.

Bibliografía

- [1] C. J. Martín Arguedas, “Técnicas de apertura sintética para la generación de imagen ultrasónica”, Tesis doctoral, Departamento de Electrónica, Escuela politécnica superior Universidad de Alcalá, Madrid, 2010.
- [2] Olympus S. I., «Tutorial sobre el ultrasonido Phased Array, Lóbulos reticulados y lóbulos laterales,» [En línea]. Disponible: <https://www.olympus-ims.com/es/ndt-tutorials/transducers/lobes/>. [Último acceso: 2022].
- [3] S. Yeasmin, «Benefits of artificial intelligence in medicine» en *2nd International Conference on Computer Applications Information Security*, Riyadh, 2019.
- [4] D. Hyun, L. Brickson, K. T. Looby y J. Dahl, «Beamforming and speckle reduction using neural networks,» *IEEE Trans*, vol. 66, nº 5, 2019.
- [5] R. JG van Sloun, R. Cohen y Y. C. Eldar «Deep Learning in Ultrasound Imaging,» *IEEE*, vol. 5, 2019.
- [6] T. D. Rossing y N. H. Fletcher, *Principles of Vibration and Sound*, New York: Springer-Verlag, 2004.
- [7] A. Vargas, L. M. Amescua Guerra, M. A. Bernal y C. Pineda, «Principios físicos básicos del ultrasonido, sonoanatomía del sistema musculoesquelético y artefactos ecográficos,» *Medigraphic*, vol. 22(6), nº Nov.-Dic: , pp. 361-373., 2008.
- [8] A. Vargas, L. M. Amescua Guerra, M. A. Bernal y C. Pineda, «Basic physical principles of ultrasonography, anatomy of the musculoskeletal system and ecographic artifacts,» *Acta ortopédica mexicana*, vol. 22, pp. 361-73, 2008.

- [9] C. G. Torre, «The wave equation in 3 dimensions,» Department of Physics, Utah State University, 2014.
- [10] H. Yao, Synthetic Aperture Methods for Medical Ultrasound Imaging, Tesis doctoral.
- [11] G. S. Kino, Acoustic Waves: Devices, Imaging, and Analog Signal Processing, Prentice-Hall, 1987.
- [12] L. Saraceno y S. Vitale, Fusión de modelos de tejido en la simulación de imágenes de ultrasonido en tiempo real (SIMECO), Trabajo Fin de Grado, Universidad Nacional del Centro de la Provincia de Buenos Aires, Buenos Aires, Argentina, 2013.
- [13] S. J. Shin y B. J. Jeong, «Principle and Comprehension of Ultrasound Imaging,» *The Journal of the Korean Orthopaedic Association*, vol. 48, n° 5, 2013.
- [14] B. A. J. Angelsen, «Ultrasound imaging. Waves, signals and signal processing.,» *Ultrasound in Medicine & Biology*, Vol. 18., pp. 401-402, 2002.
- [15] K. K. Shung y M. B. Smith, B. Tsui, Principles of Medical Imaging, Academic Press, 1992.
- [16] R. Gonzalez y R. Borrego, *Fundamentos básicos de ecografía*, Madrid: Grupo de Trabajo de Ecografía. Sociedad Española de Cuidados Intensivos Pediátricos., 2020.
- [17] J. Y. Lu, H. Zou y J. F. Greenleaf, «Biomedical Ultrasound Beamforming,» *Ultrasound in Med. and Biol.*, vol. 20, 1994.
- [18] M. P. Romero. Conformador de haces ultrasónicos mediante muestro selectivo con codificación delta, Tesis doctoral, Facultad de Informática, Universidad Politécnica de Madrid, , Madrid, 2004.
- [19] T. L. Szabo, Diagnostic Ultrasound imaging: Inside out, Hartford, Connecticut: Trinity College, 2004.

- [20] W. D'Andrea Fonseca, J. P. Ristow, D. G. Sanches, Samir N. y Y. Gerges, «Bandwidth Comparison on PSFs Simulations using Classical Beamforming,» de *Forum Acousticum*, Aalborg, Dinamarca, 2011.
- [21] D. Romero, J. Villazón, J. González y O. Martínez, «Integration of golay pulse compression in a synthetic aperture imaging system based on minimum redundancy coarray,» de *VIII Congreso Ibero-americano de acústica*, Évora - Portugal, 2012.
- [22] C. P. Seco, Detección de onda P con aprendizaje máquina y redes profundas en monitorización prolongada con wearables, Trabajo Fin de Grado, ETSIT Universidad Rey Juan Carlos. Madrid, 2020.
- [23] Margot P. (May, «Cross Validation: definición e importancia en Machine Learning,» Mayo 2022. [En línea]. Disponible: <https://datascientest.com/es/cross-validation-definicion-e-importancia>. [Último acceso: Septiembre 2022].
- [24] I. Artificial, *Notas de clase 2229037*, Universidad Rey Juan Carlos: Departamento de Teoría de la Señal y Comunicación y Sistemas Telemáticos y Computación, 2022.
- [25] J. C. Rodríguez, O. E. Gauldrón y J. Jauregui, «Desarrollo de un modelo de regresión con redes neuronales artificiales para estimar la resistencia robótica de un motor de inducción,» de *LACCEI*, Medellín, Colombia, 2011.
- [26] I. Goodfellow, Y. Bengio y A. Courville, «Deep Learning,» *MIT Press*, 2016.
- [27] Análisis de Imágenes Médicas, *Notas de clase 2229039*, Universidad Rey Juan Carlos: Departamento de Teoría de la Señal y Comunicación y Sistemas Telemáticos y Computación, 2022.
- [28] B. Luitjen, R. Cohen, F. J. de Bruijn, H. A. W. Schmeitz, M. Mischi, Y. C. Eldar y R. J. G. van Sloun, «Adaptive Ultrasound Beamforming Using Deep Learning,» *IEEE Transactions on Medical Imaging*, vol. 39, nº 12, 2020.

- [29] S. Zhan, Y. Wang, J. Jiang, J. Dong, W. Yi y W. Hou, «CNN – Based Medical Ultrasound Image Quality Assessment,» *Hindawi*, 2021.
- [30] L. Bote, S. Muñoz y A. Guerrero, «Deep learning and big data in healthcare: A double review for critical beginners,» *Applied Sciences - MDPI*, vol. 9, nº 11, 2019.
- [31] L. Guimaraes, «Técnica de enfoque de apertura sintética para la restauración de imágenes,» *NDT*, vol. 5, nº 8, 2000.
- [32] M. Schickert, «Progress in ultrasound imaging of concrete,» 2005.
- [33] R. Y. Chiao y L. J. Thomas, «Aperture formation on reduced-channel array using the transmit-receive apodization matrix,» *IEEE Ultrasonics Symposium*, vol. 2, pp. 1567-1571, 1996.
- [34] R. N. Thomson, *Ultrasonics - Transverse and longitudinal resolution of the synthetic aperture focusing technique*, Elsevier, 1984.
- [35] F. L. Tadeus Stepinski, «Synthetic Aperture Radar (EUSAR), 8th European Conference,» de *Synthetic aperture focusing techniques for ultrasonic imaging of solid objects*, 2010.
- [36] D. Romero, O. Martínez, C. J. Martín A. Ibañez y L. G. Ullate, «Paralelización de los procesos de conformación de haz para imagen ultrasónica con técnicas GPGPU,» *Revista Iberoamericana de Automática e Informática Industrial*, vol. 9, nº 2, pp. 144-151, 2012.
- [37] R. T. Hocr, S. A. Kassam, «The Unifying Role of the Coarray in Aperture Synthesis for Coherent and Incoherent Imaging,» *Proceedings of the IEEE*, vol. 78, nº 4, pp. 735-752, 1990.
- [38] S. Nikolov, *Synthetic aperture tissue and flow ultrasound imaging*, Tesis doctoral, DTU, Technical University of Denmark, Lyngby, Denmark:, 2001.
- [39] B. G. A. Nowicki, «Ultrasonic Synthetic Apertures: Review,» *Archives of Acoustics*, vol. 39, nº 4, pp. 427-438, 2014.

- [40] O. Martínez, V. Yagüe, M. Parrilla «Improving ultrasonic medical image quality by attenuation of the secondary lobes,» de *International Ultrasonics Symposium*, Glasgow, Scotland, 2019.

