



INAOE

**INSTITUTO NACIONAL DE ASTROFÍSICA,
ÓPTICA Y ELECTRÓNICA**

**MAESTRÍA EN CIENCIAS Y TECNOLOGÍAS
BIOMÉDICAS**

**OPTIMIZACIÓN DE LA RECONSTRUCCIÓN
DE IMAGEN PARA TOMOGRAFÍA
FOTOACÚSTICA UTILIZANDO
APRENDIZAJE PROFUNDO**

TESIS

**PARA OPTAR POR EL GRADO
DE MAESTRO EN CIENCIAS**

PRESENTA:

ING. JUAN JOSAFAT ZÉTERA DÍAZ

SUPERVISADA POR:

DRA. RAQUEL DÍAZ HERNÁNDEZ

**Tonanzintla, Puebla,
Febrero 2023**

© INAOE 2023

Derechos Reservados

**El autor otorga al INAOE el permiso de reproducir y
distribuir copias de esta tesis en su totalidad o en partes
mencionando la fuente.**



Agradecimientos.

Quiero agradecer a mi tutora la Dra. Raquel Díaz Hernández, por despertar mi interés en el tema de la fotoacústica y orientarme a plantear el tema de tesis, al Dr. Leopoldo Altamirano Robles por sus consejos para desarrollar la tesis, agradeciendo las facilidades, apoyo, paciencia y empatía que tuvieron durante el desarrollo del trabajo.

Al Dr. Iván Gómez Arista por su invaluable ayuda, guiándome y aconsejándome en el proceso de mi trabajo, introduciéndome en el gran campo de la fotoacústica, sin su participación este trabajo no hubiese sido posible, también al Lic. Juan José Meza Gutiérrez por su valiosa colaboración, cuando el desarrollo del trabajo se encontró con unas problemáticas me ayudo a encontrar la forma de resolverlas.

Al Instituto Nacional de Astrofísica Óptica y Electrónica por darme la oportunidad de estudiar esta maestría y ofrecerme su apoyo, a CONACyT por la beca de maestría número 1107082.

Dedicatoria

A mi familia por ser el pilar más grande en mi vida, apoyándome en todo momento a lograr las metas que me proponga, con mucho cariño a mi papá Juan Zétera Cayetano, a mi Abuelita Andrea Díaz Alcázar, a mi mamá Vianey Zétera Díaz y a mi Hermano Carlos.

INDICE

INTRODUCCIÓN	8
1.1. Antecedentes	8
1.2. Planteamiento del problema	8
1.3. Justificación	9
1.4. Objetivo general	10
1.5. Objetivos específicos	10
1.6. Hipótesis	11
MARCO TEÓRICO	11
2.1 Espectroscopia	11
2.2 Espectroscopia fotoacústica	12
2.3 Tomografía fotoacústica	14
2.4 Aprendizaje Profundo	16
2.5 Aprendizaje profundo para imágenes fotoacústicas	17
2.6 Arquitectura de las redes utilizadas	21
2.6.1 Super-Resolution Convolutional neural network (SRCNN) ,.....	22
2.6.2 Convolutional Neural Network model Rectified Linear Unit (U-net RELU)	23
2.6.3 Convolutional Neural Network model Exponential linear unit (U-net ELU)	24
2.6.4 U-net Hibrida	25
2.7.1 Reconstrucción por retroproyección lineal (LBP)	26
2.7.2 Reconstrucción Time Reversal	29
CAPÍTULO 3	30
APLICACIÓN DE LA RECONSTRUCCIÓN PAT CON RED NEURONAL	30
3.1 Entrenamiento	31
CAPÍTULO 4	35
RESULTADOS	35
4.1 Error de raíz cuadrada media (RMSE)	36
4.2 índice de similitud estructural (SSIM)	36
4.3 Correlación de Pearson (PC)	37
4.4 Análisis de resultados	37
4.4.1 Análisis de resultados para imagen palabra PAT	38
4.4.2 Análisis de resultados para imagen de Derenzo	41

4.4.3 Análisis de resultados para imagen de Vasos Sanguíneos	45
4.4.4 Análisis de resultados para imagen de Mama.....	48
4.4.5 Análisis de procesamiento de Time Reversal.....	51
CAPÍTULO 5.....	52
CONCLUSIONES Y TRABAJO A FUTURO	52
REFERENCIAS	54

LISTA DE ILUSTRACIONES

- Figura 1.** Arreglo en bloques para la formación de imágenes mediante técnica fotoacústica
- Figura 2.** Estructura convencional de una red SRCNN
- Figura 3.** Estructura de red U-net RELU
- Figura 4.** Configuración de entrenamiento para red U-net Híbrida
- Figura 5.** Proyecciones y retroproyecciones
- Figura 6.** Aplicación de filtro en retroproyecciones
- Figura 7.** Tomografía computarizada de mama
- Figura 8.** Corte axial de imagen de mama
- Figura 9.** Representación de configuración circular usada para simulación k-Wave
- Figura 10.** Comparativa de sinogramas procesados por las distintas redes
- Figura 11.** Imágenes de referencia usadas para la evaluación cuantitativa
- Figura 12.** Imágenes resultantes de la palabra PAT con LBP
- Figura 13.** Imágenes resultantes de la palabra PAT con TR
- Figura 14.** Imágenes resultantes de Derenzo con LBP
- Figura 15.** Imágenes resultantes de Derenzo con TR
- Figura 16.** Imágenes resultantes de vasos sanguíneos con LBP
- Figura 17.** Imágenes resultantes de vasos sanguíneos con TR
- Figura 18.** Imágenes resultantes de corte axial de mama con LBP
- Figura 19.** Imágenes resultantes de corte axial de mama con TR

LISTA DE TABLAS

- Tabla 1.** Artículos que utilizan técnicas de Aprendizaje profundo para las principales problemáticas que se presentan en PAI
- Tabla 2.** Arquitecturas de aprendizaje profundo mayormente utilizadas en procesamiento de PAI
- Tabla 3.** Resultados cuantitativos de imagen: palabra PAT
- Tabla 4.** Resultados cuantitativos de imagen de Derenzo
- Tabla 5.** Resultados cuantitativos de imagen de vasos sanguíneos
- Tabla 6.** Resultados cuantitativos de imagen de corte axial de mama
- Tabla 7.** Comparativa de tiempos de simulación utilizando Time Reversal

Resumen.

En esta tesis se describen los resultados obtenidos en la mejora de sinogramas de tomografía fotoacústica utilizando aprendizaje profundo y el método de reconstrucción Time Reversal . Estos resultados pueden ser de gran utilidad para generar imágenes de tomografía fotoacústica que presenten mejor calidad que las obtenidas sin usar ningún método de mejora. La investigación considera las redes mayormente utilizadas en la mejora de imágenes de tomografía fotoacústica. Las redes seleccionadas se entrenaron con el fin de disminuir el ruido que se presenta en los sinogramas de manera natural dado en la fase de adquisición de los datos (muestreo), para posteriormente reconstruir el sinograma y generar una imagen final con mayor calidad, se presenta la generación de datos sintéticos (sinogramas) generados a través de software (K-wave de Matlab) que sirven como base para realizar las pruebas con las distintas redes. Las redes que se utilizaron para este trabajo son: Super-Resolution Convolutional neural network (SRCNN), Convolutional Neural Network model Rectified Linear Unit (U-net RELU), Convolutional Neural Network model Exponential linear unit (U-net ELU), y una red U-net Híbrida.

En el trabajo se generaron sinogramas de una serie de imágenes con distintos niveles de ruido las cuales se usaron de base para probar las distintas redes y comparar las imágenes reconstruidas entre el método de retroproyección lineal y el método propuesto time reversal, y así conocer si existe una mejora de la imagen final.

La tesis muestra una mejora de la imagen reconstruida utilizando el método de reconstrucción time reversal frente al método de reconstrucción de retroproyección lineal, arrojando resultados positivos para la red SRCNN y la red U-net híbrida.

Abstract.

This thesis describes the results obtained in the enhancement of photoacoustic tomography sinograms using deep learning and the Time Reversal reconstruction method. These results can be very useful to generate photoacoustic tomography images that present better quality than those obtained without using any enhancement method. The research considers the networks mostly used in the enhancement of photoacoustic tomography images. The selected networks were trained in order to reduce the noise that is naturally present in the sinograms during the data acquisition phase (sampling), to later reconstruct the sinogram and generate a final image with higher quality, the generation of synthetic data (sinograms) generated through software (K-wave of Matlab) is presented, which serve as a basis for testing the different networks. The networks used for this work are: Super-Resolution Convolutional neural network (SRCNN), Convolutional Neural Network model Rectified Linear Unit (U-net RELU), Convolutional Neural Network model Exponential linear unit (U-net ELU), and a Hybrid U-net.

In the work were generated sinograms of a series of images with different noise levels which were used as a basis for testing the different networks and compare the reconstructed images between the linear back projection method and the proposed time reversal method, and thus to know if there is an improvement of the final image.

The thesis shows an improvement of the reconstructed image using the time reversal reconstruction method versus the linear back projection reconstruction method, yielding positive results for the SRCNN and the hybrid U-net.

CAPÍTULO 1

INTRODUCCIÓN

1.1. Antecedentes

Una de las aplicaciones más antiguas de la luz ha sido el estudio de las propiedades de las sustancias, a esto se le denomina espectroscopia y es por sí misma una ciencia completa. La forma más antigua de la espectroscopia, es la denominada espectroscopia óptica, lo que implica la interacción de fotones ópticos con la materia, su principio básico se centra en el paso de los fotones a través de la materia, estos o bien son absorbidos, transmitidos, reflejados o dispersos por ella. Se han encontrado diversas aplicaciones en la investigación de distintas propiedades de la materia en todas sus formas sin embargo presenta ciertas desventajas dependiendo del tipo de materia, en muestras muy transparentes la absorción es débil o cuando la muestra es muy opaca es difícil determinar la cantidad de absorción de los fotones. Las limitaciones se dan por la longitud de onda que es posible aplicar y el tipo de materia a analizar, es por ello que se introdujo otro tipo de técnica óptica llamada espectroscopia fotoacústica (EFA por sus siglas en inglés) [1].

Uno de los muchos usos que se le ha dado al efecto fotoacústico es su uso en la generación de imágenes médicas las cuales son de suma importancia para el diagnóstico e investigación médica, uno de ellos de gran auge en la actualidad es la imagen fotoacústica (PAI por sus siglas en inglés). La reconstrucción de estas imágenes es un proceso complicado y tardado por lo que es necesario una forma de reconstrucción eficaz, con bajo costo computacional y con resultados que permitan obtener una imagen de calidad y que presente el menor nivel de ruido posible, es por ello que se pretende utilizar técnicas de aprendizaje profundo para mejorar la PAI.

1.2. Planteamiento del problema

La imagen fotoacústica (PAI) permite obtener imágenes resueltas espacialmente de las propiedades ópticas del tejido hasta varios centímetros de profundidad, sin embargo, la extracción de parámetros de tejido relevantes a partir de los datos sin procesar requiere

la resolución de problemas de reconstrucción de imágenes inversas, que han demostrado ser extremadamente difíciles de resolver. Para obtener la imagen a partir de las señales de Tomografía fotoacústica (PAT por sus siglas en inglés) el algoritmo de reconstrucción de la imagen desempeña un papel importante. Los algoritmos convencionales de reconstrucción no iterativa (analíticos), por ejemplo, los basados en métodos de *Fast Fourier Transform* (FFT), la retroproyección filtrada (FBP por sus siglas en inglés), el retardo y la suma (DAS por sus siglas en inglés), los basados en la transformada de Radón como lo es el método de inversión de tiempo, estos prevalecen debido a su rápida velocidad, también tenemos los métodos iterativos como lo son: los métodos algebraicos, y los métodos estadísticos (mínimos cuadrados y Poisson), sin embargo, la imperfección de los algoritmos convencionales producen imágenes distorsionadas, especialmente en la configuración de vista limitada, los algoritmos iterativos tiene buena resolución, pero son muy tardados. Por otro lado, el algoritmo basado en FFT es rápido pero su resolución no es tan buena como la del iterativo. En la actualidad se investiga el problema de cómo obtener imágenes de mayor resolución y reducir el porcentaje de ruido en el proceso de reconstrucción. Por lo que es necesario desarrollar un método que combine las ventajas de los ya existentes y así poder minimizar o hasta erradicar los problemas con los que actualmente se enfrenta la reconstrucción de imágenes.

1.3. Justificación

La tomografía fotoacústica (PAT), es una técnica no invasiva y no ionizante que nos permite la adquisición y reconstrucción de objetos. Dicha reconstrucción, se basa en la detección de ondas en el rango del ultrasonido, que pueden generarse tanto en tejidos, como en materiales no biológicos que fueron previamente excitados por pulsos láser cortos del orden de unos nanosegundos. Actualmente la tomografía fotoacústica es un método que ha despertado el interés de varias áreas de la ciencia, física, medicina, ingeniería, matemáticas entre otras. Esta técnica de imagen medica tiene mucho potencial ya que permite reconstruir imágenes de estructuras pequeñas, desde unas cuantas micras, hasta objetos que midan algunos centímetros; además procesa imágenes a profundidades de hasta 7cm en tejidos por lo que es una herramienta con grandes ventajas frente al ultrasonido convencional, pudiendo ser una opción frente a

estudios como la Imagen por resonancia Magnética (IMR por sus siglas en inglés), y la Tomografía por emisión de positrones (PET por sus siglas en inglés), entre otros. Pese a las grandes ventajas que tiene la tomografía fotoacústica su uso a nivel clínico ha sido limitado debido a las problemáticas que presenta en el apartado de reconstrucción de imágenes ya que es un proceso muy lento o la calidad de imagen producida es baja debido al ruido, por lo que es de suma importancia mejorar el proceso de reconstrucción y así obtener imágenes médicas en un menor tiempo y con un mayor grado de calidad: menor ruido, mayor nitidez y contraste que puedan ser usadas a nivel clínico como una herramienta de diagnóstico.

1.4. Objetivo general

Estudio de los mecanismos para la reducción de ruido en imágenes fotoacústicas aplicando redes neuronales a los sinogramas originales.

1.5. Objetivos específicos

- Investigar los modelos de aprendizaje profundo utilizados en la reconstrucción de imágenes PAT.
- Generación de datos sintéticos mediante software (k-Wave).
- Análisis de los algoritmos más usados para el filtrado de sinogramas producidos por el proceso inherente a la tomografía foto acústica.
- Estudio de la calidad de imagen al aplicar las redes neuronales.
- Evaluación de los resultados obtenidos.

1.6. Hipótesis

La reducción de ruido utilizando redes neuronales en los sinogramas genera una mejor calidad de imagen fotoacústica.

CAPÍTULO 2

MARCO TEÓRICO

En el año de 1880 el científico inventor Alexander Graham Bell describió que cuando un sólido es iluminado con un haz de luz modulada periódicamente dentro de una celda cerrada, un sonido audible podía ser oído mediante un tubo unido a la celda [1]. Graham experimento con una variedad de materiales, solidos, líquidos y gases en ese entonces no se le dio reconocimiento a su descubrimiento y quedo olvidado [1]. Posteriormente en el año de 1930 fue reexaminado el efecto optoacústico en gases, desde entonces se ha convertido en una técnica bien establecida usada en el análisis de gases. Los fotones que son absorbidos por el gas se convierten en energía cinética de las moléculas de gas, lo que da lugar a fluctuaciones de la presión dentro de la celda, sin embargo, no se le dio uso al efecto fotoacústico en solidos por 90 años [2].

2.1 Espectroscopia

La espectroscopia es el estudio de las propiedades de las sustancias. El primer tipo de espectroscopia fue el óptico que se basa en la interacción de fotones con la materia, en donde los fotones pasan a través de la materia y son o bien absorbidos, transmitidos reflejados o dispersos por ella. La mayoría de los espectroscopios se basan en la detección y el análisis de los fotones que se transmiten o reflejan de la muestra, es una técnica versátil y no destructiva en la naturaleza, y tiene aplicaciones en la investigación de las distintas propiedades de la materia [1]. Esta técnica presenta un problema en los casos en la que la muestra es muy transparente y por lo tanto su absorción es muy débil, o cuando la muestra es muy opaca, es difícil determinar la cantidad de absorción de los fotones en el caso de muestras opacas. Técnicas como dispersión Raman [3], la reflexión

difusa [15], la reflexión total atenuada etc. han sido utilizadas para resolver este problema. Sin embargo, estas también tienen limitaciones con respecto a las regiones de longitud de onda que se podrían utilizar. La diversidad de muestras que podrían analizarse por estas técnicas no es muy grande. Por otra parte, los datos de estas técnicas no son fáciles de analizar. Para este tipo de muestras que son débilmente absorbentes, con alta dispersión u opacas, y donde las mediciones utilizando técnicas de espectroscopia convencional no pueden ser utilizadas con eficacia, se introdujo una nueva técnica óptica para la detección espectroscópica y fue llamada espectroscopia fotoacústica (EFA) [1].

2.2 Espectroscopia fotoacústica

La (EFA) se considera solo un redescubrimiento de un antiguo resultado que fue propuesto originalmente por Alexander Graham Bell en 1880 [1]. En ese entonces, se denominó como el efecto óptico-acústico. Nombres como el efecto opto-térmico y fotoacústico todavía se utilizan para describir este fenómeno. La diferencia básica de este efecto con la espectroscopia óptica convencional es que, en la espectroscopia convencional, la cantidad de energía absorbida por la muestra irradiada se mide directamente por la luz transmitida o reflejada. El efecto fotoacústico es esencialmente la generación y detección de efectos acústicos o termo-elásticos, que resultan de la absorción de cualquier tipo de radiación electromagnética modulada o pulsada. La absorción de la muestra provoca que los átomos o moléculas pasen a un nivel de excitación más alto [1]. Como consecuencia de la exposición y absorción de la muestra a una radiación modulada o pulsada, todos o parte de los átomos y moléculas excitadas pierden esta condición para regresar de nuevo al nivel inferior, a través de diversas vías de desexcitación no radiactiva. Esto hace que el calor se genere en la muestra a la misma frecuencia de modulación de la luz incidente en la muestra.

Este calor generalmente aparece como la energía cinética de los gases y como energía de vibración de iones o átomos en líquidos. En la EFA, es este calor el que detectan los sólidos, y se realiza mediante la colocación de la muestra en una celda hermética que contiene un transductor acústico sensible [1]. El calor modulado producido en la muestra se transfiere a la capa media de gas no absorbente cerca de la superficie de la muestra

contenida en la celda, haciendo que se comprima y se expanda. Por lo tanto, esta capa actúa como un pistón acústico que impulsa el resto del gas en la celda. Las variaciones de presión producidas de este modo se llevan al transductor acústico, que emite una señal eléctrica que es proporcional a la cantidad de absorción de la radiación por la muestra [1].

En el año de 1930 Viengerov consideró la utilización de una celda fotoacústica resonante para detectar la presión. Se encontró que la variación del espectro fue proporcional a la concentración de la molécula de gas absorbente. Esto se puede considerar como el inicio de la aplicación de la EFA para el análisis de trazas de compuestos en estado gaseoso [1]. Gorelik sugirió y mostro que la fase de la señal fotoacústica contenía información acerca de las tasas de transferencia de energía entre los grados de vibración y de libertad de traslación de un gas. Esto se demostró experimentalmente por Slobodskaya en 1948 [1].

Con el desarrollo del láser en 1960, se produjo otro renacimiento en el análisis de gases. Las altas potencias alcanzadas con la tecnología láser permitieron la detección y el análisis de concentraciones de gas a ultra bajas. Otras ventajas importantes del láser como fuente de luz se deben al alto grado de pureza espectral, la gran estabilidad y la reproducibilidad. Kerr y Atwood [4] en 1968 fueron los primeros en desarrollar espectrómetros de absorción iluminados por láser para el análisis de gases, uno de los cuales utilizaba un láser de rubí pulsado y el segundo un láser de CO₂ de onda corta, con detección acústica mediante un micrófono de capacitancia. Kreuzer siguió en 1971 con excitación láser de onda corta He-Ne en una célula equipada con un micrófono electret, donde demostró la detección de ultra-trazas de gas a niveles de *parts per billion (PPB)*.

En la EFA es esencial la detección de calor producido en la muestra. Por esto, la EFA puede ser referida como una técnica calorimétrica. Como la propia muestra genera la señal fotoacústica, esta técnica se puede utilizar en una amplia gama de radiaciones electromagnéticas con la única limitación de que las ventanas ópticas adecuadas que corresponden a la longitud de onda de operación puedan utilizarse en la celda fotoacústica. En comparación con otras técnicas calorimétricas, la EFA es más simple,

más rápida, no destructiva en la naturaleza, más sensible, tiene un tiempo de subida del detector superior, etc. [1]. Incluso, es capaz de detectar de manera efectiva las variaciones de presión que corresponden a variaciones de temperatura de aproximadamente $10^{-6} \text{ }^{\circ}\text{C}$ [5].

Los inconvenientes de la técnica de EFA son que, dado que la señal fotoacústica es detectada por un transductor sensible de volumen, el tiempo de respuesta se limita al tiempo de tránsito de la onda de sonido en el gas dentro de la celda y por la respuesta de baja frecuencia relativamente más débil del micrófono. Estos factores limitan el tiempo de respuesta de un sistema típico de fotoacústico-gas-micrófono al orden de aproximadamente 100 micro segundos.

En las últimas dos décadas, el efecto Fotoacústico ha encontrado muchas aplicaciones en la espectroscopia de sólidos líquidos y gases, en la medicina y la biología, el análisis de trazas, vigilancia de la contaminación, la teledetección, la física de los semiconductores y materiales amorfos, en películas delgadas, entre otras [1].

2.3 Tomografía fotoacústica

Las modalidades de obtención de imágenes ópticas tridimensionales de alta resolución comercialmente disponibles, han tenido un impacto fundamental en la biomedicina, incluida la microscopía confocal, la microscopía de dos fotones y la tomografía de coherencia óptica, Desafortunadamente, tales herramientas no pueden penetrar el tejido biológico más profundamente que el transporte óptico, lo que significa un recorrido libre de 1 mm en la piel. La tomografía fotoacústica, que combina un fuerte contraste óptico y una alta resolución ultrasónica en una única modalidad, ha superado esta limitación fundamental de la profundidad y ha logrado imágenes ópticas de alta resolución y de gran profundidad.[6]

La imagen fotoacústica (PAI) es una modalidad de imagen emergente que promete una medición en tiempo real, no invasiva y libre de radiación de las propiedades ópticas del tejido. A diferencia de otras modalidades de imágenes ópticas, PAI induce la aparición de señales acústicas para permitir la obtención de imágenes estructurales de cromóforos (estructuras moleculares que absorben la luz) hasta varios centímetros de profundidad

en el tejido. Las profundidades de penetración típicas reportadas varían desde 1 cm hasta 6 cm, dependiendo en gran medida del tejido fotografiado. Esta penetración de alta profundidad es posible porque la dispersión acústica de las ondas sonoras que surgen es órdenes de magnitud menor que la dispersión óptica de la luz incidente en el tejido biológico.[7] En esencia, una imagen fotoacústica PAI se puede considerar como una imagen de ultrasonido en la que el contraste no depende de las propiedades mecánicas y elásticas del tejido, sino de sus propiedades ópticas, específicamente la absorción óptica [8]. En el proceso de formación de una imagen fotoacústica, el tejido se irradia con pulsos láser cortos, con una longitud de onda cercana al infrarrojo (700 a 900 nm), entre otras longitudes de onda. La absorción de luz láser transfiere energía óptica al tejido, lo que provoca un aumento de presión transitoria [9]. Este aumento de presión inicial actúa como una fuente acústica que genera una onda que se propaga a través del tejido. La onda acústica se puede detectar con un transductor de ultrasonido. La imagen puede ser creada escaneando, tanto el ultrasonido como el haz óptico y deduciendo la resolución de profundidad desde el tiempo de vuelo de la onda de presión. En la Figura 1 se observa un esquema que explica el proceso básico de formación de una PAI [10].

Existen dos modalidades de PAI que presentan rápido crecimiento: la microscopía fotoacústica (PAM por sus siglas en inglés) y la tomografía computarizada fotoacústica (PACT por sus siglas en inglés). Estas técnicas obtienen imágenes de diferentes maneras: la primera es una imagen formada de escaneo punto por punto, y la segunda es una imagen reconstruida mediante la adquisición de señales Fotoacústicas en diferentes posiciones.[8] La microscopía fotoacústica de resolución óptica (OR - PAM) es una modalidad excelente para la obtención de imágenes biomédicas in vivo, ya que proporciona de forma no invasiva información morfológica y funcional de alta resolución sin la necesidad de agentes de contraste exógenos.[11]

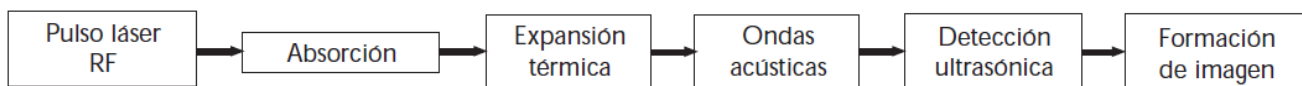


Figura 1. Arreglo en bloques para la formación de imágenes mediante la técnica fotoacústica.

2.4 Aprendizaje Profundo

En los últimos años, el aprendizaje profundo se ha desarrollado rápidamente en el área de la visión por computadora, y ha comenzado a atraer un intenso interés de investigación en problemas de reconstrucción de imágenes para la imagen médica [12-3]. La aplicación de métodos de aprendizaje profundo ha llevado a éxitos impresionantes en el contexto de las imágenes médicas y también ha desarrollado su uso en el campo de PAI. Los métodos de aprendizaje profundo poseen ventajas únicas que pueden facilitar la traducción clínica de PAI, como tiempos de cálculo extremadamente rápidos y el hecho de que pueden adaptarse a muchos problemas [7].

Podemos definir el aprendizaje profundo como una clase de técnica de aprendizaje automático, donde la información se procesa en capas jerárquicas para comprender las representaciones y características de los datos en niveles crecientes de complejidad. En la práctica, todos los algoritmos de aprendizaje profundo están conformados con redes neuronales, que comparten algunas propiedades básicas comunes. Todos están formados por neuronas interconectadas que están organizadas en capas. Donde difieren, es en la arquitectura de la red (o la forma en que se organizan las neuronas en la red) y, a veces, en la forma en que están entrenados [13]. Se puede clasificar al aprendizaje profundo como un método de representación-aprendizaje [14], el cual, mediante la composición de módulos no lineales, es capaz de aprender distintas representaciones de los datos. El punto clave de esta técnica es que no requiere de un experto humano que especifique dichos módulos, sino que la red, a través de los datos, es capaz de aprender la representación de los mismos para optimizar la tarea subsecuente, ya sea clasificación o regresión. El Aprendizaje profundo es, en esencia, una extensión de las Redes Neuronales Artificiales, las que a su vez son, una extensión del Perceptrón presentado por Frank Rosenblatt en 1958 [12].

Se puede dividir el aprendizaje automático en tres categorías de acuerdo con el proceso de entrenamiento. La primera categoría es el aprendizaje por refuerzo, en donde, la interacción entre el agente construido y el entorno logra el mayor beneficio o resuelve

problemas específicos a través de estrategias de aprendizaje. Una de sus aplicaciones más famosas es AlphaGo [3], que es un sistema de Go-playing desarrollado por DeepMind, que incluso derrota a los mejores jugadores de Go del mundo. La segunda categoría es el aprendizaje no supervisado que consiste en clasificar o agrupar en base a muestras de entrenamiento con categorías desconocidas (sin etiquetar) y descubrir patrones entre ellas. Un ejemplo común de aprendizaje no supervisado es el análisis de conglomerados. La última categoría es el aprendizaje supervisado, en el cual, los conjuntos de datos emparejados hacen que los algoritmos de aprendizaje automático convencionales se centren en ella. Entrena con los datos etiquetados para encontrar las reglas de los patrones y luego produce la etiqueta correcta en los datos invisibles. Por ejemplo, muchas imágenes médicas, incluidos los problemas de reconstrucción de imágenes PA [15,16], se basan en el aprendizaje supervisado.

El aprendizaje profundo permite que los modelos computacionales que se componen de múltiples capas de procesamiento aprendan representaciones de datos con múltiples niveles de abstracción. Estos métodos han mejorado drásticamente el estado de la técnica en reconocimiento de voz, reconocimiento de objetos visuales, detección de objetos y muchos otros dominios, como el descubrimiento de fármacos y la genómica. Las redes convolucionales profundas han producido avances en el procesamiento de imágenes, video, voz y audio, mientras que las redes recurrentes han arrojado resultados satisfactorios sobre datos secuenciales como texto y voz [17].

2.5 Aprendizaje profundo para imágenes fotoacústicas

Dado que la tomografía fotoacústica es una técnica para adquisición de imágenes 3D, las aplicaciones con aprendizaje profundo lo son aún más. En este sentido, algunos algoritmos que se utilizan actualmente para generar datos experimentales son utilizados para entrenar gran parte de los métodos dada la poca cantidad de datos clínicos con los que se cuentan.

La tomografía fotoacústica permite obtener imágenes, con alta resolución espacial, de las propiedades ópticas del tejido hasta varios centímetros de profundidad, sin embargo, la extracción de parámetros de tejido relevantes a partir de los datos sin procesar requiere la resolución de problemas de reconstrucción de imágenes inversas, que han demostrado ser extremadamente difíciles de resolver. La aplicación de métodos de aprendizaje profundo se ha disparado recientemente en popularidad, lo que ha llevado a éxitos impresionantes en el contexto de las imágenes médicas y también ha encontrado un primer uso en el campo de PAI. Los métodos de aprendizaje profundo poseen ventajas únicas como lo son tiempos de cálculo extremadamente rápidos y el hecho de que pueden adaptarse a muchos problemas [7].

Una revisión extensa de la aplicación del aprendizaje profundo en PAI se ha realizado para evaluar cuantitativamente el análisis de su desempeño, teniendo como resultado la siguiente tabla, que reporta el número absoluto y relativo de artículos que utilizan datos experimentales para la prueba o para el entrenamiento de redes.

Problema	Datos experimentales de prueba	Datos experimentales para entrenamiento
Problema inverso acústico	10(43%)	1(4%)
Post- procesamiento de la imagen	23(79%)	14(48%)
Problema inverso óptico	1(33%)	1(33%)
Anotación de imagen semántica	14(54%)	11(42%)

Tabla 1. Número de artículos que utilizan técnicas de Aprendizaje profundo para 4 de las principales problemáticas que se presentan en PAI [7].

De una lista de 83 artículos revisados que aplican técnicas de Aprendizaje profundo enfocado a PAI, se obtuvo la siguiente tabla donde se presentan las principales arquitecturas utilizadas en la actualidad:

Arquitectura	Frecuencia
U-Net	43(52%)
CNN	18(22%)
FCNN	9(11%)
ResNet	5(6%)
DenseNet	3(4%)
RNN	1(1%)
INN	1(1%)
Otras	3(4%)

Tabla 2. Arquitecturas de Aprendizaje profundo usadas actualmente para procesamiento de PAI. CNN=Red Neuronal Convolutacional, FCNN=Red Neuronal Totalmente conectada, ResNet=Red Neuronal Residual, DenseNet = Red Neuronal Densa, RNN = Red Neuronal Recurrente, INN = Red Neuronal Invertible [7].

Dado que existe una falta de datos clínicos, alrededor del 65% de los trabajos que utilizan Técnicas de Aprendizaje profundo aplicado a PAI se basan en datos simulados para el entrenamiento supervisado, pero solo el 35% de los trabajos han validado sus métodos con datos in vivo. Por ello es necesario el apoyo de software que permita generar datos sintéticos a partir de simulaciones. Actualmente se tiene disponible 5 programas de código abierto para la generación de dichos datos:

- k-Wave: es una toolbox de MATLAB que nos permite simular y reconstruir campos de ondas fotoacústicas, se ha diseñado para facilitar modelos de PAI incluyendo el modelado de dispositivos de detección de ondas. Es uno de los más utilizados y se basa en una solución pseudoespectral en el dominio del tiempo de las ecuaciones fotoacústicas.

- NIRFAST: Fue desarrollado para modelar la propagación de la luz en tejido con longitudes de onda cercanas al infrarrojo, este programa es capaz de obtener imágenes ópticas o funcionales de una o varias longitudes de onda a partir de datos simulados y medidos.
- mcxyz: es una herramienta de simulación que se basa en el modelo de Monte Carlo de transporte de luz para simular la propagación de los fotones en tejidos heterogéneos. Este método utiliza las propiedades de absorción y dispersión para encontrar las trayectorias de los fotones a través del medio; para la simulación se basa en voxeles, asignando a cada voxel un tipo particular de tejido.
- Monte Carlo *extreme* (MCX): basado en el modelo de Monte Carlo de transporte de luz, simula el transporte de fotones apoyado de procesamiento gráfico GPU, esto es una ventaja ya que acelera el procesamiento mediante una o varias GPU's.
- Toast++: es una serie de bibliotecas que ayuda a simular la propagación de la luz en medios de alta dispersión con una distribución interna heterogénea, contiene solucionadores numéricos basados en el método de elementos finitos, discretización de Galerkin discontinuo y elementos de contorno.

Estos programas permiten obtener datos con los cuales trabajar y probar los modelos, sin embargo, los modelos puramente entrenados con datos sintéticos presentan un bajo rendimiento cuando se aplican a datos clínicos. Estas diferencias entre PAI experimentales y las generadas por los modelos computacionales se conocen como brecha de dominio y puede provocar que en aplicaciones in vivo fallen; esto sucede por las distribuciones diferentes de los datos. Para que estas brechas sean menos significativas se puede abordar el problema desde distintos puntos:

- Creación de métodos para obtener simulaciones más realistas, mejorando la geometría anatómica del tejido.

- Aplicar métodos de adaptación de dominio como los usados en el campo de la visión por computadora, esto podría ayudar a traducir las imágenes del dominio de la imagen sintética a la de una PA real.
- Métodos para refinar el proceso de entrenamiento, como lo son aumentar los datos específicos para PAI, el desenredo del contenido, o los cambios de arquitectura específica del dominio.

Uno de los principales cuellos de botella de la aplicación de algoritmos de aprendizaje profundo a PAI es la falta de datos de entrenamiento fiables. Esto puede deberse, en particular, a la falta de información sobre las propiedades del tejido óptico subyacente, o a la distribución de la presión inicial al adquirir las mediciones experimentales. Esta falta de datos experimentales para entrenamiento se acentúa más en el apartado de problema inverso óptico y acústico. En el campo de las imágenes médicas, sólo unos pocos estudios prospectivos garantizan la fidelidad de los métodos de aprendizaje profundo. Uno de los principales problemas es que los algoritmos no son directamente utilizables por especialistas médicos debido a limitaciones técnicas y/o burocráticas. Dado el hecho de que la mayoría de los enfoques utilizan datos simulados para entrenar sus algoritmos, hay una alta probabilidad de que muchos de los algoritmos presentados, aunque den resultados excelentes en los datos publicados puedan fallar en un escenario clínico. Esto puede atribuirse al hecho de que los datos de entrenamiento pueden sufrir de varias deficiencias en comparación con la distribución de datos en la realidad, como una diferencia significativa en los datos (brecha de dominio), un número insuficiente de muestras, o un sesgo de selección [7].

2.6 Arquitectura de las redes utilizadas

En esta sección se explicarán de forma simplificada las redes que se consideraron importantes para el presente trabajo y que resuelven el problema de la reconstrucción de imágenes foto acústicas.

2.6.1 Super-Resolution Convolutional neural network (SRCNN),

Para este tipo de red, primeramente, se considera una imagen de baja resolución, posteriormente se escala a un tamaño deseado mediante algún tipo de interpolación, se denota la imagen interpolada como Y . El objetivo es recuperar una imagen Y , y una imagen $F(Y)$ que sea lo más similar posible a la imagen de alta resolución X , es necesario realizar un mapeo F que consiste en 3 operaciones:

Representación y extracción de parches: esta operación extrae parches (superpuestos) de la imagen de baja resolución Y , y representa cada parche como un vector, estos vectores comprenden un conjunto de mapas de características cuyo número es igual a la dimensión de los vectores.

Mapeo no lineal: esta operación mapea de forma no lineal un vector en otro vector, cada vector mapeado es conceptualmente la representación de un parche de alta resolución, estos vectores constituyen otro conjunto de mapas de características.

Reconstrucción: esta operación agrega las representaciones de alta resolución en forma de parches para generar la imagen final de alta resolución, esta imagen debería ser similar a la imagen real X .

La figura 2 muestra la representación del funcionamiento de una red SRCNN donde en la entrada ingresa una imagen de baja resolución (Y), posteriormente en la primera capa convolucional se realiza un proceso de extracción de parches de características con los cuales se crea un mapa ($N1$) de características de la imagen de entrada (Y). En la segunda capa convolucional se asignan estos mapas de características de forma no lineal a una representación de parches de alta resolución, mapa ($N2$). La última capa combina las predicciones dentro de una vecindad espacial para producir la imagen de salida de alta resolución $F(Y)$.

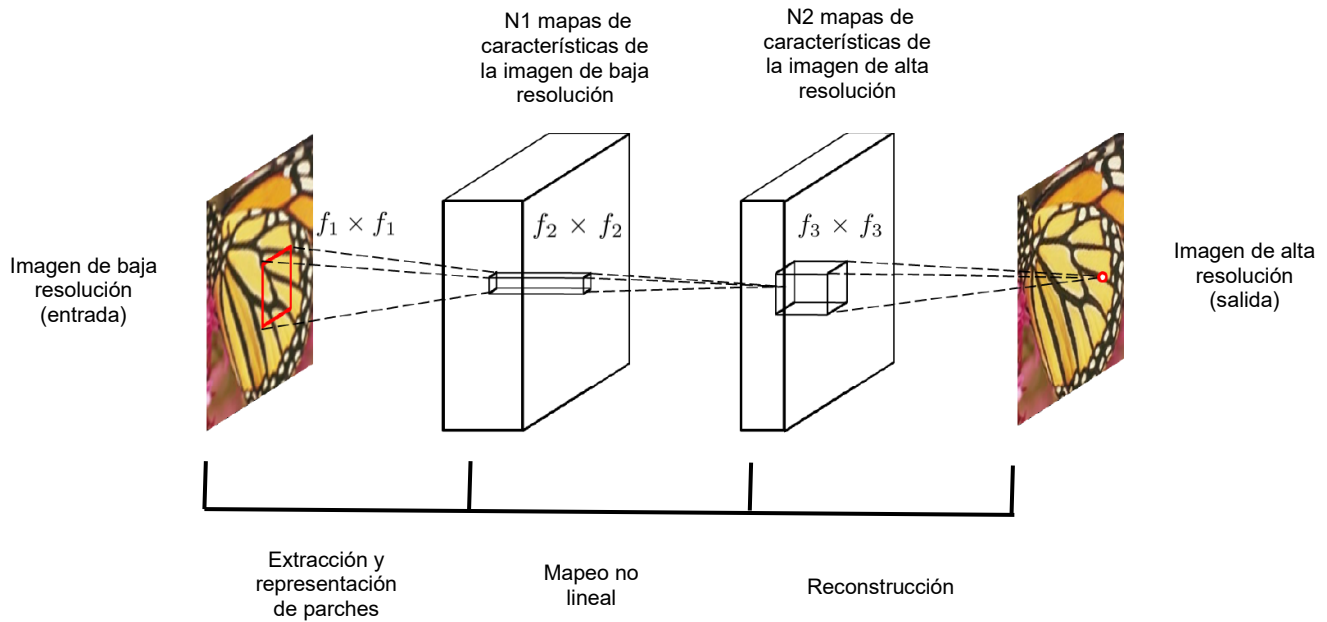


Figura 2 Estructura de una red SRCNN, dada una imagen de baja resolución, se extrae un conjunto de mapas de características. Posteriormente se asignan estos mapas de características de forma no lineal a representaciones de parches de alta resolución. Al final se realiza un proceso de combinación de predicciones dentro de una vecindad espacial para dar lugar a la imagen final de alta resolución.[18]

2.6.2 Convolutional Neural Network model Rectified Linear Unit (U-net RELU)

La U-net es uno de los modelos de *Convolutional Neural Networks* (CNN por sus siglas en inglés) propuestos para la segmentación de imágenes. Utiliza el aprendizaje residual que es una de las técnicas utilizadas al entrenar una red para que aprenda las diferencias entre los datos reales y los datos de entrada [15], este método puede hacer que la red converja más rápido y con eficiencia. Al añadir el esquema de aprendizaje residual a la red U-net mostró un mayor rendimiento en la eliminación de artefactos en imágenes médicas.

En la Figura 3 se muestra un esquema de la estructura de una red U-net RELU convencional en la cual se utilizan capas de agrupación, el uso de estas produce un resultado más rápido que el uso de una convolución basada únicamente en zancadas. En la primera capa se tiene una entrada de un único canal, posteriormente se realizan convoluciones de un tamaño de 3X3 que generan una salida de 64 canales en la primera capa, el número de canales se duplica en la capa de convolución debido a la zancada

con valor 2, hasta tener un número de 1024 canales, posteriormente se invierte el proceso haciendo deconvoluciones para generar un solo canal de salida en la estructura U-net RELU.

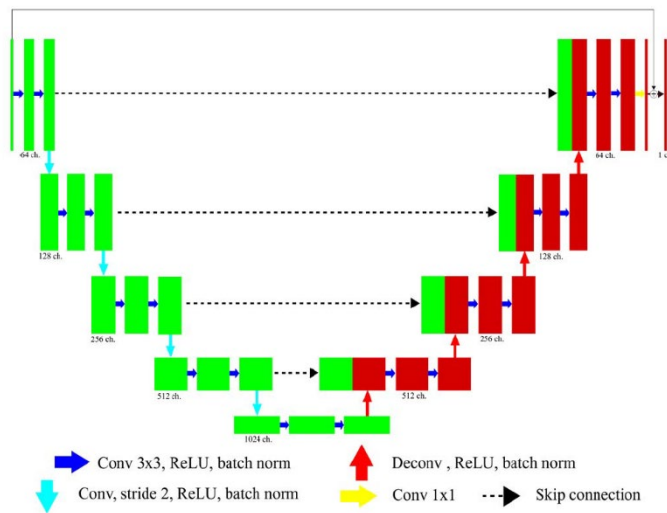


Figura 3 Estructura de la red U-Net- RELU convencional con capas de agrupación. [15]

2.6.3 Convolutional Neural Network model Exponential linear unit (U-net ELU)

La "unidad lineal exponencial" (ELU por sus siglas en inglés) acelera el aprendizaje en redes neuronales profundas y conduce a mayor precisión de clasificación. Al igual que las unidades lineales rectificadas (RELU por sus siglas en inglés), las RELUs con fugas (LRELU por sus siglas en inglés) y las RELUs parametrizadas (PReLU por sus siglas en inglés), las ELUs alivian el problema de la desaparición del gradiente a través de la identidad para valores positivos. Sin embargo, las ELUs tienen características de aprendizaje mejoradas en comparación con otras funciones de activación. En contraste con las RELUs, las ELUs tienen valores negativos, lo que les permite acercar las activaciones de las unidades medias a cero, como la normalización por lotes, pero con menor complejidad computacional.

Los desplazamientos de la media hacia cero aceleran el aprendizaje al acercar el gradiente normal al gradiente natural de la unidad debido a la reducción del efecto de desplazamiento del sesgo. Aunque los LRELU y los PReLU también tienen valores negativos, no garantizan un estado de desactivación resistente al ruido. Las ELUs saturan

a un valor negativo con entradas más pequeñas y, por lo tanto, disminuyen la variación y la información propagada. Por lo tanto, las ELU codifican el grado de presencia de fenómenos particulares en la entrada, mientras que no modelan cuantitativamente el grado de su ausencia. En los experimentos, las ELU no sólo conducen a un aprendizaje más rápido, sino también a un rendimiento de generalización significativamente mejor que las RELUs y las LRELU en redes con más de 5 capas [16].

2.6.4 U-net Híbrida

Esta red U-Net (Híbrida) es la unión de las redes U-net ELU y U-net RELU, debido a que combina las bondades de ambas, es capaz de proporcionar la generalización al modelar eficazmente los valores negativos en el sinograma a través de la red U-net ELU aplicada a las capas finales, al mismo tiempo que mantiene las ventajas de las RELUs como lo es su capacidad de converger rápidamente.

En la figura 4 se muestra la configuración de entrenamiento utilizada para la red U-net Híbrida, en la cual en el inciso a) se muestra un sinograma obtenido mediante una simulación de k-Wave con dimensiones de 256x512, la simulación se realizó utilizando 100 transductores de ultrasonido con un ancho de banda del 70%, en los sinogramas utilizados se agregó ruido variable que va en un rango de 10 a 70dB de signal-to-noise ratio (SNR por sus siglas en inglés), posteriormente se secciona en parches de 64x64 que son los datos de entrada de la red U-net híbrida, el parche de salida que tiene un tamaño de 64x64 que se utiliza para calcular la pérdida de error de raíz cuadrada media (RMSE por sus siglas en inglés), comparándolo con un parche correspondiente a una simulación realizada con 200 transductores de ultrasonido con un ancho de banda completo. Estos datos se utilizan para entrenar la red mediante retropropagación con el fin de mejorar el ancho de banda, super-resolución y la eliminación de ruido, la red comienza realizando convoluciones de un tamaño 3x3 con la red RELU y en las últimas dos capas de la red híbrida se utilizan convoluciones de 3x3 con la red ELU, la cual le permite modelar los valores negativos que contenga el sinograma. [19].

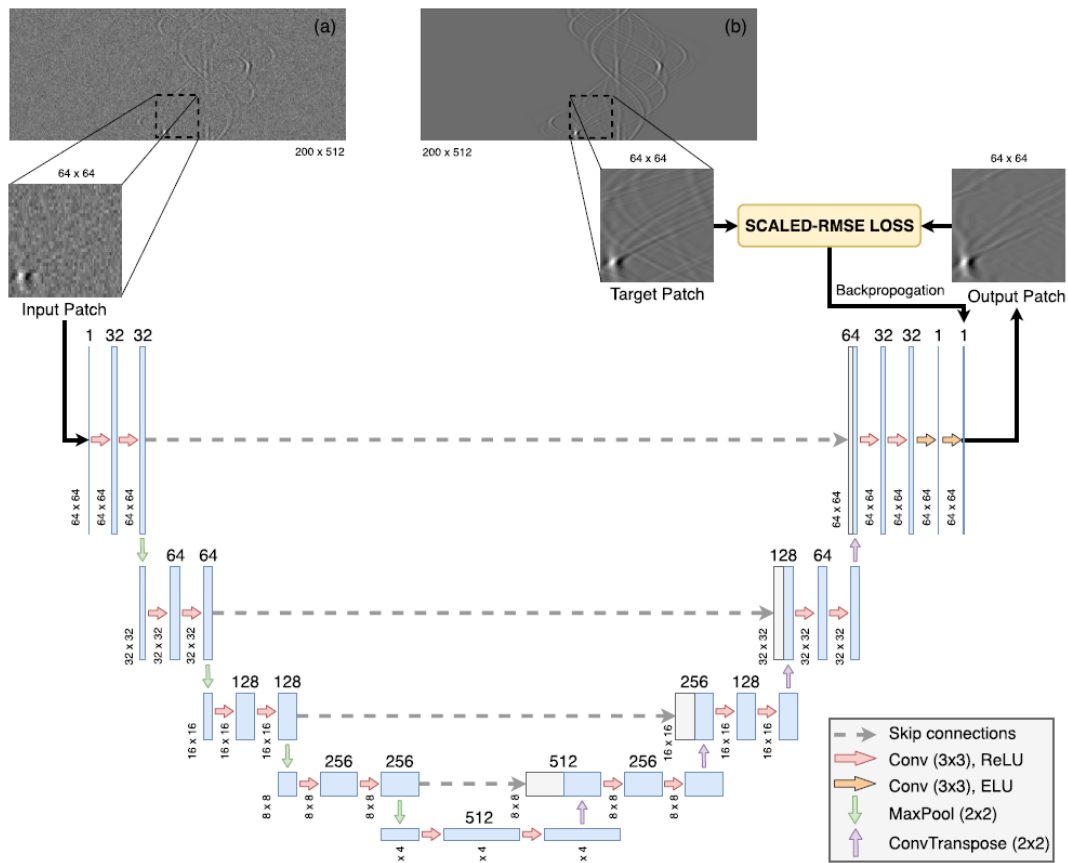


Figura 4 Configuración de entrenamiento para la arquitectura basada en U-Net híbrida. Esta comparación se utiliza para entrenar el modelo propuesto para la mejora del ancho de banda, la superresolución y la eliminación de ruido [19].

Estas arquitecturas se utilizaron para mapear primero los parches con ruido y con un ancho de banda limitado, utilizando 100 detectores, a una imagen sin ruido con ancho de banda completo utilizando 200 detectores.

2.7.1 Reconstrucción por retroproyección lineal (LBP)

El modelo de propagación de ondas fotoacústicas se puede expresar en términos de un sistema lineal de ecuaciones, por lo que el modelo se puede representar como:

$$Ax = b$$

Donde A es la matriz del sistema que contiene las respuestas al impulso de todos los píxeles en la región de las columnas de la imagen, x es el vector desconocido que representa el aumento de la presión inicial $P(0)$, y b es el vector de medición. Los datos variables en el tiempo que representan la respuesta al impulso registrada correspondiente a cada píxel se apila como un vector largo, que representa la columna de la matriz del sistema. Así, el número de columnas de la matriz del sistema (A) es igual al número de píxeles en el dominio de la imagen.

Por lo que es posible obtener una imagen X_{Lbp} mediante la reconstrucción de retroproyección lineal (LBP).

$$X_{Lbp} = A^T b$$

Donde X_{Lbp} es el vector imagen y A^T es la transposición de la matriz del sistema [19].

En el algoritmo de retroproyección se considera una imagen 2D. Se considera también un solo punto con cierto grado de intensidad como se muestra en la Figura 5 (a) donde la altura de la barra indica la intensidad del punto en el objeto. Se toman varias proyecciones desde distintos ángulos como se muestra en (b). Posteriormente es necesario reconstruir la imagen utilizando estas proyecciones. Al observar las proyecciones es posible identificar un pico. Este pico es la suma de toda la actividad a lo largo de la trayectoria de la proyección, para reconstruir la imagen es necesario redistribuir la actividad en el pico a su trayectoria original. El problema es que no se sabe dónde es que hay que poner más actividad a lo largo de la trayectoria y en que parte poner menos, por lo que se pone la misma cantidad de actividad a lo largo de todo el camino y la cantidad es el máximo del pico de la proyección. Se repite lo mismo para todas las proyecciones tomadas desde todos los ángulos como se muestra en (c), a este proceso se le llama retroproyección. Si se hace una retroproyección desde todos los ángulos se obtiene una imagen similar a (d). Después de aplicar la retroproyección la imagen no es igual a la original si no que es una versión borrosa, para afinar la imagen se aplica un filtro a las proyecciones, se introducen ramas negativas antes de la

retroproyección, esto da como resultado una imagen clara, lo que se observa en la figura 6 [20].

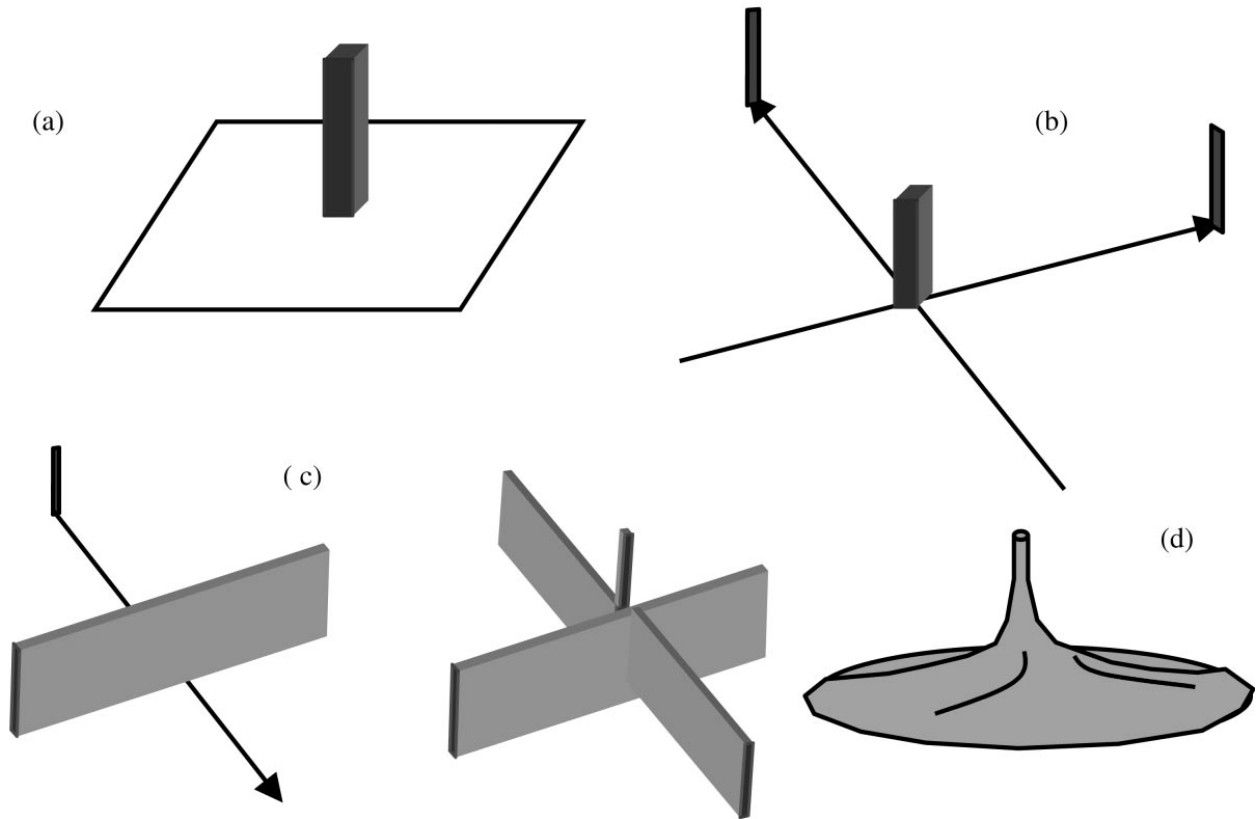


Figura 5 Proyecciones y retroproyecciones

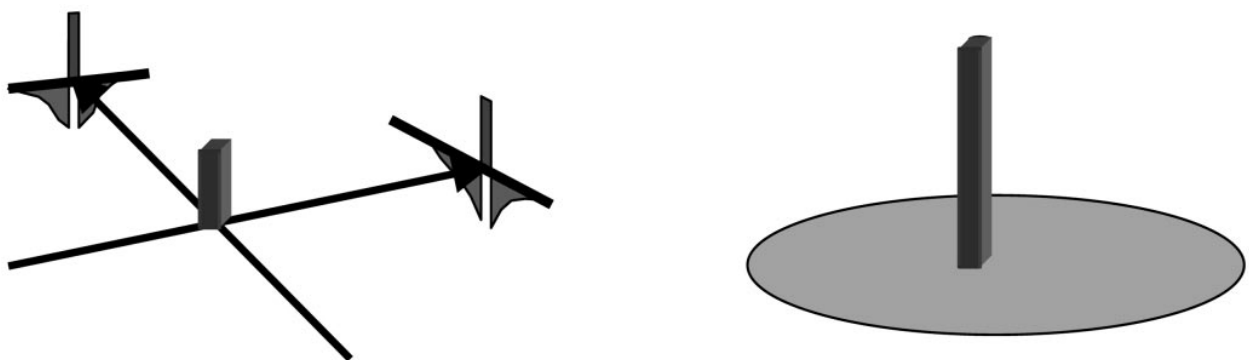


Figura 6 Aplicación de filtro para eliminar el desenfoque de la imagen

2.7.2 Reconstrucción Time Reversal

Para poder reconstruir una imagen de tomografía fotoacústica debemos conocer la función de distribución $f(x)$, la energía absorbida por el tejido provoca una expansión térmica, lo que conduce a una onda de presión ultrasónica $p(x, t)$ que se propaga a través del objeto. Esta presión se mide a través de sensores de ultrasonido situados en una superficie S que rodea al objeto, en nuestro caso con una configuración circular. El valor de $f(x)=p(x,0)$ de la presión inicial en un tiempo $t=0$ es decir que al momento de irradiar el objeto no es la función de absorción $a(x)$ si no que es proporcional a ella por lo que se debe recuperar $f(x)$.

El modelo matemático de la tomografía fotoacústica es el siguiente:

$$\begin{aligned} \rho_{tt} &= C^2(x)\Delta_x\rho, & t \geq 0, x \in \mathbb{R}^3 \\ \rho(x, 0) &= f(x), & \rho_t(x, 0) = 0, \\ \rho(y, t) &= g(y, t) & \text{para } y \in S, t \geq 0. \end{aligned} \tag{1}$$

Donde $c(x)$ es la velocidad de propagación del sonido en el tejido, $g(y,t)$ son los valores medidos por los sensores p_t, p_{tt} , denota la primera y segunda derivada de tiempo y Δx es el operador laplaciano con respecto a la variable espacial x . El problema principal puede formularse ahora como sigue: dada la velocidad del sonido $c(x)$ y los datos medidos $g(y, t)$ en (1), buscar el valor inicial $f(x)$ de la presión $p(x, t)$. Este valor inicial es la imagen.

Esta técnica fue sugerida para dimensiones impares y velocidad del sonido constante. Básicamente se basa en dos premisas, si la velocidad del sonido es constante y la dimensión espacial es impar se cumple el principio de Huygens el cual establece que para cualquier fuente inicial con un soporte acotado, la onda abandona cualquier dominio acotado en un tiempo finito, es decir, bajo esas condiciones existe un momento T (que es el tiempo más largo que tarda la onda en atravesar el dominio B) en el cual la solución $p(x,t)$ de (1) desaparece dentro de B para cualquier $t>T$ entonces se pueden imponer condiciones iniciales cero en $t=T$ y condiciones de frontera iguales a los datos medidos f y así resolver (1) en la dirección temporal inversa, llegando así en $t=0$ a la función $f(x)$.

Condiciones necesarias para utilizar time reversal: El método de time reversal funciona para geometría arbitraria de la superficie de observación cerrada S . La velocidad del sonido puede ser variable. La fuente f no tiene que estar apoyada dentro de S . De hecho, la única condición importante es que la solución dentro de S decaiga con el tiempo. Esto sucede incluso para velocidades de captura "malas", aunque el decaimiento puede ser lento, y por lo tanto puede ser necesario un mayor T . El método puede implementarse fácilmente para medios no homogéneos y superficies de observación cerradas arbitrarias, utilizando esquemas numéricos sencillos para la ecuación de onda.[21]

CAPÍTULO 3

APLICACIÓN DE LA RECONSTRUCCIÓN PAT CON RED NEURONAL

La aplicación de redes neuronales en PAI se ha extendido en los últimos años, por ejemplo, en la reconstrucción de imagen, en la segmentación de imagen, entre otras. En la reconstrucción de imagen existen diferentes métodos que se mencionaron en el capítulo 2; en la disminución de ruido se han aplicado redes neuronales en los sinogramas. En este capítulo se habla sobre la aplicación como método de reconstrucción en sinogramas post-procesados con redes neuronales.

En el capítulo 2 se presentaron las redes mayormente utilizadas, en este trabajo se implementaron 4 redes neuronales las cuales son:

- *Super-Resolution Convolutional neural network (SRCNN)* tiempo de entrenamiento 4 horas.
- *U-net Convolutional Neural Network model Exponential linear unit (U-net ELU)* tiempo de entrenamiento 12 horas.
- *U-net Convolutional Neural Network model Rectified Linear Unit (U-net RELU)* tiempo de entrenamiento 12 horas.
- *U-net Híbrida combinación entre U-net RELU y U-net ELU* tiempo de entrenamiento 12 horas.

Todos los cálculos se realizaron en una PC con Linux (Ubuntu) con una CPU AMD Radeon 4800H, Con una velocidad de reloj de 2,90 GHz, con 24 GB de RAM y una GPU NVIDIA 1650ti con 4 GB de memoria.

3.1 Entrenamiento

Para entrenar las redes y mejorar el ancho de banda (BW por sus siglas en inglés) de los datos del dominio del sinograma es necesario generar los datos de entrenamiento. Los datos de entrenamiento se generaron con la herramienta k-Wave de MATLAB. Los sinogramas obtenidos para el entrenamiento corresponden a 1000 imágenes de mama, las cuales son cortes de tomografía computarizada por ultrasonido. El sinograma se generó para todas estas imágenes con una configuración circular de 200 sensores con una presión inicial de 1kPa y una velocidad de propagación del sonido en el medio de 1500m/s. De cada imagen se genera una serie de parches, en total 105 parches con un tamaño de 64x64, que se desplaza a lo largo del sinograma el cual tiene un tamaño de 200x512 con un paso de 16x16, que daría un total de $105 \times 1000 = 105\,000$ parches, del total se divide el 20% para los datos de validación, por lo que el conjunto de datos para entrenamiento consta de 84,000 parches generados con 200 sensores y sin ruido presente , y el conjunto de validación es de 21,000 parches. Para generar los sinogramas de validación se colocan 100 sensores con un ancho de banda limitado al dominio de la imagen, posteriormente se agrega ruido gaussiano en un rango de 10-70dB. Estos datos son los utilizados como entrada para las redes. La salida consiste de sinogramas totalmente muestreados, con un ancho de banda completo que corresponderían a un sinograma con 200 sensores y sin ruido. El número total de parámetros entrenables utilizados para el entrenamiento fue de 7 759 787. Las redes se entrenaron utilizando la función de pérdida del error medio cuadrático escalado (scaled-RMSE por sus siglas en inglés) la cual se define como [19]:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \|\hat{x}_i - \phi(x_i)\|^2 \times \tau$$

Donde \hat{x}_i son las predicciones del modelo, $\phi(x_i)$ son las señales reales, dado que los datos del sinograma contienen valores del orden de 0,0001, se utilizó un factor de multiplicación de pérdidas, $\tau = 10\,000$ para garantizar una retroproyección eficaz de la

pérdida en las capas iniciales.

En la figura 7 se muestra la tomografía computarizada de mama, la cual tiene unas dimensiones de 495x615x752, de esta tomografía se generaron los cortes usados para generar los sinogramas de entrenamiento.

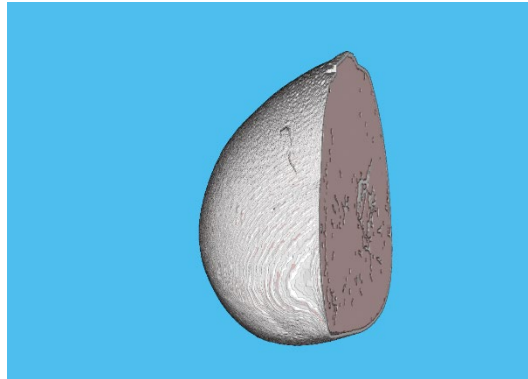


Figura 7 Tomografía computarizada de mama, utilizada para generar cortes axiales, los cuales fueron utilizados para generar los sinogramas de tomografía fotoacústica.

En la figura 8 se muestra un corte de mama de la TC de la figura 7, utilizada para generar los sinogramas de entrada para entrenar las redes.

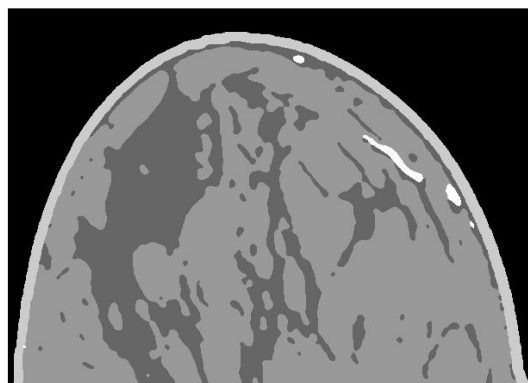


Figura 8 corte de imagen de mama a través de tomografía computarizada.

En la figura 9 se muestra la configuración utilizada en k-Wave para generar los sinogramas, consta de 100 sensores colocados en círculo alrededor del objeto (corte de Mama), a una distancia de 22.5mm del centro, el tamaño del mallado computacional es de 50.1mm x 50.1mm y el mallado utilizado para la generación y reconstrucción de la

imagen es de 20.1mmx20.1mm, la configuración utilizada para generar las imágenes de referencia tiene las mismas dimensiones, la diferencia radica en que consta de 200 sensores.

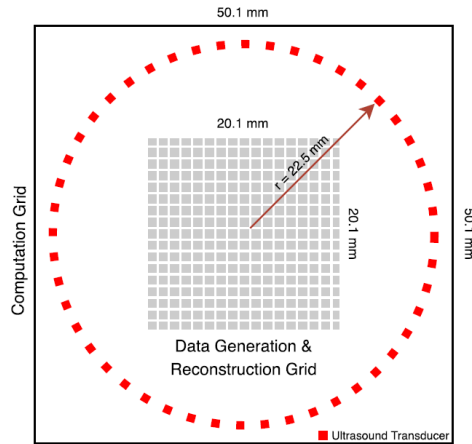
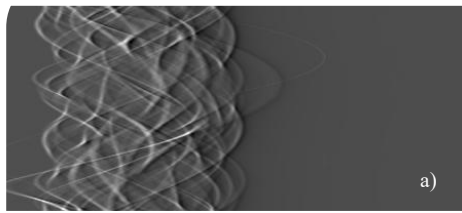


Figura 9 Representación de la configuración circular utilizada para generar los sinogramas, el mallado computacional es de 501x501 con un arreglo de 200 sensores para los datos de entrenamiento y de 100 sensores para los datos de validación, el mallado de reconstrucción es de 201x201 pixeles, los sensores se encuentran a una distancia de 22.5mm del centro [22].

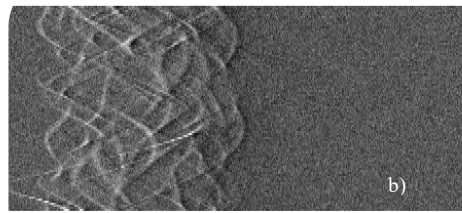
Una vez generados los datos de entrenamiento es necesario generar los datos con los cuales se harán las pruebas, estos datos son de baja resolución con 100 detectores y un ancho de banda limitado, es necesario realizar una interpolación, en este caso se utilizó la interpolación de vecino más cercano para que el sinograma resultante tenga un tamaño de 200x512, posteriormente al ingresar a las redes se agrega un relleno de tipo “reflect” para tener unas dimensiones de 512x512. Esto para que no exista un desajuste entre el tamaño de los tensores. Una vez que se tiene el sinograma procesado por las redes es necesario realizar la reconstrucción de dicho sinograma.

En la figura 10 se muestra un ejemplo de la comparación de sinogramas generados a través de las diferentes redes, en a) se muestra el sinograma de referencia, al cual no se le agrega ruido y se realizó con 200 sensores en la simulación, en b) se muestra el sinograma con un nivel de ruido de 20dB, ocupando 100 sensores, éste sinograma fue el ingresado a las distintas redes, en c) se muestra el resultado obtenido con la red SRCNN, en d) se muestra el resultado obtenido con la red U-net RELU, en e) se muestra el

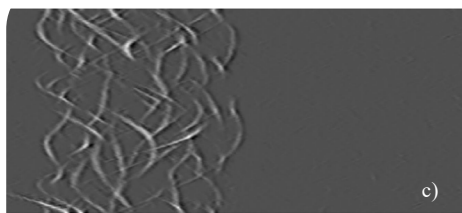
resultado obtenido con la red U-net ELU y f) muestra el resultado obtenido con la red U-net hibrida.



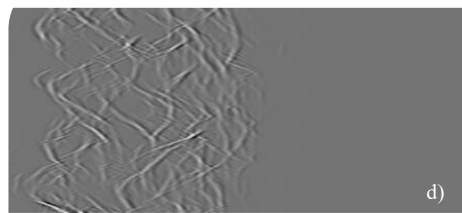
Sinograma de referencia
Sin ruido
200 Sensores



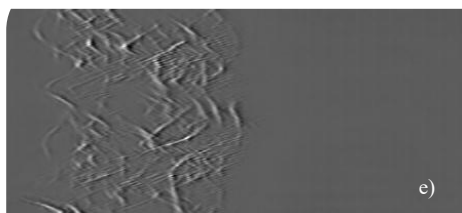
Sinograma ingresado a las redes
Ruido de 20dB
100 Sensores



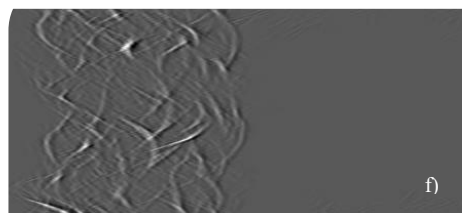
Sinograma obtenido con la red
SRCNN



Sinograma obtenido con la red
RELU



Sinograma obtenido con la red
ELU



Sinograma obtenido con la red
Hibrida

Figura 10. Resultado de un sinograma al ser procesado por las distintas redes.

CAPÍTULO 4

RESULTADOS

Una vez que las redes se encuentran entrenadas se procede a realizar el proceso de validación, para ello se seleccionaron 4 imágenes comúnmente utilizadas como referencias para pruebas de calidad de imágenes fotoacústicas, inciso (a) Siglas de PAT,(b) imagen Derenzo,(c)Imagen de vasos sanguíneos,(d) corte de tomografía de mama, del inciso (e)-(h) se muestran las reconstrucciones realizadas con el método de retroproyección lineal, del inciso (i)-(l) se muestran las reconstrucciones realizadas con el método de Time Reversal, ambas utilizando una simulación de tomografía fotoacústica sin presencia de ruido, con 200 sensores, siguiendo las especificaciones mostradas en la Figura 9.

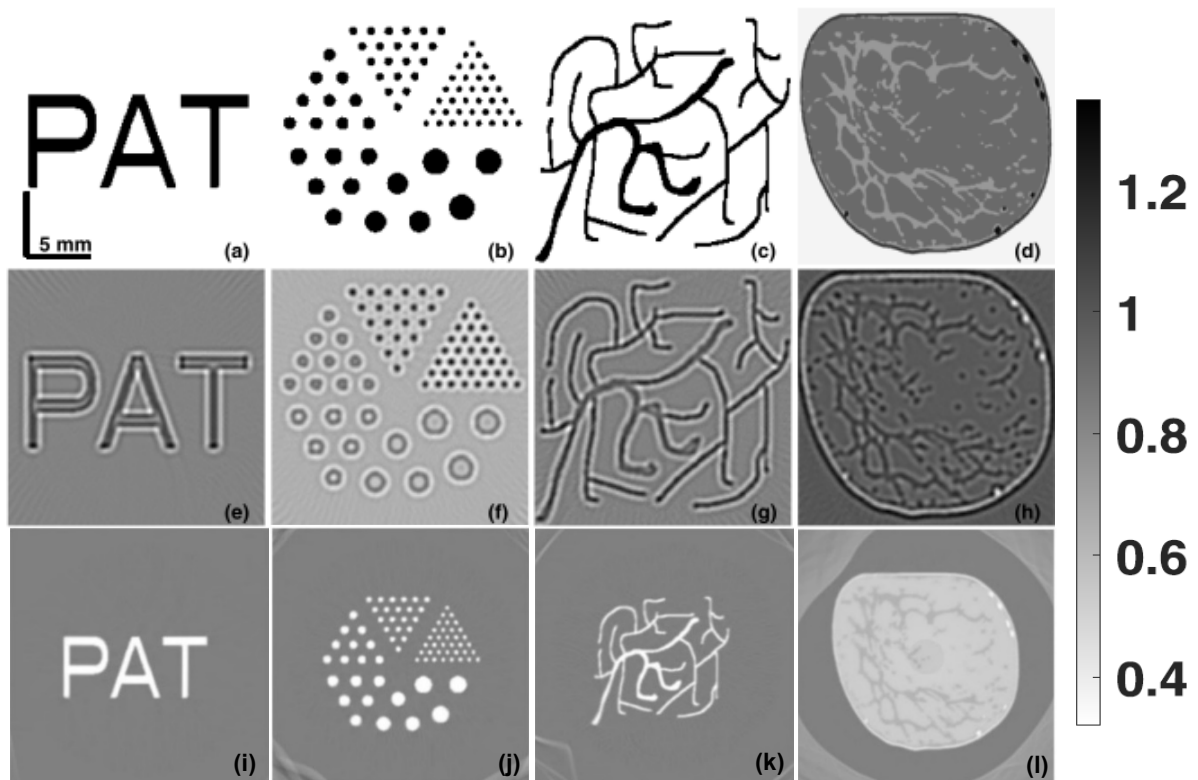


Figura 11 imágenes de referencia para evaluación cuantitativa: PAT, Derenzo, vasos sanguíneos y mama (a)-(d), correspondientes reconstrucciones utilizando (LBP), con un ancho de banda completo y sin ruido (e)-(h) [19], reconstrucción utilizando Time Reversal con un ancho de banda completo y sin ruido (i)-(l).

Para realizar la comparación de manera cuantitativa de los resultados obtenidos, es necesario utilizar métricas de evaluación [19]. Las métricas de evaluación que se utilizan son: Error de raíz cuadrada media (RMSE por sus siglas) en inglés, índice de similitud estructural (SSIM por sus siglas en inglés) y correlación de Pearson (PC por sus siglas en inglés).

4.1 Error de raíz cuadrada media (RMSE)

Este método evalúa la desviación estándar de los residuos como una medida absoluta para evaluar la calidad de la imagen reconstruida, se define como: [4-23]

$$RMSE(\tau^{objetivo}, \tau^{reconstrucción}) = \sqrt{\frac{\sum_{i,j}^N (\tau_{i,j}^{objetivo} - \tau_{i,j}^{reconstrucción})^2}{N^2}}$$

Donde $\tau^{objetivo}$ y $\tau^{reconstrucción}$ denota la imagen objetivo y la imagen reconstruida, donde cada imagen tiene un tamaño de pixel de $N \times N$. Los residuos más bajos indican una imagen de buena calidad de las imágenes reconstruidas, por lo tanto, es factible que los valores RMSE sean bajos cercanos a 0.

4.2 índice de similitud estructural (SSIM)

El índice de similitud estructural comprende 3 medidas de comparación las cuales son: luminancia (I), contraste (C) y estructura/correlación (S), para evaluar la calidad de una imagen respecto a otra de referencia [24-25]. La representación de los parámetros es:

$$I(objetivo, reconstrucción) = \frac{2\mu_{objetivo}\mu_{reconstrucción} + c_1}{\mu_{objetivo}^2 + \mu_{reconstrucción}^2 + c_1}$$

$$C(objetivo, reconstrucción) = \frac{2\sigma_{objetivo}\sigma_{reconstrucción} + c_2}{\sigma_{objetivo}^2 + \sigma_{reconstrucción}^2 + c_2}$$

$$S(objetivo, reconstrucción) = \frac{\sigma_{objetivo, reconstrucción} + c_3}{\sigma_{objetivo}\sigma_{reconstrucción} + c_3}$$

Donde $\mu_{objetivo}$ y $\mu_{reconstrucción}$ representa el promedio, $\sigma_{objetivo}$ y $\sigma_{reconstrucción}$

representan la varianza de la imagen objetivo y la imagen reconstruida, $\sigma_{objetivo,reconstrucción}$ representa la covarianza entre las dos imágenes, y C1, C2 y C3 son constantes, donde $c3 = \frac{c2}{2}$, el índice de similitud estructural se define como una combinación pondera de sus tres componentes $SSIM = [l^\alpha \cdot C^\beta \cdot S^\gamma]$, en el caso de SSIM en cuanto mayor sea el valor numérico, indica una mayor similitud entre la imagen reconstruida y la imagen objetivo.

4.3 Correlación de Pearson (PC)

La correlación de Pearson es una medida de correlación lineal entre dos objetos, la imagen objetivo y la imagen reconstruida y está definida como:[26]

$$PC(\tau^{objetivo}, \tau^{reconstrucción}) = \frac{cov(\tau^{objetivo}, \tau^{reconstrucción})}{\sigma(\tau^{objetivo}) \cdot \sigma(\tau^{reconstrucción})}$$

Donde cov representa la covarianza y σ representa la varianza. La correlación de Pearson tiene un rango que va de $[-1, 1]$, donde 1 denota una correlación lineal completa, 0 representa que no existe una correlación lineal y -1 indica una relación complementaria perfecta, por lo que lo conveniente sería que el índice de correlación de Pearson sea lo más cercano a 1.

4.4 Análisis de resultados

En esta sección se presenta el análisis de los resultados, comenzado por comparar imágenes procesadas con métodos utilizados para mejorar la calidad de una imagen y las procesadas con las distintas redes, posteriormente se presentan tablas comparando la calidad de la imagen medida a través de 3 métricas de evaluación: Error de raíz cuadrada media (RMSE), índice de similitud estructural (SSIM) y correlación de Pearson (PC), al final se presenta una tabla donde se muestran los tiempos de reconstrucción de las diferentes imágenes utilizando time reversal como método de reconstrucción.

4.4.1 Análisis de resultados para imagen palabra PAT

En la figura 12 se muestra la Imagen PAT a la cual se le realizaron diversos procesos con el fin de tener una comparativa entre el uso de redes neuronales, todas las imágenes se simularon utilizando una configuración circular de 100 sensores con un mallado computacional de 501x501 pixeles, un mallado de reconstrucción de 201x201 pixeles, los sensores están posicionados a una distancia de 22.5mm del centro, cada fila representa la imagen con distintos niveles de ruido 20dB,40dB y 60dB. En (a),(h) y (o) se muestran las imágenes sin ningún tipo de proceso con un ancho de banda limitado a 2.25 MHz, en (b),(i) y (p) se muestran las imágenes con una interpolación de vecino más cercano para mejorar la imagen, en (c),(j) y (q) se muestran las imágenes con un proceso automatizado de eliminación de ruido por ondículas¹, en (d),(k) y (r) se muestran las imágenes utilizando SRCNN, en (e),(l) y (s) se muestran las imágenes U-net RELU, en (f),(m) y (t) se muestran las imágenes utilizando U-net ELU, en (g),(n) y (u) se muestran las imágenes utilizando U-net Hibrida. Estas imágenes fueron reconstruidas utilizando la reconstrucción por retroproyección lineal [19]

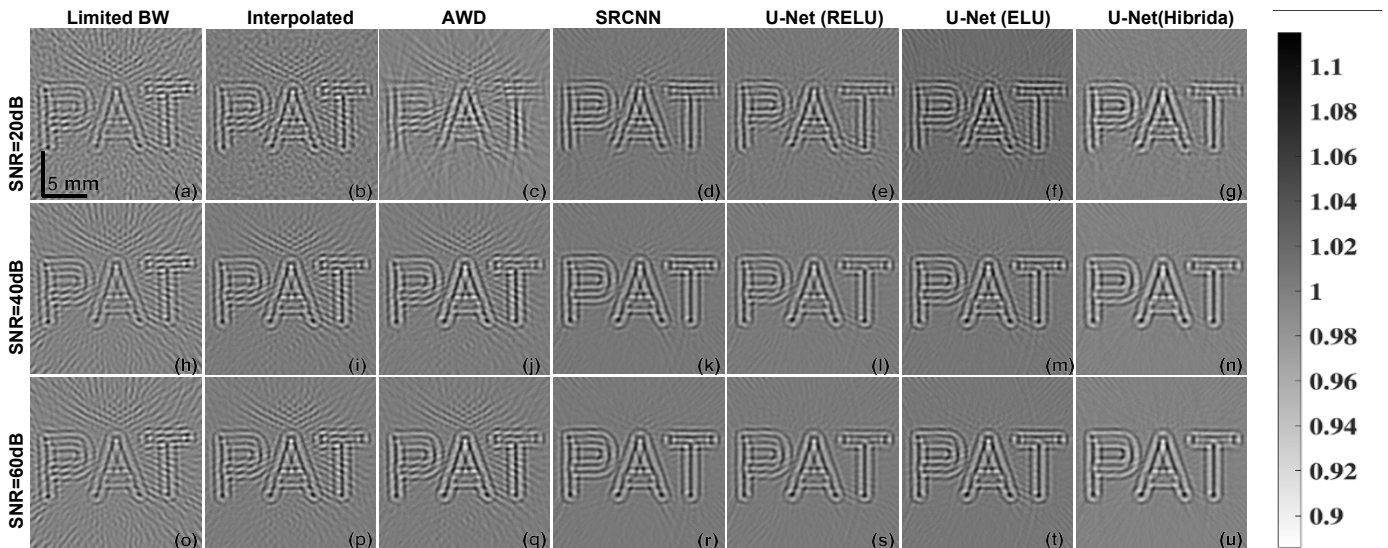


Figura 12. Imágenes reconstruidas por retroproyección lineal utilizando distintos métodos para mejorar la imagen, ancho de banda limitado (a), (h) y (o). Interpolación con vecino más cercano (b), (i) y (p). Utilizando el enfoque automatizado de eliminación de ruido por ondículas (c), (j), y (q). Resultados utilizando la red SRCNN (d), (k) y (r). Resultados utilizando la red U-net RELU (e), (l) y (s). Resultados utilizando la red U-net ELU (f), (m) y (t). Resultados utilizando la red U-net Hibrida (g), (n) y (u) [19].

En la figura 13 se muestra el conjunto de imágenes utilizando las distintas redes y reconstruyéndolas mediante el método de time reversal. En (1), (5) y (9) se muestran las imágenes PAT utilizando la red SRCNN. En (2), (6) y (10) se muestran las imágenes utilizando la red U-net RELU. En (3), (7) y (11) se muestran las imágenes utilizando la red U-net ELU. En (4), (8) y (12) se muestran las imágenes utilizando la red U-net hibrida.

En este proceso es utilizada la transformada ondícula (wavelet) la cual se basa en eliminar los cambios abruptos que puedan aparecer en la señal, es mejor que usar un filtro pasa-bajas ya que éste es incapaz de eliminar ruido que está en la misma banda de frecuencia de la señal.

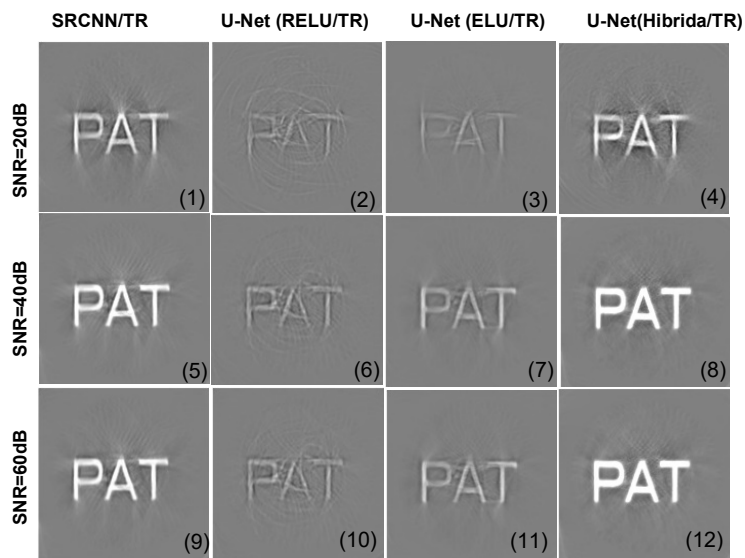


Figura 13 Imágenes reconstruidas mediante time reversal utilizando las redes SRCNN (1),(5),(9), U-net RELU (2),(6),(10), U-net ELU (3),(7),(11) y U-net Hibrida (4),(8),(12).

En la tabla 3 se muestra el desglose de los resultados aplicando las métricas de comparación de Correlación de Pearson (PC), el índice de similitud estructural (SSIM) y el error de raíz cuadrada media (RMSE) a las imágenes de la Figura 12 y 13, los valores en negrita expresan el mejor resultado posible para cada métrica. Se puede observar que en el apartado de imagen PAT con nivel de ruido de 20dB en la métrica de PC el mejor valor lo presenta la red U-net Hibrida utilizando Time Reversal como método de reconstrucción dando un valor de **0.9892** con una mejora del **25.86%** respecto al mejor valor presentado con retroproyección lineal **0.7306**. En el apartado SSIM el mejor valor

lo presenta la red U-net(ELU) con reconstrucción TR con un valor de **0.9655** dando una mejora de **26.21%** respecto al mejor valor presentado con LBP **0.7034**. En RMSE el mejor valor lo presenta la red SRCNN con **0.0373** con una mejora del **2.44%** respecto al mejor valor presentado con LBP **0.0617**. En el apartado con un nivel de ruido de 40dB en la métrica PC el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.9962** con una mejora de **22.41%** respecto al mejor valor presentado con LBP **0.7721**. En SSIM el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.9725** con una mejora de **20.16%** respecto al mejor valor presentado con LBP **0.7709**. En RMSE el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.0215** con una mejora de **3.24%** respecto al mejor valor presentado con LBP **0.0539**. En el apartado con un nivel de ruido de 60dB en la métrica PC el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.9962** con una mejora de **22.35%** respecto al mejor valor presentado con LBP **0.7727**. En SSIM el mejor valor lo presenta la red U-net (ELU) con reconstrucción TR con un valor de **0.9731** con una mejora de **20.27%** respecto al mejor valor presentado con LBP **0.7704**. En RMSE el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.0214** con una mejora de **3.25%** respecto al mejor valor presentado con LBP **0.0539**.

IMAGEN PAT									
SNR of data	20dB			40dB			60dB		
Method/Metric	PC	SSIM	RMSE	PC	SSIM	RMSE	PC	SSIM	RMSE
Limited BW	0.3077	0.2728	0.1078	0.3293	0.3262	0.1012	0.3293	0.3262	0.1012
Interpolated	0.5317	0.4386	0.0871	0.5738	0.5417	0.0822	0.5746	0.5429	0.0805
Automated Wavelet Denoising	0.5387	0.5376	0.0782	0.6367	0.5761	0.0765	0.6367	0.5761	0.0765
SRCNN	0.6712	0.6474	0.0806	0.7375	0.7096	0.0711	0.7386	0.7100	0.0685
U-Net (RELU)	0.6668	0.6416	0.0756	0.7396	0.7289	0.0637	0.7421	0.7291	0.0642
U-Net (ELU)	0.6676	0.6501	0.1083	0.7388	0.7201	0.0754	0.7419	0.7228	0.0721
U-Net (Hibrida)	0.7306	0.7034	0.0617	0.7721	0.7709	0.0539	0.7727	0.7704	0.0539
SRCNN (Time Reversal)	0.9890	0.9604	0.0373	0.9962	0.9712	0.0215	0.9962	0.9715	0.0214
U-Net	0.9240	0.8951	0.0956	0.9691	0.9509	0.0620	0.9699	0.9519	0.0609

(RELU/Time Reversal)									
U-Net (ELU/Time Reversal)	0.9671	0.9655	0.0640	0.9800	0.9725	0.0502	0.9815	0.9731	0.0482
U-Net(Hibrida/Time Reversal)	0.9892	0.9475	0.0374	0.9948	0.9636	0.0253	0.9943	0.9668	0.0264

Tabla 3 Resultados comparativos obtenidos para las imágenes reconstruidas de PAT que se muestran en la figura 12 y 13. En términos de figuras de mérito: coeficiente pc, índice SSIM y RMSE para SNR de datos de 20, 40 y 60 dB. El texto en negrita indica el mejor valor posible para la métrica utilizada.

4.4.2 Análisis de resultados para imagen de Derenzo

En la Figura 14 se muestra la Imagen de Derenzo a la cual se le realizaron diversos procesos con el fin de tener una comparativa entre el uso de redes neuronales. Todas las imágenes se simularon utilizando una configuración circular de 100 sensores con un mallado computacional de 501x501 pixeles, un mallado de reconstrucción de 201x201 pixeles, los sensores están posicionados a una distancia de 22.5mm del centro, cada fila representa la imagen con distintos niveles de ruido 20dB,40dB y 60dB. En (a),(h) y (o) se muestra la imagen sin ningún tipo de proceso con un ancho de banda limitado a 2.25 MHz. En (b), (i) y (p) se muestran las imágenes con una interpolación de vecino más cercano para mejorar la imagen. En (c), (j) y (q) se muestran las imágenes con un proceso automatizado de eliminación de ruido por ondículas (wavelets), en (d), (k) y (r) se muestran las imágenes utilizando SRCNN, en (e), (l) y (s) se muestran las imágenes U-net RELU, en (f), (m) y (t) se muestran las imágenes utilizando U-net ELU, en (g), (n) y (u) se muestran las imágenes utilizando U-net Hibrida. Estas imágenes reconstruidas utilizando la reconstrucción por retroproyección lineal [19]

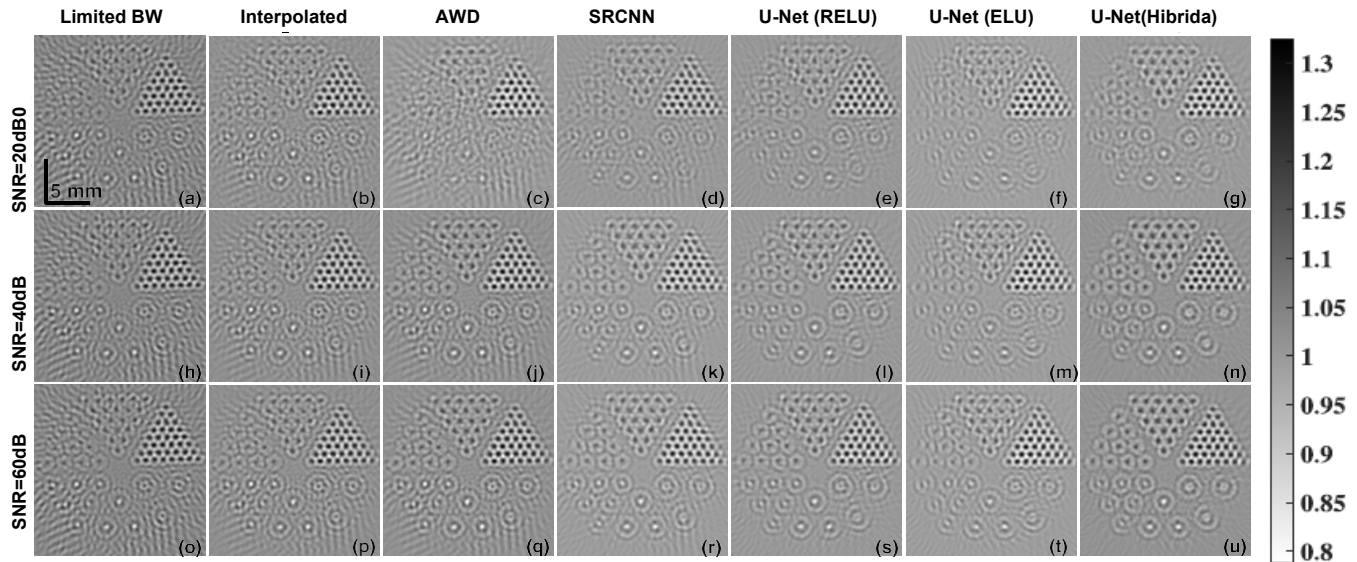


Figura 14 Imágenes reconstruidas por retroproyección lineal utilizando distintos métodos para mejorar la imagen, ancho de banda limitado (a), (h) y (o). Interpolación con vecino más cercano (b), (i) y (p). Utilizando el enfoque automatizado de eliminación de ruido por ondículas (wavelets) (c), (j), y (q). Resultados utilizando la red SRCNN (d), (k) y (r). Resultados utilizando la red U-net RELU (e), (l) y (s). Resultados utilizando la red U-net ELU (f), (m) y (t). Resultados utilizando la red U-net Híbrida (g), (n) y (u) [19].

En la figura 15 se muestra el conjunto de imágenes utilizando las distintas redes y reconstruyéndolas mediante el método de time reversal. En (1), (5) y (9) se muestran las imágenes PAT utilizando la red SRCNN. En (2), (6) y (10) se muestran las imágenes utilizando la red U-net RELU. En (3), (7) y (11) se muestran las imágenes utilizando la red U-net ELU. En (4), (8) y (12) se muestran las imágenes utilizando la red U-net híbrida.

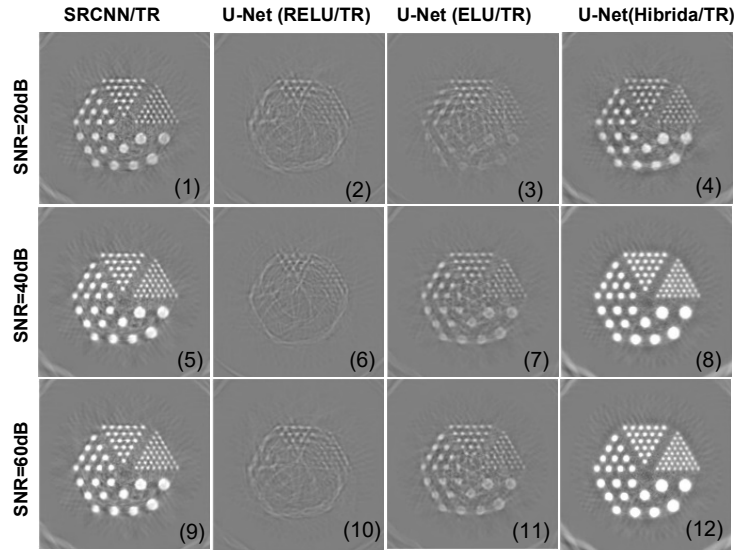


Figura 15 Imágenes reconstruidas mediante time reversal utilizando las redes SRCNN (1),(5),(9), U-net RELU (2),(6),(10), U-net ELU (3),(7),(11) y U-net Hibrida (4),(8),(12).

En la tabla 4 se muestra el desglose de los resultados aplicando las métricas de comparación de Correlación de Pearson (PC), el índice de similitud estructural (SSIM) y el error de raíz cuadrada media (RMSE) a las imágenes de la figura 14 y 15, los valores en negrita expresan el mejor resultado posible para cada métrica. Se puede observar que en el apartado de imagen de Derenzo con nivel de ruido de 20dB en la métrica de PC el mejor valor lo presenta la red U-net Hibrida utilizando Time Reversal como método de reconstrucción dando un valor de **0.9911** con una mejora del **18.91%** respecto al mejor valor presentado con retroproyección lineal **0.8020**. En el apartado SSIM el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.9373** dando una mejora de **23.58%** respecto al mejor valor presentado con LBP **0.7015**. En RMSE el mejor valor lo presenta la red U-net hibrida con **0.0339** con una mejora del **5.75%** respecto al mejor valor presentado con LBP **0.0914**. En el apartado con un nivel de ruido de 40dB en la métrica PC el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.9949** con una mejora de **16.44%** respecto al mejor valor presentado con LBP **0.8305**. En SSIM el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.9440** con una mejora de **22.45%** respecto al mejor valor presentado con LBP **0.7195**. En RMSE el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.0246** con una mejora de **6.98%** respecto al mejor valor presentado con LBP

0.0944. En el apartado con un nivel de ruido de 60dB en la métrica PC el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.9947** con una mejora de **16.42%** respecto al mejor valor presentado con LBP **0.8305**. En SSIM el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.9440** con una mejora de **22.45%** respecto al mejor valor presentado con LBP **0.7199**. En RMSE el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.0252** con una mejora de **6.88%** respecto al mejor valor presentado con LBP **0.0940**.

IMAGEN DERENZO									
SNR of data	20dB			40dB			60dB		
Method/Metric	PC	SSIM	RMSE	PC	SSIM	RMSE	PC	SSIM	RMSE
Limited BW	0.3525	0.3199	0.1953	0.3616	0.3263	0.1949	0.3616	0.3263	0.1949
Interpolated	0.6544	0.5425	0.1456	0.6661	0.5610	0.1498	0.6662	0.5612	0.1502
Automated Wavelet Denoising	0.6185	0.5110	0.1177	0.7038	0.5851	0.1426	0.7038	0.5851	0.1426
SRCNN	0.7578	0.6477	0.1103	0.7907	0.6690	0.0944	0.7912	0.6697	0.0940
U-Net (RELU)	0.7714	0.6707	0.1112	0.8015	0.6900	0.1063	0.8015	0.6905	0.1062
U-net (ELU)	0.7645	0.6728	0.0914	0.7842	0.6833	0.0952	0.7714	0.6707	0.1112
U-net(Hibrida)	0.8020	0.7015	0.1031	0.8305	0.7195	0.1240	0.8305	0.7199	0.1242
SRCNN (Time Reversal)	0.9903	0.9373	0.0347	0.9949	0.9440	0.0246	0.9947	0.9440	0.0252
U-Net (RELU/Time Reversal)	0.9598	0.9137	0.0708	0.9632	0.9183	0.0676	0.9635	0.9189	0.0674
U-Net (ELU/Time Reversal)	0.9677	0.9265	0.0638	0.9795	0.9390	0.0508	0.9799	0.9395	0.0504
U-Net(Hibrida/Time Reversal)	0.9911	0.9369	0.0339	0.9910	0.9411	0.0331	0.9903	0.9409	0.0347

Tabla 4 Resultados comparativos obtenidos para las imágenes reconstruidas de Derenzo que se muestran en la fig.14 y 15 en términos de figuras de mérito: coeficiente pc, índice SSIM y RMSE para SNR de datos de 20, 40 y 60 dB. El texto en negrita indica el mejor valor posible para la métrica utilizada.

4.4.3 Análisis de resultados para imagen de Vasos Sanguíneos

En la figura 16 se muestra la Imagen de vasos sanguíneos a la cual se le realizaron diversos procesos con el fin de tener una comparativa entre el uso de redes neuronales, todas las imágenes se simularon utilizando una configuración circular de 100 sensores con un mallado computacional de 501x501 pixeles, un mallado de reconstrucción de 201x201 pixeles. Los sensores están posicionados a una distancia de 22.5mm del centro, cada fila representa la imagen con distintos niveles de ruido 20dB,40dB y 60dB. En (a), (h) y (o) se muestran las imágenes sin ningún tipo de proceso con un ancho de banda limitado a 2.25 MHz. En (b),(i) y (p) se muestran las imágenes con una interpolación de vecino más cercano para mejorar la imagen. En (c), (j) y (q) se muestran las imágenes con un proceso automatizado de eliminación de ruido por ondículas (wavelets). En (d), (k) y (r) se muestran las imágenes utilizando SRCNN. En (e), (l) y (s) se muestran las imágenes U-net RELU. En (f), (m) y (t) se muestran las imágenes utilizando U-net ELU. En (g), (n) y (u) se muestran las imágenes utilizando U-net Híbrida. Estas imágenes fueron reconstruidas utilizando la reconstrucción por retroproyección lineal [19]

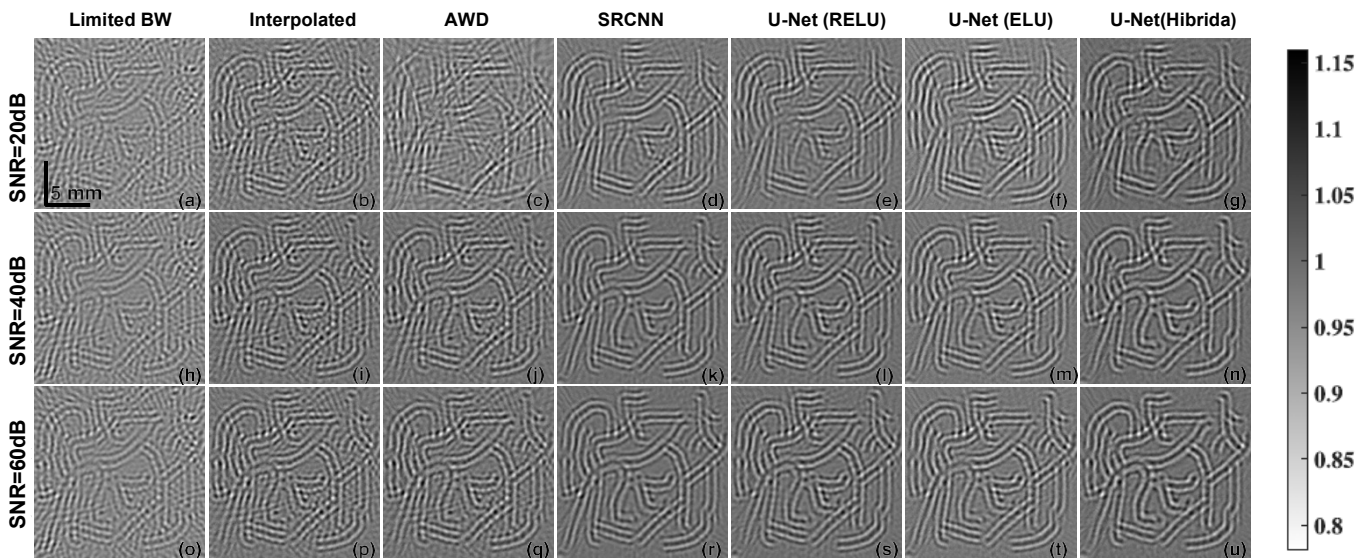


Figura 16 Imágenes reconstruidas por retroproyección lineal utilizando distintos métodos para mejorar la imagen, ancho de banda limitado (a), (h) y (o). Interpolación con vecino más cercano (b), (i) y (p). Utilizando el enfoque automatizado de eliminación de ruido por ondículas (wavelets) (c), (j), y (q). Resultados utilizando la red SRCNN (d), (k) y (r). Resultados utilizando la red U-net RELU (e), (l) y (s). Resultados utilizando la red U-net ELU (f), (m) y (t). Resultados utilizando la red U-net Híbrida (g), (n) y (u) [19].

En la figura 17 se muestra el conjunto de imágenes utilizando las distintas redes y reconstruyéndolas mediante el método de time reversal. En (1), (5) y (9) se muestran las imágenes PAT utilizando la red SRCNN. En (2), (6) y (10) se muestran las imágenes utilizando la red U-net RELU. En (3), (7) y (11) se muestran las imágenes utilizando la red U-net ELU. En (4), (8) y (12) se muestran las imágenes utilizando la red U-net hibrida.

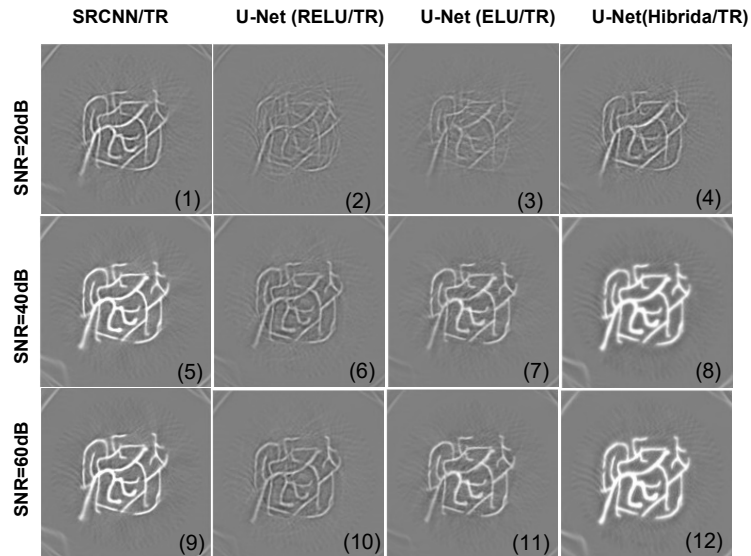


Figura 17 Imágenes reconstruidas mediante time reversal utilizando las redes SRCNN (1),(5),(9), U-net RELU (2),(6),(10), U-net ELU (3),(7),(11) y U-net Hibrida (4),(8),(12).

En la tabla 5 se muestra el desglose de los resultados aplicando las métricas de comparación de Correlación de Pearson (PC), el índice de similitud estructural (SSIM) y el error de raíz cuadrada media (RMSE) a las imágenes de la figura 16 y 17. Los valores en negrita expresan el mejor resultado posible para cada métrica, se puede observar que en el apartado de imagen de vasos sanguíneos con nivel de ruido de 20dB en la métrica de PC el mejor valor lo presenta la red SRCNN utilizando Time Reversal como método de reconstrucción dando un valor de **0.9902** con una mejora del **25.74%** respecto al mejor valor presentado con retroproyección lineal **0.7328**. En el apartado SSIM el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.9405** dando una mejora de **34.25%** respecto al mejor valor presentado con LBP **0.5980**. En RMSE el mejor valor lo presenta la red SRCNN con **0.0359** con una mejora del **6.27%** respecto al mejor valor presentado con LBP **0.0986**. En el apartado con un nivel de ruido de 40dB en la métrica PC el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor

de **0.9949** con una mejora de **20.38%** respecto al mejor valor presentado con LBP **0.7911**. En SSIM el mejor valor lo presenta la red U-net (ELU) con reconstrucción TR con un valor de **0.9579** con una mejora de **29.72%** respecto al mejor valor presentado con LBP **0.6607**. En RMSE el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.0246** con una mejora de **6.97%** respecto al mejor valor presentado con LBP **0.0943**, En el apartado con un nivel de ruido de 60dB en la métrica PC el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.9945** con una mejora de **20.23%** respecto al mejor valor presentado con LBP **0.7922**. En SSIM el mejor valor lo presenta la red U-net (ELU) con reconstrucción TR con un valor de **0.9587** con una mejora de **29.63%** respecto al mejor valor presentado con LBP **0.6624**. En RMSE el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.0256** con una mejora de **6.78%** respecto al mejor valor presentado con LBP **0.0934**.

IMAGEN VASOS									
SNR of data	20dB			40dB			60dB		
Method/Metric	PC	SSIM	RMSE	PC	SSIM	RMSE	PC	SSIM	RMSE
Limited BW	0.3124	0.2930	0.1429	0.3326	0.3136	0.1375	0.3326	0.3136	0.1375
Interpolated	0.5879	0.4591	0.1155	0.6224	0.4943	0.1200	0.6233	0.4958	0.1161
Automated Wavelet Denoising	0.4617	0.3721	0.1264	0.6540	0.5208	0.1153	0.6540	0.5208	0.1153
SRCNN	0.7094	0.5769	0.1016	0.7870	0.6607	0.1022	0.7886	0.6624	0.1031
U-Net (RELU)	0.7043	0.5767	0.1016	0.7537	0.6191	0.1084	0.7543	0.6208	0.1072
U-net (ELU)	0.7154	0.5756	0.0986	0.7621	0.6369	0.0943	0.7628	0.6374	0.0934
U-net(Hibrida)	0.7328	0.5980	0.1195	0.7911	0.6568	0.1084	0.7922	0.6581	0.1056
SRCNN (Time Reversal)	0.9902	0.9405	0.0359	0.9949	0.9526	0.0246	0.9945	0.9522	0.0256
U-Net (RELU/Time Reversal)	0.9753	0.9202	0.0559	0.9397	0.8805	0.0850	0.9846	0.9476	0.0446
U-Net (ELU/Time Reversal)	0.9780	0.9303	0.0530	0.9934	0.9579	0.0293	0.9940	0.9587	0.0280
U-Net(Hibrida/Time Reversal)	0.9843	0.9337	0.0457	0.9886	0.9412	0.0373	0.9868	0.9404	0.0405

Tabla 5 Resultados comparativos obtenidos para las imágenes reconstruidas de vasos sanguíneos que se muestran en la fig.16 y 17 en términos de figuras de mérito: coeficiente pc, índice SSIM y RMSE para SNR de datos de 20, 40 y 60 dB. El texto en negrita indica el mejor valor posible para la métrica utilizada.

4.4.4 Análisis de resultados para imagen de Mama

En la figura 18 se muestra la Imagen de mama a la cual se le realizaron diversos procesos con el fin de tener una comparativa entre el uso de redes neuronales. Todas las imágenes se simularon utilizando una configuración circular de 100 sensores con un mallado computacional de 501x501 pixeles, y un mallado de reconstrucción de 201x201 pixeles. Los sensores están posicionados a una distancia de 22.5mm del centro, cada fila representa la imagen con distintos niveles de ruido 20dB, 40dB y 60dB. En (a), (h) y (o) se muestra la imagen sin ningún tipo de proceso con un ancho de banda limitado a 2.25 MHz. En (b), (i) y (p) se muestran las imágenes con una interpolación de vecino más cercano para mejorar la imagen. En (c), (j) y (q) se muestran las imágenes con un proceso automatizado de eliminación de ruido por ondículas (wavelets), en (d), (k) y (r) se muestran las imágenes utilizando SRCNN. En (e), (l) y (s) se muestran las imágenes U-net RELU. En (f), (m) y (t) se muestran las imágenes utilizando U-net ELU. En (g), (n) y (u) se muestran las imágenes utilizando U-net Híbrida. Estas imágenes fueron reconstruidas utilizando la reconstrucción por retroproyección lineal [19]

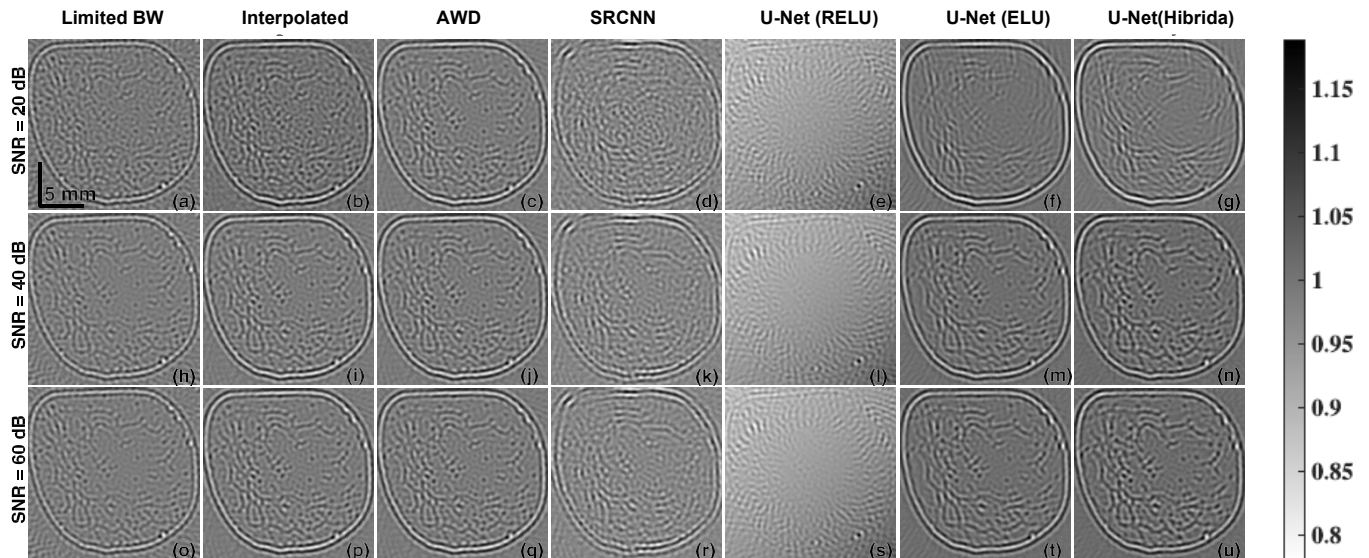


Figura 18 Imágenes reconstruidas por retroproyección lineal utilizando distintos métodos para mejorar la imagen, ancho de banda limitado (a), (h) y (o). Interpolación con vecino más cercano (b), (i) y (p). Utilizando el enfoque automatizado de eliminación de ruido por ondículas (c), (j), y (q). Resultados utilizando la red SRCNN (d), (k) y (r). Resultados utilizando la red U-net RELU (e), (l) y (s). Resultados utilizando la red U-net ELU (f), (m) y (t). Resultados utilizando la red U-net Híbrida (g), (n) y (u) [19].

En la figura 19 se muestra el conjunto de imágenes utilizando las distintas redes y reconstruyéndolas mediante el método de time reversal. En (1), (5) y (9) se muestran las imágenes PAT utilizando la red SRCNN. En (2), (6) y (10) se muestran las imágenes utilizando la red U-net RELU. En (3), (7) y (11) se muestran las imágenes utilizando la red U-net ELU. En (4), (8) y (12) se muestran las imágenes utilizando la red U-net hibrida.

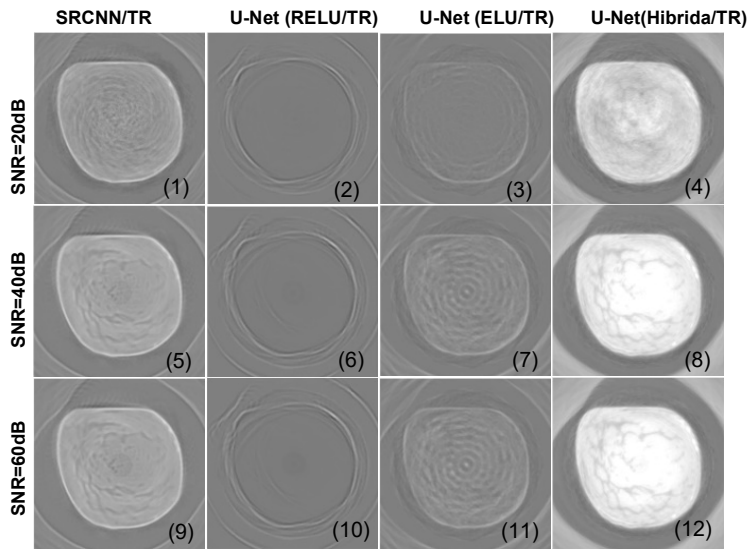


Figura 19 Imágenes reconstruidas mediante time reversal utilizando las redes SRCNN (1),(5),(9), U-net RELU (2),(6),(10), U-net ELU (3),(7),(11) y U-net Hibrida (4),(8),(12).

En la tabla 6 se muestra el desglose de los resultados aplicando las métricas de comparación de Correlación de Pearson (PC). El índice de similitud estructural (SSIM) y el error de raíz cuadrada media (RMSE) a las imágenes de la figura 18 y 19, los valores en negrita expresan el mejor resultado posible para cada métrica. Se puede observar que en el apartado de imagen de mama con nivel de ruido de 20dB en la métrica de PC el mejor valor lo presenta la red U-net Hibrida utilizando Time Reversal como método de reconstrucción dando un valor de **0.9836** con una mejora del **16.17%** respecto al mejor valor presentado con retroproyección lineal **0.8219**. En el apartado SSIM el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.9277** dando una mejora de **22.7%** respecto al mejor valor presentado con LBP **0.7007**. En RMSE el mejor valor lo presenta la red U-net Hibrida con **0.0390** con una mejora del **7.43%** respecto al mejor valor presentado con LBP **0.1133**. En el apartado con un nivel de ruido de 40dB en

la métrica PC el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.9540** con una mejora de **5.93%** respecto al mejor valor presentado con LBP **0.8947**. En SSIM el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.9561** con una mejora de **14.36%** respecto al mejor valor presentado con LBP **0.8125**. En RMSE el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.0773** con una mejora de **.79%** respecto al mejor valor presentado con LBP **0.0852**. En el apartado con un nivel de ruido de 60dB en la métrica PC el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.9533** con una mejora de **5.57%** respecto al mejor valor presentado con LBP **0.8976**. En SSIM el mejor valor lo presenta la SRCNN con reconstrucción TR con un valor de **0.9566** con una mejora de **13.78%** respecto al mejor valor presentado con LBP **0.8188**. En RMSE el mejor valor lo presenta la red SRCNN con reconstrucción TR con un valor de **0.0769** con una mejora de **.85%** respecto al mejor valor presentado con LBP **0.0854**.

IMAGEN MAMA									
SNR of data	20dB			40dB			60dB		
Method/Metric	PC	SSIM	RMSE	PC	SSIM	RMSE	PC	SSIM	RMSE
Limited BW	0.5388	0.4287	0.1515	0.5804	0.4858	0.1623	0.5804	0.4858	0.1623
Interpolated	0.7720	0.6208	0.1176	0.8156	0.6874	0.1446	0.8158	0.6877	0.1442
Automated Wavelet Denoising	0.8158	0.6877	0.1442	0.8158	0.6877	0.1442	0.8158	0.6877	0.1442
SRCNN	0.4389	0.0619	0.2406	0.4715	0.0335	0.2436	0.4714	0.0326	0.2460
U-Net (RELU)	0.0836	0.0787	0.2833	0.0387	0.1574	0.2835	0.0372	0.1586	0.2870
U-net (ELU)	0.8114	0.7007	0.1133	0.8893	0.8125	0.0947	0.8949	0.8188	0.0937
U-net(Hibrida)	0.8219	0.6898	0.1286	0.8947	0.8039	0.0852	0.8976	0.8058	0.0854
SRCNN (Time Reversal)	0.9319	0.9277	0.1008	0.9540	0.9561	0.0773	0.9533	0.9566	0.0769
U-Net (RELU/Time Reversal)	0.8489	0.8857	0.1607	0.8485	0.8812	0.1605	0.8005	0.8404	0.1704
U-Net (ELU/Time Reversal)	0.8700	0.9127	0.1476	0.8929	0.9154	0.1293	0.8951	0.9162	0.1277
U-Net(Hibrida/Time Reversal)	0.9836	0.9161	0.0390	0.9459	0.9500	0.0806	0.9468	0.9506	0.0803

Tabla 6 Resultados comparativos obtenidos para las imágenes reconstruidas de mama que se muestran en la fig. 18 y 19 en términos de figuras de mérito: coeficiente pc, índice SSIM y RMSE para SNR de datos de 20, 40 y 60 dB. El texto en negrita indica el mejor valor posible para la métrica utilizada.

4.4.5 Análisis de procesamiento de Time Reversal

En la tabla 7 se muestran los tiempos de reconstrucción utilizando Time Reversal de cada una de las imágenes usadas para validación: PAT, Derenzo, Vasos Sanguíneos y corte axial de mama, con sus respectivos niveles de ruido en cada una de las diferentes redes. Los mejores tiempos se muestran en negritas indicando el menor valor, de las 12 simulaciones realizadas con cada red se observa que la SRCNN es la que presenta los menores tiempos de reconstrucción con 9 resultados favorables, seguida de U-net RELU, con 3 de los mejores tiempos, las redes que presentan un mayor tiempo de reconstrucción son la U-net ELU y la red U-net Hibrida.

COMPARATIVA DEL TIEMPO DE SIMULACIÓN UTILIZANDO TIME REVERSAL					
Imagen	% de ruido en dB	SRCNN	RELU	ELU	HIBRIDA
Derenzo	20dB	31.6149s	35.6437s	35.0309s	43.8592s
	40dB	34.5394s	33.9143s	37.3527s	37.8523s
	60dB	34.5394s	35.5435s	39.8807s	38.7429s
PAT	20dB	44.6374s	33.114s	39.6564s	34.0664s
	40dB	31.471s	38.4355s	40.0945s	33.8547s
	60dB	32.2188s	37.048s	38.8299s	35.5876s
Vasos sanguíneos	20dB	32.7487s	35.1225s	41.9006s	45.3422s
	40dB	31.88s	38.0975s	43.4547s	42.2858s
	60dB	32.0551s	35.4127s	40.3554s	40.3796s
Mama	20dB	32.6303s	36.9838s	45.152s	40.0066s
	40dB	32.2269s	37.7902s	42.6398s	39.7267s
	60dB	36.6032s	35.322s	43.2332s	40.163s

Tabla 7 Tiempo total de simulación en k-Wave utilizando Time Reversal con las diferentes redes y a niveles de SNR de 20dB,40dB y 60dB, los mejores tiempos se muestran en negrita indicando el menor valor.

CAPÍTULO 5

CONCLUSIONES Y TRABAJO FUTURO

Al realizar el análisis de resultados comparando la calidad de imagen entre los métodos de reconstrucción de Retroproyección lineal y Time reversal, bajo las métricas cuantitativas de Correlación de Pearson (PC), índice de similitud estructural(SSIM) y error de raíz cuadrada media(RMSE) se observa que en la métrica de evaluación cuantitativa de PC de los 12 resultados 9 son favorables para Super-Resolution Convolutional neural network (SRCNN) usando time reversal, dando una mejora de entre 5.57% y 25.74%. Los 3 resultados restantes son para la red U-net hibrida dando una mejora de entre 16.17% y 25.86%. En SSIM 8 resultados son favorables para la red SRCNN/TR dando una mejora de entre 13.78% y 34.25%, los 4 resultados restantes son para la red U-net (ELU) con una mejora de entre 20.27% y 29.72%. Para RMSE 10 resultados son favorables para SRCNN/TR dando una mejora de entre .79% y 6.98%, los 2 resultados

restantes son para la red U-net Híbrida con una mejora de entre 5.75% y 7.43%, por lo que el uso de aprendizaje profundo aplicado a sinogramas para disminuir el ruido presente, junto con time reversal como método de reconstrucción mejora sustancialmente la calidad de las imágenes finales en comparación de LBP. Al hacer la comparación entre los porcentajes de mejora de las métricas de comparación y el tiempo de reconstrucción de las imágenes se concluye que la Super-Resolution Convolutional neural network (SRCNN) es una buena opción para la disminución de ruido en los sinogramas.

Respecto al trabajo a futuro se deben probar otras redes que ayuden a disminuir el ruido en los sinogramas o incluso también que permitan disminuir artefactos, además de probar otros métodos de reconstrucción que permitan mejorar la calidad de imagen al mismo tiempo que se disminuyan los tiempos de reconstrucción.

REFERENCIAS

- [1] A. V. Ravi Kumar, «Applications of laser induced photoacoustic effect for the study of gases and solids, » India, 1992.
- [2] A. Rosencwaig y A. Gersho, «Theory of photoacoustic effect with solids, » *Journal of Applied Physics*, vol. XLVII, nº 1, pp. 64-69, 1976.
- [3] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, *Mastering the game of go without human knowledge*, *Nature* 550 (7676) (2017) 354–359.
- [4] N. Gandhi, M. Allard, S. Kim, P. Kazanzides, and M. A. L. Bell, “Photoacoustic-based approach to surgical guidance performed with and without a da Vinci robot,” *J. Biomed. Opt.*, vol. 22, no. 12, Aug. 2017, Art. no. 121606.
- [5] A. Hauptmann, F. Lucka, M. Betcke, N. Huynh, J. Adler, B. Cox, P. Beard, S. Ourselin, S. Arridge, *Model-based learning for accelerated, limited-view 3-d photoacoustic tomography*, *IEEE Trans. Med. Imaging* 37 (6) (2018) 1382–1393.
- [6] L. V. Wang, «Prospects of photoacoustic tomography» *Medical Physics*, vol. 35, nº 12, pp. 5758-5767, 2008.
- [7] Janek Grohl, Melanie Schellenberg, Kris Dreher, Lena Maier-Hein. *Deep learning for biomedical photoacoustic imaging: A review*. <https://doi.org/10.1016/j.pacs.2021.100241>. *Photoacoustics* 22(2021) 100241.
- [8] Paul Beard, *Biomedical photoacoustic imaging*. *Interface Focus* (2011) 1, 602–631 doi:10.1098/rsfs.2011.0028, Published online 22 June 2011.
- [9] Karabutov, Alexander & Savateeva, Elena & Oraevsky, Alexander. (2003). *Optoacoustic tomography: New modality of laser diagnostic systems*. *Laser Physics*. 13. 711-723.
- [10] R.R ISAI, “Innovación tecnológica de un sistema fotoacústico aplicado en la caracterización y estimulación de tejidos biológicos”, tesis de licenciatura, Instituto Politécnico Nacional, Unidad Profesional Interdisciplinaria de Biotecnología, IPN, Ciudad de México, 2016
- [11] A. G. Bell, «On the production and reproduction of sound by light, » *American Journal of science*, vol. XX, nº 118, p. 305, 1880.
- [12] Rosenblatt, F. (1958). *The perceptron: A probabilistic model for information storage and organization in the brain*. *Psychological review*, 65(6):386.
- [13] I. Vasilev, D. Slater, S. Gianmario, P. Roelants y V. Zocca, *Python Aprendizaje profundo*, Birmingham: Packt Publishing, 2019.
- [14] LeCun, Y., Bengio, Y., and Hinton, G. (2015). *Aprendizaje profundo*. *Nature*, 521(7553):436–444.

[15] H. Lee, J. Lee, H. Kim, B. Cho and S. Cho, "Deep-Neural-Network-Based Sinogram Synthesis for Sparse-View CT Image Reconstruction," in *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 109-119, March 2019, doi: 10.1109/TRPMS.2018.2867611.

[16] Clevert, D., Unterthiner, T., & Hochreiter, S. (2015). *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)*. Retrieved 6 September 2022, from <https://arxiv.org/abs/1511.07289>

[17] Y. LeCun, Y. Bengio y G. Hinton, «Aprendizaje profundo» *Nature*, vol. 521, pp. 436-444, 2015.

[18] Dong, C., Loy, C. C., He, K., & Tang, X. (2016). *Image Super-Resolution Using Deep Convolutional Networks*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2), 295–307. doi:10.1109/tpami.2015.2439281

[19] N. Awasthi, G. Jain, S. K. Kalva, M. Pramanik and P. K. Yalavarthy, "Deep Neural Network-Based Sinogram Super-Resolution and Bandwidth Enhancement for Limited-Data Photoacoustic Tomography," in *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 12, pp. 2660-2673, Dec. 2020, doi: 10.1109/TUFFC.2020.2977210.

[20]. G. L. Zeng, *Medical Image Reconstruction: A Conceptual Tutorial*, Springer, New York (2010).

[21] Hristova, Yulia & Kuchment, Peter & Nguyen, Linh. (2008). "Reconstruction and time reversal in thermoacoustic tomography in acoustically homogeneous and inhomogeneous media". *Inverse Problems - INVERSE PROBL.* 24. 10.1088/0266-5611/24/5/055006. Hristova,

[22] N. Awasthi, S. K. Kalva, M. Pramanik, and P. K. Yalavarthy, "Vector extrapolation methods for accelerating iterative reconstruction methods in limited-data photoacoustic tomography," *J. Biomed. Opt.*, vol. 23, no. 4, Feb. 2018, Art. no. 071204

[23] P. P. Pai, A. De, and S. Banerjee, "Accuracy enhancement for noninvasive glucose estimation using dual-wavelength photoacoustic measurements and kernel-based calibration," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 1, pp. 126–136, Jan. 2018.

[24] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[25] S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis, "Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 568–576, Feb. 2020.

[26] J. Prakash, A. S. Raju, C. B. Shaw, M. Pramanik, and P. K. Yalavarthy, "Basis pursuit deconvolution for improving model-based reconstructed images in photoacoustic tomography," *Biomed. Opt. Express*, vol. 5, no. 5, pp. 1363–1377, Apr. 2014.