



**EDUCACIÓN**

SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO  
NACIONAL DE MÉXICO

# Tecnológico Nacional de México

Centro Nacional de Investigación  
y Desarrollo Tecnológico

## Tesis de Maestría

Segmentación de Masas sólidas benignas y malignas  
en imágenes de Ultrasonidos mediante un  
sistema de Aprendizaje Profundo

presentada por

**Alexis Vázquez Ramírez**

como requisito para la obtención del grado de  
**Maestro en Ciencias de la Computación**

Director de tesis

**Dr. Dante Mújica Vargas**

Cuernavaca, Morelos, México. Enero de 2024.



Cuernavaca, Mor.,

24/enero/2024

No. De Oficio:

SAC/008/2024

Asunto:

Autorización de  
impresión de tesis

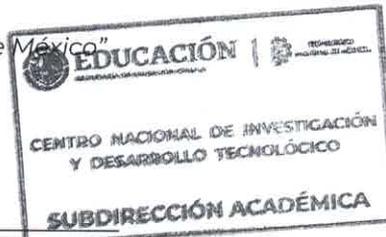
**ALEXIS VÁZQUEZ RAMÍREZ**  
**CANDIDATO AL GRADO DE MAESTRO**  
**EN CIENCIAS DE LA COMPUTACIÓN**  
**P R E S E N T E**

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado **“SEGMENTACIÓN DE MASAS SÓLIDAS MALIGNAS Y BENIGNAS EN IMÁGENES DE ULTRASONIDO MEDIANTE UN SISTEMA DE APRENDIZAJE PROFUNDO”**, ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

**ATENTAMENTE**

*Excelencia en Educación Tecnológica®*  
*“Conocimiento y tecnología al servicio de México”*



**CARLOS MANUEL ASTORGA ZARAGOZA**  
**SUBDIRECTOR ACADÉMICO**

C. c. p. Departamento de Ciencias Computacionales  
Departamento de Servicios Escolares

CMAZ/lmz



Cuernavaca, Mor., **22/enero/2024**

OFICIO No. DCC/010/2024  
Asunto: Aceptación de documento de tesis  
CENIDET-AC-004-M14-OFICIO

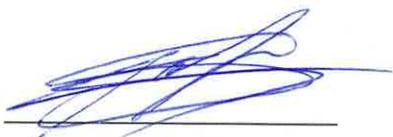
**CARLOS MANUEL ASTORGA ZARAGOZA**  
SUBDIRECTOR ACADÉMICO  
PRESENTE

Por este conducto, los integrantes de Comité Tutorial de **ALEXIS VÁZQUEZ RAMÍREZ** con número de control M21CE070, de la Maestría en Ciencias de la Computación, le informamos que hemos revisado el trabajo de tesis de grado titulado **"SEGMENTACIÓN DE MASAS SÓLIDAS MALIGNAS Y BENIGNAS EN IMÁGENES DE ULTRASONIDO MEDIANTE UN SISTEMA DE APRENDIZAJE PROFUNDO"** y hemos encontrado que se han atendido todas las observaciones que se le indicaron, por lo que hemos acordado aceptar el documento de tesis y le solicitamos la autorización de impresión definitiva.



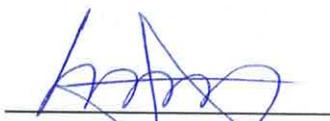
**DANTE MÚJICA VARGAS**

Director de tesis



**NOÉ ALEJANDRO CASTRO SÁNCHEZ**

Revisor 1



**JUAN GABRIEL GONZÁLEZ SERNA**

Revisor 2

C.c.p. Depto. Servicios Escolares.  
Expediente / Estudiante



## **Dedicatoria**

Dedico esta tesis a Dios  
A mis padres, Ovidio y Martha  
A mi abuela Maia  
Y a mi novia, Sofía

## **Agradecimientos**

Agradezco al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) por el apoyo económico brindado durante mis estudios de maestría. Al Tecnológico Nacional de México / Centro Nacional de Investigación y Desarrollo Tecnológico CENIDET por brindar las instalaciones y permitirme realizar los estudios de maestría.

A mi director de tesis el Dr. Dante Mújica Vargas por su apoyo, dedicación y guía brindada durante este proyecto de tesis. A mi comité revisor, el Dr. Noé Alejandro Castro Sánchez y Dr. Juan Gabriel González Serna, por su asesoramiento para el desarrollo de esta investigación. Al Mtro. Manuel Matuz por su confianza, amistad y dedicación brindada.

A mis compañeros del clan Nephilim, por su amistad y apoyo brindado durante este proyecto, los cuales llegue a considerar como parte de mi familia.

¡Gracias!

## Resumen

El cáncer de mama es la neoplasia más frecuente en mujeres, y uno de los tres cánceres más comunes en todo el mundo. Este tipo de cáncer se origina en los tejidos de la mama, específicamente en las células de los conductos o los lobulillos mamarios. Su análisis se basa en la aplicación de nuevos estándares de detección, un ejemplo de ello es el diagnóstico asistido por computadora, en tareas como la segmentación de masas mamográficas. Sin embargo, si bien el análisis funciona, no es inmune al error teniendo un margen considerable debido a la calidad de las imágenes. Para mejorar dicha calidad, se implementan técnicas de procesamiento en los pasos posteriores a la adquisición. Esto permite aumentar el contraste en la imagen mamográfica y ayuda a reducir el ruido aparente. Este estudio propone un sistema de segmentación de masas sólidas mediante tres módulos. El primero realiza un preprocesamiento de la imagen con el propósito de reducir el ruido y resaltar los límites de la masa mediante el filtro SRAD. Seguidamente, el segundo implementa el algoritmo de objetos YOLOv3 para delimitar la región de interés, evitando así posibles confusiones y asegurando que la red de segmentación se centre exclusivamente en la zona de interés previamente identificada. Finalmente, para el último módulo se aplica la red de segmentación semántica DeepLabv3+ exclusivamente a la región previamente delimitada por la red de detección. En este contexto, el objetivo principal de este trabajo es lograr la segmentación precisa de masas sólidas en imágenes de ultrasonido, independientemente de su forma, tamaño o naturaleza benigna o maligna. Esta metodología se presenta como una herramienta complementaria para los médicos durante el examen mamográficos y de acuerdo a los experimentos realizados obtiene resultados en un valor promedio de acuerdo al el índice de Jaccard igual a 0.89%, índice de similaridad de Dice igual a 0.94%, *Precision* de 0.98%, *Recall* de 0.99% y un *F1 – score* de 0.98% siendo en comparativa con la implementación del modelo original DeepLabv3+ más precisos, lo que demuestra la robustez de la metodología propuesta.

**Palabras clave:** Cáncer de mama, Ecografía, Ruido multiplicativo, SRAD, YOLOv3, DeepLabv3+.

## Abstract

Breast cancer is the most common cancer in women, and one of the three most common cancers worldwide. This type of cancer originates in the tissues of the breast, specifically in the cells of the breast ducts or lobules. Its analysis is based on the application of new detection standards, such as computer-aided diagnosis, in tasks such as mammographic mass segmentation. However, while the analysis works, it is not immune to error and has a considerable margin of error due to the quality of the images. To improve image quality, processing techniques are implemented in the post-acquisition steps. This allows for increased contrast in the mammographic image and helps to reduce apparent noise. This study proposes a solid mass segmentation system using three modules. The first one performs a pre-processing of the image in order to reduce noise and highlight the mass boundaries using the SRAD filter. Next, the second implements the YOLOv3 object algorithm to delimit the region of interest, thus avoiding possible confusion and ensuring that the segmentation network focuses exclusively on the previously identified area of interest. Finally, for the last module, the DeepLabv3+ semantic segmentation network is applied exclusively to the region previously delimited by the detection network. In this context, the main objective of this work is to achieve accurate segmentation of solid masses in ultrasound images, regardless of their shape, size or benign or malignant nature. This methodology is presented as a complementary tool for physicians during mammographic examination and according to the experiments performed, it obtains results in an average value according to the Jaccard index equal to 0.89%, Dice similarity index equal to 0.94%, Precision of 0.98%, Recall of 0.99% and a F1-Score of 0.98% being in comparison with the implementation of the original DeepLabv3+ model more accurate, which demonstrates the robustness of the proposed methodology.

**Keywords:** Breast cancer, Ultrasound, Multiplicative noise, SRAD, YOLOv3, DeepLabv3+.

# Índice General

<b>Lista de Figuras</b>	<b>viii</b>
<b>Lista de Tablas</b>	<b>x</b>
<b>1 Introducción</b>	<b>1</b>
1.1 Planteamiento del problema . . . . .	2
1.1.1 Delimitación del problema . . . . .	3
1.1.2 Complejidad del problema . . . . .	3
1.2 Objetivos . . . . .	4
1.2.1 Objetivo general . . . . .	4
1.2.2 Objetivos específicos . . . . .	4
1.3 Alcances y limitaciones . . . . .	5
1.3.1 Alcances . . . . .	5
1.3.2 Limitaciones . . . . .	5
1.4 Justificación . . . . .	6
1.5 Organización de la tesis . . . . .	6
<b>2 Marco Teórico</b>	<b>7</b>
2.1 Cáncer de Mama . . . . .	7
2.1.1 Masas Mamográficas . . . . .	8
2.2 Diagnóstico del Cáncer de Mama . . . . .	11
2.2.1 Ecografía . . . . .	11
2.3 Ruido Multiplicativo . . . . .	13
2.3.1 Modelo General del Ruido Multiplicativo . . . . .	14
2.4 Filtros para reducción de Ruido Multiplicativo . . . . .	15
2.4.1 Filtros Clásicos . . . . .	15
2.4.2 Filtros Adaptativos . . . . .	18
2.5 Red Neuronal Convolutiva . . . . .	23
2.5.1 Darknet-53 . . . . .	25
2.5.2 <i>You Only Look Once</i> . . . . .	26
2.5.3 DeepLab . . . . .	32

---

<b>3</b>	<b>Estado del Arte</b>	<b>45</b>
3.1	Antecedentes . . . . .	45
3.2	Trabajos relacionados . . . . .	50
3.3	Discusión del Estado del Arte . . . . .	62
<b>4</b>	<b>Metodología de Solución</b>	<b>65</b>
4.1	Propuesta de solución . . . . .	65
4.1.1	Módulo de Filtrado . . . . .	66
4.1.2	Módulo de Detección . . . . .	67
4.1.3	Módulo de Segmentación . . . . .	68
<b>5</b>	<b>Experimentación y Resultados</b>	<b>69</b>
5.1	Diseño Experimental . . . . .	69
5.1.1	Entorno de desarrollo . . . . .	69
5.1.2	Base de datos . . . . .	69
5.1.3	Métricas . . . . .	71
5.2	Filtrado de Ruido Multiplicativo . . . . .	74
5.3	Detección de Masas en Ecografías . . . . .	80
5.4	Segmentación de Masas en Ecografías . . . . .	85
<b>6</b>	<b>Conclusiones</b>	<b>93</b>
6.1	Objetivos y alcances logrados . . . . .	93
6.2	Resultados del trabajo . . . . .	95
6.2.1	Productos . . . . .	95
6.3	Conclusiones . . . . .	95
6.4	Trabajo futuro . . . . .	96
	<b>Referencias</b>	<b>97</b>
	<b>Apéndice A Producción</b>	<b>107</b>
	<b>Apéndice B Retribución social</b>	<b>109</b>

# Lista de Figuras

2.1	Tipos de masas sólidas en mamografías [4]. . . . .	10
2.2	Ecografía de mama [75]. . . . .	12
2.3	Modelo de bloques del filtro adaptativo [133]. . . . .	19
2.4	Arquitectura de red Darknet-53 [99]. . . . .	25
2.5	Diagrama de detección de YOLOv3 [98]. . . . .	27
2.6	Mejora propuesta por DeepLabv3+ [20]. . . . .	34
2.7	Convolución <i>atrous</i> en 1-D [18]. . . . .	35
2.8	Convolución <i>atrous</i> en 2-D [18]. . . . .	36
2.9	Convolución <i>atrous</i> con tamaño de <i>kernel</i> $3 \times 3$ y diferentes <i>rate</i> [20]. . . . .	37
2.10	Arquitectura general de DeepLabv3+ [20]. . . . .	40
2.11	<i>Atrous Separable Convolution</i> [20]. . . . .	40
2.12	Arquitectura del decodificador de DeepLabv3+ [20]. . . . .	42
3.1	Arquitectura general de modelo propuesto [77]. . . . .	46
3.2	Modelo de red neuronal y aplicación de filtros [77]. . . . .	46
3.3	Metodología propuesta [80]. . . . .	47
3.4	Flujo de procesos para las etapas de inicialización, transformación de los datos y agrupamiento [33]. . . . .	48
3.5	Flujo de procesos para las etapas de etiquetado y extracción de superpíxeles [33]. . . . .	48
3.6	Comparación de eliminación de ruido y artefactos con las tres regularización espaciales y el trabajo de [2] [85]. . . . .	52
3.7	Resultado cualitativos del modelo propuesto [94]. . . . .	56
3.8	Arquitectura FT-MTL-Net propuesta[42]. . . . .	57
4.1	Diagrama de metodología de solución propuesta. . . . .	65
4.2	Módulo de filtrado: SRAD. . . . .	66
4.3	Módulo de detección: YOLOv3. . . . .	67
4.4	Módulo de detección: DeepLabv3+. . . . .	68
5.1	<i>Ground truth</i> de masas del conjunto de datos MID [75]. . . . .	70
5.2	Ecografías y <i>ground truth</i> de la base de datos BUSI [4]. . . . .	70
5.3	Normalización del conjunto de datos [4]. . . . .	71
5.4	Resultados cuantitativos de las métricas SNR, PSNR, MSE y SSIM. . . . .	75

---

5.5	PSNR en base de datos BUSI y MID. . . . .	76
5.6	MSE en base de datos BUSI y MID. . . . .	76
5.7	SSIM en base de datos BUSI y MID. . . . .	77
5.8	Imágenes de ultrasonido utilizadas. . . . .	77
5.9	Filtrado de ecografía con masa sólida benigna. . . . .	78
5.10	Filtrado de ecografía con masa sólida maligna. . . . .	79
5.11	Etiquetado manual de masa sólida mediante Labelimg. . . . .	81
5.12	Comparación de umbrales de detección en YOLOv3. . . . .	81
5.13	Ecografías con diferentes tipos masas [44]. . . . .	82
5.14	Precisión de detección de YOLOv3. . . . .	83
5.16	Detección de masas en ecografías mediante YOLOv3. . . . .	84
5.17	Resultados de la detección de masas en ecografías de las bases de datos utilizada. . . . .	84
5.18	Proceso de segmentación mediante el sistema propuesto. . . . .	86
5.19	Resultados cuantitativos de segmentación de masas sólidas en base de datos utilizada. . . . .	87
5.20	Resultados cuantitativos en segmentación de masas sólidas benignas. . . . .	88
5.21	Resultados cuantitativos en segmentación de masas sólidas malignas. . . . .	89
5.22	Índice Jaccard en base de datos BUSI y MID. . . . .	90
5.23	Índice Dice en base de datos BUSI y MID. . . . .	90
5.24	Comparación de segmentación de masas benignas con el sistema propuesto y DeepLabv3+. . . . .	91
5.25	Comparación de segmentación de masas malignas con el sistema propuesto y DeepLabv3+. . . . .	91
A.1	Constancia de ponencia. . . . .	107
A.2	Artículo de algoritmos de filtrado. . . . .	108
B.1	Constancia de taller. . . . .	109

# Lista de Tablas

3.1	Resultados de ENL y SSI obtenidos en imágenes de ultrasonido [78]. . . .	50
3.2	Promedio de PSNR y MSE obtenido para cada clase de anormalidad [85].	51
3.3	Comparación de valores de SNR y SSIM obtenidos en una imagen sintética [95]. . . . .	54
3.4	Comparación de valores de CNR obtenidos de 4 regiones en una imagen de US [95]. . . . .	54
3.5	Comparación entre el modelo FT-MTL-Net propuesto y otros métodos de detección de masas en la base de datos INBreast [42]. . . . .	58
3.6	Comparación entre el modelo FT-MTL-Net propuesto y otros métodos de segmentación de masas en la base de datos INBreast [42]. . . . .	58
3.7	Resultados de segmentación basados en el modelo U-Net modificado basado en vista MLO [107]. . . . .	59
3.8	Comparación del rendimiento de los métodos de segmentación en imágenes del conjunto de datos MIAS [106]. . . . .	60
3.9	Comparación del rendimiento global en la tarea de delimitación entre modelo propuesto y los modelos de la literatura en base de datos INbreast [128]. . . . .	61
3.10	Resumen del estado del arte. . . . .	62
5.1	Subconjuntos de imágenes obtenidos de BUSI y MID. . . . .	71
5.2	Resultados del entrenamiento de la red. . . . .	81
5.3	Comparación del rendimiento global en la tarea de segmentación de masas en ecografías entre el sistema propuesto y los modelos de la literatura. . .	92
6.1	Objetivos cumplidos. . . . .	94
6.2	Alcances realizados. . . . .	94

# Capítulo 1

## Introducción

El diagnóstico del cáncer de mama se basa en la aplicación de nuevos estándares de detección y análisis, un ejemplo de ello es el diagnóstico asistido por computadora (CAD), como lo es la mamografía digital. Esta técnica se basa en detectar características complejas, que pueden aparecer en forma de diversas anomalías, como masas sólidas en la mamografía. La eficacia de este diagnóstico depende, en gran medida, de la precisión en la clasificación y segmentación de la masa. La segmentación de estas masas tiene como objetivo la extracción, identificación y seguimiento de sus características. Sin embargo, esto demanda complejidad en el proceso de interpretación de la imagen debido principalmente a su calidad, la variabilidad en la apariencia de las anomalías y las características del tejido mamario. Por lo que, en muchas ocasiones las anomalías no pueden ser visualizadas. Para mejorar dicho pronóstico, se implementan técnicas de procesamiento en los pasos posteriores a la adquisición, permitiendo aumentar el contraste en la imagen, reducir el ruido y resaltar los bordes de la masa.

A grandes rasgos, una delimitación y/o segmentación implica dividir una imagen en una serie de partes de acuerdo a sus características, con la finalidad de un mayor nivel de análisis y comprensión. Hasta la fecha, no existe un método general para la segmentación de imágenes, así como criterios objetivos para evaluar si la delimitación es satisfactoria. De acuerdo a la literatura, los algoritmos para realizar una segmentación de masas en imágenes mamográficas son generalmente de dos tipos, basados en píxeles y basados en regiones.

Con base en lo anterior, este trabajo propone la implementación de un sistema de delimitación de masas sólidas en ecografías mediante tres procesos. Se realiza un preprocesamiento de la imagen con el propósito de reducir el ruido y resaltar los límites de la masa. 2) Se implementa un algoritmo de objetos para delimitar la región de interés, evitando así posibles confusiones y asegurando que la red de segmentación se centre exclusivamente en la zona de interés. 3) se aplica una red de segmentación semántica exclusivamente a la región previamente delimitada por la red de detección.

## 1.1 Planteamiento del problema

La segmentación de masas en una ecografía es un proceso importante en el análisis de imágenes médicas. Una segmentación correcta de la masa durante el diagnóstico, puede ayudar a mejorar la precisión y utilidad de la información proporcionada por la imagen, lo que es sustancial para su caracterización y tratamiento, además de que en caso necesario podría ayudar a una detección temprana de cáncer de mama. Sin embargo, este método de detección tiene sus propias limitaciones debido a factores como el ruido multiplicativo aparente, la calidad del equipo de captura, o la densidad del tejido mamario que pueden afectar a la calidad de la imagen. Por tanto, se requieren métodos de análisis de imágenes para detectar y delimitar las masas, que brinden información morfológica correcta al médico con la finalidad de respaldar un diagnóstico confiable.

Por otro lado, existen otros factores propios de la mama que afectan la segmentación de la masa y por ende la precisión del diagnóstico. Estos incluyen el tipo de masa, benigna o maligna. Las masas pueden tener formas y tamaños diferentes, por ejemplo, una masa maligna puede ser irregular o tener bordes difusos, lo que hace difícil la tarea de delimitarlas con precisión; lesiones mamarias pequeñas o microcalcificaciones, que puedan ser difíciles de identificar y segmentar, especialmente si están ubicadas en áreas de tejido denso. Otro factor importante son las estructuras complejas de la mama que en algunos casos pueden generar una superposición de tejidos o estructuras, lo que hace difícil distinguir los límites de una masa, esto último es especialmente desafiante en áreas densas. Estudios revelan que durante el análisis de masas mamográficas de manera tradicional se obtiene una sensibilidad del 97%, una especificidad del 64.5%, un valor predictivo positivo de 89%, y un valor predictivo negativo de 90.9% dando una precisión del diagnóstico de solo 89.3%.

**Hipótesis:** La implementación de un sistema de varias etapas como la reducción de ruido y detección de región de interés permitirá mejorar los resultados de segmentación de masas sólidas en imágenes de ecografía.

### 1.1.1 Delimitación del problema

En este trabajo se desarrollará un sistema de tres etapas para la segmentación de masas sólidas en mamografías basado en técnicas de aprendizaje profundo, el cual segmentará una masa sólida a través de una red neuronal convolucional.

### 1.1.2 Complejidad del problema

Durante un examen mamográfico, la segmentación de las masas sólidas es crucial para el diagnóstico y tratamiento del cáncer de mama. Sin embargo, este proceso tiene cierto grado de complejidad, dado que radica en el análisis y la correcta interpretación imagenológica subjetiva que realiza el médico, al ser una tarea completamente visual. Algunos de los puntos importantes que pueden afectar a este proceso son:

- Variabilidad en la apariencia de las masas: se refiere a los diversos tipos de bordes, dado que se consideran una característica importante en la determinación de su malignidad o benignidad. Así mismo, el tamaño de la masa, puesto que algunas de las lesiones suelen ser muy pequeñas y pueden llegar a confundirse con tejido mamario normal.
- Características del tejido mamario: muchas veces, las anomalías no pueden ser visualizadas correctamente debido a que la mama está constituida por tejidos muy similares entre sí. El tipo de estructura, hace difícil distinguir los límites de una masa, debido a que la sutileza de estas lesiones podrían estar enmascaradas por tejidos homogéneos del parénquima mamario.
- Calidad de la imagen: las imágenes de ultrasonido, por su naturaleza de adquisición, son corrompidas por ruido multiplicativo. Este ruido degrada la imagen al ser un factor dominante, dificultando la interpretación del médico, dado que puede aparecer como puntos claros y oscuros, y puede alterar los bordes de la masa.

## 1.2 Objetivos

### 1.2.1 Objetivo general

Desarrollar un sistema de segmentación de masas sólidas en imágenes de ultrasonidos utilizando técnicas de aprendizaje profundo, con el propósito de mejorar la precisión y eficacia en la delimitación de las masas sólidas en el diagnóstico de cáncer de mama.

### 1.2.2 Objetivos específicos

- Investigar y comprender las técnicas de preprocesamiento de imágenes médicas y conceptos clave del aprendizaje profundo, específicamente en el contexto de filtrado, detección y segmentación de masas en imágenes de ultrasonidos.
- Recopilar y analizar una base de datos de imágenes de ultrasonidos con masas sólidas mamarias.
- Realizar un preprocesamiento en las imágenes de ultrasonidos, incluyendo normalización, reducción de ruido y mejora de la calidad, para optimizar el rendimiento del sistema.
- Diseñar y desarrollar un sistema basado en aprendizaje profundo adecuado para una segmentación precisa de masas sólidas en imágenes de ultrasonidos, considerando la reducción de ruido de la imagen, así como la detección y delimitación de los bordes de las masas.
- Entrenar y ajustar el sistema de aprendizaje profundo utilizando técnicas de entrenamiento y validación con el fin de lograr una detección y delimitación precisa de la masa.
- Evaluar el rendimiento del sistema propuesto mediante métricas como PSNR, SSIM, *Precision*, *Recall*, *Specificity*, índice de Jaccar y coeficiente de similaridad de Dice, comparando los resultados obtenidos con técnicas del estado del arte.

## 1.3 Alcances y limitaciones

### 1.3.1 Alcances

- Utilizar técnicas de reducción de ruido multiplicativo en las imágenes de la base de datos propuesta con la finalidad de mejorar su calidad.
- Desarrollar un sistema de aprendizaje profundo para la segmentación de masas sólidas en ecografías.
- Evaluar la calidad del desempeño del sistema propuesto con métricas definidas en la literatura.
- Comparar los resultados obtenidos con modelos de segmentación existentes.

### 1.3.2 Limitaciones

- Este trabajo solo se centrará en la segmentación de masas sólidas en imágenes ecográficas, las cuales contienen un *ground truth* de la masa.
- Para esta investigación solo se trabajará con imágenes ecográficas.
- Se espera que los datos estén adecuadamente clasificados, dado que el sistema propuesto no clasificara la naturaleza benigna o maligna de la masa.
- Debido a la naturaleza de las métricas especializadas en medir la calidad de reducción de ruido multiplicativo y la no existencia de una imagen ecográfica referente (sin ruido), los resultados del módulo de filtrado se comparará con el modelo de reducción de ruido dnCNN.
- Se espera que el sistema propuesto al menos iguale la calidad de segmentación de otros métodos del estado del arte en la base de datos propuesta.

## 1.4 Justificación

El diagnóstico de cáncer de mama mediante la segmentación de masas sólidas en las imágenes mamográficas es una tarea compleja debido al ruido aparente y las masas mal definidas. Sin embargo, aunque existen múltiples técnicas de segmentación de masas en ecografía mediante CAD los enfoques de desarrollo solo se centran en la segmentación de la masa en una ecografía completa, lo que puede llegar a confundir la red de segmentación dado la homogeneidad de las imágenes. Además, estas solo se centran en masas benignas, dado que estas son mejor definidas a diferencia de una masa maligna. Por otro lado, las técnicas de supresión de ruido en la literatura se centran generalmente en los ruidos aditivos que son independientes de la señal, por lo que no logran ser correctamente suprimidos o degradan la imagen.

Este trabajo propone un sistema de segmentación de masas sólidas basado en tres procesos. El primero se enfoca en un preprocesamiento de la imagen para reducir el ruido y resaltar los límites de la masa. El segundo implementa una red de detección de objetos con la finalidad de delimitar la región de interés, así evitando que la red de segmentación se confunda y delimite otros lugares en la ecografía. Y por último, una red de segmentación semántica que solo se aplica a la región antes delimitada por la red de detección. En este sentido, el objetivo de este trabajo es segmentar las masas sólidas en ecografías independientemente de su forma o tamaño, lo que puede ayudar como una segunda opinión al médico durante el examen mamográfico.

## 1.5 Organización de la tesis

Este documento de tesis está compuesto por seis capítulos, anexos y referencias. El capítulo 2 contiene el marco teórico, el cual describe conceptos importantes para el tema de investigación propuesto, como lo es el cáncer de mama, su diagnóstico, ruido multiplicativo, y técnicas de filtrado, reducir este tipo de ruido; así mismo, se describe el concepto de red neuronal convolucional seguido de la descripción de dos tipos utilizadas en tareas de detección y segmentación de objetos. En el capítulo 3 se presentan los antecedentes y el estado del arte, mencionando aquellos trabajos que fueron considerados de gran relevancia para esta investigación. En el capítulo 4 se presenta la metodología propuesta, detallando sus componentes e importancia para la investigación. En el capítulo 5 se detalla la experimentación realizada y los resultados obtenidos. Y finalmente, en el capítulo 6 se describen las conclusiones generales a las que se llegó a lo largo de la investigación, los productos desarrollados y los trabajos futuros.

# Capítulo 2

## Marco Teórico

En este capítulo se describen los conceptos básicos necesarios para el completo entendimiento y desarrollo del tema de investigación propuesto.

### 2.1 Cáncer de Mama

El cáncer de mama es la neoplasia más frecuente en mujeres, y uno de los tres cánceres más comunes en todo el mundo [86]. Este tipo de cáncer se origina en los tejidos de la mama, específicamente en las células de los conductos o los lobulillos mamarios. Al igual que otros cánceres, el cáncer de mama se considera una enfermedad progresiva, ya que corresponde a un crecimiento maligno que evoluciona a través de varias etapas de diferencia celular. Este crecimiento se inicia en los tejidos mamarios donde pueden invadir los tejidos circundantes, es decir, hacer metástasis en zonas distantes del cuerpo a través del sistema sanguíneo o linfático [61]. Sobre esta base, los cánceres que se originan en los conductos se denominan carcinomas ductales, mientras que el de los lobulillos se denominan carcinomas lobulares [122, 10].

En condiciones normales, las células del cuerpo se dividen y crecen de manera ordenada, pero en el caso del cáncer, este proceso se ve alterado, y las células comienzan a crecer de manera descontrolada, formando una masa o tumor en la mama. Esta se encuentra formada por grasa, glándulas y tejidos. Estos últimos suelen ser densos, y a medida que envejecen se transforman en tejido graso, lo que genera que sean más propensos a desarrollar tumores [61].

De acuerdo a la *American Cancer Society* [116] las principales causas del cáncer de mama son el envejecimiento, la composición de la densidad mamaria, la herencia, la menarquia temprana, la menopausia tardía y las mutaciones genéticas; así como los factores riesgo ambientales, como lo es la exposición a radiación ionizante o no ionizante. Sin embargo, al igual que otros tipos de cáncer, este se puede ser tratado de forma correcta si es detectado en una etapa temprana [122]. La clasificación de las masas pretende especificar y unificar sus características individuales, así como estados de desarrollo, con

la finalidad de un mayor impacto en las técnicas de diagnóstico y pronóstico de cáncer de mama. Existen varias clasificaciones de las masas basadas en aspectos clínicos, etiológicos, histológicos y otros que caracterizan el estado de masa y de las metástasis locales y remotas [112, 32].

### 2.1.1 Masas Mamográficas

Una masa es una lesión voluminosa que ocupa un área aleatoria con una densidad central más densa y un margen relativamente delgado. En otras palabras, una masa es una estructura tridimensional que tiene, total o parcialmente, bordes externos convexos y es evidente en al menos dos proyecciones de la mama [103, 122, 32]. Una masa puede clasificarse en benigna y maligna de acuerdo a sus características físicas como [116, 84]:

- Densidad, esta define la atenuación de los rayos X por la lesión, tomando en cuenta la disminución esperada para un volumen de tejido glandular similar. Esta característica es importante, ya que la mayoría de los cánceres de mama que forman una masa visible, son de igual o mayor densidad que un volumen similar de tejido glandular normal.
- Márgenes, se refiere a los bordes de una masa, los cuales tienen que ser cuidadosamente clasificados debido a que se considera una característica importante en la determinación de la malignidad o benignidad de la masa. Los expertos clasifican los márgenes de una masa en circunscritos, microlobulados, mal definidos y espiculados.
- Forma, una masa puede ser de diferente forma, independientemente de sus bordes. De acuerdo a [84] estas pueden ser redondas, ovaladas, lobulares o irregulares.

La etiología de las masas puede deberse a diversos factores como: quistes, fibroadenomas, tumores y calcificaciones. Estos elementos son parte importante dado que definen su naturaleza benigna o maligna [101].

Se considera que una masa es benigna cuando es detectada en una mamografía, pero no indica la presencia de tejido canceroso. Aunque este tipo de masas pueden variar en su origen y características; en muchos casos no se consideran una amenaza para el paciente, por lo que pueden no requerir un tratamiento. Sin embargo, en algunos casos, dependiendo de la situación clínica, el médico puede optar por un monitoreo continuo o realizar procedimientos para extirparlas. Este tipo de masa suele estar bien definida en comparación con las malignas y por lo general, están rodeadas por una línea radiolúcida

una denominada “*halo*” considerándose esta, una característica de su naturaleza benigna. Ejemplos de masas benignas incluyen [103, 109, 32]:

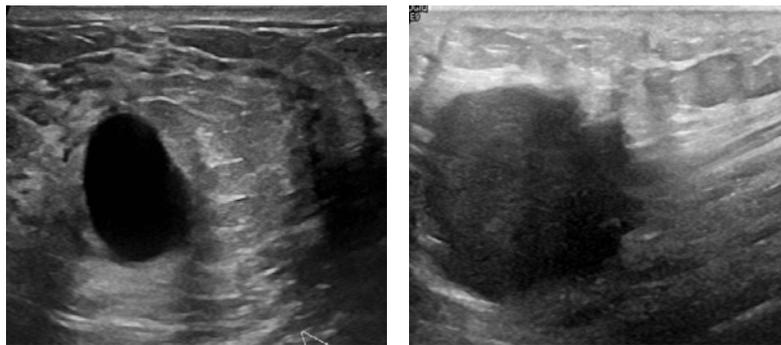
- Fibroadenoma, esta es una masa sólida compuesta por tejido glandular y conectivo, pudiendo ser móviles y de bordes bien definidos en la mamografía.
- Quiste, es una masa llena de líquido que puede aparecer como una masa redonda con bordes suaves en la mamografía y pueden variar en su tamaño.
- Adenosis, es un agrandamiento benigno de las glándulas mamarias que aparece como una masa densa en la mamografía y a menudo está asociada con cambios en el tejido mamario.
- Lipoma, es una masa compuesta por tejido graso y puede aparecer como una masa suave y bien delimitada en la mamografía.
- Hemangioma, es una masa benigna formada por vasos sanguíneos. En la mamografía, puede presentar una apariencia característica debido a la vascularización.
- Nódulo linfático aumentado, en ocasiones, los ganglios linfáticos cerca de la mama pueden aumentar de tamaño en respuesta a infecciones o inflamación y pueden parecer como masas durante la mamografía.

Por otro lado, una masa maligna se refiere a un crecimiento anormal de tejido en la mama el cual se considera indicativo de cáncer de mama. La malignidad de esta radica en que las células en la masa tienen la capacidad de invadir los tejidos circundantes y, en algunos casos, de propagarse a otras partes del cuerpo a través del sistema linfático o sanguíneo, formando metástasis. Las masas malignas pueden tener características distintivas a las benignas que ayudan a los médicos a detectarlas y reducir el riesgo de cáncer. Algunas de estas características indicativas de su malignidad incluyen, bordes irregulares o espiculados, patrones de microcalcificaciones, diferencia asimétrica en la densidad mamaria y la presencia de vasos sanguíneos anormales. Una masa maligna puede presentar una variedad de tipos histológicos, lo que significa que pueden derivar de diferentes tipos de células en la mama. Algunos de los tipos más comunes de masas malignas pueden ser [103, 109, 32]:

- Carcinoma Ductal In Situ (CDIS), en este tipo, las células cancerosas se encuentran dentro de los conductos mamarios y no han invadido los tejidos circundantes.
- Carcinoma Ductal Invasivo (CDI), este se origina en los conductos mamarios y tiene la capacidad de invadir los tejidos circundantes fuera de los conductos.

- Carcinoma Lobular Invasivo (CLI), este tipo comienza en los lóbulos de la mama y puede infiltrar los tejidos cercanos, por lo que puede ser más difícil de detectar en las ecografías.
- Carcinoma inflamatorio de mama, es poco común dado que se basa en la inflamación y el enrojecimiento de la piel en la mama, pero puede no formar una masa visible durante la ecografía.
- Carcinoma papilar, aunque en su mayoría son benignos, algunos carcinomas papilares pueden ser malignos y presentar características histológicas que indican su naturaleza.
- Carcinoma metastásico, en algunos casos, las masas malignas en la mama pueden ser metástasis de cánceres que se originaron en otros órganos.

Las Figura 2.1 muestra ejemplos de masas en una mama.



(a) Masa sólida benigna

(b) Masa sólida maligna

Figura 2.1 Tipos de masas sólidas en mamografías [4].

Actualmente todavía no es posible prevenir por completo el cáncer de mama, aunque puede tratarse de manera eficaz si aún no ha llegado a metástasis. Por tanto, la detección temprana es necesaria si se quiere tratar y controlar su progreso, al proporcionar al paciente un mejor pronóstico de vida [109]. La realización de un examen mamográfico mediante el análisis de las masas es importante para detectar signos tempranos de cáncer en pacientes que no presentan síntomas, o para confirmar la presencia de este en pacientes que sí presentan síntomas. Algunas de las técnicas más comunes para la detección del cáncer de mama son la mastografía de rayos X y la ecografía de mama [116].

## 2.2 Diagnóstico del Cáncer de Mama

Este se basa en la aplicación de nuevos estándares de detección y análisis, un ejemplo de ello es el diagnóstico asistido por computadora (CAD), como lo es la mamografía digital. Esta técnica se basan en detectar características complejas, que pueden aparecer en forma de diversas anomalías, como masas mamográficas. Este tiene como objetivo detectar síntomas no homogéneos y sutiles del cáncer de mama mediante técnicas como la ecografía y/o mastografía de rayos X. Durante el diagnóstico, el médico realiza una detección visual de cualquier tipo de anormalidad, como masas, calcificaciones, asimetrías o cambios en la densidad mamaria. Como ya se mencionó, existen diversas técnicas para la detección de cáncer de mama, pero la más importante para los médicos es el análisis de mamografías para detectar y segmentar masas en los senos [57].

La segmentación de una masa se considera un aspecto crucial para el diagnóstico de cáncer de mama. Este diagnóstico depende de la claridad con la que se pueden distinguir los bordes de una masa en estas imágenes. Dicha importancia de esta interpretación radica principalmente en los siguientes puntos [89, 13]: la diferenciación entre la naturaleza benigna o maligna de la masa, una guía para la estimación del tamaño y extensión de la masa dentro de la mama, el seguimiento y evaluación de cambios de la masa a lo largo del tiempo, y el apoyo en las decisiones del médico con respecto al tratamiento del paciente.

Sin embargo, la complejidad de este diagnóstico radica en el análisis y la correcta interpretación imagenológica que realiza el médico, dado que es una tarea visual que requiere experiencia y práctica. La literatura indica que ciertas anomalías pueden pasar desapercibidas durante una revisión de la ecografía [115]. Esto podría deberse, en gran medida, a la sutileza de las lesiones sospechosas y/o, a que estas lesiones están enmascaradas por tejidos homogéneos del parénquima mamario. Por lo que esto plantea una alta demanda de médicos expertos en análisis de imágenes médicas capaces de detectar masas casi imperceptibles [89, 13].

### 2.2.1 Ecografía

Una ecografía se basa en la emisión de pulsos cortos de energía ultrasónica a partir de elementos transductores, para luego escuchar los ecos de diferentes profundidades del tejido. En una prueba de ultrasonido, la intensidad del eco recibido se usa para controlar el brillo de la pantalla de visualización. En una imagen ultrasónica, el brillo representa una combinación compleja de propiedades ecográficas del tejido y del equipo de ultrasonido. Durante un examen ecográfico, el proceso de transmisión de los pulsos y recepción de ecos

se repite dirigiendo los pulsos ultrasónicos en diferentes direcciones. Los ecos recibidos desde todas las direcciones se combinan para crear una imagen bidimensional de los senos [112, 38].

Una imagen de ultrasonido puede revelar diferentes capas de tejido blando en el seno. Debido a que la piel es reflectante, esta aparece como una línea brillante en la imagen. Los tejidos grasos causan niveles bajos de eco y aparecen oscuros en la imagen. Por el contrario, los tejidos parenquimatosos y fibrosos reflejan fuertemente el ultrasonido y aparecen brillantes. Una imagen ecográfica de mama, Figura 2.2, suele ser bastante variable y depende, en primer lugar, de la edad del paciente y de la fase del ciclo menstrual en mujeres fértiles [112, 38].

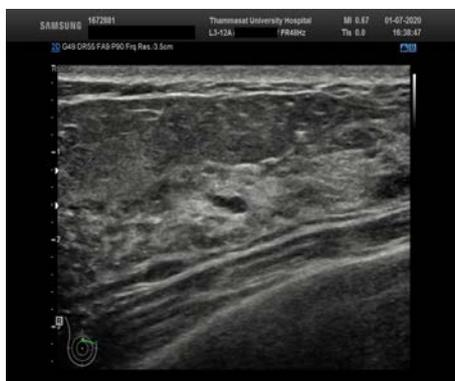


Figura 2.2 Ecografía de mama [75].

Un problema subjetivo durante el diagnóstico, de esta patología mamaria, es que puede ser malinterpretada por la experiencia insuficiente de los especialistas, lo que da lugar a errores de diagnóstico. Así mismo, suele haber casos de hiperdiagnóstico de la patología, cuando las estructuras mamarias normales se interpretan como lesiones. Esto afecta especialmente a las estructuras situadas en la periferia de la mama, en el espacio retromamario y en la zona subareolar [112]. Estudios revelan que durante el análisis de las masas empleando este tipo de técnica se obtiene una sensibilidad del 97%, una especificidad del 64.5%, un valor predictivo positivo de 89%, y un valor predictivo negativo de 90.9% dando una precisión diagnóstica de 89.3% [116, 57].

Sin embargo, la facilidad de uso y las capacidades de imágenes en tiempo real hacen que la ecografía mamaria sea el método de elección para las biopsias mamarias y otros procedimientos que requieran intervenciones no invasivas. Por lo tanto, se realiza de forma rutinaria para diferenciar quistes de nódulos muy específicos. Junto con la mastografía de rayos X, la ecografía se utiliza para determinar si un tumor sólido es benigno o

maligno. Las principales aplicaciones de la ecografía mamaria son, para diferenciar entre lesiones quísticas y sólidas, para evaluar masas palpables no visibles mamográficamente, y para evaluar pacientes jóvenes y embarazadas con masas palpables sin necesidad de exposiciones prolongadas a rayos X como en una mastografía [112].

Por otro lado, existen muchos otros factores que afectan la calidad de la imagen y por ende la capacidad del médico durante la visualización de una lesión. Estos incluyen el diseño del transductor como lo es la frecuencia, ancho y apertura de banda; características acústicas del tejido, atenuación dependiente de la frecuencia, variación regional en la velocidad del sonido, densidad del tejido, y el ruido multiplicativo debido a la interferencia entre ondas, así como el procesamiento de las señales y la naturaleza del monitor de visualización [110].

En una imagen ecográfica, las distribuciones del ruido multiplicativo pueden ser diferentes entre sí, según el número de dispersores por célula de resolución y el sistema de captura de la imagen. El principal causante de ruido ocurre cuando en una onda ultrasónica de una sola frecuencia, es dispersada por una superficie rugosa del tejido del cuerpo para formar una serie de ondas coherentes [133]. De acuerdo a la literatura, el ruido multiplicativo, proviene del hecho en que estas ondas coherentes interfieren entre sí, por ello debido a las complejas interacciones entre varios factores, como se mencionaba anteriormente, se requieren habilidades altamente especializadas para la correcta interpretación de este tipo de imágenes [133, 84].

## 2.3 Ruido Multiplicativo

En el contexto de procesamiento de señales y procesamiento de imágenes, el ruido se refiere a las variaciones no deseadas o aleatorias que se agregan a una señal o imagen y que pueden afectar la calidad de la información original. El ruido puede ser causado por diversos factores, como interferencias electromagnéticas, errores en la adquisición de datos, condiciones ambientales variables, o limitaciones en los dispositivos de captura. En un sistema de visión artificial, las imágenes adquiridas pueden estar contaminadas por diversas fuentes como el ruido. La existencia de ruido procede principalmente del proceso de captura, almacenamiento y transformación de las imágenes. En otras palabras, el ruido se refiere a las variaciones estocásticas en los valores de los píxeles de la imagen [117, 93].

En muchas aplicaciones de procesamiento de imágenes del mundo real, los ruidos multiplicativos son habituales, por ejemplo, en imágenes de láser, microscopía, imágenes de radar de apertura sintética e imágenes médicas de ultrasonido. A diferencia del ruido

adicional, este ruido multiplicativo o *speckle* es difícil de eliminar, principalmente debido a su naturaleza proliferativa, pero también a sus distribuciones típicamente no gaussianas. A continuación se describe el modelo general de este tipo de ruido y cómo este afecta a imágenes médicas de ultrasonido.

### 2.3.1 Modelo General del Ruido Multiplicativo

En un modelo de ruido multiplicativo, una imagen grabada  $f$  definida en un rectángulo  $\alpha \subset \mathbb{R}^2$ , es la multiplicación de una imagen original  $u$  y un ruido  $n$ , suponiendo que  $u$  y  $n$  son positivos, matemáticamente se puede definir como [59, 90]:

$$f = un. \quad (2.1)$$

En el contexto de imágenes médicas de ultrasonido, la varianza en una imagen o *speckle* es un ruido granulado que existe de modo inherente, por lo que está presente en toda la imagen, considerándose un factor dominante. En una imagen ecográfica, cuando la imagen posee más observaciones de una misma región “*multilook*” el ruido resultante disminuye. Este ruido en una imagen aparece como puntos claros y oscuros, y se puede comportar de forma aleatoria. Este se modela por la intensidad observada de la señal por la escena y se expresa como [125, 16]:

$$I(t) = S(t)u(t) \quad (2.2)$$

donde  $I(t)$  es Intensidad de cada pixel observado, localizado en  $t$ ,  $S(t)$  es la reflectividad de la imagen, como  $S(t) = |r(t)|^2$  y  $u(t)$  es la intensidad desarrollada completamente por el ruido. Este producto es el resultado de la multiplicación de punto a punto dentro de una imagen, donde se asume la independencia de las variables aleatorias que corresponden a cada píxel de la imagen y la dependencia de los campos  $u$  e  $n$ . La Eq. 2.1 se puede representar para cada píxel  $(i,j)$ , en una imagen de un tamaño  $M \times N$  como [90]:

$$f_{i,j} = u_{i,j}n_{i,j}, \text{ para } 0 \leq i \leq M - 1, 0 \leq J \leq N - 1. \quad (2.3)$$

En las últimas décadas, los problemas de eliminación de ruido multiplicativo se han estudiado ampliamente, sin embargo, aún no ha sido posible determinar de forma exacta los factores que los generan. La literatura indica que el enfoque variacional dedicado al ruido multiplicativo se propuso por primera vez mediante el modelo ROF como [105]:

$$\min_u \left\{ J(u) + \lambda_1 \int_{\alpha} \frac{f}{u} + \lambda_2 \int_{\alpha} \left( \frac{f}{u} - 1 \right)^2 \right\} \quad (2.4)$$

donde los dos últimos son términos de ajuste de los datos, mientras que  $\lambda_1$  y  $\lambda_2$  son los parámetros ponderados [60, 126].

Como se mencionaba en secciones anteriores, el ruido multiplicativo es un ruido correlacionado localmente, por lo que se considera un obstáculo en la interpretación de imágenes médicas por ultrasonidos. En una imagen contaminada por este ruido, el objetivo principal es reducirlo sin eliminar características importantes. Un ejemplo de ello es en tareas donde se requiere una preservación de bordes, los cuales puedan ser importante para la detección de la masa [132]. Sin embargo, la eliminación de este ruido puede ser una tarea compleja debido a que algunas características pueden ser de la misma escala que los patrones de ruido. Para estos casos en los que se desea eliminar el ruido, los filtros adaptativos locales se han convertido en una herramienta importante. En la sección posterior se describen de manera detallada diferentes tipos de filtros que pueden ser implementados en la reducción de ruido multiplicativo en el contexto de imágenes de ultrasonido.

## 2.4 Filtros para reducción de Ruido Multiplicativo

Como se ha descrito a lo largo del capítulo anterior, el ruido multiplicativo es un problema importante a la hora de analizar las imágenes. Sin embargo, actualmente no existen formas de eliminar este tipo de ruido por completo durante la adquisición de la imagen, por lo que solo se opta por técnicas de reducción mediante filtros. La mayoría de estos métodos de filtrado comparten la técnica de tomar los valores de los píxeles de una imagen original y modificarlos en función de los valores de píxeles vecinos. Los filtros simples tratan todos los píxeles de manera similar conservando [47, 132]. Un ejemplo de ello son los filtros clásicos como los basados en *Non local means* o difusión anisotrópica. Un tipo de filtros más avanzado son filtros adaptativos como *Frost* o *Lee*.

### 2.4.1 Filtros Clásicos

El proceso filtrado de ruido en una imagen busca suprimir de manera parcial o total el ruido aparente y se considera una etapa de preprocesamiento en el análisis de una imagen. La importancia de la eliminación este ruido radica en el mejoramiento de la precisión de la información durante el reconocimiento visual humano. Lo que en consecuencia disminuye el rango de error durante el proceso de extracción de características. En el contexto de las imágenes médicas, el propósito principal durante la eliminación del ruido es reducirlo sin perder información útil como bordes. El borde de una imagen se puede definir como el

límite de una región, dado que es la frontera que separa una textura de otra. Algunos filtros con estas propiedades son:

### Filtro de Difusión Anisotrópica

Este filtro, como su nombre lo indica, se basa en la técnica de difusión anisotrópica para suavizar una imagen mientras se preservan ciertas características, como bordes y detalles importantes. A grandes rasgos, este filtro ajusta la tasa de difusión según el gradiente de la imagen. Esto significa que la difusión es más fuerte a lo largo de las direcciones en las que el gradiente es bajo (generalmente donde la imagen es relativamente uniforme) y más débil en direcciones donde el gradiente es alto (cerca de bordes o detalles de la imagen). Perona en [88] propone la siguiente ecuación diferencial parcial no lineal para el suavizado de imágenes en un dominio continuo como:

$$\begin{cases} \frac{\partial I}{\partial t} = \text{div}[c(|\nabla I|) \cdot \nabla I] \\ I(t=0) = I_0 \end{cases} \quad (2.5)$$

donde  $\nabla$  es el operador de gradiente,  $\text{div}$  el operador de divergencia,  $||$  denota la magnitud,  $c(x)$  el coeficiente de difusión, y  $I_0$  la imagen inicial, lo que da dos coeficientes de difusión

$$c(x) = \frac{1}{1 + (x/k)^2} \quad (2.6)$$

y

$$c(x) = \exp[-(x/k)^2] \quad (2.7)$$

donde  $k$  es un parámetro de magnitud de borde.

En el método de difusión anisotrópica, la magnitud del gradiente se utiliza para detectar un borde o límite de la imagen como una discontinuidad del parámetro *step* en la intensidad. Si  $|\nabla I| \gg k$ , entonces  $c|\nabla I| \rightarrow 0$  se obtiene un filtro *all-pass*; si  $|\nabla I| \ll k$ , entonces  $c|\nabla I| \rightarrow 1$  por lo que se consigue una difusión isotrópica. Una forma discreta de Eq. 2.5 viene dada por:

$$I_s^{t+\Delta t} = I_s^t + \frac{\delta t}{|\bar{\eta}_s|} \sum_{p \in \bar{\eta}_s} c(\nabla I_{s,p}^t) \nabla I_{s,p}^t \quad (2.8)$$

donde  $I_s^t$  es la imagen muestreada discretamente,  $s$  denota la posición del píxel en una cuadrícula discreta bidimensional (2-D), y  $\delta t$  es el tamaño del paso de tiempo,  $\bar{\eta}_s$  representa la vecindad espacial del píxel  $s$ ,  $|\bar{\eta}_s|$  es el número de píxeles en la ventana

(normalmente cuatro, excepto en los límites de la imagen), y  $\nabla I_{s,p}^t = I_p^t - I_s^t, \nabla_p \in \bar{\eta}_s$ . Una de las ventajas de este tipo de filtro es que se obtiene un suavizado intra-regional mientras se preservan los bordes, dado que se adapta a la estructura de la imagen [132].

### Filtro de Imagen Guiada

También conocido como filtro guiado, es un filtro de imagen explícito. La salida del filtro es localmente una transformación lineal de la imagen guía. El filtro guiado contiene propiedades de suavizado que preservan los bordes, como el filtro bilateral, pero no sufre los artefactos de inversión de gradiente. Por lo que, el filtro guiado puede utilizarse más allá del suavizado. Para el filtrado primero se define un proceso general de filtrado lineal con variantes de traslación que implica una imagen de guía  $I$ , una imagen de entrada de filtrado de entrada  $p$  y una imagen de salida  $q$ . Tanto  $I$  como  $p$  se dependen según la aplicación, y pueden ser idénticas. La salida del filtrado en un píxel  $i$  se expresa como una media ponderada:

$$q_i = \sum_j W_{ij}(I) p_j \quad (2.9)$$

donde  $i$  y  $j$  son índices de píxeles. El núcleo del filtro  $W_{ij}$  es una función de la imagen guía  $I$  e independiente de  $p$ . Este filtro es lineal con respecto a  $p$ . Como definición, el filtro guiado es un modelo lineal local entre la guía  $I$  y la salida del filtro  $q$ . Suponiendo que  $q$  es una transformación lineal de  $I$  en una ventana  $\omega_k$  centrada en el píxel  $k$ :

$$q_i = a_i I_i + b_k, \forall i \in \omega_k \quad (2.10)$$

donde  $(a_k, b_k)$  son algunos coeficientes lineales que se suponen constantes en  $\omega_k$ . El filtro implementa una ventana cuadrada de un radio  $r$ . Este modelo lineal local asegura que  $q$  tiene una arista solo si  $I$  tiene un borde, dado  $\nabla_q = a \nabla_I$ . Demostrando ser un modelo útil en la mejora de resolución de imágenes, matización de imágenes y eliminación de bordes [50].

### Filtro *Non Local means*

Es una técnica de filtrado no local que aprovecha la similitud de patrones en una imagen para preservar mejor los detalles estructurales mientras reduce el ruido. A diferencia de los métodos de filtrado tradicionales, que operan localmente utilizando ventanas móviles, este filtro busca regiones en la imagen que sean similares a la región que está siendo filtrada. A grandes rasgos, este filtro promedia ponderadamente los píxeles en función de la similitud de sus vecindarios, asignando pesos más altos a regiones más similares.

Para su definición, dada una imagen discreta con ruido  $v = v(i)|i \in I$ , el valor estimado  $NL[v](i)$ , para un píxel  $i$ , se calcula como una media ponderada de todos los píxeles de la imagen,

$$NL[v](i) = \sum_{j \in I} \omega(i, j)v(j) \quad (2.11)$$

donde la familia de pesos  $w(i, j)_j$  depende de la similitud entre los píxeles  $i$  y  $j$ , y satisfacen las condiciones  $0 \leq \omega(i, j) \leq 1$  y  $\sum_j \omega(i, j) = 1$ . La similitud entre dos píxeles  $i$  y  $j$  depende de la similitud de los vectores de escala de grises  $v(N_i)$  y  $v(N_j)$ , donde  $N_k$  denota una vecindad cuadrada de tamaño fijo y centrada en un píxel  $k$ . Esta similitud se mide como una función decreciente de la distancia euclidiana ponderada,  $\|v(N_i) - v(N_j)\|_{2,a}^2$ , donde  $a > 0$  siendo la desviación estándar del núcleo gaussiano. La aplicación de la distancia euclidiana a los vecindarios ruidosos plantea la siguiente igualdad:

$$E \|v(N_i) - v(N_j)\|_{2,a}^2 = \|u(N_i) - u(N_j)\|_{2,a}^2 + 2\sigma^2. \quad (2.12)$$

Esta igualdad muestra la robustez del algoritmo dado que en expectativa la distancia euclidiana conserva el orden de similitud entre píxeles. Los píxeles con una vecindad de intensidades de grises similar a  $v(N_i)$  tienen pesos mayores en la media. Estos pesos se definen como:

$$\omega(i, j) = \frac{1}{Z(i)} e^{-\frac{\|v(N_i) - v(N_j)\|_{2,a}^2}{h^2}} \quad (2.13)$$

donde  $Z(i)$  es la constante de normalización

$$Z(i) = \sum_j e^{-\frac{\|v(N_i) - v(N_j)\|_{2,a}^2}{h^2}} \quad (2.14)$$

y el parámetro  $h$  actúa como un grado de filtrado. Este controla el decaimiento de la función exponencial y, por tanto, el decaimiento de los pesos en función de las distancias euclidianas. Por lo que se puede definir que este filtro no solo compara la escala de grises en un solo punto, sino también la configuración geométrica en toda una vecindad. Este hecho permite una comparación más robusta que los filtros de vecindad [14].

## 2.4.2 Filtros Adaptativos

Un filtro adaptativo local es un tipo de filtro que puede ajustarse automáticamente al rendimiento de la señal de entrada para el procesamiento de señales digitales. Durante el proceso adaptativo implementa un algoritmo que utiliza la función de coste, la cual determina cómo cambiar los coeficientes del filtro de manera que se reduzca el coste

del siguiente proceso de iteración. La función de coste es el criterio para determinar el rendimiento óptimo del filtro, así como la capacidad de reducir el componente de ruido en la señal de entrada. La Figura 2.3 muestra la base teórica del diagrama de bloques de un filtro adaptativo [133].

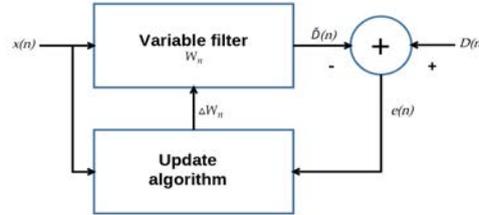


Figura 2.3 Modelo de bloques del filtro adaptativo [133].

De acuerdo a la literatura, se puede definir que los filtros adaptativos cambian su comportamiento en función de las estadísticas locales de un píxel determinado. Esto significa que los filtros adaptativos tratarán las áreas urbanas de manera diferente a las áreas rurales de una imagen. Dos filtros adaptativos muy comunes son el filtro *Frost* y el filtro *Lee*.

### Filtro *Frost*

Es un filtro adaptativo diseñado para eliminar el ruido multiplicativo de imágenes SAR [43], de forma similar al filtro *Lee*, basándose en las estadísticas locales de la imagen. La técnica de este filtro consiste en utilizar la integración no coherente para mejorar la calidad de la imagen, aumentando la relación señal-ruido con respecto a la resolución espacial. Por ejemplo, dada una imagen con ruido *speckle* modelada como [39, 133]:

$$f(x,y) = [g(x,y) \cdot n(x,y)] \cdot h(x,y) \quad (2.15)$$

donde  $f(x,y)$  corresponde a la imagen con ruido,  $g(x,y)$  es la imagen sin ruido,  $n(x,y)$  el ruido en la coordenada espacial  $(x,y)$  y  $h(x,y)$  la respuesta al impulso del sistema. La imagen libre de ruido  $g(x,y)$  se estima a partir de  $f(x,y)$  utilizando el estimador del mínimo error cuadrático medio (MMSE) y el valor de los píxeles se calcula mediante una media ponderada utilizando la vecindad píxeles en el dominio espacial. Las ponderaciones se determinan mediante las estadísticas locales de los datos:

$$\omega = k\alpha e^{-\alpha|t|} \quad (2.16)$$

donde  $\alpha \propto \sigma_f/\bar{f}$ ,  $k$  es un factor de normalización constante y  $|t|$  es la distancia geométrica espacial entre el píxel central y los píxeles vecinos [39, 133].

### Filtro *Lee*

*Lee* es un filtro para la mejora del contraste y filtrado de ruido. Este filtro es uno de los métodos más utilizados para el filtrado de ruido en imágenes SAR utilizando las características estadísticas locales de las imágenes. Para la reducción de ruido, *Lee* tiene diferentes esquemas de filtrado según las distintas propiedades del ruido en la imagen. Este filtro se caracteriza por eliminar el ruido multiplicativo preservando los bordes y las características puntuales en las imágenes. Además, está basado en un modelo lineal de ruido de *speckle* y el enfoque de diseño de error MMSE [132, 133].

Para el modelo de la Figura 2.3, *Lee* supone que la media y la varianza de la imagen con ruido son iguales a la media y la varianza de la imagen sin ruido, y que no hay correlación entre las diferentes posiciones del ruido. El modelo anterior se aproxima mediante un modelo lineal, y la estimación disponible  $\hat{g}$  de la imagen original puede obtenerse mediante la teoría de estimación del MMSE de la siguiente manera:

$$\hat{g} = E(g) + k[f - E(n)E(g)] \quad (2.17)$$

donde

$$E(g) = \frac{E(f)}{E(n)} \quad (2.18)$$

$$k = \frac{E(n)\lambda}{[E(g)]^2\sigma_n^2 + [E(n)]^2\lambda} \quad (2.19)$$

$$\lambda = \frac{\sigma_f^2 + [E(F)]^2}{\sigma_n^2 + [E(n)]^2} - [E(G)]^2 \quad (2.20)$$

$E(f)$  es la media de la imagen y  $E(n)$  es la media del ruido. Mientras que  $\sigma_f^2$  es la varianza de la imagen ruidosa y  $\sigma_n^2$  la varianza del ruido. *Lee* se caracteriza por una estimación directa basada en las propiedades estadísticas de la imagen adquirida y dado que el filtrado de una imagen mediante *Lee* no necesita realizar las transformaciones de Fourier u otras, lo hace más eficiente [66, 133]. En otras palabras, *Lee* forma una imagen de salida calculando una combinación lineal de la intensidad del píxel central en una ventana de filtro con la intensidad media de la ventana. Así, logrando un equilibrio entre el promedio directo (en regiones homogéneas) y el filtro de identidad (cuando existan bordes y características puntuales). Sin embargo, el equilibrio de este depende del coeficiente de variación dentro de la ventana móvil [132].

### Filtro SRAD

El filtro *Speckle Reducing Anisotropic Diffusion* (SRAD) es un método de filtrado adaptado a la aplicación en imágenes ultrasónicas y de radar que consiste en la difusión sensible a los bordes para imágenes con ruido multiplicativo, y del mismo modo que el filtro de difusión anisotrópica utiliza la técnica de difusión sensible a los bordes de las imágenes. Una característica importante de este filtro es que al igual que *Lee* y *Frost*, este utilizan el coeficiente de variación en el filtrado adaptativo, implementando el coeficiente instantáneo de variación, que se muestra como una función de la magnitud del gradiente local y de los operadores laplacianos, pero diferencia de estos, no solo logra preservar los bordes, sino que también los realza inhibiendo la difusión a través de los bordes y permitiendo la difusión a ambos lados del borde. SRAD utiliza el enfoque de eliminación de *speckle* basado en ecuaciones diferenciales parciales (EDP) que permite la generación de un espacio de escala de imagen, es decir, un conjunto de imágenes filtradas que varían de finas a gruesa sin sesgo debido al tamaño y la forma de la ventana de filtrado [132].

Dada una imagen de intensidad  $I_0(x, y)$  con potencia finita y sin valores cero sobre el soporte de imagen  $\Omega$ , la imagen de salida  $I(x, y, t)$  evoluciona según la siguiente EDP:

$$\begin{cases} \partial I(x, y, t) / \partial t = \text{div}[c(q)\nabla I(x, y, t)] \\ I(x, y, 0) = I_0(x, y), (\partial I(x, y, t) / \partial \vec{\eta}_s |_{\partial \Omega} = 0) \end{cases} \quad (2.21)$$

donde  $\partial \Omega$  denota el borde de  $\Omega$ ,  $\vec{\eta}$  es la normal exterior a la  $\partial \Omega$ , y

$$c(q) = \frac{1}{1 + [q^2(x, y, t) - q_0^2(t)] / [q_0^2(t)(1 + q_0^2(t))]} \quad (2.22)$$

o

$$c(q) = \exp\{-[q^2(x, y, t) - q_0^2(t)] / [q_0^2(t)(1 + q_0^2(t))]\} \quad (2.23)$$

donde  $q(x, y, t)$  es el coeficiente instantáneo de variación determinado por

$$q(x, y, t) = \sqrt{\frac{(1/2)(|\nabla I|/I)^2 - (1/4^2)(\nabla^2 I/I)^2}{[1 + (1/4)(\nabla^2 I/I)]^2}} \quad (2.24)$$

y  $q_0(t)$  es la función de escala del ruido multiplicativo. Se denomina a 2.21 la ecuación diferencial parcial SRAD. En el filtro SRAD, el coeficiente instantáneo de variación  $q(x, y, t)$  sirve como detector de bordes en imágenes con ruido, donde la función muestra valores altos en los bordes o en características de alto contraste y produce valores bajos en regiones homogéneas. Se puede observar que las expresiones del coeficiente de difusión

en Eq. 2.22 y Eq. 2.23 siguen la forma básica de Eq. 2.6 y Eq. 2.7 del filtro de difusión anisotrópica, respectivamente. Sin embargo, la modificación de estas refleja el fomento de la difusión isotrópica en regiones homogéneas de la imagen, donde  $q(x, y, t)$  fluctúa alrededor de  $q_0(t)$ . Al igual que el parámetro  $k$  en Eq. 2.6 y Eq. 2.7, la función de escala de *speckle*  $q_0(t)$  controla efectivamente la cantidad de suavizado aplicado a la imagen por el filtro. Esta se estima utilizando

$$q_0(t) = \frac{\sqrt{\text{var}[z(t)]}}{\overline{z(t)}} \quad (2.25)$$

donde  $\text{var}[z(t)]$  y  $\overline{z(t)}$  son la varianza y la media de la intensidad sobre un área homogénea en  $r$ , respectivamente.

Como se pudo observar, existen múltiples técnicas de filtrado que se pueden aplicar para reducir el ruido multiplicativo y dado que este es un factor predominante en imágenes ecográficas, la implementación de estos resultan ser una opción viable para contrarrestar la degradación y mejorar la calidad de la imagen. Por otro lado, se deduce que aunque los filtros enfocados en reducir este tipo de ruido que se caracterizan por conservar características importantes como los bordes, existen otras limitaciones en el enfoque de filtrado de ruido en general. De acuerdo a [132] se puede definir estos enfoques en 4 puntos principales:

- 1 Tamaño y forma de la ventana: los filtros pueden ser sensibles al tamaño y la forma de la ventana de filtrado. Si la ventana del filtro es demasiado grande (en comparación con la escala de interés), se producirá un exceso de suavizado y se perderán los bordes. Por otro lado, si se implementa una ventana más pequeña, esta disminuirá la capacidad de suavizado del filtro y no reducirá el ruido. En cuanto a la forma, una ventana cuadrada, como la que se aplica habitualmente, provocará, por ejemplo, el redondeo de las esquinas de los elementos rectangulares que no estén orientados en rotaciones perfectas de 90 grados.
- 2 Mejora de bordes: los filtros existentes (a excepción de algunos) no mejoran los bordes, solo inhiben el suavizado cuando están cerca. Por ejemplo, cuando alguna porción de la ventana de filtrado contiene un borde, el coeficiente de variación será alto y se inhibirá el suavizado y, por lo tanto, el ruido en las proximidades de un borde permanecerán.
- 3 Indireccionalidad: los filtros no son direccionales, por lo que en las proximidades de un borde, se excluye todo suavizado, en lugar de inhibir el suavizado en direcciones perpendiculares al borde y fomentar el suavizado en direcciones paralelas a este.

- 4 Tipo de umbral: los umbrales utilizados en los filtros mejorados, aunque motivados por argumentos estadísticos, son mejoras *ad hoc* que solo demuestran la insuficiencia de los enfoques basados en ventanas. Los umbrales “duros” que promulgan el promediado de vecindad y el filtrado de identidad en los casos extremos conducen a artefactos de manchado por el filtrado de promediado, lo que genera límites ruidosos al dejar los rasgos nítidos sin filtrar.

Con base en lo anterior es importante tomar en cuenta el tipo de filtro a utilizar para reducir el ruido multiplicativo en las ecografías, dado que este no debe afectar la calidad de la imagen y a su vez debe conservar los borde de la masa. Como el sistema propuesto requiere implementar una técnica de filtrado de ruido, este proceso es importante dado que su finalidad es mejorar la calidad de la imagen para su procesamiento en pasos posteriores. En la siguiente sección, se describe el concepto de una red neuronal convolutacional seguido de la arquitectura de una red convolutacional para la detección de objetos y una red de segmentación que se utilizaran en dicho procesamiento con la finalidad de obtener una segmentación de la masa precisa en la imagen de ultrasonido.

## 2.5 Red Neuronal Convolutacional

En el campo de la inteligencia artificial, las redes neuronales convolucionales (CNN) son una metodología utilizada frecuentemente para tareas como detección de objetos, segmentación semántica y percepción del entorno. El funcionamiento de una CNN consiste en simular el comportamiento de las neuronas humanas y se emplea para el reconocimiento de patrones. A grandes rasgos, esta consiste en múltiples capas de filtros convolucionales de una o más dimensiones que se alternan con capas de submuestreo o *max – pooling*. En términos generales, podría decirse que una CNN funciona con varias capas ocultas especializadas y con una jerarquía, esto quiere decir que las primeras capas pueden detectar líneas y curvas, especializándose hasta llegar a capas más profundas que reconocen formas complejas como un rostro o la silueta de un animal [76, 35].

Al igual que otras redes neuronales, una CNN se compone de una capa de entrada, una capa de salida y muchas capas ocultas intermedias. Estas capas intermedias realizan operaciones que alteran los datos con el objetivo de aprender las características específicas de esos mismos [76]. En una CNN, las 3 capas más frecuentes pueden definirse como:

- Capa de Convolución: esta capa hace pasar las imágenes de entrada a través de un conjunto de filtros (*Kernels*), y cada uno de estos activa determinadas características de las imágenes.
- Capa de Activación: esta permite un entrenamiento más rápido y efectivo al asignar los valores negativos a cero y mantener los valores positivos. Esto se conoce como activación, ya que solo las características activadas prosiguen su camino hacia la siguiente capa.
- Capa de Agrupación: esta capa simplifica la salida al llevar a cabo una disminución no lineal de la tasa de muestreo, lo que reduce el número de parámetros que necesita la red para aprender.

En el campo de visión por computadora, como se mencionaba una CNN, se utiliza en tareas de detección de objetos, segmentación semántica, reconocimiento de patrones y/o clasificación de imágenes. Por lo que un ámbito muy recurrente es la salud, donde las CNNs ayudan a mejorar el diagnóstico médico mediante el procesamiento de imágenes médicas, ya sea a través la detección o segmentación de objetos.

El procesamiento de imágenes mediante una CNN es un proceso que implica la extracción de características, predicción de cajas delimitadoras y clases, y posteriormente un postprocesamiento de estas predicciones. Esta tarea compleja ha dado lugar a diversas arquitecturas diseñadas para abordar el desafío. Algunos ejemplos de estos modelos son, Faster R-CNN [45], YOLO [98], SSD [70] y Mask R-CNN [49]. Por otro lado, la segmentación de imágenes mediante una CNN, consiste en asignar una etiqueta a cada píxel de la imagen, indicando a qué objeto o clase pertenece ese píxel en particular. Algunos ejemplos de CNN para este tipo de tarea son, U-Net [130], DeepLab [20], FCN [73] y SegNet [8]. A continuación se describe red neuronal Darknet-53, utilizada en esta metodología para la tarea de detección de masas sólidas en imágenes de ultrasonido a través del algoritmo YOLOv3.

### 2.5.1 Darknet-53

Es una red neuronal convolutacional utilizada en el desarrollo de YOLOv3 basada en arquitectura de red profunda, la cual se caracteriza por tener 53 capas convolucionales en su arquitectura. A diferencia de otras arquitecturas como VGG [114] o ResNet [51], Darknet-53 utiliza capas convolucionales con un tamaño de *kernel* de solo  $3 \times 3$  en lugar de *kernel* más grandes. Esto ayuda a reducir la cantidad de parámetros y a mantener una buena representación de las características de la imagen. A continuación se describe su arquitectura de acuerdo a la Figura 2.4 [99].

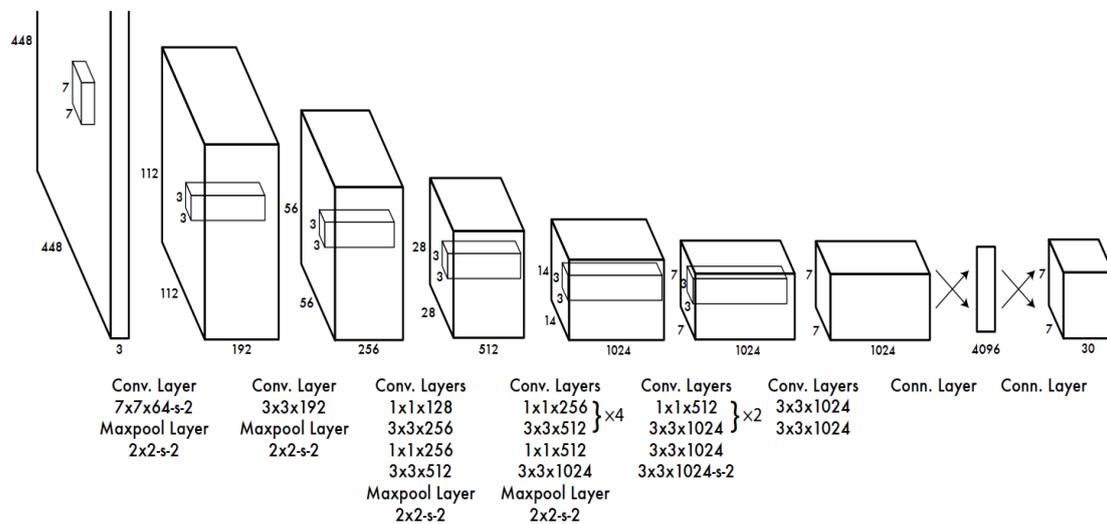


Figura 2.4 Arquitectura de red Darknet-53 [99].

De acuerdo a la Figura 2.4 la arquitectura de Darknet-53 se describe como:

- **Capa de entrada:** Darknet-53 toma una imagen de entrada de tamaño fijo y la pasa a través de una capa convolutacional inicial.
- **Bloques residuales:** A partir de la capa inicial, Darknet-53 utiliza una serie de bloques residuales. Cada bloque residual consiste en una combinación de capas convolucionales, capas de normalización (*Batch Normalization*) y activaciones no lineales (como la función de activación ReLU). Estos bloques permiten que la red aprenda representaciones más profundas y complejas de las características de la imagen.
- **Conexiones de salto (*skip connections*):** Para ayudar a mitigar el problema de la degradación del rendimiento debido a la profundidad de la red, Darknet-53 implementa conexiones de salto. Estas conexiones conectan capas anteriores con

capas posteriores de la red, permitiendo que la información se transmita directamente a través de la red sin pasar por demasiadas capas convolucionales.

- Escalas de características múltiples: Darknet-53 utiliza múltiples escalas de características en diferentes niveles de la red. Esto significa que se extraen características a diferentes resoluciones y se fusionan para obtener una representación más completa de la imagen. Esta estrategia ayuda a detectar objetos de diferentes tamaños y mejora la precisión general del modelo.
- Capa de salida: Finalmente, Darknet-53 tiene una capa de salida que produce las predicciones finales de la red. Dependiendo de la tarea específica, esta capa puede variar. En el caso de YOLOv3, se utilizan capas convolucionales adicionales para predecir los cuadros delimitadores y las clases de objetos.

Darknet-53 se caracteriza por ser implementada como una red de extracción de características en YOLOv3, la cual aprende características fundamentales de las imágenes y extraer información relevante para la detección de objetos. Cabe mencionar que esta red es una arquitectura independiente y puede utilizarse para otros fines además de la detección de objetos [97, 98].

### 2.5.2 *You Only Look Once*

YOLO del inglés *You Only Look Once* es un algoritmo de detección de objetos en tiempo real que utiliza una sola red neuronal convolutacional y plantea la detección de objetos como un problema único de regresión. Esto le permite predecir simultáneamente varios cuadros delimitadores y sus probabilidades de clase. YOLO se entrena con imágenes completas y optimiza directamente el rendimiento de detección, lo que le permite ofrecer detección de objetos precisa en tiempo real. A diferencia de otros sistemas de detección como R-CNN [46], YOLO solo necesita ver la imagen una vez para realizar la predicción utilizando una sola red neuronal.

Para la tarea de detección YOLO divide la imagen de entrada en una cuadrícula de  $S \times S$ , donde si el centro del objeto cae en una celda de cuadrícula, esa celda es responsable de detectar ese objeto. Cada celda predice los cuadros delimitadores  $B$  y las puntuaciones de “*confidence*” para ese cuadro. Estos puntajes de *confidence* reflejan que tan fiable es el modelo, es decir, que el cuadro delimitador realmente contenga el objeto y la precisión con la que se predice. La Figura 2.5 muestra el diagrama de detección de YOLO.

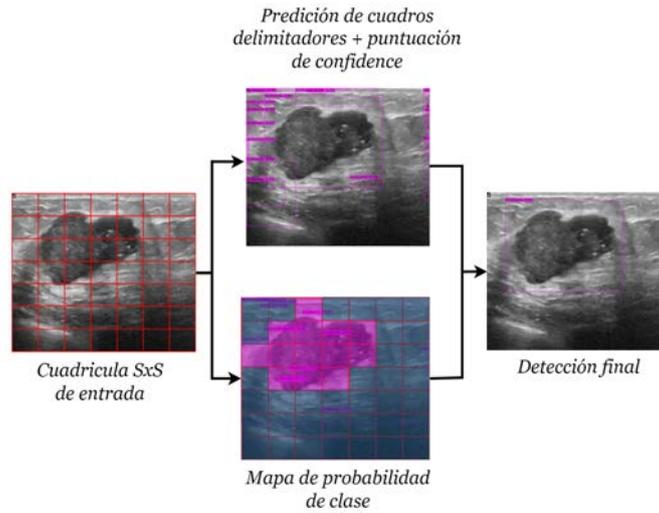


Figura 2.5 Diagrama de detección de YOLOv3 [98].

Formalmente se define *confidence* como  $Pr(Object) * IoU_{pred}^{truth}$ . Donde, si no existe un objeto en esa celda, las puntuaciones de *confidence* deben ser 0. De lo contrario, se busca que la puntuación de *confidence* sea igual a la intersección sobre la unión (*IoU*) entre el cuadro predicho y el cuadro definido (*ground truth*). Cada cuadro delimitador consta de 5 predicciones:  $x$ ,  $y$ ,  $w$ ,  $h$  y *confidence*. Las coordenadas  $(x,y)$  representan el centro del cuadro en relación con los límites de la celda de cuadrícula. El ancho ( $w$ ) y la altura ( $h$ ) se predicen en relación con toda la imagen. Finalmente, la predicción de *confidence* representa la *IoU* entre el cuadro predicho y cualquier cuadro definido.

Cada celda de la cuadrícula también predice  $C$  probabilidades de clases condicionales,  $Pr(class_i|Object)$ . Estas probabilidades están condicionadas en la celda de cuadrícula que contiene un objeto. Solo se predice un conjunto de probabilidades de clase por celda de cuadrícula, independientemente del número de cajas  $B$ . Durante la evaluación, multiplica las probabilidades de clase condicionales y las predicciones de confianza de cada cuadro individual, como se puede ver en la Eq. 2.26:

$$Pr(class|Object) \cdot Pr(Object) \cdot IoU_{pred}^{truth} = Pr(class_i) \cdot IoU_{pred}^{truth} \quad (2.26)$$

lo que da puntuaciones de *confidence* específicas de la clase para cada cuadro delimitador. Estos puntajes definen tanto la probabilidad de la clase que aparece en el cuadro como el cuadro predicho que se ajusta al objeto [98]. La arquitectura de detección está basada en 24 capas convolucionales seguidas de 2 capas completamente conectadas.

Donde las capas de convolución se van alternando para utilizar una reducción  $1 \times 1$  con la finalidad de reducir la profundidad de los mapas de características. Para la última capa de convolución, genera un tensor con forma  $(7, 7, 1024)$  y posteriormente este se aplana. Usando 2 capas completamente conectadas como una forma de regresión lineal, genera  $7 \times 7 \times 30$  parámetros para después cambiar a la forma  $(7, 7, 30)$ , es decir, 2 predicciones de cuadro de límite por ubicación. YOLO solo usa la función de activación lineal para la capa final y en todas las demás utilizan la función de activación Leaky RELU, la cual permite valores negativos, pero acentúa en los positivos [98]. En la Eq. 2.27 se puede ver su expresión matemática:

$$f(x) = \begin{cases} x \times \alpha \rightarrow x > 0 \\ x \rightarrow x \leq 0 \end{cases} \quad (2.27)$$

Durante el entrenamiento se implementa una función de pérdida para calcular la discrepancia entre las predicciones del modelo y las etiquetas de referencia (*ground truth*). Durante este procedimiento se realiza un proceso de *Back Propagation* es decir, se propagan hacia atrás los gradientes calculados a través de la función de pérdida antes mencionada, esto para ajustar los pesos de la red y mejorar las predicciones. El objetivo de este proceso es minimizar la pérdida total, que es una combinación ponderada de la pérdida de localización, la pérdida de clasificación y la pérdida de confianza. Por otro lado, durante la fase de inferencia, la función de pérdida no se utiliza directamente. En cambio, el modelo ya entrenado se aplica a nuevas imágenes para realizar predicciones de cuadros, delimitadores y clasificaciones de objetos. En resumen, la función de pérdida se utiliza durante el entrenamiento de la red neuronal para guiar el proceso de aprendizaje y mejorar las predicciones del modelo. Durante la inferencia en tiempo real, el modelo ya entrenado se utiliza para realizar predicciones sin la función de pérdida. A continuación se describe matemáticamente este proceso de la función de pérdida.

### Función de Pérdida

Como se describió anteriormente para definir el cuadro delimitador correcto dentro de las múltiples predicciones, en cada celda se hace uso de la *IoU* con respecto a la predicción definida. Esta estrategia garantiza la inferencia correcta entre las predicciones del cuadro delimitador. YOLO utiliza el error de suma cuadrada entre las predicciones y el *ground truth* para calcular la pérdida, esta función se compone de la suma de la pérdida de clasificación, pérdida de localización y pérdida de confianza.

La función pérdida de clasificación se encarga de medir la precisión en la predicción de las clases de objetos presentes en los cuadros delimitadores. Se utiliza la función de pérdida de entropía cruzada (*Cross-Entropy Loss*) para comparar las probabilidades de las clases predichas y las clases de referencia. La pérdida de clasificación penaliza la discrepancia entre las probabilidades de las clases predichas y las clases de referencia, asignando un mayor peso a las predicciones incorrectas. Durante la pérdida de clasificación, si se detecta un objeto, la pérdida en cada celda es el error al cuadrado de las probabilidades condicionales para cada clase, como se puede ver en la Eq. 2.28:

$$\sum_{i=0}^{S^2} \mathbf{1}_i^{obj} \sum_{c \in \text{classes}} (P_i(c) - \hat{p}_i(c))^2 \quad (2.28)$$

donde:

$\mathbf{1}_i^{obj} = 1$  si un objeto aparece en la celda, de lo contrario será igual a 0.

$\hat{p}_i(c)$  denota la probabilidad de clase condicional para la clase  $c$  en la celda  $i$ .

La función de pérdida de localización, se refiere a la precisión de los cuadros delimitadores (*Bounding Boxes*) predichos por el modelo en términos de su posición y tamaño. Esta utiliza la distancia de la *IoU* para medir la superposición entre el cuadro delimitador predicho y el cuadro delimitador de referencia (*ground truth*). La pérdida de localización penaliza la diferencia entre las coordenadas (centro, ancho y alto) del cuadro delimitador predicho y del cuadro delimitador de referencia, de manera proporcional a la distancia *IoU*. En resumen, esta función mide los errores en las ubicaciones y tamaños de los cuadros delimitadores previstos. Solo contando la casilla encargada de detectar el objeto, Eq. 2.29 y Eq. 2.30:

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbf{1}_{ij}^{obj} (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \quad (2.29)$$

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbf{1}_{ij}^{obj} (\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \quad (2.30)$$

donde:

$\mathbf{1}_{ij}^{obj} = 1$  si el cuadro de límite  $j$  en la celda  $i$  es responsable de detectar el objeto, de lo contrario será igual a 0.

$\lambda_{coord}$  aumentar el peso de la pérdida en la coordenada del cuadro delimitador.

Para no ponderar los errores absolutos en los cuadros grandes y pequeños por igual (es decir, un error de 2 píxeles en un cuadro grande será el mismo que en un cuadro pequeño), YOLO predice la raíz cuadrada del ancho y alto del cuadro delimitador en lugar del ancho y alto. Además, para poner más énfasis en la precisión del cuadro delimitador, multiplica la pérdida por  $\lambda_{coord}$  siendo su valor predeterminado 0.5.

La función pérdida de confianza se enfoca en la precisión de la confianza (*confidence*) asignada a cada cuadro delimitador predicho. Se utiliza la función de pérdida de entropía cruzada para comparar la confianza predicha, es decir, la probabilidad de que el cuadro delimitador contenga un objeto y la confianza de referencia, siendo 1 para cuadros delimitadores que se superponen significativamente con objetos y 0 para cuadros delimitadores de fondo. La pérdida de confianza penaliza la discrepancia entre la confianza predicha y la confianza de referencia, enfocándose en asegurar que los cuadros delimitadores se asignen adecuadamente a objetos y fondos. De acuerdo a la función de pérdida de confianza si se detecta un objeto en el cuadro, la pérdida de confianza (midiendo la objetividad del cuadro) se puede expresar como:

$$\sum_{i=0}^{S^2} \sum_{j=0}^B \mathbf{1}_{ij}^{obj} (C_i - \hat{C}_i)^2 \quad (2.31)$$

donde:

$\hat{C}_i$  es la puntuación de confianza del cuadro  $j$  en la celda  $i$ .

$\mathbf{1}_{ij}^{obj} = 1$  si el límite  $j$  en la celda  $i$  es responsable de detectar el objeto, de lo contrario será igual a 0.

De lo contrario, si no se detecta un objeto en el cuadro, la pérdida de confianza será:

$$\lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbf{1}_{ij}^{noobj} (C_i - \hat{C}_i)^2 \quad (2.32)$$

donde:

$\mathbf{1}_{ij}^{noobj}$  es el complemento de  $\mathbf{1}_{ij}^{obj}$

$\hat{C}_i$  es la puntuación de confianza del cuadro  $j$  en la celda  $i$ .

$\lambda_{noobj}$  pondera la pérdida al detectar el fondo.

Dado que la mayoría de los cuadros delimitadores no contienen ningún objeto, se puede llegar a provocar un problema de desequilibrio de clases, es decir, entrenamos el modelo

para que detecte el fondo con más frecuencia que la detección de objetos. Para remediar esto, YOLO pondera esta pérdida por un factor  $\lambda_{noobj}$  siendo 0.5 su valor predeterminado.

Para la función de pérdida total en YOLOv3 esta es una combinación ponderada de las tres pérdidas: pérdida de localización, pérdida de clasificación y pérdida de confianza [98, 99, 35, 97]. Donde estos componentes trabajan en conjunto para entrenar el modelo y optimizar la precisión de la detección de objetos, tal y su expresión general se puede ver en Eq. 2.33:

$$\begin{aligned}
& \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbf{1}_{ij}^{obj} (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \\
& + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbf{1}_{ij}^{obj} (\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \\
& \quad + \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbf{1}_{ij}^{obj} (C_i - \hat{C}_i)^2 \\
& \quad + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbf{1}_{ij}^{noobj} (C_i - \hat{C}_i)^2 \\
& \quad + \sum_{i=0}^{S^2} \mathbf{1}_i^{obj} \sum_{c \in classes} (P_i(c) - \hat{p}_i(c))^2 \tag{2.33}
\end{aligned}$$

YOLOv3 es un algoritmo de detección de objetos eficiente y preciso. Este, como se mencionaba anteriormente, utiliza la red neuronal convolutacional Darknet-53 para extraer características, realiza predicciones de cuadros, delimitadores y clases, y filtra las detecciones redundantes. Durante el entrenamiento, la red utiliza una función de pérdida que incluye la pérdida de localización, la pérdida de clasificación y la pérdida de confianza. Esta función de pérdida se utiliza para retropropagar los gradientes y ajustar los pesos de la red neuronal, con el objetivo de minimizar la pérdida total y mejorar la precisión de la detección de objetos. En cuanto a la inferencia en tiempo real, una vez que el modelo YOLOv3 ha sido entrenado, realizar predicciones de cuadros, delimitadores y clases en nuevas imágenes sin la función de pérdida. Durante la detección, el algoritmo se aplica a la imagen en una sola pasada, generando las detecciones de objetos en tiempo real. Este proceso de detección ayuda a reducir la región de interés del objeto, lo cual lo hace ideal para reducir la región de búsqueda del objeto a una sola parte con respecto a toda la imagen si se requiriera realizar algún tipo de procesamiento del objeto detectado como lo es una segmentación.

A continuación se describe el modelo de red neuronal convolutacional DeepLabv3+ para la segmentación de imágenes médicas que se aplicará en la tarea de delimitación de masas sólidas como paso posterior a la detección de la misma mediante el algoritmo ya antes visto YOLOv3. La segmentación de objetos mediante una CNN es una técnica habitual en el campo de la visión por computadora, que se utiliza para dividir una imagen en una serie de partes, propiedades o subconjuntos según el objeto de interés, con la finalidad de un mayor nivel de análisis y comprensión. En el campo de la salud, esta práctica es esencial para llevar a cabo un análisis más profundo y preciso de las imágenes médicas [12, 72]. En este contexto, la segmentación de una masa sólida en una ecografía mediante CNN no solo contribuye a una mayor precisión en el diagnóstico, sino que también juega un papel crucial en la planificación y monitorización de la masa en respuesta con la terapia propuesta por el médico.

### 2.5.3 DeepLab

Es un modelo de aprendizaje profundo para la segmentación semántica de imágenes, mediante la asignación de etiquetas de clase a nivel píxel, lo que permite una delimitación directa la estructura del objeto en la imagen. Este tipo de técnica es crucial en la captura de detalles a diferentes escalas, arrojando resultados más precisos en la segmentación de objetos y áreas en imágenes. Este modelo demuestra ser sustancial en aplicaciones médicas, desde la detección y segmentación de objetos en imágenes como masas sólidas, hasta realidad aumentada para simulaciones. Este emplea una arquitectura de redes neuronales convolucionales, basándose directamente en la red Xception [21] en su versión más reciente, DeepLabv3+.

Una característica principal de este tipo de modelo es el uso de *Dilated Convolutions*, también conocidas como convoluciones de paso amplio o *atrous*. Este tipo de convoluciones permiten expandir el campo receptivo de las capas convolucionales sin reducir el tamaño espacial de las características, dado que insertan “huecos” entre los elementos del *kernel* de la convolución con la finalidad de aumentar la cantidad de información que una neurona en la capa de convolución puede ver, esto sin la necesidad de aumentar el tamaño del *kernel* o su profundidad. Así mismo, con la finalidad de no aumentar excesivamente la cantidad de parámetros durante las convoluciones *atrous*, el modelo utiliza el módulo de *Atrous Spatial Pyramid Pooling* (ASPP) para capturar información contextual a diferentes escalas mejorando el rendimiento en la segmentación del objeto en las imágenes [17, 18].

Otro componente importante de esta red, es la implementación de un decodificador para mejorar la resolución espacial de las características. Este utiliza operaciones de *Upsampling* con la finalidad de aumentar la resolución espacial de las características extraídas mediante las convoluciones *atrous*, lo que permite generar una máscara de segmentación con la misma resolución de la imagen de entrada mejorando la calidad de segmentación. Donde está asigna a cada píxel una etiqueta que indica a qué clase pertenece en el contexto de la tarea de segmentación [17, 18, 20].

Desde su presentación en 2014, la arquitectura DeepLab ha ido evolucionando. Por ejemplo, DeepLabv1, la primera versión, utilizo convoluciones *atrous* para extraer información contextual, siendo uno de los primeros enfoques en utilizar esta técnica en la segmentación semántica [17]. DeepLabv2, por otro lado, introducido en 2016, mejoró su precisión al incorporar convoluciones *atrous* espacialmente separables [18]. DeepLabv3, presentado en 2017, por su parte introdujo el uso de ASPP, lo cual permitió extraer información contextual a múltiples escalas espaciales de manera más efectiva [19]. Y por último DeepLabv3+ presentada en 2018, siendo la versión más reciente, combina la arquitectura de DeepLabv3 con un decodificador para mejorar tanto la precisión y como la resolución de las segmentaciones [20]. A continuación se detalla más ampliamente esta última versión del modelo.

### DeepLabv3+

Como se mencionaba anteriormente, esta arquitectura es una evolución de las versiones anteriores de DeepLab y de acuerdo a la literatura, demuestra ser más efectiva en tareas de segmentación semántica en una variedad de aplicaciones, como la segmentación de objetos en imágenes médicas [20]. DeepLabv3+ combina el módulo de *Atrous Spatial Pyramid Pooling* y la estructura codificador-decodificador [65], que se utilizan en redes neuronales profundas para tareas de segmentación semántica. Donde el primero codifica información contextual multiescala sondeando las características entrantes con filtros u operaciones de *pooling* a múltiples velocidades y múltiples campos de visión, mientras que el segundo captura los límites del objeto más nítidos, recuperando gradualmente la información espacial como se puede ver en la Figura 2.6.

De acuerdo a la Figura 2.6 se deduce que la mejora propuesta puede controlar arbitrariamente la resolución de las características extraídas del codificador mediante la convolución *atrous* equilibrando la precisión y el tiempo de ejecución durante el proceso de segmentación. A grandes rasgos se puede decir que DeepLabv3+ amplía el modelo original DeepLabv3 añadiendo un módulo decodificador para refinar los

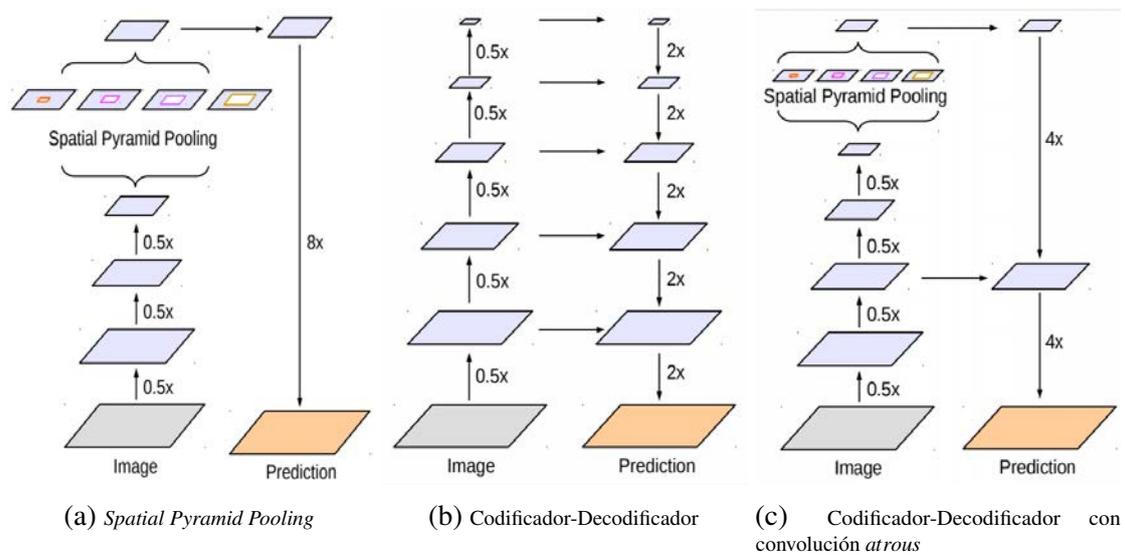


Figura 2.6 Mejora propuesta por DeepLabv3+ [20].

resultados de la segmentación, especialmente a lo largo de los límites del objeto. Así mismo implementa el modelo Xception a la tarea de segmentación y aplica la convolución *atrous* tanto a los módulos ASPP como al módulo decodificador, dando como resultado una red codificador-decodificador más rápida y potente que sus versiones anteriores. A continuación se describe de manera general el contexto de los componentes antes mencionados.

**Convolución Atrófica:** si bien el uso de CNN en tareas de segmentación semántica abordan de forma sencilla y parcialmente precisa la segmentación de objetos, se ha demostrado que la combinación repetida de *max – pooling* y *downsampling (striding)* en capas consecutivas reduce significativamente la resolución espacial de los mapas de características resultantes, normalmente por un factor de 32 en cada dirección en las CNN. Es decir, que da como resultado mapas de características con una resolución espacial significativamente reducida [73, 18]. Por ello, Deeplabv3+ propone eliminar el operador de *downsampling* de las últimas capas de *max – pooling* de las CNN y en su lugar realizar un *upsampling* de los *kernels* en las capas convolucionales subsiguientes, lo que resulta en mapas de características con una tasa de muestreo más alta. Esta propuesta consiste en insertar agujeros (*trous* en francés) entre las tomas de filtro distintas de cero. Esta técnica utilizada originalmente en el procesamiento de señales, desarrollada para el cálculo eficiente de la transformada de *Wavelet* sin decimar en un esquema, se denomina algoritmo *atrous* [52]. El autor implementa el término Convolución Atrófica o *atrous* como abreviatura de convolución con filtros sobremuestreados [18].

De acuerdo a [18], en la práctica, se pueden recuperar mapas de características de resolución completa mediante la combinación de convolución *atrous*, dado que calcula mapas de características más densos, seguido de una interpolación bilineal simple de las respuestas de características al tamaño original de la imagen. Por lo que se considera una alternativa sencilla pero potente al uso de capas deconvolucionales o convolución transpuesta [82, 73] en tareas de predicción más densas. Otra ventaja de esta técnica propuesta es, que en comparación con la convolución regular con filtros más grandes, permite ampliar eficazmente el campo de visión de los filtros sin aumentar el número de parámetros ni la cantidad de cálculo, lo que resulta factible para la aplicación en arquitecturas de hardware poco robustas. En resumen, esta técnica permite calcular las respuestas de cualquier capa a cualquier resolución deseable y puede aplicarse tanto en *post-hoc*, una vez que la red ha sido entrenada, como durante el entrenamiento [18, 20]. Para su mejor comprensión se plantea el siguiente ejemplo: considérese primero las señales unidimensionales, la salida  $y[i]$  de la convolución *atrous* de una señal de entrada unidimensional  $x[i]$  con un filtro  $w[k]$  de longitud  $k$  que se define como:

$$y[i] = \sum_{k=1}^k x[i + r \cdot k] w[k]. \quad (2.34)$$

Donde el parámetro de *rate*  $r$  corresponde al *stride* con la que se muestrea la señal de entrada. La convolución estándar es un caso especial para la *rate*  $r = 1$ , véase Figura 2.7 donde se muestra el algoritmo en 1-D. La Figura 2.7(a) muestra la extracción de características dispersas con una convolución estándar sobre un mapa de características de entrada de baja resolución. Mientras que la Figura 2.7(b) muestra la extracción de características densas con convolución *atrous* con *rate*  $r = 2$ , aplicada a un mapa de características de entrada de alta resolución.

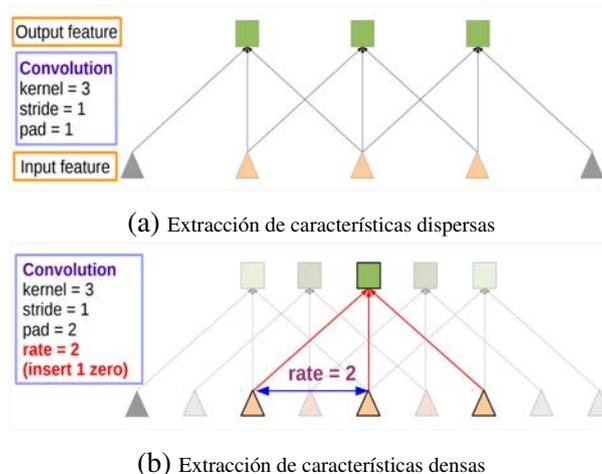


Figura 2.7 Convolución *atrous* en 1-D [18].

La Figura 2.8 muestra el funcionamiento del algoritmo, pero ahora en 2-D. Donde dada una imagen, se supone que primero se realiza una operación de *downsampling* que reduce la resolución en un factor de 2 y, a continuación, se efectúa una convolución con un *kernel*, en este caso, la derivada vertical de Gauss. Por lo que si se implementa el mapa de características resultante en las coordenadas de la imagen original, solo se obtienen respuestas en  $\frac{1}{4}$  de las posiciones de la imagen. En el cuadro superior de la figura se muestra la extracción de rasgos dispersos con convolución estándar en un mapa de características de entrada de baja resolución. Y el cuadro inferior muestra la extracción de rasgos densos con convolución *atrous* con *rate*  $r = 2$ , aplicada a un mapa de características de entrada de alta resolución.

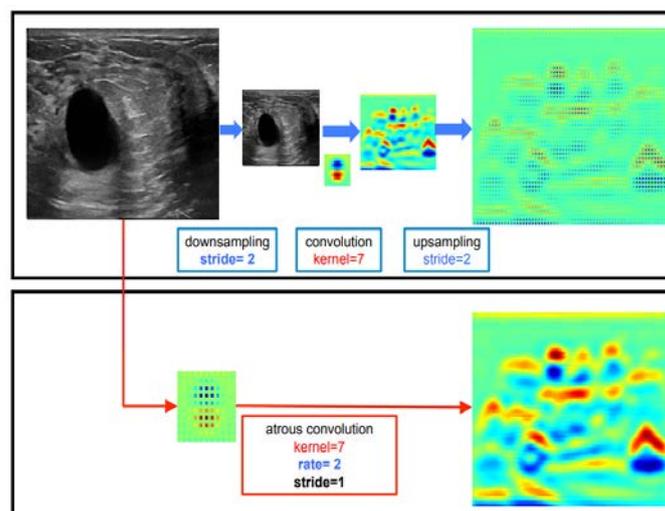


Figura 2.8 Convolución *atrous* en 2-D [18].

De acuerdo a lo anterior se puede deducir que si se aplica una convolución a la imagen de resolución completa con un filtro “*con agujeros*” de la convolución *atrous* se puede calcular las respuestas en todas las posiciones de la imagen, en el que se aumenta la muestra del filtro original en un factor de 2 y se introduce ceros entre los valores del filtro. Aunque el tamaño efectivo del filtro aumenta, solo se toma en cuenta los valores de filtro distintos de cero, por lo que tanto el número de parámetros de filtro como el número de operaciones por posición no cambian. El esquema resultante permite controlar la resolución espacial de las respuestas de las características de la red neuronal. En el contexto de las CNN, se puede utilizar la convolución *atrous* en una cadena de capas, lo que permite calcular las respuestas finales de la red con una resolución arbitrariamente alta [18].

Como se mencionaba, la convolución *atrous* también permite ampliar arbitrariamente el campo de visión de los filtros en cualquier capa de la red. Las CNN más modernas suelen emplear *kernels* espacialmente pequeños (normalmente de  $3 \times 3$ ) para reducir tanto el cálculo como el número de parámetros. La Figura 2.9 muestra el funcionamiento de la convolución *atrous* donde la convolución estándar corresponde a la convolución *atrous* con  $rate = 1$ . Se puede observar que el empleo de un valor elevado de la tasa de atrocidad (*rate*) amplía el campo de visión del modelo, lo que permite codificar objetos a múltiples escalas.

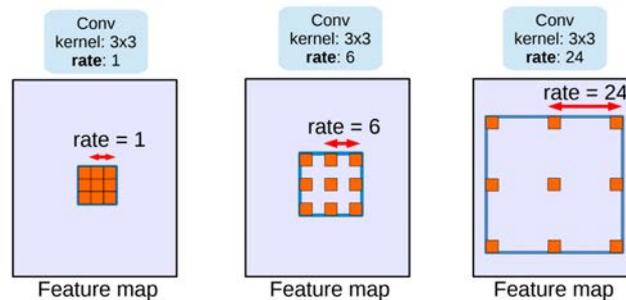


Figura 2.9 Convolución *atrous* con tamaño de *kernel*  $3 \times 3$  y diferentes *rate* [20].

La convolución *atrous* con *rate*  $r$  introduce  $r - 1$  ceros entre valores de filtro consecutivos, lo que amplía el tamaño del núcleo de un filtro  $k \times k$  a  $k_e = k + (k - 1)(r - 1)$  sin aumentar el número de parámetros ni la cantidad de cálculo. De este modo, ofrece un mecanismo eficaz para controlar el campo de visión y encuentra el mejor equilibrio entre localización precisa (campo de visión pequeño) y asimilación del contexto (campo de visión grande) [17, 18, 20].

***Spatial Pyramid Pooling:*** en este contexto, un término importante es *Image Pyramid* que se refiere a un mismo modelo, normalmente con pesos compartidos, el cual se aplica a entradas multiescala. Donde las respuestas de las entradas a pequeña escala codifican el contexto de largo alcance, mientras que las entradas a gran escala conservan los detalles de los objetos pequeños. Un ejemplo es el modelo Segnet [8], que transforma la imagen de entrada mediante una pirámide laplaciana, alimentan cada entrada de escala a una CNN y fusionan los mapas de características de todas las escalas. Sin embargo, la principal desventaja de este tipo de modelos es que no se adapta bien a redes más grandes y profundas, como por ejemplo ResNet [51] o Dilated CNN [131], debido a la limitada memoria de la GPU, por lo que solo suele aplicarse durante la etapa de inferencia [18, 19].

Un modelo ASPP como su nombre lo indica, emplea la agrupación de pirámides espaciales [48] para capturar esta información en diferentes rangos. El modelo DeepLabv2 propuso la aplicación de ASPP, en las capas de convolución atrófica (*atrous*) paralelas con diferentes velocidades, con la finalidad de capturar información multiescala [19]. Por otro lado, DeepLabv3+ explora principalmente este tipo de convolución atrófica como módulo contextual y herramienta para la agrupación de pirámides espaciales, incluyendo varias convoluciones *atrous* en paralelo en el módulo ASPP. Un detalle importante es que los módulos ASPP se pueden aplicar directamente sobre los mapas de características en lugar de sobre los mapas de confianza. De igual manera, para capturar mejor el contexto global, DeepLabv3+ propone aumentar el ASPP con características a nivel de imagen, de forma similar a trabajos como [71] y [136].

***Depthwise Separable Convolution:*** las convoluciones *depthwise* separable son una técnica de operación de convolución utilizada para reducir el coste computacional y el número de parámetros manteniendo un rendimiento óptimo. Debido a que se utilizan convoluciones separadas en profundidad, que requieren menos parámetros en comparación con las convoluciones tradicionales, lo que hace que las redes sean más ligeras y utilicen menos memoria. Esta técnica es particularmente útil en aplicaciones de visión por computadora, donde la eficiencia computacional es crucial, por lo que se ha adoptado en muchos modelos donde se requiere un proceso acelerado tanto en entrenamiento como inferencia. Un claro ejemplo son las aplicaciones en tiempo real y en dispositivos con recursos limitados [53, 20]. Así mismo, dado que este tipo de modelos manejan un menor número de recursos, el riesgo de *overfitting* es menor. Una convolución *depthwise* separable constan de dos etapas [21, 53]:

- 1 Convolución en profundidad (*Depthwise Convolution*): En se aplica una convolución separada a cada canal de entrada (profundidad) de la imagen. En otras palabras, para cada canal de entrada, se utiliza un *kernel* independiente. Esto significa que cada canal de entrada es convolucionado con su propio *kernel*, y el resultado se combina en la salida.
- 2 Convolución en punto (*Pointwise Convolution*): Después de la convolución en profundidad, se aplica una convolución de punto, que es una convolución  $1 \times 1$  (un solo píxel de ancho y alto). Esto se hace para combinar los resultados de las convoluciones en profundidad y crear una representación final de la imagen.

DeepLabv3+ define a la convolución resultante como convolución separable *atrous* (*Atrous Separable Convolution*), la cual tiene como finalidad reducir significativamente la complejidad computacional del modelo, manteniendo un rendimiento similar y en algunos casos mejor. A continuación se describe la arquitectura general propuesta por DeepLabv3+

**Encoder-Decoder:** las redes codificador-decodificador (*Encoder – Decoder*), son un tipo de arquitectura de red neuronal utilizada en visión por computadora y otras tareas de procesamiento de imágenes. Estas redes están diseñadas para aprender representaciones eficientes de datos, especialmente útiles en tareas de reconstrucción, segmentación y generación de imágenes. Generalmente, este tipo de modelo consta de dos partes [18, 20, 111]:

- **Codificador:** donde la dimensión espacial de los mapas de características se reduce gradualmente y, por tanto, la información de mayor alcance se captura más fácilmente en la salida del codificador más profundo. En otras palabras, este módulo codificador al reducir gradualmente los mapas de características captura mayor información semántica. Esta parte de la red se encarga de transformar la entrada, que suele ser una imagen de grandes dimensiones, en una representación de características menores. El objetivo del codificador es capturar las características más importantes y relevantes de la entrada. Por lo general, consta de capas convolucionales y de reducción de dimensionalidad, como capas de *pooling* o capas de convolución con *stride*. La salida del codificador es una representación de características de menor dimensión, a menudo llamado espacio de características latentes. Este “*espacio*” contiene una representación comprimida de la entrada original, donde cada dimensión puede considerarse como una característica relevante aprendida por la red.
- **Decodificador:** donde los detalles del objeto y la dimensión espacial se recuperan gradualmente. Este se encarga de reconstruir la entrada original a partir de la representación de características latentes. El decodificador toma la representación de características del espacio latente y la transforma de nuevo en una imagen o una salida de alta dimensionalidad que se asemeje a la entrada original. El decodificador suele estar compuesto por capas deconvolucionales y/o capas de aumento de dimensionalidad.

De acuerdo a la literatura, las redes codificador-decodificador son una arquitectura fundamental en tareas de visión por computadora y se han aplicado en tareas como clasificación de imágenes [63], aumento de resolución [104], segmentación de imágenes [130], detección de objetos [69] entre otras. Partiendo de esta idea, DeepLabv3+ utiliza

la red Xception en lugar de ResNet como módulo codificador y añade un módulo decodificador sencillo para obtener segmentaciones más nítidas.

### Arquitectura Codificador-Decodificador con Convoluciones Atróficas

A continuación se presenta la arquitectura general de la red DeepLabv3+:

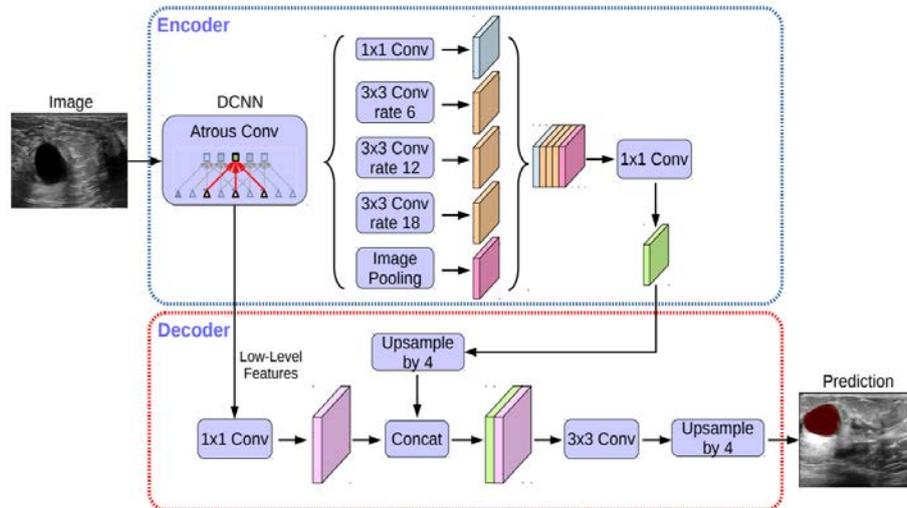


Figura 2.10 Arquitectura general de DeepLabv3+ [20].

De acuerdo a la Figura 2.10 DeepLabv3+ propone un modelo *Encoder-Decoder with Atrous Convolution* como arquitectura, el cual consta de 3 partes:

#### 1. *Depthwise Separable Convolution*

De acuerdo a [1] en la implementación de una convolución separable en profundidad, la convolución *atrous* se combina con la convolución *depthwise* como se puede ver en la Figura 2.11.

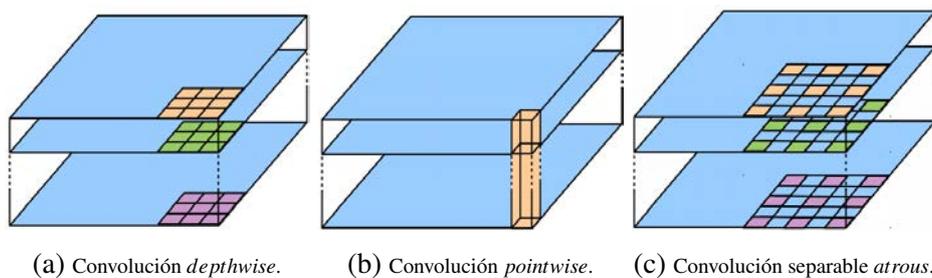


Figura 2.11 *Atrous Separable Convolution* [20].

Donde esta convolución *depthwise* de  $3 \times 3$  descompone una convolución estándar en una convolución en profundidad aplicando un único filtro para cada canal de entrada, Figura 2.11a y una convolución puntual combinando las salidas de la convolución en profundidad a través de los canales, Figura 2.11a. En concreto este modelo explora la convolución separable *atrous* con  $rate = 2$ , que es la combinación de la convolución *atrous* y la convolución *depthwise*, Figura 2.11c.

## 2. Codificador

Como se ha descrito en párrafos anteriores, el modelo original (DeepLabv3) utiliza la convolución *atrous* para extraer característica a una resolución arbitraria mediante una CNN. Este denota el *stride* de salida (*output stride*) como la relación entre la resolución espacial de la imagen de entrada y la resolución de salida final, es decir, antes de la agrupación global o de la capa totalmente conectada. Donde, para la tarea de clasificación de imágenes, la resolución espacial de los mapas de características finales suele ser 32 veces menor que la resolución de la imagen de entrada y, por tanto, el intervalo de salida (*output stride*) es igual a 32. Mientras que para la tarea de segmentación semántica, se puede adoptar una longitud de salida igual 16 u 8 dependiendo de qué tan densa se requiera la extracción de características, eliminando la longitud de salida en los últimos bloques y aplicando la convolución *atrous* correspondiente, por ejemplo, con un  $rate = 2$  y un  $rate = 4$  a los dos últimos bloques respectivamente para una longitud de salida igual 8.

El modelo original amplía el módulo ASPP, que examina las características convolucionales a múltiples escalas, aplicando la convolución *atrous* con diferentes tamaños de *rate*, con las características a nivel de imagen. Tomando en cuenta esto, DeepLabv3+ utiliza este último mapa de características antes del *logits* en el modelo original como salida del codificador en su arquitectura como un modelo de codificador-decodificador.

### Codificador basado en el modelo Xception

A diferencia del modelo original en lugar de utilizar ResNet, DeepLabV3+ elige el modelo Xception como su red codificadora. Xception es una arquitectura de red neuronal profunda que se caracteriza por el uso de convoluciones *atrous* en lugar de convoluciones convencionales, por lo que ha demostrado resultados prometedores en el campo de la clasificación de imágenes [18, 21, 92]. DeepLabv3+ adapta este modelo a la tarea de segmentación semántica de imágenes de acuerdo a:

- 1 Modelo Xception más profundo igual que en [92] pero sin modificaciones a la estructura de la red de flujo de entrada para un cálculo rápido y mejor eficiencia sin demasiado consumo de memoria.
- 2 Todas las operaciones de *max – pooling* son reemplazadas por convolución *atrous* en profundidad con *striding*, lo que permite aplicar la convolución separable *atrous* para extraer mapas de características a una resolución arbitraria (otra opción es extender el algoritmo *atrous* a las operaciones de *max pooling*).
- 3 La normalización extra de lotes y la activación ReLU se añaden después de cada convolución *depthwise*  $3 \times 3$ , de forma similar al diseño de MobileNet [53].

En resumen, en DeepLabv3+ utiliza el modelo Xception como codificador para extraer características de la imagen de entrada, y luego combina convoluciones *atrous*, el módulo ASPP y un decodificador para producir segmentaciones semánticas precisas. El uso de Xception como codificador en lugar de ResNet se basa en su mejor eficacia para capturar características de manera precisa con menos parámetros que las arquitecturas convencionales.

### 3. Decodificador

En el modelo original las características del codificador se calculan normalmente con un *stride* de salida igual a 16, por tanto, las características se sobremuestran bilinealmente por un factor de 16, por lo que podría considerarse un módulo decodificador ingenuo. Sin embargo, este módulo puede no recuperar con éxito los detalles de segmentación del objeto. Por ello, DeepLabv3+ propone el siguiente módulo decodificador de la Figura 2.12:

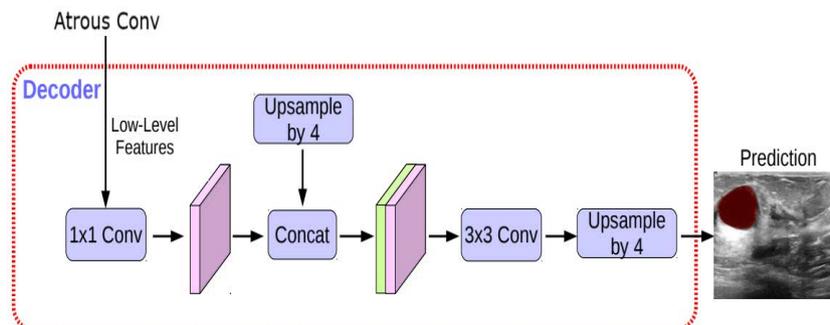


Figura 2.12 Arquitectura del decodificador de DeepLabv3+ [20].

De acuerdo a la Figura 2.10 el módulo *Encoder* codifica la información contextual multiescala aplicando una convolución *atrous* a múltiples escalas, mientras que el módulo *Decoder*, refina los resultados de la segmentación a lo largo de los límites de los objetos. Esta arquitectura codificador-decodificador mejora los resultados obtenidos con el modelo original, dado que, en primer lugar, las características del codificador se sobremuestran bilinealmente por un factor de 4 y, a continuación, se concatenan con las correspondientes características de bajo nivel de la red principal que tienen la misma resolución espacial (por ejemplo, *Conv2* antes del *striding* en ResNet-101). Luego aplica otra convolución  $1 \times 1$  en las características de bajo nivel para reducir el número de canales, dado que estas suelen contener un gran número de canales de entre 256 y 512, lo que puede ser perjudicial al momento de superponerse a las características importantes del codificador de solo 256 canales en DeepLabv3+ dificultando su entrenamiento. Tras la concatenación, se aplican unas cuantas convoluciones  $3 \times 3$  para refinar las características, seguidas de otro muestreo bilineal simple por un factor de 4. Donde a grandes rasgos el proceso de segmentación mediante DeepLabv3+ se puede definir como:

- 1 El modelo toma la imagen de entrada y utiliza como *encoder* a la arquitectura de red neuronal profunda (DCNN) Xception en lugar de ResNet como en el modelo original, reduciendo el número de parámetros y la carga computacional.
- 2 Al igual que el modelo original, DeepLabv3+ combina la DCNN con convoluciones *atrous* en el codificador, lo que permite capturar información contextual a diferentes escalas sin reducir la resolución espacial de las características. Este proceso es importante dado que ayuda a capturar mejor las características relevantes en la imagen.
- 3 DeepLabv3+ utiliza el módulo ASPP en el codificador, el cual se basa en convoluciones *atrous* con diferentes tasas de dilatación. Esto permite capturar información contextual a múltiples escalas y es fundamental para la precisión de la segmentación.
- 4 En el decodificador DeepLabv3+ incluye un módulo de agrupación piramidal espacial y operaciones de *upsampling*. Esto ayuda a restaurar la resolución espacial y los detalles en las predicciones. Es decir, este módulo se encarga de combinar información a diferentes escalas espaciales antes de realizar el *upsampling*, lo que contribuye a obtener predicciones más precisas.
- 5 Por último, DeepLabv3+ mantiene conexiones residuales entre el codificador y el decodificador, lo que ayuda a preservar la información relevante.

Como se pudo observar, aunque este modelo utiliza a Xception como modelo decodificador, sigue implementando la misma arquitectura general del modelo original. Sin embargo, la principal diferencia radica en ese mismo modelo y cómo este se combina con el decodificador para realizar la segmentación. DeepLabv3+ toma el enfoque central de DeepLabv3 al utilizar convoluciones *atrous* y el módulo ASPP para capturar información contextual, pero mejora significativamente el decodificador al incorporar un módulo de agrupación piramidal espacial y operaciones de *upsampling* dando como resultados segmentaciones más precisas.

Basándose en lo anterior, se propone combinar el modelo de segmentación DeepLabv3+ con la red de detección de objetos YOLOv3 para la tarea de detección de masas sólidas en imágenes de ecografía, donde este toma como único punto de delimitación la región previamente delimitada por YOLO, evitando que este se confunda y delimite zonas erróneas al confundir el objeto de interés en este caso, una masa sólida.

## **Discusión**

Este capítulo presentó un panorama general sobre los temas relacionados con el trabajo de investigación que se desarrolla en este tema de tesis, abarcando los conceptos básicos respecto al cáncer de mama y su diagnóstico, así como sus principales desafíos como la variabilidad en las masas y el ruido multiplicativo. De igual manera se describe cada uno de los conceptos implementados en la propuesta de solución con la finalidad de tener un mejor conocimiento del funcionamiento de la misma.

# Capítulo 3

## Estado del Arte

En esta sección se presentan brevemente investigaciones que anteceden al tema de tesis, así como trabajos relacionados con la detección y delimitación de objetos en imágenes de médicas.

### 3.1 Antecedentes

Algunas de las investigaciones que anteceden a la actual propuesta de tesis son:

#### **Detección de anomalías en mamografías utilizando la red neuronal convolucional Alexnet [77]**

En este trabajo se presenta una metodología computacional que permite detectar anomalías, las cuales pueden ser un indicio de un incipiente cáncer de mama. En esta investigación, se contempló una red neuronal convolucional debido a su alto nivel de abstracción, dada su factibilidad en el procesamiento gráfico. La metodología utiliza el conjunto de datos de la base de dominio público MIAS, el cual contiene un total de 322 mamografías; así como la base de datos DDSM de 2,500 estudios, de los cuales se seleccionó una muestra de 250 imágenes divididas en cinco categorías; para la implementación, evaluación, comparación de resultados y clasificación.

La Figura 3.1 muestra la metodología propuesta, la cual se basa en dos etapas, la primera de preprocesamiento, que consiste en pasar la imagen de entrada, atravesar de una etapa de extracción de la región de interés para después realizar una conversión basada en una paleta de colores. En la segunda fase, se observa la composición de la red convolucional empleada en forma de bloques. Mientras que la Figura 3.2 muestra el modelo de red neuronal convolucional utilizado mostrando el resultado del preprocesamiento manual y la aplicación de los filtros de las capas de convolución y capas de max-pooling.

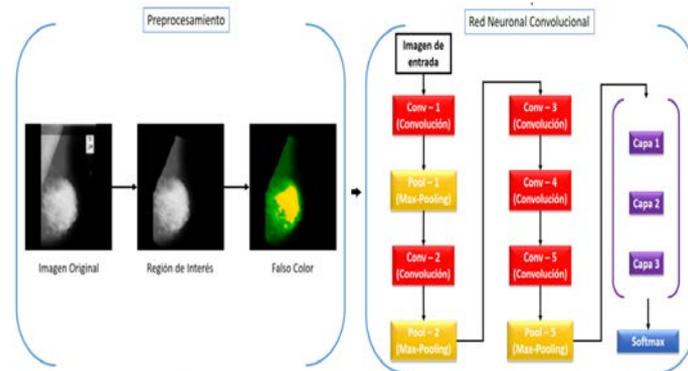


Figura 3.1 Arquitectura general de modelo propuesto [77].

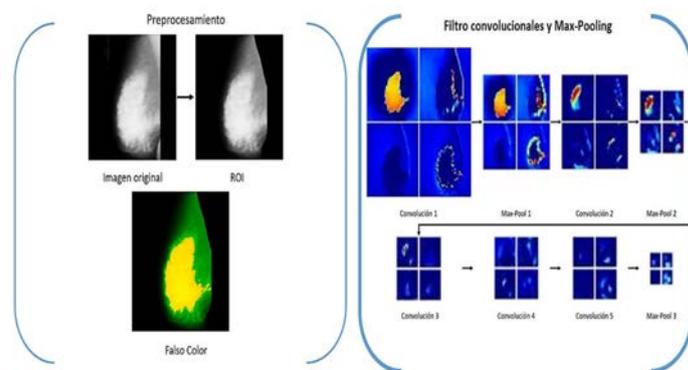


Figura 3.2 Modelo de red neuronal y aplicación de filtros [77].

De acuerdo a las Figuras 3.1 y 3.2 se puede observar que la primera fase consistió en el preprocesamiento de las imágenes con el fin de mejorar el rendimiento del clasificador; y la segunda parte, se enfocó en el diseño y caracterización de la red convolucional para la detección de anomalías en mamografías basada en la arquitectura AlexNet. Las pruebas realizadas para probar el rendimiento de la metodología propuesta, están basadas en la caracterización del modelo, pruebas del modelo empleando el banco de imágenes MIAS y rendimiento del modelo empleando la base de datos DDSM. Evaluando la sintonización de parámetros necesarios, la red neuronal convolucional logró una efectividad del 94% al clasificar a las mamografías como sanas o con alguna anomalía.

### **Segmentación no paramétrica de tejidos cerebrales mediante una arquitectura paralela de redes neuronales convolucionales [80]**

Este trabajo describe la propuesta, desarrollo e implementación de un método de segmentación no paramétrica de tejidos cerebrales, basada en una arquitectura paralela de redes neuronales convolucionales. Utilizando como modelo base utilizado para la adaptación de una arquitectura paralela U-Net. La arquitectura propuesta se compuso

de cuatro redes neuronales convolucionales, a las cuales les fue realizado un ajuste paramétrico, con la finalidad de realizar un entrenamiento individual. Cada uno de los modelos utilizados en la arquitectura logró realizar la segmentación binaria de un tejido cerebral; mediante la unificación de tejidos, obteniendo una segmentación completa de la imagen evaluada. La Figura 3.3 muestra la metodología propuesta en este trabajo.

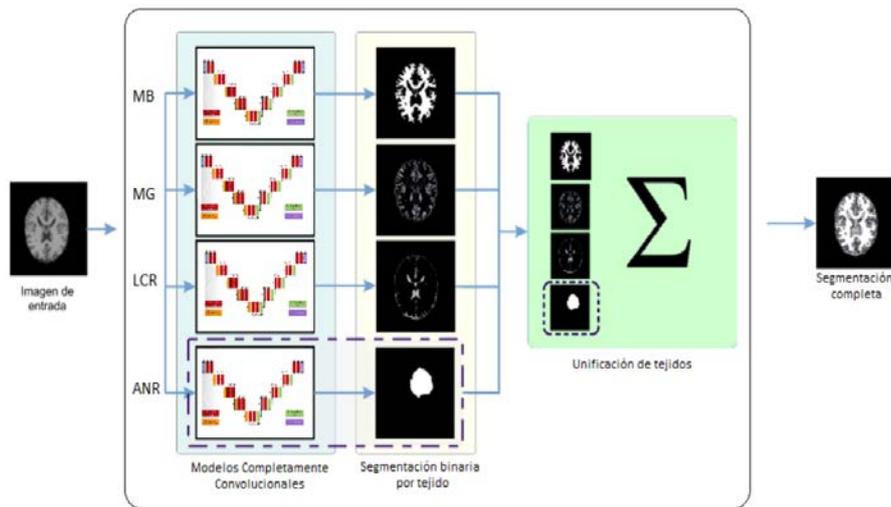


Figura 3.3 Metodología propuesta [80].

La metodología propuesta utilizó dos repositorios de imágenes: BrainWeb para entrenamiento y BraTS 2017 para validación y pruebas. Donde a los estudios de resonancia magnética utilizados para entrenamiento les fue realizado un preprocesamiento para la supresión de tejidos no blandos, seguido de una transformación a formato TIF (*Tagged Image File Format*) para el uso en los framework Keras y TensorFlow; finalmente para contar con un número aceptable de imágenes para la realización de un entrenamiento desde cero, fue realizado un aumento de datos.

De acuerdo a los experimentos realizados en promedio, los resultados obtenidos con el repositorio BrainWeb fueron de 0.9653 con la métrica Jaccard, 0.9438 para coeficiente de similitud Dice. La segmentación de acuerdo al área bajo la curva para la materia blanca fue de 0.90, en la materia gris 0.87, un 0.93 para el líquido cefalorraquídeo y 0.88 para el fondo. En cuanto al repositorio BraTS 2017 se obtuvo un 0.9692 con la métrica Jaccard, 0.9619 en el coeficiente de similitud Dice, los valores del área bajo la curva fueron de 0.93 para la materia blanca, 0.86 materia gris, 0.89 líquido cefalorraquídeo, 0.87 para la anormalidad y 0.95 para el fondo.

### Extracción de superpíxeles y fusión local mediante un algoritmo de agrupamiento difuso intuitivo [33].

Este trabajo propone un método para la extracción de superpíxeles utilizando un enfoque de agrupamiento difuso intuitivo (*Intuitionistic Fuzzy C-Means*, IFCM). Donde para el manejo de la incertidumbre, se aborda el problema del tratamiento de imágenes complejas, en específico, las imágenes de resonancia magnética cerebral. Dado que este tipo de imágenes son difíciles de tratar debido a que usualmente presentan condiciones inherentes como la presencia de ruido, bajo contraste, inhomogeneidad de la intensidad, etc.

La metodología propuesta realiza una adaptación del algoritmo IFCM para generar superpíxeles, considerando las propiedades básicas de los mismos, como la adaptación a los bordes naturales de la imagen y la homogeneidad en los superpíxeles. La inicialización del método propuesto se basó en algoritmo SLIC (*Simple Linear Iterative Clustering*) al limitar el espacio de búsqueda en regiones proporcionales al tamaño del superpíxel. La Figura 3.4 muestra el flujo de procesos para la extracción de superpíxeles con el algoritmo IFCM, en las etapas de inicialización, transformación de los datos y agrupamiento. Y la Figura 3.5 muestra el flujo de procesos para la extracción de superpíxeles con el algoritmo IFCM, en las etapas de etiquetado y extracción de superpíxeles.

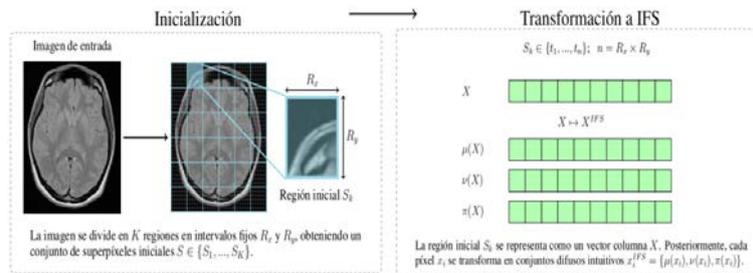


Figura 3.4 Flujo de procesos para las etapas de inicialización, transformación de los datos y agrupamiento [33].

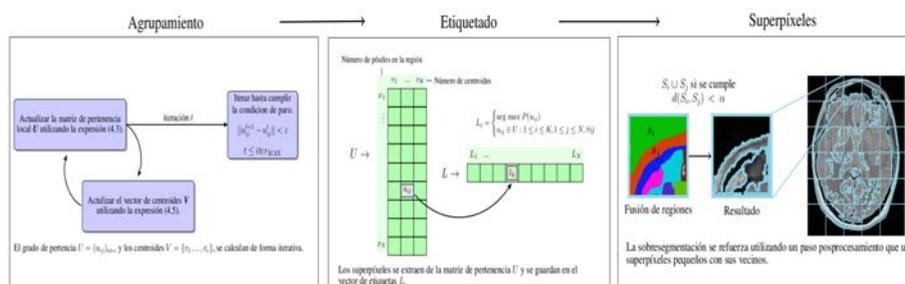


Figura 3.5 Flujo de procesos para las etapas de etiquetado y extracción de superpíxeles [33].

El método propuesto se comparó con cinco algoritmos para extraer superpíxeles con implementaciones de código abierto. Para la evaluación se utilizaron dos estudios de resonancia magnética cerebral, obtenidos del simulador BrainWeb, con diferentes porcentajes de ruido. Adicionalmente, se contempló el conjunto de datos BSDS500 para la evaluación con imágenes a color. La eficiencia del método propuesto se evaluó con respecto a las métricas de recuperación de los límites como: Rec (Recuerdo de límites), UE (Error de sub-segmentación) y EV (Variación Explicada). Como resultados en los límites Rec la metodología obtuvo un 0.998, para UE un 0.046 y para EV 0.988 con un número de superpíxeles  $K$  igual a 1,450 en el conjunto de datos de BrainWeb.

### **Agrupamiento difuso titubeante para tareas de segmentación de imágenes y reconocimiento de patrones [100]**

En este trabajo se propone un algoritmo de agrupamiento difuso mediante la incorporación de la lógica de funcionamiento de los conjuntos difusos titubeantes aplicable al reconocimiento de patrones y segmentación de imágenes. Las principales aportaciones de esta investigación fueron:

- La justificación matemática del algoritmo de agrupamiento basado en los conjuntos difusos titubeantes, el algoritmo Hesitant Fuzzy C-Means.
- Tres propuestas de mejora al algoritmo propuesto, la primera (PHFCM) para acelerar su tiempo de ejecución mediante una técnica de paralelización utilizando la biblioteca de OpenMP; la segunda (AHFCM) para dar solución al problema de la inicialización de parámetros en la agrupación, donde el modelo sugiere el uso del índice de Calinski-Harabasz para seleccionar el número óptimo de clusters en los que se agruparán los datos. Y la tercera (FAHFC) para proponer una alternativa para generar la información difusa titubeante, que según la literatura sugiere inicializar la matriz difusa titubeante y no hacer uso de otros algoritmos de agrupamiento.

Los resultados obtenidos fueron comparados con algunos algoritmos mencionados en la literatura, los cuales mostraron que el algoritmo AHFCM tiene un mejor rendimiento en comparación con los demás. De acuerdo al primer experimento, esta metodología demostró una mayor rapidez con un tiempo medio de 0.028 segundos, además de converger en pocas iteraciones. En cuanto a las métricas para evaluar la calidad del agrupamiento, el algoritmo AHFCM obtuvo un mejor rendimiento sobre los otros algoritmos, respecto a la métrica ARI tuvo una diferencia de 0.021 sobre el algoritmo FRFCM, para ACC y Pureza la diferencia sobre el algoritmo FCM-ELPSO fue de 0.009 y 0.004, respectivamente; finalmente respecto a la métrica NMI, la diferencia con FCM fue de 0.007.

## 3.2 Trabajos relacionados

### *Improved non-local self-similarity measures for effective speckle noise reduction in ultrasound images [78]*

Este trabajo presenta un método de eliminación de ruido *speckle* mediante la autosimilitud no local para la restauración de imágenes de ultrasonido. El método propuesto implementa tres etapas. Primero utiliza un filtro *Bayesian Non-Local Means* (OBNLM) en el que el parche de píxeles se representa mediante una forma vectorial para obtener una estimación preliminar de la imagen libre de ruido. Después, calcula un índice denominado índice de redundancia de cada parche de píxeles (*Redundancy Index of each Pixel Patch*) para determinar qué áreas de la imagen tienen baja redundancia. Y por último, utiliza otra forma vectorial para representar el parche de píxeles en las zonas con baja redundancia obtenidas en la segunda etapa para recalcular la salida filtrada para luego superponerse a la estimación preliminar y generar una imagen filtrada. Esta técnica hace uso de los operadores de foveation de [37] y la distancia de Pearson de [24] para calcular la similitud, pero sin la implementación en sentido de las agujas del reloj.

Para la experimentación del filtrado en imágenes de ultrasonido se seleccionaron dos imágenes ecográficas reales con lesiones de pacientes con miomas uterinos y pacientes con adenomiosis. La Tabla 3.1 muestra los resultados obtenidos respecto a las métricas *Equivalent number of looks* (ENL) [41] y *Speckle Suppression Index* (SSI) [113].

Tabla 3.1 Resultados de ENL y SSI obtenidos en imágenes de ultrasonido [78].

Metodología	ENL		SSI	
	Mioma	Adenomiosis	Mioma	Adenomiosis
Imagen original	14.9	35.0	1.0000	1.0000
Kuan [64]	20.2	61.6	0.8597	0.7537
SRAD [132]	20.1	59.2	0.8603	0.7688
NL [15]	24.3	75.7	0.7835	0.6801
ELM [87]	24.9	74.3	0.7738	0.6867
OBNLM [24]	24.8	85.6	0.7748	0.6396
BM3D [26]	24.9	84.8	0.7742	0.6427
Metodología propuesta	26.7	110.9	0.7474	0.5618

***Performance analysis of pre-cancerous mammographic image enhancement feature using non-subsampled shearlet transform [85]***

Este trabajo se propone el modelo *Non Subsampled Shearlet Transform* (NSST), una versión invariante de cambio de la transformada shearlet para eliminar el ruido en imágenes mamográficas. Esta metodología convierte una imagen codificada en una función suave en diferentes rangos de frecuencia mediante el análisis multiescalar y multidireccional que da como resultado una aproximación detallada y exacta. A grandes rasgos, el modelo convierte en información los coeficientes NSST, mientras que los coeficientes del ruido se eliminan a través umbrales adaptativos. NSST se puede definir como la combinación de filtros piramidales laplacianos no sub-muestreados (NSPL) y un filtro de corte. El cual proporciona una mejor estabilidad de desplazamiento que la transformada Shearlet discreta y una selectividad direccional más flexible que NSCT. Los componentes de alta y baja frecuencia de las imágenes de mamografía se obtienen mediante la descomposición con la pirámide laplaciana no sub-muestreada. Luego, el parámetro Shearlet y sus sub-bandas se obtienen usando filtros direccionales. El NSST se forma mapeando la expansión que consiste en el sistema afín bidimensional.

Para evaluar el rendimiento de este algoritmo, se utilizan las métricas *Mean Square Error* (MSE) [123] y la relación señal/ruido máximo (PSNR) en imágenes de las bases de datos Jansons MRI [58] y MIAS [118]. La Tabla 3.2 presenta un resumen general de los valores promedio de estas métricas.

Tabla 3.2 Promedio de PSNR y MSE obtenido para cada clase de anormalidad [85].

Base de datos	Tipo de anormalidad en la mamografía	Transformada de Shearlet		Modelo propuesto	
		PSNR	MSE	PSNR	MSE
MIAS	Normal	37.991	0.0368	41.549	0.0278
MIAS	Con clasificaciones	40.310	0.0409	44.331	0.0225
MIAS	Con asimetría	41.282	0.0396	47.498	0.0221
MIAS	Con masas espiculadas	41.244	0.0384	46.497	0.0215
MIAS	Con masas definidas por ILL	40.218	0.0370	45.024	0.0204
MIAS	Con masas circunscritas	40.413	0.0338	44.265	0.0227
Jansons MRI	Normal	39.592	0.0399	42.929	0.0266
Jansons MRI	Con lesiones anormales	38.513	0.0327	42.694	0.0265

### ***Multiplicative noise removal and blind inpainting of ultrasound images based on a new variational framework [34]***

Este artículo presenta un método para eliminar el ruido multiplicativo y marcas en la imagen (artefactos), cuando se desconoce la ubicación de los píxeles dañados por estos. Para suprimir el ruido y reparar las intensidades perdidas, este método aplica una regularización espacial a la imagen deseable, y una regularización de norma  $\gamma_0$  a los artefactos. La metodología introduce tres regularizaciones espaciales: *Total Variation* (TV), *Second Order Total Generalized Variation* (TGV) y *Fractional Order Total Variation* (FOTV) para el suavizado de imágenes. Así mismo, y debido a la no-convexidad y a la no-diferenciabilidad del problema de minimización propuesto, introduce variables auxiliares adicionales para simplificar el problema original, y después, utiliza el método de los multiplicadores de dirección alterna para resolverlo.

El conjunto de experimentos con imágenes sintéticas e imágenes médicas reales de ultrasonidos demuestran que el método propuesto es capaz de reducir el ruido multiplicativo y rellenar los píxeles faltantes. De acuerdo al autor, en comparación con otros métodos similares, la regularización TGV y FOTV no solo puede preservar los bordes y los detalles de la textura de la imagen, sino también evita el efecto escalera. Los resultados obtenidos para imágenes sintéticas fueron en PSNR y SSIM de 30.9510 y 0.9989 respectivamente. La Figura 3.6 muestra la comparación de filtrado y eliminación de artefactos en una imagen ecográfica con los modelos propuestos y una metodología de la literatura.

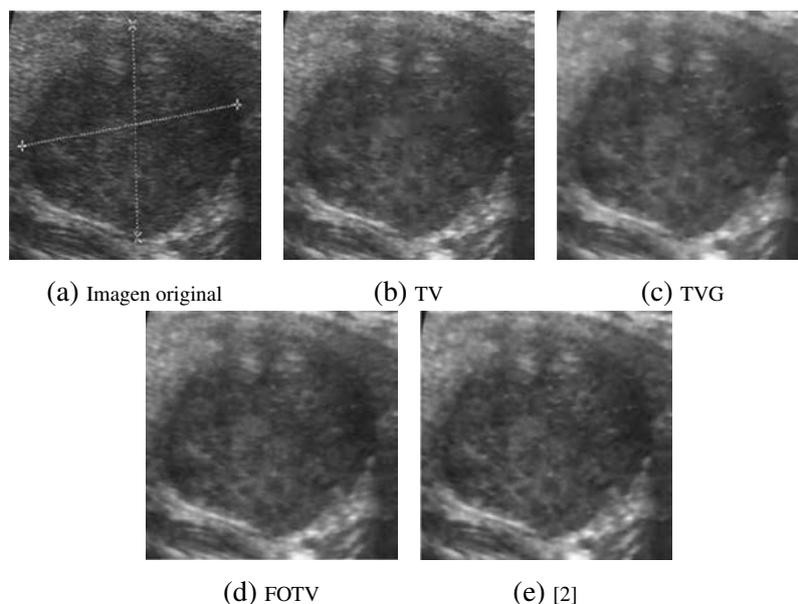


Figura 3.6 Comparación de eliminación de ruido y artefactos con las tres regularización espaciales y el trabajo de [2] [85].

### ***An optimized non-local LMMSE approach for speckle noise reduction of medical ultrasound images [95]***

En este artículo se presenta dos metodologías NLMMSE y ONLMMSE para la reducción de ruido multiplicativo en imágenes médicas de ultrasonidos (US) basadas en un estimador *Linear Minimum Mean Square Error* (LMMSE) modificado. Dado que la redundancia de datos existe de forma natural en las imágenes de ultrasonidos y que el método LMMSE original [7] se basa en el uso de estas estadísticas locales, este trabajo propone su modificación en función de explotar mejor la similitud entre píxeles donde la variación de intensidad es alta. En este sentido, utilizando las características de los estadísticos locales de segundo orden y la distancia de Pearson, el método propuesto selecciona un conjunto óptimo de píxeles similares, consiguiendo un equilibrio entre el mantenimiento de los pequeños detalles y la reducción del ruido *speckle* en diferentes regiones de las imágenes US.

El primer esquema propuesto, LMMSE no local (NLMMSE), intenta eliminar el ruido sin sentido en la ventana local y preseleccionar bloques similares en un área de búsqueda no local. Sin embargo, debido a la presencia de detalles estructurales duplicados en las imágenes de US, así como a la deformación de cada píxel individual por ruido, las estadísticas de similitud existentes en la imagen mediante un enfoque basado *enac patch-based* se utilizan para encontrar píxeles similares y, en consecuencia, mejorar el rendimiento del estimador NLMMES. Sin embargo, el resultado puede seguir conteniendo muestras inadecuadas entre las muestras seleccionadas. Por lo tanto, este trabajo propone otro esquema de NLMMSE optimizado (ONLMMSE). El cual tiene como objetivo conseguir una clase de muestras óptimas considerando el efecto del ruido *speckle* en el proceso de selección de muestras mediante otro criterio que utiliza la distancia de Pearson, que es más compatible con el modelo de ruido *speckle* presente en las imágenes US.

La Tabla 3.3 muestra la comparación de SNR y SSIM con filtros de la literatura en una imagen sintética con diferentes intensidades de ruido multiplicativo ( $\sigma$ ). Mientras que la Tabla 3.4 muestra la comparación de *Contrast to Noise Ratio* (CNR) con filtros de la literatura en una imagen de US.

Tabla 3.3 Comparación de valores de SNR y SSIM obtenidos en una imagen sintética [95].

Metodología	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.4$	$\sigma = 0.8$
Imagen original	53.11, 0.4368	39.37, 0.2503	26.01, 0.1388	14.17, 0.0662
LMMSE [25]	70.60, 0.9016	59.03, 0.7442	47.92, 0.4960	37.01, 0.2813
OBNLM [23]	79.80, 0.9574	68.31, 0.9456	55.62, 0.7488	45.49, 0.6256
NPD [124]	65.00, 0.7579	53.93, 0.5656	42.77, 0.3397	33.16, 0.2041
ADMSS [96]	54.29, 0.4782	53.15, 0.7854	39.02, 0.4584	37.26, 0.6462
UNLM [119]	78.72, 0.9586	67.16, 0.9250	53.84, 0.7791	45.48, 0.6153
HWF [36]	53.96, 0.9278	53.34, 0.8968	51.24, 0.8338	47.37, 0.7665
NLMMSE	79.85, 0.9514	68.89, 0.9452	58.15, 0.8997	49.92, 0.7832
ONLMMSE	81.90, 0.9610	69.58, 0.9531	58.64, 0.9016	49.92, 0.7832

Tabla 3.4 Comparación de valores de CNR obtenidos de 4 regiones en una imagen de US [95].

Metodología	Región 1	Región 2	Región 3	Región 4
Imagen original	1.1358	1.7531	0.3418	1.4085
LMMSE [25]	2.0602	3.7690	0.4900	3.1016
OBNLM [23]	3.0935	5.9212	0.4474	3.6105
NPD [124]	1.7523	2.8307	0.5412	2.5217
ADMSS [96]	1.9445	3.4504	0.5271	2.9480
UNLM [119]	2.9943	5.8142	0.4452	3.5016
HWF [36]	2.9004	6.0656	0.5039	3.7862
NLMMSE	3.7888	9.9413	0.5429	5.8905
ONLMMSE	3.8203	9.4884	0.7324	6.3843

### ***Medical breast ultrasound image segmentation by machine learning [129]***

Este trabajo utiliza una CNN para la segmentación de imágenes de ecografía mamaria tridimensional en cuatro clases: piel, tejido fibroglandular, masa y tejido graso. Este trabajo diseña una CNN de 8 capas para realizar el etiquetado de píxeles en tres planos ortogonales de la imagen, incluyendo capas de convolución 1-3, capas de agrupación 1-3, una capa totalmente conectada (FC), y una capa *Softmax*. En los experimentos la imagen de ultrasonido de la mama se dividió en conjunto de entrenamiento y prueba. Los resultados se compararon con los de radiólogos, con el fin de observar la similitud visual de las formas. Las métricas para la evaluación de los resultados de la segmentación incluyeron *Accuracy*, *Precision*, *Recall*, y *F1 – score*, los cuales alcanzaron más del 80%. Así mismo, se midió el índice de Jaccard el cual obtuvo un valor del 85.1%.

### ***Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model [55]***

Esta investigación propone un método automático para la segmentación de tumores en imágenes de ecografía de mamas (BUS) combinando *Dilated Fully Convolutional Network* (DFCN) con un modelo *Phase Based Active Contour* (PBAC). DFCN es una red neuronal totalmente convolucional mejorada con convolución dilatada en capas más profundas, con menos parámetros y técnicas de normalización por lotes, por lo que consigue un campo receptivo más amplio pudiendo separar los tumores del fondo. La normalización por lotes en la red permitió la aplicación de mayores tasas de aprendizaje y acelera el proceso de entrenamiento. Para resolver los problemas que plantean los límites borrosos, realiza un paso de optimización, el cual logra un resultado más preciso.

La combinación de DFCN y PBAC logra segmentar automáticamente imágenes de BUS con variaciones en el tamaño y la forma del tumor. Por último, basándose en los resultados de segmentación del método DFCN+PBAC, se extraen 460 características BI-RADS de alto rendimiento para diferenciar las lesiones benignas y malignas. Las predicciones realizadas por la DFCN son relativamente aproximadas debido a los límites borrosos, por lo que se aplica el modelo PBAC, que añade funciones de energía basadas en la región y en la fase para mejorar los resultados de la segmentación. El modelo DFCN se entrena y prueba en un conjunto de datos propio, que contiene 570 imágenes de BUS de 89 pacientes.

En la evaluación del método propuesto, se utilizó un conjunto de 170 imágenes BUS de prueba, con métricas de *Dice Similarity coefficient* (DSC), *Hausdorff Distance* (HD) y *Mean Absolute Deviation* (MAD) en comparación con el *ground truth* de las imágenes. Los resultados muestran que el método propuesto obtiene un DSC de 88.9710.01%, HD de 35.5429.70 píxeles, y un MAD de 7.676.67 píxeles.

### ***An attention-supervised full-resolution residual network for the segmentation of breast ultrasound images [94]***

Este trabajo propone el método *Attention Supervised Full-Resolution Residual Network* (ASFRRN) para segmentar los tumores de las imágenes de ecografía mamaria. El método propuesto utiliza *Global Attention Upsample* (GAU) y *Deep supervision* en una red residual de resolución completa (*full-resolution residual network* (FRRN)), donde GAU aprende a fusionar características a diferentes niveles con atención para la supervisión profunda.

Para la evaluación, este trabajo implementa dos conjuntos de datos, el conjunto A que consiste en 163 imágenes de ecografía mamaria con 53 tumores malignos y 110 benignos del Centro de Diagnóstico [127], y el conjunto de datos B que está conformado por 980 imágenes de ecografía mamaria con 595 tumores malignos y 385 benignos del Centro Oncológico de la Universidad Sun Yat-sen [120]. Los tumores de ambos conjuntos de datos fueron segmentados manualmente por médicos. Para la evaluación, se utilizó coeficiente de similitud de Dice (DSC), el índice de Jaccard (DI) y  $F1 - score$  (F1).

Como resultados se obtuvo en el conjunto de datos A, un DC de  $84.3 \pm 10.0\%$ , para DI un  $75.2 \pm 10.7\%$  y para F1 un  $84.3 \pm 10.0\%$ . Mientras que para el conjunto de datos B, el método propuesto obtuvo un DSC de  $90.7 \pm 13.0\%$ , un  $83.7 \pm 14.8\%$  en DSC y  $90.7 \pm 13.0\%$  para  $F1 - score$ . En el conjunto de datos A+B, el método obtuvo en DSC  $90.5 \pm 13.1\%$ , DI de  $83.3 \pm 14.8\%$  y F1 de  $90.5 \pm 13.1\%$ . En la Figura 3.7 se muestra la comparación de los resultados cualitativos obtenidos con la metodología propuesta y métodos del estado del arte.

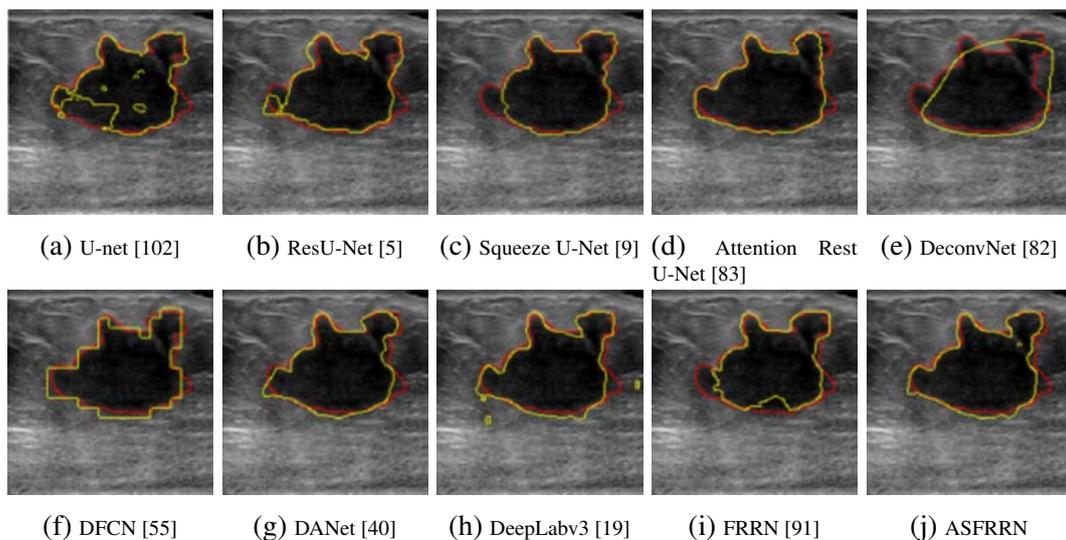


Figura 3.7 Resultado cualitativos del modelo propuesto [94].

### ***A feature transfer enabled multi-task deep learning model on medical imaging [42]***

El aprendizaje profundo multitarea (MTL) aborda la tarea de detección y segmentación, ventajas como el ahorro de costo computacional y mejora de la robustez frente al sobreajuste. Dado que los modelos profundos multitarea existentes comienzan aprendiendo cada tarea como un objetivo individual en paralelo y luego integran las tareas al final de la arquitectura con una función de coste. En este trabajo se propone una arquitectura llamada

FT-MTL-Net que es a grandes rasgos un modelo MTL habilitado por transferencia de características. Para validar la idea de la transferencia de características en la arquitectura MTL propuesta, se explora la transferencia de características de una tarea de segmentación a una tarea de clasificación de acuerdo a 2 puntos:

- 1 El objetivo de la mayoría de las aplicaciones de imágenes médicas es diagnosticar/estadificar con precisión la enfermedad, es decir, un problema de clasificación.
- 2 Aunque la segmentación y la detección están estrechamente vinculadas a la clasificación, las características utilizadas en la segmentación, la detección y la clasificación difieren.

En otras palabras, la clasificación y la detección requieren características de baja resolución para la representación abstracta, mientras que la segmentación necesita características de alta resolución para la predicción de píxeles. La arquitectura del modelo FT-MTL-Net propuesto se muestra en la Figura 3.8.

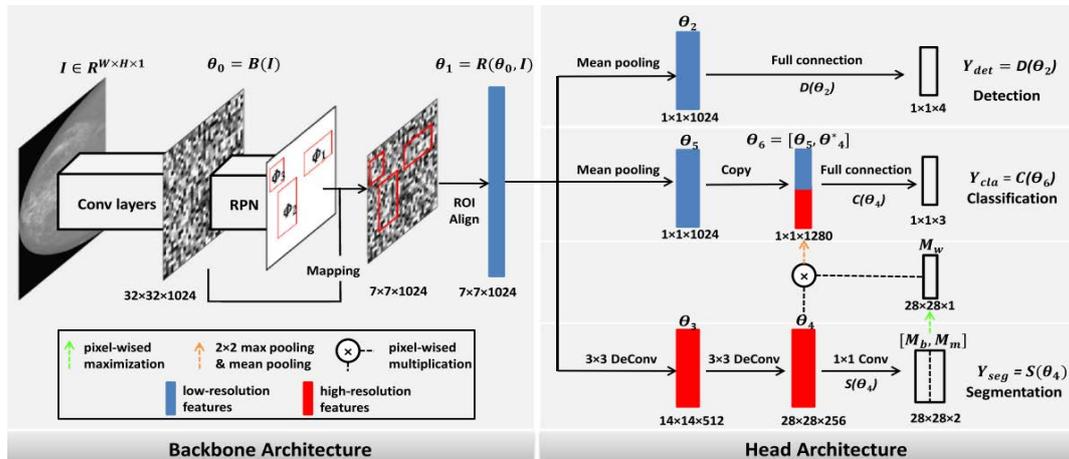


Figura 3.8 Arquitectura FT-MTL-Net propuesta[42].

Para evaluar la validez del enfoque propuesto, este se compara con modelos de la literatura, utilizando un conjunto de datos de mamografías digitales INBreast [81]. La Tabla 3.5 muestra los resultados de detección de acuerdo a las métricas *True Positive Rate* (TPR) y *false Positive Rate per Image* (FPI) y la Tabla 3.6 muestra los resultados de segmentación de acuerdo al coeficiente de similitud de Dice (DSC).

Tabla 3.5 Comparación entre el modelo FT-MTL-Net propuesto y otros métodos de detección de masas en la base de datos INBreast [42].

Metodología	Configuración	TPR	FPI
<i>Adaptive Thresholding</i> -ML [62]	Monotarea	0.84	3.67
<i>Cascaded DL-Random Forests</i> [27]	Monotarea	0.78	3.67
<i>Random Forest-CNN</i> [30]	Por etapas	0.87	5.00
<i>Multi-view Residual Net</i> [31]	Por etapas	0.9 ± 0.03	0.80
Mask-RCNN [49]	Multitarea	0.85 ± 0.07	3.67 0.85 ± 0.07
FT-MTL-Net	Multitarea	0.91 ± 0.05	3.67 0.91 ± 0.05

Tabla 3.6 Comparación entre el modelo FT-MTL-Net propuesto y otros métodos de segmentación de masas en la base de datos INBreast [42].

Metodología	Configuración	DSC
FrCNN [3]	Monotarea	92.67
<i>Random Forest-CNN</i> [30]	Por etapas	0.85 ± 0.02
Mask-RCNN [49]	Multitarea	0.79 ± 0.02
FT-MTL-Net	Multitarea	0.76 ± 0.03

### ***Deep learning in mammography images segmentation and classification: Automated CNN approach [107]***

Este artículo propone un modelo U-Net modificado preentrenado y diferentes modelos de redes de clasificación InceptionV3 [121], DenseNet121 [56], ResNet50 [51], VGG16 [114] y Mobile-NetV2 [108]; para la clasificación y segmentación de imágenes mamográficas de los conjuntos de datos MIAS [118], DDSM [67] y CBIS-DDSM en masas benignas y malignas. El marco de trabajo propuesto se basa en imágenes Mediolateral Oblique (MLO) y Cráneo Caudal (CC) para mejorar el rendimiento obtenido del sistema.

El rendimiento del diagnóstico se evalúa en términos de (IoU), *Dice Similarity Coefficients* (DSC), *Accuracy* (ACC), *Recall* (R), *Precision* (P), *Area Under the Curve* (AUC), *F1-score* (F1) y tiempo computacional (Tiempo). Para el entrenamiento se inicializa los pesos de la red ya entrenada con otro conjunto de datos y luego se implementa la técnica de aumento de datos, se para resolver las deficiencias de los datos e introducir variedad al mismo, con la finalidad de mejorar la capacidad de generalización de la red y eliminar el sobre entrenamiento. En la Tabla 3.7 se muestran los resultados obtenidos en la segmentación de imágenes utilizando este modelo propuesto en imágenes de mamografías en vista MLO.

Tabla 3.7 Resultados de segmentación basados en el modelo U-Net modificado basado en vista MLO [107].

Conjunto de datos	IoU	DSC
DDSM	92.99%	91.89%
MIAS	90.78%	89.99%
CBIS-DDSM	87.96%	88.65%

En cuanto a la clasificación del modelo U-Net y el modelo InceptionV3 logran el mejor rendimiento alcanzando un 98.87% de *Accuracy*, 98.88% de AUC, 98.98% de *Recall*, 98.79% de *Precision*, 97.99% de puntuación F1 y 1.2134 segundos de tiempo computacional en los conjuntos de datos DDSM. El marco de trabajo propuesto, al utilizar la combinación entre la vista MLO y CC, logra un mejor rendimiento que utilizando MLO solamente, donde las métricas se mejoran a 99.43% de *Accuracy*, 99.22% de AUC, 99.12% de *Recall*, 98.99% *Precision* y 98.98% de puntuación F1.

### ***Automatic suspicions lesions segmentation based on variable-size windows in mammography images [106]***

Este trabajo propone un método de umbral adaptativo basado en ventanas de tamaño variable para la segmentación automática de lesiones sospechosas en mamografías. El propósito del trabajo es estimar la ubicación de la masa y luego determinar la ubicación exacta del tejido canceroso para reducir los falsos positivos. Esta metodología implementa características a escala de grises, dada la factibilidad en algoritmos de detección de masas, especialmente en algoritmos de umbralización adaptativa. Esta característica incluye escalas grises que pueden utilizarse para encontrar el umbral adecuado para cada uno de los dos pasos del método propuesto. En el primer paso obtiene un umbral utilizando la función PDF (*Probability Density Function*), y luego, para reducir la tasa de falsos positivos y detectar la forma y la ubicación exacta de la masa, utiliza el umbral final basado en ventanas de tamaño variable.

En el trabajo de detección automática de la masa, la metodología utiliza histogramas y sus máximos relativos para calcular el umbral inicial y estimar su localización. Primero utiliza dos ventanas que contienen información alrededor de cada píxel y su tamaño varía en función del valor medio de cada imagen para preservar la información importante. Después utiliza dos ventanas más para el umbral final con el fin de descubrir la ubicación y forma exacta de la masa.

Para su evaluación, la metodología propuesta se aplicó a 170 imágenes de la base de datos MIAS. Donde obtuvo una sensibilidad del 96.7% y una tasa de falsos positivos del 0.79. La Tabla 3.8 muestra una comparación de los resultados obtenidos comparado con [54] y [6] para los siguientes tipos de masa: CIRC = Masas bien definidas/circunscritas, SPIC = Masas espiculadas, MISC = Otras masas mal definidas, ARCH = Distorsiones arquitectónicas, ASYM = Asimetría y NORM = Normal.

Tabla 3.8 Comparación del rendimiento de los métodos de segmentación en imágenes del conjunto de datos MIAS [106].

Tipo de masa	Masas / Imágenes	ATBMA [54]		DuSAT [6]		Método propuesto	
		Masas	Sensibilidad	Masas	Sensibilidad	Masas	Sensibilidad
CIRC	24 / 24	23	95.8	24	100	24	100
SPIC	19 / 19	15	78.9	16	84.9	18	94.7
ARCH	19 / 19	18	94.7	17	89.5	18	94.7
ASYM	15 / 15	14	93.3	14	93.3	15	100
MISC	14 / 15	14	93.3	15	100	14	93.3
TOTAL	89 / 92	84	91.3	86	93.5	89	96.7

### ***Study on Data Partition for Delimitation of Masses in Mammography [128]***

Este trabajo propone un sistema integrado de delimitación de masas en mamografías asistida por computadora basado en métodos de visión artificial y aprendizaje automático. En este método, las imágenes de mamografía del conjunto de datos Inbreast [81], se convierten primero en pseudocolor y luego se utilizan para entrenar y evaluar la red neuronal profunda Mask R-CNN. El enfoque de este trabajo es una comparación de los resultados obtenidos con respecto a la importancia de dividir el conjunto de datos adecuadamente, evitando el sobreajuste. Durante el entrenamiento de una CNN el enfoque más común consiste en comenzar con un conjunto de datos y dividir las imágenes en un subconjunto de entrenamiento y otro de prueba de forma aleatoria. Sin embargo, dado que a menudo hay dos o más imágenes del mismo caso en el conjunto de datos, la forma en que se divide el conjunto de datos puede tener un impacto en los resultados. Partiendo de esto, los experimentos de este trabajo se realizaron con particiones del conjunto de datos en función del caso y con partición aleatoria. En todos los casos, el conjunto de datos se dividió en 280 imágenes para el entrenamiento, 65 imágenes para validación y 65 imágenes para pruebas. Para la evaluación del rendimiento del método propuesto, las métricas utilizadas fueron, para la tarea de detección de masas: tasa de verdaderos positivos (TPR) y tasa de falsos positivos por imagen (FPPI); y para la segmentación de las masas: Índice de Similitud Dice (DSC).

Los experimentos realizados demostraron que la partición aleatoria de los datos puede producir un entrenamiento poco fiable, por lo que el conjunto de datos debe dividirse utilizando la partición por casos para obtener resultados más estables. En los resultados cuantitativos, utilizando la partición aleatoria, este método alcanza un TPR de  $0.936 \pm 0.063$  y  $0.908 \pm 0.002$  utilizando la partición por casos. En la Tabla 3.9 se muestra la comparación global de resultados entre el método propuesto y otros del estado del arte.

Tabla 3.9 Comparación del rendimiento global en la tarea de delimitación entre modelo propuesto y los modelos de la literatura en base de datos INbreast [128].

Metodología	TPR	FPPI	DSC
<i>Cascaded DL-Random Forests</i> [27]	$0.87 \pm 0.14$	0.8	n.a.
Deep MIL [22]	0.76	0.48	n.a.
SSVM [29]	n.a.	n.a.	0.88
CRF - SSVM [28]	n.a.	n.a.	$0.90 \pm 0.06$
<i>Adversarial deep FCN-CRF</i> [137]	n.a.	n.a.	0.9097
DL-CNN [74]	n.a.	n.a.	0.90
U-ResNet [135]	n.a.	n.a.	0.85
CGANs [68]	n.a.	n.a.	0.91
<i>Random Forest-CNN</i> [30]	$0.90 \pm 0.02$	1.3	$0.85 \pm 0.02$
FrCNN [3]	n.a.	n.a.	0.9269
FTMTL-Net [42]	$0.91 \pm 0.05$	1.5	$0.76 \pm 0.03$
PCMs+Mask R-CNN [79]	$0.90 \pm 0.05$	0.9	$0.88 \pm 0.10$
Método propuesto	0.909	0.769	$0.891 \pm 0.05$

### 3.3 Discusión del Estado del Arte

Los trabajos mencionados en el estado del arte abarcan diversos temas referentes a la segmentación de imágenes que son de importancia para esta investigación, entre los temas de interés se encuentran: Reducción de ruido multiplicativo en imágenes mamográficas, delimitación de regiones de interés (RIO), detección de masas mamográficas, segmentación de masas en mamografías, métricas para la evaluación de resultados y conjuntos de datos públicos de imágenes de mamografías. A continuación se presenta la Tabla 3.10 donde se muestra el resumen de cada uno de los artículos presentados y su utilidad para esta investigación.

Tabla 3.10 Resumen del estado del arte.

Título	Algoritmos o Técnica	Objetivo	Resultados / Base de datos	Utilidad
<i>Improved non-local self-similarity measures for effective speckle noise reduction in ultrasound images</i> [78]	Filtro Bayesian Non-Local Means (OBNLM) y Redundancy Index of each Pixel Patch.	Eliminación del ruido <i>speckle</i> mediante la autosimilitud no local para la restauración de imágenes de ultrasonido corruptas.	Miomas: ENL = 26.7 y SSI=0.7474. Adeniosis: ENL = 110.9 y SSI = 0.5618 / Base de datos propia.	Análisis de técnicas de eliminación de ruido mediante filtros NL y métricas cuantitativas.
<i>Performance analysis of pre-cancerous mammographic image enhancement feature using non-sampled shearlet transform</i> [85]	Filtro Non Subsampled Shearlet Transform (NSST).	Reducción de ruido en imágenes mamográficas.	MIAS: PSNR = 41.549 y MSE = 0.0278. Jasons MRI: PSNR = 42.929 y MSE=0.0266.	Análisis de técnicas de eliminación de ruido y métricas cuantitativas.
<i>Multiplicative noise removal and blind inpainting of ultrasound images based on a new variational framework</i> [34]	Total Variation (TV), Second Order Total Generalized Variation (TGV) y Fractional Order Total Variation (FOTV).	Suprimir el ruido y reparar las intensidades perdidas en una imagen ecográfica.	PSNR = 30.9510 y SSIM = 0.9989 / Base de datos propia.	Análisis de técnicas de eliminación de ruido, marcas en la mamografía y métricas cuantitativas.
<i>An optimized non-local LMMSE approach for speckle noise reduction of medical ultrasound images</i> [95]	Minimum Mean Square Error (LMMSE), LMMSE no local (NLMSE) y NLMSE optimizado (ONLMSE).	Reducción de ruido multiplicativo en imágenes médicas de ultrasonidos mediante un estimador LMMSE modificado.	NLMSE: $\sigma = 0.2$ , SNR = 68.89 y SSIM = 0.9452. ONLMSE: $\sigma = 0.2$ , SNR = 69.58 y SSIM = 0.9531 en imágenes sintéticas. Imagen US: NLMSE obtuvo un valor CNR promedio en 4 regiones de 5.04 y ONLMSE 5.10 / Base de datos propia.	Análisis de técnicas de eliminación de ruido y métricas cuantitativas.

Titulo	Algoritmos o Técnica	Objetivo	Resultados / Base de datos	Utilidad
<i>Medical breast ultrasound image segmentation by machine learning</i> [129]	Convolutional Neural Networks (CNN).	Segmentación de una imagen ecográfica 3D en cuatro clases: piel, tejido fibroglandular, masa y tejido graso.	Los resultados de la segmentación incluyeron <i>Accuracy</i> , <i>Precision</i> , <i>Recall</i> , y <i>F1 – score</i> , los cuales alcanzaron más del 80%. Para el índice de Jaccard se obtuvo un 85.1% / Base de datos propia.	Análisis de técnicas de segmentación basadas en CNN y métricas cuantitativas.
<i>Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model</i> [55]	Método automático para la segmentación de tumores en imágenes de ecografía de mamas combinando una red <i>Dilated Fully Convolutional Network</i> y un modelo <i>Phase Based Active Contour</i> .	Segmentación automática de tumores en imágenes de ecografía de mamas.	DSC de $88.97 \pm 10.01\%$ , HD de $35.54 \pm 29.70$ pixeles, y un MAD de $7.67 \pm 6.67$ pixeles / Base de datos propia.	Combinación de varias metodologías para la segmentación de tumores en imágenes de ecografía. Análisis de métricas cuantitativas.
<i>An attention-supervised full-resolution residual network for the segmentation of breast ultrasound images</i> [94]	Attention Supervised Full-Resolution Residual Network (ASFRRN).	Segmentar tumores en imágenes de ecografía mamaria.	DSC = $90.5 \pm 13.1\%$ , DI = $83.3 \pm 14.8\%$ y F1 = $90.5 \pm 13.1\%$ / Base de datos propia.	Análisis de metodología basada en una red residual para la segmentación de imágenes de ultrasonido.
<i>A feature transfer enabled multi-task deep learning model on medical imaging</i> [42]	FTMTL-Net	Detección y segmentación de objetos de imágenes médicas de ultrasonido.	Detección: TPR = $0.91 \pm 0.05\%$ y FPI = $3.67 \pm 0.91 \pm 0.0\%$ . Segmentación: DSC = $0.76 \pm 0.03\%$ / Base de datos INBreast.	Análisis de técnicas de segmentación de imágenes médicas con aprendizaje profundo multitarea.
<i>Deep learning in mammography images segmentation and classification: Automated CNN approach</i> [107]	U-Net modificado, InceptionV3, DenseNet121, ResNet50, VGG16 y Mobile-NetV2.	Clasificación y segmentación de imágenes mamográficas.	Segmentación: IoU = $92.99\%$ y DSC $91.89\%$ en DDSM. IoU = $90.78\%$ y DSC = $89.99\%$ en MIAS. IoU = $87.96\%$ y DSC = $88.65\%$ en CBIS-DDSM. Clasificación: U-Net y InceptionV3 logran el mejor rendimiento de ACC = $98.87\%$ , AUC = $98.88\%$ , R = $98.98\%$ , P = $98.79\%$ , F1 = $97.99\%$ y Tiempo = $1.2134$ sg en DDSM	Descripción general de métricas cuantitativas para la segmentación y clasificación de imágenes mamográficas.
<i>Automatic suspicious lesions segmentation based on variable-size windows in mammography images</i> [106]	Umbral Adaptativo PDF (Probability Density Function) basado en Ventanas de Tamaño Variable.	Estimar la ubicación de la masa y luego determinar la ubicación exacta del tejido canceroso para reducir los falsos positivos.	R = $96.7\%$ y FPR = $0.79$ / Base de datos MIAS.	Análisis de técnicas de detección de masas en mamografías.

Titulo	Algoritmos o Técnica	Objetivo	Resultados / Base de datos	Utilidad
<i>Study on Data Partition for Delimitation of Masses in Mammography</i> [128]	Mask R-CNN entrenado con un conjunto de datos dividido adecuadamente para evitar el sobreajuste.	Delimitación de masas en mamografías.	Partición aleatoria: TPR = $0.936 \pm 0.063$ . Partición por casos: TPR = $0.908 \pm 0.002$ . DSC global = $0.891 \pm 0.05$ / Base de datos Inbreast.	Análisis de la importancia de las técnicas de entrenamiento de una red neuronal convolucional para evitar el sobreajuste.
<i>Deep learning in mammography images segmentation and classification: Automated CNN approach</i> [107]	U-Net modificado, InceptionV3, DenseNet121, ResNet50, VGG16 y Mobile-NetV2.	Clasificación y segmentación de imágenes mamográficas.	Segmentacion: IoU = 92.99% y DSC 91.89% en DDSM. IoU = 90.78% y DSC=89.99% en MIAS. IoU = 87.96% y DSC = 88.65% en CBIS-DDSM. Clasificación: U-Net y InceptionV3 logran el mejor rendimiento de ACC = 98.87%, AUC = 98.88%, R = 98.98%, P = 98.79%, F1 = 97.99% y Tiempo = 1.2134 sg en DDSM	Descripción general de métricas cuantitativas para la segmentación y clasificación de imágenes mamográficas.
<i>Automatic suspicions lesions segmentation based on variable-size windows in mammography images</i> [106]	Umbral Adaptativo PDF (Probability Density Function) basado en Ventanas de Tamaño Variable.	Estimar la ubicación de la masa y luego determinar la ubicación exacta del tejido canceroso para reducir los falsos positivos.	R = 96.7% y FPR = 0.79 / Base de datos MIAS.	Análisis de técnicas de detección de masas en mamografías.

# Capítulo 4

## Metodología de Solución

### 4.1 Propuesta de solución

De acuerdo con el objetivo principal de esta investigación, se plantea la siguiente metodología para la segmentación de masas sólidas en imágenes de ecografía mediante un sistema de aprendizaje profundo. El cual se basa en la aplicación de un filtro de ruido multiplicativo y dos redes neuronales, donde una se utiliza para la detección de la región de interés en la ecografía (masa) y otra para la segmentación de la misma. El diagrama del sistema propuesto se puede ver en la Figura 4.1.

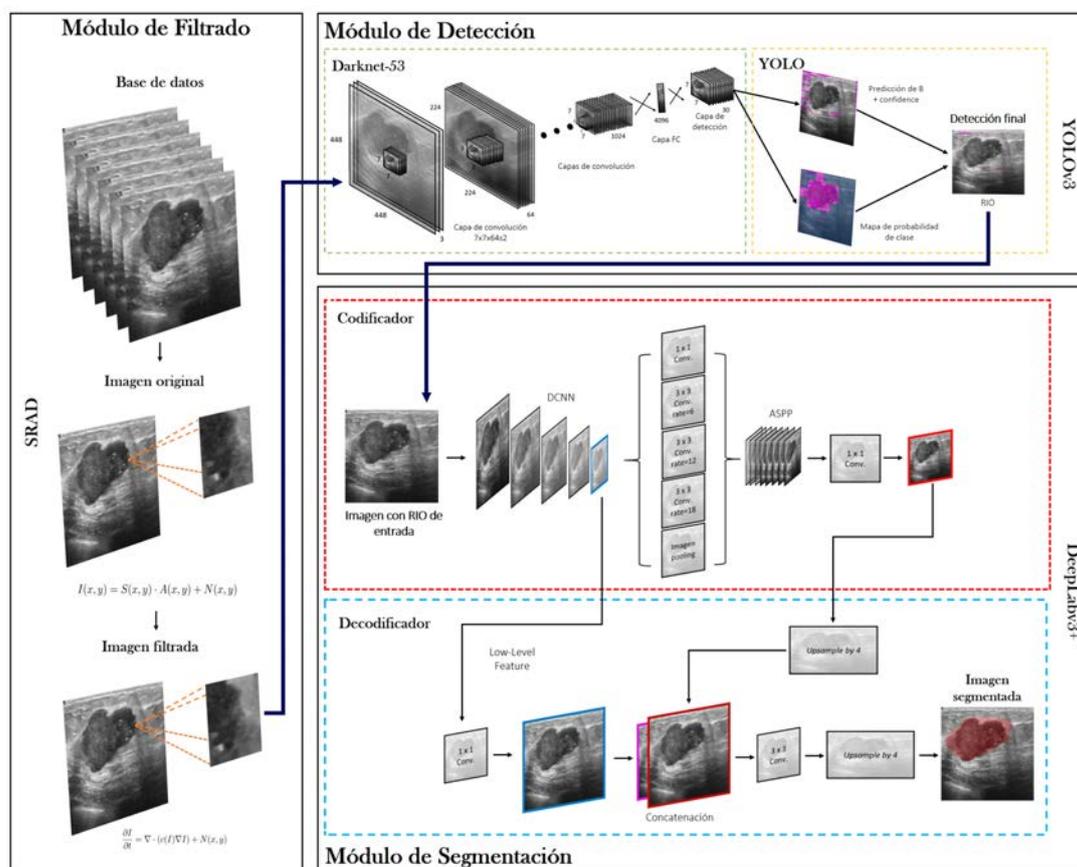


Figura 4.1 Diagrama de metodología de solución propuesta.

A continuación se describe cada módulo de manera general, así como su función en la metodología.

#### 4.1.1 Módulo de Filtrado

Este se centra en el preprocesamiento de la imagen para su mejora de calidad. Donde se busca reducir el ruido multiplicativo y resaltar los bordes de la masa mediante el filtro *Speckle Reducing Anisotropic Diffusion*. La Figura 4.2 muestra el diagrama del proceso de mejora de la imagen.

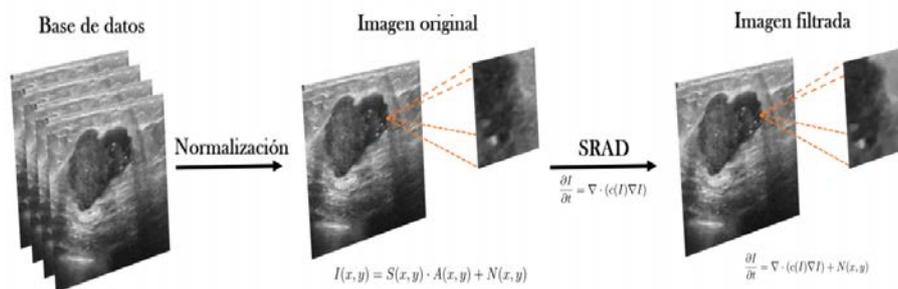


Figura 4.2 Módulo de filtrado: SRAD.

De acuerdo a la Figura 4.2 una imagen ecográfica con ruido definida como:

$$I(x, y) = S(x, y) \cdot A(x, y) + N(x, y) \quad (4.1)$$

donde  $I(x, y)$  es la imagen observada,  $S(x, y)$  es la señal original,  $A(x, y)$  es la atenuación y  $N(x, y)$  es el ruido multiplicativo; es tratada con el filtro SARD que se define como:

$$\frac{\partial I}{\partial t} = \nabla(c(I)\nabla I) \quad (4.2)$$

donde  $\frac{\partial I}{\partial t}$  representa la evolución temporal de la imagen con el filtro SRAD, y  $\nabla(c(I)\nabla I)$  la parte del filtro que controla la difusión anisotrópica. Esto devuelve una imagen filtrada definida como:

$$\frac{\partial I}{\partial t} = \nabla(c(I)\nabla I) \quad (4.3)$$

### 4.1.2 Módulo de Detección

Debido a que las masas en una ecografía son difíciles de analizar, tanto por el ruido aparente como a sus bordes poco definidos, esta investigación propone un módulo para reducir la región de búsqueda de la masa (RIO), el cual tiene como objetivo indicarle a la red de segmentación la parte de donde se ubica la masa así evitando que está segmente regiones erróneas. La Figura 4.3 muestra la red de detección utilizada en este módulo, el cual se basa en el algoritmo de detección YOLOv3.

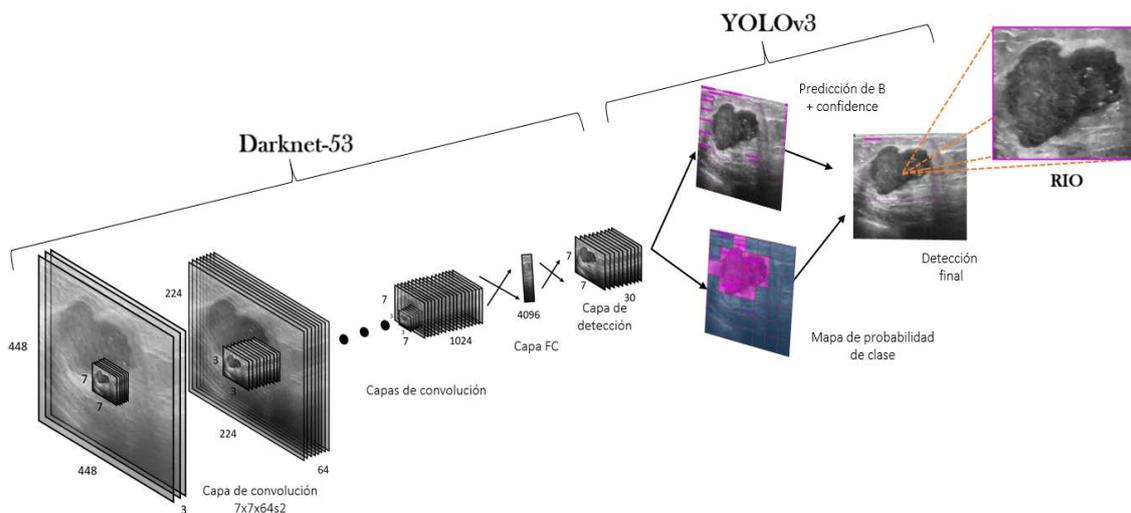


Figura 4.3 Módulo de detección: YOLOv3.

Como se puede observar este utiliza la red neuronal Darknet53 como base y para la última capa aplica el algoritmo YOLOv3, el cual se encarga de predecir los cuadros delimitadores donde se ubica la masa, los cuales tomará como referencia la red de segmentación DeepLabv3+.

### 4.1.3 Módulo de Segmentación

Este se utiliza para hacer la segmentación de la masa en la RIO obtenida en el módulo anterior con YOLOv3. Este módulo utiliza como red de segmentación la red DeepLabv3+ el cual es se basa en un auto-encoder. Donde este toma la imagen de entrada (RIO) y asigna etiquetas a cada pixel para realizar la segmentación de la masa mediante convoluciones *atrous*.

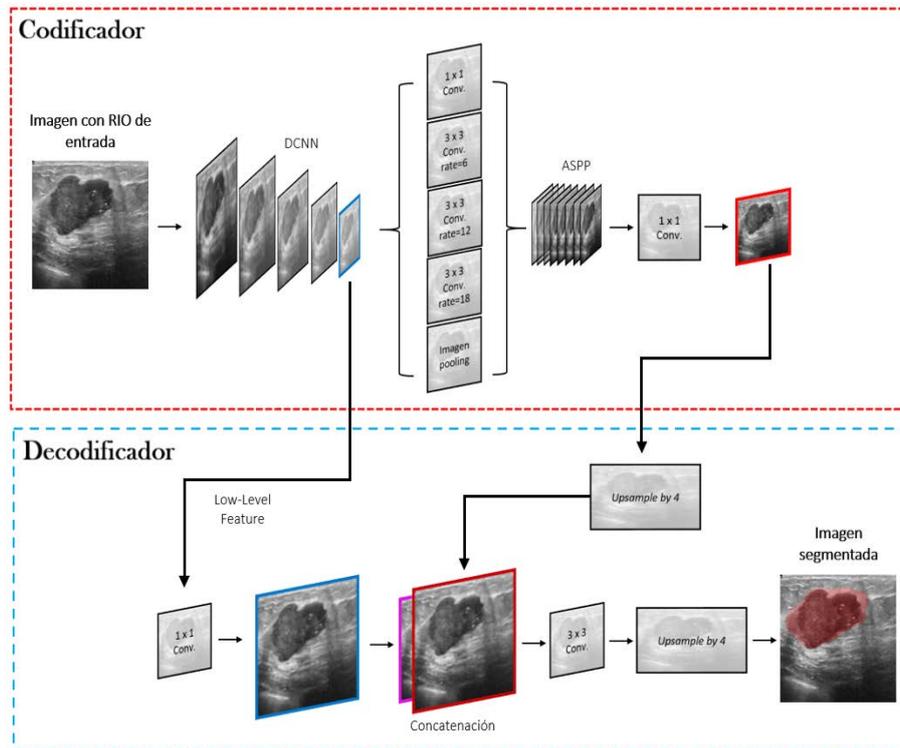


Figura 4.4 Módulo de detección: DeepLabv3+.

En general, el sistema propuesto pretende mejorar el proceso de segmentación de una masa sólida en ecografía, dado que no solo implementa un preprocesamiento en la imagen como la mayoría de lo trabajo de la literatura, sino que también hace una delimitación de la región de interés para indicar al la red de segmentación donde debe aplicarse así evitando que esta se confunda y delimite otros lugares en la ecografía.

# Capítulo 5

## Experimentación y Resultados

### 5.1 Diseño Experimental

En esta sección se detallan las métricas, bases de datos, experimentos realizados y resultados obtenidos con la metodología propuesta.

#### 5.1.1 Entorno de desarrollo

Para los experimentos propuestos se utilizó un equipo de cómputo con las siguientes especificaciones: Arquitectura de hardware: Procesador 11th Gen Intel(R) Core(TM) i7-11800H @ 2.30GHz, 16 GB de memoria RAM y GPU NVIDIA Geforce RTX 3060 16GB. Arquitectura de software: Sistema Operativo Ubuntu 22.04 LTS, MATLAB R2023b, gThumb Software, LabelImg Software y librería CUDA 11.

#### 5.1.2 Base de datos

Para la experimentación se utilizaron dos bases de datos de imágenes ecográficas con presencia de masas anormales en la mama. La primera base de datos, *Medical Image Database* (MID) fue obtenida del *Biomedical Engineering Research Unit of SIIT* [75], la cual consta de 763 imágenes de ecografía, clasificada en cinco conjuntos; masas sólidas benignas, masas sólidas malignas, quistes, fibrodemas y sin clasificación, las cuales a su vez se dividen en tres subconjuntos, imagen original, imagen original recortada e imágenes *ground truth* delimitada por radiólogo. Las Figuras 5.1a y 5.1b muestra el *ground truth* de una masa benigna y una maligna delimitada por un radiólogo respectivamente.

La segunda base de datos utilizada *Breast Ultrasound Images Dataset* (BUSI) fue obtenida de [4] la cual contiene 1312 imágenes de ecografías divididas en 4 subconjuntos, ecografías con masas malignas, *ground truth* de ecografías con masas malignas delimitadas por radiólogo, ecografías de masas benignas y *ground truth* de ecografías con masas benignas delimitadas por radiólogo. Cabe mencionar que el *ground thrut* de esta base de datos, a diferencia de la base de datos anterior, está en formato de máscara binaria. Las Figuras 5.2a y 5.2b muestra dos imágenes del conjunto de datos BUSI.

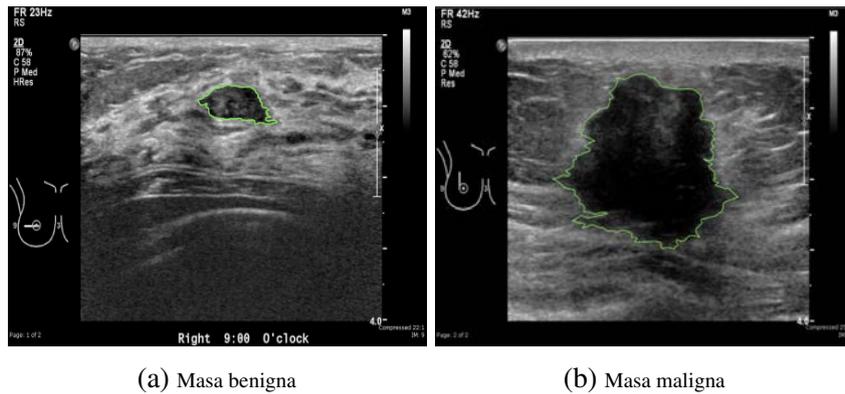


Figura 5.1 *Ground truth* de masas del conjunto de datos MID [75].

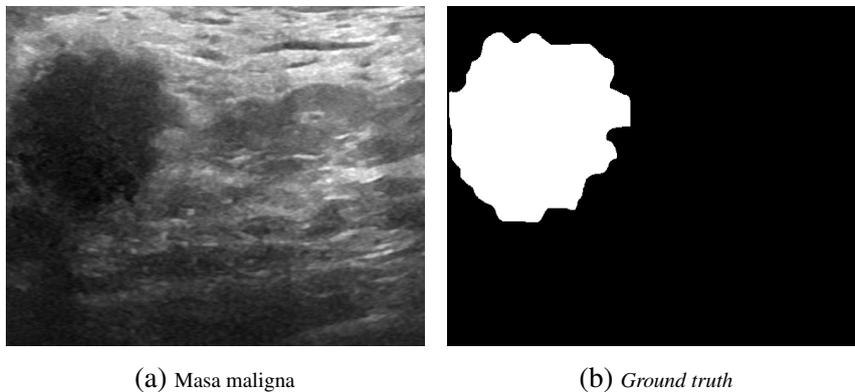
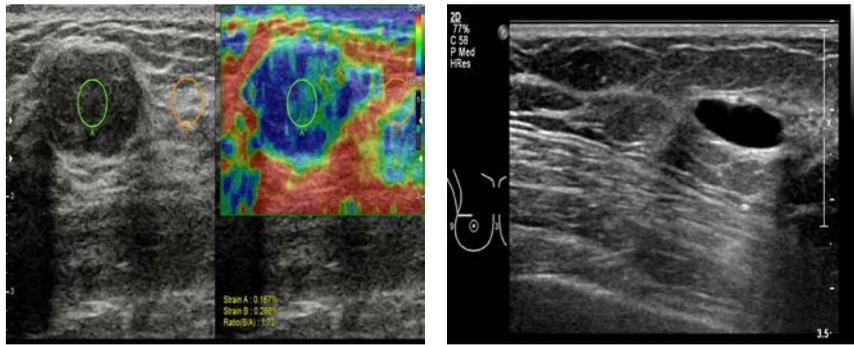


Figura 5.2 Ecografías y *ground truth* de la base de datos BUSI [4].

### Normalización de la base de datos

Con la finalidad de obtener subconjuntos de imágenes ecográficas con masas sólidas de ambas bases de datos, se realizó la extracción solo de imágenes de ecográficas con masas sólidas malignas y benignas y sus respectivos *ground truth*. Así mismo, la normalización del conjunto de datos se llevó, acabó descartando imágenes con anotaciones realizadas por el radiólogo, Figura 5.3a. De igual manera, se realizó un recorte en la imagen mediante el software gThumb con la finalidad de eliminar partes de la imagen que contuviera algún borde realizado por el dispositivo de captura, Figura 5.3b. La Tabla 5.1 muestra los subconjuntos obtenidos después de este proceso de normalización.



(a) Anotaciones realizadas por el radiólogo

(b) Bordos del dispositivo de captura

Figura 5.3 Normalización del conjunto de datos [4].

Tabla 5.1 Subconjuntos de imágenes obtenidos de BUSI y MID.

Base de datos	Conjunto Utilizado	Cantidad
BUSI	Masas malignas (M-BUSI)	210
	Masas benignas (B-BUSI)	437
MID	Masas malignas (M-MID)	108
	Masas benignas (B-MID)	25

### 5.1.3 Métricas

A continuación se muestran las métricas utilizadas para los diferentes experimentos realizados.

#### Reducción de ruido

Para medir cuantitativamente la efectividad y robustez de los filtros de reducción de ruido, se utilizaron las siguientes métricas:

**Signal-to-Noise Ratio (SNR):** es la relación de potencia entre la fuerza de señal y el nivel de ruido, y está dada por:

$$SNR = \frac{P_{signal}}{P_{noise}} \quad (5.1)$$

**Peak Signal-to-Noise Ratio (PSNR):** es relación entre el valor máximo posible (potencia) de una señal y la potencia del ruido distorsionador que afecta la calidad de su representación.

$$PSNR = 20 \log_{10} \left( \frac{MAX_f}{\sqrt{MSE}} \right). \quad (5.2)$$

Donde el **MSE** (*Mean Squared Error*) se define como:

$$MSE = \frac{1}{mn} \sum_0^{m-1} \sum_0^{n-1} \| f(i, j) - g(i, j) \|^2 \quad (5.3)$$

**Structural Similarity Index (SSIM):** es una métrica perceptual que cuantifica la degradación de la calidad de la imagen causada por el procesamiento.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (5.4)$$

### Detección de objetos

Para determinar la eficiencia y precisión del algoritmo de detección de objetos fue necesaria una evaluación con las siguientes métricas:

#### Matriz de Confusión (ConfMtr):

De acuerdo a [11] una matriz de confusión se define como:

$$ConfMtr = \begin{bmatrix} VP & FN \\ FP & VN \end{bmatrix} \quad (5.5)$$

donde,

$FP$  = objetos que no son de la clase detectados.

$FN$  = objetos de la clase no detectados.

$VP$  = objetos de la clase detectados.

$VN$  = objetos que no son de la clase no detectados.

Derivado de 5.5 se utilizaron las siguientes métricas [11]:

**Precision (P):** esta mide la precisión de los cuadros delimitadores teniendo en cuenta la relevancia o exactitud de su predicción.

$$P = \frac{VP}{VP + FP} \quad (5.6)$$

**Recall (R):** esta calcula cuántos objetos son seleccionados en los cuadros delimitadores, permitiendo conocer la eficiencia del sistema en la detección del objeto de la clase de interés. También se le conoce como sensibilidad o *True Positive Rate* (TPR).

$$R = \frac{VP}{VP + FN} \quad (5.7)$$

**Specificity (S):** es la tasa de verdaderos negativos, es decir, el número correcto de detecciones negativas obtenidas en la red.

$$S = \frac{VN}{VN + FN} \quad (5.8)$$

**F1-score (F1):** es la estadística media de la *Precision* y *Recall*.

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (5.9)$$

**False Negative Rate (FNR):** representa la proporción de instancias positivas que fueron incorrectamente clasificadas como negativas en relación con el total de instancias positivas reales.

$$FNR = \frac{FN}{FN + VP} \quad (5.10)$$

### Segmentación de objetos

Las métricas implementadas para evaluar los resultados de la delimitación de las masas sólidas mediante la red utilizada fueron:

**Jaccard Index (DI):** Es una estadística utilizada para comparar la similitud de dos imágenes y se define como el tamaño de la intersección dividido por el tamaño de la unión de las imágenes.

$$DI = \frac{2|A \cap B|}{|A| + |B|} \times 100 \quad (5.11)$$

Donde  $A$  y  $B$  son dos imágenes. Si el valor del coeficiente es 1, entonces muestra la alta similitud entre  $A$  y  $B$ . De lo contrario, si su valor es 0, entonces no hay similitud entre  $A$  y  $B$ .

**Jaccard similarity Coefficient (DSC):** También conocido como índice de Sørensen-Dice o simplemente coeficiente de similaridad de Dice, es una métrica estadística que mide la similitud entre dos conjuntos de datos. Matemáticamente se define como:

$$DC = 2 \frac{|A \cap B|}{|A| + |B|} \quad (5.12)$$

Dónde A y B son dos conjuntos con barras verticales a ambos lados se refiere a la cardinalidad del conjunto, es decir, al número de elementos de ese conjunto, por ejemplo,  $|A|$  significa el número de elementos del conjunto A.  $\cap$  se utiliza para representar la intersección de dos conjuntos, y significa los elementos que son comunes a ambos conjuntos.

## 5.2 Filtrado de Ruido Multiplicativo

Con el objetivo de evaluar el desempeño de las diferentes técnicas de filtrado se realizaron los siguientes experimentos. El primer experimento consistió en la sintonización de los filtros mediante una imagen de muestra previamente contaminada con ruido multiplicativo con diferentes intensidades del filtro. El segundo experimento consistió en la implementación de los filtros previamente sintonizados en imágenes ecográficas de la base de datos utilizada.

### Experimento 1. Sintonización de los Filtros

Dada la naturaleza de adquisición de una imagen ecográfica, esta siempre contendrá una cierta cantidad de ruido multiplicativo, por lo que no es posible evaluar el proceso de filtrado con métricas de la literatura que basan sus resultados en comparar una imagen original sin ruido y una imagen filtrada previamente contaminado con ruido. Por ello es necesario realizar la sintonización de los filtros con una imagen sin ningún tipo de ruido y una imagen igual previamente corrompida. A continuación se evalúan los filtros antes mencionado a diferentes intensidades con métricas de SNR, PSNR, MSE y SSIM.

Las imágenes utilizadas para la sintonización fue “*peppers.png*” original a escala de grises y “*peppers.png*” con una densidad de ruido *speckle* de 0.05. En la Figura 5.4 se muestran las gráficas de los resultados cuantitativos obtenidos con las métricas SNR, PSNR, MSE y SSIM. Donde FA es el filtro de difusión anisotrópica, FG es el filtro de imagen guiada, NL es el filtro de *Non Local means*, FF el filtro *Frost*, FL el filtro *Lee* y FS el filtro SRDA.

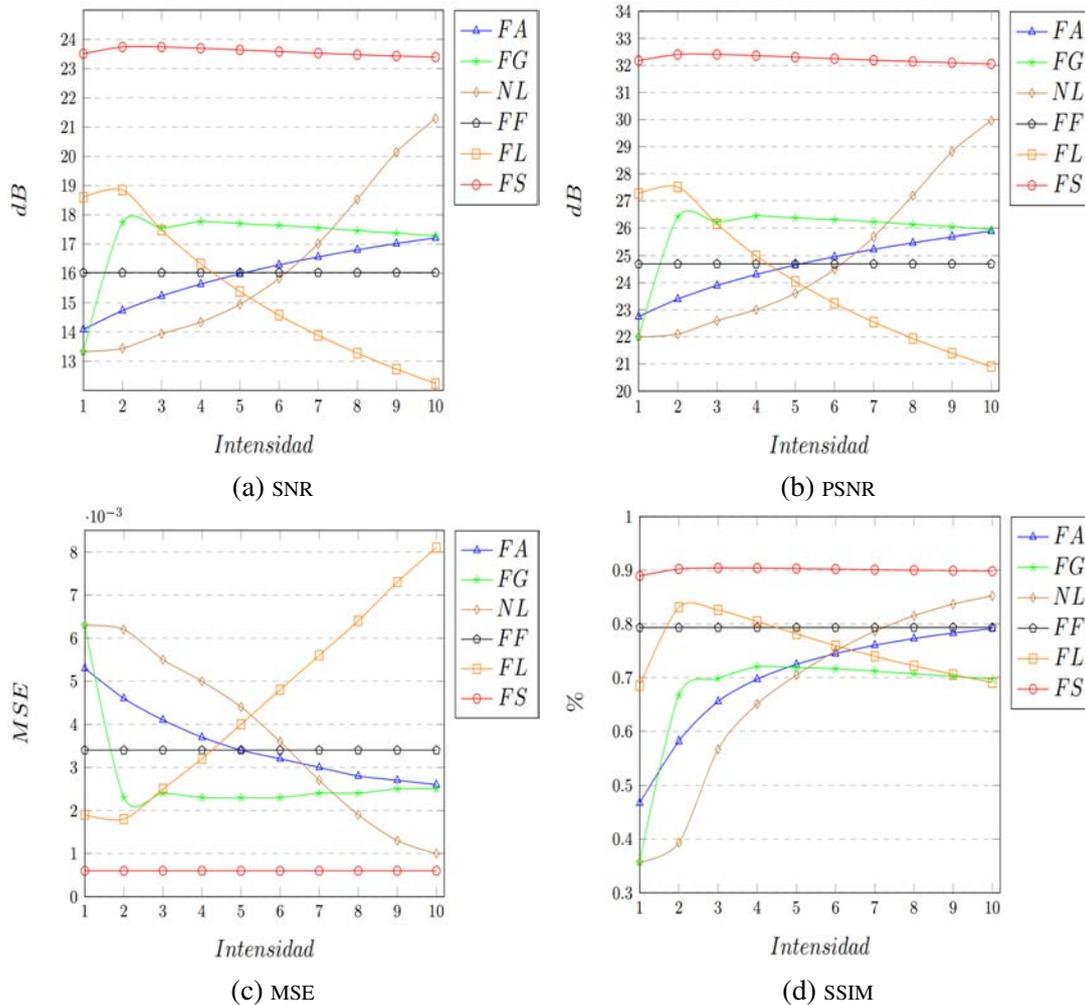


Figura 5.4 Resultados cuantitativos de las métricas SNR, PSNR, MSE y SSIM.

De acuerdo a los experimentos realizados, se observó que el filtro FS logra obtener resultados muy por encima de los demás filtros en los 10 niveles de intensidad que fue aplicado, siendo 2 el que mejores resultados obtuvo. Por otro lado, los mejores resultados para los otros filtros fueron: 10 para FA, 2 para FG, 10 para NL y para FL, el que tuvo peor desempeño, dado que no logró reducir adecuadamente el ruido multiplicativo y generó una mayor degradación de la calidad en la imagen sin importar la intensidad con la que fue aplicado. Los resultados obtenidos proporcionan un panorama más amplio para comprender las ventajas y limitaciones de cada filtro en general, así como para seleccionar la técnica de filtrado más adecuada para el procesamiento de imágenes ecográficas. Sin embargo, es importante tener en cuenta que este experimento se realizó con el objetivo de ajustar los parámetros de los filtros, por lo que, los resultados obtenidos pueden variar al aplicarlos a imágenes ecográficas reales, debido a la diversidad de los tipos de imagen y complejidad estructural.

## Experimento 2. Filtrado de imágenes de Ecografía

Dada su naturaleza de adquisición, una ecografía siempre contendrá una cierta cantidad de ruido con una densidad desconocida, por lo que no es posible evaluar el filtrado con métricas de la literatura de manera directa. Por ello, para este experimento se propone realizar un filtrado previo con la red de eliminación de ruido dnCNN [134] con la finalidad de obtener una imagen de referencia. Para este experimento se compararon los filtros FA, FG, NL, FF, FL y FS con las intensidades mejor puntuadas en el experimento anterior. Se utilizaron imágenes de ecografías reales con presencia de masas sólidas extraídas de las bases de datos MID y BUSI. Las Figuras 5.5, 5.6 y 5.7 muestran los resultados obtenidos en los cuatro subconjuntos de la base de datos utilizada en la Tabla 5.1 respecto a las métricas PSNR, SSIM y MSE.

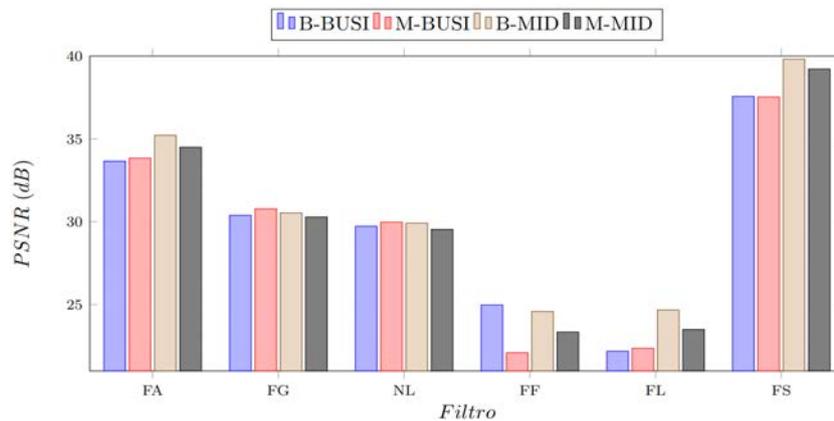


Figura 5.5 PSNR en base de datos BUSI y MID.

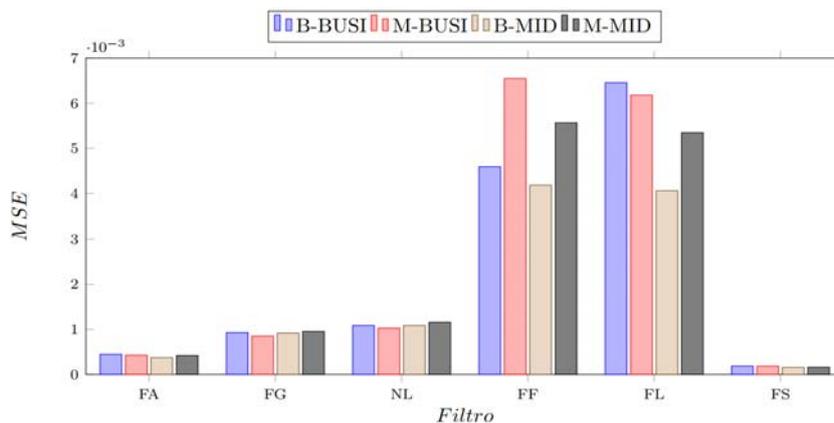


Figura 5.6 MSE en base de datos BUSI y MID.

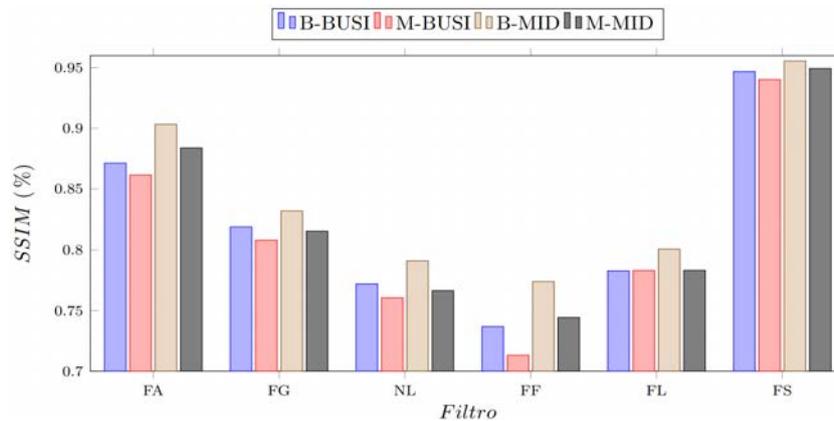
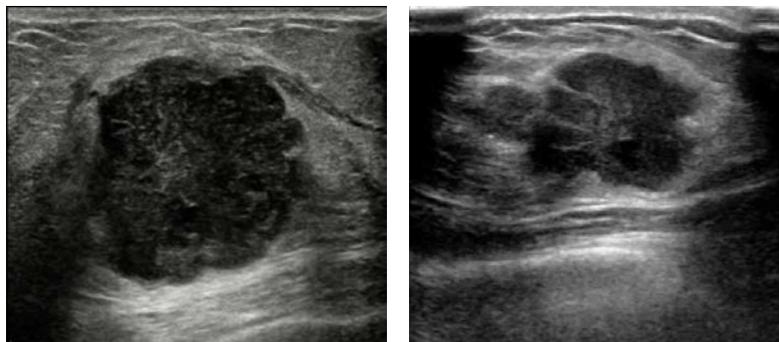


Figura 5.7 SSIM en base de datos BUSI y MID.

De acuerdo a los resultados de las figuras anteriores, se puede apreciar que el filtro con mejores resultados es FS tanto en imágenes de la base de datos BUSI como en MID, obteniendo resultados muy por encima de los otros filtros en todas las métricas. Así mismo, de acuerdo a la gráfica de MSE se puede apreciar como el filtro con menor error cuadrático medio después de FS es el filtro FA, dado que FS es una variante del filtro FA. Por otro lado, los filtros con mayor discrepancia entre la imagen original y la filtrada de acuerdo a SSIM fueron FF y FL teniendo su puntaje más bajo en las imágenes de la base de datos BUSI.

A continuación se presentan los resultados cualitativos de acuerdo a la calidad de los bordes de la masa y menor intensidad de ruido. La Figura 5.8 muestra las 2 imágenes utilizadas para este experimento, donde MB es una imagen con masa benigna y MM es una imagen con masa maligna. Debido a que la base de datos MID contiene mayor grado de ruido en la imagen, se seleccionaron para este experimento 2 imágenes ecográficas, una con masa maligna, Figuras 5.8b y otra con masa benigna 5.8a.



(a) Masa benigna (MB)

(b) Masa maligna (MM)

Figura 5.8 Imágenes de ultrasonido utilizadas.

Para la imagen MB se obtuvieron los siguientes resultados, Figura 5.9.

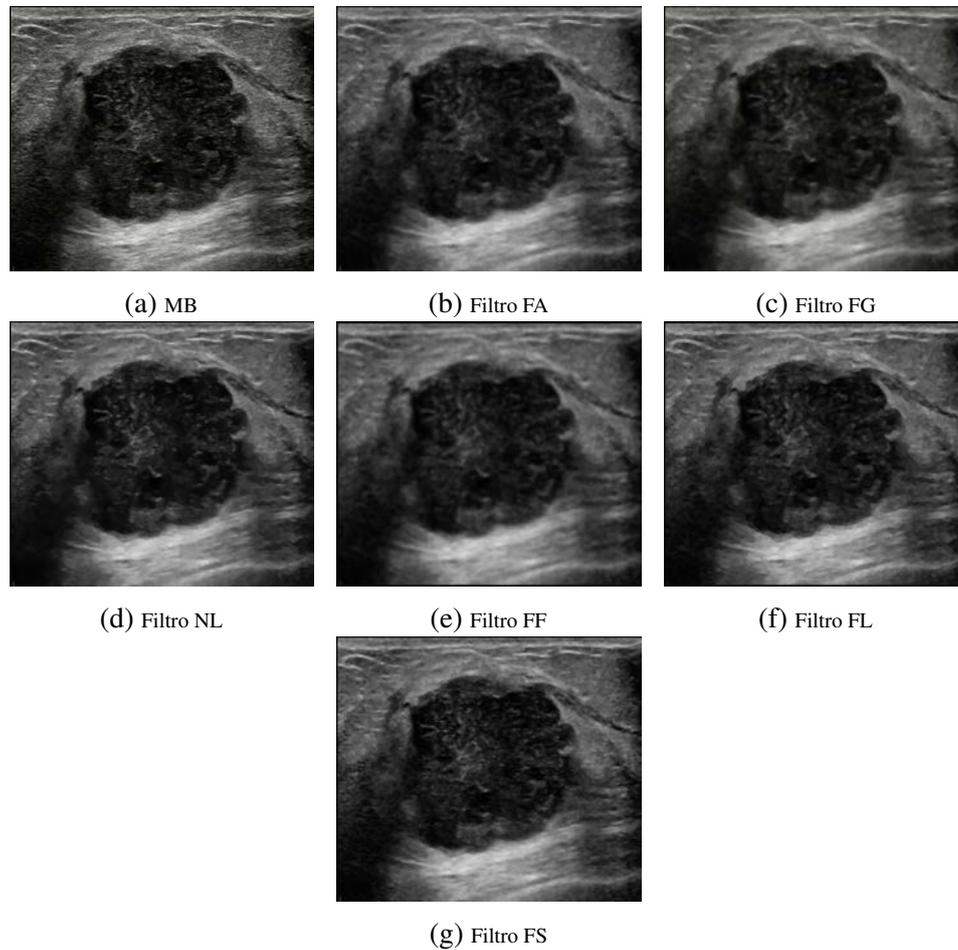


Figura 5.9 Filtrado de ecografía con masa sólida benigna.

De acuerdo a los resultados de la Figura 5.9 al ser una imagen con menor complejidad en sus bordes, se puede observar que en el proceso de reducción del ruido multiplicativo por parte de los seis filtros obtienen buenos resultados, dado que estos a grandes rasgos preservan de bordes; sin embargo, la reducción de ruido es poca con excepción del filtro FS. La Figura 5.9g muestra como el filtro con mejor preservación de bordes y reducción de ruido en una imagen de masa benigna es el filtro FS. Por otro lado, los filtros FA, FG, NL, FF y FL, si bien reducen la intensidad del ruido de igual manera degradan los bordes de la masa.

Para la imagen MM se obtuvieron los siguientes resultados, Figura 5.10.

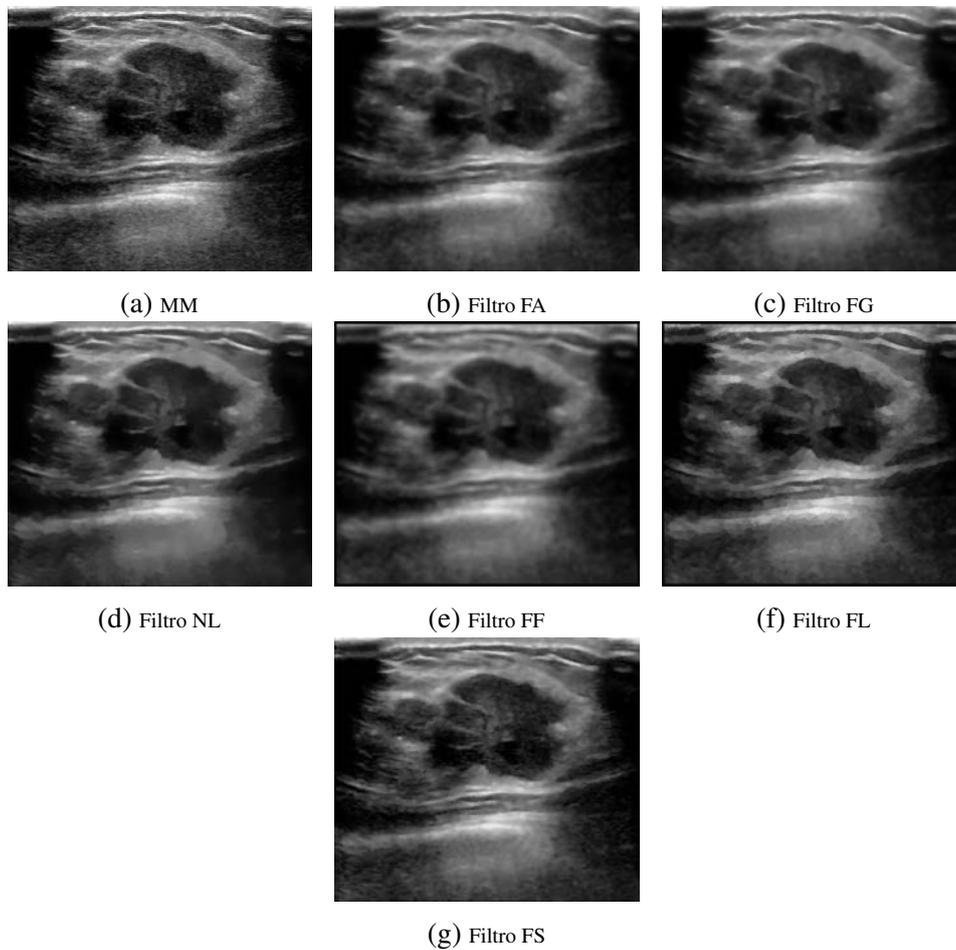


Figura 5.10 Filtrado de ecografía con masa sólida maligna.

De acuerdo a la Figura 5.10 a diferencia de la imagen anterior, la ecografía de la Figura 5.9 presenta un mayor grado de complejidad al analizar los bordes de la masa, por lo que hace más difícil el proceso de filtrado de ruido conservando los bordes. De acuerdo a la Figura 5.10g, el filtro con mejores resultados fue el filtro FS, seguido del filtro de NL, Figura 5.10d y FA, Figura 5.10b. Para los filtros FG, FF, y FL los resultados fueron poco favorables al perder mayor calidad en los bordes de la masa. La justificación de esta serie de experimentos se centró, como primer punto, en la evaluación cuantitativa de los resultados del proceso de filtrado en imágenes ecográficas, específicamente en términos de SSIM. Los resultados demuestran que todos los filtros superan el umbral del 70% en ambas bases de datos, siendo el filtro FS el que muestra una mejora superior del 5% en comparación con los demás. El objetivo principal fue determinar el impacto del proceso de filtrado en la imagen, analizando cómo se modifica en relación con la imagen original.

De acuerdo a los resultados obtenidos de las imágenes de masa maligna, se puede concluir que el filtro SRAD demostró una mejor preservación de bordes y reducción de ruido, seguido por el filtro FA y el filtro NL. Por otro lado, los filtros de FG y FF y FL tuvieron los resultados menos factibles dado que degradaron los bordes de la masa. En el caso de las imágenes de masa benigna, nuevamente el filtro SRAD mostró la mejor preservación de bordes y reducción de ruido, seguido por el filtro NL y el filtro FA. Por otro lado, los filtros FG, FF y FL de igual manera obtuvieron un rendimiento inferior en la eliminación de ruido, lo que resultó en una degradación de los bordes de la masa.

En resumen, la finalidad de este experimento consistió en analizar el comportamiento de los filtros utilizados en imágenes ecográficas y determinar su viabilidad en el preprocesamiento durante la reducción de ruido. La importancia de este experimento radica en la necesidad de seleccionar el filtro más adecuado para mejorar la calidad de las imágenes ecográficas, ya que esto puede tener un impacto significativo en el diagnóstico y análisis con técnicas de deep learning. Al comprender y cuantificar el efecto de cada filtro en términos de cómo la SSIM, MSE y PSNR se proporciona una base sólida para la optimización del proceso de filtrado y la selección de la técnica más adecuada para la reducción de ruido.

### 5.3 Detección de Masas en Ecografías

Esta serie de experimentos se basa en la aplicación del detector de objetos YOLO para extraer la región que contiene la masa sólida, reduciendo así el área de análisis para mejorar la eficiencia computacional y precisión en un paso posterior de delimitación de la masa.

#### Experimento 3. Reentrenamiento de la red

Como primer experimento se realizó el reentrenamiento de la red Darknet-53 para que esta pudiera detectar de manera correcta una masa sólida a través del algoritmo YOLOv3. La arquitectura para el reentrenamiento se basó en una CPU Intel Xeon con 13 GB de RAM y una GPU Nvidia Tesla T4. El conjunto de datos de entrenamiento empleado constó de 2,000 imágenes de ecografía, obtenidas de la base de datos [44] la cual se dividió en dos subconjuntos: entrenamiento (90%) y validación (10%). Para iniciar el entrenamiento de la red fue necesario descargar los pesos iniciales del algoritmo proporcionados por el autor a manera de usarlos para la “*transferencia de aprendizaje*” con la finalidad de reducir el tiempo y costo computacional. Así mismo se realizó el etiquetado manual de la masa sólida en la ecografía por medio del software Labelimg, Figura 5.11. Los resultados del reentrenamiento de la red se muestran en la Tabla 5.2.

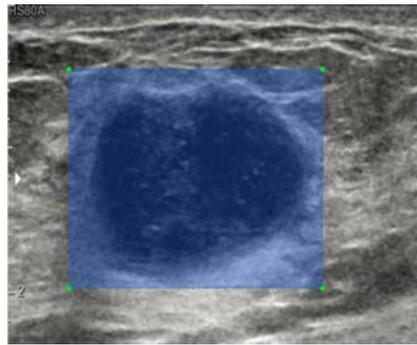


Figura 5.11 Etiquetado manual de masa sólida mediante Labelimg.

Tabla 5.2 Resultados del entrenamiento de la red.

Iteraciones	Pérdida promedio	Tasa de aprendizaje	Tiempo	Imágenes utilizadas
10,000	0.23, 0.26	$1 e^{-5}$	68 hr.	640,000

#### Experimento 4. Sintonización de la red

Para este experimento se utilizaron 100 imágenes de la base de datos utilizada, las cuales fueron procesadas a diferentes umbrales de detección. Dado que este umbral controla el valor de *confidence* de los cuadros delimitadores predichos por el algoritmo, se realizó la detección de masas con la finalidad de encontrar el que mejores resultados obtuviera de acuerdo a las métricas de evaluación de detección de objetos.

La Figura 5.12 muestra los resultados obtenidos con respecto a las métricas utilizadas de acuerdo a 10 umbrales de detección. Las imágenes procesadas fueron de un tamaño de  $416 \times 416$  a escala de grises.

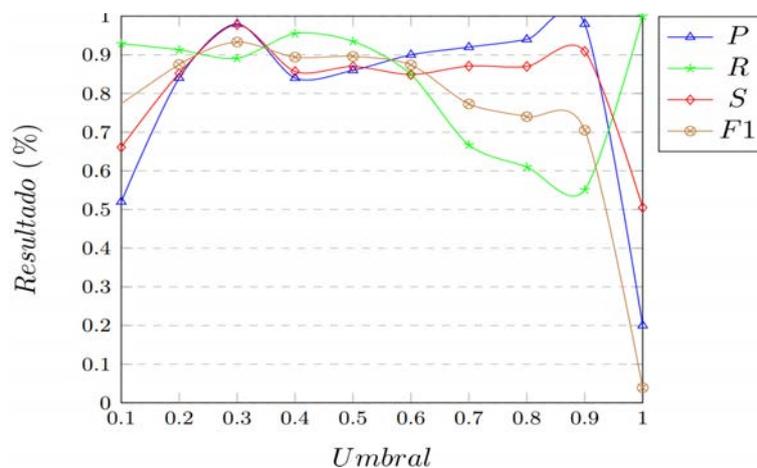


Figura 5.12 Comparación de umbrales de detección en YOLOv3.

Como se puede observar en la gráfica anterior, el umbral de detección que mejores resultados obtuvo fue 0.25, teniendo una sincronía en las métricas empleadas, alcanzando valores arriba del 90% en las 4 métricas. Si bien el umbral 0.35 obtiene una sincronía total en las métricas del 90% la red realiza detecciones duplicadas para el mismo objeto por lo que se descarta para futuros experimentos.

### Experimento 5. Detección de diferentes tipos de Masas

Para este experimento se utilizaron imágenes con diferentes tipos y estados de masas en ecografías obtenidas de la base de datos [44]: Masa Benigna Adenosis (MBA), Masa Benigna Fibroadenoma (MBF), Masa Benigna Vascularizada (MBV), Masa Maligna de Carcinoma Ductal en Situ (MMS), Masas Maligna Carcinoma Lobular Invasivo (MML) y Masas Maligna Carcinoma Metastásico (MMM). En la Figura 5.13 se pueden ver estos tipos de masas.

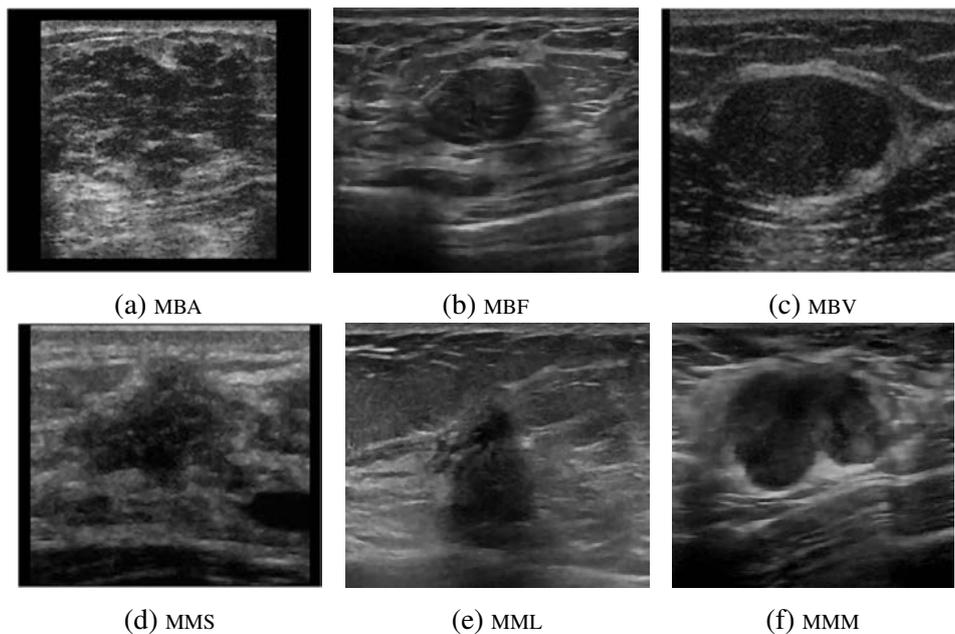


Figura 5.13 Ecografías con diferentes tipos masas [44].

Los resultados cuantitativos con las métricas seleccionadas se muestran en la gráfica de la Figura 5.14. Donde se compara la precisión del algoritmo en 10 imágenes de cada tipo de masa.

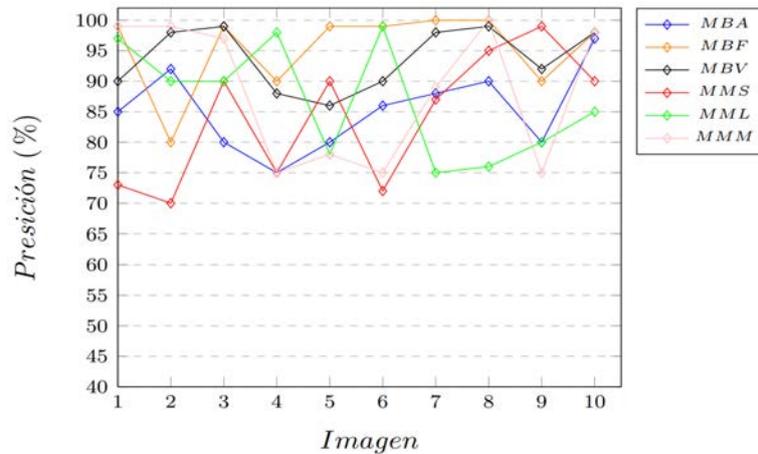
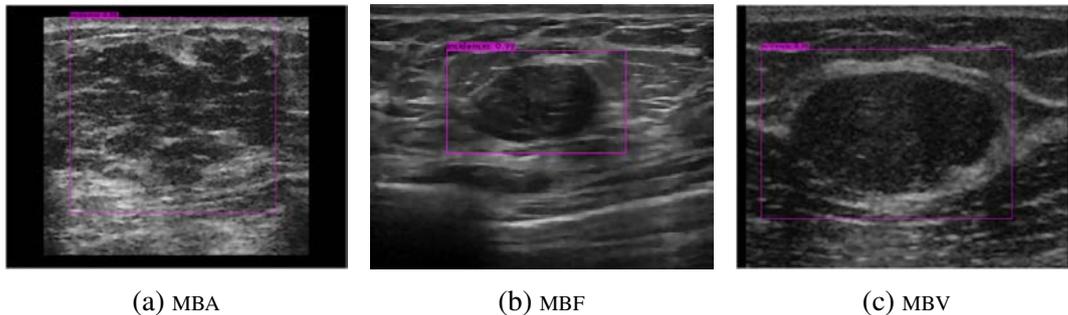


Figura 5.14 Precisión de detección de YOLOv3.

De acuerdo a los resultados obtenidos en la gráfica de la Figura 5.14, se observa una detección mejor en las masas tipo MBII y CDS, logrando una precisión del 100% en algunos casos. Estos resultados resaltan la relevancia de las características distintivas presentes en este tipo de masa benigna, donde los bordes son más nítidos, lo cual facilita su detección en la ecografía. Por otro lado, los resultados obtenidos demuestran la efectividad del algoritmo en la detección de masas sólidas en ecografías, demostrando la robustez y eficacia de la aplicación de este tipo de redes neuronales en el procesamiento de imágenes médicas. En cuanto a los resultados cualitativos obtenidos, se puede observar la detección precisa de los diferentes tipos de masas en las imágenes de ecografías utilizadas, pudiendo detectar masas pobremente delimitadas y ecografías corrompidas con ruido. La Figura 5.16 muestra estos resultados obtenidos durante la detección de las masas en las imágenes utilizadas.



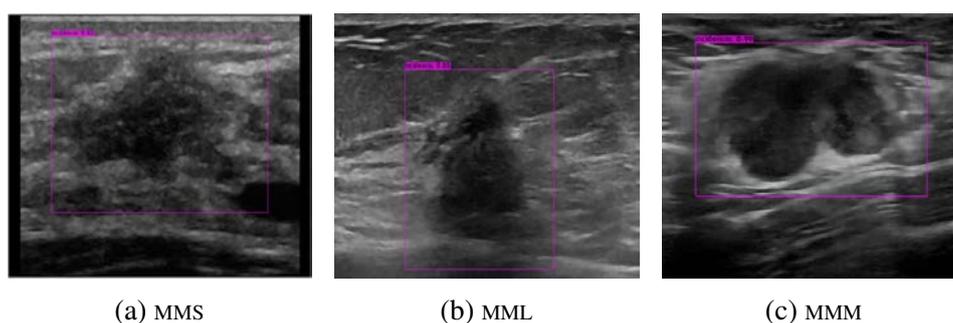


Figura 5.16 Detección de masas en ecografías mediante YOLOv3.

De acuerdo a la Figura 5.16 se puede observar que YOLO demuestra una eficiencia en la detección de masas en imágenes ecográficas, siendo capaz de detectar masas con delimitación deficiente y presencia de ruido. Es importante mencionar que estos resultados se basan en el conjunto de imágenes de ecografías de la base de datos utilizada. Por lo que la aplicación del algoritmo en diferentes conjuntos de datos y escenarios clínicos puede variar. No obstante, los resultados obtenidos respaldan la efectividad del algoritmo y proporcionan una base sólida para la implementación del sistema propuesto. Como último experimento para evaluar este algoritmo se realiza una detección de masas sólidas en la base de datos utilizada.

### Experimento 6. Detección de Masas en la base de datos utilizada

Para este experimento se evaluó la detección de masas benignas y malignas de imágenes de la base de datos utilizada. La evaluación se llevó a cabo con las métricas de *Precision*, *Recall*, *Specificity* y *F1 – Score*. Los resultados obtenidos se muestran en la gráfica de la Figura 5.17.

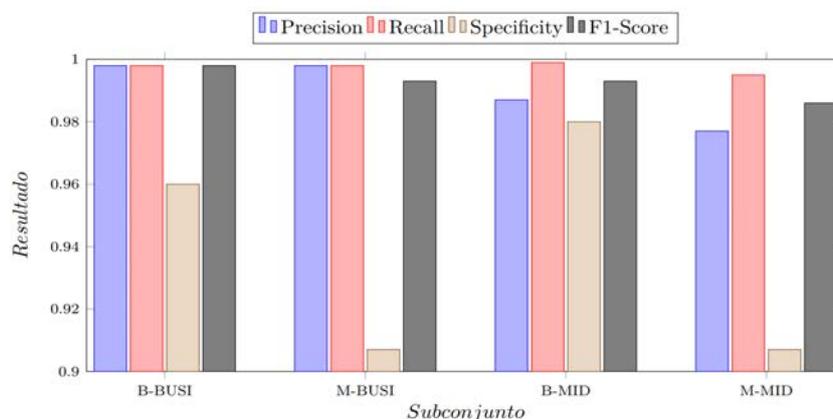


Figura 5.17 Resultados de la detección de masas en ecografías de las bases de datos utilizada.

En la gráfica anterior se observa que hubo una mejor detección en masas benignas de acuerdo a las métricas establecidas, dado que como se mencionaba anteriormente, estas cuentan con un mayor nivel de detalle en los bordes a diferencia de las masas malignas. De igual manera, se puede apreciar que la red obtiene resultados por encima del 94% en todas sus métricas, lo que demuestra su efectividad en la detección de diferentes tipos de masas. Por ello se puede deducir que los resultados reflejan la efectividad y confiabilidad del modelo utilizado en la detección de masas, independientemente de su naturaleza benigna o maligna. Los resultados de este experimento demuestran la efectividad que tiene la implementación de algoritmos de detección de objetos como YOLO en el campo de la detección de masas sólidas en ecografías, ya que respaldan el uso de redes neuronales convolucionales como una herramienta confiable en el diagnóstico y evaluación de incidencias en imágenes médicas.

El propósito de este experimento fue utilizar YOLO para reducir el área de búsqueda de la masa en la imagen, permitiendo al algoritmo de delimitación enfocarse únicamente en una sola región de interés (RIO) en un paso posterior. En otras palabras, este trabajo de investigación pretende tomar solo la región delimitada por el cuadro de predicción proporcionado por YOLOv3 como región para buscar y delimitar la masa en la ecografía.

## 5.4 Segmentación de Masas en Ecografías

La segmentación precisa de masas sólidas en imágenes médicas, como la ecografía, desempeña un papel fundamental en el diagnóstico y tratamiento temprano del cáncer. En particular, la segmentación precisa de las masas puede proporcionar información crucial sobre su forma, tamaño y ubicación, lo que ayuda a los médicos a tomar mejores decisiones y brindar el tratamiento adecuado. A continuación se presenta los experimento realizados para la segmentación de masas mamográficas mediante el sistema propuesto, el cual integra el preprocesamiento de la imagen mediante el filtro SRAD y la detección de región de interés proporcionada por YOLOv3, al modelo de segmentación de objetos DeepLabv3+. La Figura 5.18 muestra el proceso de una imagen de ecografía por el cual pasa para realizar la delimitación de la masa.

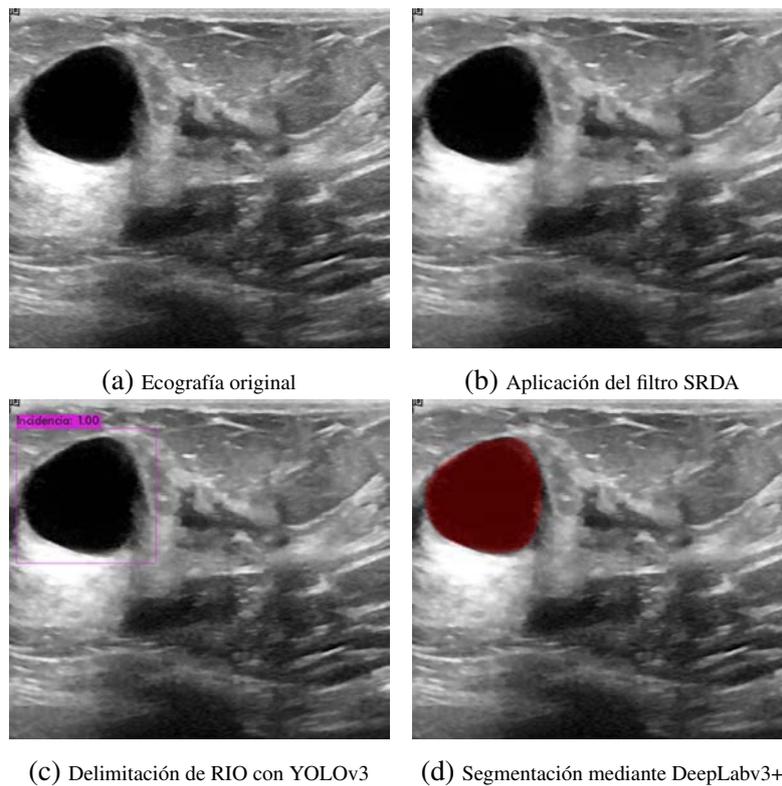


Figura 5.18 Proceso de segmentación mediante el sistema propuesto.

Este sistema de aprendizaje profundo diseñado para la segmentación de masas en imágenes médicas, el cual utiliza una combinación de técnicas, como el preprocesamiento de la imagen a través del filtro SRDA, la delimitación de la región de interés con el algoritmo de detección de objetos YOLOv3 y la segmentación semántica con la red de segmentación DeepLabv3+.

### Experimento 7. Segmentación de Masas sólidas en Ecografías con el sistema propuesto

Como primer experimento se realiza la segmentación de la base de datos completa con el sistema propuesto con la finalidad de comparar el comportamiento del sistema, dado que la red de segmentación fue entrenada solo con imágenes de la base de datos BUSI y no de MID. La Figura 5.19 muestra la gráfica de los resultados obtenidos con respecto a las métricas *Jaccar-Index*, *Sørensen-Dice*, *Precision*, *Recall*, *Specificity* y *F1 – score*. Donde B-BUSI Y M-BUSI corresponde a una masa benigna y maligna respectivamente de la base de datos BUSI, mientras que B-MID y M-MID sin las masas malinas y benignas de la base de datos MID.

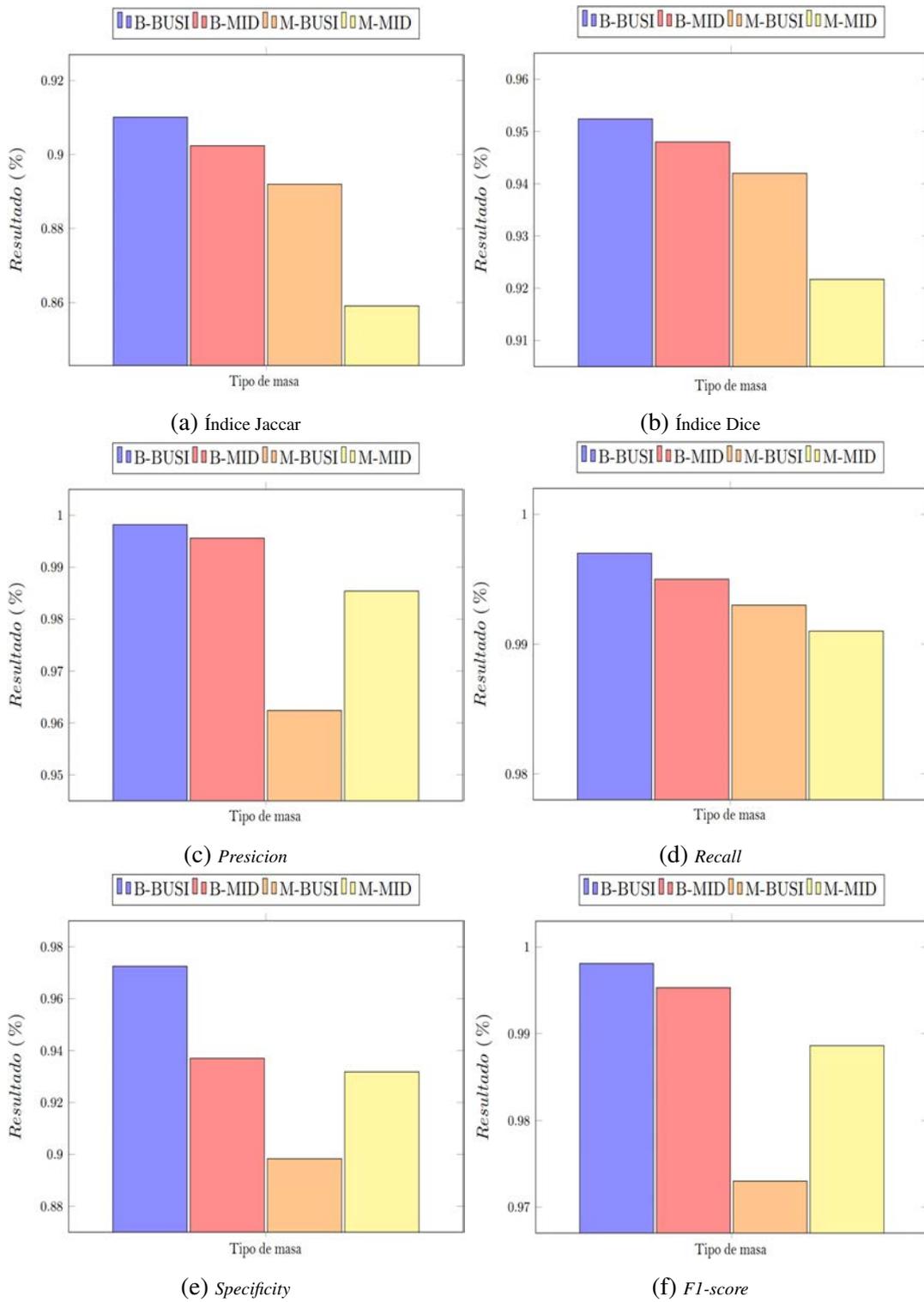


Figura 5.19 Resultados cuantitativos de segmentación de masas sólidas en base de datos utilizada.

Como se puede observar en la Figura 5.19 el sistema propuesto logra tener un buen desempeño en la segmentación de masas sólidas para ambas bases de datos. Para las masas benignas, el sistema es capaz de segmentar con una puntuación por encima del 90% en todas las métricas, siendo las masas benignas de la base de datos BUSI las que mejor puntuación obtienen. Por otro lado, si bien la puntuación de la segmentación de masas malignas están por debajo en comparación con las benignas, estas logran una puntuación de alrededor del 90% lo cual resulta bastante bueno dado que este tipo de masas tiene bordes poco definidos y son difíciles de analizar. Otro aspecto importante a notar fue el comportamiento del sistema al segmentar masas de ecografías que no pertenecen a la base de datos con la que fue entrenado, lo que refleja la versatilidad del sistema, dado que este es capaz de trabajar de manera correcta con imágenes con características de captura totalmente diferentes al que fue entrenado.

A continuación se presentan los resultados cualitativos obtenidos con este experimento. La Figura 5.20 muestra la segmentación de masas benignas a través del sistema propuesto donde la primera columna corresponde a la imagen original, la segunda al *ground truth* de la masa y la tercera a la delimitación realizada por el sistema.

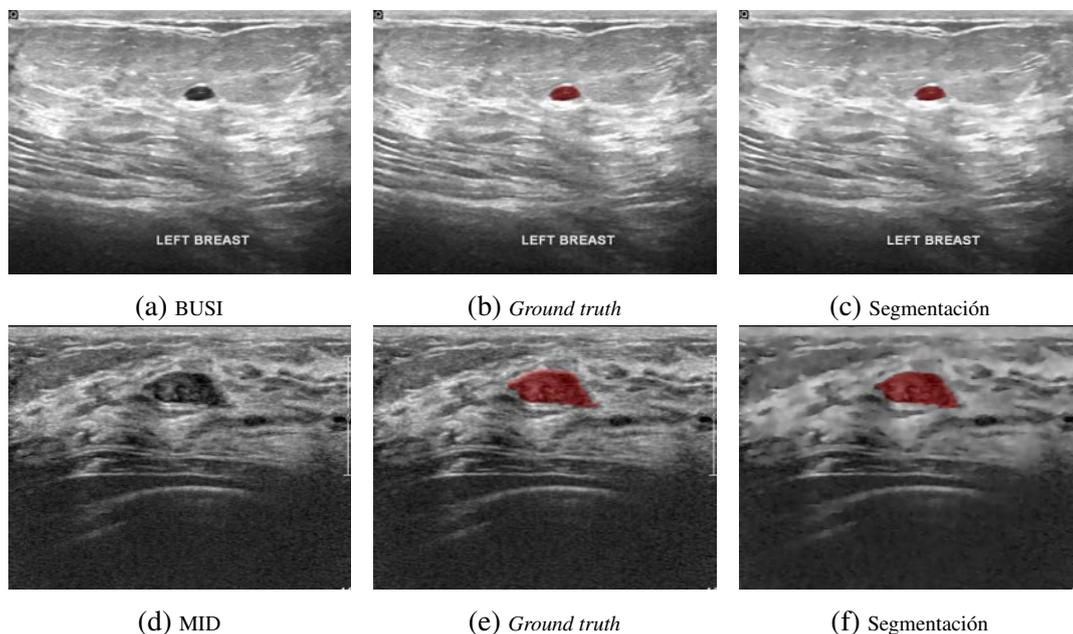


Figura 5.20 Resultados cualitativos en segmentación de masas sólidas benignas.

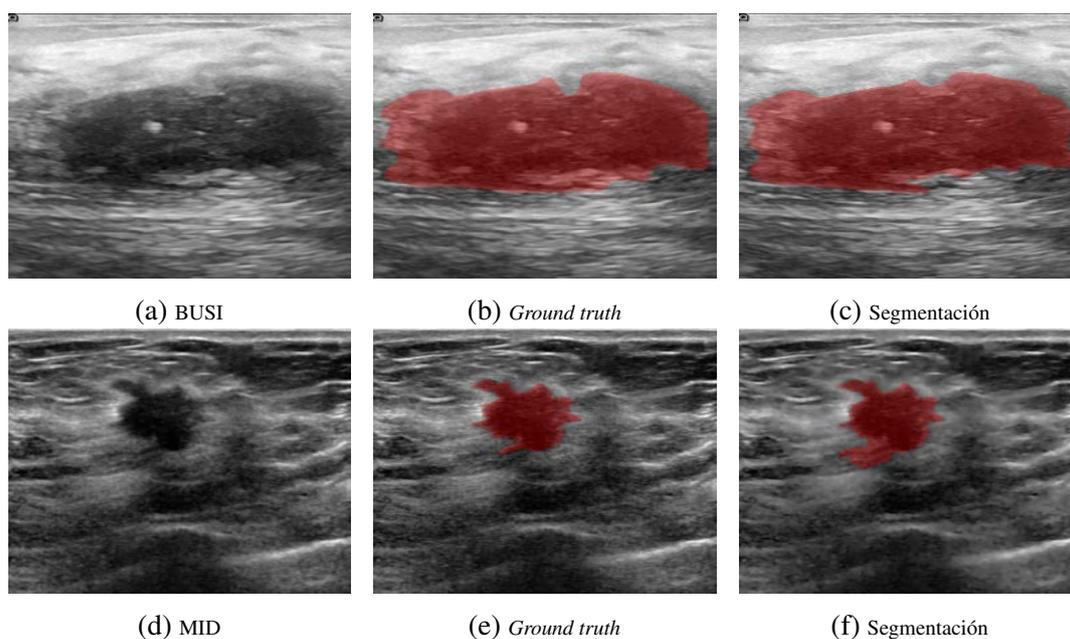


Figura 5.21 Resultados cuantitativos en segmentación de masas sólidas malignas.

Con base en los resultados de las Figuras 5.20 y 5.21 se puede apreciar como el sistema propuesto logra hacer segmentaciones casi similares al *ground truth* lo que se refleja en la precisión alcanzada en las métricas de la gráficas de la Figura 5.19

### Experimento 8. Comparación con algoritmo base DeepLabv3+

Para la realización de este experimento se seleccionaron imágenes de ecografías con masas sólidas de las bases de datos BUSI y MID. Los resultados se compararon con el modelo base del sistema DeepLabv3+. La Primera columna representa las masas de la base de datos BUSI y la segunda las masas de la base de datos MID.

En las Figuras 5.22 y 5.23 se muestra como el sistema propuesto mejora los resultados obtenidos con DeepLabv3+ lo que demuestra la importancia del preprocesamiento de los datos y reducción de la región de interés durante la segmentación. Así mismo se puede apreciar cómo el algoritmo supera hasta por un 10% al modelo original, donde en algunos casos este solo logra alcanzar puntuaciones de solo 0.8, mientras que el sistema propuesto se mantiene por encima de esta cifra tanto en masas benignas como malignas de ambas bases de datos.

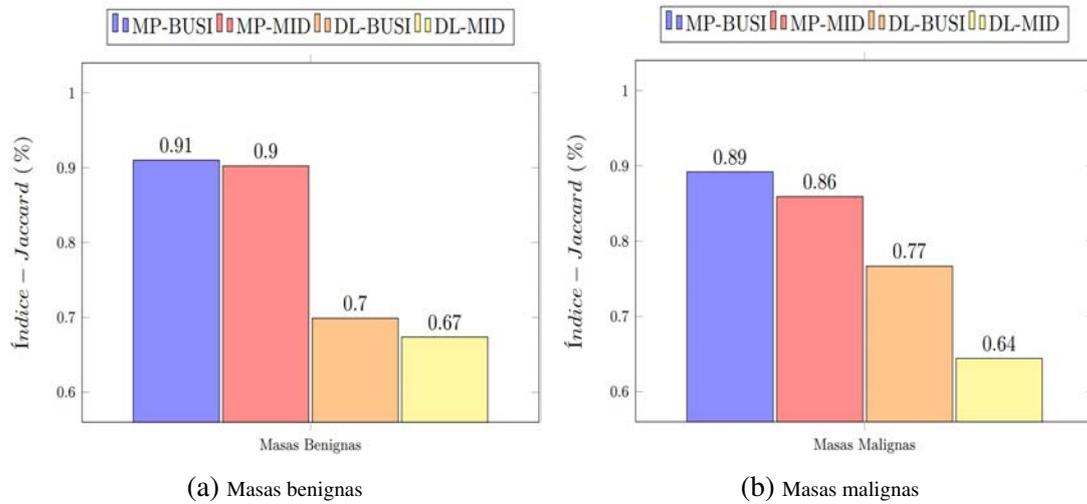


Figura 5.22 Índice Jaccard en base de datos BUSI y MID.

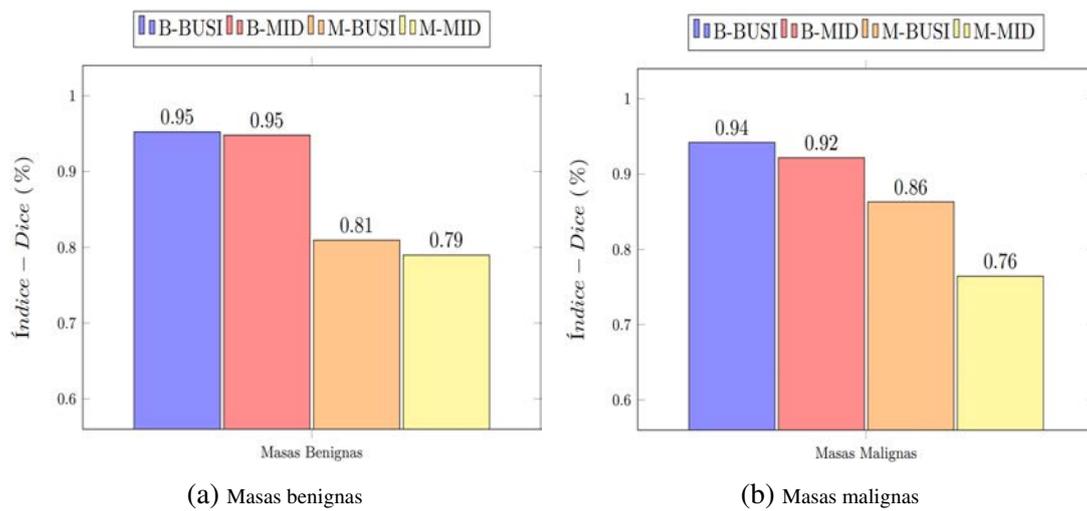


Figura 5.23 Índice Dice en base de datos BUSI y MID.

A manera de tener una mejor perspectiva de los resultados obtenidos durante esta comparación, en la Figura 5.24 y 5.25 se puede apreciar los resultados cualitativos obtenidos en este experimento. Donde la primera columna es la imagen original, la segunda el *ground truth* de la masa, la tercera la segmentación con DeepLabv3+ y la cuarta con el sistema propuesto.

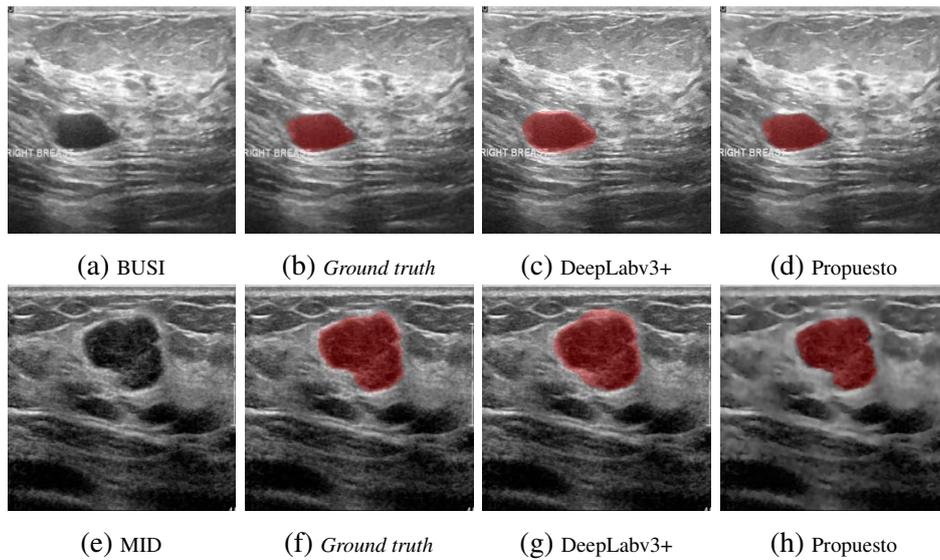


Figura 5.24 Comparación de segmentación de masas benignas con el sistema propuesto y DeepLabv3+.

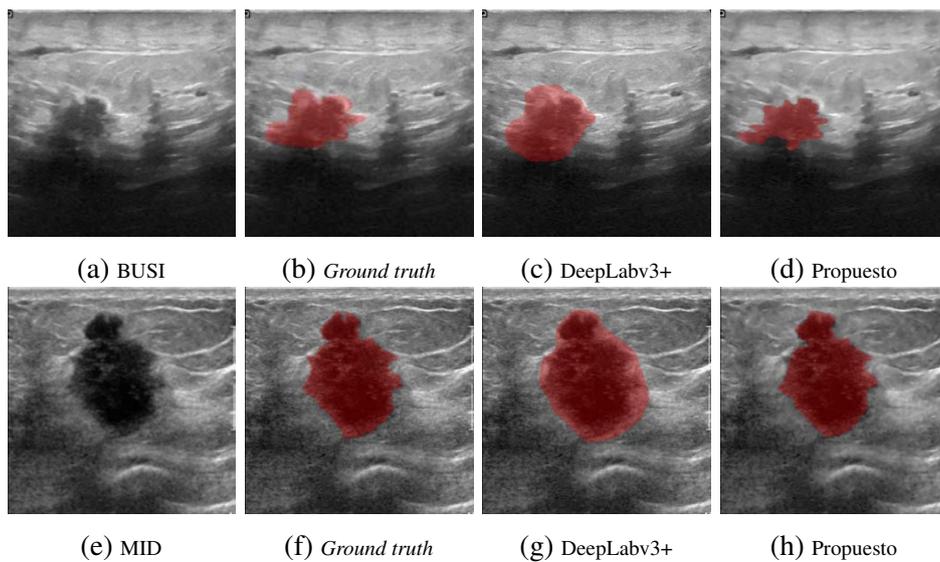


Figura 5.25 Comparación de segmentación de masas malignas con el sistema propuesto y DeepLabv3+.

## Experimento 9. Comparación de Resultados con métodos del Estado del Arte

Para este experimento se realizó la comparación de los resultados obtenidos con trabajos del estado del arte respecto a las métricas: *Jaccar-Index* (DI), *Sørensen-Dice* (DSC), *Precision* (P), *Recall* (R), *F1 – score* (F1) y *False Negative Rate* (FNR) de acuerdo a la Tabla 5.3.

Tabla 5.3 Comparación del rendimiento global en la tarea de segmentación de masas en ecografías entre el sistema propuesto y los modelos de la literatura.

Metodología	BD	DI	DSC	P	R	F1	FNR
CNN [129]	Propia	85.10%	n.a.	80.00%	80.00%	80.00%	n.a.
Dilated Fully Convolutional Network [55]	Propia	n.a.	88.97 ± 10.01%	n.a.	n.a.	n.a.	n.a.
ASFRRN [94]	Propia	83.3 ± 14.8%	90.5 ± 13.1%	n.a.	n.a.	90.5 ± 13.1%	n.a.
FTMTL-Net [42]	INBreast	n.a.	0.76 ± 0.03%	n.a.	0.91 ± 0.05%	n.a.	n.a.
U-Net mod. [107]	DDSM	n.a.	91.89%	n.a.	n.a.	n.a.	n.a.
	MIAS	n.a.	89.99%	n.a.	n.a.	n.a.	n.a.
	CBIS - DDSM	n.a.	88.65%	n.a.	n.a.	n.a.	n.a.
Umbral Adaptativo PDF [106]	MIAS	n.a.	n.a.	n.a.	97.7%	n.a.	n.a.
Mask R-CNN [128]	Inbreast	n.a.	89.10 ± 05.00%	n.a.	93.6 ± 0.63%	n.a.	n.a.
DeepLabv3+	BUSI	73.28 ± 6.79%	83.62 ± 5.37%	99.85 ± 0.28%	97.94 ± 1.94%	98.88 ± 1.13%	02.36 ± 2.55%
	MID	65.89 ± 2.97%	77.70 ± 2.53%	98.66 ± 0.37%	96.04 ± 0.64%	97.30 ± 0.54%	03.65 ± 1.25%
Método propuesto	BUSI	90.10 ± 1.89%	94.72 ± 1.04%	98.03 ± 3.58%	99.56 ± 0.45%	98.55 ± 2.50%	00.44 ± 0.45%
	MID	88.07 ± 4.32%	93.49 ± 2.63%	99.05 ± 1.02%	99.37 ± 0.26%	99.19 ± 0.66%	00.63 ± 0.26%

De acuerdo a los resultados de la tabla anterior se puede ver como el sistema propuesto supera a los demás modelos, lo que demuestra la eficiencia del sistema propuesto. En términos de DI el algoritmo alcanza un  $90.10 \pm 1.89\%$  mientras que el trabajo mejor puntuado de la tabla solo alcanza un  $85.10\%$ . Por otro lado, en términos de DSC el modelo obtiene un  $94.72 \pm 1.04\%$ , mientras que en métricas de P, R y F1 los resultados estuvieron por encima del  $98\%$  con un FNR de solo el  $00.44 \pm 0.45\%$ .

# Capítulo 6

## Conclusiones

En este capítulo, después de haber realizado las evaluaciones correspondientes, se hace un análisis de las conclusiones de la investigación a partir de los resultados obtenidos y teniendo en cuenta los objetivos planteados. Además, se describen los productos generados y las principales aportaciones. Por último, se complementa con ideas y sugerencias para trabajos futuros.

### 6.1 Objetivos y alcances logrados

Como aportaciones al tema de investigación propuesto en este trabajo segmentación de masas sólidas en imágenes de ecografías se detallan en la Tabla 6.1 en concordancia con los objetivos específicos establecidos y en la Tabla 6.2 de acuerdo a los alcances y limitaciones establecidos.

Tabla 6.1 Objetivos cumplidos.

Objetivo	Actividad
Investigar y comprender las técnicas de preprocesamiento de imágenes médicas y conceptos clave del aprendizaje profundo, específicamente en el contexto de filtrado, detección y segmentación de masas en imágenes de ultrasonidos.	Se realizó el estudio del marco teórico y estado del arte de técnicas de preprocesamiento de imágenes médicas enfocadas en el filtrado de ruido multiplicativo y mejora de la calidad de la imagen, así como métodos de aprendizaje profundo para la detección y segmentación de masas en imágenes de ultrasonidos.
Recopilar y analizar una base de datos de imágenes de ultrasonidos con masas sólidas mamarias.	Se recopiló y analizó 2 bases de datos de imágenes ecográficas con masas sólidas, así como sus respectivos <i>groun truth</i> para la evaluación del módulo de segmentación del sistema propuesto. Así mismo se recopiló una tercera base de datos con masas sólidas en ecografías para el entrenamiento del módulo de detección.
Realizar un preprocesamiento en las imágenes de ultrasonidos, incluyendo normalización, reducción de ruido y mejora de la calidad, para optimizar el rendimiento del sistema.	Se realizó el preprocesamiento de las imágenes de la base de datos propuesta mediante 6 filtros de reducción de ruido con la finalidad de reducir el ruido y mejorar los detalles en la imagen.
Diseñar y desarrollar un sistema basado en aprendizaje profundo adecuado para una segmentación precisa de masas sólidas en imágenes de ultrasonidos, considerando la reducción de ruido de la imagen, así como la detección y delimitación de los bordes de las masas.	Se desarrolló un sistema basado en aprendizaje profundo mediante 3 módulos: Módulo de filtrado, el cual reduce el ruido multiplicativo y mejorar los detalles en los bordes de la masa; Módulo de detección, el cual delimita la masa en una sola región de interés mediante un cuadro delimitador; y módulo de segmentación el cual toma las region previamente delimitadas y realiza la segmentación de la masa.
Entrenar y ajustar el sistema de aprendizaje profundo utilizando técnicas de entrenamiento y validación con el fin de lograr una detección y delimitación precisa de la masa.	Se entrenó y validó el módulo de detección del sistema propuesto con métricas de <i>Precisin, Recall, Specificity</i> y <i>F1 – score</i> con resultados de detección de masas por encima del 90%.
Evaluar el rendimiento del sistema propuesto mediante métricas como PSNR, SSIM, <i>Precisin, Recall, Specificity</i> , índice de Jaccar y coeficiente de similaridad de Dice, comparando los resultados obtenidos con técnicas del estado del arte.	Se evaluó el sistema con las métricas propuestas y se hizo una comparación de los resultados con 7 metodologías del estado del arte así como el algoritmo base DeepLabv3+.

Tabla 6.2 Alcances realizados.

Alcance	Actividad
Utilizar técnicas de reducción de ruido multiplicativo en las imágenes de la base de datos propuesta con la finalidad de mejorar su calidad.	Se realizó un filtrado de ruido multiplicativo en las imágenes de la base de datos propuesta y se publicó un artículo con los resultados obtenidos.
Desarrollar un sistema de aprendizaje profundo para la segmentación de masas sólidas en ecografías.	Se desarrolló un sistema de aprendizaje profundo basado en tres módulos para la segmentación de masas sólidas en ecografías.
Evaluar la calidad del desempeño del sistema propuesto con métricas definidas en la literatura.	Se evaluó los resultados obtenidos mediante el sistema propuesto con diferentes métricas de la literatura. Para filtrado de ruido multiplicativo: SNR, PSNR, MSE y SSIM. Para detección de RIO de la masa: <i>Precision, Recall, Specificity</i> y <i>F1 – score</i> . Y para la segmentación de la masa: <i>Jaccar-Index, Sørensen-Dice, Precision, Recall, F1 – score</i> y FNR.
Comparar los resultados obtenidos con modelos de segmentación existentes.	Se realizó una Tabla de comparación de resultados obtenidos con diferentes métodos existentes de acuerdo a las métricas de <i>Jaccar-Index, Sørensen-Dice, Precision, Recall, F1 – score</i> y <i>False Negative Rate</i> .

## 6.2 Resultados del trabajo

### 6.2.1 Productos

En esta investigación se ha considerado entregar como producto final lo siguiente:

- Reporte del estado del arte.
- Implementación del sistema propuesto.
- Presentación de Resultados.
- Reporte de resultados.
- Artículo de congreso.
- Constancia de ponencia.
- Documento de tesis.

## 6.3 Conclusiones

Conforme a los resultados obtenidos en los experimentos realizados, se puede deducir que el sistema propuesto cumple con el objetivo principal de la investigación, obtenidos resultados bastante favorables. De acuerdo a los experimentos, se puede deducir que para el módulo de filtrado, el cual utiliza el filtro *Speckle Reducing Anisotropic Diffusion* (SRAD) se demostró eficacia de esta técnica al reducir el ruido multiplicativo y resaltar los bordes de la masa en las imágenes ecográficas. Los resultados cuantitativos demuestran un valor promedio en PSNR de 38.535, MSE de 0.00016 y SSIM de 0.9479. Por otro lado, el módulo de detección mediante el algoritmo YOLOv3, demostró una gran capacidad para identificar y delimitar las regiones de interés que contenían las masas en las imágenes, logrando distinguir entre los diferentes tejidos masas malignas, la cual por su forma generalmente con complejas de detectar. Donde de acuerdo a las métricas propuestas se obtuvo resultados de *Precision* igual al 97%, *Recall* igual al 99%, *Specificity* igual al 90% y *F1 – score* igual a 98%. Y finalmente, para el módulo de segmentación basado en DeepLabv3+ y con ayuda de los módulos anteriores, se ha alcanzado resultados con un valor promedio en el índice de Jaccard igual a 0.89%, índice de similaridad de Dice igual a 0.94%, *Precision* de 0.98%, *Recall* de 0.99% y un *F1 – score* de 0.98% siendo en comparativa con la implementación del modelo original DeepLabv3+ más altos. Por ende, de acuerdo a los resultados, se concluye que la capacidad de DeepLabv3+ para llevar a cabo una segmentación precisa de la masa es gracias a las regiones de interés definidas por

YOLOv3 y el preprocesamiento de la imagen por parte del filtro SRAD. Lo cual se traduce en un FNR bajo, lo que demuestra la efectividad del sistema propuesto al minimizar los falsos negativos en la tarea de segmentación.

En resumen, los resultados obtenidos a través de la implementación del sistema propuesto son alentadores, la combinación los tres módulos desempeñan un papel crucial en el proceso de segmentación y provee de un mejor resultado que al utilizar solo la red de segmentación a las imágenes en bruto. Los hallazgos obtenidos a lo largo de los experimentos no solo contribuyen en el avance de la aplicación de técnicas de aprendizaje profundo en el análisis clínico, sino que también señalan posibles soluciones para la evolución y mejora de aplicaciones clínica de tipo de sistemas.

## **6.4 Trabajo futuro**

Como trabajo futuro derivado de este proyecto de investigación, se plantea experimentar con otro tipo de filtros para reducir el ruido multiplicativo, con la finalidad resaltar aún más los bordes de las masas, así mismo se plantea implementar otro tipo de red de detección de la masa más enfocada a imágenes médicas, el cual pueda clasificar la masa de acuerdo a su benignidad o malignidad. De igual manera, se considera reentrenar tanto el módulo de detección como el módulo de segmentación con otras bases de datos, todo esto con la finalidad de poder extraer un mayor número de características y así realizar una segmentación más precisa.

# Referencias

- [1] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., et al. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.
- [2] Afonso, M. V. and Sanches, J. M. R. (2015). Blind inpainting using gama and total variation regularization. *IEEE Transactions on Image Processing*, 24(7):2239–2253.
- [3] Al-Antari, M. A., Al-Masni, M. A., Choi, M.-T., Han, S.-M., and Kim, T.-S. (2018). A fully integrated computer-aided diagnosis system for digital x-ray mammograms via deep learning detection, segmentation, and classification. *International journal of medical informatics*, 117:44–54.
- [4] Al-Dhabyani, W., Gomaa, M., Khaled, H., and Fahmy, A. (2020). Dataset of breast ultrasound images. *Data in brief*, 28:104863.
- [5] Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., and Asari, V. K. (2018). Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*.
- [6] Anitha, J., Peter, J. D., and Pandian, S. I. A. (2017). A dual stage adaptive thresholding (dusat) for automatic mass detection in mammograms. *Computer methods and programs in biomedicine*, 138:93–104.
- [7] Argenti, F. and Torricelli, G. (2003). Speckle suppression in ultrasonic images based on undecimated wavelets. *EURASIP Journal on Advances in Signal Processing*, 2003:1–9.
- [8] Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495.
- [9] Beheshti, N. and Johnsson, L. (2020). Squeeze u-net: A memory and energy efficient image segmentation network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 364–365.
- [10] Benson, J. R., Gui, G. P., and Tuttle, T. (2013). *Early breast cancer: from screening to multidisciplinary management*. CRC Press.
- [11] Berzal, F. (2019). *Redes neuronales & deep learning: Volumen II*. Independently published.
- [12] Bounceur, A., Bezoui, M., and Euler, R. (2018). *Boundaries and hulls of Euclidean graphs: From theory to practice*. Chapman and Hall/CRC.
- [13] Brooks, S. E., Hembree, T. M., Shelton, B. J., Beache, S. C., Aschbacher, G., Schervish, P. H., and Dignan, M. B. (2013). Mobile mammography in underserved populations: analysis of outcomes of 3,923 women. *Journal of community health*, 38(5):900–906.

- [14] Buades, A., Coll, B., and Morel, J.-M. (2005a). A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65. IEEE.
- [15] Buades, A., Coll, B., and Morel, J.-M. (2005b). A review of image denoising algorithms, with a new one. *Multiscale modeling & simulation*, 4(2):490–530.
- [16] Bustos, O., Palacio, M., and Frery, A. (2002). Filtros interactivos reductores de ruido speckle en imagenes. *Revista de Teledeteccion*, 17:61–70.
- [17] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2014). Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*.
- [18] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017a). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848.
- [19] Chen, L.-C., Papandreou, G., Schroff, F., and Adam, H. (2017b). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.
- [20] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818.
- [21] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258.
- [22] Choukroun, Y., Bakalo, R., Ben-Ari, R., Akselrod-Ballin, A., Barkan, E., and Kisilev, P. (2017). Mammogram classification and abnormality detection from nonlocal labels using deep multiple instance neural network. In *VCBM*, pages 11–19.
- [23] Coupé, P., Hellier, P., Kervrann, C., and Barillot, C. (2008a). Bayesian non local means-based speckle filtering. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1291–1294. IEEE.
- [24] Coupé, P., Hellier, P., Kervrann, C., and Barillot, C. (2009). Nonlocal means-based speckle filtering for ultrasound images. *IEEE transactions on image processing*, 18(10):2221–2229.
- [25] Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., and Barillot, C. (2008b). An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images. *IEEE transactions on medical imaging*, 27(4):425–441.
- [26] Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2007). Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095.
- [27] Dhungel, N., Carneiro, G., and Bradley, A. P. (2015a). Automated mass detection in mammograms using cascaded deep learning and random forests. In *2015 international conference on digital image computing: techniques and applications (DICTA)*, pages 1–8. IEEE.

- [28] Dhungel, N., Carneiro, G., and Bradley, A. P. (2015b). Deep learning and structured prediction for the segmentation of mass in mammograms. In *International Conference on Medical image computing and computer-assisted intervention*, pages 605–612. Springer.
- [29] Dhungel, N., Carneiro, G., and Bradley, A. P. (2015c). Deep structured learning for mass segmentation from mammograms. In *2015 IEEE international conference on image processing (ICIP)*, pages 2950–2954. IEEE.
- [30] Dhungel, N., Carneiro, G., and Bradley, A. P. (2017a). A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Medical image analysis*, 37:114–128.
- [31] Dhungel, N., Carneiro, G., and Bradley, A. P. (2017b). Fully automated classification of mammograms using deep residual neural networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 310–314. IEEE.
- [32] DiSaia, P. J., Creasman, W. T., Mannel, R. S., McMeekin, D. S., and Mutch, D. G. (2018). *Oncología ginecológica clínica*. Elsevier Health Sciences.
- [33] Domínguez, D. (2019). *Extracción de Superpíxeles y Fusión Local Mediante un Algoritmo de Agrupamiento Difuso Intuitivo*. Tesis, Centro Nacional de Investigación y Desarrollo Tecnológico.
- [34] Dong, F. and Li, N. (2021). Multiplicative noise removal and blind inpainting of ultrasound images based on a new variational framework. *Machine Vision and Applications*, 32(4):1–15.
- [35] Du, J. (2018). Understanding of object detection based on cnn family and yolo. In *Journal of Physics: Conference Series*, volume 1004, page 012029. IOP Publishing.
- [36] Farouj, Y., Freyermuth, J.-M., Navarro, L., Clausel, M., and Delachartre, P. (2016). Hyperbolic wavelet-fisz denoising for a model arising in ultrasound imaging. *IEEE Transactions on Computational Imaging*, 3(1):1–10.
- [37] Foi, A. and Boracchi, G. (2016). Foveated nonlocal self-similarity. *International Journal of Computer Vision*, 120:78–110.
- [38] Fornage, B. D. (1994). Interventional ultrasound of the breast. In *Breast Ultrasound Update*, pages 260–264. Karger Publishers.
- [39] Frost, V. S., Stiles, J. A., Shanmugan, K. S., and Holtzman, J. C. (1982). A model for radar images and its application to adaptive digital filtering of multiplicative noise. *IEEE Transactions on pattern analysis and machine intelligence*, (2):157–166.
- [40] Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., and Lu, H. (2019). Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3146–3154.
- [41] Gagnon, L. and Jouan, A. (1997). Speckle filtering of sar images: a comparative study between complex-wavelet-based and standard filters. In *Wavelet Applications in Signal and Image Processing V*, volume 3169, pages 80–91. SPIE.

- [42] Gao, F., Yoon, H., Wu, T., and Chu, X. (2020). A feature transfer enabled multi-task deep learning model on medical imaging. *Expert Systems with Applications*, 143:112957.
- [43] Gao, G. (2010). Statistical modeling of sar images: A survey. *Sensors*, 10(1):775–795.
- [44] Geertsma, T. (2022). Ultrasound cases.
- [45] Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448.
- [46] Girshick, R., Donahue, J., Darrell, T., Malik, J., and Mercan, E. (2014). R-cnn for object detection. In *IEEE Conference*.
- [47] Gong, Y., Liu, B., Hou, X., and Qiu, G. (2018). Sub-window box filter. In *2018 IEEE Visual Communications and Image Processing (VCIP)*, pages 1–4. IEEE.
- [48] Grauman, K. and Darrell, T. (2005). The pyramid match kernel: Discriminative classification with sets of image features. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 1458–1465. IEEE.
- [49] He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969.
- [50] He, K., Sun, J., and Tang, X. (2012). Guided image filtering. *IEEE transactions on pattern analysis and machine intelligence*, 35(6):1397–1409.
- [51] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- [52] Holschneider, M., Kronland-Martinet, R., Morlet, J., and Tchamitchian, P. (1990). A real-time algorithm for signal analysis with the help of the wavelet transform. In *Wavelets: Time-Frequency Methods and Phase Space Proceedings of the International Conference, Marseille, France, December 14–18, 1987*, pages 286–297. Springer.
- [53] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- [54] Hu, K., Gao, X., and Li, F. (2010). Detection of suspicious lesions by adaptive thresholding based on multiresolution analysis in mammograms. *IEEE Transactions on Instrumentation and Measurement*, 60(2):462–472.
- [55] Hu, Y., Guo, Y., Wang, Y., Yu, J., Li, J., Zhou, S., and Chang, C. (2019). Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model. *Medical physics*, 46(1):215–228.
- [56] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- [57] Institute, N. C. (2021). National cancer institute.

- [58] Jansons, K. M. and Alexander, D. C. (2003). Persistent angular structure: new insights from diffusion magnetic resonance imaging data. *Inverse problems*, 19(5):1031.
- [59] Jin, Z. and Yang, X. (2011). A variational model to remove the multiplicative noise in ultrasound images. *Journal of Mathematical Imaging and Vision*, 39(1):62–74.
- [60] Kaplan, D. and Ma, Q. (1994). On the statistical characteristics of log-compressed rayleigh signals: Theoretical formulation and experimental results. *The Journal of the Acoustical Society of America*, 95(3):1396–1400.
- [61] Kopans, D. B. (2007). Breast imaging. *Philadelphia, Pa*, 1998:55–100.
- [62] Kozegar, E., Soryani, M., Minaei, B., and Domingues, I. (2013). Assessment of a novel mass detection algorithm in mammograms. *Journal of cancer research and therapeutics*, 9(4):592–600.
- [63] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- [64] Kuan, D. T., Sawchuk, A. A., Strand, T. C., and Chavel, P. (1985). Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE transactions on pattern analysis and machine intelligence*, (2):165–177.
- [65] Lazebnik, S., Schmid, C., and Ponce, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, volume 2, pages 2169–2178. IEEE.
- [66] Lee, J.-S. (1980). Digital image enhancement and noise filtering by use of local statistics. *IEEE transactions on pattern analysis and machine intelligence*, (2):165–168.
- [67] Lee, R. S., Gimenez, F., Hoogi, A., and Rubin, D. (2016). Curated breast imaging subset of dds. *The cancer imaging archive*, 8:2016.
- [68] Liang, D., Pan, J., Yu, Y., and Zhou, H. (2019). Concealed object segmentation in terahertz imaging via adversarial learning. *Optik*, 185:1104–1114.
- [69] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125.
- [70] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2016). Ssd: Single shot multibox detector. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 21–37. Springer.
- [71] Liu, W., Rabinovich, A., and Berg, A. C. (2015). Parsenet: Looking wider to see better. *arXiv preprint arXiv:1506.04579*.
- [72] Liu, X., Song, L., Liu, S., and Zhang, Y. (2021). A review of deep-learning-based medical image segmentation methods. *Sustainability*, 13(3):1224.

- [73] Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.
- [74] Lu, L., Zheng, Y., Carneiro, G., and Yang, L. (2017). Deep learning and convolutional neural networks for medical image computing. *Advances in computer vision and pattern recognition*, 10:978–3.
- [75] Makhanov, P. S. (2001). Medical image database. Head Biomedical Engineering Research Unit.
- [76] Mathworks (N.D.). Deep learning: Tres cosas que es necesario saber.
- [77] Matuz, M. (2017). *Detección de Anormalidades en Mamografías Utilizando la Red Neuronal Convolutiva AlexNet*. Tesis, Centro Nacional de Investigación y Desarrollo Tecnológico.
- [78] Mei, F., Zhang, D., and Yang, Y. (2020). Improved non-local self-similarity measures for effective speckle noise reduction in ultrasound images. *Computer methods and programs in biomedicine*, 196:105670.
- [79] Min, H., Wilson, D., Huang, Y., Liu, S., Crozier, S., Bradley, A. P., and Chandra, S. S. (2020). Fully automatic computer-aided mass detection and segmentation via pseudo-color mammograms and mask r-cnn. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1111–1115. IEEE.
- [80] Morales, M. (2018). *Segmentación no Paramétrica de Tejidos Cerebrales Mediante una Arquitectura Paralela de Redes Neuronales Convolutivas*. Tesis, Centro Nacional de Investigación y Desarrollo Tecnológico.
- [81] Moreira, I. C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M. J., and Cardoso, J. S. (2012). Inbreast: toward a full-field digital mammographic database. *Academic radiology*, 19(2):236–248.
- [82] Noh, H., Hong, S., and Han, B. (2015). Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1520–1528.
- [83] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., et al. (2018). Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.
- [84] Oliver Malagelada, A. et al. (2007). *Automatic mass segmentation in mammographic images*. Universitat de Girona.
- [85] Padmavathy, T., Vimalkumar, M., Nagarajan, S., Babu, G. C., and Parthasarathy, P. (2021). Performance analysis of pre-cancerous mammographic image enhancement feature using non-sampled shearlet transform. *Multimedia Tools and Applications*, 80(18):26997–27012.
- [86] PAHO, T. P. A. H. O. (2021). World cancer day 2021: I am and i will.
- [87] Parekh, A. and Selesnick, I. W. (2016). Enhanced low-rank matrix approximation. *IEEE Signal Processing Letters*, 23(4):493–497.

- [88] Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on pattern analysis and machine intelligence*, 12(7):629–639.
- [89] Pisano, E. D., Yaffe, M. J., and Kuzmiak, C. M. (2004). *Digital mammography*. Lippincott Williams & Wilkins.
- [90] Plazas, J. A. P., Hernández, J. T., and De la Rosa, M. F. (2015). Estudio de reducción del ruido en imágenes sar para el análisis de inundaciones. *Ingenium Revista de la facultad de ingeniería*, 16(32):9–18.
- [91] Pohlen, T., Hermans, A., Mathias, M., and Leibe, B. (2017). Full-resolution residual networks for semantic segmentation in street scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4151–4160.
- [92] Qi, H., Zhang, Z., Xiao, B., Hu, H., Cheng, B., Wei, Y., and Dai, J. (2017). Deformable convolutional networks—coco detection and segmentation challenge 2017 entry. In *ICCV COCO Challenge Workshop*, volume 15, pages 764–773.
- [93] Qidwai, U. and Chen, C.-h. (2009). *Digital image processing: an algorithmic approach with MATLAB*. Chapman and Hall/CRC.
- [94] Qu, X., Shi, Y., Hou, Y., and Jiang, J. (2020). An attention-supervised full-resolution residual network for the segmentation of breast ultrasound images. *Medical physics*, 47(11):5702–5714.
- [95] Rahimizadeh, N., Hasanzadeh, R. P., and Janabi-Sharifi, F. (2021). An optimized non-local Immse approach for speckle noise reduction of medical ultrasound images. *Multimedia Tools and Applications*, 80:9231–9253.
- [96] Ramos-Llordén, G., Vegas-Sánchez-Ferrero, G., Martín-Fernandez, M., Alberola-López, C., and Aja-Fernández, S. (2014). Anisotropic diffusion filter with memory based on speckle statistics for ultrasound images. *IEEE transactions on Image Processing*, 24(1):345–358.
- [97] Redmon, J. (2013–2016). Darknet: Open source neural networks in c. <http://pjreddie.com/darknet/>.
- [98] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788.
- [99] Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
- [100] Rincón, V. V. V. (2022). *Agrupamiento difuso titubeante para tareas de segmentación de imágenes y reconocimiento de patrones*. Tesis, Centro Nacional de Investigación y Desarrollo Tecnológico.
- [101] Rodríguez López, V. et al. (2012). Análisis de imágenes de mamografía para la detección de cáncer de mama. *REPOSITORIO NACIONAL CONACYT*.

- [102] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer.
- [103] Rovere, G. d., Warren, R., and Benson, J. (2006). Early breast cancer from screening to multidisciplinary management. *Taylor & Francis Group, Florida, USA*.
- [104] Ruan, H., Tan, Z., Chen, L., Wan, W., and Cao, J. (2022). Efficient sub-pixel convolutional neural network for terahertz image super-resolution. *Optics Letters*, 47(12):3115–3118.
- [105] Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268.
- [106] Sadeghi, B., Karimi, M., and Mazaheri, S. (2021). Automatic suspicious lesions segmentation based on variable-size windows in mammography images. *Health and Technology*, 11(1):99–110.
- [107] Salama, W. M. and Aly, M. H. (2021). Deep learning in mammography images segmentation and classification: Automated cnn approach. *Alexandria Engineering Journal*, 60(5):4701–4709.
- [108] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- [109] Scully, O. J., Bay, B.-H., Yip, G., and Yu, Y. (2012). Breast cancer metastasis. *Cancer genomics & proteomics*, 9(5):311–320.
- [110] Sehgal, C. M., Weinstein, S. P., Arger, P. H., and Conant, E. F. (2006). A review of breast ultrasound. *Journal of mammary gland biology and neoplasia*, 11(2):113–123.
- [111] Seif, G. (2019). Semantic segmentation with deep learning: A guide and code. *Medium*.
- [112] Sencha, A. N., Evseeva, E. V., Mogutov, M. S., and Patrunov, Y. N. (2014). *Breast ultrasound*. Springer Science & Business Media.
- [113] Sheng, Y. and Xia, Z.-G. (1996). A comprehensive evaluation of filters for radar speckle suppression. In *IGARSS'96. 1996 International Geoscience and Remote Sensing Symposium*, volume 3, pages 1559–1561. IEEE.
- [114] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [115] Skaane, P. (2009). Studies comparing screen-film mammography and full-field digital mammography in breast cancer screening: updated review. *Acta radiologica*, 50(1):3–14.
- [116] Society, A. C. (2022). Cancer.org. <https://www.cancer.org>. 12 mayo 2022.
- [117] Solomon, C. and Breckon, T. (2011). *Fundamentals of Digital Image Processing: A practical approach with examples in Matlab*. John Wiley & Sons.

- [118] Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S., et al. (2015). Mammographic image analysis society (mias) database v1. 21.
- [119] Sudeep, P., Palanisamy, P., Rajan, J., Baradaran, H., Saba, L., Gupta, A., and Suri, J. S. (2016). Speckle reduction in medical ultrasound images using an unbiased non-local means method. *Biomedical Signal Processing and Control*, 28:1–8.
- [120] Sun Yat-sen, U. C. (2023). Sun yat-sen university cancer center. <https://www.uicc.org/membership/sun-yat-sen-university-cancer-center>. Octubre 2023.
- [121] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- [122] Tarver, T. (2012). Cancer facts & figures 2012. american cancer society (acs) atlanta, ga: American cancer society, 2012. 66 p., pdf.
- [123] Thompson, M. (1981). Digital image processing by rafael c. gonzalez and paul wintz. *Leonardo*, 14(3):256–257.
- [124] Tian, J. and Chen, L. (2011). Image despeckling using a non-parametric statistical model of wavelet coefficients. *Biomedical Signal Processing and Control*, 6(4):432–437.
- [125] Touzi, R. (2002). A review of speckle filtering in the context of estimation theory. *IEEE Transactions on Geoscience and Remote Sensing*, 40(11):2392–2404.
- [126] Tur, M., Chin, K.-C., and Goodman, J. W. (1982). When is speckle noise multiplicative? *Applied optics*, 21(7):1157–1159.
- [127] UDIAT, D. C. (2023). Udiat diagnostic center. <https://www.tauli.cat/en/udiat/>. Octubre 2023.
- [128] Viegas, L., Domingues, I., and Mendes, M. (2021). Study on data partition for delimitation of masses in mammography. *Journal of Imaging*, 7(9):174.
- [129] Xu, Y., Wang, Y., Yuan, J., Cheng, Q., Wang, X., and Carson, P. L. (2019). Medical breast ultrasound image segmentation by machine learning. *Ultrasonics*, 91:1–9.
- [130] Yin, X.-X., Sun, L., Fu, Y., Lu, R., Zhang, Y., et al. (2022). U-net-based medical image segmentation. *Journal of Healthcare Engineering*, 2022.
- [131] Yu, F. and Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*.
- [132] Yu, Y. and Acton, S. T. (2002). Speckle reducing anisotropic diffusion. *IEEE Transactions on image processing*, 11(11):1260–1270.
- [133] Zhang, J. and Cheng, Y. (2020). *Despeckling Methods for Medical Ultrasound Images*. Springer.
- [134] Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. (2017). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155.

- 
- [135] Zhang, R., Zhang, H., and Chung, A. C. (2018). A unified mammogram analysis method via hybrid deep supervision. In *Image Analysis for Moving Organ, Breast, and Thoracic Images: Third International Workshop, RAMBO 2018, Fourth International Workshop, BIA 2018, and First International Workshop, TIA 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16 and 20, 2018, Proceedings 3*, pages 107–115. Springer.
- [136] Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890.
- [137] Zhu, W. and Xie, X. (2016). Adversarial deep structural networks for mammographic mass segmentation. *Biorxiv*, page 095786.

# Apéndice A

## Producción

Se realizó la publicación y presentación del artículo “Análisis comparativo de técnicas de filtrado para reducción de ruido en imágenes de Ecografía”, en el 12° Congreso Internacional de Computación (CICOM 2022) en la Ciudad de México, el 7 de octubre de 2022 por la universidad Autónoma de Guerrero. La Figura A.1 muestra la constancia de presentación y la Figura A.2 la primera página del artículo publicado.



Figura A.1 Constancia de ponencia.



## Análisis comparativo de técnicas de filtrado para reducción de ruido en imágenes de Ecografía

Ing. Alexis Vázquez Ramírez  
TecNM/ CENIDET  
Depto. de Ciencias Computacionales  
+52 962 107 9387  
alexxis.97@outlook.com

M. C. Arturo Rendon Castro  
TecNM/ CENIDET  
Depto. de Ciencias  
Computacionales  
angel.rendon18ce@cenidet.edu.mx

Dr. Dante Mújica Vargas  
TecNM/ CENIDET  
Depto. de Ciencias Computacionales  
dante.mv@cenidet.tecnm.mx

M. C. Manuel de Jesús Matuz  
Cruz  
TecNM Campus Tapachula  
Depto. de Sistemas  
Computacionales  
mjmatuz@tapachula.tecnm.mx

### RESUMEN

Las imágenes por ultrasonido son utilizadas en gran medida como una herramienta de diagnóstico médico. Sin embargo, este tipo de imágenes están corrompidas por ruido y no se tiene conocimiento a priori de su densidad, este oculta características importantes, por tanto, su reducción es esencial para un diagnóstico preciso y confiable. Este trabajo se propone la comparación de algoritmos de filtrado para la reducción de ruido *speckle* en ecografías de mama. Se utilizan los filtros de difusión anisotrópica, filtro de imagen guiado, filtro *Non-local Means*, filtro *Frost* y filtro *Lee*, para la reducción de ruido y preservación de bordes en imágenes médicas de ultrasonido, obtenidas de las bases de datos Medical Image Database del Biomedical Engineering Research Unit of SIIT and Breast Ultrasound Images Dataset. De acuerdo a las métricas de PSNR, MSE y SSIM, el filtro con mejores los resultados obtenidos fue el filtro de difusión anisotrópica seguido del filtro de imagen guiado.

### ABSTRACT

Ultrasound images are used to a large extent as a medical diagnostic tool. However, this type images are corrupted with noise and there is no prior knowledge of their density, this hides important features, therefore, their reduction is essential for a precise and accurate diagnosis. This paper proposes a comparison of filtering algorithms for speckle noise reduction in breast ultrasound. Anisotropic

El permiso para hacer copias digitales o impresas en parte o en la totalidad de este artículo, se otorga sin tener que cubrir una contribución financiera, siempre y cuando sea para uso personal o en el aula, que las copias que se realicen o se distribuyan no sean con fines de lucro o ventaja comercial y que las copias conserven este aviso y los datos de la cita completa en la primera página. Para otro tipo de copias, o volver a publicar el artículo, para almacenarlos en servidores o redistribuirlos en listas de correo, se requiere una autorización previa de los autores y/o una posible cuota financiera.

12o. Congreso Internacional de Computación CICOM 2022, (del 5 al 7 de octubre del 2022), Sede: Ciudad de México. Copyright 2022 Universidad Autónoma de Guerrero.

diffusion filter, Guided image filter, Non-local Means filter, Frost

filter and Lee filter are used for the noise reduction and edge preservation of medical ultrasound images from the Medical Image Database of the Biomedical Engineering Research Unit of SIIT and Breast Ultrasound Images Dataset. According to PSNR, MSE and SSIM metrics, the best performing filters were the Anisotropic diffusion filter followed by Guided image filter.

### Categorías y Descriptores Temáticos

**Artificial vision:** Image filtering, Image processing, Denoising.

**Visión artificial:** Filtrado de imágenes, procesamiento de imágenes, reducción de ruido.

### Términos Generales

Visión artificial, Procesamiento de imágenes, Programación.

### Palabras clave

*Speckle*, Ecografía, Filtro Difusión Anisotrópica, Filtro de Imagen Guiado, Filtro de *Non-local Means*, Filtro *Frost*, Filtro *Lee*.

### Keywords

Speckle, Ultrasound, Anisotropic Diffusion Filter, Guided Image Filter, Non-local Means Filter, Frost Filter, Lee Filter.

### INTRODUCCIÓN

A diferencia del ruido adicional, este ruido multiplicativo o *speckle* es difícil de reducir, debido a sus características proliferante y sus distribuciones típicamente no gaussianas [1-4]. En un modelo de ruido multiplicativo, una imagen grabada  $f$  definida en un rectángulo  $\alpha \subset \mathbb{R}^2$ , es la multiplicación de una imagen original  $u$  y un ruido  $n$ , suponiendo que  $u$  y  $n$  son positivos [2],[3]:

$$f = un \quad (1)$$

La importancia de la eliminación de ruido radica en la mejora de la precisión durante el reconocimiento visual humano. En el caso de las imágenes de ultrasonido, dicho ruido produce incertidumbre en el diagnóstico, debido a la pérdida de los detalles, como bordes

Figura A.2 Artículo de algoritmos de filtrado.

# Apéndice B

## Retribución social

Se impartió un taller sobre técnicas de clasificación y delimitación en el CICOM 2022 los días 5, 6 y 7 de octubre de 2022. La Figura B.1 muestra la constancia obtenida.



Figura B.1 Constancia de taller.