



Institución Universitaria

**Síntesis de imágenes de mama con
Deep Learning para la ayuda al
diagnóstico de cáncer**

Sara Cañaverál Uribe

Instituto Tecnológico Metropolitano
Facultad de Ingeniería
Medellín, Colombia
2023

Síntesis de imágenes de mama con Deep Learning para la ayuda al diagnóstico de cáncer

Sara Cañaverál Uribe

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:
Magíster en Automatización y Control Industrial

Directores:

Ph.D Carlos Andrés Mera Banguero

Ph.D(c) Rubén Darío Fonnegra

Línea de Investigación:

Procesamiento de datos de alta dimensión

Grupo de Investigación:

Máquinas Inteligentes y Reconocimiento de Patrones – MIRP

Instituto Tecnológico Metropolitano

Facultad de Ingeniería

Medellin, Colombia

2023

”No iba a ser fácil. Nada importante lo es.”

Los asesinos del emperador
Santiago Posteguillo

A mis padres y hermanos, que tuvieron el amor y la paciencia para acompañarme en este proceso. Y a Kevin Osorno y Juan Pablo Restrepo, que me brindaron su guía y conocimiento. Sin todos ellos, no hubiese sido posible.

Agradecimientos

Quiero agradecer a mi familia, quienes entendieron mis ausencias, quienes me acompañaron en los malos y buenos momentos que tuve durante el desarrollo de la investigación. Adicional, quiero agradecer al laboratorio Máquinas Inteligentes y Reconocimiento de Patrones (MIRP) del Instituto Tecnológico Metropolitano, por la vinculación al proyecto de investigación titulado *”Generación de respuesta sintética a la aplicación de medios de contraste en imágenes médicas para mejorar el diagnóstico automático de cáncer de mama”*; y al Instituto Universitario Pascual Bravo por facilitar los recursos tecnológicos para el desarrollo de la investigación. Agradezco al Ministerio de Ciencias por la financiación del proyecto en el que se enmarca esta investigación. Además agradezco a los tutores de la tesis Ph.D Carlos Andrés Mera Banguero y Ph.D(c) Rubén Darío Fonnegra.

Resumen

El cáncer de mama es una de las principales causas de muerte en mujeres en el mundo, por lo que su detección de forma temprana se ha convertido en una prioridad para salvar vidas. Para el diagnóstico de este tipo de cáncer existen técnicas instrumentales como la mamografía y la resonancia magnética contrastada (DCE-MRI). La resonancia magnética usa un medio de contraste que permite realzar anomalías en el tejido de la mama y apoya la detección y caracterización de tumores que pueden estar asociados al desarrollo de cáncer. Generalmente, los estudios que requieren el uso de medios contrastados se emplean en mujeres que tienen una alta probabilidad de desarrollar cáncer de mama o en aquellas en las que otros métodos no son concluyentes, esto se debe a las limitaciones que presenta, entre las que se encuentra el alto costo que de estos exámenes y la poca disponibilidad de personal y equipos para realizarlos. También, se ha encontrado que los medios de contraste pueden generar efectos adversos por reacciones alérgicas, además, algunos estudios han mostrado rastros de los medios de contraste alojados en la base del cerebro, cuando su uso es constante. Considerando lo anterior, este trabajo propone un modelo basado en aprendizaje profundo, que considerando las regiones de realce de contraste, genera imágenes sintéticas de DCE-MRI postcontraste a partir de imágenes precontraste, lo que puede apoyar la disminución del uso de medios contrastados en los exámenes diagnósticos de cáncer de mama. Para evaluar la eficacia del modelo propuesto se usaron las métricas de evaluación comúnmente usadas en el estado del arte (PSNR, SSIM, y MAE). Los resultados obtenidos evidencian que el modelo preserva la mayor similitud entre métricas basadas en píxeles para las imágenes sintéticas y las imágenes reales en comparación con los modelos evaluados de base en el estado del arte.

Palabras clave: Cáncer de mama, imagen diagnóstica, síntesis de imagen, redes neuronales.

Abstract

Breast cancer is one of the leading causes of death in women worldwide, so its early detection has become a priority to save lives. For the diagnosis of this type of cancer there are instrumental techniques such as mammography and contrasted magnetic resonance imaging (DCE-MRI). MRI uses a contrast medium to enhance abnormalities in breast tissue and supports the detection and characterization of tumors that may be associated with the development of cancer. Generally, studies that require the use of contrast media are used in women who have a high probability of developing breast cancer or in those in whom other methods are not conclusive, this is due to the limitations it presents, among which is the high cost of these tests and the limited availability of personnel and equipment to perform them. Also, it has been found that contrast media can generate adverse effects due to allergic reactions, in addition, some studies have shown traces of contrast media lodged in the base of the brain, when its use is constant. Considering the above, this work proposes a model based on deep learning, which considering contrast enhancement regions, generates synthetic post-contrast DCE-MRI images from pre-contrast images, which can support the decrease in the use of contrasted media in breast cancer diagnostic examinations. To evaluate the efficacy of the proposed model, commonly used state-of-the-art evaluation metrics (PSNR, SSIM, and MAE) were used. The results obtained evidence that the model preserves the highest similarity between pixel-based metrics for synthetic images and real images compared to the models evaluated on a state of the art.

Contenido

Agradecimientos	7
Resumen	9
1 Introducción	13
1.1 Objetivos	17
1.1.1 Objetivo General	17
1.1.2 Objetivos Específicos	17
1.2 Alcances	17
2 Marco Teórico y Estado del arte	18
2.1 Marco teórico	18
2.1.1 El cáncer de mama y su diagnóstico	18
2.1.2 Síntesis de imagen médica	20
2.1.3 Aprendizaje profundo para síntesis de imágenes médicas	21
2.2 Estado del arte en síntesis de imágenes médicas de mama	29
2.2.1 Aumento de datos	29
2.2.2 Superresolución	31
2.2.3 Traducción de imágenes o cambio de dominio	32
3 Desarrollo Metodológico	41
3.1 Base de datos	41
3.2 Modelos de línea base	42
3.2.1 RiedNet: Residual Inception Encoder Decoder Neural Network	42
3.2.2 Pix2Pix	43
3.2.3 Ea-GAN: Edge-Aware Generative Adversarial Networks	46
3.3 Modelos ensamblados	48
3.3.1 RiedGAN: ensamble entre RiedNet y Pix2Pix	48
3.3.2 gEa-RiedGAN y dEa-RiedGAN: dos ensambles entre Ea-GAN y RiedGAN	49
3.4 Modelos propuestos	50
3.4.1 Detección de zonas con realce de contraste	51
3.4.2 Función de pérdida para las regiones con realce de contraste	52
3.4.3 G-RiedGAN	54

3.4.4	D-RiedGAN	55
4	Resultados y Discusión	58
4.1	Métricas de evaluación	58
4.1.1	Error Medio Absoluto (MAE)	58
4.1.2	Medida de Proporción Señal a Ruido (PSNR)	58
4.1.3	Índice de Medida de Similitud Estructural (SSIM)	59
4.1.4	Mapas de diferencia	59
4.2	Configuración experimental	60
4.3	Evaluación comparativa	61
5	Conclusiones y Trabajo Futuro	67
6	Anexos	70
	Referencias	72

1 Introducción

El cáncer es una enfermedad crónica no transmisible y uno de los principales problemas de investigación en salud pública en el mundo debido a sus altas tasas de incidencia [1]. En general, esta enfermedad se puede dar en diferentes lugares del cuerpo, y se debe a alteraciones a nivel del ADN, las cuales afectan la división y crecimiento normal de las células de un tejido en específico.

Es importante entender que las células tienen un proceso de proliferación habitual, y, aunque se presentan alteraciones, el proceso de proliferación no se ve interrumpido, sino que se proliferan como lo haría una célula sana, por lo tanto se forman tumores malignos que degradan la salud de quienes los padecen causando la muerte en algunos de los casos. En la Figura 1-1 se visualiza el proceso de multiplicación de células malignas luego de que una célula sana tuvo un cambio dentro de su ADN.

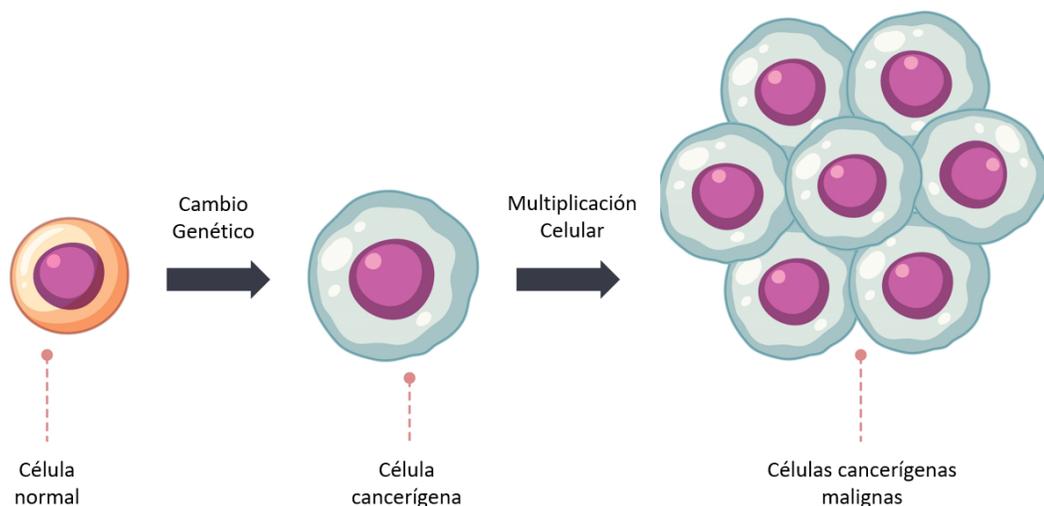


Figura 1-1: Proceso ilustrado sobre cómo las células malignas provienen de una célula normal que tuvo un cambio genético. Fuente: recurso electrónico de libre acceso

Como se menciona antes, el cáncer se puede dar en diversas partes del cuerpo y los tejidos mamarios no están exentos. Estas son zonas propensas para la producción de cáncer. Como enfermedad, el cáncer de mama se presenta tanto en hombres como en mujeres; sin embargo,

en los hombres, se presenta una incidencia de solo el 1% de los casos registrados de cáncer de mama a nivel mundial [2]. De acuerdo con la Organización Panamericana de la Salud (OPS), el cáncer de mama es el más común y es la principal causa de muerte por cáncer en las mujeres en América [3]. Esto ha llevado a que esta enfermedad sea uno de los principales problemas de investigación en salud pública en el mundo, buscando disminuir su alta incidencia y tasa de mortalidad [4].

Existen diversos factores de incidencia de esta enfermedad, entre estos, factores hereditarios, la exposición a agentes cancerígenos, la edad avanzada, mala alimentación, o malos hábitos, que incluyen fumar o ingerir alcohol de manera frecuente [1]. De acuerdo con el Instituto Nacional de la Salud y el Observatorio Mundial del Cáncer (GLOBOCAN), para el año 2020 se registraron más de 2.2 millones de nuevos casos (12.5% de la totalidad de nuevos casos de cáncer para ambos sexos [2]. Además, para ese mismo año se registró un total de 684,996 muertes por cáncer de mama correspondientes al 6.9% del total de muertes por cáncer en el mundo. En Colombia el panorama es similar, el cáncer de mama es el más frecuente en el país con una incidencia de 15,509 nuevos casos en el 2020, además se ha registrado como la primera causa de muerte entre quienes padecen cáncer en el país, ocasionando que 4,411 personas mueran a causa de esta patología. Estas cifras han prendido las alarmas a nivel mundial puesto que el cáncer de mama es el tipo de cáncer más frecuentemente diagnosticado y es la segunda causa de muerte relacionada con el cáncer en mujeres en todo el mundo [2].

Como patología, el cáncer de mama es tratable y puede tener un buen pronóstico si se diagnostica e interviene en etapas tempranas [5]. Existen diferentes métodos que permiten a los especialistas hacer un diagnóstico temprano de la enfermedad, principalmente basados en imagenología dada su naturaleza no invasiva. Entre estos está la mamografía, considerada el método de tamizaje estándar para la detección de cáncer de mama en mujeres mayores de 40 años. Este examen utiliza una dosis baja de rayos X para generar una imagen del tejido mamario en la que se pueden detectar crecimientos anormales y cambios en el tejido relacionados con el desarrollo del cáncer. No obstante, a pesar de su utilidad, la mamografía ha mostrado tener un desempeño pobre en la detección del cáncer en pacientes con mamas con una alta densidad tisular. Otra limitación de la mamografía es que no se recomienda su uso en pacientes jóvenes debido los efectos adversos que puede generar la radiación a largo plazo [6].

A diferencia de la mamografía, la ecografía es un método de tamizaje para la detección del cáncer de mama que no usa ningún tipo de radiación. En su lugar usa ondas de ultrasonido para crear las imágenes del tejido mamario. Por esta razón, la ecografía se recomienda como herramienta para el tamizaje de cáncer de mama en mujeres menores de 40 años o en estado de embarazo. Al igual que la mamografía, la ecografía ha mostrado tener un desempeño bajo en la detección del cáncer de mama cuando se practica a pacientes con mamas densas, además de ser altamente dependiente de la experiencia del operador que obtiene la imagen ya que una falta de examinación podría evitar la visualización correcta del tejido [7].

La alta densidad tisular, conocida como mamas densas, hace referencia a la cantidad relativa de tejido epitelial o elementos grasos en toda la zona de la mama, además de la piel. Estos elementos hacen parte del tejido de sostén que rodea y cubre las glándulas mamarias. Estas últimas son de especial interés puesto que son el foco del análisis de los exámenes de ayuda diagnóstica del cáncer de mama. Cuando hay alta densidad tisular se pierde la visibilidad de estas glándulas en las imágenes generadas por la mamografía y la ecografía, lo que dificulta el diagnóstico temprano del cáncer de mama [8; 9].

Con el fin de mejorar la capacidad para el diagnóstico temprano del cáncer de mama se han desarrollado otros métodos de diagnóstico como la resonancia magnética contrastada (DCE-MRI –*Dynamic Contrast-Enhanced Magnetic Resonance Imaging*) y la mamografía espectral con contraste (CEDM –*Contrast-Enhanced Digital Mammography*). Estos métodos utilizan un compuesto químico moderadamente tóxico denominado agente de contraste que es aplicado vía intravenosa. El principal propósito del medio de contraste es aprovechar el acelerado comportamiento metabólico de las células tumorales, las cuales poseen mayor capacidad de absorción. La absorción acelerada del medio de contraste se manifiesta en forma de incremento hiperintenso en las imágenes, permitiendo visualizar no solo la ubicación sino además la estructura interna de los tejidos hipercaptantes para posterior análisis [10].

En el caso de los estudios de DCE-MRI estos consisten en la adquisición de una serie de imágenes de resonancia magnética antes y después de inyectar el agente de contraste. Una de estas imágenes se captura justo antes de proporcionar el medio de contraste, mientras que las otras se capturan después de un período de tiempo determinado para evaluar el comportamiento de los tejidos a medida que se elimina del cuerpo el medio de contraste. La Figura 1-2 muestra dos de las imágenes de un estudio de DCE-MRI, la imagen de la izquierda fue tomada antes de aplicar el contraste y la imagen de la derecha después de aplicar el agente de contraste [11].

Por su parte, los estudios de CEDM aplican un agente de contraste en la toma de una mamografía convencional, lo que mejora la visibilidad de las lesiones para su diagnóstico. En concreto, después de la aplicación del agente de contraste esta técnica adquiere dos imágenes: una imagen de baja energía que es similar a una mamografía convencional en la que se revela la densidad de los tejidos, y una imagen de alta energía que capta el realce de contraste. Posteriormente, se procesan ambas imágenes para suprimir la señal del fondo y crear una imagen recombinada que mantiene las zonas de mayor captación del agente de contraste [12].

Aunque los estudios de DCE-MRI y CEDM son más precisos que la mamografía y el ultrasonido, estos presentan algunas limitaciones. Por ejemplo, para el caso de la DCE-MRI, hay baja disponibilidad de los equipos para la realización de estos estudios en los centros de salud, incluso en las ciudades principales de muchos países; su costo, incrementado por el uso del medio de contraste y el tiempo de adquisición de las imágenes, es mucho mayor que el de

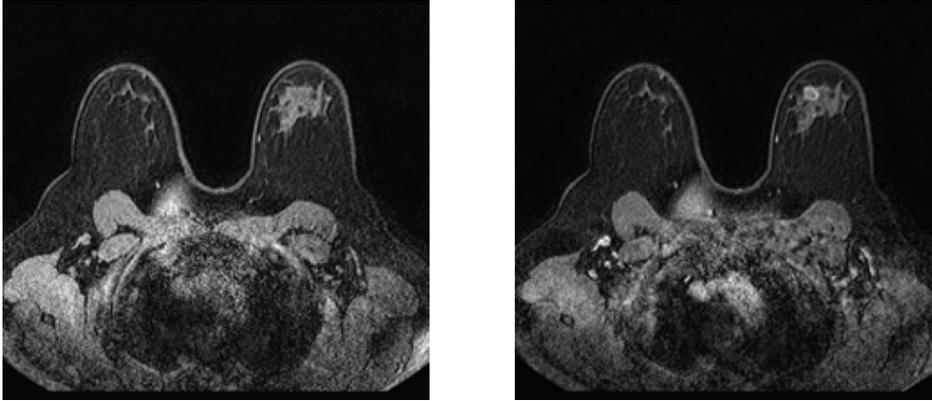


Figura 1-2: Estudio de DCE-MRI: la imagen de la izquierda corresponde a la imagen tomada antes de la inyección del agente de contraste, mientras que la imagen de la derecha fue tomada posterior a la inyección del agente de contraste. Fuente: base de datos del experimento

una mamografía o ecografía. Para el caso de la CEDM, su mayor limitante es la cantidad de radiación que se recibe en el proceso de adquisición de las imágenes. Además, según reportes en la literatura, se han encontrado pacientes que pueden desarrollar reacciones alérgicas a los agentes de contraste; e incluso se han reportado casos en los que estos agentes se quedan alojados en la base del cerebro o en la piel, y en los que aún se desconoce cuáles son los riesgos a largo plazo que esto puede generar [13] [14].

Estas limitantes han llevado a la búsqueda de alternativas que permitan realizar un diagnóstico temprano del cáncer de mama, manteniendo la precisión de los estudios de DCE-MRI y CEDM, pero reduciendo el costo y el tiempo asociados a la realización del examen y evitando la exposición del paciente a los agentes de contraste y sus posibles riesgos para la salud. Este es un tema de investigación que aún está abierto y vigente en la literatura.

Específicamente, para esta línea de trabajo, existen algunas propuestas orientadas al desarrollo de algoritmos que permitan la generación de imágenes sintéticas [15] [16] [17] [18]. En general, la síntesis de imágenes es el proceso de crear imágenes artificiales con un contenido específico a partir de una imagen de referencia, o la descripción de la imagen que se desea. En el contexto de los estudios de DCE-MRI y CEDM, la síntesis de imágenes puede utilizarse, por ejemplo, para generar imágenes sintéticas contrastadas a partir de las imágenes obtenidas antes de la inyección del agente de contraste.

Considerando la necesidad de reducir las limitaciones mencionadas para los estudios de DCE-MRI, en este trabajo se propone una estrategia basada en aprendizaje profundo que permite la generación de imágenes sintéticas postcontraste, a partir de imágenes precontraste. Las imágenes postcontraste sintéticas simulan el realce de contraste en los tejidos, sin utilizar un agente de contraste, manteniendo una calidad similar a las imágenes postcontraste reales,

tanto en los tejidos de la mama, como en las zonas de realce del contraste.

1.1 Objetivos

1.1.1 Objetivo General

Proponer un modelo basado en aprendizaje profundo que a partir de imágenes precontrastadas de la mama permita la generación de imágenes sintéticas similares a las imágenes reales obtenidas con el uso de medios de contraste.

1.1.2 Objetivos Específicos

1. Implementar una línea base de las técnicas de síntesis de imágenes considerando la información diagnóstica de los pacientes.
2. Desarrollar un modelo computacional basado en aprendizaje profundo para generar imágenes sintéticas postcontraste del tejido de la mama
3. Evaluar el desempeño del modelo computacional propuesto mediante la comparación de los datos sintetizados con los datos reales, con el fin de validar su correcto comportamiento.

1.2 Alcances

En el marco de este trabajo la evaluación se realiza sobre la calidad de las imágenes en términos de similitud a nivel de píxeles usando métricas computacionales. Dado que este trabajo se encuentra en etapa exploratoria, no se realiza evaluación por aparte de personal experto médico debido por las limitaciones de recursos humanos y económicos, principalmente orientados a la disponibilidad y alto costo de tiempo de servicio de personal médico del área de radiología.

2 Marco Teórico y Estado del arte

2.1 Marco teórico

2.1.1 El cáncer de mama y su diagnóstico

El cáncer de mama es uno de los tipos de cáncer con más prevalencia en todo el mundo. Al igual que otros tipos de cáncer, éste afecta tanto a hombres como mujeres. Sin embargo, quienes más lo padecen es la población femenina. El cáncer de mama se distingue de otros tipos de cáncer porque se desarrolla a partir de las células mamarias [19]; específicamente, suele originarse en los lobulillos, que son las glándulas productoras de leche, o en los conductos, que son las vías que transportan la leche desde los lobulillos hasta el pezón. También, el cáncer de mama puede originarse en los tejidos estromales, que incluyen a los tejidos conjuntivos grasos y fibrosos de la mama [5]. A modo de ilustración, la Figura 2-1 muestra las zonas mencionadas.

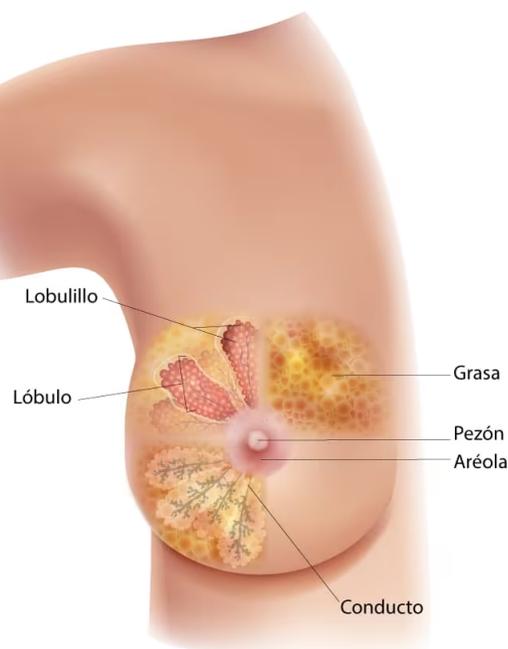


Figura 2-1: Anatomía de la mama. Fuente: [5]

Si bien el cáncer de mama tiene una alta incidencia de muertes, su tratamiento puede ser eficaz, especialmente cuando se detecta y se interviene en etapas tempranas. No obstante, para formular el tratamiento adecuado se requiere de un diagnóstico que los especialistas generan a partir del estudio de diferentes exámenes de detección. Entre esos, la mamografía es uno de los exámenes más empleados y consiste en la aplicación de una dosis de baja potencia de rayos X sobre las mamas para generar una imagen en la que se pueden identificar y localizar las lesiones asociadas al cáncer de mama [4]. La ecografía es otra técnica de diagnóstico cuyo principio de funcionamiento es el ultrasonido y a través del cual se generan imágenes de los tejidos de la mama para identificar tumores, quistes y masas sólidas. Estos métodos tienen varias limitaciones; sin embargo, una de las más importantes es que ambas tienen un desempeño bajo cuando se practican a pacientes con mamas densas, es decir, pacientes cuyas mamas tienen mucho contenido de fibras, grasa o piel, lo que dificulta la detección de lesiones asociadas al cáncer [7].

Para los casos en los que la mamografía y la ecografía no se pueden usar, o no son concluyentes, se utilizan otras modalidades de examen basadas en imágenes, como la resonancia magnética contrastada (DCE-MRI) y la mamografía espectral con contraste (CEDM). Ambas técnicas utilizan un agente de contraste, aplicado vía intravenosa, que permite resaltar las posibles lesiones en la mama a partir del comportamiento metabólico de las estructuras celulares en los tejidos tumorales [10].

En cuanto a la DCE-MRI, esta ha mostrado ser extremadamente sensible en la detección del cáncer de mama y no está limitada por la densidad del tejido mamario. Esto se logra gracias a que el resonador tiene la capacidad de manipular la magnitud y velocidad de los protones presentes en las células de los tejidos, a través de un pulso de radiofrecuencia que este emite, haciendo que los protones liberen energía, que posteriormente es captada por el resonador. Este proceso permite generar múltiples imágenes en serie de la mama [20]. Durante el proceso de adquisición de una resonancia magnética se capturan dos imágenes:

- Imagen ponderada en T1, que se genera en el tiempo de relajación en el que los protones de las células están alineados paralelamente al campo magnético del resonador. Esta imagen muestra de forma óptima la anatomía normal del tejido blando y la grasa [21].
- Imagen ponderada en T2, que indica el tiempo de relajación en el que los protones están alineados perpendicularmente al campo magnético del resonador. Esta imagen muestra de manera óptima los líquidos y las alteraciones, por ejemplo: tumores, inflamaciones y traumatismos en los tejidos [21].

En la práctica clínica, las imágenes correspondientes a los tiempos de relajación T1 y T2, proporcionan información complementaria, por lo que ambas son importantes para la detección y caracterización de las alteraciones en la mama.

En un estudio de DCE-MRI se realizan alrededor de seis adquisiciones que se usan para

evaluar, en las imágenes ponderadas en T1, como es el comportamiento dinámico, en el tiempo, del medio de contraste sobre los tejidos. Esto es importante puesto que el medio de contraste se absorbe más rápido por los tejidos que presentan alteraciones, facilitando su identificación en las imágenes obtenidas [20]. Así, la primera adquisición se realiza antes de suministrar el medio de contraste y las demás se toman después de aplicado este, una tras otra.

A partir de cada secuencia de imágenes es posible analizar el comportamiento de la captación de medio de contraste considerando los cambios en la intensidad de los píxeles al rededor de una zona. Este comportamiento dinámico es comúnmente denominado como curva de captación y es usado junto con las imágenes de T1 y T2 para determinar un diagnóstico concluyente.

Por su parte, la CEDM es un método alternativo a la DCE-MRI. Este método evalúa dos elementos importantes de la mama: su anatomía y la vascularidad de la zona. Esta evaluación se da mediante dos imágenes: una imagen de baja energía (LE –*Low Energy*) en la que se evidencia la anatomía de la mama y la cual es equivalente a una mamografía digital de campo completo; y una imagen de alta energía (HE –*High Energy*) la cual se toma posterior a la inyección del agente de contraste. A partir de la substracción de estas dos imágenes se genera una imagen recombinada en la que se suprime el tejido normal y se realzan las zonas de comportamiento tumoral [22].

DCE-MRI y CEDM son estudios que tienen una precisión alta en la detección del cáncer de mama, comparados con los métodos convencionales como la ecografía y la mamografía [20]. Las limitaciones que se mencionaron con anterioridad sobre estos métodos, mantienen abiertas las posibilidades de investigación en el campo y han llevado a que se explore el uso de estrategias de inteligencia artificial para generar imágenes contrastadas artificiales que no requieran la aplicación de un agente de contraste, con lo que además de evitar las reacciones adversas ayudaría a disminuir los tiempos de adquisición y los altos costos de estos exámenes.

A continuación se describe en qué consiste la generación de imágenes artificiales (denominada **síntesis de imágenes**) y se hace una breve descripción de los métodos de aprendizaje profundo utilizados para esta tarea. Adicionalmente, se presenta el estado del arte en síntesis de imágenes de DCE-MRI para el diagnóstico de cáncer de mama.

2.1.2 Síntesis de imagen médica

La síntesis de imagen es el proceso de creación de imágenes artificiales a partir de alguna descripción de su contenido. Esa descripción se puede suministrar como texto, audio o a partir de otras imágenes [23]. En los últimos años la tarea de crear imágenes artificiales ha ganado relevancia en múltiples dominios de aplicación como en la edición de imágenes [24] [25], producción de imágenes a partir de texto [26] [27], generación de video [28], traducción

de imagen a imagen [15] [16] [29], entre otros.

En el campo de la medicina, la síntesis de imágenes se ha utilizado para múltiples propósitos; por ejemplo, para aumentar los conjuntos de datos de imágenes que entrenan modelos para el diagnóstico de enfermedades como el cáncer [30]; para mejorar la resolución de imágenes obtenidas con modalidades de imágenes de baja resolución [31]; para obtener la imagen de una modalidad a partir de los datos de otra modalidad [32]; o para segmentar regiones de interés en la imagen [33], entre otros. La importancia del uso de la síntesis de imágenes médicas para diversos objetivos, radica en su potencial para replicar las estructuras y las texturas de los órganos y tejidos presentes en las imágenes médicas.

2.1.3 Aprendizaje profundo para síntesis de imágenes médicas

El aprendizaje profundo ha revolucionado las aplicaciones de visión artificial al ofrecer capacidades avanzadas de procesamiento para el reconocimiento y clasificación de objetos en imágenes de múltiples dominios de aplicación. Como modelo de aprendizaje, las redes neuronales han permitido avances significativos en tareas como clasificación de imágenes [34], la detección de rostros [35] hasta llegar al reconocimiento de objetos individuales en las imágenes [36]. Esta tecnología no solo han mejorado la precisión en la identificación de objetos y características visuales, sino que también han allanado el camino para aplicaciones más sofisticadas, como el seguimiento de objetos en tiempo real, la generación de contenido visual realista y la interpretación de escenas complejas. Además, en campos como la medicina, las redes neuronales han facilitado el diagnóstico a través de la interpretación de imágenes médicas, mientras que en la industria automotriz han contribuido al desarrollo de vehículos autónomos. Estos avances han posicionado a las redes neuronales como herramientas fundamentales para potenciar aplicaciones diversas y sofisticadas en el ámbito de la visión artificial.

Más específicamente, en el área de la medicina, la síntesis de imágenes se suele realizar a partir de modelos de aprendizaje profundo, debido a que poseen la capacidad para aprender patrones complejos, lo cual es importante debido a que las imágenes médicas son datos de alta complejidad estructural. Además, estos modelos, tiene la capacidad de generar imágenes completamente nuevas, lo cual es útil, por ejemplo, para tareas en las que se quiera eliminar la dependencia a ciertos tipos de datos de difícil acceso. Finalmente, este tipo de modelos, se usan para síntesis de imagen médica debido a que pueden adaptarse a una amplia variedad de tareas, como la superresolución, la eliminación de ruido o la traducción de imágenes entre dominios [37]. El aprendizaje profundo es una rama del aprendizaje de máquinas que se basa en la formulación de redes neuronales artificiales, las cuales tratan de emular el comportamiento de las redes neuronales humanas, principalmente su proceso de aprendizaje.

Este tipo de modelos se construye a partir de capas de neuronas artificiales interconectadas

entre si, lo que les da la capacidad de auto ajustarse en función de la entrada y la cantidad de datos que estos procesan [38]. Al simular la estructura de las neuronas humanas, cada neurona que compone la red neuronal artificial contribuye con el aprendizaje a través de transformaciones matemáticas no lineales operadas en cadena, las cuales permiten separar o agrupar los datos de entrada en un espacio dimensional diferente al original [39].

Toda red neuronal está compuesta de secciones denominadas capas. En la Figura 2-2 se muestra una estructura visual de una red neuronal sencilla compuesta por una capa de entrada, dos capas ocultas y una capa de salida. La primera capa, denominada capa de entrada, es la capa encargada de recibir los datos que serán procesados. La última capa de la red se denomina capa de salida, y es la que entrega el resultado para el cual fue diseñada la red. Las capas internas, llamadas capas ocultas, se encargan de procesar los datos en diferentes niveles de abstracción [40]. Cada capa de la red neuronal está compuesta por un número determinado de neuronas, y la salida de cada neurona en una capa se conecta con la entrada de otra neurona en la siguiente capa.

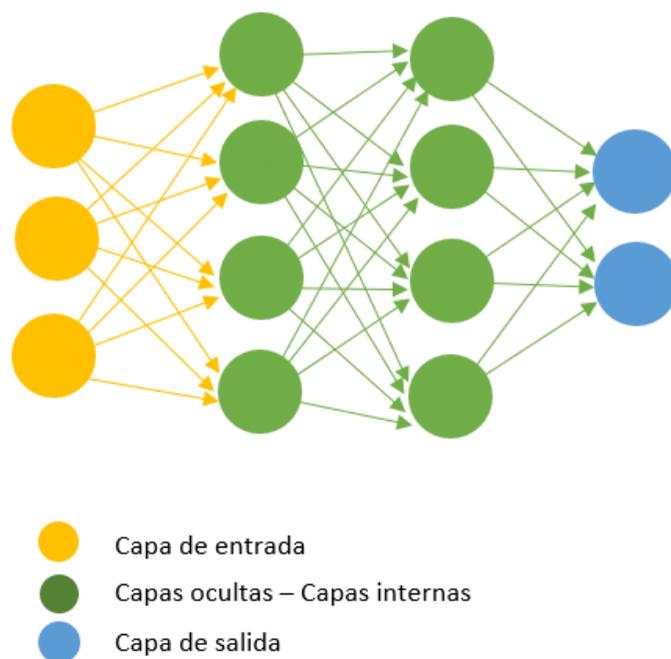


Figura 2-2: Esquema básico de una red neuronal con una capa de entrada, dos capas ocultas y una capa de salida. Fuente: elaboración propia

Diferentes investigaciones han usado múltiples arquitecturas de redes neuronales para abordar la síntesis de imágenes médicas. En general, esas arquitecturas consisten en modelos de redes neuronales convolucionales (CNN –*Convolutional Neural Networks*) en una configuración de *Autoencoders*, compuestos de una parte codificadora y otra decodificadora, o como redes generativas adversarias (GANs –*Generative Adversarial Networks*).

Redes neuronales convolucionales

La implementación de una red neuronal convolucional consiste en la aplicación sucesiva de convoluciones en forma de capas a una imagen para reducir su dimensionalidad extrayendo y conservando la información relevante de esta [23]. En general, la convolución es una operación que pondera los píxeles de la imagen a partir de la importancia especificada en un filtro, denominada kernel de convolución. La particularidad de este filtro es que sus valores se optimizan en función de la salida esperada a través del algoritmo de gradiente descendente. La Figura 2-3 muestra el proceso de la convolución. La fórmula básica para calcular la salida de una capa de convolución en una red neuronal convolucional (CNN) se puede expresar como se ve en la Ecuación 2-1.

$$\text{Salida}(i, j, k) = \sum_{l=1}^C \sum_{m=1}^{H_f} \sum_{n=1}^{W_f} (\text{Entrada}(i+m-1, j+n-1, l) \cdot \text{Filtro}(m, n, l, k)) + \text{Desplazamiento}(k) \quad (2-1)$$

Donde $\text{Salida}(i, j, k)$ es el valor en la posición (i, j) del canal k en la salida de la capa de convolución; $\text{Entrada}(i + m - 1, j + n - 1, l)$ es el valor en la posición $(i+m-1, j+n-1)$ del canal l en la entrada; $\text{Desplazamiento}(k)$ es el valor del filtro en la posición (m, n) del canal l al canal k ; C es el número de canales en la entrada; H_f y W_f son el alto y ancho del filtro de convolución; i y j son índices que recorren el alto (alto de la salida) y ancho (ancho de la salida) de la salida, respectivamente; k es el índice que recorre los canales en la salida.

Esta fórmula matemática describe cómo se realiza la operación de convolución entre la entrada, el filtro y el sesgo para calcular los valores de la capa de convolución. La convolución se aplica en cada posición (i, j) de la salida y para cada canal de la misma. La operación de suma acumula el resultado de la multiplicación entre la entrada y el filtro para cada posición del filtro y para cada canal, y luego se suma el sesgo correspondiente. La operación de convolución se repite para cada posición (i, j) en la salida y para cada canal de la misma, generando así la salida completa de la capa de convolución.

La salida de una capa de convolución es una representación enriquecida de las características detectadas en la entrada a través del filtro de convolución. Es importante tener en cuenta que en una CNN típica, a menudo hay varias capas de convolución apiladas en una secuencia, y cada capa puede tener diferentes tamaños de filtro, relleno y pasos. Además, es común aplicar una función de activación, como ReLU, a la salida de la capa de convolución para introducir no linealidad en la red.

Por la naturaleza de las redes neuronales, la salida de la convolución de una capa de la red se propaga a la siguiente capa después de aplicar una operación de *pooling* o submuestreo. Es así como capa tras capa se van aplicando convoluciones sucesivas a la imagen que junto

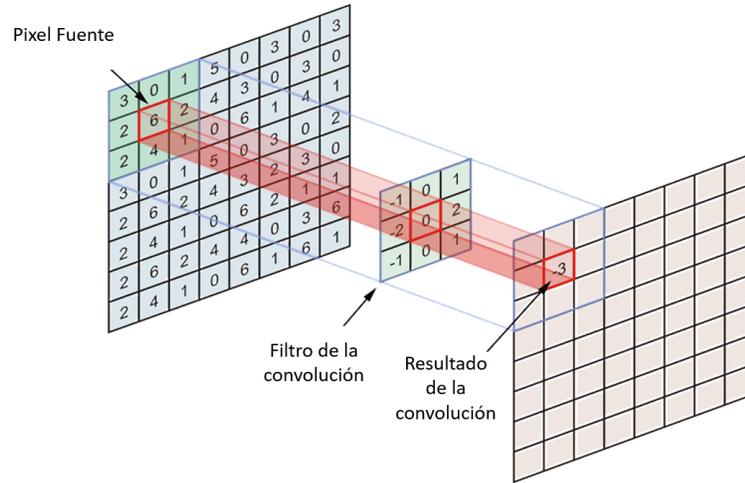


Figura 2-3: Operación de convolución. En azul la imagen de entrada, en verde el kernel de convolución y en caki la imagen de salida. El valor de cada píxel en la imagen de salida se calcula como una suma ponderada entre los píxeles de la imagen de entrada usando los pesos del kernel. Fuente [41]

con la operación de submuestreo, reducen la dimensionalidad de la imagen manteniendo la información relevante. En la Figura 2-4 se visualiza un proceso de submuestreo: se conserva la información de mayor relevancia dentro del filtrado inicial del kernel [23].

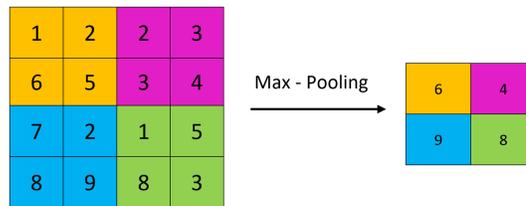


Figura 2-4: Descripción gráfica de submuestreo. La malla que contiene los píxeles de la imagen se reduce a través de la operación del submuestreo. Este es un ejemplo de un submuestreo de tamaño 2x2, donde se conserva la información más importante en cada bloque. Fuente: [41]

Dado la formulación de las capas convolucionales, estas son empleadas usualmente para la extracción de características abstractas relevantes en las imágenes a partir de la información espacial contenida en ellas. Estas características son usadas como entrada de una red neuronal completamente conectada para obtener una estimación final que se ajusta de acuerdo al valor esperado. De esta forma, las redes que incluyen capas convolucionales pueden entrenarse de inicio a fin optimizadas por el algoritmo de gradiente descendente.

Otro elemento de las redes neuronales es la función de activación, la cual permite aplicar

transformaciones no lineales a la salida de las neuronas [23]. Por otro lado, el proceso de ponderación y propagación de la información entre las capas convolucionales de la red está orientando iterativamente por una función de pérdida ($\mathcal{L}(x, y)$) la cual orientará el proceso de aprendizaje con el fin de obtener el valor real esperado de salida. La función de pérdida es la base del proceso de aprendizaje puesto que la red neuronal usa el algoritmo de gradiente descendente para minimizarla a medida que transcurre el proceso de entrenamiento [42].

Existen varios tipos de redes neuronales convolucionales según su configuración. Para el caso de la síntesis de imágenes médicas, dos tipos de configuración de redes han cobrado importancia: los autoencoders y las redes generativas antagónicas.

Autoencoders

Los *Autoencoders* (AE) fueron propuestos originalmente por Hinton et al. [43]. Son arquitecturas diseñadas con el fin de reconstruir una observación de entrada con el mínimo error posible. Típicamente involucra un codificador el cuál cumple con la tarea de comprimir los datos de entrada, preservando los de mayor importancia, pasando de un espacio de gran dimensión a un espacio de menor dimensión. Además, involucra un decodificador, el cual realiza la labor contraria del codificador reconstruyendo la información de un espacio dimensional mayor, a partir de la información de un espacio de menor dimensión [44].

El objetivo principal detrás de los autoencoders es aprender una representación abstracta (o latente) de la distribución original de los datos de entrada, independiente de su dimensionalidad. Logrando esto, es posible aumentar o reducir la dimensionalidad de manera eficiente para resolver tareas para la toma de decisiones, empleándose así para clasificación, agrupamiento, reconstrucción, segmentación, entre otras tareas [44].

En los autoencoders, el codificador se diseña para reducir la dimensionalidad de los datos de entrada. Este proceso produce una pérdida de información que está ligada a la profundidad del codificador y a la aplicación sucesiva de la operación de submuestreo [44]. Para contrarrestar esta limitación de los autoencoders originales en su formulación original se creó una nueva arquitectura denominada U-Net [45], cuyo nombre se deriva de la forma como se conecta el codificador y decodificador incluyendo conexiones entre las capas intermedias o *skip connections*, para mitigar la pérdida de la información (ver la Figura 2-5).

A diferencia de los autoencoders, la arquitectura U-Net copia la información de las capas del codificador a las capas del decodificador generando una dependencia entre estos. Esto permite que el decodificador reconstruya la información no solo usando las capas anteriores sino también usando la información que proviene directamente del espacio de los codificadores. Como tal, la U-Net es una arquitectura que fue propuesta en el contexto de generación de información sintética a partir de imágenes médicas [44; 45]. Sin embargo, esta solución que aborda la pérdida de información en el contexto de segmentación semántica posee una no-

table desventaja en otros contextos como la generación de imágenes. Esto es que la función de costo (basada en píxeles) permite sintetizar las componentes fundamentales de las imágenes, las cuales corresponden a las frecuencias altas en los cambios de intensidad. No obstante, esta formulación no apunta directamente a los cambios en frecuencias bajas en términos de intensidad, las cuales son importante en el contexto de imágenes médicas con medio de contraste debido a las estructuras microscópicas (células) que pueden tener captación. Por este motivo, los modelos basados en arquitecturas U-Net son reconocidos por generar imágenes que reflejan la estructura general del tejido pero con notables regiones borrosas [16].

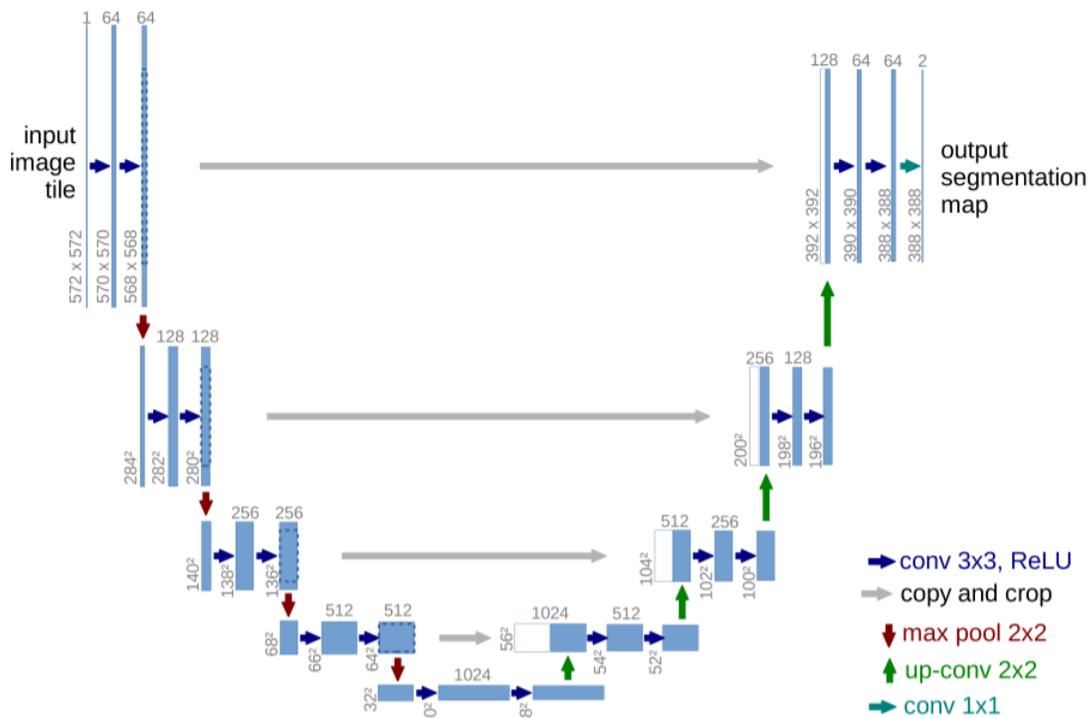


Figura 2-5: Ejemplo de una arquitectura tipo U-Net. En el codificador los bloques azules son los mapas de características, las flechas azules representan las convoluciones y las flechas rojas definen las operaciones de submuestreo. En el decodificador, las flechas verdes definen el proceso de expansión del mapa de características, mientras que las flechas grises que van del codificador al decodificador representan la operación de concatenación de información. Fuente: [45]

Por razones como estas, se han propuesto redes con diferentes niveles de profundidad, o de diferentes distribuciones, y también, se han implementado combinaciones entre ellas, como es el caso de los modelos generativos

Redes generativas adversarias

Las redes neuronales generativas adversarias, (GANs – *Generative Adversarial Networks*), se han convertido en uno de los modelos de generación de datos sintéticos más utilizados. Este tipo de arquitectura está compuesta por dos modelos de redes neuronales que se entrenan confrontándose el uno con el otro. Estos modelos son llamados generador (G) y discriminador (D) y su estructura se muestra en la Figura 2-6.

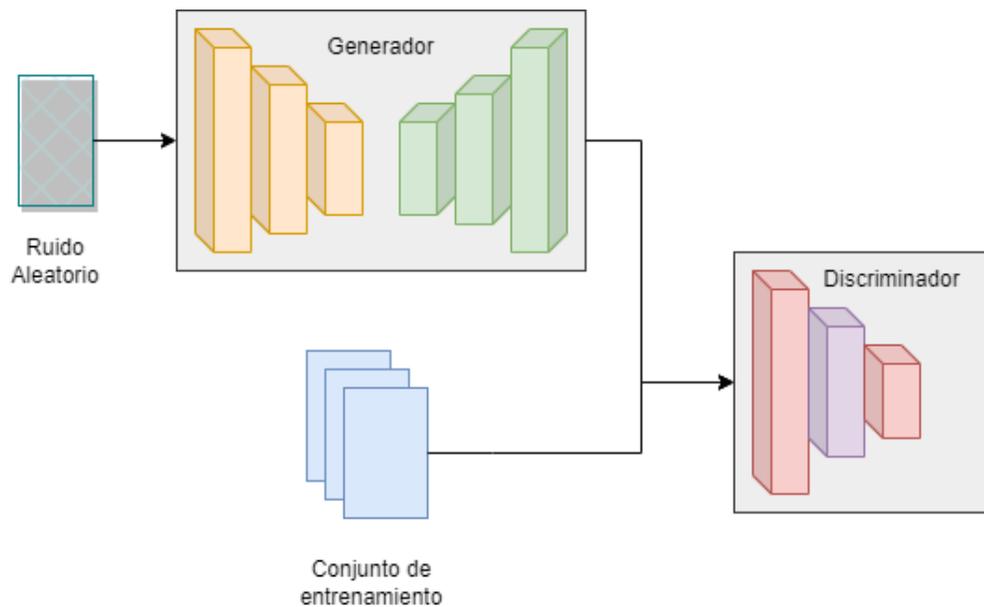


Figura 2-6: Estructura general de una GAN. El generador crea imágenes sintéticas a partir de un vector de información aleatoria, mientras que el discriminador trata de discernir entre las imágenes reales y falsas. Fuente: Elaboración propia

Por un lado, el generador es una red convolucional que intenta aprender la distribución de los datos reales para generar información sintética a partir de una muestra de ruido aleatorio. Por otro lado, el discriminador es una arquitectura complementaria que se comporta como “experto” y cuya tarea es diferenciar entre la información real y la generada. El entrenamiento de ambas redes se realiza de manera simultánea y adversaria, lo cual le otorga el nombre a esta arquitectura [46]. De este modo, el generador G y el discriminador D , compiten el uno contra el otro, siguiendo una aproximación Min-Max. El proceso de entrenamiento finaliza cuando el discriminador es incapaz de diferenciar entre la información real y la información generada.

Durante el proceso de entrenamiento ambas redes usan para el aprendizaje una función de pérdida, la cual se calcula a partir de la habilidad del discriminador para diferenciar entre la información real y la información generada. Así, el generador (G) intenta minimizar dicha función, mientras que el discriminador intenta maximizarla. La función de pérdida propuesta

en la publicación original de las GANs [46] se describe de manera simplificada en la ecuación 2-2.

$$L_{GAN}(G, D) = \mathbb{E}_x [\log D(x)] + \mathbb{E}_z [\log(1 - D(G(z)))] \quad (2-2)$$

Donde $D(x)$ es la estimación del discriminador de la probabilidad de que la instancia x de datos real sea real, \mathbb{E}_x es el valor esperado para todas las instancias de datos reales, $G(z)$ es la salida del generador para el ruido z , $D(G(z))$ es la estimación del discriminador para que una instancia falsa sea real, y \mathbb{E}_z es el valor esperado en todas las entradas aleatorias del generador.

Una de las variaciones de la GAN es la DCGAN (*Deep Convolutional GAN*) [47]. Este tipo de arquitectura fue propuesta en el campo de la medicina para enfocar el proceso de generación en las zonas de interés de lesiones de diversas enfermedades, tratando de dar importancia a los tejidos de interés que podrían contener dichas lesiones. A diferencia de una GAN convencional, la DCGAN cuenta con un mayor número de capas convolucionales en el generador para extraer más información, esto, bajo la suposición de que a mayor cantidad de capas convolucionales, mayor es la extracción de datos relevantes de las imágenes que se están sintetizando [47]. La estructura de la arquitectura es igual a la de una GAN convencional; sin embargo, la DCGAN reemplaza las capas de *pooling* por capas convolucionales con un tamaño de paso mayor, además agrega un bloque llamado *batch normalization* tanto en el generador como en el discriminador y cambia las funciones de activación del discriminador.

Otra de las variaciones de la arquitectura GAN es es la *conditional GAN* (cGAN), la cual agrega tanto el generador como al discriminador una entrada adicional que es una etiqueta o vector de condición [48]. Dicho vector de condición ayuda a que el generador se centre en generar datos de una clase específica, mientras que el discriminador debe diferenciar entre si el dato ingresado es un dato real o falso con respecto a la clase especificada.

Existe otro tipo de GAN que se conoce como CycleGAN (*Cycle-Consistent Adversarial Networks*) [49]. La principal aplicación de CycleGAN es la traducción de imágenes de un dominio a otro de manera no supervisada. Esto significa que no se requiere un conjunto de datos que se encuentre emparejado, es decir, que contenga imágenes correspondientes en ambos dominios para entrenar el modelo [49]. La característica distintiva de CycleGAN es su capacidad para mantener la consistencia cíclica. Esto significa que, dado un par de dominios A y B, si una imagen se traduce de A a B y luego de B a A, debería ser similar a la imagen original. Esto asegura que el proceso de traducción sea reversible y que no se pierda información importante en el proceso.

CycleGAN utiliza dos generadores y dos discriminadores en su arquitectura para realizar estas traducciones bidireccionales. Los generadores intentan traducir imágenes de un dominio

al otro, mientras que los discriminadores evalúan la autenticidad de las imágenes generadas en comparación con las imágenes reales en su respectivo dominio.

En general, existen diferentes tipos de arquitecturas de aprendizaje profundo que se pueden emplear para el proceso de síntesis de imagen médica. En la siguiente sección se describen algunos de estos trabajos, agrupados por la aplicación que se da a la imágenes sintéticas.

2.2 Estado del arte en síntesis de imágenes médicas de mama

En la literatura se han propuesto diferentes modelos para la generación sintética de imágenes médicas. Para el caso específico de las modalidades de imagen usadas como apoyo para el diagnóstico de cáncer de mama, los métodos de síntesis propuestos se han orientado para una de tres tareas específicas: como estrategia de aumento de datos para el apoyo al entrenamiento de modelos de inteligencia artificial; como estrategia para mejorar la resolución de las imágenes, también conocido como un método de superresolución; y como estrategia para el cambio de dominio entre diferentes modalidades de imagen.

A continuación se presentan algunos de los trabajos relevantes asociados a la síntesis de imágenes médicas para el apoyo en el diagnóstico de cáncer de mama.

2.2.1 Aumento de datos

Las estrategias para el aumento de datos surgen como una alternativa para contrarrestar la limitación en cantidad de los datos médicos públicos disponibles para el entrenamiento de los modelos basados en aprendizaje profundo. El aumento de datos consiste en generar nuevos datos a partir de un conjunto de entrada con el fin de aumentar la cantidad y variabilidad de datos durante el entrenamiento. A diferencia de las estrategias tradicionales de aumento de datos, las cuales suelen aplicar transformaciones afines como rotaciones o traslaciones y modificaciones al brillo y contraste de las imágenes, las nuevas muestras suelen ser muy similares a las originales y poco diversas, lo que limita la capacidad de generalización de muchos métodos de aprendizaje profundo. Por ello, los modelos de generación de imágenes pueden presentar múltiples ventajas además de mitigar el sobre-ajuste en el entrenamiento como mejorar el desempeño en términos de la calidad de las imágenes, mejorar la capacidad de generalización, reducir aumentar la variabilidad de las muestras y mejorar el desempeño al realizar tareas de bajo nivel (clasificación) [30; 50].

Una de las arquitecturas usadas para ese propósito es la DCGAN. Por ejemplo, Dhivya, et al. [51] compararon el resultado de las técnicas de aumento de datos tradicionales contra el aumento de datos a partir de una DCGAN en el desempeño de la clasificación de lesiones

mamarias. Como conclusión, los autores encontraron que el uso de la DCGAN permitió la generación de imágenes realistas pero diferentes a las de la base de datos, lo que ayudó a mejorar la precisión del clasificador entrenado.

Por otro lado, Shen et al. [48] usan una cGAN para sintetizar imágenes de mamografía como estrategia de aumento de datos en el contexto de segmentación de lesiones en la mama. En su trabajo los autores manifiestan que en la literatura existen pocas bases de datos de imágenes con lesiones mamarias etiquetadas y por ello proponen una arquitectura cGAN para aprender la distribución de las lesiones mamarias en la mamografía a partir de sus máscaras de segmentación correspondientes. Como resultado el modelo entrenado permite generar imágenes con lesiones sintéticas a partir de una mamografía sin lesiones y una máscara que indica la ubicación deseada de la lesión. De acuerdo con los autores, el uso de las imágenes sintéticas aumentó, respecto al conjunto de datos original, la precisión de un modelo U-Net para segmentar lesiones en imágenes de mamografía.

Otra de las investigaciones que ha usado redes neuronales para el aumento de datos es el trabajo de Sandfort et al. [14]. En este los autores evalúan la capacidad de una CycleGAN [49] para aumentar un conjunto de datos de imágenes de tomografía para la tarea de segmentación de órganos. En el estudio se concluyó que el modelo de segmentación entrenado con el conjunto de datos aumentado con los datos generados con la CycleGAN logró un mejor desempeño comparado al modelo entrenado con el conjunto de datos original, dejando en evidencia que las imágenes aumentadas por este medio presentan mejor calidad [14].

Otra de las aplicaciones en las que se ha utilizado el aumento de datos es en la detección de lesiones mamarias [50]. En el trabajo de Guan et al. [50] los autores proponen una estrategia de aumento de datos basada en una GAN progresiva multicanal restringida por textura, a la que denominan TMP-GAN. De acuerdo con los autores, la arquitectura propuesta, que usa una función de pérdida basada en textura, genera imágenes sintéticas de baja calidad a las cuales, progresivamente, se les mejora su calidad hasta obtener imágenes sintéticas con una textura continua en el área de fusión de las lesiones. Al igual que en el trabajo de Lin et al. [52], los autores muestran que el desempeño de una U-Net en la detección de lesiones mejora cuando se usa el conjunto de datos aumentado con la arquitectura propuesta.

En cuanto al trabajo de K. Balaji [53], el autor propone trabajar con segmentación de los tumores de mama en imágenes de DCE-MRI. Sin embargo, para que una red tenga la habilidad de detectar las zonas de interés para segmentarlas, tiene que tener gran cantidad de datos etiquetados con sus respectivas lesiones para aprender de ellas. Hacer eso, anotar manualmente los tumores, es un proceso que requiere tiempo y conocimientos médicos específicos. Por ende, deciden aumentar la base de datos con aprendizaje profundo. Para esta labor, emplean una salida de codificador automático variacional (VAE) de forma complementaria para renovar la imagen que va de entrada a la red de segmentación. Es importante destacar que los VAE son un tipo de modelo generativo que contiene un codificador automático que

se encarga de mapear datos de entrada, a un espacio de características latentes. La característica distintiva de un VAE es que el espacio de características está modelado como una distribución de probabilidad, lo que permite que el modelo capture la incertidumbre en la representación. Su salida es una distribución de probabilidad, donde cada dimensión de esta distribución representa una característica latente del dato de entrada. Se emplea una red autoencoder para realizar el proceso de segmentación de las lesiones. Este tipo de anclaje a la red, permite reforzar el hallazgo de lesiones, pues el codificador empleado aumenta los datos de entrada del autoencoder [53]

2.2.2 Superresolución

La superresolución consiste en mejorar la calidad de una imagen en función de la resolución. En términos de imágenes médicas, la superresolución busca mejorar la calidad de las imágenes diagnósticas para proveer a los expertos de herramientas de mayor calidad diagnóstica, considerando que los equipos de adquisición de altas resoluciones son considerablemente más costosos y con menor disponibilidad. Sin embargo, el caso particular de imágenes médicas plantea un problema mayor en términos de procesamiento de imágenes dado que la calidad de las mismas no se limita únicamente al tamaño espacial (como en el caso de las imágenes naturales) sino a la potencia del escáner que permite identificar estructuras con mayor precisión y variación técnica en términos de la intensidad de píxeles [54].

Entre los trabajos que usan aprendizaje profundo para mejorar la calidad de imágenes médicas está el de Campaz et al. [55]. En este, los autores proponen una arquitectura tipo U-Net, *Convolutional Autoencoder*, para mejorar la calidad de imágenes de DCE-MRI capturadas con un resonador de 1.5 Teslas. De acuerdo con los resultados presentados, los autores muestran que la arquitectura propuesta genera imágenes en las que existe una correspondencia positiva entre la estructural y la distribución de los píxeles de las imágenes sintetizadas, respecto a las imágenes reales.

La DCE-MRI también fue estudiada por Wang et al. [56]. En esta investigación, buscaron la segmentación de tumores de mama, visibles a partir de la DCE-MRI. El problema al que se enfrentan es que las formas y tamaños de los posibles tumores de mama son muy variables, además el fondo de la imagen puede no ser homogéneo. Elementos así, dificultan la tarea de segmentación. Para poder mejorar esta situación, se propone un módulo de síntesis que sea sensible a los tumores, para que se pueda agregar al proceso de segmentación. La idea es que este módulo retroalimente al segmentador, para que mejore la calidad de la imagen al momento de ser segmentada [56]

2.2.3 Traducción de imágenes o cambio de dominio

La traducción de imágenes, o cambio de dominio, consiste en transformar una imagen x , que pertenece a una distribución X , a otra imagen similar y que pertenece a una distribución Y . Ambas imágenes, x e y , deben compartir atributos semánticos y de contenido, aunque sean significativamente diferentes [32]. El cambio de dominio en el contexto de imágenes médicas ocurre cuando para un diagnóstico se requieren imágenes complementarias de dos modalidades de imagen diferente que necesitan de dos o más procesos de adquisición. Un ejemplo, es el par de imágenes de emisión de positrones (PET) y tomografía computarizada (CT) donde la imagen CT se utiliza con fines técnicos de corrección de atenuación del estudio PET [57]. Otro ejemplo de cambio de dominio es el que se presenta entre las imágenes precontraste y postcontraste de un estudio de DCE-MRI. Entre las aplicaciones de cambio de dominio, Gao et al. [16] proponen el uso de una arquitectura de red neuronal poco profunda, denominada *Shallow-Deep Convolutional Neural Network* (SD-CNN). En este trabajo los autores usan la arquitectura, que se muestra en la Figura 2-7.

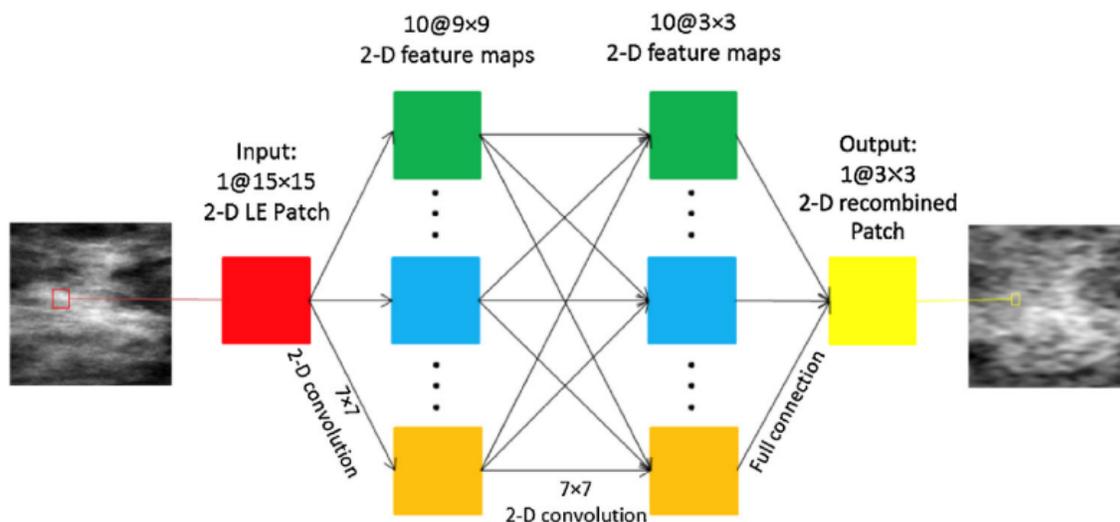


Figura 2-7: Arquitectura de 4 capas de una SD-CNN para la reconstrucción de parches de imágenes re combinadas. Fuente: [16].

Esta arquitectura se usa para aprender la relación entre parches de imágenes de baja energía (LE) y los correspondientes parches de imágenes re combinadas de estudios de mamografía contrastada CEDM. Así, la arquitectura propuesta aprende a sintetizar parches de imágenes re combinadas a partir de parches de imágenes de baja energía. Los autores evaluaron la capacidad de generación de la red a partir la precisión de un clasificador entrenado para identificar si un parche de una mamografía tiene o no una lesión asociada al cáncer. Entre las limitaciones de este tipo de arquitectura, es que es una arquitectura ligera que tiene poca utilidad en la síntesis de imágenes completas, principalmente porque no tiene suficientes

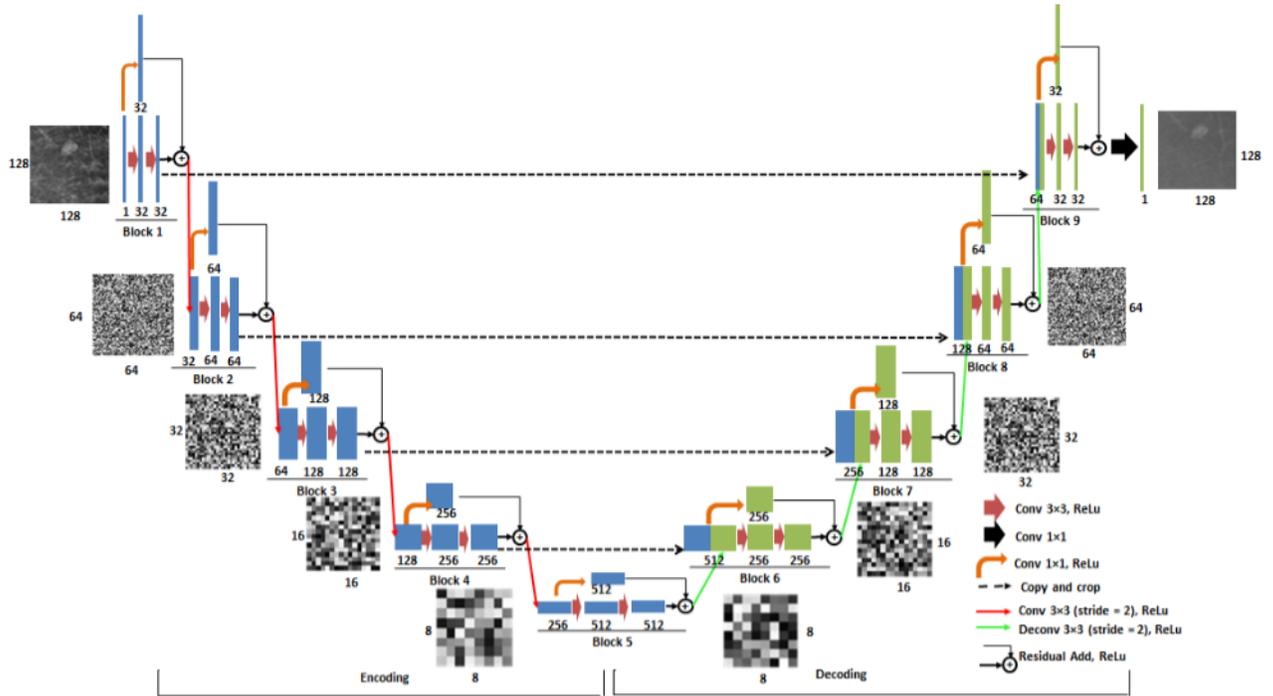


Figura 2-8: La Arquitectura RiedNet está compuesta por un codificador que se utiliza para extraer las características relevantes de los objetos y un decodificador que restaura los detalles de la imagen. Las *skip connections* se usan para pasar la información de las características de los objetos del codificador al decodificador para contrarrestar la pérdida de la propagación del gradiente. Fuente: [17].

niveles de abstracción para su reconstrucción.

Como una continuación del trabajo anterior, Gao et al. propusieron en [17] una arquitectura tipo U-Net denominada *Residual Inception Encoder-Decoder Network* (RiedNet). Esta arquitectura se compone de distintos bloques que se encargan de comprimir la información en la imagen de entrada y reconstruir la imagen a partir de la información comprimida. Además de estos bloques, la RiedNet tiene una ruta de expansión que captura la información contextual en el codificador y la transmite hasta el decodificador. Por último, la RiedNet introduce capas convolucionales y deconvolucionales para reemplazar las típicas capas *pooling* que se implementan para reducir la complejidad computacional de la U-Net. La razón es que incluir filtros convolucionales permite aprender los valores de los filtros, de modo que la información de píxeles se pueda rastrear tanto en la codificación para reducir la resolución espacial de los mapas de características, como en la decodificación para aumentar la resolución espacial de la imagen [17]. La Figura 2-8 describe la arquitectura general de esta red.

A pesar de que en los experimentos los autores de esta arquitectura presentan en su trabajo los resultados de síntesis de imagen completa, la arquitectura fue diseñada y entrenada para

sintetizar bloques de 128x128. Es decir, que para la síntesis de la imagen completa se usó un proceso de síntesis de bloques de 128x128 con un paso de 8. En las zonas superpuestas entre los bloques se aplicó el promedio entre todos los bloques sintetizados.

En cuanto a estudios de DCE-MRI, los autores de [18] propusieron una arquitectura basada en una cGAN para generar imágenes ponderadas en T1, con supresión de grasa, a partir de imágenes con saturación. Esta arquitectura propuesta a partir del modelo de Isola et al. [15] consiste de un generador tipo U-Net y un modelo discriminador tipo PatchGAN. Para evaluar el resultado del proceso de generación los autores emplean métricas cuantitativas y cualitativas. De acuerdo a los autores, las imágenes sintéticas obtuvieron puntuaciones que se establecen de malas a buenas verificando el grado de supresión de la grasa. Estos resultados dieron errores en imágenes que contenían mamas densas, lo que significa que, al intentar sintetizar tejidos de alta densidad tisular, la red puede ser limitada.

Otro de los trabajos es presentado por Ming et al. [58]. Aquí se trabajó con imágenes de DCE-MRI, donde resaltan una de las mayores limitantes, la cual es el gran tiempo que toma la adquisición de todas las imágenes del estudio. Proponen un enfoque de síntesis de características que se basó en una red generativa adversarial (GAN) paramétrica cruzada (CPGANFS) para generar características DCE-MRI discriminativas, las cuales parten de imágenes T2 saturadas en grasa (T2W). La idea es tratar de repotenciar la capacidad sintética de la GAN para recrear características puntuales de la imagen. Emplean una GAN de Wasserstein con una penalización de gradiente para diferenciar las características generadas de las características reales extraídas de DCE-MRI. El modelo sintetizado logró un rendimiento estable. La visualización del modelo muestra que el método empleado mejora el poder predictivo al hacer resaltar la atención a la lesión [58].

Uno de los trabajos en el contexto de cambio de dominio entre modalidades de imagen son los de Jiang et al. en el que se combina la mamografía de campo completo (FFDM) con la tomosíntesis digital (DBT) [59; 60]. De acuerdo con los autores, en algunas ocasiones se presentan resultados en los que se dan efectos positivos sin ser realmente positivos, a este tipo de respuestas se les conoce como falsos positivos. Usar la combinación de estos estudios, ayuda a reducir la tasa de falsos positivos en la detección del cáncer de mama. Sin embargo, esta estrategia es de uso complejo, debido a las altas dosis de radiación que deben recibir los pacientes. Así, para reducir la dosis de radiación que requiere el uso de estas dos modalidades de imagen los autores proponen el uso de modelos generativos para sintetizar mamografías digitales a partir de estudios de tomosíntesis.

Zhang et al [61] tratan de abordar el desafío de generar imágenes de tomografía computarizada (TC) a partir de imágenes de cono de haz (CBCT) en pacientes con cáncer de mama, utilizando un enfoque de aprendizaje multimodal de desenredo de representación no supervisada (MURD) que descompone la información en las imágenes y mejora la calidad de las imágenes sintéticas generadas. Los experimentos realizados con un conjunto de datos de

cáncer de mama demuestran que el enfoque MURD logra un buen rendimiento al generar imágenes de TC sintéticas con alta precisión y calidad visual [61].

La estrategia propuesta por los autores en [59; 60] consiste en una arquitectura denominada GGGAN (*Gradient Guided cGAN*) cuya base es una arquitectura Pix2PixHD [62] modificada en la que se agrega a la entrada del discriminador los gradientes de la mamografía digital, calculados con un filtro Sobel. Adicionalmente, la función de pérdida del generador se modifica para considerar la pérdida entre los mapas de características generados por el discriminador entre la imagen real y la imagen sintetizada. De acuerdo con los autores, la arquitectura propuesta mejora el problema de suavizado que produce una DCNN y preserva mejor los bordes, comparada con una Pix2PixHD.

Una de las propuestas que incluye el uso de DCE-MRI es la de Wang et al. [56]. El objetivo de este trabajo es desarrollar un modelo de aprendizaje profundo para sintetizar las primeras fases de secuencias dinámicas (FP-Dyn) con el propósito de compensar la falta de información en las exploraciones de resonancia magnética de mama que no utilizan contraste. Se desarrolló un modelo llamado "Enhance Border Lifelike Synthesize" (EDLS) en el conjunto de entrenamiento, y se utilizó para generar imágenes de FP-Dyn a partir de las imágenes T1WI en el conjunto de validación. El modelo EDLS superó en rendimiento a los modelos convencionales, según los resultados de las métricas cuantitativas. Los resultados subjetivos mostraron una consistencia visual entre las imágenes FP-Dyn generadas y las imágenes originales [56].

Otro tipo de arquitectura que se ha propuesto en la literatura para el transformación de imágenes médicas es la *Collaborative Adversarial Network* (CAN) [33]. Este tipo de arquitectura se compone de tres redes que trabajan de manera colaborativa: un generador, un discriminador y un inversor. Específicamente, el generador usa una máscara de segmentación de una lesión y sintetiza el parche correspondiente a la lesión en la mamografía; el inversor hace el proceso contrario, dado un parche de la mamografía, este segmenta la lesión presente en dicho parche; por su parte el discriminador juega el papel regular de una GAN, determinar si los pares de parches y sus correspondientes máscaras de segmentación son reales o no. La importancia de esta arquitectura es que usa más de dos redes de manera simultánea, abriendo paso a arquitecturas que son aún más complejas. La Figura 2-9 muestra la estructura general de la arquitectura CAN.

Otra investigación que usa la Pix2Pix como base para el cambio de dominio entre modalidades de imagen fue desarrollada por Funjioka et al. [63]. En ese trabajo los autores buscan mejorar la calidad de un tipo de imagen de mama conocido como tomografía por emisión de positrones de recuento bajo (dbPET). Específicamente, la arquitectura Pix2Pix fue entrenada para generar imágenes de alta calidad, capturadas con tiempos de adquisición altos, a partir de imágenes de baja calidad, capturadas con tiempos de adquisición bajos. De acuerdo a la evaluación, tanto cualitativa como cuantitativa, la red entrenada puede generar

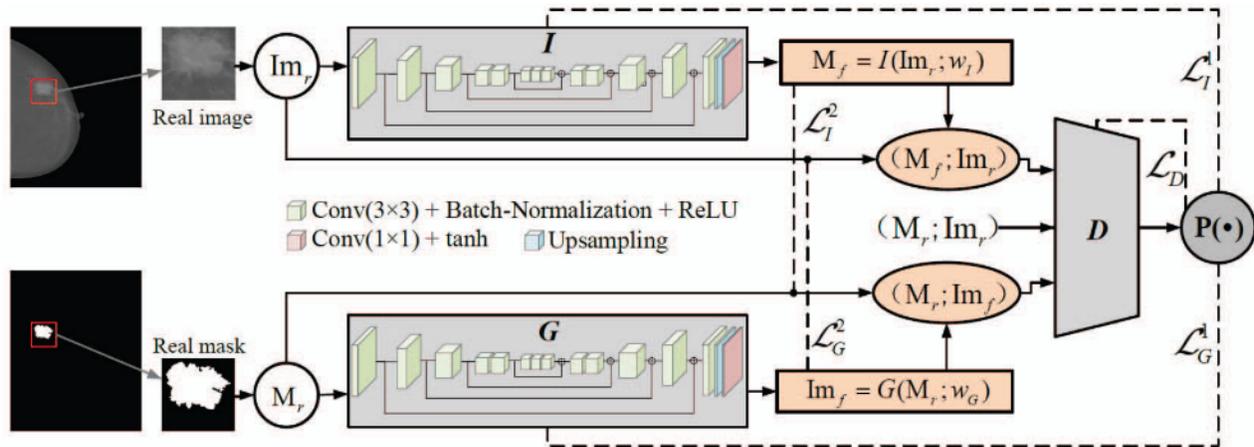


Figura 2-9: Estructura general de la arquitectura CAN. El discriminador toma un par de parches, el parche generado por el generador a partir de una máscara, y el parche de segmentación, generado a partir de un parche de la imagen real. Fuente:[33]

imágenes de alta calidad desde la perspectiva de radiólogos expertos.

Respecto a estudios de imágenes de CEDM, Jiang et al. [64] proponen el uso de una arquitectura basada en CycleGAN [49] para generar imágenes recombinadas a partir de imágenes de baja energía. La arquitectura propuesta en ese trabajo es denominada cGSNT y consta de dos generadores tipo U-Net que buscan aprender a transformar una imagen de baja energía en una imagen recombinada y viceversa; y dos discriminadores que buscan determinar si los pares de imágenes son reales o falsos. Los dos generadores de la arquitectura usan tanto la vista craneocaudal (CC), como la vista mediolateral ablicua (MLO) de las imágenes de CEDM en el proceso de generación de las imágenes recombinadas. Cada generador usa dos codificadores, uno para codificar el contenido de las imágenes (la estructura espacial) y otro para codificar los atributos de las imágenes (la información de las intensidades). El decodificador del se encarga de fusionar la información de los dos codificadores en el proceso de generación. De acuerdo con los autores, la arquitectura propuesta produce imágenes recombinadas de alta calidad que resaltan de manera prominente las lesiones del cáncer.

Otra implementación de la síntesis de imágenes de diferentes modalidades fue propuesta por Kim et al. [65]. Este trabajo se centra en desarrollar un método para obtener imágenes de DCE-MRI en mama sin utilizar agentes de contraste, con el fin de reducir el riesgo para los pacientes. Se propone el uso de una red generativa adversaria (GAN) para crear imágenes de resonancia magnética con contraste mejorado. Esta red utiliza un enfoque especializado que se centra en la región tumoral en particular, además de toda la mama. Se incluye una red de segmentación que realiza una tarea relacionada de identificación de la región tumoral.

También se utilizan mapas de bordes para proporcionar información detallada sobre la forma y la estructura de las imágenes [65].

En un contexto diferente a las imágenes de mama, Yu et al. [29] presentan un modelo que utiliza las características de las texturas de los tejidos en imágenes de resonancia magnética de cerebro. A diferencia de los trabajos anteriores, este se centra en minimizar la diferencia entre las intensidades en el espacio de los píxeles [29] usando la información de los bordes de los tejidos. La arquitectura propuesta se denominada red generativa adversaria sensible a los bordes o Ea-GAN (*Edge-Aware Generative Adversarial Network*) y consta de un generador tipo U-Net, un discriminador y un extractor de bordes que utiliza las máscaras del operador Sobel. Existen dos aproximaciones en esta arquitectura: gEa-GAN, en la que la información de los bordes sólo se considera en la función objetivo del generador, lo que obliga a que la imagen sintetizada tenga un mapa de bordes similar a la imagen real; y dEa-GAN, en la que la información de bordes se incorpora tanto en el generador como en el discriminador, lo que permite que la información de bordes se aprenda de manera adversa. La figura 2-10, describe de manera general la arquitectura de ambas aproximaciones [29].

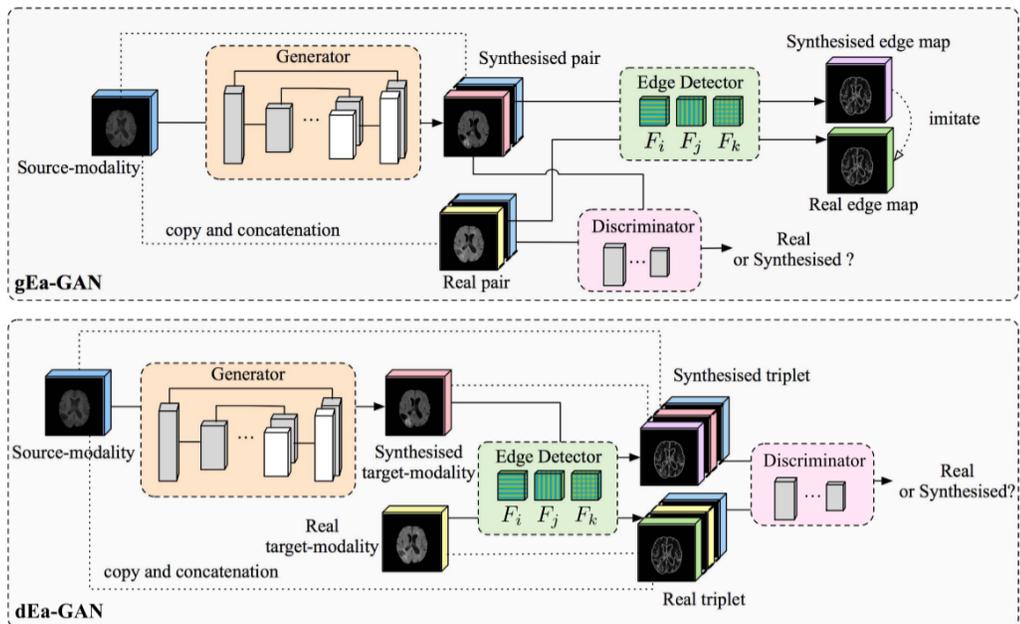


Figura 2-10: Arquitectura Ea-Gan. En la versión gEa-GAN la información de los bordes se integra en el generador para apoyar el proceso de aprendizaje; y en la dEa-GAN la información de los bordes se integra tanto al generador como en el discriminador. Fuente: [29]

A modo de resumen las Tablas 2.1 y 2.2 enumeran los principales modelos descritos en la literatura para la generación imágenes médicas artificiales. Ahora, a pesar de que en la

literatura existen diferentes modelos propuestos para la síntesis de imágenes médicas, aún se evidencian algunas limitaciones:

- A diferencia de las imágenes de cerebro, las imágenes de la mama tienen mayor variabilidad en las texturas y en el tipo de estructuras microscópicas que se presentan en el tejido. Esto implica que las funciones objetivo convencionales (como la norma L1 o L2) no puedan generar adecuadamente estas estructuras, lo que resulta en imágenes borrosas o con valor diagnóstico impreciso.
- Las imágenes de mama poseen una mayor variabilidad en densidad de tejido, lo que implica que la visualización del medio de contraste, en función de la intensidad de los píxeles, disminuye lo que hace poco viable usar los algoritmos de detección de bordes porque se vuelven menos efectivos.
- A pesar de que incluir información adicional en los modelos parezca una tarea trivial, ello conlleva un ajuste importante de las arquitecturas y de las estrategias de entrenamiento para lograr la convergencia de los modelos.

Estas limitaciones motivaron el uso de tres modelos de referencia del estado del arte (Pix2Pix, RIED-Net y Ea-GAN⁹) para crear tres nuevos modelos ensamblados (RiedGAN, gEa-RiedGAN, dEa-RiedGAN) y proponer dos nuevos modelos (G-RiedGAN y D-RiedGAN) para abordar las particularidades de las imágenes contrastadas de la mama. La descripción de cada uno de estos modelos se presenta en el capítulo siguiente.

Red	Año	Tipo Imagen	Resultado principal	Referencia	N° Citaciones
DCGAN	2020	Mamografía	Generación de imágenes realistas funcionales para aumento de datos.	[51]	11
cGAN	2019	Mamografía	Generación de imágenes con lesiones a partir de imágenes sin lesiones.	[48]	32
cGAN	2020	DCE-MRI	El proceso presentó algunos inconvenientes al intentar sintetizar tejidos de alta densidad tisular donde se presentó como limitada.	[18]	27
CycleGAN	2019	Tomografía	Generación de imágenes para segmentación de lesiones	[14]	354
TMP-GAN	2022	Mamografía	genera imágenes sintéticas de baja calidad a las cuales se les mejora su calidad hasta obtener imágenes sintéticas con una textura continua en el área de la lesione	[50]	86
VAE	2023	DCE-MRI	Buen hallazgo de lesiones gracias a la calidad en el aumento de datos.	[53]	2
U-net	2021	DCE-MRI	Generación de imágenes en las que existe una correspondencia positiva entre la parte estructural y la distribución de los píxeles	[55]	2
SD-CNN	2018	CEDM	La arquitectura propuesta aprende a sintetizar parches de imágenes recombina- das a partir de parches de imágenes de baja energía	[16]	171

Tabla 2.1: Modelos para la generación de imágenes médicas sintéticas

Red	Año	Tipo Imagen	Resultado principal	Referencia	N°Citaciones
RiedNet	2020	CEDM	La arquitectura genera parches de imágenes recombina- das a partir de parches de imágenes de baja energía	[17]	56
CPGANFS	2023	DCE-MRI	La visualización del modelo muestra que el método em- pleado mejora el poder pre- dictivo al hacer resaltar la atención a la lesión	[58]	1
MURD	2023	Tomografía	Generación de imágenes de TC sintéticas con alta pre- cisión y calidad visual	[61]	8
GGGAN	2019/2021	CEDM	Se solucionan problemas de suavizado que pueden darse por los generadores con ca- pas muy profundas	[59][60]	13 y 25
EDLS	2021	DCE-MRI	Consistencia visual entre lo generado y lo predicho	[56]	19
CAN	2020	Mamografía	La importancia de esta ar- quitectura es que usa más de dos redes de manera si- multánea, abriendo paso a arquitecturas que son aún más complejas	[33]	32
Pix2Pix	2022	Tomografía	De acuerdo con las evalua- ciones cualitativas y cuan- titativas, la red entrenada puede generar imágenes de alta calidad desde la per- spective de radiólogos ex- pertos.	[63]	3
Ea-GAN	2019	MRI(cerebro)	La imagen sintetizada con- sigue un mapa de bordes similar a la imagen real	[29]	165

Tabla 2.2: Modelos para la generación de imágenes médicas sintéticas

3 Desarrollo Metodológico

La metodología de desarrollo de este trabajo partió por la implementación de una línea base de tres modelos con el objetivo de sintetizar imágenes postcontraste a partir de imágenes precontraste en estudios de DCE-MRI. Después, estos modelos base se ensamblaron para analizar la mejora en el proceso de síntesis de las imágenes postcontraste. Finalmente, se propusieron dos modelos adicionales que buscan sacar provecho de los modelos de línea base en la mejora de la síntesis de las imágenes postcontraste.

Como el objetivo de este trabajo es generar imágenes sintéticas postcontraste que repliquen las características de las estructuras anatómicas, los tejidos y las zonas de captación del contraste presentes en las imágenes reales en estudios de DCE-MRI, es necesario describir en este Capítulo la base de datos de imágenes usada para el entrenamiento de los modelos, ahondar en la descripción e implementación de los modelos de la línea base, de los modelos ensamblados y, por supuesto, de los modelos propuestos.

3.1 Base de datos

Partiendo del objetivo de este trabajo, los modelos de aprendizaje profundo que se usan en este trabajo se entrenan para que a partir de una imagen precontrastada, (x) , se generen imágenes postcontraste sintéticas, $(G(x))$, tal que estas sean similares a las imágenes reales (y) obtenidas con el uso del medio de contraste en estudios de DCE-MRI. Específicamente, como la reacción al medio de contraste se analiza en distintos instantes de tiempo, en este trabajo se usa el tipo de imagen que contiene la respuesta temprana al medio de contraste, ya que es allí donde se manifiesta la reacción en su mayor intensidad [40].

Teniendo en cuenta lo anterior, en este trabajo utiliza una base de datos privada, retrospectiva, y anonimizada de estudios de DCE-MRI de diversos pacientes que fue obtenida a partir del proyecto P20213 de Instituto Tecnológico Metropolitano de Medellín [66]. La base de datos se compone de un conjunto de 150 estudios, que incluyen imágenes estructurales potenciadas en T1 y T2, imágenes de difusión (DWI), y de DCE (6 dinámicos). Para esta investigación, se tomó la respuesta en la secuencia T1 con saturación de grasa antes de la aplicación de medio de contraste, como imagen precontraste (x) , correspondiente al dinámico 0, y esta misma imagen con la aplicación posterior del medio de contraste (y) , cor-

respondiente a un dinámico 2. Para la adquisición de los estudios se usaron distintos tipos de resonadores de 1.5T junto con compuestos basados en gadolinio con dosis entre 0.014 y 0.016 *ml/mol*.

Es importante resaltar que la preparación de la base de datos consistió en generar un subconjunto de datos, con un número balanceado de estudios normales, es decir, sin respuesta a contraste; benignos, con lesiones que contrastan ante el medio de contraste pero categorizadas como benignas; y malignos, con lesiones que responden a la aplicación del medio de contraste y que fueron categorizados como malignas. Además, esta base de datos retrospectiva cuenta con regiones anotadas de lesiones en el tejido y lecturas por expertos radiólogos. De estos estudios se extrajeron únicamente las imágenes con las lesiones anotadas y según cada estudio, se tomaron slices posteriores y anteriores en donde estas mismas eran visibles para los radiólogos.

Esto deriva en un subconjunto de datos de 937 imágenes, de las cuales 718 imágenes se destinan para el entrenamiento de los modelos de aprendizaje y 219 imágenes para validarlos. Inicialmente, estas imágenes cuentan con tamaño variados, que iba de 480x480 píxeles hasta 512x512 píxeles. En términos de preprocesamiento de las imágenes, estas fueron inicialmente reescaladas a 240x240 píxeles usando interpolación bicúbica. Después, los valores de intensidad de las imágenes fueron normalizados en el rango $[-1, 1]$.

3.2 Modelos de línea base

Inicialmente se partió por la implementación de la línea base de tres modelos para la síntesis de imágenes postcontraste. Esos modelos se describen en esta sección y fueron seleccionados para su implementación por sus notables resultados en la síntesis de imágenes médicas en distintas modalidades.

3.2.1 RiedNet: Residual Inception Encoder Decoder Neural Network

La RiedNet es una red construida sobre una arquitectura tipo U-Net que consta de un codificador para comprimir la información y extraer las características salientes de los objetos en la imagen y un decodificador para reconstruir la imagen a partir de los mapas de características comprimidos. A diferencia de la U-Net, la RiedNet usa filtros convolucionales y deconvolucionales para comprimir de descomprimir la información en el codificador y decodificador, respectivamente. Adicionalmente, la RiedNet considera el uso de bloques residuales después de la primera convolución de cada bloque, a fin de reducir el problema de desvanecimiento del gradiente que se produce por la profundidad de la red [16], la estructura de este bloque residual se evidencia en la Figura 3-1. Como fue definida en su implementación original,

la RiedNet se entrenó para sintetizar parches de imágenes; sin embargo, en este trabajo la RiedNet fue entrenada para sintetizar imágenes completas de estudios de DCE-MRI.

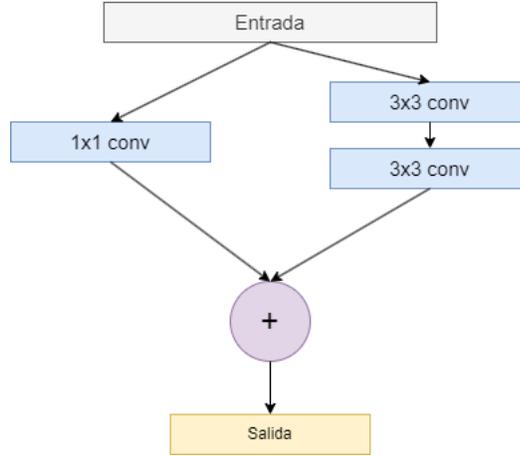


Figura 3-1: Bloque de inicio residual implementado para la RiedNet. Fuente: Elaboración propia

Para adaptar dicha red al tipo de imágenes de este trabajo, se modificaron las funciones de activaciones en algunas capas. La función de activación de las capas más externas y cercanas a la capa de entrada era originalmente tipo ReLU, la cual elimina los valores negativos y mantienen los valores positivos. Esta función fue reemplazada por Leaky ReLU, la cual introduce una pendiente pequeña en el dominio negativo de la función a razón de no eliminarla completamente [67]. También se modificó la función de activación de la capa de salida y en lugar de una función sigmoide, se utilizó la función tangente hiperbólica. Igual que el cambio anterior, esto se hizo para mantener en las imágenes de salida el rango de valores de las imágenes de entrada. La Figura 3-2 muestra el detalle de la arquitectura RiedNet implementada, mientras que la ecuación 3-1 define la función de pérdida que guía el aprendizaje de la red, el cuál está basado en el error medio absoluto o *mean absolute error* (MAE).

$$F(y, x) = \frac{1}{N} \|y_i - G(x_i)\| \quad (3-1)$$

En esta ecuación, y corresponde a la imagen real con contraste; x corresponde a la imagen sin contraste y $G(x)$ es la imagen sintetizada.

3.2.2 Pix2Pix

El modelo Pix2Pix fue propuesto por Isola et al. en [15]. Pix2Pix es un modelo con una arquitectura GAN condicional que fue propuesto para la traducción de imágenes naturales

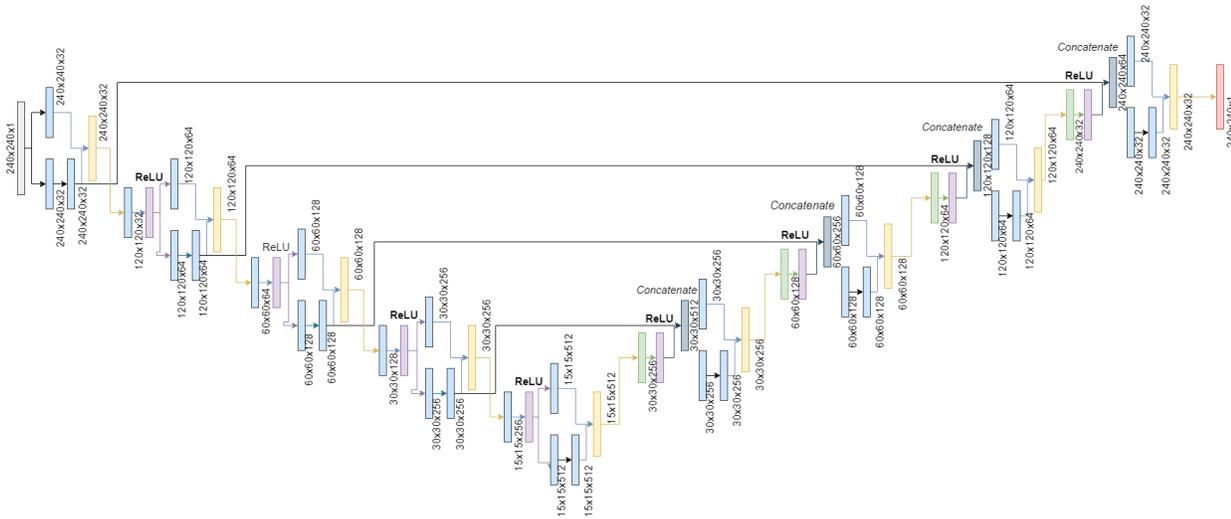


Figura 3-2: Arquitectura RiedNet implementada. En esta arquitectura los bloques azules representan las capas de convoluciones, los bloques morados corresponden a las funciones de activación, los bloques amarillos representan la operación de adición, la cual es una operación que se hace sobre la información que proviene del bloque de convolución completo antes de pasar al siguiente bloque, los bloques azul oscuro son el momento de concatenación. Fuente: elaboración propia.

y cuyo uso se ha extendido en el dominio de imágenes médicas [18; 63]. Este modelo fue el primero en incluir un generador condicional tipo U-Net, que hace uso de los autoencoders. De acuerdo con los autores, esto permite que las imágenes generadas sean notablemente coherentes con la condición establecida en el dominio espacial, ya que combina la función de pérdida adversaria junto con la función de pérdida al nivel de los píxeles de los autoencoders. La Figura 3-3 ilustra la arquitectura tipo U-Net del generador de la Pix2Pix implementada en este trabajo.

El modelo Pix2Pix también introduce el uso de un discriminador Markoviano [15] que pondera y valida el realismo de la imagen a partir de su contenido local usando parches. Cada cadena de parches obtiene un índice que es verificado y optimizado para obtener regiones más realistas en comparación con el discriminador convencional de una GAN. A este discriminador es al que se le denomina PatchGAN y su estructura general, adaptada a este trabajo, se muestra en la Figura 3-4.

Como la arquitectura general de la Pix2Pix es una GAN, esta usa una función de pérdida adversaria que mide, en cada iteración, la precisión del generador y la capacidad del discriminador diferenciando entre imágenes reales y falsas. La formulación de la función de pérdida adversaria para la Pix2Pix implementada está definida en las ecuaciones 3-2 y 3-3.

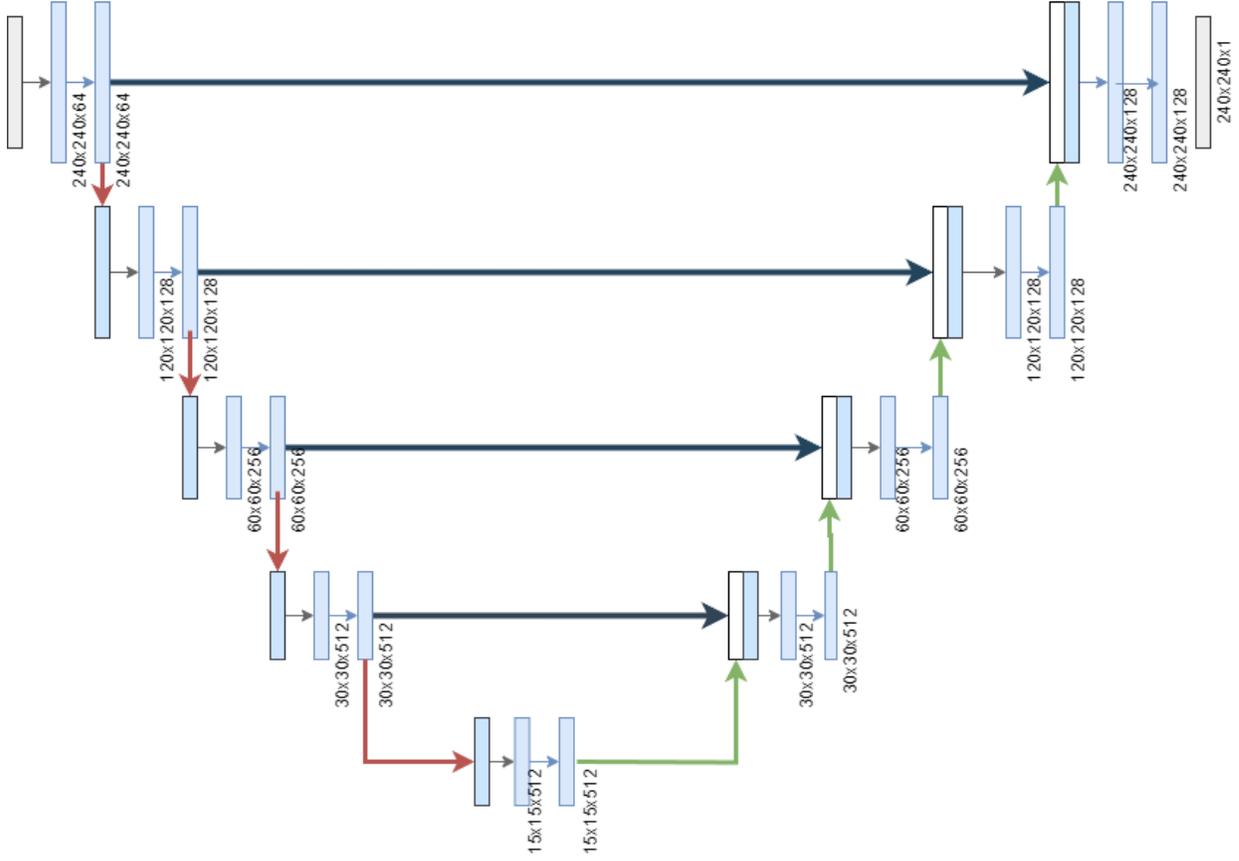


Figura 3-3: Detalle de la U-Net implementada en la arquitectura Pix2Pix de este trabajo. Los bloques azules representan las convoluciones, las flechas rojas el proceso de *maxpooling*, las flechas verdes representan el proceso de expansión, y las flechas azules la concatenación. Fuente: elaboración propia.

$$\mathcal{L}_{cGAN}^G = \mathbb{E}_x [\log(1 - D(x, G(x)))] + \lambda_{L1} \mathbb{E}_{x,y} [||y - G(x)||_1] \quad (3-2)$$

$$\mathcal{L}_{cGAN}^D = -\mathbb{E}_{x,y} [\log D(x, y)] - \mathbb{E}_x [\log(1 - D(x, G(x)))] \quad (3-3)$$

Por un lado, la Ecuación 3-2 calcula la pérdida del generador: el primer término de la ecuación establece la pérdida adversaria a partir de la capacidad del discriminador (D) para identificar si el conjunto de imágenes $[x, y]$ y $[x, G(x)]$ son reales o falsas. El segundo término de la Ecuación 3-2 define la pérdida, a bajo nivel, entre y y $G(x)$. En este término, λ_{L1} corresponde al factor de ponderación de la pérdida de píxeles a la función de costo total. Por otro lado, la Ecuación 3-3 establece la pérdida del discriminador

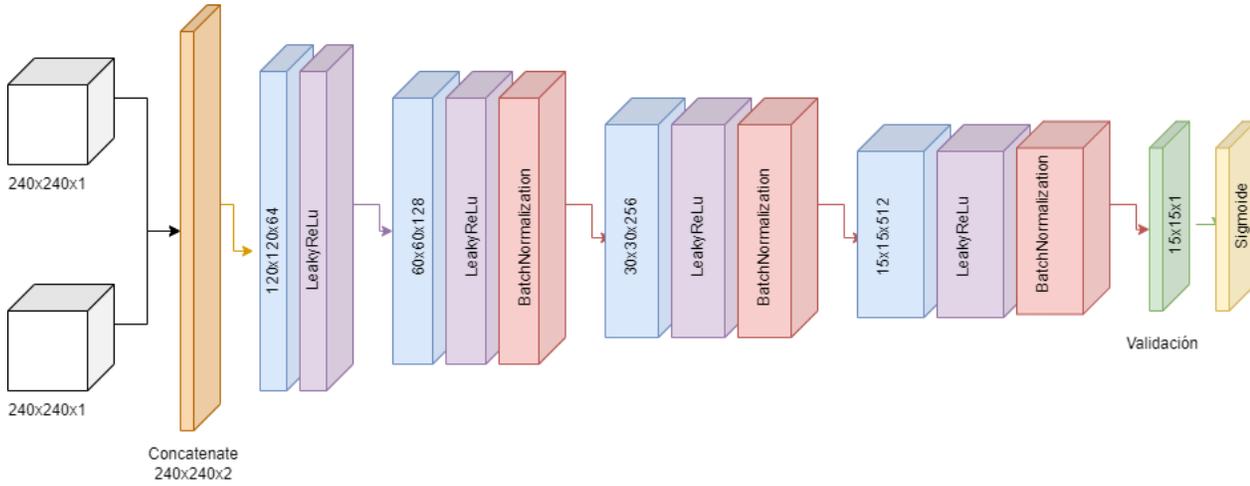


Figura 3-4: Arquitectura PatchGAN. Los bloques blancos iniciales representan los datos de entrada de la red, los cuales se concatenan a través del bloque amarillo. Los bloques azules representan las convoluciones; los bloques morados representan la función de activación y los bloques la normalización de los datos. Fuente: elaboración propia.

3.2.3 Ea-GAN: Edge-Aware Generative Adversarial Networks

Este modelo está basado en una Pix2Pix, por lo que incluye un generador tipo U-Net y un discriminador tipo PatchGAN. La mayor contribución de este modelo es la integración de las diferencias entre los bordes de la imagen sintetizada y la imagen real, calculados con un filtro Sobel. En [29], los autores definen tres filtros para el cálculo de los bordes, F_i , F_j y F_k , donde F hace alusión al filtro aplicado y las letras i , j y k corresponden a las tres dimensiones de una imagen, largo, ancho y profundidad. Estos filtros se aplican a la imagen y y se combinan en un mapa de bordes $S(x)$ a partir de la Ecuación 3-4.

$$S(y) = \sqrt{(F_i * y)^2 + (F_j * y)^2 + (F_k * y)^2} \quad (3-4)$$

En esta ecuación, el operador $*$ representa la convolución entre la imagen y y cada uno de los filtros Sobel establecidos en las direcciones i , j , y k . Para integrar de forma óptima la información de los bordes, se proponen dos variaciones para la Ea-GAN: la primera variación formula que la integración del filtro sea inducida a través del aprendizaje el generador (gEa-GAN), la segunda variación formula que dicha información sea inducida por el discriminador (dEa-GAN).

gEa-GAN: información adicional inducida por el Generador

Con el fin de incluir la información de los bordes en este modelos los autores modificaron pertinentemente la función de pérdida de la red. Para el caso de la gEa-GAN, se modifica la función de pérdida del generador unicamente para que este considere las diferencias entre el mapa de bordes de la imagen generada $S(G(x))$ y el mapa de bordes de la imagen real $S(y)$. La Ecuación 3-5 muestra la función de pérdida del generador de la Pix2Pix, considerando la pérdida entre los mapas de bordes que integra la gEa-GAN. En esta ecuación el término λ_{edge} determina el factor de ponderación de la pérdida de bordes con respecto a la función de costo total.

$$\begin{aligned} \mathcal{L}_{gEa-GAN}^G = \mathbb{E}_x [\log(1 - D(x, G(x)))] + \lambda_{L1} \mathbb{E}_{x,y} [\|y - G(x)\|_1] \\ + \lambda_{edge} \mathbb{E}_{x,y} [\|S(y) - S(G(x))\|_1] \end{aligned} \quad (3-5)$$

dEa-GAN: información adicional inducida por Discriminador

A diferencia del esquema anterior, la dEa-GAN considera la información de los bordes tanto al generador como del discriminador a través de la función adversaria. Esto teniendo en cuenta que el discriminador también puede utilizar los bordes para validar el realismo entre las imágenes reales y sintéticas, lo que a su vez obliga al generador a producir mejores detalles de los bordes durante la síntesis. Así, el discriminador se modifica y recibe, además de las imágenes de entrada y de salida, el mapa de bordes de la imagen de salida $S(\cdot)$. En términos de la función objetivo, el generador de la dEa-GAN mantiene la función de la gEa-GAN, considerando que el discriminador recibe adicionalmente el mapa de bordes, como muestra la ecuación 3-6. Por otro lado, la ecuación 3-7 describe la ecuación de la función de pérdida del discriminador.

$$\begin{aligned} \mathcal{L}_{dEa-GAN}^G = \mathbb{E}_x [\log(1 - D(x, G(x), S(G(x))))] + \lambda_{L1} \mathbb{E}_{x,y} [\|y - G(x)\|_1] \\ + \lambda_{edge} \mathbb{E}_{x,y} [\|S(y) - S(G(x))\|_1] \end{aligned} \quad (3-6)$$

En estas dos funciones, y es la imagen postcontraste real, x es la imagen precontraste condicional, $S(y)$ es el mapa de bordes de la imagen postcontraste real, $G(x)$ es la imagen postcontraste sintetizada a partir de x , y $S(G(x))$ es el mapa de bordes de la imagen sintetizada.

$$\mathcal{L}_{dEa-GAN}^D = -\mathbb{E}_{x,y} [\log(D(x, y, S(y)))] - \mathbb{E}_x [\log(1 - D(x, G(x), S(G(x))))] \quad (3-7)$$

3.3 Modelos ensamblados

Durante el desarrollo de este trabajo, se consideró la posibilidad de modificar los modelos de línea base para que pudiesen funcionar específicamente en el tipo de problema que se aborda en esta investigación, y el tipo de imagen que se trata. Pero, se tuvo en cuenta que si se modifican estos modelos de línea base de manera estructural, para adecuarlos a imágenes en el dominio DCE-MRI, podría generar una tarea que no sería trivial y requeriría de otros estudios y etapas de optimización. Por lo cual, no representó una opción para lograr cumplir con los objetivos dentro del trabajo.

Por este motivo, se optó por validar el proceso sintético para DCE-MRI a partir de otros puntos de vista, como por ejemplo, la combinación de las contribuciones más importantes de cada modelo usando múltiples estrategias para abordar la tarea de síntesis. En esta sección describimos aquellas combinaciones, las cuales hemos denominado como *modelos ensemble*.

3.3.1 RiedGAN: ensamble entre RiedNet y Pix2Pix

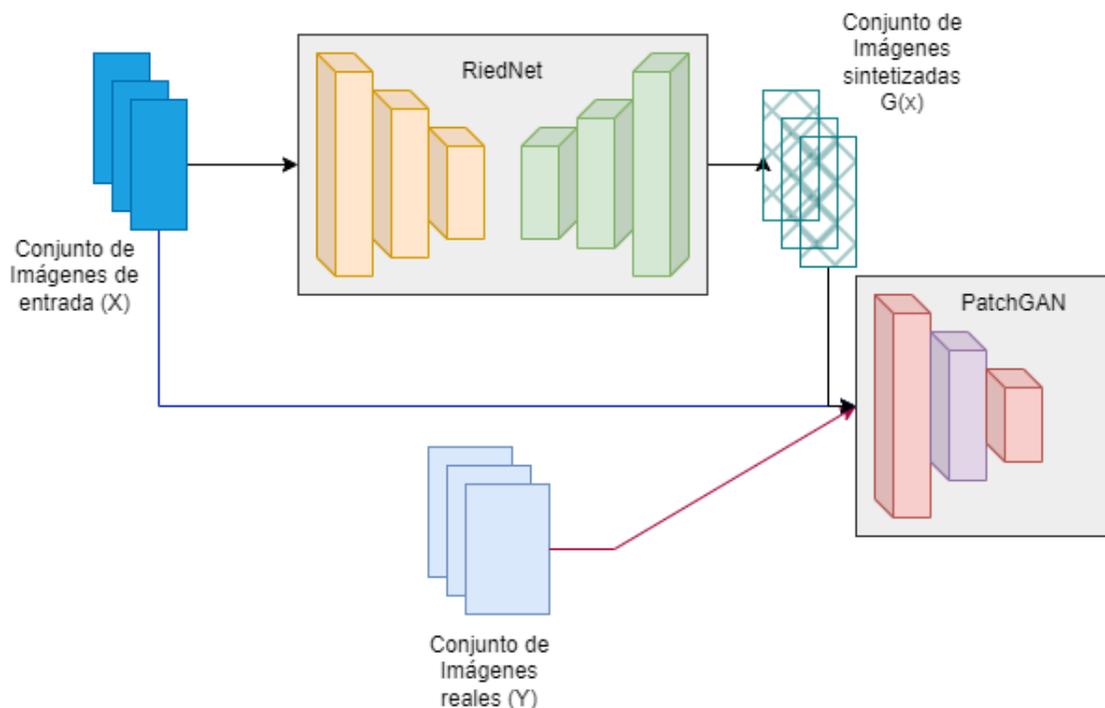


Figura 3-5: Arquitectura RiedGAN. El generador, tipo U-Net corresponde al generador de la RiedNet. Este recibe una imagen precontraste x y genera una imagen post-contraste $G(x)$. El discriminador tiene una arquitectura PatchGAN. Fuente: Elaboración propia

La característica principal de RiedNet es que ha sido originalmente planteada para la síntesis de parches de imágenes, por lo que para sintetizar una imagen completa se debe usar una estrategia de composición de la imagen a partir de parches. Para abordar la limitación sobre la síntesis de una imagen completa, se explora la integración de un discriminador tipo PatchGAN a la RiedNet con el fin de reducir la borrosidad y mejorar la percepción de las imágenes completas a partir del aprendizaje adversario. El ensamble de estas dos redes se ha denominado RiedGAN. Esta red sigue la distribución de una cGAN convencional y su comportamiento es similar al de la Pix2Pix, a diferencia de la estructura del generador.

La arquitectura de la RiedGAN se muestra en la Figura 3-5. En esta se usa el generador de la RiedNet, la cual, al igual que los experimentos anteriores, recibe como dato de entrada a la imagen sin contraste. Además, se usa como discriminador la arquitectura PatchGAN del modelo Pix2Pix. El discriminador recibe un par de imágenes reales que son la imagen con contraste y la imagen sin contraste como condición, y el par generado que es la imagen generada postcontraste y la imagen sin contraste. Además, este modelo se optimiza teniendo en cuenta la función de pérdida adversaria 3-2 y la función de pérdida de píxeles 3-3.

3.3.2 gEa-RiedGAN y dEa-RiedGAN: dos ensambles entre Ea-GAN y RiedGAN

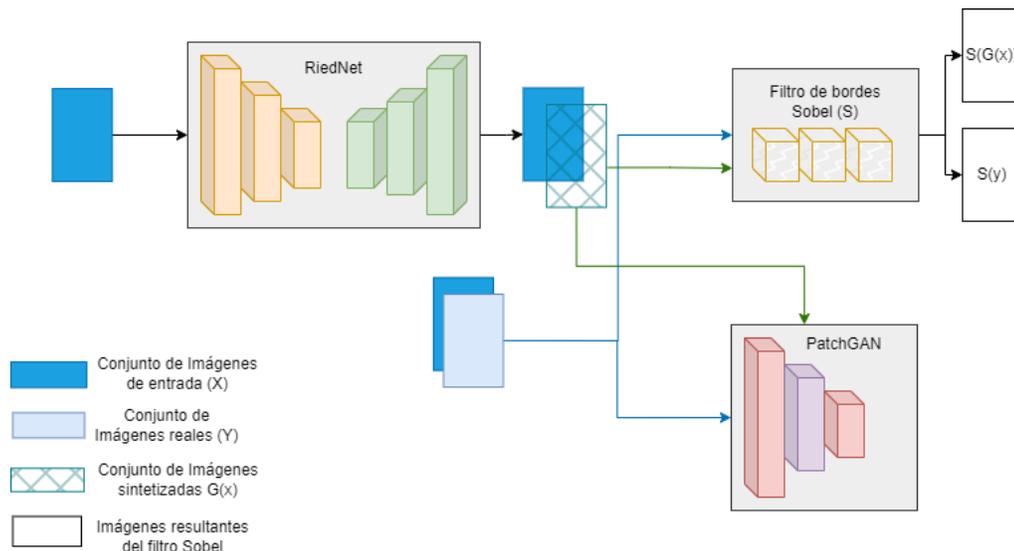


Figura 3-6: Arquitectura gEa-RiedGAN. El generador corresponde a la RiedNet. El discriminador corresponde a la PatchGAN. El filtro de bordes corresponde al filtro Sobel. El generador recibe una imagen precontraste x y genera una imagen postcontraste $G(x)$. El discriminador recibe la imagen real y y la imagen generada $G(x)$. El filtro Sobel genera los filtros de las imágenes generada $G(x)$ y real y . Fuente: Elaboración propia

La Ea-GAN es una red que establece un aprendizaje adversario, característico de las GAN, con la particularidad que integra en el proceso de aprendizaje información sobre los bordes de los tejidos. A fin de mejorar la síntesis de las imágenes, y en particular del efecto del medio de contraste, se propone integrar la información de los bordes a la RiedGAN, de la misma forma que se hace a la Ea-GAN. Así, la arquitectura gEa-RiedGAN agrega el mapa de bordes de la imagen en el generador de la RiedGAN, mientras que la arquitectura dEa-RiedGAN considera la información del mapa de bordes tanto en el generador como en el discriminador. En ambos casos, las funciones de pérdida para ambas arquitecturas son las definidas en las dos variaciones originales de la Ea-GAN. Es decir, la Ecuación 3-5 define la función de pérdida del generador de la gEa-RiedGAN y las Ecuaciones 3-6 y 3-7 definen la función de costo para el generador y le discriminador, respectivamente, para dEa-RiedGAN. En la Figura 3-6 se puede apreciar la arquitectura gEa-RiedGAN, mientras que en la Figura 3-7 se aprecia la arquitectura de la dEa-RiedGAN.

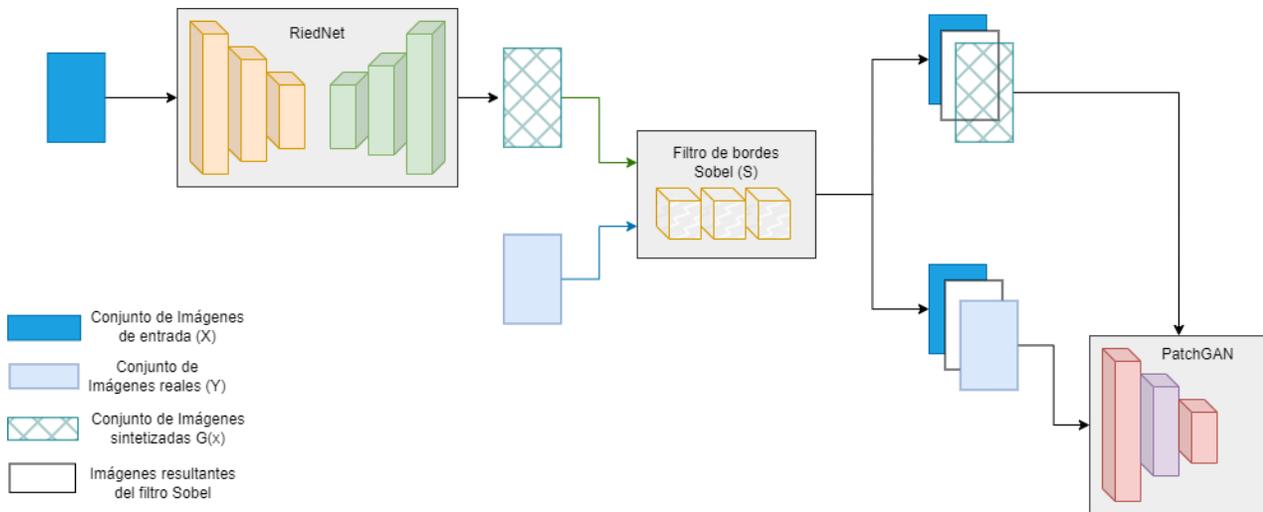


Figura 3-7: Arquitectura dEa-RiedGAN. El generador corresponde a la RiedNet. El discriminador corresponde a la PatchGAN. El filtro de bordes corresponde al filtro Sobel. El generador recibe una imagen precontraste x y genera una imagen postcontraste $G(x)$. El filtro Sobel genera los filtros de las imágenes generada $G(x)$ y real y . El discriminador recibe la imagen real y y la imagen generada $G(x)$ y lo obtenido mediante los filtros. Fuente: Elaboración propia

3.4 Modelos propuestos

Los modelos presentados hasta el momento han sido aquellos que en la literatura han mostrado resultados adecuados en la tarea de generación de imágenes médicas, pero que han sido formulados para la tarea específica de este trabajo. Adicionalmente, se describieron

tres ensambles entre esos modelos base del estado del arte a fin de considerar su uso en la síntesis de imágenes postcontraste en estudios de DCE-MRI.

A partir de los modelos ensamblados descritos, en esta sección se proponen dos nuevos modelos implementados específicamente para la generación de imágenes sintéticas postcontraste de imágenes de DCE-MRI de mama. Los dos modelos propuestos siguen un esquema de entrenamiento adversario con una arquitectura en la que el generador usa la estructura de la RiedNet, y el discriminador consiste en una PatchGAN. A estos modelos propuestos se les agrupa bajo el nombre de RiedGAN. La configuración de la RiedGAN está basada en la arquitectura de la RiedNet y su tarea es aprender a sintetizar tanto las estructuras de los tejidos como el realce de contraste en de las imágenes postcontraste. Esta modificación hace la diferencia con la Ea-GAN en la cual se utiliza la información de los bordes de los tejidos, pero no información sobre el realce de contraste. Es importante notar la capacidad sintética de un modelo que se apoya de información adicional como la Ea-GAN.

3.4.1 Detección de zonas con realce de contraste

La Ea-GAN aborda el problema de la pérdida del detalle en las texturas de los tejidos usando un mapa de bordes que se obtiene tras la aplicación de un filtro Sobel. En el caso de este trabajo, el objetivo no solo consiste en sintetizar los tejidos de la mama, sino también sintetizar la respuesta al medio de contraste. Es así que el modelo propuesto necesita de un filtro que permita detectar las zonas en las se está produciendo un realce de contraste en la imagen postcontraste, respecto a la imagen precontraste, similar a como lo hace la Ea-GAN para obtener información de los bordes de los tejidos.

A nivel de la intensidad de los píxeles, el realce producido por el agente de contraste se puede identificar a partir de los píxeles que tienen una alta intensidad en la imagen postcontraste. Si bien el proceso de generación se lleva a cabo a partir de una imagen precontraste, durante el entrenamiento se cuenta con la imagen postcontraste, por lo que es posible usar la información del realce de contraste en el proceso de aprendizaje del generador. Aunque en la imagen de entrada de la red no hay contraste, es importante igualmente que el filtro dé preferencia a los puntos de alta intensidad, permitiendo poder orientar los posibles puntos de contraste.

Así, para detectar las zonas de realce del contraste en las que el generador debe prestar atención, se propone el uso de un método de segmentación por umbralización sobre la imagen postcontraste. La umbralización permite crear una máscara binaria en la que se identifican y mantienen los píxeles cuyo valor de intensidad es mayor que un valor dado umbral t , tal

como indica la Ecuación 3-8.

$$FI_y(i, j) = \begin{cases} 1 & \text{si } y(i, j) \geq t \\ 0 & \text{en otro caso.} \end{cases} \quad (3-8)$$

El elemento FI_y denota el filtro empleado. De acuerdo con la Ecuación 3-8, la definición del parámetro t puede cambiar por completo el resultado de la umbralización, es por eso que es un parámetro sensible. En el caso de las imágenes de DCE-MRI, no es posible establecer un valor de umbral fijo t que nos permita identificar las zonas de realce de contraste, esto porque los niveles de intensidad de los píxeles en las imágenes puede variar substancialmente dependiendo del proceso de captura de los estudios.

Para abordar la variabilidad entre el cambio de las intensidades entre estudios, el valor de umbral t se calcula con base en el percentil 90 sobre el histograma de píxeles de la imagen a fin de mantener el 10% de los píxeles con mayor intensidad. El valor del percentil se estableció de manera heurística, pero este parámetro es susceptible de optimización, a pesar de que ello se encuentra fuera del alcance de este trabajo.

Posterior a la umbralización, se aplicaron las operaciones morfológicas de cierre y apertura, en ese orden. La operación de cierre, además de suavizar los contornos de las zonas de realce de contraste detectadas permite eliminar pequeños agujeros en las zonas detectadas y juntar aquellas zonas separadas por pequeñas brechas. Por otro lado, la operación de apertura nos permite eliminar pequeñas salientes de las zonas de realce detectadas y además eliminar zonas de contraste aisladas muy pequeñas. Ambas operaciones se aplicaron utilizando un elemento estructurante con forma elipse de 7×7 . En la Figura 3-8 se muestran el resultado de aplicar la detección del medio de contraste para un paciente aleatorio en la imagen real y sintetizada.

3.4.2 Función de pérdida para las regiones con realce de contraste

Una vez aplicada la detección de zonas de contraste, se pretende minimizar la diferencia entre estas zonas en la imagen real y sintetizada. Dada la naturaleza binaria de las zonas de contraste, es posible emplear algunas métricas de diferencia como el índice de Jaccard [68]. Esta métrica calcula la relación entre la interceptación de las zonas segmentadas en ambas imágenes y la unión de esas mismas zonas, tal y como se ilustra en la ecuación 3-9, en donde FI_y y $FI_{G(x)}$ corresponde a las zonas de realce de contraste real y sintetizados, respectivamente.

$$J_c(FI_y, FI_{G(x)}) = \frac{|FI_y \cap FI_{G(x)}|}{|FI_y| + |FI_{G(x)}| - |FI_y \cap FI_{G(x)}|} \quad (3-9)$$

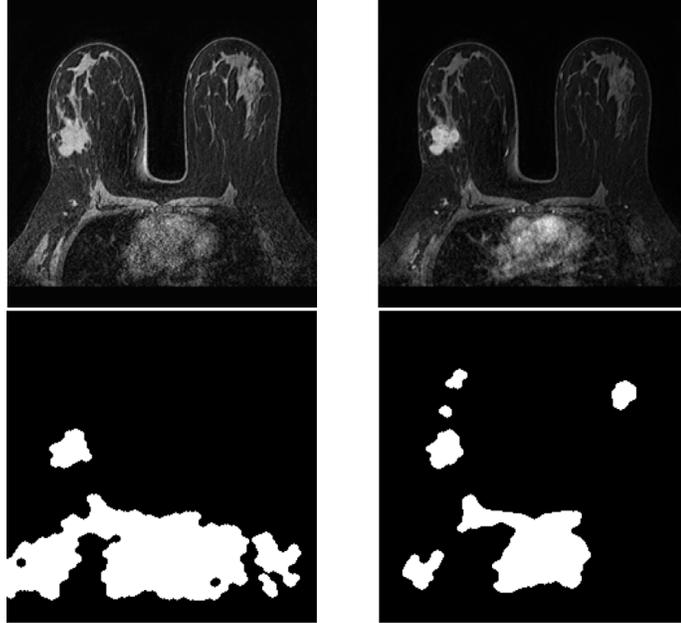


Figura 3-8: Imagen de muestra de detección de zonas de contraste. Esta segmentación binaria es usada como guía en el modelo de aprendizaje para generar las zonas de contraste adecuadas en la imagen sintética.

El resultado del índice de Jaccard es un valor entre 0 y 1, donde un valor cercano a 1 indica una superposición alta con la región de referencia, mientras que un valor cercano a 0 sugiere una falta de coincidencia con la región de referencia.

Sin embargo, el principal inconveniente de esta métrica es que no es una función convexa, por lo que su optimización usando el gradiente descendente durante el entrenamiento puede llevar a soluciones subóptimas o errores de convergencia. Estas limitaciones son abordadas en el trabajo presentado por [68], en el cual los autores utilizan la extensión convexa para funciones discretas. Esta función submodular es estimada a partir del subconjunto de píxeles predichos erróneamente y que en el caso del índice Jaccard han mostrado ser calculable en tiempo polinómico, por lo que su solución corresponde con su extensión de Lovasz (*Lovasz Loss*) [69]. Esta aproximación convexa y continua representa un reemplazo soluble de la función discreta usando métodos de optimización de primer orden, como el algoritmo del gradiente descendente que comúnmente usan las redes neuronales. De esta forma, con la extensión Lovasz ($\overline{\Delta J}_c(K_y, K_{G(x)})$) para el conjunto de predicciones erróneas usando Jaccard ($\Delta J_c(K_y, K_{G(x)}) = m(c)$ por simplicidad) se estima utilizando la ecuación 3-10.

$$\text{Lovasz}(f(c)) = \frac{1}{|C|} \sum_{i=1}^c \overline{\Delta J}_c(m(c)) \quad (3-10)$$

En donde $f(c)$ es la densidad de probabilidad estimada usando la función *Softmax* estimada en la ecuación 3-11 y c es el número de clases en la imagen de segmentación. z corresponde a las salidas de una capa antes de aplicar la función Softmax.

Por último adoptamos la configuración sugerida por los autores, quienes mencionan que el *Lovasz loss* obtiene sus mejores resultados cuando es entrenado en conjunto con la entropía cruzada binaria o binary cross-entropy (*BCE*) la cual se expresa en la ecuación 3-12.

$$f(c) = \frac{e^{z(c)}}{\sum_{c' \in C} e^{z(c')}} \quad (3-11)$$

$$BCE(K_y, K_{G(x)}) = -\frac{1}{C} \sum_{i=1}^c (K_y \cdot \log(K_{G(x)}) + (1 - K_y) \cdot \log(1 - K_{G(x)})) \quad (3-12)$$

3.4.3 G-RiedGAN

Con base en los elementos antes descritos, la Figura 3-9 muestra la arquitectura del primer modelo RiedGAN, denominado G-RiedGAN. Esta arquitectura integra, después del generador, el filtro para la detección de zonas de realce del contraste con el que se busca retroalimentar al generador y guiar su aprendizaje hacia la replica del realce de contraste. En este caso se mantiene inmutable el discriminador PatchGAN el cual debe identificar si el par de imágenes pre y poscontraste que recibe son reales o sintéticas.

La retroalimentación del filtro de detección de las zonas de contraste en las imágenes post-contraste se realiza integrando la pérdida entre las zonas de realce de la imagen real contra las zonas de realce obtenidas en la imagen sintética.

La ecuación 3-13 muestra la función de pérdida propuesta para el generador de la G-RiedGAN. De esta ecuación E es el índice de verosimilitud entre la información expuesta, x es la imagen sin contraste, $G(x)$ es la imagen sintetizada, y es la imagen real o contrastada. El primer término de la ecuación corresponde al cálculo orientado sobre el discriminador, basado en MSE. El segundo término de la ecuación corresponde al cálculo orientado sobre el generador, basado en MAE. Mientras que el último término es el cálculo de la función de costo Lovász. Es importante resaltar de λ son los pesos o puntos de importancia que tiene cada término sobre el resultado en general. En este caso, se le dan pesos de importancia al término Lovász y al término orientado en MAE.

$$\mathcal{L}_{G-RiedGAN}^G = \mathbb{E}_x[\log(1 - D(x, G(x)))] + \lambda_{l1} \mathbb{E}_{x,y}[\|y - G(x)\|_1] + [(1 - \lambda_{lovasz})BCE + \lambda_{lovasz}Lovasz] \quad (3-13)$$

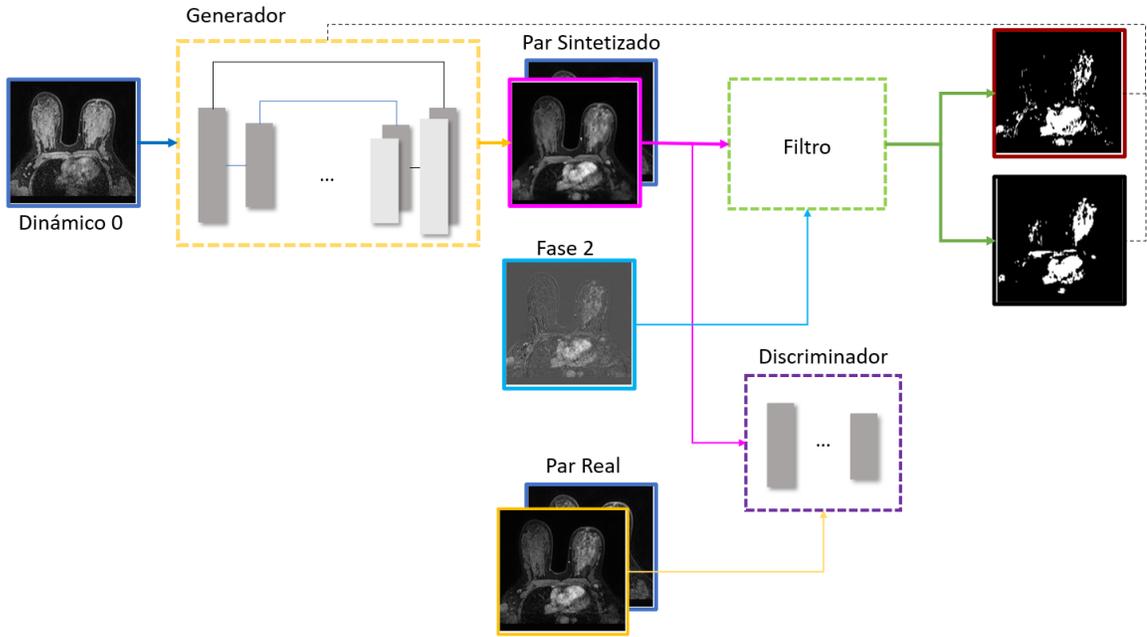


Figura 3-9: Arquitectura G-RiedGAN. Esta arquitectura incluye un filtro para la detección de las zonas de realce de contraste en las imágenes postcontraste que retroalimenta al generador a través de la función de costo. Fuente: Elaboración propia

Como se indicó, se mantiene la función de pérdida del discriminador PatchGAN, la cual se referencia de nuevo en la ecuación 3-14.

$$\mathcal{L}_{G-RiedGAN}^D = -\mathbb{E}_{x,y}[\log D(x, y)] - \mathbb{E}_x[\log(1 - D(x, G(x)))] \quad (3-14)$$

3.4.4 D-RiedGAN

El segundo modelo propuesto se denomina D-RiedGAN. A diferencia del modelo anterior, la información que proviene del filtro de las zonas de realce de contraste se integra tanto al generador como al discriminador. Esto permite que el generador centre más su atención en la síntesis de las zonas de realce de contraste al considerarlas en la contraparte adversaria, lo que en general ayuda a mejorar la calidad de las síntesis del modelo. La arquitectura D-RiedGAN se describe en la Figura 3-10.

Esto permite que el proceso verificador integre en el aprendizaje la existencia de estas intensidades, para que su proceso orientador para el generador, potencia su capacidad. La Figura 3-10 deja en evidencia cómo se visualiza la propuesta.

Note que para realizar el aprendizaje adversario, el discriminador de la D-RiedGAN se modifica para recibir una tripleta de imágenes que son la imagen de entrada, la imagen sintética o real y las zonas de realce de contraste de la imagen real o de la imagen sintética. Las ecuaciones 3-15 y 3-16 muestran las funciones de pérdida del generador y del discriminador, respectivamente.

La función de pérdida del generador D-RiedGAN es la misma que la del generador de la G-RiedGAN. Por otro lado, la función de pérdida del discriminador se modifica y considera también lo calculado a través de Lovász.

$$\mathcal{L}_{D-RiedGAN}^G = \mathbb{E}_x[\log(1 - D(x, G(x), F(G(x)))))] + \lambda_{l1} \mathbb{E}_{x,y}[||y - G(x)||_1] + [(1 - \lambda_{lovasz})BCE + \lambda_{lovasz}Lovasz] \quad (3-15)$$

$$\mathcal{L}_{D-RiedGAN}^D = -\mathbb{E}_{x,y}[\log D(x, y, F(y))] - \mathbb{E}_x[\log(1 - D(x, G(x), F(G(x)))))] \quad (3-16)$$

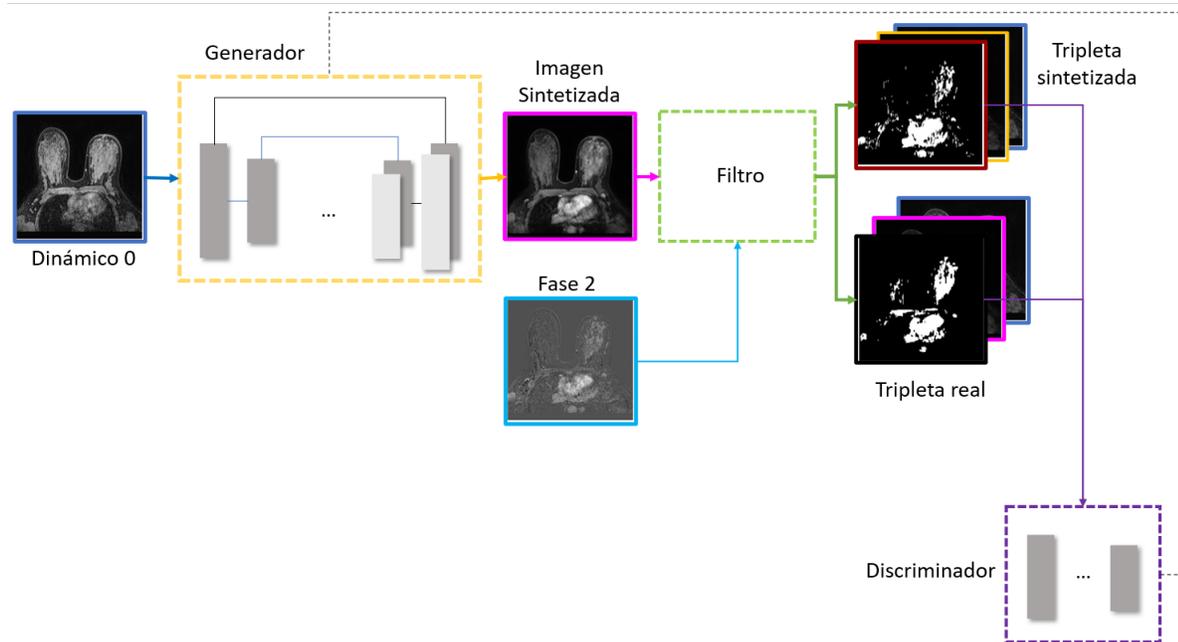


Figura 3-10: Arquitectura D-RiedGAN. La información de las zonas de realce del contraste obtenidas con el filtro retroalimentan al discriminador y al generador a través de las funciones de pérdida. Fuente: Elaboración propia

La arquitectura RiedGAN que se propone este trabajo integra elementos de las arquitecturas RIED-Net y la Pix2Pix bajo la estructura de la EaGAN. A diferencia de esta última,

la RiedGAN integra un filtro para la detección de las zonas de realce de contraste que son las zonas clave de la síntesis en el problema de análisis de imágenes de DCE-MRI para la detección y diagnóstico del cáncer de mama. Las zonas de contraste detectadas en las imagen sintética se usan para guiar el aprendizaje de la red a partir de la función de pérdida Lovász integrada a la función de pérdida del generador y del discriminador.

4 Resultados y Discusión

En esta sección se presentan los resultados de los *modelos de línea base*, los *modelos ensamblados* y los *modelos propuestos*. Previo a ello se describe las métricas de evaluación empleadas en la validación de los resultados.

4.1 Métricas de evaluación

Para validar la calidad de las imágenes sintéticas generadas se utilizaron las métricas cuantitativas más comunes en el estado del arte: *Mean Absolute Error* (MAE), *Peak Signal to Noise Ratio* (PSNR) y *Structural Similarity Index Measure* (SSIM).

4.1.1 Error Medio Absoluto (MAE)

El MAE es una métrica de estimación de la diferencia entre un par de observaciones que describen un fenómeno y que se encuentran en la misma escala [70]. En el caso concreto en imágenes, el MAE estima la diferencia píxel a píxel entre las intensidades de un par de imágenes calculando el promedio entre esas diferencias. Dada su formulación, el MAE tiende a tomar el valor de cero cuando la diferencia entre las dos observaciones es mínima; por el contrario, a medida que la diferencia entre las observaciones es mayor, el valor del MAE aumenta. La ecuación 4-1 muestra la fórmula para el cálculo del MAE entre una imagen generada $G(x)$ y la imagen real y , ambas de tamaño $m * n$ píxeles.

$$MAE(y, G(x)) = \frac{1}{m * n} \sum_{i=1}^m \sum_{j=1}^n ||y_{(i,j)} - G(x)_{(i,j)}|| \quad (4-1)$$

4.1.2 Medida de Proporción Señal a Ruido (PSNR)

El PSNR es una métrica empleada para definir la relación entre la máxima energía que emite una señal de cualquier tipo, y el ruido que afecta la representación de dicha señal [71]. A diferencia del MAE, el PSNR considera la proporción de información que es semejante entre dos imágenes. Esta métrica se expresa en decibeles (dB) y entre mayor es su valor, mayor

es la similitud entre dos imágenes. La ecuación 4-2 presenta el cálculo de esta métrica, considerando la imagen real (y) y la imagen sintetizada ($G(x)$). En esta ecuación MAX_I es el valor de intensidad máximo posible para las imágenes y el MSE es el error cuadrático medio entre las dos imágenes, el cual se calcula como indica la ecuación 4-3

$$PSNR(y, G(x)) = 20 \cdot \log_{10} \left(\frac{MAX_I}{\sqrt{MSE(y, G(x))}} \right) \quad (4-2)$$

$$MSE(y, G(x)) = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n [y_{(i,j)} - G(x)_{(i,j)}]^2 \quad (4-3)$$

4.1.3 Índice de Medida de Similitud Estructural (SSIM)

El SSIM es una métrica que calcula la similitud a nivel estructural entre dos imágenes. El SSIM es considerada una métrica más cercana a la percepción humana, comparada con el PSNR y el MAE. La información estructural considerada por el SSIM considera los píxeles tienen fuertes interdependencias, en especial cuando están espacialmente cerca. Estas dependencias contienen información sobre la luminancia, el contraste y la estructura de los objetos en la imagen y se pueden estimarse de manera conjunta como indica la ecuación 4-4 [71].

$$SSIM(y, G(x)) = \frac{(2\mu_{G(x)}\mu_y + c_1)(2\sigma_{G(x),y} + c_2)}{(\mu_{G(x)}^2 + \mu_y^2 + c_1)(\sigma_{G(x)}^2 + \sigma_y^2 + c_2)} \quad (4-4)$$

En esta ecuación, los términos representados por: $\mu_{G(x)}$, μ_y , $\sigma_{G(x)}$, σ_y y $\sigma_{G(x),y}$, corresponden a las medias locales, desviaciones estándar y la covarianza entre las imágenes $G(x)$ y y , respectivamente.

4.1.4 Mapas de diferencia

Calcular los mapas de diferencia entre una imagen generada y una imagen real implica comparar píxeles individuales para evaluar la discrepancia entre ellas. Para este trabajo, se consideró importante calcular estos mapas debido a que mostrarían qué tantas diferencias existen entre la imagen generada y la imagen real.

Para calcular estos mapas, se empleó la ecuación 4-5. Esta fórmula describe un proceso en el que se analizó cada píxel de las imágenes, y se calculó la diferencia en intensidad entre el

mismo píxel en otra imagen. Esto se hace para entender cuánto difieren las dos imágenes en cada punto.

En términos más simples, cada píxel en una imagen tiene un valor que representa su intensidad, la comparación de píxeles implica restar el valor del píxel correspondiente en una imagen del valor del mismo píxel en otra imagen. El propósito de esta comparación es cuantificar y visualizar las diferencias entre las imágenes.

$$DF = ||y_{(i,j)} - G(x)_{(i,j)}|| \quad (4-5)$$

4.2 Configuración experimental

Con el fin de evaluar los modelos implementados, se utilizan una serie de hiperparámetros que representan los valores usados para controlar el proceso de aprendizaje. Se presentaron hiperparámetros que no cambiaron de un modelo a otro, debido a que representaban puntos cruciales a la hora de hacer seguimiento entre un modelo y otro, como por ejemplo, el número de las épocas de entrenamiento. Los hiperparámetros que no cambiaron están descritos en la Tabla 4.1. Pero además de estos hiperparámetros, se presentaron otros en los que sí se dieron diferentes variaciones que fueron necesarias para la convergencia de los modelos. En la Tabla 4.2 se pueden visualizar estos hiperparámetros.

Hiperparámetro	Valor
Época	100
Optimizador	Adam
Taza de aprendizaje	0.0002
Tamaño de entrada	240x240

Tabla 4.1: Hiperparámetros que contienen el mismo valor para cada modelo computacional implementado

Modelo	Batch	Valor de λ_{L1}	Valor de λ_{edge} o λ_{Lovasz}	Activación Salida
RiedNet	4	N/A	-	Linear
PixToPix	1	100	-	TanH
gEa-GAN	1	300	300	Sigmoide
dEa-GAN	1	300	300	Sigmoide
RiedGAN	4	100	-	TanH
gEa-Ried-GAN	1	100	300	TanH
dEa-Ried-GAN	1	100	300	TanH
G-RiedGAN	1	100	150	TanH
D-RiedGAN	1	100	150	TanH

Tabla 4.2: Hiperparámetros que variaron dependiendo del modelo

Por ejemplo, vale la pena mencionar que las variaciones en el *batch_size*, el cual define cuántas muestras se utilizan en cada paso de entrenamiento, dependen de la capacidad de cómputo limitada para realizar los experimentos. Otro de los hiperparámetros que varía en el entrenamiento de cada red son los valores de λ_{L1} y λ_{edge} o λ_{lovasz} , los cuales brindan el nivel de importancia que tiene cada una de las métricas dentro del entrenamiento. Es importante resaltar que este valor se aplica sobre modelos tipo GAN, ya que en ellos se evalúan de forma paralela varias métricas al tiempo, generalmente dos, una métrica para el generador, y otra para el discriminador, por lo que es importante diferenciar cuál métrica tiene mayor relevancia en el entrenamiento. Para la RiedNet, por ejemplo, este hiperparámetro no aplica, ya que este es un modelo convolucional.

En la mayoría de los casos, los valores propuestos de λ_{L1} se dió en un rango de 100, variando únicamente para los casos de las implementaciones de la eA-GAN, donde se tomó un valor de 300. Además, para todos los casos se dió como valor 1 el peso para la función de costo del discriminador. Finalmente, para las implementaciones que contaron con el uso de información adicional, que implica el uso de una función de costo, también se requería proporcionar un peso para dicha métrica de evaluación. Para el caso de los modelos que emplearon extracción de bordes con el filtro sobel, se implementó un valor de λ_{edge} de 300. Para el valor de las implementaciones en las que se usó extracción de puntos de intensidad, evaluado mediante Lovász, se implementó con un λ_{lovasz} de 150.

Además de las variaciones anteriormente mencionadas que se presentaron en los implementaciones, la capa de activación en la salida también varió dependiendo del modelo. Para el caso de la RiedNet se implementó una función de activación lineal, mientras que para las implementaciones de la eA-GAN se usó la función de activación sigmoide. Finalmente, para el caso de las implementaciones propuestas y los modelos ensamblados se empleó tangente hiperbólica, esto con el fin de poder contener información tanto de valores negativos como de valores positivos.

4.3 Evaluación comparativa

En términos de los resultados cuantitativos, la Figura 4-1 muestra un gráfico de dispersión en el cual se compara el PSNR y el SSIM obtenido por los modelos sobre el conjunto de validación. Además, se evidencian los valores de MAE obtenidos por cada uno de los modelos.

De acuerdo con los resultados obtenidos, el modelo D-RiedGAN obtiene mejor desempeño en términos del PSNR y el SSIM, seguido de la G-RiedGAN. En cuanto al MAE, la RiedGAN obtiene imágenes con el menor valor, seguido de la D-RiedGAN. Sin embargo, a pesar de que la RiedGAN obtiene el mejor desempeño en MAE, su resultado en PSNR y SSIM no son comparables con otros modelos. Esto podría deberse a que la función adversaria de la RiedGAN incrementa el realismo de las imágenes a través de las componentes de alta.

Además, puede notarse también en las imágenes generadas como la RiedNet obtiene imágenes más borrosas en contraste con D-RiedGAN, típico de los modelos que emplean únicamente la métrica de píxeles como función de costo.

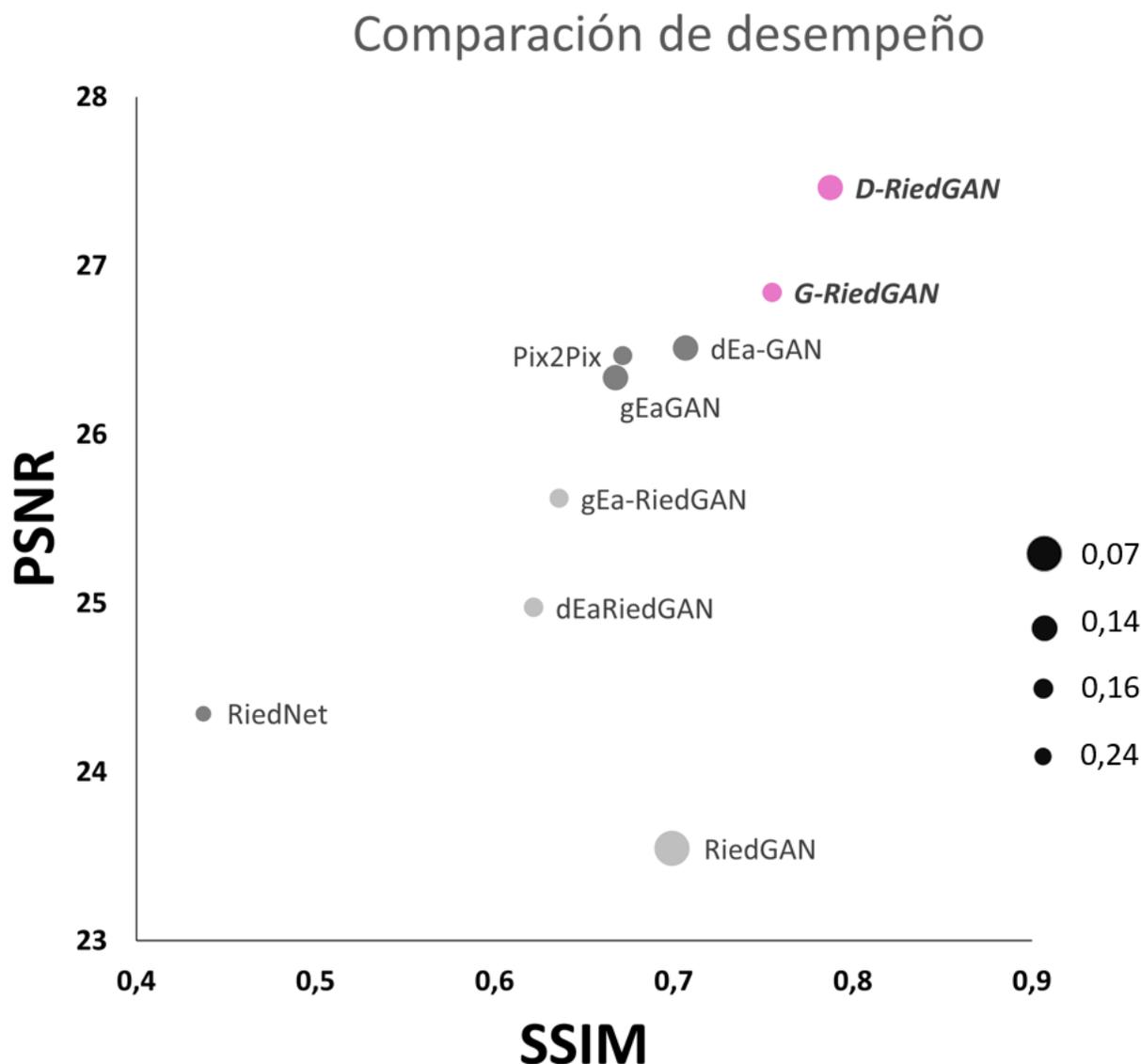


Figura 4-1: Gráfico de dispersión del PSNR y el SSIM de los modelos base (círculo gris oscuro), los modelos ensamblados (círculo gris claro) y los modelos propuestos (círculo rosa). Para las métricas PSNR y SSIM, cuanto más alto es el valor, mejor es el resultado. Al lado izquierdo, se evidencia el MAE, que cuanto más bajo es el valor, mejor es el resultado. Fuente: Elaboración propia.

Es posible observar que D-RiedGAN presenta mayor desempeño en la generación de imágenes en comparación con los demás modelos. Además, es importante resaltar que el desempeño

obtenido por los modelos ensamble obtienen un desempeño incluso menor que los modelos base. Esto sugiere que el problema de síntesis de imágenes de DCE-MRI no se resuelve de manera trivial, tan solo combinando componentes o módulos de los modelos existentes. Para evaluar también la calidad visual de las imágenes generadas, en la Figura 4-2 se reportan algunas muestras reales y generadas por los modelos de síntesis.



Figura 4-2: Resultados comparativos de todos los modelos. Las proyecciones aumentadas corresponden a regiones extraídas en donde se presenta captación por medio de contraste. Mejor visto al hacer zoom.

Es posible evidenciar que todos los modelos pueden generar eficientemente las estructuras corporales más prominentes como el tórax, el corazón y la forma del órgano. Sin embargo, ellos se diferencian principalmente en las pequeñas estructuras microscópicas internas, o en las zonas que cuentan con captación de medio de contraste. Este detalle es importante debido a que la información importante para el diagnóstico se encuentra usualmente en estructuras microscópicas encontradas en el tejido que presentan captación al medio de contraste. En este sentido, algunos de los modelos base obtienen resultados en los cuales la visualización

de tejido carece de detalles como texturas o componentes de baja frecuencia. Esto debido a que ellos dependen principalmente de la función basada en píxeles, la cual se centra en el promedio de las componentes encontradas en las imágenes originales. Esta tendencia se mantiene en los modelos ensamblados, los cuales demuestran que este problema no depende únicamente de componentes basados en píxeles o en bloques de arquitectura particulares. Sin embargo, los modelos propuestos basados en las componentes diseñadas para resolver el problema permiten obtener resultados superiores a los modelos base y ensamblados, lo cual mostrando su efectividad. Ello puede verse en los mapas de diferencia, específicamente en las regiones ampliadas; en las cuales la diferencia absoluta es menor para el modelo D-RiedGAN.

También se reportan los mapas de diferencia en la Figura 4-3, con el fin de evidenciar aquellas zonas en donde las estructuras de tejido o de contraste se diferencian en las imágenes reales y generadas.

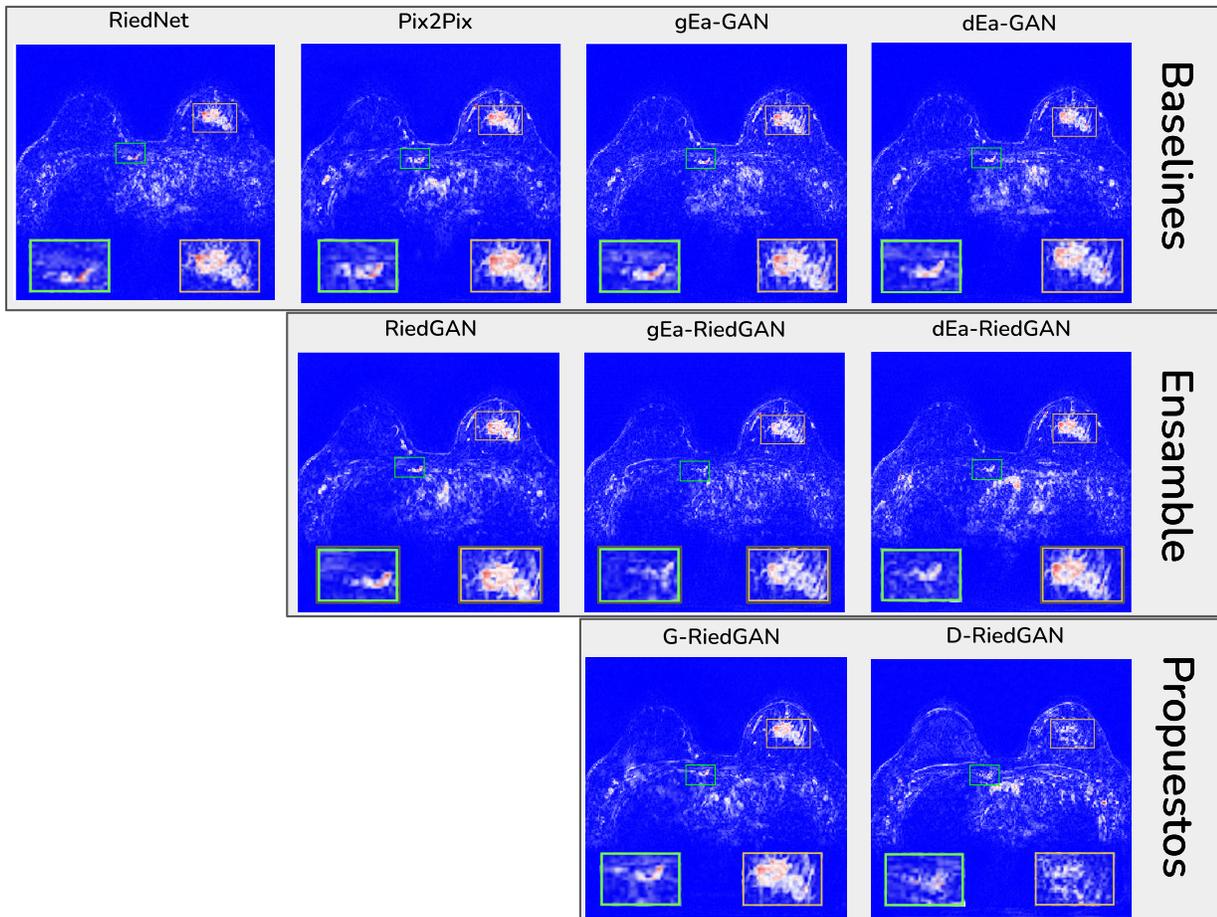


Figura 4-3: Mapas de diferencia entre las imágenes reales y generadas. Las regiones aumentadas corresponden con las mismas en la figura anterior. Mejor visto al hacer zoom

Existen varios puntos a resaltar dentro de las figuras. En primera instancia, es posible notar que en las imágenes sintetizadas todos los modelos logran generar de manera eficiente las estructuras anatómicas de mayor tamaño, tales como el tórax, el corazón o la forma de los órganos. Sin embargo, los mapas de diferencia muestran que los modelos base generan múltiples borrosidades en las imágenes, que se traducen en imprecisiones en estructuras microscópicas o diferencias significativas en intensidad de captación que podrían ser importantes para el diagnóstico de la enfermedad. En primera instancia, las borrosidades pueden deberse a que la naturaleza de las imágenes de DCE-MRI presentan, en sí mismas, variaciones microscópicas entre ellas debidas al agente de contraste, que podrían generar dificultad en el proceso sintético sin importar el tipo de modelo.

Además, con el fin de confirmar que el incremento en desempeño obtenido por el modelo D-RiedGAN es producido por la función de costo propuesta, se incluyen algunos experimentos investigando la variación al modificar el valor de λ en las funciones de costo. De esta forma se modifican los valores de $\lambda_{L1} = 20$ y $\lambda_{Lovasz} = 30$, $\lambda_{L1} = 40$ y $\lambda_{Lovasz} = 60$; y por último $\lambda_{L1} = 100$ y $\lambda_{Lovasz} = 150$ y se reentrena D-RiedGAN bajo los parámetros mencionados anteriormente. Los resultados de estos experimentos en función de MAE y PSNR se muestran en la Figura 4-4, en la cual es posible evidenciar un incremento evidente en las métricas al incrementar los valores de ponderación de la función de costo empleada por la D-RiedGAN. Este resultado es relevante debido a que muestra la eficiencia de la función de costo propuesta para mejorar los resultados del modelo de síntesis.

Sin embargo, este trabajo presenta algunas limitaciones. Primero, a pesar de que este estudio se centra en el análisis de la calidad cualitativa y cuantitativa de las imágenes es importante considerar la validez clínica de los estudios en escenarios de diagnóstico reales. Sin embargo, dada la naturaleza del trabajo y la ausencia de talento humano y recursos económicos para el desarrollo de este proyecto, esta evaluación sale del alcance del trabajo por lo que se pondrá como trabajo posterior. Por otra parte, el modelo ha sido entrenado usando una base de datos privada proveniente de una entidad de salud, debido a las bajas disponibilidades de datos clínicos que contengan secuencias pre y postcontraste de DCE-MRI. Dado que es conocida la limitación de los modelos de aprendizaje profundo para aprender en subconjuntos reducidos, incrementar la cantidad de muestras para el entrenamiento y validación de los modelos podría a su vez incrementar su desempeño. Por último, teniendo en cuenta el creciente cierre de brechas sociales en el mundo, es importante la validación del modelo considerando múltiples condiciones de adquisición (diferentes resonadores, medios de contraste, dosis, variaciones técnicas, etc) a su vez que de condiciones o variables biológicas (aspectos raciales, densidad de tejido, edad, etc) con el fin de propender por el desarrollo de modelos robustos ante estas variaciones.

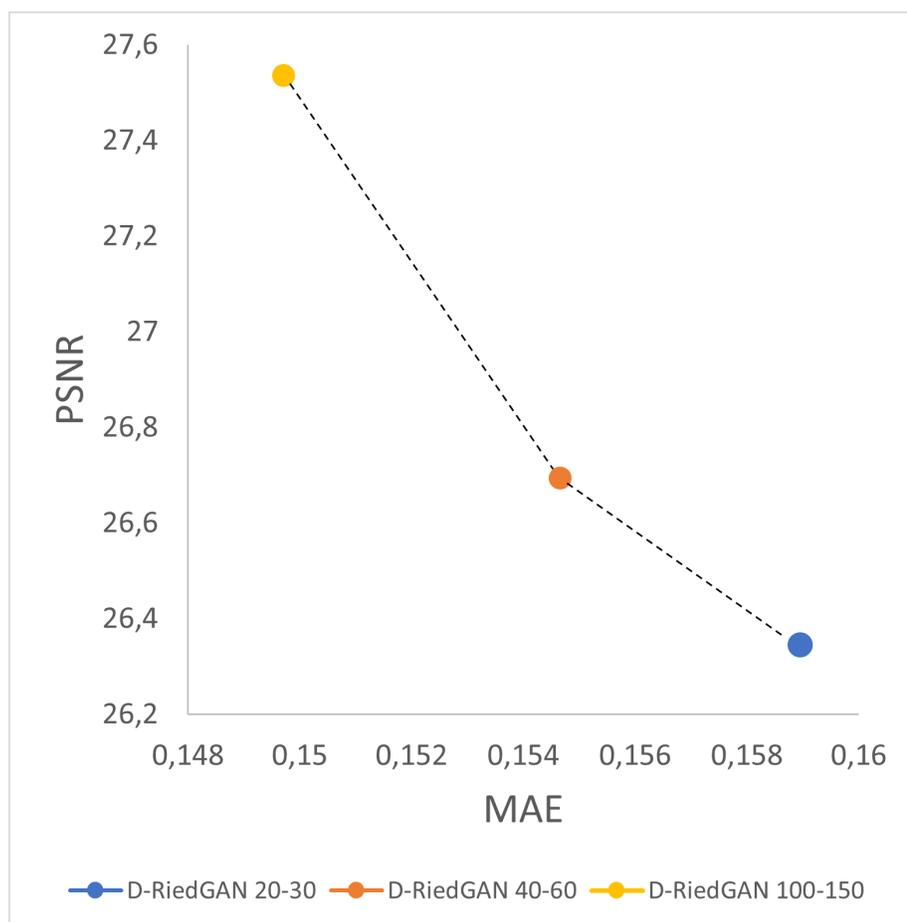


Figura 4-4: Prueba para establecer la mejor distribución de los niveles de importancia entre el MAE y el PSNR

5 Conclusiones y Trabajo Futuro

En este trabajo se presentó un modelo de aprendizaje profundo que permite la generación de imágenes sintéticas de postcontraste temprano de DCE-MRI de mama a partir de las imágenes precontraste. El desarrollo de este trabajo partió de la caracterización e implementación de tres modelos de línea base (Pix2Pix, RiedNet y Ea-GAN), los cuales fueron usados entrenados para generar imágenes sintéticas postcontraste de DCE-MRI, dando cumplimiento al Objetivo Específico 1. Los resultados iniciales de esta línea base indicaron que sin bien los modelos podían replicar las estructuras anatómicas, estos tenían limitaciones en replicar tanto las texturas como las zonas en las que había un realce de contraste. A partir de esa línea base, y para dar cumplimiento al Objetivo Específico 2, se consideraron los elementos clave de los modelos implementados y se creó una nueva línea de modelos ensamblados, esto a fin de evaluar si al combinar las fortalezas de los modelos de base se podían mejorar los resultados cuantitativos de la síntesis de imágenes.

Posteriormente, y continuando con el cumplimiento del Objetivo Específico 2, se propusieron dos nuevos modelos, G-RiedGAN y D-RiedGAN. Estos modelos integran en sus funciones de costo la información asociada a las zonas de captación de contraste que genera el agente de contraste tras su aplicación en el paciente. Esas zonas son uno de los elementos clave para el análisis de las imágenes postcontraste de DCE-MRI para el diagnóstico del cáncer de mama y es la razón por la cual se consideró su inclusión en los modelos propuestos.

Finalmente, y con el fin de dar cumplimiento al Objetivo Específico 3, se realizó una evaluación cuantitativa de los modelos implementados usando algunas de las métricas de evaluación comunes en el estado del arte: PSNR, SSIM y MAE. Los resultados cuantitativos, a nivel de las diferencias entre los píxeles en las imágenes generadas y las imágenes reales, mostraron que los dos modelos propuestos (G-RiedGAN y D-RiedGAN) mejoran los resultados obtenidos tanto por los modelos de línea base como por los modelos ensamblados. Adicionalmente, se evidenció (visualmente) que estos modelos disminuyen las borrosidades en las texturas de las estructuras anatómicas y sobretodo en las zonas de las mamas, que son el objeto de interés de este trabajo. La misma percepción se tuvo respecto a las zonas de realce de contraste, las cuales fueron mejor identificadas por el modelo para su réplica en las imágenes sintéticas.

Entre los resultados más importantes destacan algunos aspectos. Los experimentos propuestos para todos los modelos de ensamble obtuvieron resultados significativamente más

bajos que los resultados de los modelos base. Esto sugiere que el problema de síntesis para imágenes postcontraste de DCE-MRI no es trivial ni puede resolverse eficientemente mediante la combinación o modificación de los componentes más importantes de los modelos usados. Además, estos sugiere que este problema debe tratarse utilizando información de interés específica contenida en las imágenes precontraste y postcontraste. Por otra parte, los resultados de los experimentos con los modelos propuestos, que involucran algunos componentes relevantes de los modelos de línea base, obtienen un desempeño superior en comparación con los modelos ensamble de línea base en función de las métricas de evaluación. También, se realizaron algunos experimentos para determinar el impacto de la función de costo propuesta en el desempeño de la síntesis de las imágenes. Los resultados de estos experimentos indicaron que la mayor contribución en el desempeño de los nuevos modelos proviene por la función de costo, ya que al incrementar la ponderación de esta (λ_G y λ_D) el desempeño incrementa de manera significativa. Estos aspectos demuestran la eficiencia de los modelos propuestos para el problema de síntesis de imágenes postcontraste, y permiten validar que incluir información característica del medio de contraste mejora el desempeño de los modelos de síntesis. Todos los elementos presentados dan razón del cumplimiento del Objetivo General de este trabajo.

Ahora, es importante mencionar que estas estrategias presentan algunas limitaciones. En general, es posible que los modelos de síntesis para la reconstrucción de imágenes puedan generar imágenes borrosas, esto sucede porque las funciones de costo suelen optimizarse usando el promedio de las intensidades de los píxeles en toda la imagen. Es así que los modelos propuestos en este trabajo aún presentan algunos efectos de borrosidades, aunque estos son significativamente menores que los generados por los modelos de la línea base y los modelos ensamblados. Por otra parte, todos los modelos considerados en este trabajo fueron entrenados con una base de datos de imágenes privada de estudios de DCE-MRI que tienen una cantidad limitada de estudios, lo que puede limitar el poder de generación de las imágenes postcontraste. Esto se debe a que la base de datos utilizada hace parte de un proyecto de investigación en el que se enmarcó el desarrollo de este trabajo de maestría y que busca dar respuesta al problema de la síntesis de imágenes usando los equipos con los que se cuentan en la ciudad. En este sentido, consideramos que si los modelos son entrenados con una base datos de mayor tamaño los resultados puede mejorar considerablemente, sobretodo por la inclusión de la detección de las zonas de realce de contraste durante el entrenamiento del modelo.

Por último, este trabajo plantea algunas cuestiones como trabajo futuro. Por ejemplo, la evaluación realizada de los modelos sólo incluye métricas cuantitativas que usan la información de las intensidades de los píxeles; sin embargo, se debe realizar una validación cualitativa con radiólogos expertos en este tipo de imágenes para determinar la validez diagnóstica de las imágenes sintéticas generadas. Esta cuestión quedó por fuera del alcance de este trabajo debido a la falta de talento humano que pudiera realizar dicha evaluación y, también, por

la falta de recursos económicos que nos permitiera contratar a un radiólogo experto en el tema. Otra consideración es que a futuro se debe realizar un análisis de como los modelos de línea base y los modelos propuestos responden ante el desafío de entrenarse con bases de datos de imágenes adquiridas bajo condiciones heterogéneas con imágenes de distintas calidades (0.5T, 1.5T, 3T y 7T), adquiridas usando dosis y medios de contraste diferentes, y con marcas de resonadores diferentes.

6 Anexos

El entrenamiento en una red neuronal artificial, es la etapa en la que se dedica a aprender la distribución de las características de las imágenes, lo que para el caso de la presente investigación, involucra el aprendizaje de la forma en la que se da el contraste en las imágenes. En esta sección se presenta la Figura 6-1, en la que se evidencia el proceso de aprendizaje de los diferentes modelos propuestos a lo largo de la investigación.

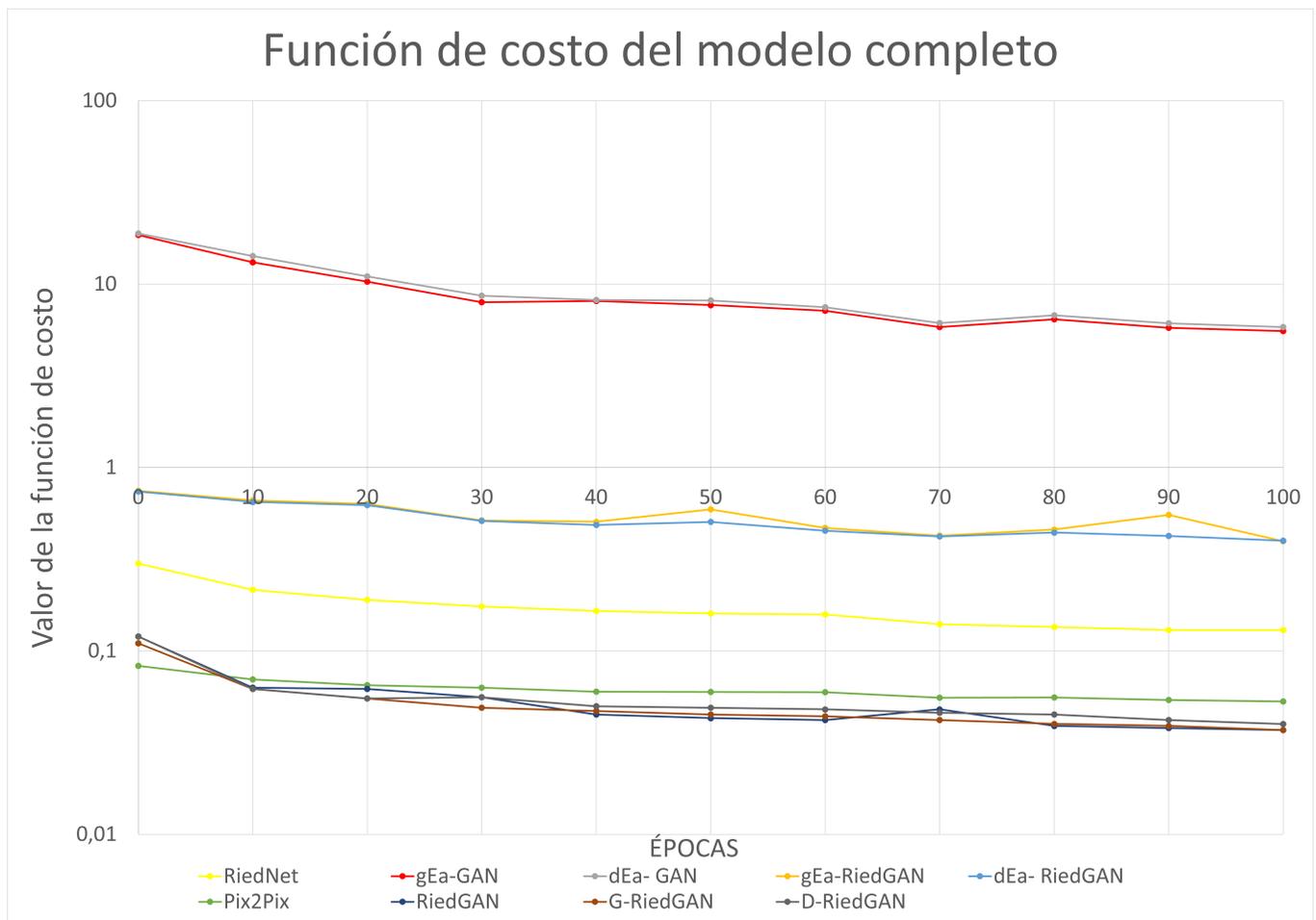


Figura 6-1: Función de costo de cada uno de los modelos implementados a lo largo de la investigación.

Se logran evidenciar todos los modelos empleados para la realización de la tesis de investigación, están los modelos de línea base (RiedNet, Pix2Pix, gEa-GAN y dEa-GAN), los modelos ensamblados (RiedGAN, gEa-RiedGAN y dEa-RiedGAN) y los modelos propuestos (G-RiedGAN y D-RiedGAN). Dentro de esta gráfica se muestran las curvas sobre cómo es el proceso de aprendizaje de las diferentes redes empleadas, indicando que, mientras más decae la curva, menor diferencia se dió entre lo sintetizado y lo generado, dejando en evidencia que los modelos propuestos, además del modelo ensamblado RiedGAN, junto con Pix2Pix son los que evidencian mejor desempeño al momento de evaluar las diferencias entre la imagen real y la imagen sintetizada por el modelo.

Bibliography

- [1] Jimenez Herrera, M.P.: Cáncer de Mama y Cuello Uterino. Informe De Evento **03** (2018) 2–15
- [2] The Global Cancer Observatory: GLOBOCAN 2020: International Agency Research on Cancer. **509** (2020) 1–2
- [3] OPS, O.: Cáncer de mama. <https://www.paho.org/es/temas/cancer-mama> (2018)
- [4] Martín, M., Herrero, A., Echavarría, I.: El cáncer de mama. *Arbor* **191** (2015)
- [5] de División Prevención Control del Cáncer, C.: ¿qué es el cáncer de mama? (2022)
- [6] InstitutoNacionalDelCancer: Tratamiento del cáncer de seno (mama) en adultas (PDQ®)–Versión para pacientes - Instituto Nacional del Cáncer. Sitio Web del Departamento de Salud y Servicios Humanos de Estados Unidos (2015) 1–37
- [7] Macias, S.G.: Ecografía mamaria. (2019) 1–20
- [8] Freer, P.E.: Mammographic breast density: Impact on breast cancer risk and implications for screening. *Radiographics* **35** (2015) 302–315
- [9] Naranjo, I.D., de Vega, V.M., González, S.L., Mollinedo, D., Álvarez Perez, L., Alonso, M.R.: Automated breast ultrasound for the detection of breast lesions: comparison with hand-held ultrasound. *Revista de Senología y Patología Mamaria* **31** (2018) 108–113
- [10] Sartori, P., Rizzo, F., Taborda, N., Anaya, V., Caraballo, A., Saleme, C., Carrizo, R., Cayo, M., Peña, A.: Medios de contraste. *Revista Argentina de Radiología* **77** (2013) 49–62
- [11] Pérez-Zúñiga, I., Villaseñor-Navarro, Y., Pérez-Badillo, M.P., Cruz-Morales, R., Pavón-Hernández, C., Aguilar-Cortázar, L.: Resonancia magnética de mama y sus aplicaciones. *Gaceta Mexicana de Oncología* **11** (2012) 268–280
- [12] Rodríguez Marcano, Y.M., González, I., Palencia, H., Sandoval, M., León, L.: Mamografía espectral con realce de contraste. Nuestra experiencia. *Revista Venezolana de Oncología* **26** (2014) 743–751
- [13] Valenzuela, R., Arevalo, O., Tavera, A., Riascos, R., Bonfante, E., Patel, R.: Imágenes

- del depósito de gadolinio en el sistema nervioso central. *Revista Chilena de Radiología* **23** (2017) 59–65
- [14] Sandfort, V., Yan, K., Pickhardt, P.J., Summers, R.M.: Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks. *Scientific Reports* **9** (2019)
- [15] Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. (2016)
- [16] Gao, F., Wu, T., Li, J., Zheng, B., Ruan, L., Shang, D., Patel, B.: SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis. *Computerized Medical Imaging and Graphics* **70** (2018) 53–62
- [17] Gao, F., Wu, T., Chu, X., Yoon, H., Xu, Y., Patel, B.: Deep Residual Inception Encoder-Decoder Network for Medical Imaging Synthesis. *IEEE Journal of Biomedical and Health Informatics* **24** (2020) 39–49
- [18] Mori, M., Fujioka, T., Katsuta, L., Kikuchi, Y., Oda, G., Nakagawa, T., Kitazume, Y., Kubota, K., Tateishi, U.: Feasibility of new fat suppression for breast MRI using pix2pix. *Japanese Journal of Radiology* **38** (2020) 1075–1081
- [19] Stratton, M.R., Campbell, P.J., Futreal, P.A.: The cancer genome. *Nature* **458** (2009) 719–724
- [20] Carrasco, L.F., Costa, E.Y.: MAMA GUIADOS POR RESONANCIA MAGNÉTICA : NUESTRA EXPERIENCIA MRI Breast Guided Procedures : Our Experience. (2019) 5210–5215
- [21] Pons Porrata, L.M., de la Cruz de Oña, A., Álvarez Cobas, I., Daudinot Gutiérrez, Y., Salomón López, J.: Caracterización del uso de gadolinio en pacientes con indicación de resonancia magnética. *Medisan* **16** (2012) 498–503
- [22] Gluskin, J., Saccarelli, C.R., Avendano, D., Marino, M.A., Bitencourt, A.G., Pilewskie, M., Sevilimedu, V., Sung, J.S., Pinker, K., Jochelson, M.S.: Contrast-enhanced mammography for screening women after breast conserving surgery. *Cancers* **12** (2020) 1–14
- [23] S.S., Le, T..N.T.B.: Image synthesis: a review of methods, datasets, evaluation metrics, and future outlook. *Artif Intell Rev* (2023)
- [24] Chen, J., Shen, Y., Gao, J., Liu, J., Liu, X.: Language-based image editing with recurrent attentive models. (2017)
- [25] García, J.I.C., Tanco, M.M., Olmedo, M.G.: Aprendizaje profundo para el procesamiento de imágenes-optimización del conjunto de datos de entrenamiento. (2021)

-
- [26] Zhu, X., Goldberg, A.B., Eldawy, M., Dyer, C.R., Strock, B.: A text-to-picture synthesis system for augmenting communication. In: AAAI. Volume 7. (2007) 1590–1595
- [27] Mendoza, J.: Herramienta para segmentación de líneas de texto en imágenes basada en deep learning. Grado en Ingeniería de las Tecnologías de Telecomunicación (2020)
- [28] Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., Lee, H.: Generative adversarial text to image synthesis. (2016)
- [29] Yu, B., Zhou, L., Wang, L., Shi, Y., Fripp, J., Bourgeat, P.: Ea-gans: Edge-aware generative adversarial networks for cross-modality mr image synthesis. *IEEE Transactions on Medical Imaging* **38** (2019) 1750–1762
- [30] Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. *Journal of Big Data* **6** (2019)
- [31] Beers, A., Brown, J., Chang, K., Peter Campbell, J., Ostmo, S., Chiang, M.F., Kalpathy-Cramer, J.: High-resolution medical image synthesis using progressively grown generative adversarial networks. arXiv (2018)
- [32] Pang, Y., Lin, J., Qin, T., Chen, Z.: Image-to-image translation: Methods and applications. *IEEE Transactions on Multimedia* **24** (2022) 3859–3881
- [33] Shen, T., Gou, C., Wang, J., Wang, F.Y.: Collaborative adversarial networks for joint synthesis and segmentation of x-ray breast mass images. *Proceedings - 2020 Chinese Automation Congress, CAC 2020* (2020) 1743–1747
- [34] S, D., S, M., S, K., S, S., R, M.: Gan based data augmentation for enhanced tumor classification, *IEEE* (2020) 1–5
- [35] García, J.M.J., Sánchez, J.P.A., Rodríguez, M.V., del Río, S.J., Moctezuma, R., Cayetano, C.S.: Reconocimiento facial basado en aprendizaje profundo para aplicaciones de seguridad automotriz deep learning-based face recognition for automotive security applications. (2023)
- [36] Rull, M.V.: Reconocimiento de objetos usando deep learning (2016)
- [37] Suárez, A., Tutoras, L., Begoña, ., Piñero, A., Gotarredona, C.S.: (Aplicación de técnicas de aprendizaje profundo (deep learning) a clasificación de imágenes histológicas)
- [38] Huang, D., Feng, M.: Understanding Deep Convolutional Networks for Biomedical Imaging: A Practical Tutorial. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* (2019) 857–863
- [39] Currie, G., Hawk, K.E., Rohren, E., Vial, A., Klein, R.: Machine learning and deep

- learning in medical imaging: Intelligent imaging. *Journal of Medical Imaging and Radiation Sciences* **50** (2019) 477–487
- [40] CHAGAS, E.T.D.O.: Aprendizaje profundo y sus aplicaciones hoy. **04** (2019) 5–26
- [41] KeepCoding, R.: ¿qué es la convolución? <https://keepcoding.io/blog/que-es-la-convolucion/> (2023)
- [42] Mandal, M.: Introduction to convolutional neural networks (cnn). <https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/> (2021)
- [43] Hinton, G.E., Krizhevsky, A., Wang, S.D.: (Transforming auto-encoders)
- [44] Anwar, A.: Difference between autoencoder (ae) and variational autoencoder (vae). <https://towardsdatascience.com/difference-between-autoencoder-ae-and-variational-autoencoder-vae-ed7be1c038f2> (2021)
- [45] Ronneberger, O.: INet: Convolutional Networks for Biomedical Image Segmentation. *IEEE Access* **9** (2015) 16591–16603
- [46] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. (2014)
- [47] Alyafi, B., Diaz, O., Elangovan, P., Vilanova, J.C., del Riego, J., Marti, R.: Quality analysis of dcgan-generated mammography lesions. Volume 11513. (2020) 19
- [48] Shen, T., Gou, C., Wang, F.Y., He, Z., Chen, W.: Learning from adversarial medical images for X-ray breast mass segmentation. *Computer Methods and Programs in Biomedicine* **180** (2019) 1–13
- [49] Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. (2017)
- [50] Guan, Q., Chen, Y., Wei, Z., Heidari, A.A., Hu, H., Yang, X.H., Zheng, J., Zhou, Q., Chen, H., Chen, F.: Medical image augmentation for lesion detection using a texture-constrained multichannel progressive gan. *Computers in Biology and Medicine* **145** (2022) 105444
- [51] Dhivya, S., Mohanavalli, S., Karthika, S., Shivani, S., Mageswari, R.: Gan based data augmentation for enhanced tumor classification, Institute of Electrical and Electronics Engineers Inc. (2020)
- [52] Lin, C., Tang, R., Lin, D.D., Liu, L., Lu, J., Chen, Y., Gao, D., Zhou, J.: Breast mass detection in mammograms via blending adversarial learning. (2019) 52–61
- [53] Balaji, K.: Image augmentation based on variational autoencoder for breast tumor segmentation. *Academic Radiology* (2023)

- [54] Dar, S.U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., Cukur, T.: Image synthesis in multi-contrast mri with conditional generative adversarial networks. *IEEE transactions on medical imaging* **38** (2019) 2375–2388
- [55] Campaz-Usuga, P., Fonnegra, R.D., Mera, C.: Quality enhancement of breast dce-mri images via convolutional autoencoders. *2021 IEEE 2nd International Congress of Biomedical Engineering and Bioengineering (CI-IBBI)* (2021) 1–4
- [56] Wang, S., Sun, K., Wang, L., Qu, L., Yan, F., Wang, Q., Shen, D.: Breast tumor segmentation in dce-mri with tumor sensitive synthesis. *IEEE Transactions on Neural Networks and Learning Systems* (2021) 1–12
- [57] Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S., Yang, B.: Medgan: Medical image translation using gans. *Computerized Medical Imaging and Graphics* **79** (2020) 101684
- [58] Fan, M., Huang, G., Lou, J., Gao, X., Zeng, T., Li, L.: Cross-parametric generative adversarial network-based magnetic resonance image feature synthesis for breast lesion classification. *IEEE Journal of Biomedical and Health Informatics* (2023) 1–11
- [59] Jiang, G., Lu, Y., Wei, J., Xu, Y.: Synthesize Mammogram from Digital Breast Tomosynthesis with Gradient Guided cGANs. Volume 11769 LNCS. Springer International Publishing (2019)
- [60] Jiang, G., Wei, J., Xu, Y., He, Z., Zeng, H., Wu, J., Qin, G., Chen, W., Lu, Y.: Synthesis of Mammogram from Digital Breast Tomosynthesis using Deep Convolutional Neural Network with Gradient Guided cGANs. *IEEE Transactions on Medical Imaging* **0062** (2021) 1–12
- [61] Zhang, Y., Li, C., Dai, Z., Zhong, L., Wang, X., Yang, W.: Breath-hold cbct-guided cbct-to-ct synthesis via multimodal unsupervised representation disentanglement learning. *IEEE Transactions on Medical Imaging* **42** (2023) 2313–2324
- [62] Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, A., Kautz, J., Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional gans. (2017)
- [63] Fujioka, T., Satoh, Y., Imokawa, T., Mori, M., Yamaga, E., Takahashi, K., Kubota, K., Onishi, H., Tateishi, U.: Proposal to improve the image quality of short-acquisition time-dedicated breast positron emission tomography using the pix2pix generative adversarial network. *Diagnostics* **12** (2022)
- [64] Jiang, Y., Zheng, Y., Jia, W., Song, S., Ding, Y.: Synthesis of contrast-enhanced spectral mammograms from low-energy mammograms using cgan-based synthesis network. (2021) 68–77
- [65] Kim, E., Cho, H.H., Kwon, J., Oh, Y.T., Ko, E.S., Park, H.: Tumor-attentive

- segmentation-guided gan for synthesizing breast contrast-enhanced mri without contrast agents. *IEEE Journal of Translational Engineering in Health and Medicine* **11** (2023) 32–43
- [66] Fonnegra, R.D., María, L.N., Hernández, P.C.C.: (Proyecto de investigación: Generación de respuesta sintética a la aplicación de medios de contraste en imágenes médicas para mejorar el diagnóstico automático de cáncer de mama)
- [67] Xu, B., Wang, N., Chen, T., Li, M.: Empirical evaluation of rectified activations in convolutional network. (2015)
- [68] Berman, M., Rannen, A., Matthew, T., Blaschko, B.: The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks. (2018)
- [69] Lovász, L.: Submodular functions and convexity. *Mathematical Programming The State of the Art* (1983) 235–257
- [70] Schneider, P., Xhafa, F.: Anomaly detection. *Anomaly Detection and Complex Event Processing over IoT Data Streams* (2022) 49–66
- [71] Hore, A., Ziou, D.: Image quality metrics: Psnr vs. ssim, *IEEE* (2010) 2366–2369