

Mejora de resolución en imágenes de ultrasonido de seno usando redes GAN

Juan Carlos Solano Torres

Edward Andrés Vega Cortés

Trabajo de Grado para Optar el Título de Ingeniero Electrónico.

Directora

Ana Beatriz Ramirez Silva

Doctora en Ingeniería Electrónica

Co-Director

Sergio Alberto Abreo Carrillo

Doctor en Ingeniería Electrónica

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones

Bucaramanga

2023

Dedicatoria

A Dios, por darme la perseverancia y persistencia que siempre necesité a lo largo de toda mi vida académica.

A mi tío Jovanny, pues sus palabras sabias, el apoyo constante y su fe en mí han sido fundamentales en mi crecimiento y éxito. Siempre has estado ahí, brindándome orientación y ánimo cuando más lo necesitaba.

A mi madre Marisol, pues su amor incondicional y apoyo constante han sido la fuerza motriz detrás de mis logros. Esta tesis, como un pequeño reflejo de mi camino, está dedicada a ella con el más profundo agradecimiento y amor.

A mi padre Juan Carlos, por demostrarme que la fuerza de voluntad es la chispa que enciende nuestros propios deseos de superación.

Al regalo más hermoso que mis padres me han dado, mi hermano Andrés, este logro a ti como un recordatorio de lo agradecido que estoy por tu presencia en mi vida.

A mi abuela Leonor, porque aunque ya no te pueda tener presente físicamente, tu presencia en mi corazón me acompañará en cada logro y momento especial de mi vida.

A mi novia Juliana, que por su forma de ser ha iluminado mis días y ha hecho que los desafíos sean más fáciles de superar.

A mis familiares y amigos, que han estado presentes apoyándome tanto en lo académico, como en lo personal.

Juan Carlos Solano Torres

Dedicatoria

Finalmente, quiero expresar mi agradecimiento a mi familia y amigos por su apoyo incondicional durante esta etapa de mi vida académica. Su paciencia y aliento constante han sido un motor fundamental para perseverar en este proyecto.

Mi familia, mi querida nona Nelly, mi madre y mi padre que han sido una base y apoyo incondicional, durante todo este proceso y parte de mi vida. Sin ellos o sin su ayuda no estaría en donde estoy y por este motivo dedico cada uno de mis triunfos.

No podría desperdiciar esta oportunidad para agradecer infinitamente a mi hermosa novia, Stefany. Junto con mi familia representa un pilar en mi vida, ella conoce muy bien mi proceso durante esta etapa de mi vida.

Edward Andrés Vega Cortés

Agradecimientos

En el proceso de elaboración de esta tesis de pregrado, nos gustaría expresar nuestro sincero agradecimiento al Grupo de Investigación CPS y en particular a la profesora Ana Beatriz Ramírez Silva y al profesor Sergio Alberto Abreo Carrillo, por su invaluable apoyo y colaboración a lo largo de esta travesía académica.

¡Gracias!

Tabla de Contenido

| | |
|--|-----------|
| Introducción | 15 |
| 1 Objetivos | 18 |
| 1.1 Objetivo General | 18 |
| 1.2 Objetivos Específicos | 18 |
| 2 Marco Conceptual | 20 |
| 2.1 Escenario Biomédico y principios físicos | 20 |
| 2.1.1 Método de Resonancia Magnética Nuclear (RMN) | 21 |
| 2.1.2 Método de ultrasonido o ecografía | 21 |
| 2.1.2.1 Generación de imágenes | 22 |
| 2.2 Modelo neuronal artificial | 23 |
| 2.2.1 Red neuronal artificial: conceptos y generalidades | 23 |
| 2.2.2 Red neuronal Residual: ResNet | 24 |
| 2.2.3 Red tipo GAN | 25 |
| 3 Modelo propuesto: SRGAN o GAN de Súper resolución. | 28 |
| 3.1 Arquitectura y topología del modelo Generador o SRResNet | 29 |
| 3.2 Arquitectura y topología del modelo Discriminador | 31 |

| | | |
|----------|---|-----------|
| 3.3 | Entrenamiento del modelo SRGAN | 32 |
| 3.4 | Funciones de pérdida del modelo SRGAN | 33 |
| 3.4.1 | Función de pérdida de contenido | 34 |
| 3.4.2 | Función de pérdida adversaria | 36 |
| 3.4.3 | Función de pérdida del discriminador | 36 |
| 3.5 | Especificaciones y ajuste de hiperparámetros | 37 |
| 3.5.1 | Hiperparámetros para SRResNet | 37 |
| 3.5.2 | Hiperparámetros para SRGAN | 38 |
| 4 | Generación de la base de Datos | 40 |
| 4.1 | Antecedentes y trabajo previo | 40 |
| 4.2 | Fase 1: Vóxeles RMN a información de tejido mamario | 42 |
| 4.3 | Fase 2: Extracción de imágenes | 42 |
| 4.4 | Fase 3: Data augmentation | 43 |
| 4.5 | Fase 4: Procesamiento para la uniformidad de imágenes | 45 |
| 4.5.1 | Padding | 45 |
| 4.5.2 | Centrado | 46 |
| 4.6 | Fase 5: Submuestreo de las imágenes de alta resolución | 46 |
| 4.7 | Fase 6: Transformación de datos RMN a información de ultrasonido mediante FWI | 48 |
| 4.8 | Fase 7: Cuantificación de las imágenes en formato de 8 bits | 48 |
| 4.9 | Fase 8: Selección final de imágenes para la base de datos | 50 |

| | | |
|----------|---|-----------|
| 4.9.1 | Base de datos balanceada (Tipo A) | 50 |
| 4.9.2 | Base de datos desbalanceada en imágenes (Tipo B) | 51 |
| 4.9.3 | Base de datos desbalanceada en tipos de imágenes y atributos (Tipo C) | 52 |
| 5 | Entorno de desarrollo | 54 |
| 5.1 | Hardware de alto rendimiento | 54 |
| 5.1.1 | Hardware usado | 54 |
| 5.2 | Software de aprendizaje profundo | 55 |
| 5.2.1 | Librerías y APIs usados | 55 |
| 5.2.1.1 | PyTorch (torch) | 55 |
| 5.2.1.2 | TorchVision (torchvision) | 56 |
| 5.2.1.3 | TensorFlow (tf) | 56 |
| 5.2.1.4 | Numpy (np) | 56 |
| 5.2.1.5 | TensorBoard (torch.utils.tensorboard) | 56 |
| 5.2.1.6 | TensorFlow-Keras | 56 |
| 5.2.1.7 | PIL | 56 |
| 5.2.1.8 | OS | 57 |
| 5.2.1.9 | Funciones y clases específicas importadas desde TensorFlow-Keras | 57 |
| 5.2.2 | VGG: Modelo neuronal preentrenado | 58 |
| 5.2.3 | Entorno virtual | 58 |
| 6 | Presentación de resultados | 59 |

| | | |
|----------|---|-----------|
| 6.1 | Resultados referentes al modelo SRGAN | 59 |
| 6.1.1 | Análisis cuantitativo | 60 |
| 6.1.2 | Análisis cualitativo | 62 |
| 6.2 | SRResNet vs SRGAN | 65 |
| 6.2.1 | Resultados referentes al modelo SRResNet con base de datos Tipo B | 65 |
| 6.2.2 | Resultados referentes al modelo SRGAN con base de datos Tipo B | 67 |
| 6.2.3 | Análisis cuantitativo | 69 |
| 6.2.4 | Análisis cualitativo | 70 |
| 6.2.5 | Discusión de resultados | 72 |
| 6.3 | Pruebas multiparamétricas | 73 |
| 6.3.1 | Análisis cuantitativo | 75 |
| 6.3.2 | Análisis cualitativo | 75 |
| 7 | Conclusiones | 78 |
| 8 | Recomendaciones para trabajos futuros | 79 |
| | Referencias Bibliográficas | 81 |

Lista de Tablas

| | | |
|---------|--|----|
| Tabla 1 | Algunos hiperparámetros de arquitectura de los modelos adversarios | 38 |
| Tabla 2 | Rangos y tamaño del set de imágenes RMN | 44 |
| Tabla 3 | Conjunto final de imágenes reconstruidas (FWI) | 49 |
| Tabla 4 | Valores y parámetros asociados a los tipos de tejidos | 49 |
| Tabla 5 | Cantidades de cada grupo de imágenes (Tipos y atributos) | 51 |
| Tabla 6 | Valores de métricas del modelo SRGAN respecto a cada tipo de base de datos | 62 |
| Tabla 7 | Comparación de métricas entre SRResNet y SRGAN | 70 |
| Tabla 8 | Resultados de pruebas multiparamétricas, al variar LR y alpha | 74 |

Lista de Figuras

| | | |
|-----------|--|----|
| Figura 1 | Esquema de un circuito para la generación de imágenes mediante ultrasonido | 22 |
| Figura 2 | Esquema de conexión secuencial de capas con <i>skip connection</i> , base del modelo ResNet. | 25 |
| Figura 3 | Ejemplo de arquitectura general del modelo GAN | 27 |
| Figura 4 | Esquema de la arquitectura y topología de la red generadora del modelo SRGAN | 29 |
| Figura 5 | Esquema de la arquitectura de la red discriminadora del modelo SRGAN | 31 |
| Figura 6 | Tipos de Senos en la base de datos | 41 |
| Figura 7 | Distribución espacial de datos RMN y su contorno | 43 |
| Figura 8 | Imágenes consecuentes con el tipo de corte | 43 |
| Figura 9 | Ejemplo de imágenes de diferente tamaño | 45 |
| Figura 10 | Ejemplo de centralización de imágenes. | 46 |
| Figura 11 | Par de imágenes RMN | 47 |
| Figura 12 | Bases de datos generadas | 53 |
| Figura 13 | Métricas (PSNR, SSIM y MSE) de validación sobre los tres tipos de datos | 61 |
| Figura 14 | Comparación de resultados cualitativos de la red SRGAN con imágenes en escala de grises. | 64 |

| | | |
|-----------|---|----|
| Figura 15 | Función de pérdida y métricas (PSNR, SSIM y MSE) del modelo SRResNet-Tipo B | 66 |
| Figura 16 | Resultados visuales del modelo SRResNet-Tipo B | 67 |
| Figura 17 | Funciones de pérdida y métricas (PSNR, SSIM y MSE) del modelo SRGAN-Tipo B | 68 |
| Figura 18 | Resultados visuales del modelo SRGAN-Tipo B | 69 |
| Figura 19 | Comparación cualitativa entre resultado de lo modelos Tipo B, imágenes a color. | 71 |
| Figura 20 | Comparación de métricas | 74 |
| Figura 21 | Imagen HR RMN (Ground Truth) | 76 |
| Figura 22 | Resultados visuales de las 9 pruebas | 77 |

Glosario

Arquitectura se refiere a la estructura, forma, contenido funcional y modo en que se conecta a nivel lógico una red neuronal.

Perceptrón unidad básica de procesamiento de información del modelo neuronal. Análogo a neurona.

Red neuronal estructura conformada por interconexiones variadas entre nodos (neuronas) con el fin de conformar un sistema complejo.

Sobreajuste fenómeno no deseado del aprendizaje automático en el cual el modelo solo efectúa predicciones sobre datos conocidos.

Topología forma y estructura a nivel físico que define el modo en que se conecta y distribuye algún tipo de red.

Resumen

Título: Mejora de resolución en imágenes de ultrasonido de seno usando redes GAN. *

Autores: Juan Carlos Solano Torres y Edward Andrés Vega Cortés. **

Palabras Clave: Ultrasonido, seno, red, neuronal, GAN, IA, resolución.

Descripción: El cáncer de seno es una de las principales causas de muerte en mujeres. La detección temprana del cáncer de seno representa un factor fundamental en la lucha por eliminar este mal y esto se realiza mediante un correcto análisis diagnóstico en función de las pruebas o métodos preventivos. La radiografía de mama o mamografía, el examen de ultrasonido y la resonancia magnética nuclear, son los métodos que se emplean para el diagnóstico temprano de cáncer de seno.

El enfoque en este trabajo apunta hacia el método de ultrasonido, particularmente sobre sus imágenes. La baja resolución en las imágenes de ultrasonido de seno, constituye la principal falencia al momento de efectuar un correcto diagnóstico por parte del especialista médico. La corrección o la mejora de resolución de dichas imágenes representa el reto a resolver en este trabajo de investigación, válido como trabajo de grado. Por este motivo, se plantea implementar un modelo de red neuronal artificial basado en arquitecturas de tipo GAN (Generative Adversarial Network por sus siglas en inglés) (Goodfellow *et al.*, 2014), que logre entregar imágenes de ultrasonido de seno con mejor resolución. El diseño parte de una arquitectura y topología de red base ya existente desarrollada por (Ledig *et al.*, 2017), la cual consiste en una red SRGAN o GAN de súper resolución. Este modelo presenta resultados de inferencia en imágenes a un factor x4 de ampliación, con una valoración en sus métricas de desempeño más cercanas al resultado de las imágenes fotorrealistas de entrada, comparadas con cualquier otro modelo generativo.

* Trabajo de Grado.

** Facultad de Ingeniería Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y Telecomunicaciones. Directora: Ana Beatriz Ramírez Silva; Co-Director: Sergio Alberto Abreo Carrillo

Abstract

Title: Resolution enhancement in breast ultrasound images using GAN networks. *

Authors: Juan Carlos Solano Torres and Edward Andrés Vega Cortés. **

Keywords: Ultrasound, breast, Network, neural, GAN, AI, resolution.

Description: Breast cancer is one of the leading causes of death in women. The early detection of cancer represents a highly empowering factor in the fight to eliminate this illness and this is done through a correct diagnostic analysis based on tests or preventive methods. Breast radiography or mammography, ultrasound examination and nuclear magnetic resonance are the preventive methods used for the diagnosis of breast cancer.

The focus in this work is towards the ultrasound method, particularly on its images. The low resolution in the breast ultrasound images constitutes the main shortcoming when making a correct diagnosis by the medical specialist. The correction or improvement of the resolution of these images represents the challenge to be solved in this valid research work as a degree project. For this reason, it is proposed to implement an artificial neural network model based on GAN (Generative Adversarial Network) architectures (Goodfellow *et al.*, 2014), which manages to deliver breast ultrasound images with better resolution. The design builds on an existing net-based architecture and topology used by (Ledig *et al.*, 2017), which consists of a super-resolution SRGAN or GAN network. This model presents inference results in images at a magnification factor x4, with an assessment in its performance metrics closer to the result of the input photorealistic images compared to any other generative model.

* Undergrade Work

** Faculty of Engineering Physicomechanics. School of Electrical, Electronic and Telecommunications Engineering.
Advisor: Ana Beatriz Ramirez Silva; Co-Advisor: Sergio Alberto Abreo Carrillo

Introducción

A nivel mundial el cáncer de seno es el más mortal y más común entre la población femenina (Servick, 2014). Para mujeres con alto riesgo de contraer cáncer de seno, la Sociedad Americana contra el Cáncer aconseja una resonancia magnética nuclear (RMN) y una mamografía cada año, por lo general a partir de los 30 años (Oeffinger *et al.*, 2015). La tasa de mortalidad debida al cáncer de seno podría disminuir si se efectúan métodos para la detección temprana. En general, existen varios factores que afectan el proceso de detección temprana de cáncer, como por ejemplo la alta densidad relativa del seno, esto dificulta la detección en pruebas como la mamografía o mastografía. DeSantis *et al.* (2015) afirman que cuando se encuentra tejido mamario denso, las detecciones hechas por el método de la mamografía necesitan ser confirmadas mediante pruebas complementarias de ecografía o ultrasonido.

Múltiples estudios han demostrado la importancia de realizar un análisis complementario al método de la mamografía en donde el ultrasonido de seno juega un papel importante. La imposibilidad de realizar correctos diagnósticos empleando únicamente el método de mamografía en donde por ejemplo predominan senos densos, es uno de los factores que condiciona dicha prueba para este grupo poblacional en particular (Scheel *et al.*, 2014). A diferencia, el ultrasonido detecta daños cancerígenos invasivos adicionales en etapa temprana cuando se combina con la mamografía (Vourtsis y Berg, 2019). No obstante, existen aspectos que limitan el potencial de esta técnica, especialmente relacionados con el personal capacitado para efectuar dichas pruebas.

En general, las imágenes por ultrasonido de seno tienen una baja resolución en comparación con las imágenes obtenidas por mamografía o por resonancia. En este aspecto en particular se va a centrar el propósito y la justificación de este trabajo, es decir, el interés radica sobre la problemática de la baja resolución en las imágenes de ultrasonido de seno. Al reconocer y entender las problemáticas generales que surgen con la técnica de ultrasonido, es necesario resaltar que el enfoque que se aborda en este trabajo se orienta en intentar realizar una mejora en la resolución de las imágenes.

La alternativa de implementación en este trabajo de investigación se basa especialmente en el uso de inteligencia artificial para realizar una mejora de resolución sobre nuestras imágenes de ultrasonido de seno. Ledig *et al.* (2017) han presentado un estudio enfocado en la generación de muestras de súper resolución soportado por un modelo de red neuronal de tipo GAN. Su modelo se denomina SRGAN¹ que especialmente genera imágenes fotorrealistas de súper resolución (SR). El enfoque esencial que surge en este trabajo se orienta en la aplicación y el uso del modelo SRGAN para uso particular sobre imágenes de ultrasonido de seno.

Otro aporte diferencial en este proyecto se orienta en el desarrollo de la base de datos para entrenamiento y validación de la red. Las imágenes están precedidas por la base de datos pública OA-Breast, los datos de OA-Breast están representados por información netamente asociada a angiografías por resonancia magnética (ARM) que fueron generadas por Lou *et al.* (2017), inves-

¹ Super Resolution Generative Adversarial Network. Por sus siglas en inglés

tigadores de la universidad de Washington en St. Louis. Las imágenes que inicialmente son de ARM² no representan ninguna relación con las características de ultrasonido, por esta razón fue necesario realizar una serie de procesos de transformación de datos de tejido caracterizado por resonancia magnética, a rangos de valores de velocidad del sonido particular de cada tipo de tejido mamario. Finalmente la base de datos generada constituye el recurso primordial de la red neuronal en cuestión, de la cual el 80 % representa las imágenes de entrenamiento y el 20 % para validación.

Es importante mencionar que la información estructural de los tejidos en las imágenes, no contienen características ni patrones asociados con tumores malignos. En síntesis, las muestras que componen la base de datos representan senos sanos.

² La angiografía por resonancia magnética (ARM) es un método RMN enfocado en la caracterización de vasos sanguíneos

1. Objetivos

1.1. Objetivo General

Diseñar y/o implementar un modelo de red GAN para la mejora de resolución en imágenes de ultrasonido de seno.

1.2. Objetivos Específicos

Identificar el modelo (arquitectura y topología) de red GAN que se ajusta y brinda una solución deseada. Existen múltiples modelos de redes neuronales que se basan en algoritmos de tipo GAN que proporcionan resultados de alta y súper resolución, es necesario realizar las correspondientes pruebas e implementaciones para encontrar un modelo base para eventualmente reproducir y diseñar un modelo propio.

Preparar y/o generar una base de datos apropiada para el entrenamiento y validación de la red.

Adecuar un entorno de desarrollo basado en el lenguaje orientado a objetos como Python, uso de librerías de inteligencia artificial como TensorFlow, PyTorch y enfocadas a las redes neuronales como Keras.

Evaluar los resultados de la red con las imágenes reservadas para validación. La validación corresponde al proceso de evaluar de manera cuantitativa y cualitativa los resultados de la red,

comparándolos con datos reales. Se tendrá que seleccionar un lote de imágenes de la base de datos distintas a las de entrenamiento. Posteriormente se decide si los resultados sintéticos generados son aceptados como descriptores de los datos reales.

2. Marco Conceptual

A continuación se hará una introducción referente al contexto base que conforma el marco teórico del trabajo. En primer lugar será necesario entender las definiciones y conceptos básicos relacionados al ámbito biomédico y su base científica que soportan el argumento y justificación del proyecto. Las definiciones científicas obedecen a fenómenos físicos como *resonancia magnética nuclear* (RMN) y perturbaciones de ultrasonido, todo esto enfocado en técnicas y métodos de generación de imágenes sobre tejidos mamarios. La obtención de muestras o imágenes mediante RMN y ultrasonido será explicado de mejor manera en el capítulo 4.

2.1. Escenario Biomédico y principios físicos

El ámbito biomédico en general se centra en el proceso salud-enfermedad. Desde el enfoque moderno, la biomedicina ha logrado grandes avances si se tiene en cuenta la participación directa de la ingeniería y todo su desarrollo tecnológico asociado (Gismondi Glave, 2010).

El contexto biomédico de este proyecto se enfoca en los métodos generativos de imágenes, particularmente los que se basan en el fenómeno de ultrasonido y resonancia magnética nuclear (RMN). Para una comprensión adecuada del contexto físico y biomédico, es esencial abordar y explicar los conceptos relacionados con estos dos métodos de generación de imágenes.

2.1.1. Método de Resonancia Magnética Nuclear (RMN)

El método RMN es una técnica de generación de imágenes que se basa en el fenómeno físico de resonancia magnética nuclear, el cual mediante generación de campos magnéticos busca alinear el momento magnético nuclear (MMN)³ de los núcleos de hidrógeno del agua presentes en el cuerpo (Brown *et al.*, 2014). Es una técnica no invasiva que busca información sobre la estructura y composición interna del órgano o parte del cuerpo que se desea analizar.

2.1.2. Método de ultrasonido o ecografía

A diferencia del método RMN, el ultrasonido se basa en la generación de una serie de pulsos u ondas mecánicas, que como su nombre lo indica, emplea frecuencias mayores al rango audible por el oído humano. El límite varía dependiendo de la persona y es de unos 20 KHz aproximadamente. Los equipos de ultrasonido operan con frecuencias más elevadas aunque la mayoría de los transductores actualmente empleados operan a frecuencias mucho más altas (MHz) (Vargas *et al.*, 2008).

Las ondas o perturbaciones ultrasónicas como cualquier otro tipo de onda, también pueden ser caracterizadas por su velocidad de propagación. Su velocidad de propagación depende esencialmente del medio material en el que viaja la onda. El valor de rapidez en medios acuosos como por ejemplo solución salina, es de aproximadamente 1540 m/s, en tejido corporal blando aumenta

³ Momento magnético que poseen los núcleos atómicos asociado al movimiento del núcleo al girar sobre sí mismo

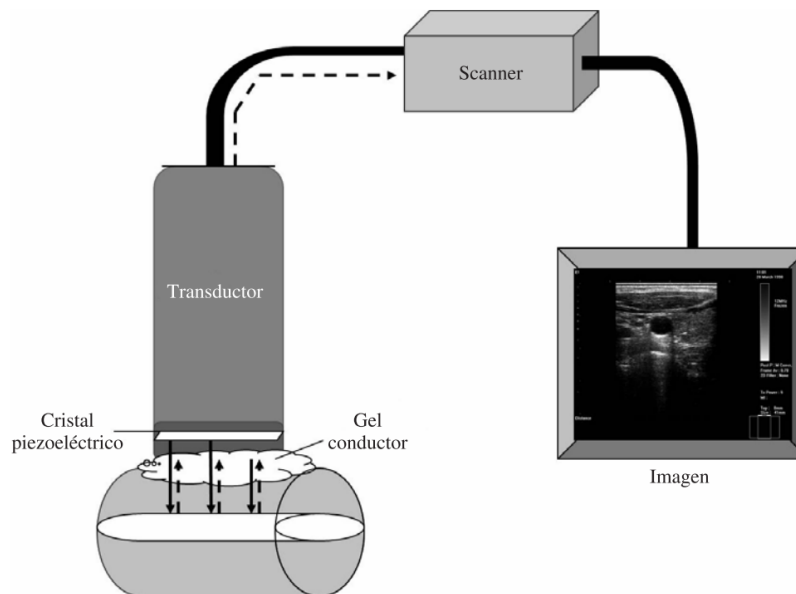
entre un 1 % y 2% y se ralentiza un 10% en tejido graso (Kossoff, 2000).

2.1.2.1. Generación de imágenes. Inicialmente un circuito eléctrico (transmisor) genera un tren de pulsos eléctricos orientados a un transductor. El dispositivo transductor convierte la energía eléctrica en energía ultrasónica y dicha energía en forma de onda, incide sobre el tejido a caracterizar. Las ondas sonoras se atenúan y reflejan, la energía reflejada regresa al dispositivo transductor el cual transforma las vibraciones reflejadas en corriente eléctrica.

Existe un circuito receptor que se encarga de cuantificar la amplitud y el tiempo de transmisión total de la onda ultrasónica. La amplitud de la perturbación que retorna, determina la imagen en escala de grises del tejido o sección afectada.

Figura 1.

Esquema de un circuito para la generación de imágenes mediante ultrasonido



Nota. Ejemplificación de un modelo generador de imágenes empleando ultrasonido. Tomado de (Vargas *et al.*, 2008).

2.2. Modelo neuronal artificial

En esta etapa es necesario abordar las nociones generales sobre el concepto de red neuronal artificial, algoritmo y red tipo GAN, modelo generador, discriminador, arquitectura, topología de la red y se explicarán los aspectos sobre implementación del trabajo y modelo SRGAN.

2.2.1. Red neuronal artificial: conceptos y generalidades

El proceso de transformación y desarrollo de avances tecnológicos se asocia con la capacidad de crear e innovar, específicamente en función de las necesidades y exigencias del mundo actual. Dichas exigencias sugieren establecer desarrollos basados en tecnologías automatizadas o lo que recientemente se ha popularizado como lo es la inteligencia artificial (IA). "Buscamos software inteligente para automatizar el trabajo rutinario, comprender el habla o las imágenes, realizar diagnósticos en medicina y respaldar la investigación científica básica"(Goodfellow *et al.*, 2016).

Dentro del amplio espectro de alternativas basadas en IA, sobresalen los modelos y técnicas representativas del deep learning. Uno de los pilares del aprendizaje profundo son las redes neuronales artificiales, las cuales hacen parte de un paradigma inspirado en el modelo biológico cerebral del ser humano. Su funcionamiento se basa en interconexiones de unidades de computo (neuronas) que buscan efectuar y emular el comportamiento analítico, descriptivo, asociativo e interpretativo que posee el ser humano.

La explicación detallada sobre el modo en que funciona una red neuronal o el concepto matemático asociado no es el objetivo principal de este proyecto. En este caso se pretende abordar

el tema de la red neuronal mediante la explicación mucho más amplia del modelo y algoritmo GAN.

2.2.2. Red neuronal Residual: ResNet

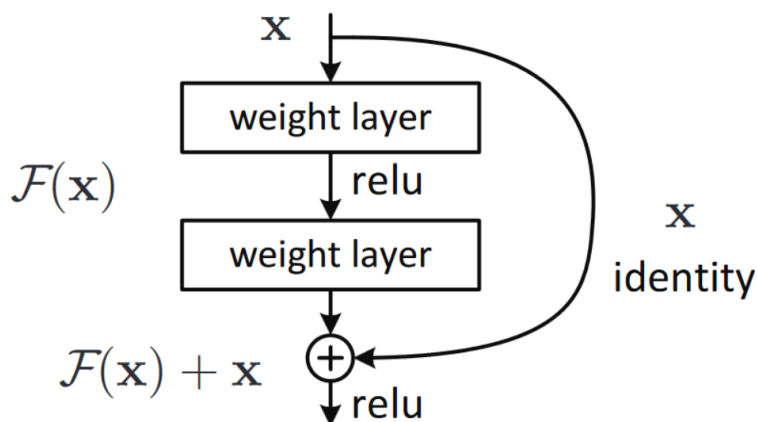
El modelo o arquitectura de red neuronal residual se basa en el uso de conexiones de salto o residuales, las cuales sirven para optimizar modelos especialmente profundos.

El concepto de red residual fue propuesto y desarrollado inicialmente por He *et al.* (2016) y representa un novedoso avance en el campo del deep learning. Los autores resaltan una problemática que denominan *degradación de la precisión del entrenamiento*, presente en arquitecturas multicapa o profundas (He *et al.*, 2016). Este problema de degradación genera irregularidades en el proceso de entrenamiento, debido a un desbordamiento de la capacidad de procesamiento de la red.

El modelo base que se observa en la figura 2, representa el ejemplo de conexión residual mediante la función *identidad* x , esta función mejora y optimiza el modelo profundo evitando el problema de degradación.

Figura 2.

Esquema de conexión secuencial de capas con *skip connection*, base del modelo ResNet.



Nota. Imagen tomada de (He *et al.*, 2016).

2.2.3. Red tipo GAN

Dentro de la extensa variedad de arquitecturas y topologías de redes neuronales propuestas en la literatura, la red GAN ha demostrado interesantes y positivos resultados enfocados particularmente en el procesamiento, generación y manipulación de imágenes. El algoritmo GAN posee grandes ventajas debido a la enorme variedad de contextos de aplicación, adaptabilidad en cuanto características de arquitectura y la posibilidad ser un modelo generativo de alta fidelidad si se establece un correcto proceso de entrenamiento (Salimans *et al.*, 2016).

El término GAN representa las siglas en inglés de *Generative Adversarial Network* y su concepto fue desarrollado y presentado inicialmente por Goodfellow *et al.* (2014). Esta red neuronal basa su funcionamiento en el desarrollo constante de un entrenamiento simultáneo de dos

subredes, una red que se denomina *Generadora* y otra *Discriminadora*. El modelo generador se encarga de entregar muestras cercanas a la realidad, estos datos cercanos a la realidad son muestras sintéticas falsas respecto a la información ground truth. El objetivo principal del generador es producir datos cada vez mas realistas. El discriminador actúa como juez o evaluador respecto a las muestras sintéticas desarrolladas por el generador, basándose en estimaciones de probabilidad. En síntesis, los modelos actuarán como adversarios en donde el objetivo del generador será engañar con datos falsos muy realistas al discriminador.

La imagen de la figura 3 representa solo un ejemplo del esquema general o más simple de un modelo GAN. En esta imagen se puede apreciar como la red generadora recibe información aleatoria y la información de error que proviene de las funciones de pérdidas o de costo de la red discriminadora. El generador se retroalimenta de la actualización en el estado de la función de costo establecida en el discriminador.

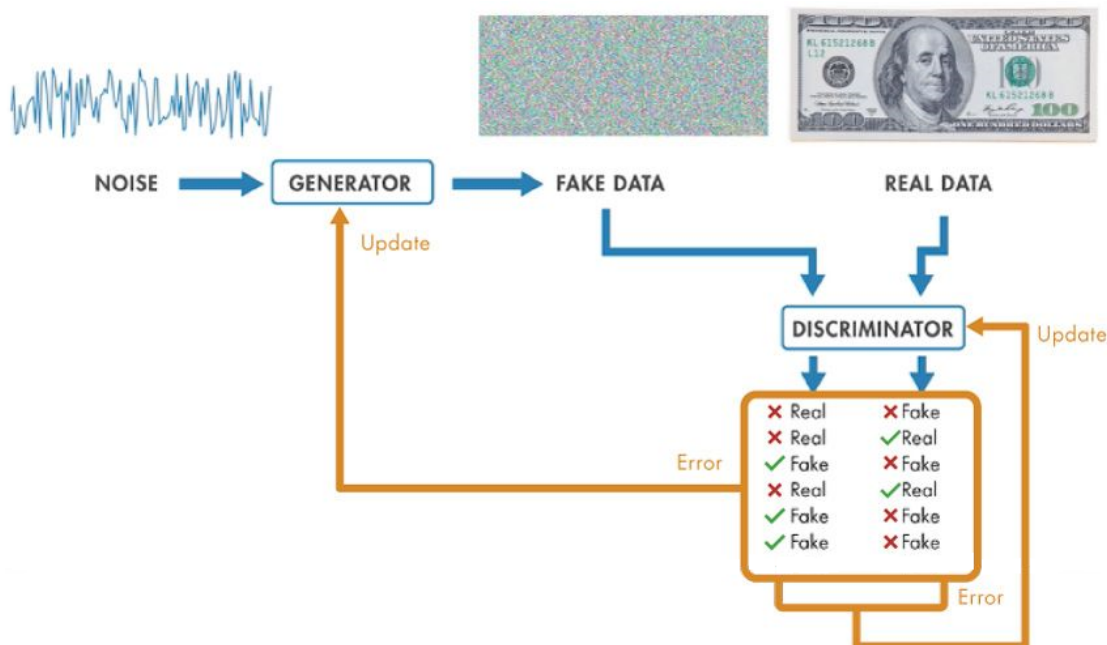
Un objetivo desde el punto de vista de entrenamiento del modelo GAN, es encontrar el valor mínimo posible de las funciones de costo de cada modelo (generador y discriminador), sin afectarse mutuamente. La función de costo del modelo generador se define como $J^{(G)}(\Theta^{(G)}, \Theta^{(D)})$ y la del discriminador es $J^{(D)}(\Theta^{(G)}, \Theta^{(D)})$, donde $\Theta^{(G)}$ y $\Theta^{(D)}$ representan las distribuciones paramétricas⁴ de los modelos neuronales del generador y discriminador respectivamente (Salimans *et al.*, 2016). Las funciones paramétricas no están correlacionadas entre sí pero las funciones de costo

⁴ Concepto relacionado con los valores o pesos sinápticos que toman las neuronas en el modelo neuronal

si dependen de estos parámetros. "Desafortunadamente, una modificación de $\Theta^{(D)}$ que reduce $J^{(D)}$ puede aumentar $J^{(G)}$, y una modificación de $\Theta^{(G)}$ que reduce $J^{(G)}$ puede aumentar $J^{(D)}$ " (Salimans *et al.*, 2016).

Figura 3.

Ejemplo de arquitectura general del modelo GAN



Nota. Imagen tomada de (MathWorks., sf)

El proceso de confrontación entre las subredes puede ser visto desde el punto de vista de la teoría de juegos no cooperativos (Salimans *et al.*, 2016). En este tipo de juego los jugadores priorizan sus beneficios personales sobre los intereses de los demás participantes. Pero es posible que desde la búsqueda de objetivos individuales se llegue a un punto en donde cada jugador obtenga su mejor beneficio, esto se conoce como equilibrio de *Nash* (Lillo, 2006).

Al extrapolar el concepto de equilibrio de Nash sobre los modelos antagónicos, el interés radica en desarrollar y establecer una metodología para el entrenamiento de las dos subredes buscando dicho equilibrio. Esto en la práctica resulta tedioso y representa un trabajo amplio dentro de los desarrollos de algoritmos para el entrenamiento.

3. Modelo propuesto: SRGAN o GAN de Súper resolución.

El modelo utilizado en este trabajo de investigación se basa en el desarrollo propuesto por Ledig *et al.* (2017), en donde revelan resultados interesantes enfocados en la generación de imágenes de súper resolución sobre detalles de textura y escalamiento de imágenes. Los autores proponen una red con conexiones residuales para generar imágenes fotorrealistas de superresolución (SR).

La red procesa las imágenes como tensores que dependen de sus dimensiones alto, ancho y dimensión de características. Los tensores I^{LR} (baja resolución) y I^{HR} (alta resolución), representativos de imágenes con C canales de color, tienen dimensiones $W \times H \times C$. El tensor I^{SR} (superresolución) tiene dimensiones $rW \times rH \times C$ (Ledig *et al.*, 2017).

La diferencia primordial entre este proyecto y el modelo base, radica en el set de datos para entrenamiento, validación e inferencia que se empleó. El desarrollo y generación de dicho conjunto de muestras se explica con más detalle en los próximos capítulos.

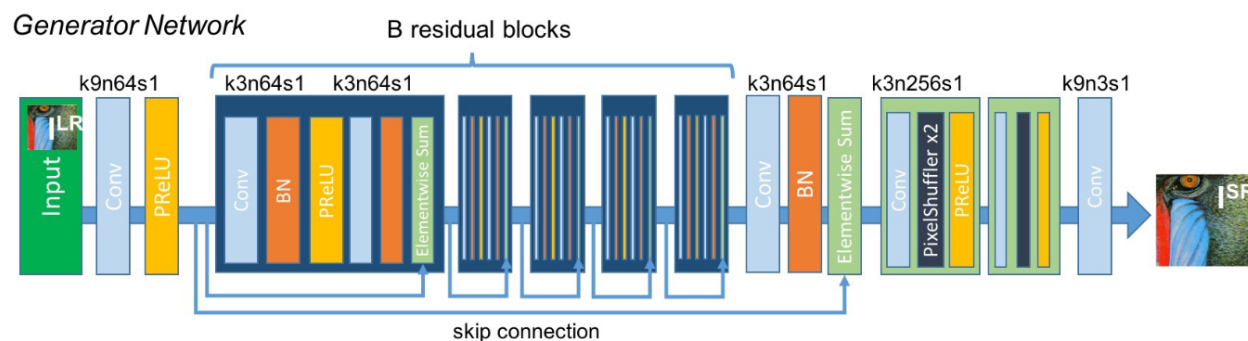
3.1. Arquitectura y topología del modelo Generador o SRResNet

Siguiendo de cerca el trabajo de Ledig *et al.* (2017), se entrenará el modelo SRResNet con la base de datos conformada en este proceso investigativo, la misma usada para entrenar la red SRGAN. El propósito del modelo SRResNet será el de usarse como modelo inicial para el generador de la red SRGAN, esto con el fin de evitar de otra forma que el modelo SRGAN converja en algún mínimo local.

La arquitectura del generador representativa del modelo SRGAN se basa en una estructura secuencial de bloques residuales. Como se mencionó anteriormente, el modelo generador es el mismo SRResNet y por lo tanto la imagen de la figura 4 representa su arquitectura y topología.

Figura 4.

Esquema de la arquitectura y topología de la red generadora del modelo SRGAN



Nota. Generador con bloques residuales. (k#) tamaño del kernel convolucional, (n#) cantidad de filtros de características, (s#) stride del proceso convolucional. Tomada de (Ledig *et al.*, 2017).

De la figura 4 se observa que en la entrada se tiene una imagen de baja resolución segui-

da de una primera capa convolucional (Conv⁵ con un Kernel de tamaño 9, 64 filtros y stride 1). Después de la capa convolucional sigue una capa de activación *Parametric Relu*⁶ (PRelu) que pasa su resultado a una secuencia de 16 bloques residuales idénticos (la cantidad de bloques residuales corresponde con la recomendación hecha por Ledig *et al.* (2017)). Los bloques están conformados por una capa convolucional (Kernel de tamaño 3, 64 filtros y stride 1), seguida de una capa de *Batch normalization*⁷ (BN) que normaliza los datos y los prepara para una capa de activación PRelu. Finalmente los datos entran a otra capa de convolución (Kernel de tamaño 3, 64 filtros y stride 1), para pasar a otro BN y terminar el bloque residual con una capa *Elementwise Sum*⁸ que recibe un *Skip connection*⁹ desde la entrada del bloque residual.

Posterior a los bloques residuales, se agregan: una capa convolucional (Kernel de tamaño 3, 64 filtros y stride 1), una capa BN y una capa Elementwise Sum que recibe una conexión de salto desde la entrada del primer bloque residual.

⁵ La capa convolucional se encarga de generar una matrix con características importantes de los datos de entrada.

⁶ La capa PRelu tiene la característica de permitir un alpha dinámico para cada neurona y que se mejore la adquisición de algunas características de los datos modificando la pendiente de activación para valores negativos.

⁷ Normalización por lotes. Es un proceso en el que su objetivo principal es acelerar el entrenamiento y hacer que la red sea más robusta al normalizar las entradas de cada capa para que tengan una media cercana a cero y una desviación estándar cercana a uno.

⁸ Suma elemento a elemento entre dos estructuras de tipo lista o tensor.

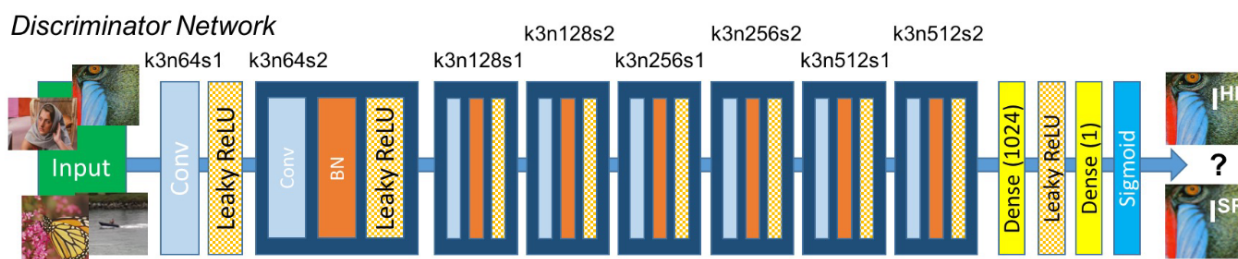
⁹ El Skip connection se define como un salto que conecta dos puntos de la red y que en este caso llega siempre a una capa Elementwise Sum en la cuál se suman los datos de entrada del primer punto con los resultados del procesamiento de las capas anteriores al Elementwise Sum.

Luego se agregan dos bloques denominados *upsample*¹⁰, cada bloque contiene una capa convolucional (Kernel de tamaño 3, 256 filtros y stride 1), una capa *PixelShuffler*¹¹ y una capa PRelu. La arquitectura del generador finaliza con una capa convolucional (Kernel de tamaño 9, 3 filtros y stride 1).

3.2. Arquitectura y topología del modelo Discriminador

Figura 5.

Esquema de la arquitectura de la red discriminadora del modelo SRGAN



Nota. En este modelo no se emplean bloques residuales, se usa la función de activación *Leaky Relu* y una capa densa de 1024 neuronas. Tomada de (Ledig *et al.*, 2017)

El modelo de red para el discriminador será entrenado para discernir si la imagen de salida SR (Súper resolución) puede considerarse real o falsa al compararla con una imagen HR (Alta resolución) o ground truth.

¹⁰ Agrupación de bloques para el aumento de la resolución (dimensiones de la imagen de entrada).

¹¹ La capa PixelShuffler usa la función *depth to space* de *TensorFlow* (Abadi *et al.*, 2015), para reorganizar los valores de la dimensión de profundidad del tensor de entrada, localizándolos en las dimensiones espaciales (alto y ancho), esto permite aumentar la resolución de una imagen al redistribuir los valores de píxeles existentes.

De la figura 5 es posible apreciar que inicialmente existe una capa de entrada con dos imágenes: la imagen SR y la imagen HR. Se observa que la red se compone de 8 capas convolucionales que poseen un Kernel de tamaño 3 y que varían en su cantidad de filtros, los cuales aumentan en un factor de dos cada dos capas de convolución empezando con $n=64$ y terminando con un $n=512$. También varían en los strides con un $s=1$ en las capas 1,3,5,7 y $s=2$ en las capas 2,4,6,8. De la figura 5 se examina que después de la primer capa Leaky Relu¹² las capas de se repiten, estos bloques de capas están compuestos por una capa de convolución, una capa BN y una capa Leaky Relu, en total son 7 bloques seguidos cuyo resultado se conecta con una capa densa¹³ de 1024 neuronas. Esta capa densa entrega sus datos a una capa Leaky Relu que continúa a una capa de una sola neurona y finaliza con una capa de activación *sigmoide*.¹⁴.

3.3. Entrenamiento del modelo SRGAN

La estructura SRGAN obedece a los principios y fundamentos generales que se abordaron en la subsección 2.2.3. Referente al entrenamiento del modelo, los autores Ledig *et al.* (2017) plantean un entrenamiento de las subredes partiendo del enfoque adversario característico del modelo GAN, que fue propuesto por Goodfellow *et al.* (2014).

¹² Leaky Relu a diferencia de PRelu posee un alpha fijo para todas sus neuronas, normalmente es un valor que se encuentra entre 0.1 y 0.3

¹³ La capa densa se encarga de hacer las comparaciones binarias de las características extraídas hasta el momento para si cada dato generado se puede determinar como verdadero o falso.

¹⁴ La capa sigmoide se encarga de reducir todas las probabilidades en una sola respuesta: verdadero o falso.

El modelo generador SRResNet se entrena para que logre establecer muestras sintéticas con características HR (imágenes ground truth), a partir de datos de entrada LR (imágenes de baja resolución). El generador tiene una estructura base de capas convolucionales y conexiones de bloques residuales, parametrizado por θ_G , en donde $\theta_G = \{W_{1:L}; b_{1:L}\}$ son los pesos y sesgos de la red con L capas. Durante el proceso de entrenamiento se ajustan y actualizan los parámetros en función del objetivo generativo sujeto a los procesos de optimización de la función de costo.

El discriminador se encarga de evaluar y analizar los datos que le entrega el generador y los compara con los datos reales. Este proceso comparativo e iterativo, constituye el modo en que el discriminador entrena y optimiza su rol.

3.4. Funciones de pérdida del modelo SRGAN

Para intentar contrarrestar y minimizar el error en el proceso de entrenamiento, es necesario establecer una función de pérdida. En este caso se adoptó la función de pérdida propuesta por Ledig *et al.* (2017), que representa una estrategia para intentar un equilibrio de Nash en el modelo GAN y se basa en intentar optimizar lo que se denomina *función de pérdida perceptual* l_{Per}^{SR} .

$$l_{Per}^{SR} = l_{Con}^{SR} + 10^{-3} l_{Gen}^{SR} \quad (1)$$

El modelo matemático de la ecuación 1 representa la suma de dos funciones: una función de *Pérdida de contenido* (l_{Con}^{SR}) y el modelo de *Pérdida adversaria* (l_{Gen}^{SR}).

3.4.1. Función de pérdida de contenido

Como su nombre lo indica, es una función que evalúa el desempeño del modelo desde el punto del contenido o información característica de los datos. La función de costo basada en el MSE es una variante y es empleada comúnmente en el campo de procesamiento de imágenes. Con el MSE se logra un PSNR relativamente alto pero su solución es precaria respecto a información de alta frecuencia, lo que genera falencias perceptivas (Ledig *et al.*, 2017).

La pérdida MSE se calcula píxel a píxel, mediante la siguiente expresión:

$$l_{MSE}^{SR} = \frac{1}{r^2WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} \left(I_{x,y}^{HR} - G_{\theta_G} (I^{LR})_{x,y} \right)^2 \quad (2)$$

En este trabajo se usa una función que se basa en una expresión similar al MSE, pero enfocado en los mapas de características y su objetivo es mejorar la *percepción de semejanza*. Para esto se probó la estrategia de Ledig *et al.* (2017) utilizando un modelo VGG preentrenado que fue desarrollado por Simonyan y Zisserman (2015). Con este modelo se realiza el mapeo de características en el proceso convolucional. De Ledig *et al.* (2017) sabemos que el modelo de pérdida está “basado en la distancia euclidiana entre mapas de características extraído de la red VGG19”. Estos mapas de características se extraen de las muestras sintéticas generadas ($G_{\theta_G} (I^{LR})$) y las imágenes de referencia (I^{HR}).

$$l_{VGG/i,j}^{SR} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\phi_{i,j} (I^{HR})_{x,y} - \phi_{i,j} (G_{\theta_G} (I^{LR}))_{x,y} \right)^2 \quad (3)$$

$$Error = (\phi_{i,j} (I^{HR})_{x,y} - \phi_{i,j} (G_{\theta_G} (I^{LR}))_{x,y}) \quad (4)$$

El término $\phi_{i,j}$ representa el mapa de características resultado de la j -ésima convolución posterior a la función de activación y antes de la i -ésima capa de *maxpooling*. $W_{i,j}$ y $H_{i,j}$ son las dimensiones referentes a los respectivos mapas de características del modelo VGG19 (Ledig *et al.*, 2017).

Como característica se empleó el modelo VGG de $i=5$ (5 capas de convolución) y $j=4$ (4 capas de maxpooling) pues ha sido uno de los mejores resultados respecto de la semejanza en percepción destacado en (Ledig *et al.*, 2017). Se tiene en cuenta que si se desea comparar las funciones de contenido $l_{VGG/5,4}^{SR}$ y l_{MSE}^{SR} , es necesario cambiar la escala de los mapas de $l_{VGG/5,4}^{SR}$ multiplicando por un factor de $\frac{1}{12.75}$ como se explica en (Ledig *et al.*, 2017).

$$l_{Con}^{SR} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\frac{\phi_{i,j}}{12.75} (I^{HR})_{x,y} - \frac{\phi_{i,j}}{12.75} (G_{\theta_G} (I^{LR}))_{x,y} \right)^2 \quad (5)$$

En la ecuación 5 se muestra la función de pérdida de contenido.

3.4.2. Función de pérdida adversaria

La función de pérdida adversaria se relaciona directamente con el modelo GAN. Esta función se encarga de promover el proceso generativo de la red al intentar reproducir muestras plausibles que logren engañar al discriminador.

”La pérdida generativa l_{Gen}^{SR} se define en función de las probabilidades del discriminador $D_{\theta_D}(G_{\theta_G}(I^{LR}))$ sobre todas las muestras de entrenamiento” (Ledig *et al.*, 2017).

$$l_{Gen}^{SR} = \sum_{n=1}^N -\log D_{\theta_D}(G_{\theta_G}(I^{LR})) \quad (6)$$

La función $G_{\theta_G}(I^{LR})$ representa la imagen reconstruida por el modelo generador.

3.4.3. Función de pérdida del discriminador

Además de las funciones usadas en el modelo generador de la red, también es importante mencionar que el modelo discriminador posee una función de pérdida. La función en mención tiene el objetivo de medir el desempeño del modelo según las etiquetas realizadas por el mismo, en donde los datos reales o verdaderos son I^{HR} (alta resolución) y los falsos o generados se conocen como I^{SR} (súper resolución).

Los dos resultados de la red discriminadora serán sumados conformando el valor en la

función de pérdida para el discriminador como se muestra en la ecuación 7.

$$l_{Dis} = D_{\theta_D} (G_{\theta_G} (I^{LR})) + D_{\theta_D} (I^{HR}) \quad (7)$$

3.5. Especificaciones y ajuste de hiperparámetros

Con del modelo SRGAN se realizaron una serie de pruebas y ajustes sobre las distintas variables controlables dentro de la red. Los hiperparámetros son los valores del modelo que pueden ser ajustables y controlados a voluntad. En el trabajo de Ledig *et al.* (2017), se mostraron y validaron algunos hiperparámetros los cuales son soportados por sus pruebas realizadas y se observan en la tabla 1.

Existen otros tipos de hiperparámetros relacionados con los métodos y técnicas de entrenamiento, como por ejemplo el número de *iteraciones* del algoritmo de entrenamiento, el número de *épocas* y la tasa de aprendizaje o *learning rate*. Estos hiperparámetros se asocian con el algoritmo de optimización *Adam* con un $\beta_1 = 0.9$ para minimizar la función de error durante el entrenamiento.

3.5.1. Hiperparámetros para SRResNet

Las pruebas realizadas sobre el modelo SRResNet se profundizan en la sección 6.2.1 usando los mismos hiperparámetros de la tabla 1, recordando que éste modelo representa la arquitectura y topología del generador dentro la red SRGAN (explicado en la subsección 3.1).

Tabla 1
Algunos hiperparámetros de arquitectura de los modelos adversarios

| Bloque | Capa | Hiperparámetros | | | | | | | |
|---------------------------------|-------------------------|-----------------|---------|---------------------------|----------|------------|----------------------|----------------|-----------|
| | | #Kernel | Strides | #mapas de características | α | $momentum$ | $scale, block_size$ | $shared_axes$ | #neuronas |
| | -Conv (1er capa Gen) | 9 | 1 | 64 | - | - | - | - | - |
| | -PReLU | - | - | - | - | - | - | [1, 2] | - |
| | -Conv (última capa Gen) | 9 | 1 | 3 | - | - | - | - | - |
| | -Conv (1er capa Disc) | 3 | 1 | 64 | - | - | - | - | - |
| | -Leaky ReLU | - | - | - | 0.2 | - | - | - | - |
| | -Densa1 | - | - | - | - | - | - | - | 1024 |
| | -Densa2 | - | - | - | - | - | - | - | 1 |
| Residual* (Generador) | -Conv | 3 | 1 | 64 | - | - | - | - | - |
| | -PReLU | - | - | - | - | - | - | [1, 2] | - |
| | -BN | - | - | - | - | 0.8 | - | - | - |
| | -Element wise Sum | - | - | - | - | - | - | - | - |
| Upsample (Generador) | -Conv | 3 | 1 | 256 | - | - | - | - | - |
| | -PixelShuffler | - | - | - | - | - | 2 | - | - |
| | -PReLU | - | - | - | - | - | - | [1, 2] | - |
| Secuencial** (Discriminador) | -Conv1 | 3 | 2 | 64 | - | - | - | - | - |
| | -Conv2 | 3 | 1 | 128 | - | - | - | - | - |
| | -Conv3 | 3 | 2 | 128 | - | - | - | - | - |
| | -Conv4 | 3 | 1 | 256 | - | - | - | - | - |
| | -Conv5 | 3 | 2 | 256 | - | - | - | - | - |
| | -Conv6 | 3 | 1 | 512 | - | - | - | - | - |
| | -Conv7 | 3 | 2 | 512 | - | - | - | - | - |
| | -BN | - | - | - | - | 0.8 | - | - | - |
| -Leaky ReLU | - | - | - | 0.2 | - | - | - | - | |

*Nota. * En cada bloque residual hay un par de capas: convolucional, BN y PReLU, como se muestra en la figura 4. En estas capas se repiten hiperparámetros. ** A diferencia del generador, en el discriminador cada capa de convolución tiene una distinta configuración de hiperparámetros.*

Se ha decidido adoptar los valores de los demás hiperparámetros al igual que en Ledig *et al.* (2017), teniendo así 10^6 iteraciones, 1000 épocas y un learning rate (Lr) de 10^{-4} .

3.5.2. Hiperparámetros para SRGAN

Respecto al modelo SRGAN se realizaron pruebas de las cuales se mostrarán en 6.1 y se usaron los mismos hiperparámetros de la tabla 1.

Se ha decidido adoptar los valores de los demás hiperparámetros al igual que en Ledig *et al.* (2017), teniendo así un entrenamiento con $2 * 10^5$ iteraciones, 200 épocas y un learning rate (Lr) que inicialmente es de 10^{-4} y que 10^5 iteraciones después de iniciado el entrenamiento decae a 10^{-5} .

4. Generación de la base de Datos

En este capítulo se darán las generalidades sobre todo lo relacionado con la obtención, configuración y generación de los datos (imágenes) que nutren el modelo de red neuronal. Se establece un marco en el que se mencionan el origen de las imágenes crudas RMN y las diferentes etapas de transformación a datos de ultrasonido.

4.1. Antecedentes y trabajo previo

Como se mencionó anteriormente, la generación de las muestras para el entrenamiento y validación de la red representan un recurso único y propio del desarrollo de este trabajo de investigación. El set final de imágenes se consolida a partir del conjunto de muestras que inicialmente estaban caracterizadas mediante el método de resonancia magnética nuclear, particularmente sobre datos clínicos de angiografías por resonancia magnética (ARM). Este set de datos se denomina OA-Breast⁴ y fue desarrollado por el laboratorio de Ciencias de Imágenes Computacionales, de la Universidad de Illinois en Urbana-Champaign. Los datos recopilados por OA-Breast son precedidos por los desarrollos e investigaciones realizadas por Lou *et al.* (2017).

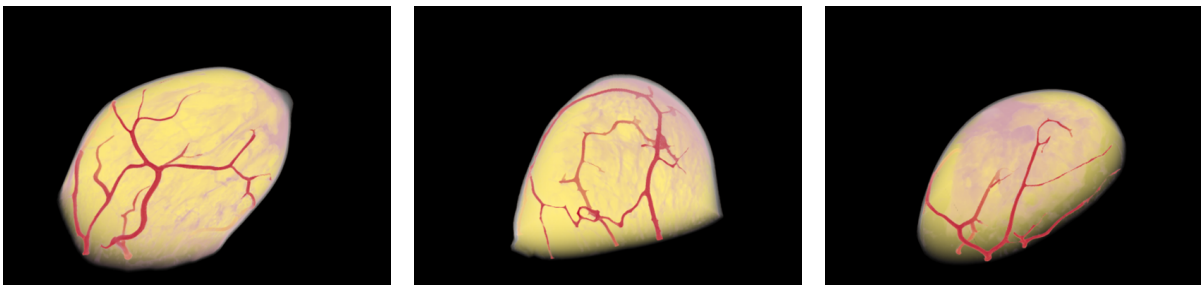
El set está compuesto de angiografías por resonancia magnética de tres pacientes con diferentes características de tejido mamario. Para facilitar el uso de los términos de lenguaje, se definen

⁴ OA-Breast: Base de datos pública que representa una colección de maniqués de senos para su uso en diversos estudios.

las muestras tipo 1 para los datos del paciente 1 con mamas fibroglandulares dispersas, tipo 2 para los datos del paciente 2 con senos heterogéneamente densos y tipo 3 para los datos del paciente 3 con senos extremadamente densos. Los tres tipos de ejemplares se aprecian en la Figura 6.

Figura 6.

Tipos de Senos en la base de datos



(a) Mamas fibroglandulares dispersas (Tipo 1) (b) Senos heterogéneamente densos (Tipo 2) (c) Senos extremadamente densos (Tipo 3)

Nota. Imagen tomada de (Lou *et al.*, 2017)

Nuestro desarrollo se basa en la mejora de resolución de imágenes sintéticas caracterizadas mediante datos de velocidad del sonido. En ese orden de ideas, las imágenes base de OA-Breast no representan un recurso idóneo según nuestros intereses. Por este motivo se realizaron una serie de etapas para la adecuación y la transformación final de las imágenes de RMN a información de contenido mamario representado por valores de ultrasonido. El proceso general de formación y generación de muestras artificiales basadas en datos de ultrasonido se puede resumir en 8 fases.

4.2. Fase 1: Vóxeles RMN a información de tejido mamario

Las imágenes crudas generadas a partir de datos RMN (Resonancia Magnética Nuclear) contienen información estructurada en vóxeles.⁵ Esta unidad de información cúbica no almacena valores distinguibles de cada tipo de tejido presente en las muestras y además mantiene una estructura tridimensional generando mayor costo de procesamiento. Posteriormente, las unidades de volumen de información son codificadas en cinco niveles asociados a un tipo de tejido mamario particular, el fondo (generalmente agua) recibe el valor de 0, el tejido fibroglandular es 2, la grasa es 3, la piel es 4 y a los vasos sanguíneos le corresponde el valor de 5. Este proceso es realizado exclusivamente por los autores de OA-Breast.

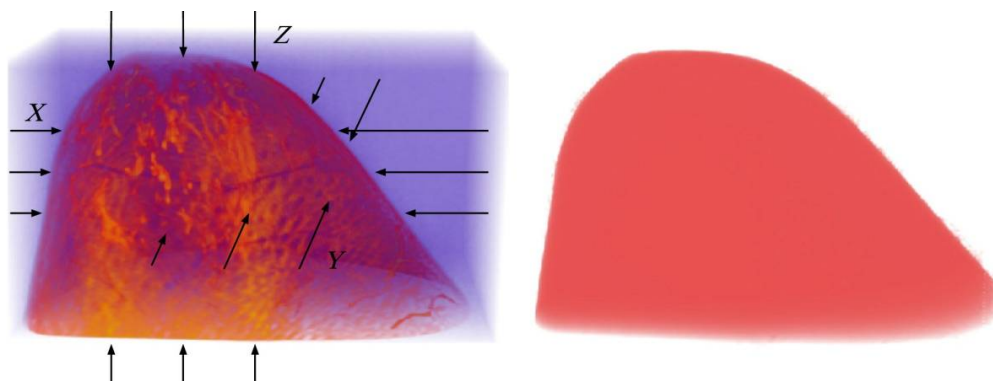
En la imagen de la Figura7 se puede notar un ejemplo de vóxel, con sus respectivos ejes de coordenadas.

4.3. Fase 2: Extracción de imágenes

De la fase inicial se obtienen 3 volúmenes de datos, de los cuales cada uno contiene un tipo de seno asociado. Teniendo en cuenta las direcciones que se pueden observar en la Figura7, de cada volumen de datos se pueden sacar imágenes en diferentes planos como se muestra en la Figura8 dependiendo de la dirección de corte. Mediante un script hecho para ejecutar en MATLAB, se realizaron cortes en dirección Z, paralelos al plano X-Y respectivamente como se puede observar

⁵ Vóxel: Es la unidad mínima procesable de una matriz tridimensional. Análogo al píxel bidimensional.

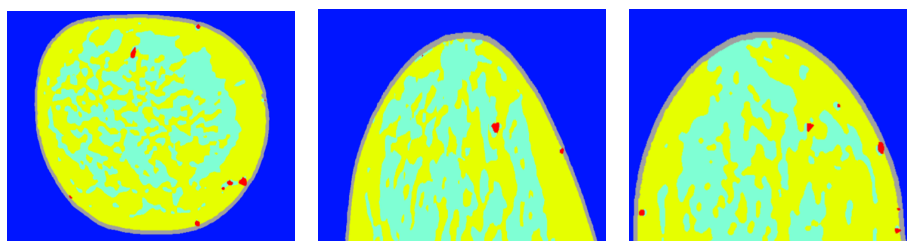
Figura 7.
Distribución espacial de datos RMN y su contorno



Nota. Imagen tomada de (Lou *et al.*, 2017)

en (a) de la Figura 8. Este proceso se realizó para los tres volúmenes y finalmente se consolidan las imágenes de alta resolución RMN.

Figura 8.
Imágenes consecuentes con el tipo de corte



(a) Dirección Z, plano X-Y (b) Dirección Y, plano X-Z (c) Dirección X, plano Y-Z

Nota. Imagen tomada de (Lou *et al.*, 2017)

4.4. Fase 3: Data augmentation

De los tres volúmenes se obtiene un total de 1395 imágenes RMN de alta resolución. Para lograr constituir un conjunto de datos de mayor tamaño, se realizó un aumento en la cantidad de

muestras o de imágenes mediante el proceso de data augmentation. El aumento de datos o data augmentation es un método común en el desarrollo y entrenamiento de sistemas basados en *deep learning*. Su uso se fundamenta esencialmente para evitar el fenómeno de sobre ajuste (overfitting⁶ en inglés) (Shorten y Khoshgoftaar, 2019). El aumento de las imágenes se realizó mediante procesos de manipulaciones básicas basadas en transformaciones geométricas, como rotación e inversión, estas técnicas se realizaron mediante el uso de MATLAB.

Es importante mencionar que el proceso de data augmentation se realizó previo a la transformación de ultrasonido. Es decir, el grupo de imágenes aumentadas fueron las representadas mediante RMN. Con el proceso de esta fase se logra obtener una base de datos compuesta por 4185 imágenes RMN de alta resolución.

Tabla 2

Rangos y tamaño del set de imágenes RMN

| Imágenes RMN (atributos) | Seno tipo I | Seno tipo II | Seno tipo III | Total ** |
|--------------------------|---------------|---------------|---------------|----------|
| Set original | [1 - 616] | [617 - 900] | [901 - 1395] | 1395 |
| Rotadas | [1396 - 2011] | [2012 - 2295] | [2296 - 2790] | 1395 |
| Reflejadas | [2791 - 3406] | [3407 - 3690] | [3691 - 4185] | 1395 |
| Total * | 1848 | 852 | 1485 | 4185 |

*Nota. * Total de imágenes referente a cada tipo de seno. ** Total de imágenes referente al set original y cada tipo de data augmentation.*

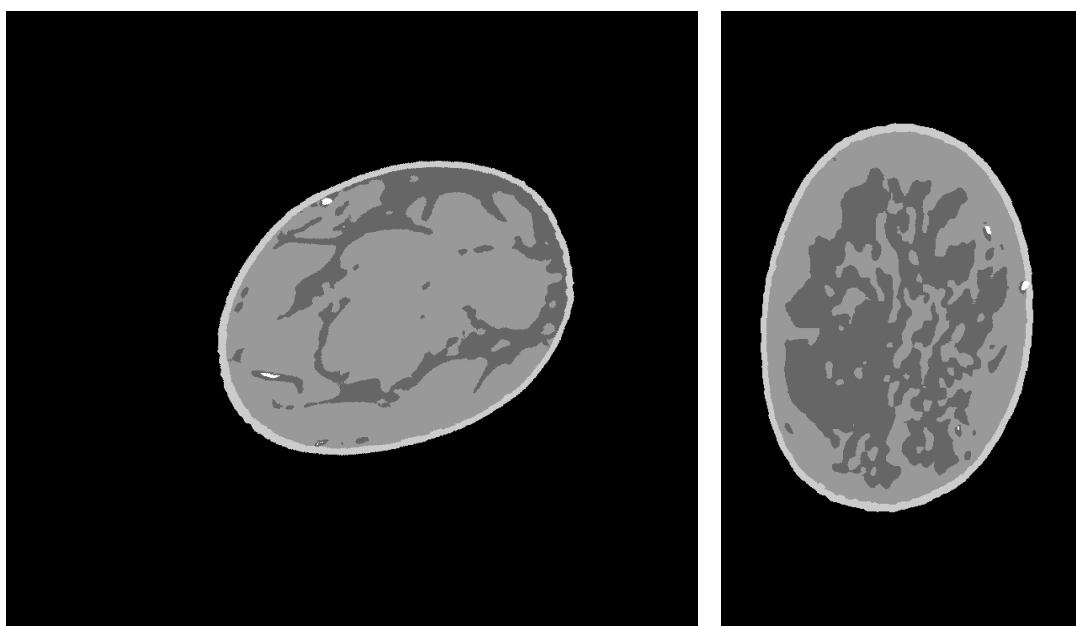
⁶ Fenómeno de sobre ajuste o sobre entrenamiento de un modelo, el cual genera incapacidad predictiva.

4.5. Fase 4: Procesamiento para la uniformidad de imágenes

Hasta el momento las imágenes poseen diferencias tanto en tamaño, como en distribución espacial. Las diferencias de tamaño se presentan debido a que son tres tipos de senos diferentes y las diferencias de distribución se presentan como consecuencia del data augmentation (4.4). Se procede a solucionar dichas diferencias de la siguiente manera:

Figura 9.

Ejemplo de imágenes de diferente tamaño



(a) Tamaño de imagen 854*773 Píxeles (imagen 1533)

(b) Tamaño de imagen 436*736 Píxeles (imagen 2138)

4.5.1. Padding

Se realiza un proceso de padding (relleno) para dejar todas las imágenes de un tamaño único 1000x1000 píxeles. El valor de dicho relleno es 0 (fondo), debido a que los demás valores

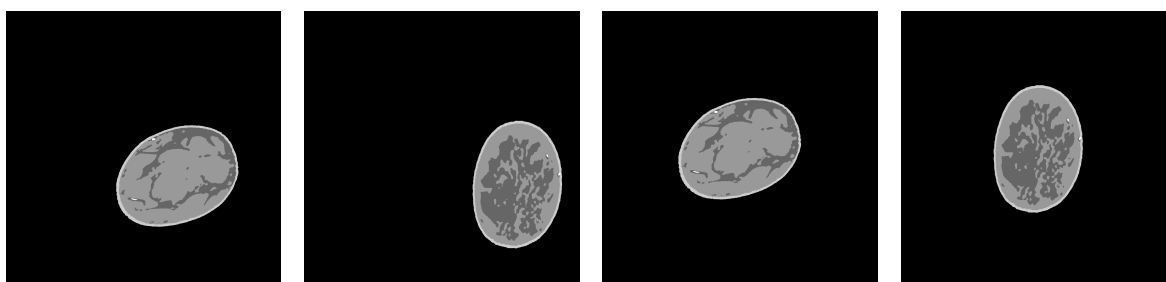
corresponden a los tejidos y se pueden generar datos falsos.

4.5.2. Centrado

Para centrar las imágenes se hallan sus centroides correspondientes teniendo en cuenta solamente los tejidos (ignorando el fondo de valor 0). Después de hallar los centroides se procede a calcular las distancias necesarias para desplazar la imagen de los tejidos y así centrarla reubicando sus dos centroides en la posición 500 tanto para el eje X, como para el eje Y.

Figura 10.

Ejemplo de centralización de imágenes.



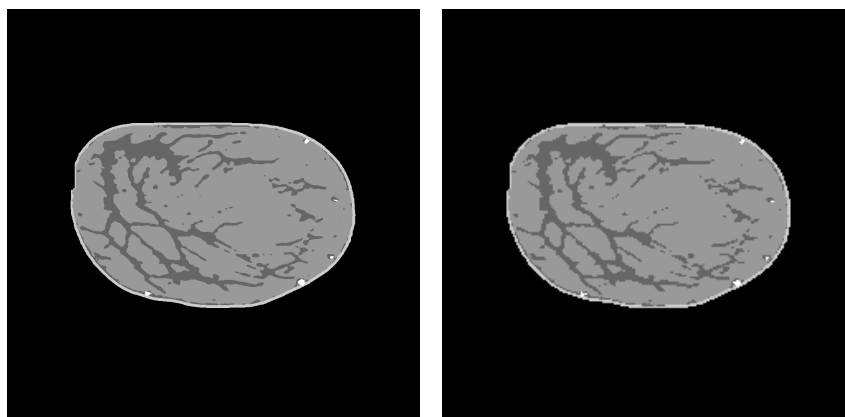
(a) Tamaño de imagen 1000x1000 píxeles (imagen 1533) (b) Tamaño de imagen 1000x1000 píxeles (imagen 2138) (c) Imagen de tamaño 1000x1000 píxeles centrada (imagen 1533) (d) Imagen de tamaño 1000*1000 píxeles centrada (imagen 2138)

4.6. Fase 5: Submuestreo de las imágenes de alta resolución

En esta fase las imágenes ofrecidas por OA-Breast aún están caracterizadas mediante el método RMN. En esta etapa las imágenes tienen un tamaño de 1000x1000 píxeles y son consideradas de alta resolución.

El modelo de red neuronal usado es un modelo generativo que entrega imágenes de salida ampliadas en un factor x4, por este motivo se procedió a realizar un submuestreo o diezmado a

Figura 11.
Par de imágenes RMN



(a) Imagen RMN de alta resolución (1000x1000 píxeles) (b) Imagen RMN diezmada de baja resolución (250x250 píxeles).

1/4 del tamaño original, es decir, las nuevas imágenes tienen un tamaño de 250x250 píxeles y para nuestra base de datos serán consideradas como imágenes RMN de baja resolución. Al finalizar esta fase se obtienen 4185 imágenes RMN de baja resolución y 4185 imágenes RMN de alta resolución.

Las imágenes de la figura 11 representan un par de alta resolución (a) y su correspondiente imagen diezmada (b).

Con las imágenes RMN de alta resolución se establece el conjunto de datos *ground truth* o datos catalogados como verdaderos, estas muestras sirven de base para realizar comparaciones y evaluaciones del modelo generativo.

4.7. Fase 6: Transformación de datos RMN a información de ultrasonido mediante FWI

Las imágenes RMN de baja resolución se transforman en una equivalente representación de datos caracterizados mediante ultrasonido. Ramirez *et al.* (2022) presentan resultados que respaldan el uso de la técnica de inversión de onda completa (FWI por sus siglas en inglés) para la generación de datos sintéticos de ultrasonido, con dicha técnica es posible generar muestras con información cuantitativa asociada a las características del tejido mamario. Con el método FWI ultrasónico se generaron las imágenes sintéticas de baja resolución representativas del seno, que en su totalidad fueron empleadas para el set de entrenamiento de la red.

En el proceso de transformación el algoritmo presentó fallas frente a algunas imágenes, arrojando algunos errores de procesamiento y por tanto no fue posible reconstruir todas las imágenes que resultaron de la fase 5 (4.6). Entre las posibles causas de fallo se encuentra que el área a reconstruir es muy pequeña o en otro caso que el tejido está muy cerca de las fuentes de disparo ultrasónico.

Como resultado de este proceso FWI se obtiene una cantidad total de 2416 imágenes de baja resolución y por tanto de aquí en adelante la base de datos se compone de 2416 imágenes para alta resolución (RMN), como para baja resolución (FWI).

4.8. Fase 7: Cuantificación de las imágenes en formato de 8 bits

Finalmente las imágenes reconstruidas a partir de FWI fueron cuantificadas píxel a píxel con 8 bits en un rango de valores de 0 a 255. Este proceso de determinación de valores se asoció a

Tabla 3
Conjunto final de imágenes reconstruidas (FWI)

| Tipo de seno | # Imágenes | # Imágenes | Entrenamiento | Validación |
|--------------|-------------|-------------|---------------|------------|
| | RMN * | FWI ** | 80 % | 20 % |
| Tipo I | 1848 | 1365 | 1092 | 273 |
| Tipo II | 852 | 392 | 314 | 78 |
| Tipo III | 1485 | 659 | 527 | 132 |
| Total | 4185 | 2416 | 1933 | 483 |

Nota. * Se consideran todas las imágenes: Set original y aumentadas. ** Imágenes reconstruidas correctamente. Los valores para entrenamiento y validación corresponden exclusivamente a las imágenes FWI.

un tipo de tejido particular en la escala de 0 a 5, que así mismo se relaciona con un valor específico de velocidad del sonido comprendido en un rango entre 1437 a 1577 m/s. En la tabla 4 se resume esta fase.

Tabla 4
Valores y parámetros asociados a los tipos de tejidos

| Tipo de tejido | Valor asociado [0-5] | Cuantificación | Velocidad del sonido m/s |
|-----------------------|----------------------|-------------------|-----------------------------|
| | | [0-255] 8 bits | |
| Fondo (Agua) | 0 | 0 | 1520 |
| Piel | 4 | [154 - 204] | 1577 |
| Grasa | 3 | [103 - 153] | 1437 |
| Vasos sanguíneos | 5 | [205 - 255] | 1570 |
| Tejido Fibroglandular | 2 | [1 - 102] | 1540 |

4.9. Fase 8: Selección final de imágenes para la base de datos

Para la consolidación final de los conjuntos de entrenamiento (80%) y validación (20%), se definen tres conceptos o características asociadas al set de muestras FWI. A continuación se explican brevemente los tipos de bases generadas.

Las características de selección para la conformación de los grupos de imágenes, se definen en *tipos de densidad* (densidad tipo I,II y III) y el grupo de imágenes que comparten el mismo *atributo* (imágenes iniciales u originales e imágenes rotadas).

De manera experimental se decide crear tres conjuntos de datos para observar el comportamiento de la red. Como resultado de esta fase 8 resultan las siguientes bases de datos:

4.9.1. Base de datos balanceada (Tipo A)

Contiene un balance en proporción de imágenes para cada tipo de densidad y los atributos original y rotado. En esta selección se omitieron las imágenes reflejadas y la idea principal de este conjunto es establecer un set de datos con la misma cantidad de imágenes, referente a cada clasificación (tipos de senos y atributo). En síntesis, se seleccionó el grupo de imágenes con menor cantidad y sobre cada grupo restante se eligió de manera aleatoria dicha cantidad de imágenes.

El grupo con la menor cantidad de muestras es el de las imágenes rotadas del seno tipo 2, con un total de 195 imágenes. El total de imágenes de esta base de datos es de 1170, con 80% (936 imágenes) para entrenamiento y 20% (234 imágenes) para validación.

Tabla 5
 Cantidades de cada grupo de imágenes (Tipos y atributos)

| Seno tipo I | | | |
|---------------|-------|-------------------|----------------|
| Atributo | Total | Entrenamiento 80% | Validación 20% |
| Original | 473 | 378 | 95 |
| Rotado | 387 | 310 | 77 |
| Reflejado | 505 | 404 | 101 |
| Seno tipo II | | | |
| Atributo | Total | Entrenamiento 80% | Validación 20% |
| Original | 197 | 158 | 39 |
| Rotado | 195 | 156 | 39 |
| Reflejado | NE | NE | NE |
| Seno tipo III | | | |
| Atributo | Total | Entrenamiento 80% | Validación 20% |
| Original | 361 | 289 | 72 |
| Rotado | 298 | 238 | 60 |
| Reflejado | NE | NE | NE |

Nota. La expresión NE hace referencia al conjunto de imágenes que no fueron tomadas en cuenta o sencillamente no existen.

4.9.2. Base de datos desbalanceada en imágenes (Tipo B)

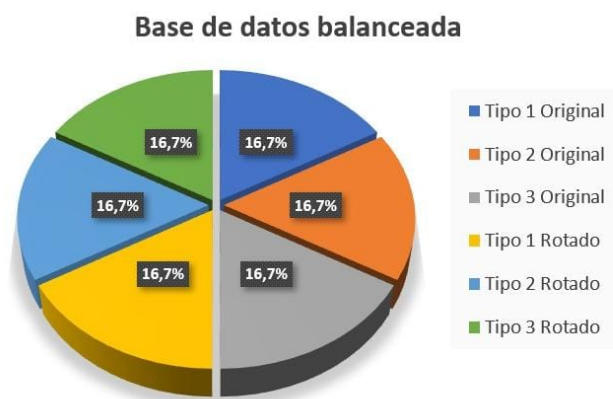
Contiene una desproporción de imágenes para cada tipo de densidad y los tipos de atributo. Se tiene un total de 1911 imágenes, con 80% (1529 imágenes) para entrenamiento y 20% (382 imágenes) para validación.

4.9.3. Base de datos desbalanceada en tipos de imágenes y atributos (Tipo C)

Contiene una desproporción de imágenes para cada tipo de densidad y también desproporción en atributos. agregando a la base de datos desbalanceada en imágenes las imágenes del atributo reflejado del seno tipo 1. Con un total de imágenes de 2416, con 80% (1933 imágenes) para entrenamiento y 20% (483 imágenes) para validación.

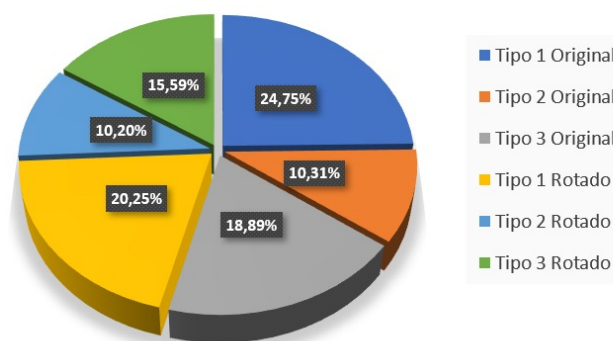
En la tabla 5 se puede observar las cantidades de imágenes para cada tipo de seno y cada atributo diferente. Además en la misma tabla se procede a calcular el 80% y 20% para cada una de las cantidades de imágenes, para conocer el número de imágenes que se pueden tomar de cada tipo de densidad y cada tipo de atributo.

Figura 12.
Bases de datos generadas



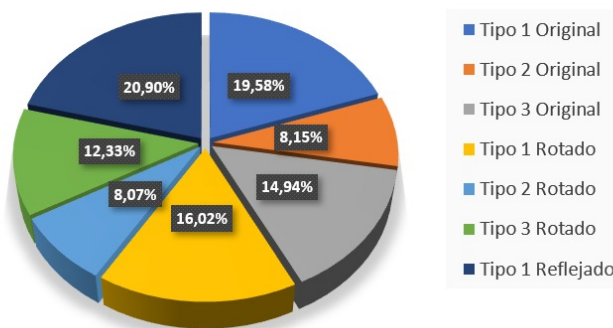
(a) Balanceada (Tipo A)

Base de datos desbalanceada en imágenes



(b) Desbalanceada en imágenes (Tipo B)

Base de datos desbalanceada en imágenes y atributos



(c) Desbalanceada en imágenes y atributos (Tipo C)

5. Entorno de desarrollo

El entorno de desarrollo para entrenar el modelo de red neuronal propuesto es un componente esencial en este proceso de implementación de inteligencia artificial y aprendizaje profundo. Para lograr un entrenamiento óptimo en los modelos propuestos, es crucial adecuar el entorno de desarrollo tanto en hardware, como en software. Es necesario tener en cuenta que el trabajo propuesto requiere del uso de diferentes herramientas para ejecutar aplicaciones de inteligencia artificial, como la visión por computadora, el procesamiento de imágenes y el reconocimiento de patrones.

5.1. Hardware de alto rendimiento

Para acelerar el proceso de entrenamiento, se requiere hardware de alto rendimiento, como unidades de procesamiento gráfico (GPU) de última generación o unidades de procesamiento tensorial (TPU). Estos recursos permiten la paralelización de cálculos intensivos y reducen significativamente el tiempo de entrenamiento.

5.1.1. Hardware usado

Para el entrenamiento de los modelos planteados se hace uso de un clúster perteneciente al grupo de investigación CPS (Connectivity and Signal Processing) de la Universidad Industrial de Santander, el cual tiene una capacidad de almacenamiento de 28 TB (terabytes) y dos nodos de

acceso, Tokyo y Kioto. Del clúster se hace uso del nodo Kioto que posee una capacidad de memoria RAM de 252GB, 64 procesadores (CPU's) y 8 GPU's (más potentes que las disponibles en Tokyo). En Kioto se emplea una de las GPU's disponibles para reducir el tiempo de entrenamiento. Las GPU's de Kioto son de la referencia Nvidia Tesla V100 de 16GB.

5.2. Software de aprendizaje profundo

Para el desarrollo y entrenamiento de los modelos se realiza una combinación de bibliotecas y frameworks de aprendizaje profundo de dos ecosistemas diferentes: PyTorch y TensorFlow. A continuación en la subsección 5.2.1, se realiza una descripción de las bibliotecas y los componentes específicos utilizados en el código.

Para el uso de todas las bibliotecas y funciones en general dentro del clúster, fue necesario hacer uso de un entorno virtual creado en Anaconda y que será explicado en 5.2.3.

5.2.1. Librerías y APIs usados

Estas herramientas ofrecen una amplia gama de funcionalidades, desde la definición de la arquitectura del modelo hasta la optimización y la evaluación del rendimiento.

5.2.1.1. PyTorch (torch). PyTorch es una biblioteca de código abierto de aprendizaje profundo ampliamente utilizada para la creación y entrenamiento de modelos de redes neuronales. Proporciona una estructura de tensores y herramientas para la construcción de modelos de manera flexible y eficiente.

5.2.1.2. TorchVision (torchvision). TorchVision es una parte de PyTorch que ofrece conjuntos de datos y modelos específicos para tareas de visión por computadora. Incluye herramientas para cargar y preprocesar imágenes, así como modelos preentrenados para tareas comunes de clasificación, detección de objetos, etc.

5.2.1.3. TensorFlow (tf). TensorFlow es otra biblioteca de aprendizaje profundo de código abierto que se utiliza en este trabajo en combinación con PyTorch. TensorFlow es conocido por su versatilidad y rendimiento, especialmente para la implementación de redes neuronales profundas (Abadi *et al.*, 2015).

5.2.1.4. Numpy (np). NumPy es una biblioteca fundamental para la manipulación de matrices y arreglos multidimensionales. Es ampliamente utilizado en tareas de preprocesamiento de datos y cálculos numéricos.

5.2.1.5. TensorBoard (torch.utils.tensorboard). TensorBoard es una herramienta de visualización desarrollada por TensorFlow que permite supervisar y visualizar métricas y resultados de entrenamiento. En este caso, se utiliza la versión adaptada de TensorBoard para PyTorch.

5.2.1.6. TensorFlow-Keras. TensorFlow-Keras es la API de alto nivel de TensorFlow para la construcción y entrenamiento de modelos de aprendizaje profundo. Se utilizan clases y funciones relacionadas con Keras para definir modelos, optimizadores, funciones de pérdida y métricas.

5.2.1.7. PIL. PIL es una biblioteca para el procesamiento de imágenes en Python. En este contexto, se utiliza para cargar y manipular imágenes.

5.2.1.8. OS. Se utiliza para realizar tareas relacionadas con el sistema, como la manipulación de archivos y directorios, la gestión de rutas de archivos y la configuración del entorno.

5.2.1.9. Funciones y clases específicas importadas desde TensorFlow-Keras. En esta sección serán nombradas las funciones empleadas en el entorno.

Adam: Es un optimizador que se utiliza comúnmente para el entrenamiento de redes neuronales. Ajusta automáticamente las tasas de aprendizaje durante el proceso de entrenamiento. Es una función adaptada del algoritmo de optimización desarrollado por (Kingma y Ba, 2015).

PiecewiseConstantDecay: Es una función de keras que se utiliza para definir una tasa de aprendizaje que cambia en pasos específicos durante el entrenamiento.

MeanSquaredError, BinaryCrossentropy, MeanAbsoluteError: Estas son funciones de pérdida comunes utilizadas en problemas de regresión y clasificación, respectivamente.

Model: Se utiliza para construir modelos de aprendizaje profundo personalizados utilizando la API funcional de Keras.

Mean, Accuracy: Métricas utilizadas para evaluar el rendimiento del modelo durante el entrenamiento.

TensorBoard: Callback de Keras que permite la integración de TensorBoard para el seguimiento de métricas y visualización.

5.2.2. VGG: Modelo neuronal preentrenado

VGG19: Es una arquitectura de red neuronal convolucional preentrenada que se utiliza para extracción de características. Esta arquitectura tiene 19 capas profundas y su modelo base fue propuesto originalmente por (Simonyan y Zisserman, 2015).

Existen varias arquitecturas y topologías consecuentes con el modelo base VGG que en su mayoría se emplean para procesos de visión por computadora y procesamiento digital de imágenes.

5.2.3. Entorno virtual

En este trabajo se hace uso de Anaconda, la cual es una plataforma de código abierto que se utiliza principalmente para la gestión y distribución de paquetes y entornos de desarrollo en Python y otros lenguajes de programación. Como Anaconda permite la creación y gestión de entornos virtuales independientes, en este proyecto se creó un entorno virtual con una versión de Python 3.9, además de instalar las librerías tensorflow y tensorflow-gpu.

6. Presentación de resultados

En este apartado se mostrarán los resultados referentes a los procesos de entrenamiento, validación e inferencia, sobre los modelos SRGAN y SRResNet (modelo contra el cual se harán las comparaciones). Estos modelos fueron entrenados usando los tres tipos de bases de datos que se mencionaron en la sección 4.9. Las pruebas se realizaron siguiendo las recomendaciones y especificaciones de diseño propuestas por (Ledig *et al.*, 2017) y mencionadas en la subsección 3.5. Posteriormente, se realizaron observaciones de los resultados para concluir con el conjunto de datos y modelo más conveniente, con el fin de realizar una prueba multiparamétrica, cambiando algunos hiperparámetros de entrenamiento.

Para la visualización y generación de gráficos referentes a las métricas de desempeño de los modelos, se usó la herramienta *TensorBoard* una API de *TensorFlow* (Abadi *et al.*, 2015).

6.1. Resultados referentes al modelo SRGAN

Sobre el modelo SRGAN se realizaron pruebas de entrenamiento y validación con los tres tipos de bases de datos (mencionadas en 4.9). Se tuvieron en cuenta las métricas como *MSE* (Error cuadrático medio), *SSIM* (índice de similitud estructural) y el *PSNR* (Proporción Máxima de Señal a Ruido).

Respecto a la medida *SSIM*, es posible asociarla con la métrica *accuracy*, pues el *SSIM* mide la similitud estructural entre la imagen sintética (generada por el modelo) y la imagen real.

La exactitud o accuracy mide el porcentaje de predicciones correctas realizadas por el modelo y al igual que el SSIM, este valor se aproxima a 1 si el modelo es altamente eficiente.

Se hace importante mencionar que las métricas que se mostrarán serán la media del total de las predicciones de validación en cada época de entrenamiento.

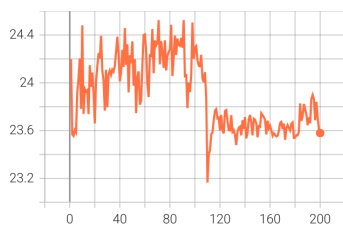
Dentro de la estructura SRGAN existen 3 posibles modelos asociados a cada tipo de base de datos usada. El modelo SRGAN que ha sido entrenado con la base de datos tipo A se denomina SRGAN-Tipo A y de igual manera con los datos tipo B y C, corresponden los modelo SRGAN-Tipo B Y SRGAN-Tipo C respectivamente. Estos submodelos SRGAN se entrenaron siguiendo las especificaciones de diseño e hiperparámetros que se mencionaron en la subsección 3.5.

6.1.1. Análisis cuantitativo

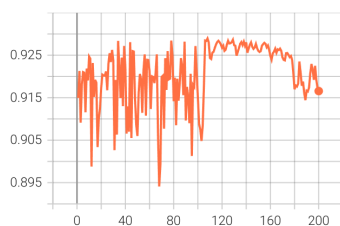
El análisis y evaluación de resultados se enfoca en las métricas evaluadas sobre los conjuntos de validación de cada tipo de base de datos. Por tanto, en la figura 13 se muestran los comportamientos de las métricas del modelo SRGAN respecto al proceso de validación, referente a los tres tipos de datos.

Figura 13.

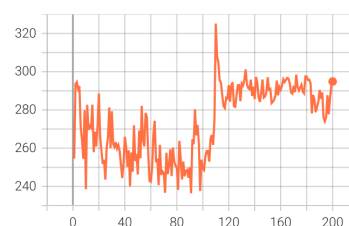
Métricas (PSNR, SSIM y MSE) de validación sobre los tres tipos de datos



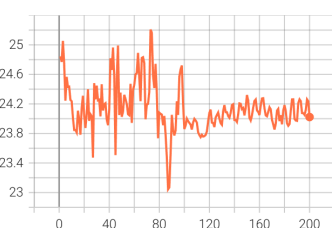
(a) PSNR-Datos Tipo A



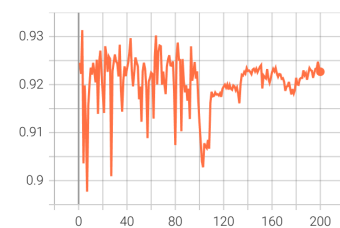
(b) SSIM-Datos Tipo A



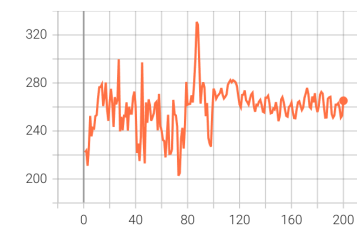
(c) MSE-Datos Tipo A



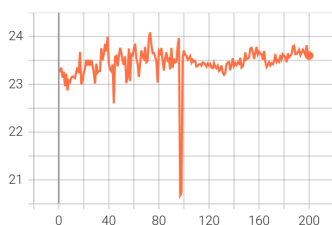
(d) PSNR-Datos Tipo B



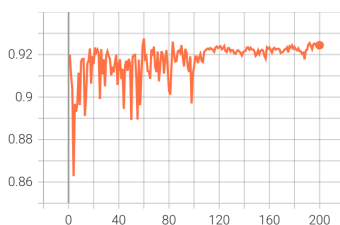
(e) SSIM-Datos Tipo B



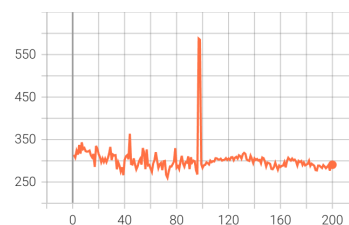
(f) MSE-Datos Tipo B



(g) PSNR-Datos Tipo C



(h) SSIM-Datos Tipo C



(i) MSE-Datos Tipo C

Los gráficos muestran la tendencia de cada una de las métricas en función de la cantidad de épocas. Se logra observar que en la mayoría de casos, el comportamiento de las medidas se intenta estabilizar en un rango de valores específico alrededor de la época 120, esto se debe al cambio del Learning rate que sucede después de las 100 épocas.

Tabla 6

Valores de métricas del modelo SRGAN respecto a cada tipo de base de datos

| Base de datos | PSNR | SSIM | MSE |
|---------------|-------|--------|-------|
| Tipo A | 23.58 | 0.9166 | 294.9 |
| Tipo B | 24.02 | 0.9227 | 265.2 |
| Tipo C | 23.6 | 0.9244 | 290.7 |

Nota. Estos valores corresponden al cómputo de la época 200

En cuestión de análisis cuantitativo es necesario recalcar que se pretende que las métricas sean observadas de la siguiente manera: el PSNR debería ser alto y por tanto a mayor magnitud mejor resultado, MSE por el contrario mejora a menor magnitud y SSIM es mejor cuanto más cerca esté de 1.

Desde el punto de vista cuantitativo, el modelo SRGAN-Tipo B, tiene los mejores valores de PSNR y SSIM. En este caso se presenta un PSNR más alto, el SSIM se aproxima a 1 en gran medida y el MSE es el más bajo.

Basados en la tabla 6 se concluye de manera preliminar que el modelo SRGAN-Tipo B, constituye una buena opción como modelo SRGAN con relación a los resultados de las métricas de desempeño.

6.1.2. Análisis cualitativo

Hasta el momento solo se han realizado comparaciones cuantitativas basadas en métricas referentes al modelo SRGAN, entrenado con las tres bases de datos. Para realizar observaciones

y evaluaciones del modelo desde un punto de vista cualitativo, se dispone de las imágenes de la figura 14, la cual muestra los resultados de los modelos correspondientes a cada tipo de base (a,b,c) y la imagen original o de referencia (Ground Truth).

Desde un análisis visual se reconocen patrones y formas representativas de los tipos de tejidos presentes en las muestras. Es necesario recordar y tener en cuenta los valores y rangos de cuantificación asociados a cada tipo de tejido mamario. En la tabla 4, se muestran los valores en los rangos de [0 - 5] y [0 - 255], referentes a una escala de grises que está sujeta a un tipo de tejido particular.

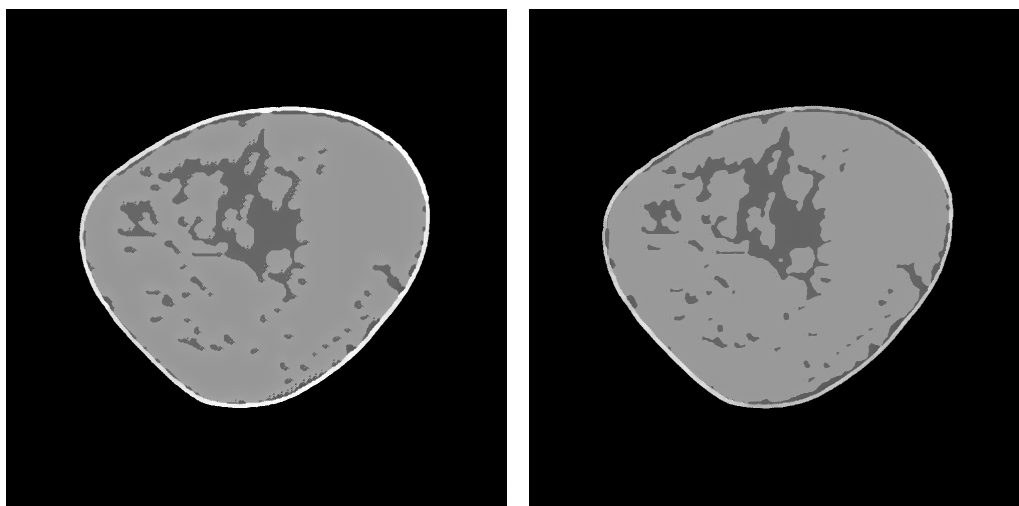
En la imagen b) de la figura 14 se muestra la salida o respuesta del modelo SRGAN-Tipo B. De manera perceptiva se aprecia una imagen sintética más aproximada a la realidad y se observa una reconstrucción de patrones sobresaliente.

Existen características de la imagen como por ejemplo el tipo de tejido asociado a los vasos sanguíneos, que se representa mediante el rango de [205 - 255] y visualmente se caracteriza con el color blanco dentro del espectro de grises. La principal falencia en las muestras generadas se orienta en la incapacidad de reconstruir de manera eficiente dichos tejidos.

Finalmente se optó por establecer el modelo SRGAN-Tipo B como el más apropiado, en consecuencia a los resultados cuantitativos y cualitativos anteriormente mencionados.

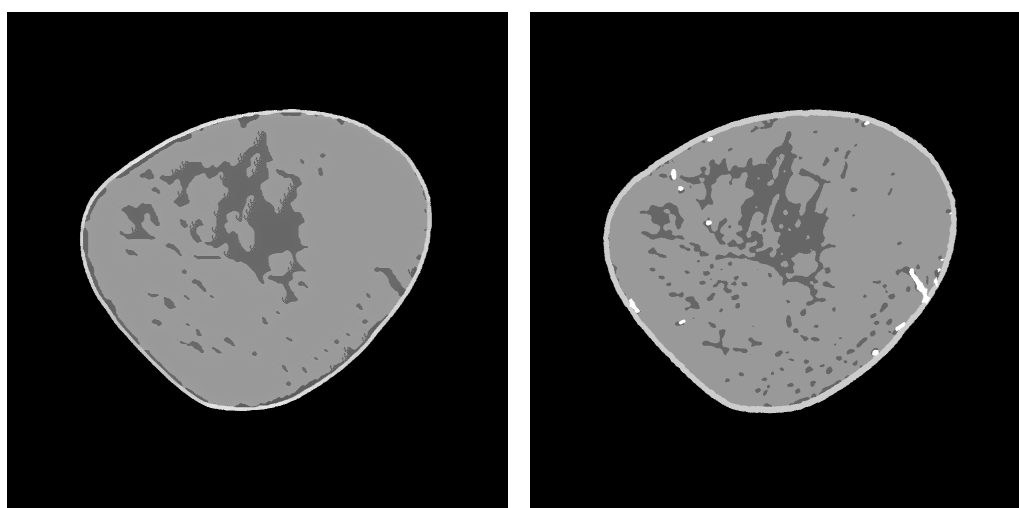
Figura 14.

Comparación de resultados cualitativos de la red SRGAN con imágenes en escala de grises.



(a) Tipo A

(b) Tipo B



(c) Tipo C

(d) Original

6.2. SRResNet vs SRGAN

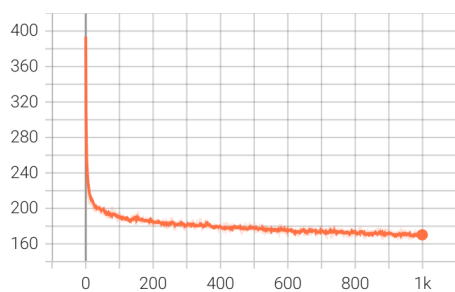
Teniendo en cuenta los resultados de los análisis propuestos en las subsecciones 6.1.1 y 6.1.2 en donde se ha elegido el modelo SRGAN-Tipo B como aquel que mejor desempeño obtuvo, se presentarán entonces los resultados de los modelos que fueron entrenados con la misma base de datos (Tipo B) y con los hiperparámetros propuestos en 3.5.

6.2.1. Resultados referentes al modelo SRResNet con base de datos Tipo B

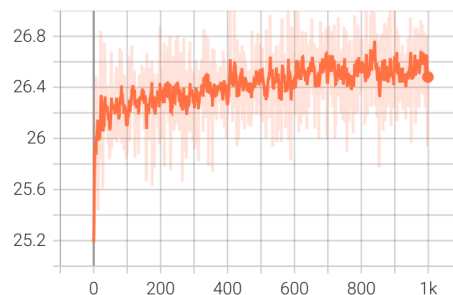
Se procede a presentar los resultados tanto para análisis cuantitativos en la Figura 15, como para análisis cualitativos Figura 16.

Figura 15.

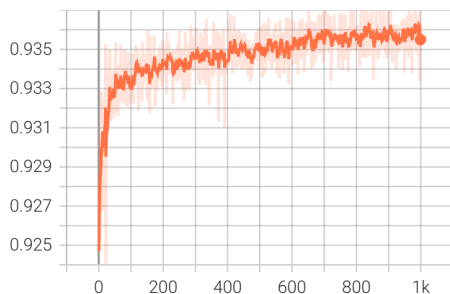
Función de pérdida y métricas (PSNR, SSIM y MSE) del modelo SRResNet-Tipo B



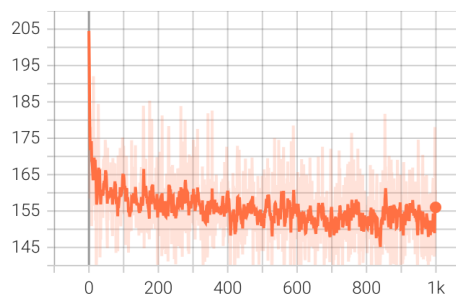
(a) Loss



(b) PSNR



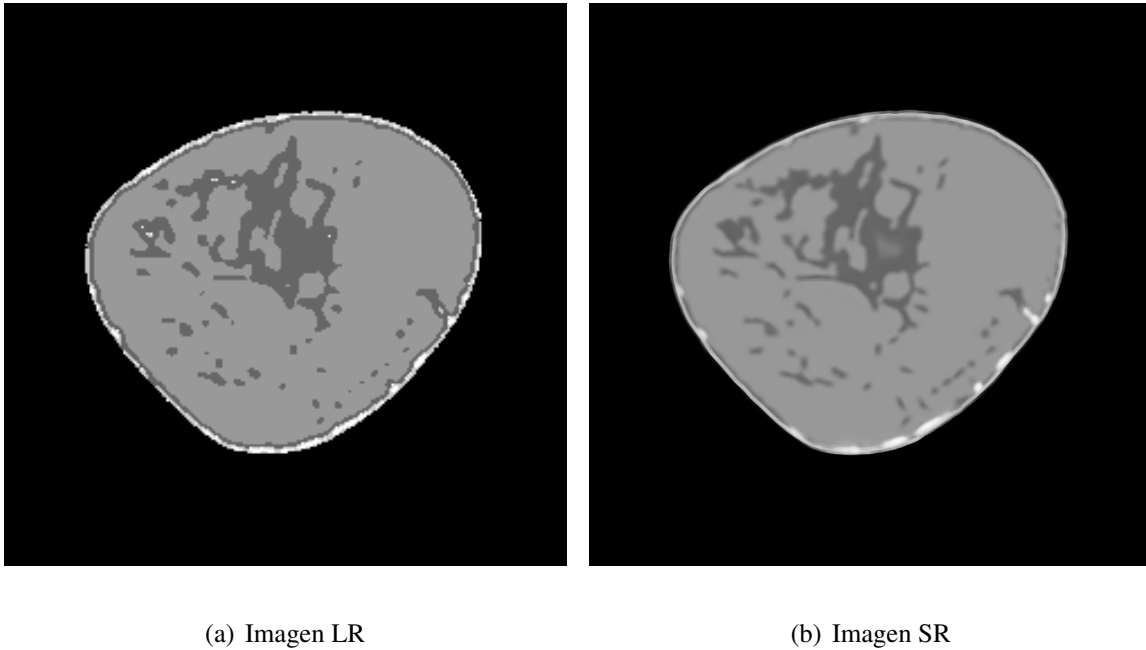
(c) SSIM



(d) MSE

De la figura 15 es posible encontrar ilustraciones de los comportamientos tanto de la función de pérdida, como de las métricas PSNR, SSIM y MSE evaluadas en cada una de las 1000 épocas de entrenamiento. Es importante mencionar que las métricas fueron calculadas con los datos de validación.

Figura 16.
Resultados visuales del modelo SRResNet-Tipo B



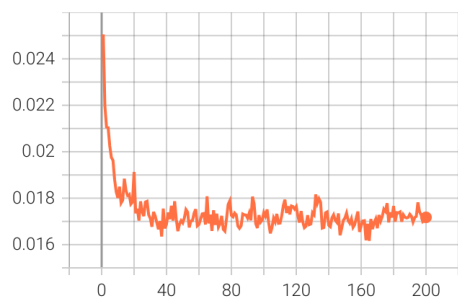
De la Figura 16 es factible visualizar un ejemplo del resultado al usar el modelo SRResNet. En (a) se muestra una imagen de entrada o de baja resolución (LR) y en (b) se observa la salida ampliada por un factor de 4.

6.2.2. Resultados referentes al modelo SRGAN con base de datos Tipo B

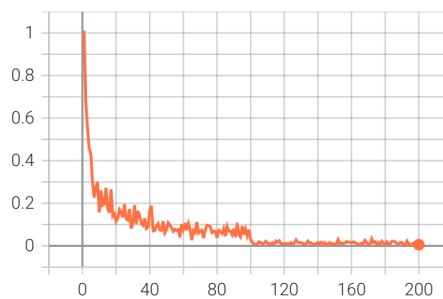
Al igual que en el apartado 6.2.1, se presentan los resultados para los análisis cuantitativos en la figura 17, como para los análisis cualitativos en la figura 18.

Figura 17.

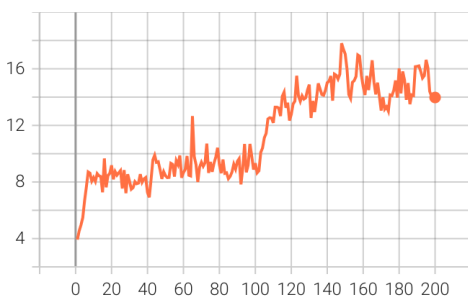
Funciones de pérdida y métricas (PSNR, SSIM y MSE) del modelo SRGAN-Tipo B



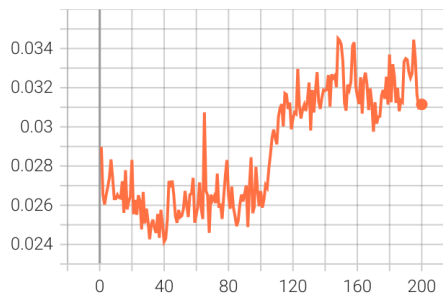
(a) Loss C



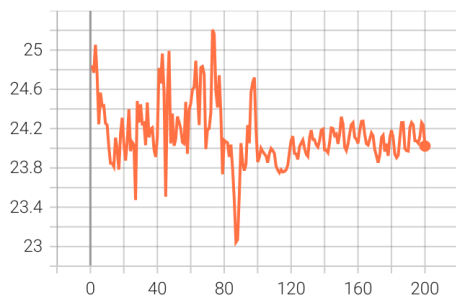
(b) Loss D



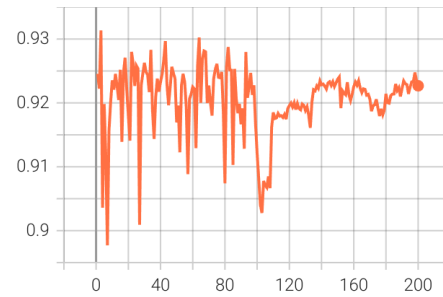
(c) Loss G



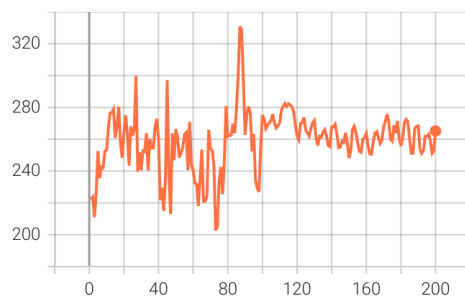
(d) Loss P



(e) PSNR



(f) SSIM

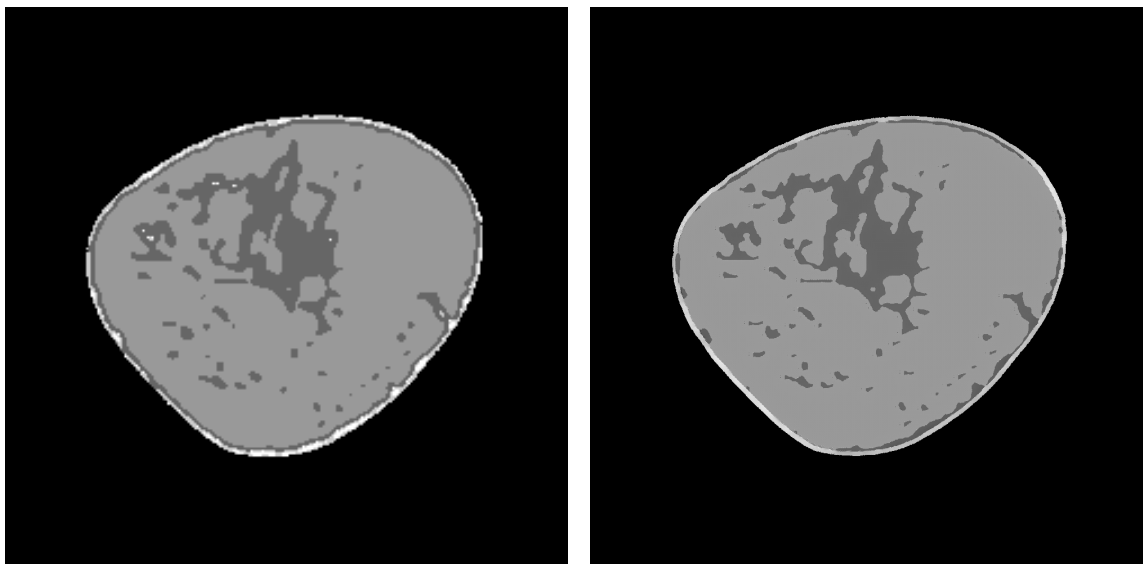


(g) MSE

De la figura 17 se advierten las diferentes funciones de pérdida que se evalúan en el modelo SRGAN como se especificó en 3.4 y además se logran observar las métricas PSNR, SSIM y MSE evaluadas con los datos de validación en cada una de las 200 épocas de entrenamiento.

Figura 18.

Resultados visuales del modelo SRGAN-Tipo B



(a) Imagen LR

(b) Imagen SR

Se observa en la figura 18 un ejemplo del resultado (b), de ingresar una imagen de baja resolución (a) al modelo SRGAN y ser escalada en un factor X4.

6.2.3. Análisis cuantitativo

Se observa la tendencia de las métricas para ambos modelos en las figuras 15 y 17, además de la tabla 7 con la información de validación en la última época.

El análisis cuantitativo puede ser rápido y directo, pues de la tabla 7 se puede observar que el modelo SRResNet-Tipo B ha superado en todas las métricas al modelo SRGAN-Tipo B.

Tabla 7

Comparación de métricas entre SRResNet y SRGAN

| Modelo | PSNR | SSIM | MSE |
|-----------------|-------|--------|-------|
| SRResNet-Tipo B | 26.33 | 0.9348 | 160.8 |
| SRGAN-Tipo B | 24.02 | 0.9227 | 265.2 |

Nota. Estos valores corresponden al cómputo de la última época en cada modelo.

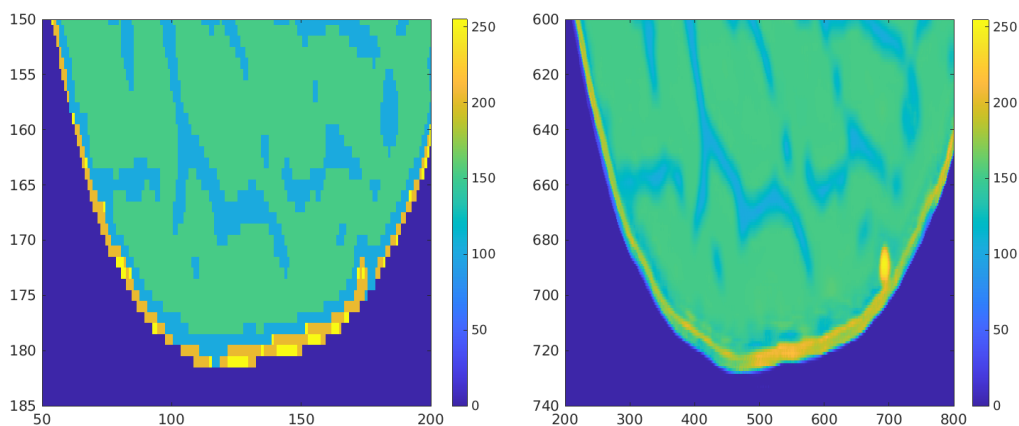
6.2.4. Análisis cualitativo

En este apartado se hace un análisis más minucioso, pues se pueden notar cosas interesantes a destacar en los modelos a manera de percepción visual. Se tienen como referencia para la comparación las figuras 16 y 18.

Con el fin de hacer un análisis cualitativo más detallado, se decide hacer un enfoque en una zona particular del seno (se usó la imagen LR con etiqueta 2695x4.png de la base de datos) en donde se puedan apreciar los diferentes tipos de tejidos. La imagen (a) de la figura 19 es la entrada de baja resolución, se puede observar el problema a resolver, pues al acercarse la imagen se obtiene una pérdida de resolución presentando una imagen pixelada y restringida en información valiosa. En (d) de la figura 19 se muestra el dato real y representa el objetivo desde el punto de vista del modelo generativo. Posteriormente se realiza una comparación cualitativa y cuantitativa entre los modelos enfrentados SRResNet (b) de la figura 19 y SRGAN (c) de la figura 19.

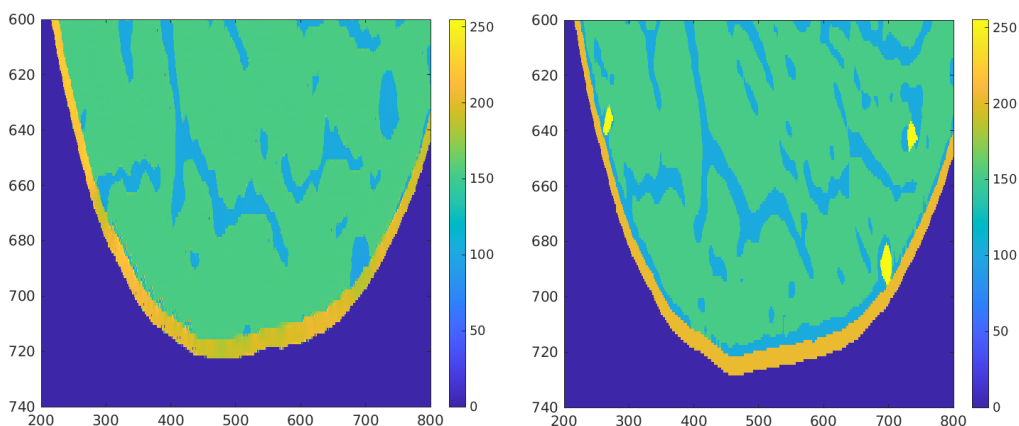
Figura 19.

Comparación cualitativa entre resultado de lo modelos Tipo B, imágenes a color.



(a) Imagen de baja resolución

(b) Imagen de red SRResNet-Tipo B



(c) Imagen de red SRGAN-Tipo B

(d) Imagen Ground Truth

Haciendo observaciones sobre las figuras 16, 18 y 19, es posible notar de la red tipo SRResNet una respuesta desde la percepción visual que genera un suavizado en la imagen, pero que como punto fuerte para este caso en particular puede generar todos los tipos de tejido, esto haciendo énfasis en los tejidos de vasos sanguíneos (los valores en el rango dinámico para diferenciar los tejidos

se encuentran en la tabla 4) los cuales no fueron generados en la respuesta del modelo SRGAN (c) de la figura 19 en ninguna de las tres zonas existentes como lo muestra (d) de la figura 19. Contrastar la respuesta del modelo SRResNet contra el ground truth puede dar la sensación de una visión miope de la imagen deseada, además de que suele generar la unión de tejidos en algunas zonas en las cuales no debería existir.

A diferencia del resultado del modelo SRResNet, sobre la salida de la red SRGAN se aprecia una imagen más consecuente con el ground truth, se muestra una imagen más clara y con pixeles más definidos que no presentan ningún tipo de difuminación o desvanecimiento. Se notan mejores detalles de definición, mejor reconstrucción de tejidos a nivel de contornos y cumple el objetivo por el cual fue diseñada, entregar imágenes reconstruidas con mejor resolución. En síntesis se puede afirmar que la respuesta de SRGAN nos puede ofrecer una mejor calidad perceptiva.

6.2.5. Discusión de resultados

A modo de resumen de lo expresado en 6.1.1 y 6.1.2, se puede decir que el modelo más adecuado por métricas de desempeño sería el modelo SRResNet, pero como en este proyecto es importante sobre todo un resultado visual, esto teniendo en cuenta que el profesional de la salud quien al fin de cuentas será el usuario del modelo escogido, basa sus análisis sobre su percepción de la imagen, observando detalles como la forma de algunos tejidos, su geometría, sus contornos, sus distribuciones, etc.

Finalmente se concluye que el modelo SRGAN-Tipo B presenta soluciones generativas

de percepción con mejor resultado que el modelo SRResNet-Tipo B, por este motivo se decide realizar las pruebas multiparamétricas y procesos de inferencia con dicho modelo SRGAN, a pesar de los resultados cuantitativos que se muestran en la tabla 7.

6.3. Pruebas multiparamétricas

Las pruebas multiparamétricas corresponden a una serie de entrenamientos secuenciales que se realizaron con el fin de hacer una sintonización de la red para poder verificar qué hiperparámetros son los más convenientes para el entrenamiento del modelo y así obtener una visión más amplia de las posibles soluciones.

Con la base de datos Tipo B se realizaron 9 pruebas en total, variando los hiperparámetros de Learning rate y alpha de la función Leaky Relu. Cada una de las pruebas fue nombrada con un número que va del 1 al 9 y se pueden observar en la tabla 8 junto con los hiperparámetros asignados para cada una de ellas.

Además de la información ofrecida en la tabla 8, se puede obtener una representación gráfica (figura 20) sobre los resultados de estas pruebas multiparamétricas. El gráfico representa un seguimiento de los distintos resultados de rendimiento, en función de los hiperparámetros empleados. Esto se visualiza gracias a TensorBoard.

Tabla 8

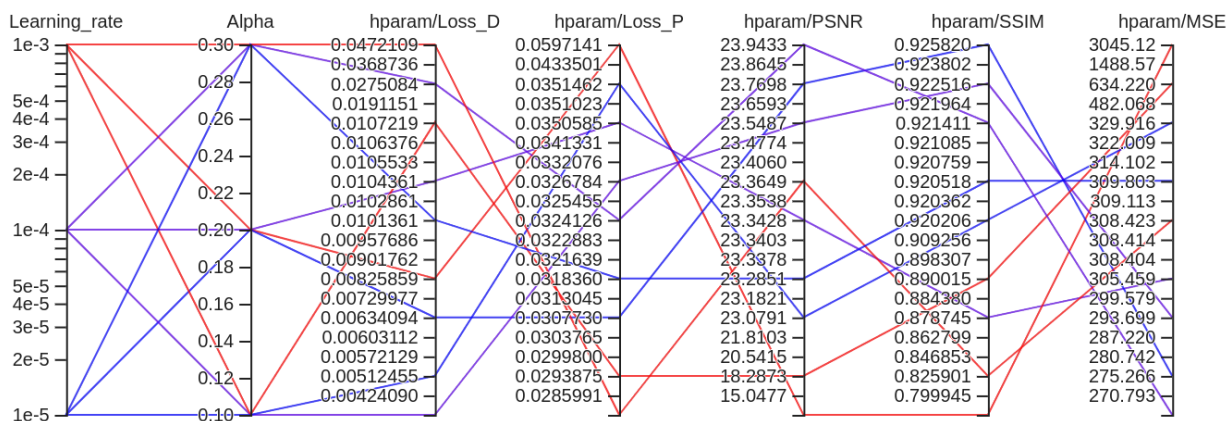
Resultados de pruebas multiparamétricas, al variar LR y alpha

| #Prueba | Learning_rate | Alpha | Loss_D | Loss_P | PSNR | SSIM | MSE |
|---------|------------------------------------|-------|---------|--------|------|-------|------|
| 1 | $1 \cdot 10^{-4}, 1 \cdot 10^{-3}$ | 0.2 | 0.0105 | 0.0351 | 23.3 | 0.879 | 308 |
| 2 | $1 \cdot 10^{-3}, 1 \cdot 10^{-2}$ | 0.2 | 0.00874 | 0.0761 | 11.8 | 0.774 | 4600 |
| 3 | $1 \cdot 10^{-5}, 1 \cdot 10^{-4}$ | 0.2 | 0.00634 | 0.0308 | 23.8 | 0.928 | 278 |
| 4 | $1 \cdot 10^{-4}, 1 \cdot 10^{-3}$ | 0.1 | 0.00336 | 0.0327 | 23.5 | 0.923 | 294 |
| 5 | $1 \cdot 10^{-3}, 1 \cdot 10^{-2}$ | 0.1 | 0.0107 | 0.0298 | 19.9 | 0.893 | 710 |
| 6 | $1 \cdot 10^{-5}, 1 \cdot 10^{-4}$ | 0.1 | 0.00557 | 0.0352 | 23.1 | 0.92 | 330 |
| 7 | $1 \cdot 10^{-4}, 1 \cdot 10^{-3}$ | 0.3 | 0.0317 | 0.0324 | 24 | 0.921 | 266 |
| 8 | $1 \cdot 10^{-3}, 1 \cdot 10^{-2}$ | 0.3 | 0.0575 | 0.0278 | 23.4 | 0.839 | 308 |
| 9 | $1 \cdot 10^{-5}, 1 \cdot 10^{-4}$ | 0.3 | 0.0101 | 0.0321 | 23.3 | 0.921 | 310 |

Nota. Es posible observar dos valores de Learning Rate para cada una de las pruebas, recordando que según lo explicado en 3.5, el primer valor hace referencia a las primeras 100 épocas y el segundo para las 100 posteriores.

Figura 20.

Comparación de métricas



Nota. El valor de Learning Rate corresponde al valor de las primeras 100 épocas, a manera de identificación, pues las pruebas se realizaron tal cual se indica en la tabla 8.

6.3.1. Análisis cuantitativo

A manera de análisis cuantitativo de la tabla 8 y de la figura 20 se aprecian las métricas de interés con las cuales podemos hacer un primer filtrado de modelos, viendo en los resultados que las métricas de validación PSNR, SSIM y MSE se ven afectadas negativamente en todas las pruebas que usaron su Learning Rate inicial de $1 \cdot 10^{-3}$ (pruebas 2,5,8), dejando claro que aún falta el análisis cualitativo para descartar por completo dichas soluciones. Las pruebas 7, 3 y 4 presentan mejores resultados en sus métricas de desempeño.

6.3.2. Análisis cualitativo

En cuanto al análisis de los resultados en cada uno de los modelos, se tiene como referencia la figura 21 para el resultado deseado y la figura 22 con los resultados de cada uno de los 9 modelos entrenados. A grandes rasgos es posible confirmar en este análisis el hecho de descartar los modelos 2,5 y 8 propuesto en el análisis cuantitativo 6.3.1, pues en 2 y 5 se pierde por completo la generación de tejidos internos, mientras que en 8 se generan tejidos internos sin un sentido lógico. Los resultados visuales de los modelo 3 y 7 respaldan los valores de las métricas de rendimiento vistas en la tabla 8, mientras que la prueba 4 aunque tuvo unas buenas métricas, a manera de percepción visual el resultado es insatisfactorio. Con respecto a resultados satisfactorios a manera visual y generación de tejidos, podemos proponer las pruebas 1,3,6,7 y 9. Teniendo en cuenta que particularmente la prueba 7 ofrece los mejores resultados cuantitativos y una percepción cualitativa satisfactoria.

Figura 21.
Imagen HR RMN (Ground Truth)

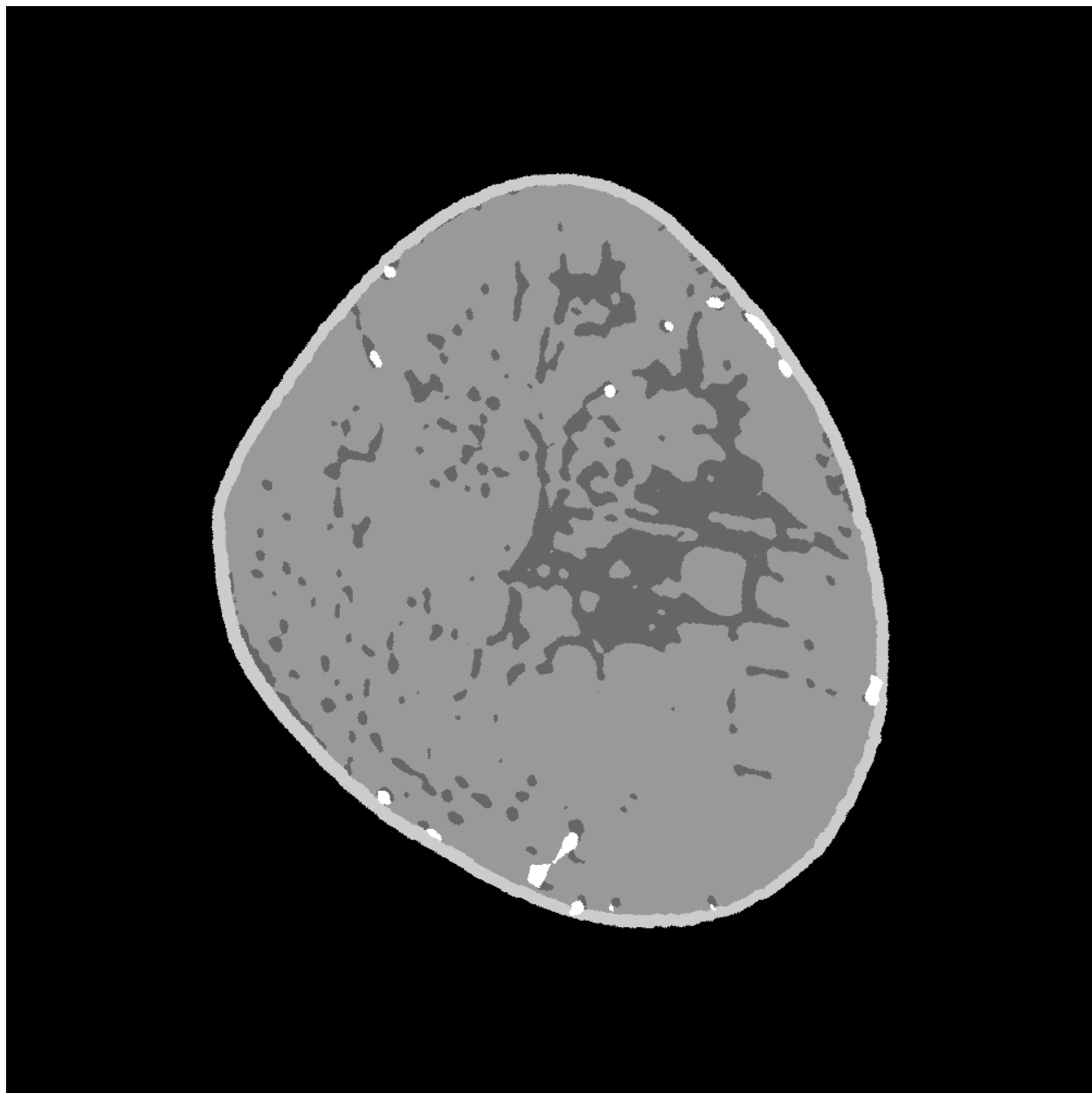
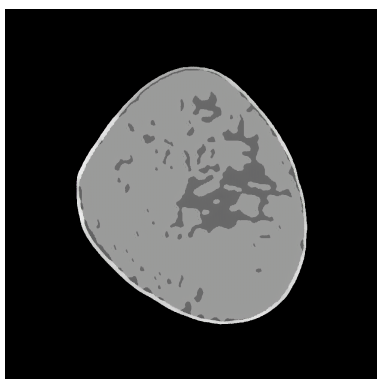
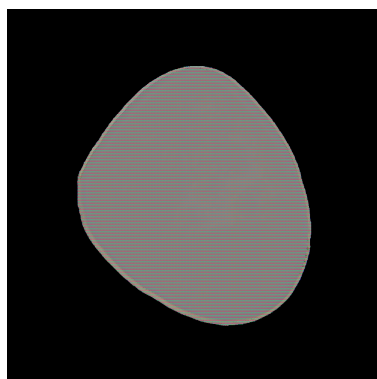


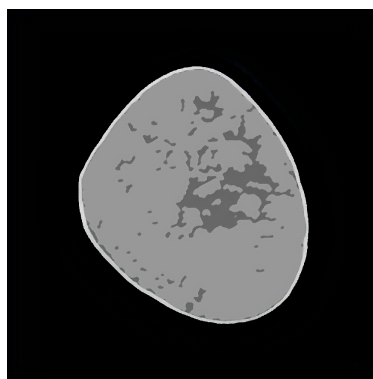
Figura 22.
Resultados visuales de las 9 pruebas



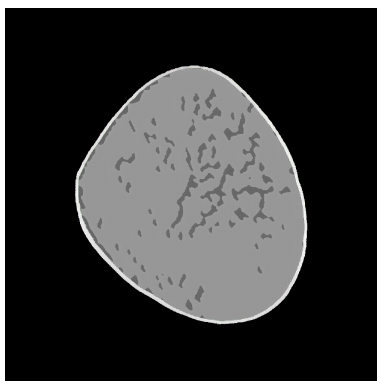
(a) Prueba 1



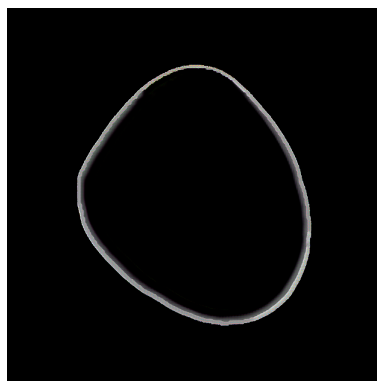
(b) Prueba 2



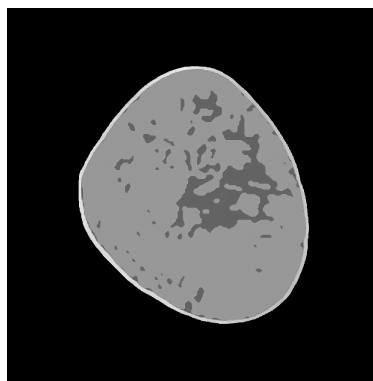
(c) Prueba 3



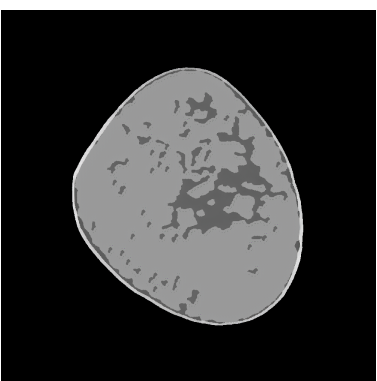
(d) Prueba 4



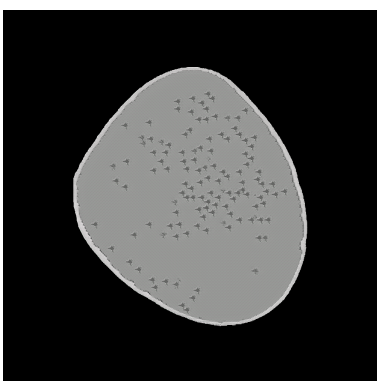
(e) Prueba 5



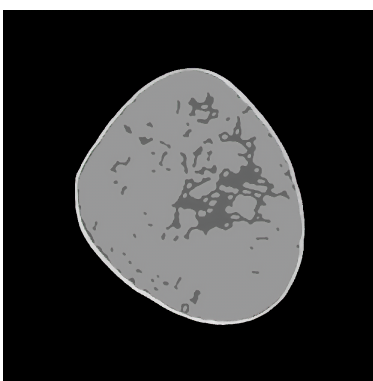
(f) Prueba 6



(g) Prueba 7



(h) Prueba 8



(i) Prueba 9

7. Conclusiones

Después de una investigación en el estado del arte, se logra identificar y proponer un modelo GAN que se ajusta a los requerimientos y objetivos del proyecto. La arquitectura SRGAN proporciona en gran medida resultados positivos enfocados en la mejora de resolución, además de imágenes de súper resolución, generando una imagen ampliada en un factor X4 respecto de las dimensiones de entrada.

Por otra parte se ha realizado un trabajo significativo en la creación y preparación de una base de datos apropiada para el entrenamiento y validación de la red. Esta base de datos incluye una amplia variedad de imágenes de alta y baja resolución, cuidadosamente seleccionadas y etiquetadas. La calidad y diversidad de los datos son fundamentales para garantizar un entrenamiento efectivo de la red. Además, se han aplicado técnicas de preprocesamiento para mejorar la consistencia y la calidad de los datos. Como producto final se ofrece una base de datos total que se puede dividir en tres bases de datos como fue expuesto en éste trabajo.

El entorno de desarrollo se ha configurado de manera integral utilizando lenguajes orientados a objetos como Python y librerías de inteligencia artificial, como TensorFlow, PyTorch y Keras. Esta configuración proporciona una base sólida para la implementación y experimentación con diferentes modelos de redes neuronales. Además, se ha optimizado para aprovechar al máximo los recursos computacionales disponibles, lo que permite un proceso de entrenamiento eficiente y

escalable.

Se ha llevado a cabo una evaluación exhaustiva de los resultados generados por la red utilizando un conjunto de imágenes reservadas específicamente para la validación. Esta etapa crítica del proceso ha permitido una evaluación cuantitativa y cualitativa de la calidad de las imágenes de alta resolución generadas en comparación con los datos reales. Los criterios de aceptación se han definido con precisión para determinar si los resultados sintéticos cumplen con los estándares deseados en términos de fidelidad y detalle.

En conjunto, estos logros representan un sólido progreso hacia el desarrollo de un sistema de alta y súper resolución basado en redes GAN. Los conocimientos adquiridos y los resultados obtenidos hasta el momento sientan las bases para futuros desarrollos y la eventual creación de un modelo personalizado que cumpla con los requisitos específicos de la tarea de alta y súper resolución de imágenes.

8. Recomendaciones para trabajos futuros

Existe una amplia variedad de desarrollos que podrían proceder de la base de este proyecto realizado. Las posibilidades se pueden orientar desde una configuración y generación de base de datos de entrada del modelo mucho más extensa y variada, disponer de muestras con características extras, como por ejemplo datos con información referente de tejido canceroso.

Desde el punto de vista de la configuración y creación del modelo neuronal, se plantean de igual manera múltiples opciones respecto al diseño de arquitectura y topología de la red. Cada

diseño cumple con un objetivo o enfoque particular y depende en gran medida del proyecto o aplicación.

Siguiendo con el modelo SRGAN es posible probar y experimentar con distintos modelos o configuraciones de las subredes. Examinar otro modelo generativo distinto al SRResNet y/o confrontar resultados con uno o varios modelos en el estado del arte.

Referencias Bibliográficas

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., & Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.
- Brown, R. W., Cheng, Y.-C. N., Mark Haacke, E., Thompson, M. R., & Venkatesan, R. (2014). *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Wiley-Blackwell.
- DeSantis, C., Bray, F., Ferlay, J., Lortet-Tieulent, J., O Anderson, B., & Jemal, A. (2015). International variation in female breast cancer incidence and mortality rates. *Cancer Epidemiology, Biomarkers Prevention*, 24(10):1495–1506.
- Gismondi Glave, G. (2010). Ingeniería biomédica. *Revista Ciencia y Cultura*, (24):99–118.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. C., & Bengio, Y. (2014). Generative adversarial nets. En *Advances in Neural Information*

- Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pp. 2672–2680.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. En *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778.
- Kingma, D. P. & Ba, J. (2015). Adam: A method for stochastic optimization. En Bengio, Y. & LeCun, Y., editores, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- Kossoff, G. (2000). Basic physics and imaging characteristics of ultrasound. *World Journal of Surgery*, 24(2):134–142.
- Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. En *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 105–114.
- Lillo, P. (2006). *Economía, política y otros juegos: una introducción a los juegos no cooperativos*. UNAM, Facultad de Ciencias.
- Lou, Y., Weimin, Z., Matthews, P. T., Catherine, M. A., & Mark, A. A. (2017). Generation of anatomically realistic numerical phantoms for photoacoustic and ultrasonic breast imaging. *Journal of Biomedical Optics*, 22(4):41015.

- MathWorks. (s.f.). *Redes generativas antagónicas (GAN)*. Recuperado el 20 de septiembre de 2023 de <https://la.mathworks.com/discovery/generative-adversarial-networks.html>.
- Oeffinger, K. C., Fontham, E. T. H., Etzioni, R., Herzig, A., Michaelson, J. S., Shih, Y.-C. T., Walter, L. C., Church, T. R., Flowers, C. R., LaMonte, S. J., Wolf, A. M. D., DeSantis, C., Lortet-Tieulent, J., Andrews, K., Manassaram-Baptiste, D., Saslow, D., Smith, R. A., Brawley, O. W., & Wender, R. (2015). Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. *JAMA*, 314(15):1599–1614.
- Ramirez, A. B., Abreo, S. A., & Van Dongen, K. (2022). Time versus frequency domain full waveform inversion for ultrasound imaging. En *2022 IEEE International Ultrasonics Symposium (IUS)*, pp. 1–4.
- Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training gans.
- Scheel, J. R., Lee, J. M., Sprague, B. L., Lee, C. I., & Lehman, C. D. (2014). Screening ultrasound as an adjunct to mammography in women with mammographically dense breasts. *American Journal of Obstetrics and Gynecology*. <https://doi.org/10.1016/j.ajog.2014.06.048>.
- Servick, K. (2014). Breast cancer: A world of differences. *Science*, 343(6178):1452–1453. <https://www.science.org/doi/epdf/10.1126/science.343.6178.1452>.
- Shorten, C. & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*. <https://doi.org/10.1186/s40537-019-0197-0>.

Simonyan, K. & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. pp. 1–14. Computational and Biological Learning Society.

Vargas, A., Amescua-Guerra, L. M., Bernal, M. A., & Pineda, C. (2008). Principios físicos básicos del ultrasonido, sonoanatomía del sistema musculoesquelético y artefactos ecográficos. *Acta Ortopédica Mexicana*, 22(6):361–373.

Vourtsis, A. & Berg, W. A. (2019). Breast density implications and supplemental screening. *European Radiology*, 29(4):1762–1777.