



Pontificia Universidad
JAVERIANA
Cali

Super-resolución en estudios de MRI mediante técnicas de aprendizaje profundo

Carlos Manuel Arcos Ramírez
Rafael Giovanni Ortega Solarte
Charles Erasmo Daza Malagón

Proyecto Aplicado para optar al título de
Magíster en Ciencia de Datos

Director:
Ing. Hernán Darío Vargas Cardona. Ph.D.

Facultad de Ingeniería y Ciencias
Maestría en Ciencia de Datos
Santiago de Cali, 9 de junio de 2025

Índice general

1. Definición del Problema	8
1.1. Planteamiento del Problema	8
1.1.1. Formulación del Problema	9
1.2. Objetivos	9
1.2.1. Objetivo General	9
1.2.2. Objetivos Específicos	9
1.3. Justificación	9
2. Desarrollo Teórico	12
2.1. Marco Teórico	12
2.1.1. Estudios MRI	12
2.1.2. Aprendizaje Automático	13
2.1.3. Super-resolución	14
2.1.4. Métricas de Evaluación	15
2.1.5. Modelos de Super-resolución	16
2.2. Antecedentes	26
3. Desarrollo Metodológico	29
3.1. Obtención del Conjunto de Datos	29
3.2. Exploración de Datos	29
3.3. Preprocesamiento de Imágenes MRI	30
3.3.1. Preprocesamiento de Imágenes MRI 3D	30
3.3.2. Preprocesamiento de Imágenes MRI 2D	31
3.4. Selección de Hiperparámetros	32
3.5. Implementación de Modelos	32
3.5.1. Modelos 3D	32
3.5.2. Modelos 2D	34
4. Resultados y Discusión	41
4.1. Modelos 3D	41
4.1.1. UNET 3D	41
4.1.2. EDSR 3D	43
4.1.3. cGAN 3D	44
4.1.4. SRCNN 3D	46
4.2. Modelos 2D	47
4.2.1. Autoencoder 2D	47
4.2.2. VDSR 2D	49
4.2.3. SRGAN 2D	50
4.2.4. DRCN 2D	51
4.2.5. UNET 2D	53
4.2.6. SRResNet 2D	54
4.2.7. SRDenseNet 2D	55
4.2.8. SR3 2D	57

4.3. Resumen Comparativo	60
4.3.1. Modelos 3D	60
4.3.2. Modelos 2D	60
5. Conclusiones	61
6. Trabajos Futuros	63
Bibliografía	63
Anexos	67

Índice de figuras

2.1. Arquitectura CNN de 7 capas para reconocimiento de caracteres. Imagen tomada de [16]. . .	14
2.2. Arquitectura original de U-Net. Imagen tomada de [18].	16
2.3. Arquitectura EDSR. Imagen tomada de [19].	17
2.4. Estructura básica de una cGAN. Imagen tomada de [20].	18
2.5. Arquitectura original de SRCNN. Imagen tomada de [21].	19
2.6. Estructura básica de una arquitectura Autoencoder. Imagen tomada de [22].	20
2.7. Estructura básica de una arquitectura VDSR. Imagen tomada de [23].	20
2.8. Estructura básica de una arquitectura SRGAN. Imagen tomada de [24].	21
2.9. Estructura básica de una arquitectura DRCN. Imagen tomada de [23].	22
2.10. (a) Bloque residual ResNet, (b) Bloque Residual SRResNet. Imagen tomada de [19].	22
2.11. Arquitectura SRResNet. Imagen tomada de [25].	23
2.12. Bloque de conexión densa de cinco capas. Imagen tomada de [27].	23
2.13. Arquitectura DenseNet profunda con tres bloques densos. Imagen tomada de [27].	24
2.14. Tres arquitecturas diferentes de SRDenseNet. (a) SRDenseNet H que únicamente utiliza características de alto nivel para la reconstrucción de imágenes HR. (b) SRDenseNet HL que utiliza características de alto y bajo nivel para la reconstrucción de imágenes HR. (c) SRDenseNet All que utiliza todos los niveles de características combinados con <i>skip connections</i> para la reconstrucción de imágenes HR. Imagen tomada de [28].	24
2.15. El proceso de difusión hacia adelante q (de izquierda a derecha) agrega gradualmente ruido gaussiano a la imagen de referencia. El proceso de inferencia inversa p (de derecha a izquierda) elimina de manera iterativa el ruido de la imagen de referencia. Imagen tomada de [29]. . . .	25
2.16. Refinamiento iterativo para la reconstrucción de la imagen a partir de ruido. Imagen tomada de [30].	26
3.1. Ejemplo de estudio MRI 3D de corte sagital.	30
3.2. Ejemplo 1 de MRI con resolución original y resolución reducida.	31
3.3. Ejemplo 2 de MRI con resolución original y resolución reducida.	32
3.4. Bloque convolucional en SR3.	38
3.5. Degradación progresiva de una imagen MRI.	40
4.1. Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo UNET 3D.	42
4.2. Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo UNET 3D.	42
4.3. Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo EDSR 3D.	43
4.4. Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo EDSR 3D.	43
4.5. Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo cGAN 3D.	45
4.6. Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo cGAN 3D.	45

4.7. Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo SRCNN 3D.	46
4.8. Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo SRCNN 3D.	46
4.9. Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo Autoencoder 2D.	47
4.10. Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo Autoencoder 2D.	48
4.11. Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo VDSR 2D.	49
4.12. Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo VDSR 2D.	49
4.13. Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo SRGAN 2D.	50
4.14. Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo SRGAN 2D.	51
4.15. Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo DRCN 2D.	52
4.16. Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo DRCN 2D.	52
4.17. Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo UNET 2D.	53
4.18. Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo UNET 2D.	53
4.19. Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo SRResNet 2D.	54
4.20. Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo SRResNet 2D.	55
4.21. Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo SRDenseNet 2D.	56
4.22. Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo SRDenseNet 2D.	56
4.23. Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo SR3 2D.	57
4.24. Pérdida de entrenamiento en las marcas temporales 500, 400, 300, 200, 100, 50, 30, 10 y 1 para el modelo SR3 2D.	58

Índice de cuadros

4.1. Análisis comparativo <i>Gold Standard</i> con estudios MRI de prueba para el modelo UNET 3D. .	42
4.2. Análisis comparativo <i>Gold Standard</i> con estudios MRI de prueba para el modelo EDSR 3D. .	44
4.3. Análisis comparativo <i>Gold Standard</i> con estudios MRI de prueba para el modelo cGAN 3D. .	45
4.4. Análisis comparativo <i>Gold Standard</i> con estudios MRI de prueba para el modelo SRCNN 3D.	47
4.5. Análisis comparativo <i>Gold Standard</i> con imágenes MRI de prueba para el modelo Autoencoder 2D.	48
4.6. Análisis comparativo <i>Gold Standard</i> con imágenes MRI de prueba para el modelo VDSR 2D.	50
4.7. Análisis comparativo <i>Gold Standard</i> con imágenes MRI de prueba para el modelo SRGAN 2D.	51
4.8. Análisis comparativo <i>Gold Standard</i> con imágenes MRI de prueba para el modelo DRCN 2D.	52
4.9. Análisis comparativo <i>Gold Standard</i> con imágenes MRI de prueba para el modelo UNET 2D.	54
4.10. Análisis comparativo <i>Gold Standard</i> con imágenes MRI de prueba para el modelo SRResNet 2D.	55
4.11. Análisis comparativo <i>Gold Standard</i> con imágenes MRI de prueba para el modelo SRDenseNet 2D.	57
4.12. Análisis comparativo <i>Gold Standard</i> con imágenes MRI de prueba para el modelo SR3 2D. .	58
4.13. Resumen Comparativo Modelos 3D.	60
4.14. Resumen Comparativo Modelos 2D.	60

Introducción

Las imágenes de resonancia magnética (MRI) son una herramienta fundamental en el diagnóstico y tratamiento de diversas patologías. A diferencia de otras técnicas de imágenes médicas, como la radiografía o la tomografía computarizada, las MRI no utilizan radiación ionizante, lo que las hace más seguras para los pacientes. Además, ofrecen una resolución espacial y un contraste de tejidos superiores, lo que permite obtener imágenes más detalladas y precisas de los órganos y tejidos del cuerpo.

Sin embargo, la calidad de las MRI puede verse comprometida por diversas limitaciones; como la duración de la exposición del paciente, que debe ser minimizada para evitar incomodidades y riesgos; las capacidades técnicas de los equipos, específicamente la fuerza del campo magnético; y factores humanos como la claustrofobia y los movimientos involuntarios durante el escaneo. Estas limitaciones pueden resultar en imágenes de baja resolución, afectando la precisión de los diagnósticos y tratamientos.

Como respuesta a estos desafíos, en la actualidad se han desarrollado múltiples técnicas computacionales basadas en aprendizaje profundo (*Deep Learning*) para mejorar la resolución de las MRI. Estas técnicas conocidas como super-resolución, utilizan algoritmos avanzados para aumentar la calidad de las imágenes, proporcionando más detalles de los tejidos y facilitando una detección más precisa de patologías. Este enfoque no solo reduce los costos asociados con la repetición de estudios, sino que también disminuye los riesgos para los pacientes, al minimizar la necesidad de exposiciones prolongadas.

En este proyecto se implementaron y evaluaron técnicas de super-resolución 2D y 3D basadas en aprendizaje profundo, con el objetivo de mejorar la resolución espacial de imágenes de resonancia magnética (MRI) tipo T1. Para ello, se gestionaron estudios anonimizados provenientes de bases de datos públicas y se entrenaron múltiples arquitecturas de redes neuronales convolucionales, incluyendo modelos deterministas, generativos y de difusión.

El proceso incluyó el preprocesamiento de los datos, la implementación de los modelos y su entrenamiento con imágenes emparejadas de baja y alta resolución. Posteriormente, se estableció un protocolo riguroso de evaluación cuantitativa, basado en métricas ampliamente aceptadas como PSNR, SSIM y MSE, con el fin de comparar el desempeño de las distintas arquitecturas utilizando imágenes de referencia (*Gold Standard*). Los resultados obtenidos permiten identificar las arquitecturas más eficaces tanto en dominios bidimensionales como tridimensionales, contribuyendo así al desarrollo de soluciones automáticas para la mejora de calidad en estudios médicos de resonancia magnética.

Capítulo 1

Definición del Problema

Los estudios de Imágenes de Resonancia Magnética (MRI, por sus siglas en inglés) constituyen una herramienta poderosa y ampliamente reconocida para el diagnóstico y tratamiento de diversas condiciones médicas. Utilizada en la evaluación de una gran variedad de objetivos morfológicos y funcionales, la MRI es fundamental para el estudio de procesos patológicos y fisiológicos complejos [1].

La MRI ofrece imágenes directas en planos sagital, coronal, oblicuo y axial, permitiendo visualizar estructuras corporales internas sin intervenciones invasivas. Una de sus principales ventajas es su alta resolución de contraste, que es la capacidad de distinguir tejidos blandos adyacentes con gran precisión [2].

1.1. Planteamiento del Problema

Debido a restricciones impuestas en protocolos clínicos, los estudios de MRI deben realizarse en periodos de tiempo limitados. Estas restricciones buscan proteger a los pacientes de la exposición prolongada a campos electromagnéticos, así como también se busca reducir el malestar asociado con la inmovilidad, el ruido y la claustrofobia [3]. Además, debe considerarse la eficiencia operativa, ya que en instalaciones de salud con alta demanda es primordial maximizar el uso de los equipos de MRI, lo que implica realizar sesiones lo más breves posible.

La duración del estudio es crítica porque podría limitar la cantidad de datos que se necesitan recopilar, afectando directamente la resolución espacial de las imágenes y, por lo tanto, el nivel de detalle que estas pueden ofrecer. Otros aspectos importantes que podrían afectar la resolución y la calidad de la imagen son el movimiento del paciente y señales externas o propias del equipo MRI [4]. Imágenes con una resolución espacial insuficiente pueden dificultar aplicaciones diagnósticas clave, como la segmentación de tejidos, la detección de patologías y el análisis volumétrico.

Existen diversas limitaciones técnicas y económicas que dificultan mejorar la resolución de las imágenes de MRI. Una mayor intensidad del campo magnético podría mejorar la resolución en las imágenes, pero técnicamente es muy complicado el desarrollo de máquinas de resonancia magnética que alcancen una intensidad de campo superior a 4 teslas. Esto se debe, principalmente, a las complicaciones asociadas con el manejo de las propiedades físicas de los materiales superconductores utilizados en la construcción de los imanes [5].

Otro aspecto para tener en cuenta es el económico. Máquinas sofisticadas con tecnologías avanzadas y mayor capacidad de intensidad de campo magnético, implican costos muy elevados en términos de adquisición, instalación y operación. Se estima que el gasto asociado a la adquisición e instalación de equipos MRI puede alcanzar aproximadamente un millón de dólares por tesla [5].

Existen formas de mejorar la resolución de imágenes MRI con técnicas de posprocesamiento, como la interpolación; sin embargo, estos pueden distorsionar la información en algún grado. En este contexto, han

surgido enfoques que ofrecen soluciones prometedoras, como los modelos de aprendizaje profundo (*Deep Learning*). Estos modelos, especialmente aquellos diseñados para tareas de super-resolución de imágenes, son capaces de aprender patrones complejos y extraer características significativas, lo que permite mejorar la resolución espacial de las imágenes de MRI sin introducir distorsiones significativas.

1.1.1. Formulación del Problema

Surge entonces la pregunta de investigación: ¿Cómo mejorar la resolución espacial en estudios MRI mediante métodos basados en aprendizaje profundo, de manera que los resultados puedan ser validados en comparación con *Gold Standard*? Por lo tanto, se formulan además los siguientes interrogantes: ¿Cómo deben gestionarse y tratarse los datos de las imágenes de MRI para lograr un acertado proceso de mejora de resolución?, ¿Qué algoritmos de *Deep Learning* deberían implementarse y evaluarse para alcanzar el objetivo de mejorar la resolución espacial?, ¿Cómo se puede validar el mejoramiento de la resolución espacial en estudios MRI?

1.2. Objetivos

1.2.1. Objetivo General

Implementar técnicas de super-resolución 2D y 3D basadas en aprendizaje profundo para mejorar la resolución espacial en estudios MRI que sean validadas con métricas de comparación con *Gold Standard*.

1.2.2. Objetivos Específicos

1. Gestionar estudios MRI anatómicos T1 a partir de bases de datos públicas y anonimizadas.
2. Implementar y entrenar diferentes algoritmos de super-resolución (SR) basados en aprendizaje profundo para mejorar la resolución espacial de estudios MRI.
3. Evaluar el desempeño de los modelos implementados mediante métricas establecidas en la literatura de comparación con *Gold Standard* e identificar los mejores métodos.

1.3. Justificación

En el contexto de imágenes por resonancia magnética, alcanzar una alta resolución espacial es crucial para permitir diagnósticos precisos y detallados por parte de los profesionales de la salud. Aunque esta calidad es posible con los equipos MRI modernos y los protocolos adecuados, muchos centros de salud, especialmente en países en desarrollo, enfrentan limitaciones significativas. Estos incluyen la falta de acceso a tecnología actualizada y la presencia de equipos obsoletos que pueden afectar negativamente la calidad de las imágenes obtenidas. Este enfoque es especialmente relevante debido a las limitaciones físicas inherentes a los equipos de MRI, como restricciones de hardware y artefactos, que pueden degradar la calidad de las imágenes obtenidas [6].

Es en este escenario donde las técnicas de super-resolución basadas en aprendizaje profundo emergen como una innovación clave para abordar estos desafíos. Mejorar la resolución espacial no solo facilita la detección temprana y el seguimiento de enfermedades neurodegenerativas como el Parkinson y diversos tipos de cáncer [7], sino que también impulsa avances en la investigación médica al proporcionar imágenes más claras y detalladas de estructuras anatómicas y microestructuras cruciales. Además, el uso de estas técnicas podría permitir que los escáneres MRI convencionales ofrezcan imágenes de calidad superior sin necesidad de grandes inversiones adicionales. La implementación de estas técnicas tiene el potencial de mejorar significativamente la calidad de la atención médica en entornos con recursos restringidos.

La viabilidad del proyecto está relacionada con varios factores clave. La existencia y acceso gratuito a bases de datos públicas y anonimizadas de estudios MRI constituye un punto crucial, ya que la obtención de

información de calidad es una de las etapas que más tiempo consume en este tipo de proyectos. Ejemplos de bases de datos confiables son la ADNI (*Alzheimer's Disease Neuroimaging Initiative*) y OASIS (*Open Access Series of Imaging Studies*) de la Washington University School of Medicine in St. Louis [8] [9]. Estas bases de datos han sido ampliamente utilizadas y validadas por la comunidad científica.

Para garantizar una cantidad adecuada de datos para el entrenamiento de los modelos, se planea trabajar con al menos cinco estudios completos de resonancia magnética. Esta cantidad se considera suficiente debido a que el entrenamiento se realiza a nivel de vóxeles, permitiendo obtener una gran cantidad de datos de entrenamiento a partir de cada estudio individual. Un estudio de resonancia magnética típicamente contiene millones de vóxeles. Al trabajar con múltiples estudios, se asegurará una diversidad de datos que es esencial para entrenar modelos robustos y generalizables.

Los avances recientes en técnicas de aprendizaje profundo, como las redes neuronales convolucionales y las redes adversarias generativas, han demostrado ser altamente efectivas en la mejora de la resolución de imágenes médicas [10], facilitando la implementación de estos métodos mediante bibliotecas y frameworks de código abierto. La disponibilidad de recursos computacionales avanzados, tanto físicos como remotos a través de diversos servicios en la nube, permite un entrenamiento eficiente de los modelos que requieren un alto nivel de procesamiento. Además, la dirección del profesor Hernán Darío Vargas Cardona, PhD en Ingeniería, quien cuenta con amplia experiencia en modelos empleados para la mejora de la resolución espacial de imágenes por resonancia magnética será fundamental.

Desde el punto de vista económico, el proyecto es muy viable. No se requiere una gran inversión inicial gracias al acceso a datos públicos y gratuitos, herramientas y bibliotecas de código abierto, y la posibilidad de utilizar servicios en la nube en caso de que el poder de cómputo físico disponible no sea suficiente. Este enfoque minimiza los costos de adquisición de datos y software, y reduce la necesidad de infraestructura de hardware costosa. Además, los servicios en la nube ofrecen modelos de pago por uso, lo que permite gestionar los costos de manera eficiente y escalar los recursos según las necesidades del proyecto.

El desarrollo de este proyecto tiene impactos importantes. En cuanto a la esfera social, la mejora en la calidad de las imágenes MRI puede llevar a diagnósticos más precisos y tempranos de diversas enfermedades, lo que permite emplear tratamientos más efectivos y mejorar la calidad de vida de los pacientes. La detección temprana ha demostrado ser crucial en la gestión efectiva de enfermedades crónicas, como el cáncer, asegurando intervenciones terapéuticas oportunas que mejoran los resultados de salud a largo plazo [11]. Además, se ha destacado la importancia de la detección precoz en el contexto de enfermedades cardiovasculares, subrayando cómo el diagnóstico temprano puede prevenir complicaciones severas y reducir la mortalidad entre los pacientes [12]. También se beneficia significativamente al personal de salud, al mejorar la atención al paciente y reducir la incertidumbre en los diagnósticos. Asimismo, al disminuir la necesidad de repetir estudios, se reduce la carga sobre los sistemas de salud pública, permitiendo beneficiar a un mayor número de pacientes en un menor periodo de tiempo.

Los impactos ambientales y económicos están estrechamente ligados en este proyecto. Al reducir el consumo energético de los equipos MRI mediante la mejora de la calidad de las imágenes, no solo se disminuye la huella de carbono y los costos operativos relacionados con la energía, sino que también se promueve una gestión más eficiente de los recursos dentro de las instalaciones médicas. Esto puede reducir los gastos de funcionamiento, mejorar la sostenibilidad financiera de los centros de salud y permitir una asignación más efectiva de recursos a otras áreas de atención médica y desarrollo tecnológico.

Finalmente, el proyecto tiene impactos significativos en el avance de la imagenología médica mediante técnicas de super-resolución basadas en aprendizaje profundo. Esto incluye la implementación de algoritmos avanzados de inteligencia artificial en contextos clínicos, promoviendo el desarrollo teórico y metodológico en el procesamiento de imágenes médicas y facilitando la colaboración interdisciplinaria entre la ciencia de datos, la ingeniería biomédica y la medicina. Este enfoque promete generar nuevas investigaciones y publicaciones académicas que amplíen la comprensión de cómo las técnicas de aprendizaje profundo pueden mejorar la calidad y precisión de las imágenes de resonancia magnética, contribuyendo al avance científico y clínico en el

diagnóstico y tratamiento de enfermedades. Además, para nosotros como estudiantes de Maestría en Ciencia de Datos, este proyecto representa una oportunidad única para aplicar y expandir nuestros conocimientos en medicina computacional, un campo emergente de gran impacto.

Capítulo 2

Desarrollo Teórico

2.1. Marco Teórico

En esta sección se revisarán los fundamentos teóricos de los modelos de super-resolución basados en aprendizaje profundo. Asimismo, se analizarán aplicaciones específicas de dichos modelos en el contexto de estudios de imágenes de resonancia magnética cerebral.

2.1.1. Estudios MRI

La imagen de resonancia magnética (MRI, por sus siglas en inglés) es una técnica de imagen médica que utiliza un campo magnético y ondas de radio para obtener imágenes de alta resolución de los órganos y tejidos internos del cuerpo humano. La MRI se emplea tanto para el diagnóstico como para el seguimiento de diversas patologías, incluyendo neoplasias, esclerosis múltiple, accidentes cerebrovasculares (ictus), aneurismas e infecciones. Además, puede utilizarse para el estudio de la estructura y función del cerebro, el corazón, las articulaciones y los vasos sanguíneos.

Las imágenes por MRI pueden adquirirse mediante distintas modalidades, dependiendo de los parámetros que se deseen resaltar o potenciar. Las modalidades más comunes son [13]:

- **Imágenes potenciadas en densidad protónica (D):** Aprovechan la difusión molecular del agua en los tejidos biológicos para obtener información sobre su microestructura. La difusión del agua es sensible a las barreras y restricciones que encuentra en su camino, lo que permite inferir detalles sobre la organización interna de los tejidos.
- **Imágenes potenciadas en T1:** Se basan en el tiempo de relajación longitudinal de los protones de hidrógeno en los tejidos. El tiempo de relajación longitudinal es el tiempo que tardan los núcleos en recuperar su magnetización longitudinal después de un pulso de radiofrecuencia. Las imágenes T1 ofrecen un alto contraste entre tejidos con diferentes densidades de protones, lo que permite diferenciar estructuras como la materia gris, la materia blanca o el líquido cefalorraquídeo. En este tipo de imágenes, la materia gris se ve más brillante y la blanca más oscura.
- **Imágenes potenciadas en T2:** Utilizan el fenómeno de la relajación transversal de los protones de hidrógeno en los tejidos para obtener información sobre su estructura y composición. El tiempo de relajación transversal (T2) es la medida del tiempo que tardan los protones en perder su magnetización después de un pulso de radiofrecuencia. Las imágenes T2 son sensibles a las diferencias en el entorno molecular de los tejidos, lo que permite obtener imágenes con alto contraste entre tejidos con diferentes propiedades de relajación T2. Por ejemplo, la materia gris y la materia blanca del cerebro tienen tiempos de relajación T2 diferentes: la materia gris se muestra de un tono intermedio debido a su contenido de agua, mientras que la materia blanca aparece brillante debido a la mayor movilidad de los protones de mielina.

2.1.2. Aprendizaje Automático

Aprendizaje Profundo

Durante varias décadas, las técnicas convencionales de aprendizaje automático presentaron limitaciones para procesar datos en su forma cruda. Por esta razón, los sistemas de reconocimiento de patrones requerían un exhaustivo desarrollo de extractores de características que transformaran los datos crudos en representaciones internas adecuadas para su procesamiento. Con la introducción de los métodos de aprendizaje de representaciones, se logró que un sistema interprete automáticamente las representaciones necesarias para la detección o clasificación a partir de los datos crudos de entrada [14].

El enfoque basado en el aprendizaje de representaciones dio origen al aprendizaje profundo, el cual utiliza una cascada de múltiples capas de unidades de procesamiento no lineal para la extracción y transformación progresiva de características [15]. Cada capa, o nivel de representación, transforma la salida de la capa anterior en una representación más abstracta y compleja. Las capas inferiores, que reciben directamente los datos crudos, aprenden características simples, mientras que los niveles superiores construyen características de mayor complejidad a partir de las características de las capas inferiores, estructurando la composición de una arquitectura multicapa de representación de características jerárquica útil para analizar grandes cantidades de datos de diferentes fuentes [15].

En el caso práctico de una imagen, donde los datos se representan como una matriz de valores de píxeles, las características aprendidas en la primera capa de representación generalmente representan la presencia o ausencia de bordes en orientaciones y ubicaciones. La segunda capa, a su vez, detecta patrones identificando características particulares en los bordes, sin importar que existan variaciones en la posición. La tercera capa puede combinar patrones que corresponden a algunas partes de objetos similares, mientras que las capas posteriores integran dichas partes para reconocer objetos completos. El aspecto clave del aprendizaje profundo es que las capas de extracción de características aprenden a partir de datos utilizando un procedimiento de aprendizaje genérico [14].

Aprendizaje Supervisado

El entrenamiento de modelos de aprendizaje supervisado consiste en utilizar datos etiquetados para entrenar la red por medio de un procedimiento de retroalimentación [16].

Durante el proceso de entrenamiento se calcula la función objetivo, o función de pérdida, que cuantifica el error (o distancia) entre las predicciones del modelo y los valores reales esperados. A partir de esta medida, el modelo ajusta sus parámetros internos (pesos) con el objetivo de minimizar el error de predicción. Estos parámetros ajustables son los que determinan la función que transforma las entradas en salidas dentro del modelo [14].

Retropropagación

El modelo de aprendizaje profundo calcula un vector gradiente que indica, para cada parámetro ajustable, en qué cantidad aumentaría o disminuiría el error si el peso se incrementa en una pequeña cantidad. Esta información permite ajustar el vector de pesos en la dirección que minimiza el error. En el proceso de retropropagación, se calcula la derivada (o gradiente) de la función de pérdida con respecto a las entradas de un módulo de procesamiento, partiendo desde el gradiente asociado a su salida (o a la entrada del siguiente módulo). Este cálculo se aplica de manera iterativa desde la capa de salida (la predicción) hasta la capa de entrada, propagando así los gradientes a través de todos los módulos de la red [14].

Redes Neuronales Convolucionales

Son una arquitectura especializada en el procesamiento de datos con estructura espacial. Están compuestas por una secuencia de capas: capas convolucionales, capas de agrupamiento (*pooling*) y capas completamente conectadas (*fully connected*).

Las primeras capas de una red convolucional suelen ser capas convolucionales y capas de agrupamiento. Las unidades de las capas convolucionales están organizadas en mapas de características (*feature maps*), que aplican una serie de filtros a los datos de entrada para extraer características relevantes y detectar patrones locales. Cada unidad de estas capas está conectada a regiones locales de la capa anterior mediante un conjunto de pesos conocidos como banco de filtros. El resultado de la operación de convolución se somete a una función de activación no lineal, comúnmente la ReLU (*Rectified Linear Unit*), que introduce no linealidad al modelo.

Después de una capa de convolución suele incorporarse una capa de agrupamiento (*pooling*), cuya función principal es reducir la dimensionalidad espacial de los mapas de características y mantienen la información más relevante [14]. Las capas finales de la red son las capas completamente conectadas (*fully connected layers*), que se encargan de tomar las características extraídas por las capas anteriores y de combinarlas para realizar la tarea de clasificación o regresión, según el objetivo del modelo [16].

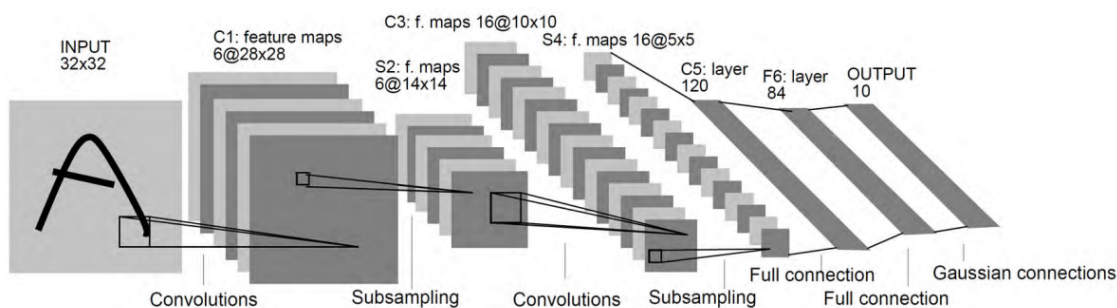


Figura 2.1: Arquitectura CNN de 7 capas para reconocimiento de caracteres. Imagen tomada de [16].

2.1.3. Super-resolución

La super-resolución es una técnica consistente en aumentar la resolución de una imagen a partir de una o varias imágenes de baja resolución. En el primero de los casos, se utiliza una sola imagen de baja resolución como entrada y genera una imagen de alta resolución como salida, mediante el uso de redes generativas o convolucionales. Para el caso de multi-imágenes, se utilizan varias imágenes de baja resolución, capturadas desde diferentes ángulos o con diferentes niveles de *zoom*, como entrada y fusionan la información de todas ellas para generar una imagen de alta resolución, mediante el uso de redes recurrentes o de atención.

Ambos tipos de métodos tienen sus ventajas y desventajas. Los métodos basados en una sola imagen son más simples y rápidos, pero pueden generar imágenes menos detalladas o más borrosas. Los métodos basados en múltiples imágenes son más complejos y lentos, pero pueden generar imágenes más nítidas y realistas [13].

Las técnicas de super-resolución se pueden aplicar a las MRI con diferentes ponderaciones, como D (densidad de protones), T1 y T2, aprovechando las características específicas de cada una para obtener mejores resultados.

- **Ponderación D (densidad de protones):** El objetivo es mejorar la visualización de la anatomía general y la diferenciación entre tejidos con diferentes densidades de protones. Las técnicas basadas en la interpolación espacial, como el vecino más cercano o la interpolación bicúbica, pueden ser efectivas para mejorar la resolución espacial de las imágenes D. La superresolución de imágenes D puede ser útil para visualizar mejor la estructura ósea, los músculos y la grasa.
- **Ponderación T1:** El objetivo es mejorar la visualización de la materia gris y la materia blanca del cerebro, así como de otras estructuras con diferentes tiempos de relajación longitudinal (T1). Las técnicas basadas en el aprendizaje automático, como las redes neuronales convolucionales (CNN), pueden ser más adecuadas para aprovechar las características complejas de las imágenes T1 y obtener

una superresolución más precisa. La superresolución de imágenes T1 puede ser útil para detectar anomalías en la materia gris o la materia blanca del cerebro, como tumores o esclerosis múltiple [13].

- **Ponderación T2:** El objetivo es mejorar la visualización de los detalles finos en los tejidos, como la microestructura del tejido cerebral o la morfología de los tumores. Las técnicas basadas en la combinación de diferentes métodos de superresolución, como la interpolación espacial y el aprendizaje automático, pueden ser más efectivas para capturar la complejidad de las imágenes T2. La superresolución de imágenes T2 puede ser útil para caracterizar mejor los tumores y otras lesiones, o para estudiar la microestructura del tejido cerebral en enfermedades neurodegenerativas.

2.1.4. Métricas de Evaluación

Mean Squared Error (MSE)

Es una de las métricas más utilizadas para evaluar la calidad de la reconstrucción de una imagen. El MSE se calcula promediando las diferencias al cuadrado píxel a píxel entre la imagen generada y la correspondiente imagen de referencia. El MSE entre dos imágenes se define de la siguiente manera [17]:

$$MSE = \frac{1}{WH} \sum_{i=1}^W \sum_{j=1}^H (\hat{x}_r(i, j) - x_g(i, j))^2 \quad (2.1)$$

Donde W indica el ancho de la imagen y H el alto de la imagen. (i, j) define la posición del píxel, mientras que \hat{x}_r representa la imagen de referencia y x_g la imagen generada.

Peak Signal-to-Noise Ratio (PSNR)

Es una métrica comúnmente usada para medir la calidad de reconstrucción y es inversamente proporcional al logaritmo del error cuadrático medio (MSE) entre la imagen de referencia y la imagen generada. El PSNR se expresa como [17]:

$$PSNR = 20 \cdot \log_{10} \left(\frac{MAX_I}{RMSE(\hat{x}_r, x_g)} \right) \quad (2.2)$$

Donde MAX corresponde al máximo valor de píxel posible. De manera general, un valor alto de PSNR indica una mejor calidad en la reconstrucción de la imagen, lo que implica una gran similitud entre ambas imágenes.

Structural Similarity Index Measure (SSIM)

Es una métrica diseñada para evaluar la calidad de la reconstrucción de una imagen basándose en una perspectiva perceptual. A diferencia del MSE y el PSNR que estiman errores absolutos, el SSIM proporciona errores basados en la percepción y la saliencia, teniendo en cuenta características como la luminancia, el contraste y los cambios en la información estructural. La idea clave al considerar los cambios en la información estructural es que los píxeles están fuertemente correlacionados, especialmente cuando están espacialmente cerca. El SSIM se puede definir de la siguiente manera [17]:

$$SSIM = \frac{(2\mu_{\hat{x}_r}\mu_{x_g} + c_1)(2\sigma_{\hat{x}_r x_g} + c_2)}{(\mu_{\hat{x}_r}^2 + \mu_{x_g}^2 + c_1)(\sigma_{\hat{x}_r}^2 + \sigma_{x_g}^2 + c_2)} \quad (2.3)$$

Donde $\mu_{\hat{x}_r}$ y μ_{x_g} representan las medias de la imagen original y la imagen reconstruida, respectivamente. En consecuencia, $\sigma_{\hat{x}_r}$ y σ_{x_g} son la desviación estándar de \hat{x}_r y x_g . Por lo tanto, $\sigma_{\hat{x}_r x_g}$ denota la covarianza entre ambas imágenes, mientras c_1 y c_2 son constantes establecidas para evitar inestabilidad [17].

2.1.5. Modelos de Super-resolución

UNET

La U-Net es una arquitectura de redes neuronales convolucionales propuesta por Ronneberger et al. [18], diseñada originalmente para la segmentación de imágenes biomédicas. Su nombre proviene de la forma en U del grafo computacional, que combina un camino de contracción (encoder) y un camino de expansión (decoder). La arquitectura se compone de dos bloques principales:

- **Encoder (contracción):** Consiste en varias capas de convolución (3×3) seguidas de operaciones de *max pooling* (2×2). En cada etapa se reduce la resolución espacial de la imagen mientras se incrementa el número de filtros, lo que permite capturar características contextuales de mayor nivel.
- **Decoder (expansión):** Emplea operaciones de upsampling (mediante convoluciones transpuestas) para recuperar la resolución espacial original. En cada etapa del decoder, se concatenan las características correspondientes del encoder a través de conexiones de salto (skip connections), lo que permite preservar detalles espaciales finos esenciales para la tarea de reconstrucción.

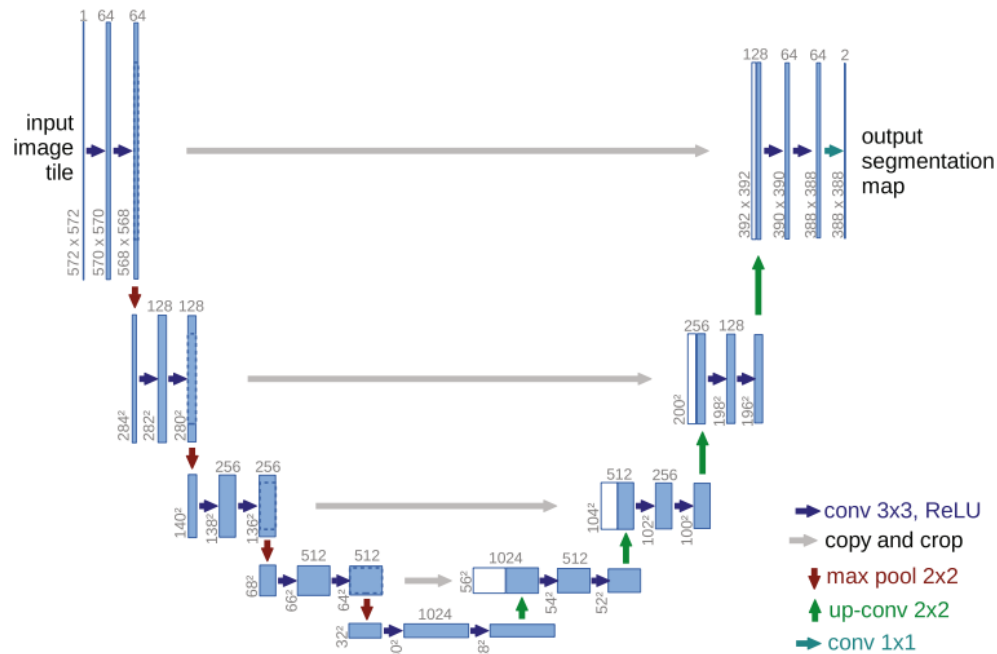


Figura 2.2: Arquitectura original de U-Net. Imagen tomada de [18].

Aunque la U-Net fue concebida inicialmente para tareas de segmentación semántica en imágenes biomédicas, su arquitectura basada en una estructura simétrica encoder-decoder con conexiones de salto le permite preservar y reconstruir información espacial detallada. Estas características la hacen especialmente útil en tareas de super-resolución, donde es fundamental recuperar estructuras finas y coherentes a partir de entradas degradadas.

Para el problema de super-resolución de imágenes médicas, la U-Net puede adaptarse tanto en su versión 2D como en su variante 3D. En la versión 2D, opera sobre cortes individuales de imágenes, mientras que en su versión 3D extiende las operaciones convolucionales a tres dimensiones, permitiendo procesar directamente volúmenes completos como estudios de resonancia magnética (MRI), mejorando la calidad espacial en los tres ejes.

EDSR

La arquitectura *Enhanced Deep Super-Resolution Network (EDSR)* representa un avance significativo frente a modelos previos de super-resolución basados en redes convolucionales profundas. Su objetivo principal es mejorar la calidad de las imágenes generadas mediante una arquitectura más profunda y precisa [19].

Una de sus principales innovaciones consiste en eliminar las capas de normalización por lotes (*batch normalization*). Aunque estas capas son útiles para estabilizar el entrenamiento en otras tareas, en super-resolución pueden limitar la capacidad de la red para modelar relaciones precisas entre los píxeles. Además, EDSR introduce bloques residuales simplificados, lo que permite construir redes más profundas sin degradar el rendimiento.

La arquitectura general de EDSR se compone de una capa convolucional inicial, seguida por múltiples bloques residuales sin *batch normalization*, y finaliza con una etapa de reconstrucción, la cual incluye una operación de *upsampling* para generar la imagen de alta resolución.

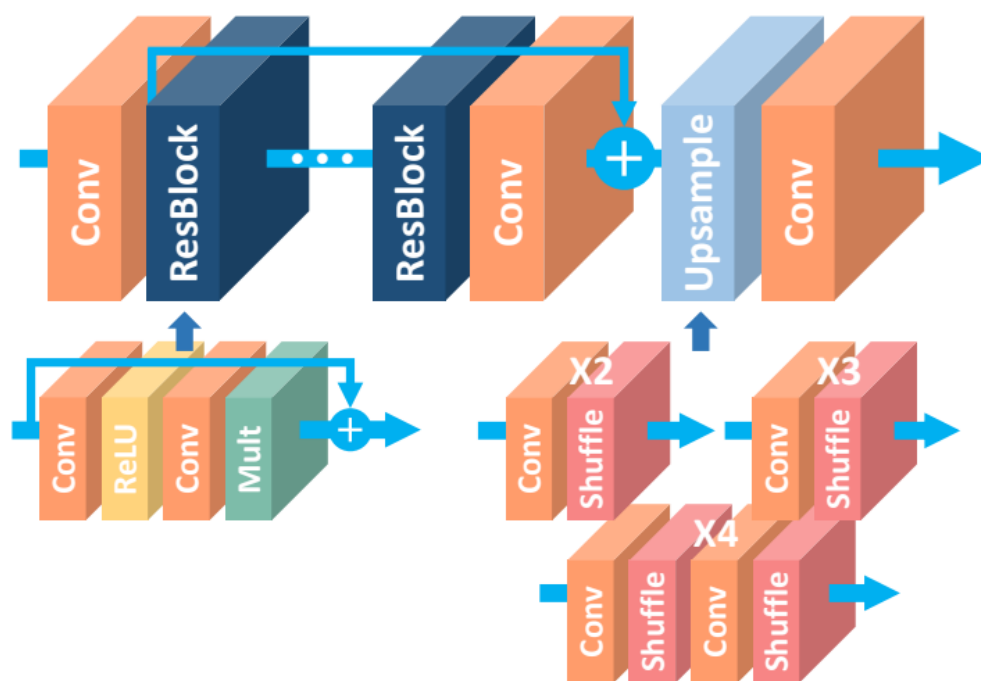


Figura 2.3: Arquitectura EDSR. Imagen tomada de [19].

Aunque fue diseñada originalmente para trabajar con imágenes 2D, su estructura modular permite una adaptación sencilla a contextos tridimensionales mediante el reemplazo de operaciones 2D por sus equivalentes 3D. Esta capacidad de adaptación la convierte en una alternativa adecuada para tareas de super-resolución en imágenes médicas volumétricas, como los estudios de resonancia magnética.

cGAN

Las *Conditional Generative Adversarial Nets (cGAN)* son una extensión de las GAN tradicionales, en las que tanto el generador como el discriminador reciben información adicional como condición [20]. En el contexto de super-resolución, esta condición suele ser la versión de baja resolución de la imagen que se desea mejorar. Estas redes aprenden a generar imágenes de alta resolución coherentes con su contraparte de baja resolución, lo que permite preservar detalles estructurales importantes.

La arquitectura general de una cGAN enfocada en super-resolución se compone de dos redes principales: el generador, que intenta convertir una imagen de baja resolución en una versión de alta resolución; y el discriminador, que aprende a distinguir entre imágenes reales de alta resolución y las generadas sintéticamente. La Figura 2.4 ilustra la estructura básica de una red condicional adversarial, donde tanto el generador como el discriminador reciben información condicional (por ejemplo, la imagen de baja resolución) para guiar el proceso de generación y evaluación.

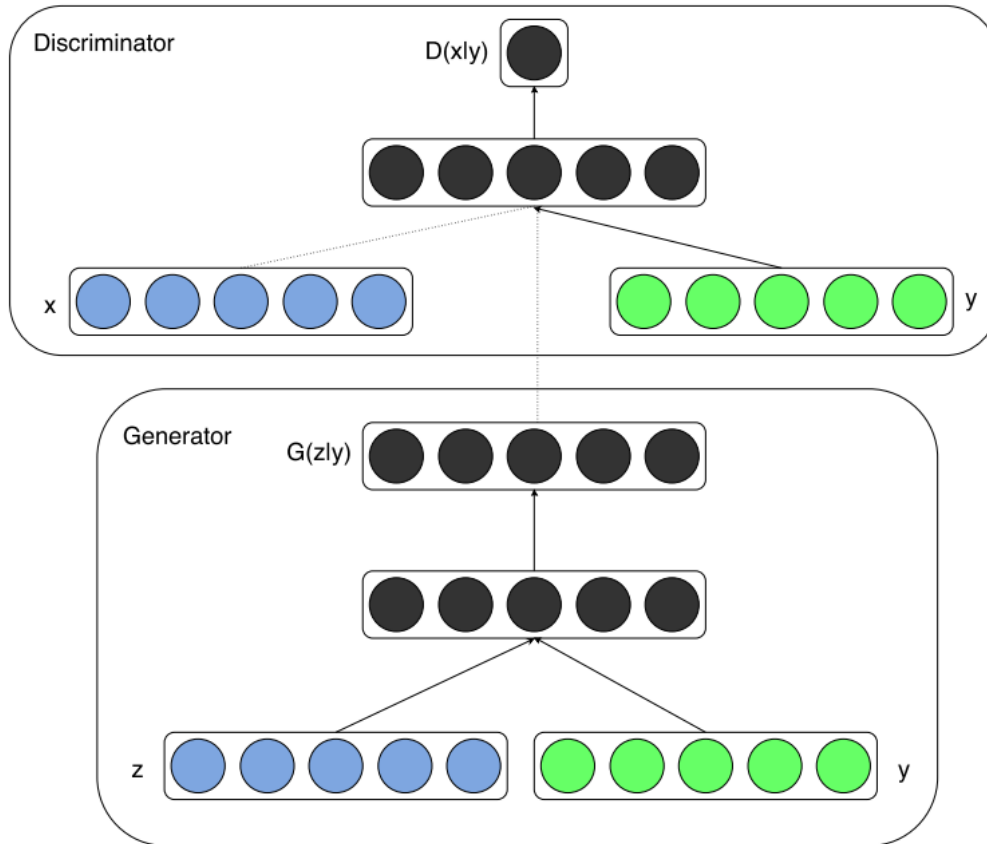


Figura 2.4: Estructura básica de una cGAN. Imagen tomada de [20].

Gracias a su flexibilidad, este tipo de arquitectura puede adaptarse fácilmente tanto a imágenes bidimensionales como a volúmenes tridimensionales, mediante el uso de capas convolucionales 2D o 3D según el caso, lo que la hace especialmente adecuada para tareas de super-resolución en imágenes médicas como las resonancias magnéticas cerebrales.

SRCNN

La *Super-Resolution Convolutional Neural Network (SRCNN)* fue pionera en el uso de redes neuronales profundas aplicadas exitosamente a la super-resolución de imágenes. Esta arquitectura marcó el inicio del uso de técnicas de aprendizaje profundo para incrementar la resolución espacial de imágenes de forma precisa. Su funcionamiento se basa en aprender una relación directa entre imágenes de baja y alta resolución a partir de datos de entrenamiento, eliminando la necesidad de métodos clásicos de interpolación [21]. La red original consta de tres capas convolucionales secuenciales:

- *Extracción de características:* Se aplica una convolución con un kernel de gran tamaño (por ejemplo, 9×9) para extraer representaciones de la imagen de baja resolución.

- *Mapeo no lineal*: Esta capa transforma las características extraídas hacia un espacio de representación de mayor dimensión, utilizando kernels más pequeños (como 1×1 o 5×5).
- *Reconstrucción*: Se emplea una tercera convolución que sintetiza la imagen de alta resolución a partir de las características transformadas.

Este diseño permite que la red aprenda de manera efectiva cómo refinar detalles y estructuras presentes en la imagen degradada. La Figura 2.5 muestra el flujo de procesamiento de la arquitectura original de SRCNN, incluyendo las etapas de extracción, mapeo no lineal y reconstrucción.

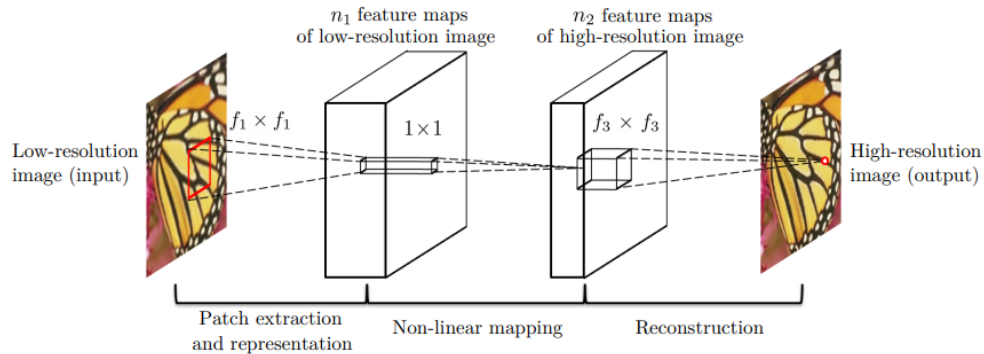


Figura 2.5: Arquitectura original de SRCNN. Imagen tomada de [21].

Si bien la arquitectura SRCNN fue diseñada originalmente para imágenes 2D, su estructura simple y modular ha facilitado su extensión al dominio tridimensional mediante el uso de convoluciones 3D, lo que permite abordar tareas de super-resolución en estudios médicos volumétricos, como los de resonancia magnética (MRI), donde es necesario preservar la coherencia espacial en los tres ejes.

Autoencoder

Un autoencoder es un tipo de red neuronal artificial utilizada para el aprendizaje no supervisado, cuyo objetivo principal es aprender codificaciones eficientes de los datos sin necesidad de etiquetas. Funciona codificando la entrada en una representación comprimida y luego decodiéndola para reconstruir la entrada original. Este proceso obliga a la red a extraer características útiles y relevantes de los datos, facilitando así tareas como la reducción de dimensionalidad y la detección de anomalías [22].

La arquitectura básica de un autoencoder consta de dos partes: el codificador, que transforma los datos de entrada en una representación de menor dimensión llamada espacio latente o cuello de botella, y el decodificador, que intenta reconstruir los datos originales a partir de esta representación. Entre sus ventajas destacan la capacidad para aprender automáticamente características relevantes, manejar relaciones no lineales y detectar anomalías. Sin embargo, también presenta desventajas como el riesgo de sobreajuste, la complejidad en algunas arquitecturas, la dependencia de la calidad y cantidad de datos de entrenamiento, la falta de interpretabilidad de las características aprendidas y una generalización limitada a datos muy diferentes del conjunto original [22].

VDSR

VDSR (Red Convolutiva Profundamente Recursiva para Super-Resolución) es una red neuronal convolutiva profunda diseñada específicamente para la super-resolución de una sola imagen (SISR). Su objetivo principal es aprender la correspondencia entre imágenes de baja resolución y sus versiones de alta resolución, utilizando una arquitectura muy profunda compuesta por pequeños filtros convolucionales, típicamente de tamaño 3×3 . Esta red aprende directamente el residuo entre la imagen de baja resolución, previamente aumentada al tamaño deseado, y la imagen de alta resolución objetivo, lo que permite enfocarse en los detalles de alta frecuencia necesarios para mejorar la calidad visual [23].

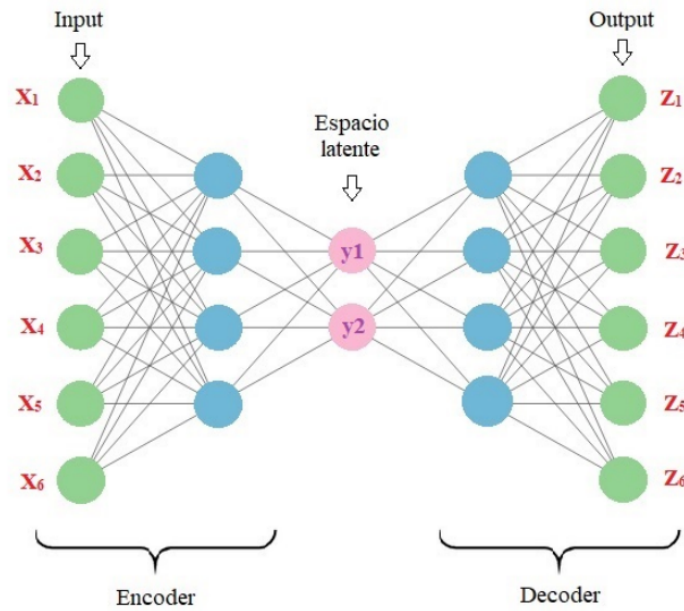


Figura 2.6: Estructura básica de una arquitectura Autoencoder. Imagen tomada de [22].

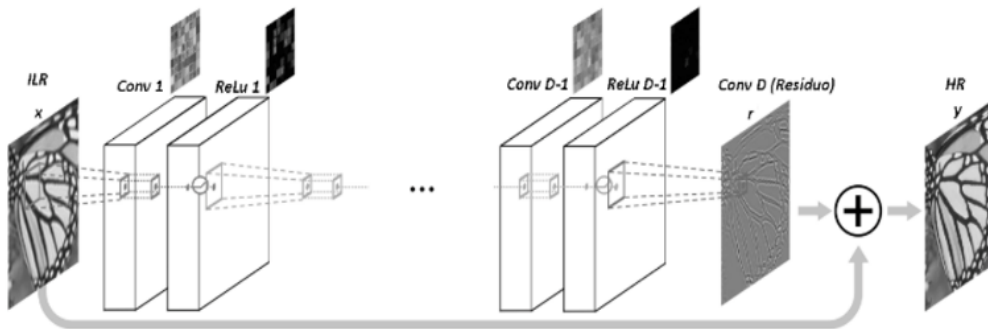


Figura 2.7: Estructura básica de una arquitectura VDSR. Imagen tomada de [23].

La arquitectura de VDSR suele tener alrededor de 20 capas y emplea técnicas como el aprendizaje residual y el recorte de gradiente para estabilizar el entrenamiento en redes tan profundas. Además, es capaz de manejar múltiples factores de aumento de escala (por ejemplo, 2x, 3x, 4x) dentro de un solo modelo, lo que la hace versátil para diferentes aplicaciones. Entre sus ventajas destacan el aprendizaje efectivo del residuo, la capacidad para extraer características complejas gracias a su profundidad, y la super-resolución multiescala. Sin embargo, también presenta desventajas como un alto costo computacional durante el entrenamiento y la inferencia, y una fuerte dependencia de grandes y diversos conjuntos de datos para lograr un buen rendimiento [23].

SRGAN

Las *Super-Resolution Generative Adversarial Networks* están compuestas por dos redes principales que se entrenan de manera antagónica: el generador y el discriminador. La red generadora toma una imagen de baja resolución como entrada y busca producir una versión de alta resolución que sea visualmente similar

a las imágenes reales y que además conserve una buena reconstrucción de la entrada original. Esta red generalmente utiliza una arquitectura convolucional profunda con bloques residuales. Por otro lado, la red discriminadora recibe imágenes que pueden ser generadas o reales y su objetivo es distinguir correctamente entre ambas. Durante el entrenamiento, el generador intenta engañar al discriminador, mientras que este se entrena para mejorar su capacidad de clasificación, lo que resulta en la generación de imágenes con super-resolución cada vez más realistas [24].

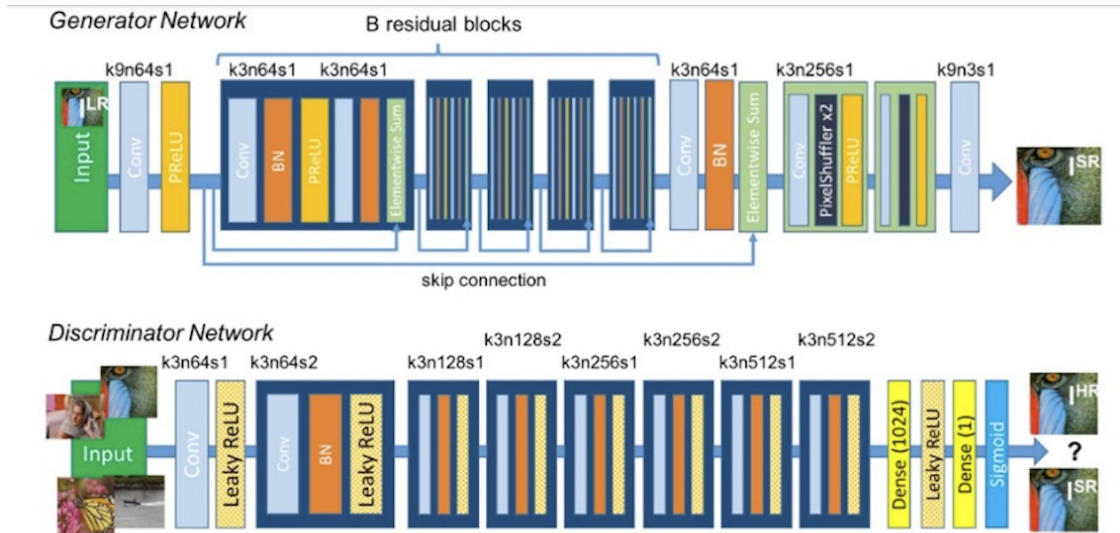


Figura 2.8: Estructura básica de una arquitectura SRGAN. Imagen tomada de [24].

Entre las ventajas de las SRGAN destacan su capacidad para generar imágenes con detalles realistas y una calidad perceptiva superior, gracias a la combinación de la pérdida antagónica con la pérdida de contenido. Esto hace que los resultados sean preferidos por observadores humanos en comparación con métodos que solo optimizan la precisión a nivel de píxel. Además, el marco GAN permite que el modelo genere detalles plausibles que podrían faltar en la imagen original de baja resolución, y el aprendizaje se realiza de extremo a extremo, facilitando la generalización a datos no vistos. Sin embargo, el entrenamiento puede ser inestable y costoso computacionalmente debido a la estructura de doble red y el proceso antagónico. También existe el riesgo de que el modelo introduzca artefactos o detalles irreales en las imágenes generadas [24].

DRCN

DRCN (Red Convolutiva Profundamente Recursiva) es una arquitectura diseñada para la super-resolución de una sola imagen que enfatiza el uso de la recursión para aumentar eficazmente la profundidad de la red sin incrementar significativamente el número de parámetros. Esto permite a la red aprender características más complejas necesarias para mejorar la calidad de las imágenes de alta resolución. La estructura básica de DRCN incluye una red de incrustación que extrae características de la imagen de baja resolución, seguida de múltiples bloques recursivos donde las capas convolucionales dentro de cada bloque comparten los mismos pesos, lo que mantiene el modelo manejable en términos de parámetros. Finalmente, una red de reconstrucción genera la imagen de alta resolución a partir de las características aprendidas, y a menudo se utilizan varias redes de reconstrucción cuyas salidas se combinan para producir el resultado final [23].

Entre las ventajas de DRCN destaca su capacidad para lograr una gran profundidad efectiva con un número reducido de parámetros, lo que mejora el rendimiento de super-resolución en comparación con redes menos profundas. Además, es especialmente eficaz para capturar estructuras finas en las imágenes. Sin embargo, esta arquitectura presenta ciertas desventajas, como la mayor complejidad en su comprensión e implementación debido a la naturaleza recursiva, así como los desafíos asociados al entrenamiento de redes muy profundas y recursivas, que pueden afectar la optimización y la convergencia del modelo. Los diagramas

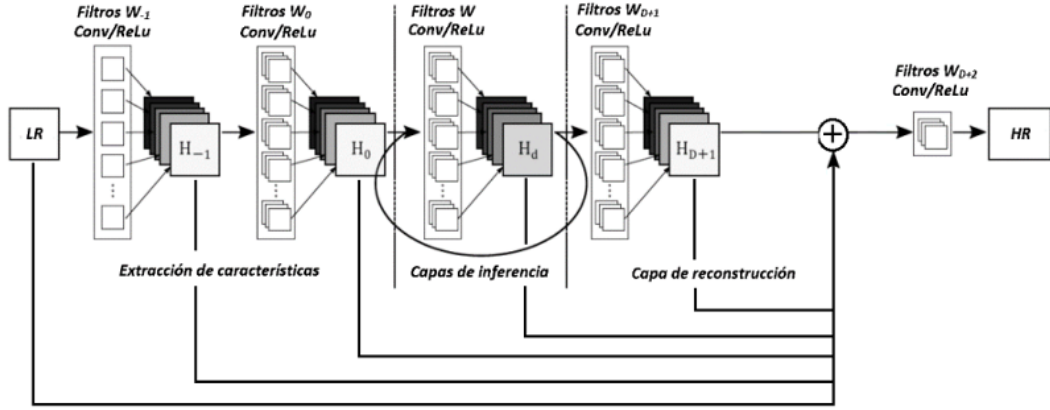


Figura 2.9: Estructura básica de una arquitectura DRCN. Imagen tomada de [23].

arquitectónicos de DRCN suelen destacar los bloques convolucionales recursivos con pesos compartidos y la combinación de múltiples redes de reconstrucción [23].

SRResNet

El objetivo principal de la arquitectura ResNet es permitir que una red neuronal convolucional sea muy profunda, evitando el problema del gradiente que desaparece (*vanishing gradient problem*) durante el proceso de entrenamiento [25]. Fue en [26] donde se propuso inicialmente abordar el problema de la degradación del gradiente usando un marco de aprendizaje residual, en el cual las capas de ResNet incorporan conexiones residuales (*skip connections*) que omiten una o más capas, realizando un mapeo identidad que permite sumar directamente la entrada a la salida.

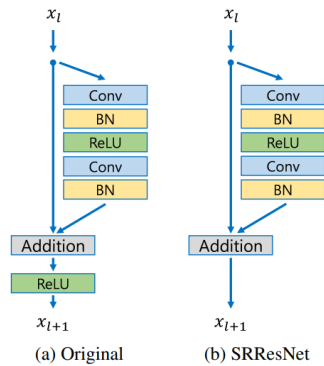


Figura 2.10: (a) Bloque residual ResNet, (b) Bloque Residual SRResNet. Imagen tomada de [19].

A partir de ResNet, se han desarrollado arquitecturas optimizadas específicamente para tareas de super-resolución y reconstrucción de imágenes, como la *Super Resolution Residual Network (SRResNet)*. En la Figura 2.10 se puede observar cómo varía un bloque residual optimizado para tareas de super-resolución, mientras que en la Figura 2.11 se presenta una arquitectura con *skip connections* que conectan las capas de *downsampling* con las capas de *upsampling* para la reconstrucción de la imagen. Este tipo de arquitecturas han demostrado ser compactas, reduciendo tanto el tamaño del modelo como el tiempo de entrenamiento [19].

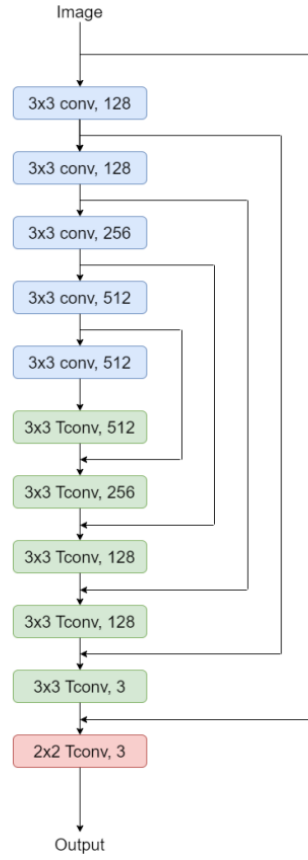


Figura 2.11: Arquitectura SRResNet. Imagen tomada de [25].

SRDenseNet

La arquitectura *Dense Convolutional Network (DenseNet)* propone un enfoque de conectividad que garantiza un flujo máximo de información entre las capas, conectando directamente todas ellas entre sí. Para mantener el flujo de información hacia adelante, cada capa recibe como entrada la salida de todas las capas anteriores y transmite su propia salida a todas las capas posteriores [27].

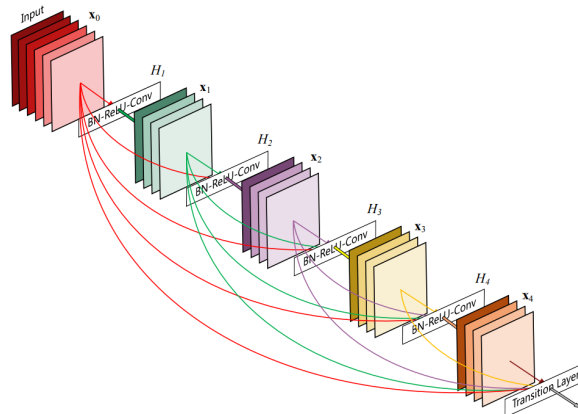


Figura 2.12: Bloque de conexión densa de cinco capas. Imagen tomada de [27].

La Figura 2.12 ilustra esquemáticamente el diseño de un bloque densamente conectado, en el cual los mapas de características se combinan mediante concatenación. De este modo, la capa l –ésima recibe l entradas, correspondientes a los mapas de características de todas las capas convolucionales anteriores. A su vez, sus propios mapas de características se transmiten a todas las $L - l$ capas siguientes, lo que da como resultado $L(L + 1)/2$ conexiones en una red de L capas. La Figura 2.13 muestra una arquitectura profunda DenseNet, en donde la red se divide en múltiples bloques densos. Entre estos bloques se insertan capas de transición que permiten reducir la dimensionalidad (*downsampling*) mediante operaciones de convolución y agrupamiento (*pooling*) [27].

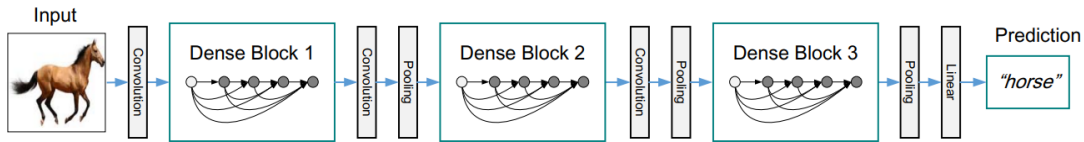


Figura 2.13: Arquitectura DenseNet profunda con tres bloques densos. Imagen tomada de [27].

En la DenseNet, cada capa tiene acceso directo tanto a los gradientes de la función de pérdida como la señal de entrada original, lo que permite una supervisión profunda implícita. Esto facilita el proceso de entrenamiento y contribuye además en la reducción del sobreajuste [27].

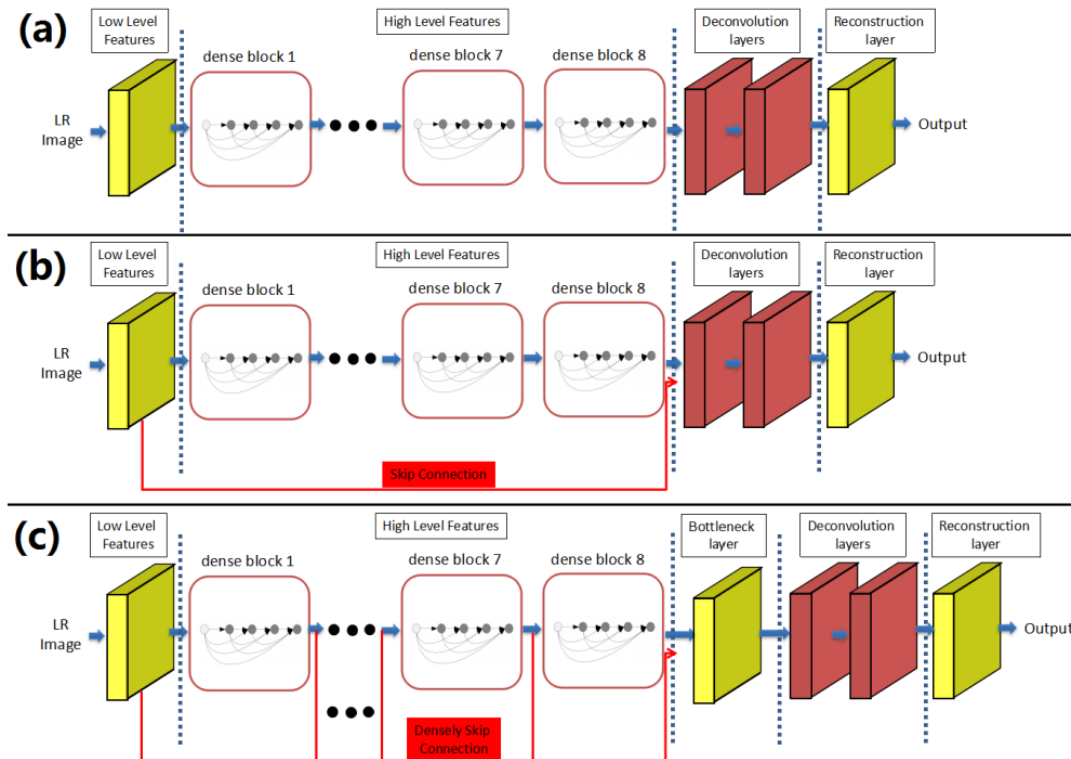


Figura 2.14: Tres arquitecturas diferentes de SRDenseNet. (a) SRDenseNet H que únicamente utiliza características de alto nivel para la reconstrucción de imágenes HR. (b) SRDenseNet HL que utiliza características de alto y bajo nivel para la reconstrucción de imágenes HR. (c) SRDenseNet All que utiliza todos los niveles de características combinados con *skip connections* para la reconstrucción de imágenes HR. Imagen tomada de [28].

Como se observa en la Figura 2.14, una arquitectura DenseNet diseñada para tareas de super-resolución

(SRDenseNet) puede dividirse en varias fases: una capa de convolución para el aprendizaje de características de bajo nivel, múltiples bloques densamente conectados para la extracción de características de alto nivel, capas de deconvolución para el aumento de escala y una capa de reconstrucción para generar la imagen en alta resolución (HR). DenseNet reduce el número de parámetros gracias a la reutilización de características, lo que permite un entrenamiento más eficiente al requerir menos memoria y capacidad de cómputo. Además, se ha demostrado que la combinación de características a diferentes niveles mejora el rendimiento en tareas de supe-resolución [28].

Super-Resolución mediante Refinamiento Repetido (SR3)

SR3 es un enfoque propuesto en [29] para la generación condicional de imágenes, el cual está inspirado en los modelos *Denoising Diffusion Probabilistic Models (DDPM)* introducidos en [30]. Los modelos probabilísticos de difusión aprenden a aproximar un proceso estocástico inverso para revertir una serie de pasos de difusión ruidosa, típicamente modelada como una cadena de Markov gaussiana (*forward process*) [30]. Estos modelos han demostrado un buen desempeño en tareas de super-resolución gracias a su efectividad en la modelación de distribuciones complejas, lo que ha permitido mejoras significativas en la calidad perceptual [31].

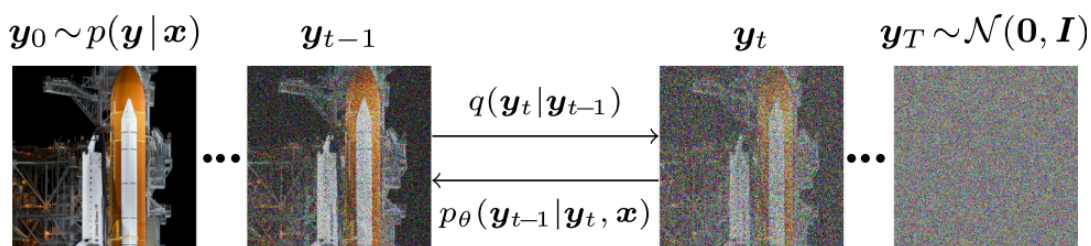


Figura 2.15: El proceso de difusión hacia adelante q (de izquierda a derecha) agrega gradualmente ruido gaussiano a la imagen de referencia. El proceso de inferencia inversa p (de derecha a izquierda) elimina de manera iterativa el ruido de la imagen de referencia. Imagen tomada de [29].

- **Proceso de Difusión Gaussiana:** Inicialmente se realiza un proceso de difusión markoviano hacia adelante durante el entrenamiento, el cual agrega gradualmente ruido gaussiano a una imagen de alta resolución y_0 durante T iteraciones hasta que se obtenga una imagen completamente ruidosa [29]:

$$q(\mathbf{y}_{1:T} | \mathbf{y}_0) = \prod_{t=1}^T q(\mathbf{y}_t | \mathbf{y}_{t-1}) \quad (2.4)$$

$$q(\mathbf{y}_t | \mathbf{y}_{t-1}) = \mathcal{N}(\mathbf{y}_t | \sqrt{\alpha_t} \mathbf{y}_{t-1}, (1 - \alpha_t) \mathbf{I})$$

En la anterior definición, los escalares $\alpha_{1:T}$ son hiperparámetros que determinan la varianza del ruido agregado en cada iteración, como se observa en la Figura 2.15.

- **Modelo de Eliminación de Ruido (*Denoising*):** Con el objetivo de revertir el proceso de difusión y recuperar la imagen objetivo sin ruido y_0 , se utiliza una imagen de referencia en baja resolución, junto con una versión ruidosa generada a partir de la imagen de alta resolución (proceso de difusión), y las estadísticas suficientes del ruido añadido (varianza del ruido). Con esta información, el modelo aprende a predecir el vector de ruido presente en la imagen ruidosa [29].
- **Inferencia a través de Refinamiento Iterativo:** La inferencia en SR3 es un proceso Markoviano inverso, que comienza desde la imagen con ruido Gaussiano y avanza hacia atrás en la dirección opuesta al proceso de difusión [29], como se observa en la Figura 2.15 y como se define a continuación:

$$\begin{aligned}
p_{\theta}(\mathbf{y}_{0:T} | \mathbf{x}) &= p(\mathbf{y}_T) \prod_{t=1}^T p_{\theta}(\mathbf{y}_{t-1} | \mathbf{y}_t, \mathbf{x}) \\
p(\mathbf{y}_T) &= \mathcal{N}(\mathbf{y}_T | \mathbf{0}, \mathbf{I}) \\
p_{\theta}(\mathbf{y}_{t-1} | \mathbf{y}_t, \mathbf{x}) &= \mathcal{N}(\mathbf{y}_{t-1} | \mu_{\theta}(\mathbf{x}, \mathbf{y}_t, \gamma_t), \sigma_t^2 \mathbf{I}).
\end{aligned}
\tag{2.5}$$

En este proceso, el modelo hace una predicción del ruido presente en la imagen ruidosa para posteriormente generar una versión menos ruidosa. De forma iterativa, se reconstruye así la imagen de alta resolución [29].

- **Arquitectura del Modelo SR3:** La arquitectura de SR3 es una U-Net con bloques residuales y *skip connections* en el *upsampling*. Esta es la red convolucional encargada de aprender a predecir el ruido agregado en cada paso del proceso de difusión [29].

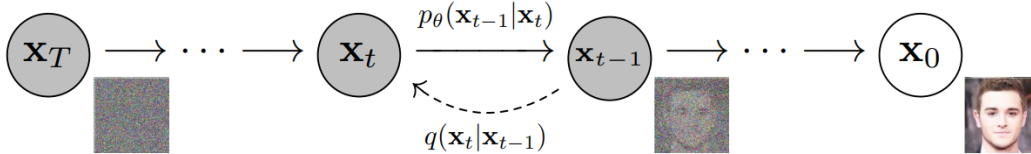


Figura 2.16: Refinamiento iterativo para la reconstrucción de la imagen a partir de ruido. Imagen tomada de [30].

2.2. Antecedentes

Se realizó una revisión de la bibliografía disponible con el objetivo de identificar trabajos previamente realizados que sean relevantes para la mejora de la resolución espacial en estudios de resonancia magnética (MRI) mediante el uso de métodos basados en aprendizaje profundo. Como resultado de esta búsqueda, se identificaron los siguientes estudios de revisión:

1. **Super Resolución de Imágenes de Resonancia Magnética Mediante Aprendizaje Profundo:** En el artículo se realiza un estudio detallado sobre cómo el uso de redes neuronales profundas mejora la resolución de las imágenes de resonancia magnética (MRI). Los autores adaptaron dos modelos avanzados de redes neuronales, originalmente desarrollados para la mejora de imágenes de tomografía computarizada (CT), y los aplicaron a la superresolución de imágenes MRI, logrando una mejora de hasta dos veces la resolución. En su metodología, utilizaron un conjunto de datos de imágenes T2 ponderadas de cerebro y crearon versiones de baja resolución para entrenar y probar los modelos. Este trabajo es significativo ya que sugiere que la implementación de estas técnicas podría permitir que los escáneres MRI convencionales ofrezcan imágenes de calidad superior sin necesidad de costosas actualizaciones de hardware, mejorando potencialmente la accesibilidad y eficacia de las imágenes médicas [32].
2. **Diseño y desarrollo de un método de super-resolución en imágenes de RM usando Deep Learning:** En este trabajo de grado, los autores proponen un método de super-resolución y clasificación de imágenes de MRI basado en redes neuronales artificiales, para mejorar el diagnóstico de pacientes con enfermedad de Alzheimer. Este método se divide en dos fases, la primera se desarrolla y entrenan redes neuronales convolucionales para transformar imágenes de baja resolución de T1 a imágenes de super-resolución T1 y T2. La segunda fase emplea las imágenes de la fase 1 en un modelo de redes neuronales de clasificación para clasificar al individuo como sano o con la enfermedad. En este documento, se corrobora la utilidad de las técnicas de super-resolución para mejorar los diagnósticos clínicos basados en las MRI [13].

3. **Estudio comparativo de algoritmos de super-resolución de una sola imagen basados en aprendizaje profundo:** El estudio evalúa cuatro algoritmos de super-resolución basados en aprendizaje profundo: Red Neuronal Convolutiva para Super-Resolución SRCNN, Red Neuronal Convolutiva Rápida para Super Resolución FSRCNN, Red muy profunda para super-resolución VDSR y Red profunda y resolutive para super-resolución DRCN; utilizando métricas de calidad de imagen en bases de datos de referencia y diferentes escalas de magnificación (2x, 3x y 4x). Los resultados demuestran que las arquitecturas profundas VDSR y DRCN, que emplean aprendizaje residual, superan a los métodos poco profundos SRCNN y FSRCNN en términos de métricas de relación señal-ruido PSNR e índice de similitud estructural SSIM. Esta mejora se atribuye a la reducción de la complejidad del entrenamiento, la minimización del desvanecimiento del gradiente y el mayor procesamiento de información. Los métodos poco profundos ofrecen menor complejidad computacional, pero las redes profundas son la mejor opción si se prioriza la calidad de la imagen [23].
4. **Reconstrucción de Imágenes Acelerada por Aprendizaje Profundo en Resonancia Magnética de la Órbita Ocular:** En el documento se lleva a cabo un estudio detallado sobre cómo el uso de técnicas de aprendizaje profundo (DL) puede mejorar la calidad de las imágenes de resonancia magnética (MRI) de la órbita ocular, reducir el tiempo de adquisición y minimizar los artefactos. Los autores evaluaron métodos basados en DL aplicados a secuencias T2 ponderadas y T1 ponderadas. Los resultados demostraron que el uso de DL redujo significativamente el ruido y los artefactos, mejoró la nitidez y disminuyó el tiempo de escaneo en un 69 % [33].
5. **Superresolución efectiva de resonancia magnética cerebral con aprendizaje profundo utilizando datos de entrenamiento simulados:** El objetivo de esta investigación es indagar sobre el potencial de los datos simulados de imágenes de resonancia magnética del cerebro para entrenar modelos de aprendizaje profundo basados en redes de superresolución. Los autores simularon un conjunto amplio y anatómicamente diverso de volúmenes cerebrales sin artefactos, alineados por vóxel y con diferentes resoluciones, con el fin de entrenar cuatro redes de superresolución: FSRCNN, RRDBnet, DeepUResnet y ESRGAN. Estas redes fueron evaluadas posteriormente con datos reales de múltiples fuentes. Los resultados muestran que los modelos entrenados con datos simulados lograron mejorar de forma significativa la nitidez de las imágenes de baja resolución. Aunque su rendimiento fue ligeramente inferior cuando se evaluaron con imágenes reales obtenidas de un único conjunto de datos, las redes entrenadas con datos simulados mostraron una mejor capacidad de generalización al ser evaluadas sobre múltiples conjuntos de datos reales. Estos hallazgos respaldan el uso de datos simulados como una estrategia efectiva para entrenar y robustecer redes de superresolución en MRI cerebral [34].
6. **Superresolución de imágenes de resonancia magnética cerebral:** En este estudio se propone la utilización de redes neuronales convolucionales profundas para mejorar la resolución espacial de imágenes de resonancia magnética del cerebro. En particular, se emplea una red SRCNN entrenada con imágenes de baja resolución, con el objetivo de reconstruir imágenes de alta resolución con mayor claridad y fidelidad que los métodos tradicionales, como la interpolación bicúbica. Los resultados obtenidos demuestran que esta red neuronal supera a los métodos clásicos en métricas como PSNR, SSIM y MSE, lo que confirma la eficacia de las técnicas de aprendizaje profundo para la reconstrucción de imágenes médicas [35].
7. **Superresolución de imágenes por resonancia magnética utilizando una red generativa adversarial cíclica tridimensional (3D CycleGAN):** En el documento se propone un enfoque basado en aprendizaje profundo que emplea una red 3D CycleGAN para mejorar la resolución espacial de imágenes de resonancia magnética sin necesidad de contar con datos emparejados. A diferencia de otros métodos supervisados que requieren pares exactos de imágenes de baja y alta resolución, se presenta una arquitectura denominada SRCycleGAN, que permite realizar el entrenamiento de forma no supervisada. Esta red aprovecha bloques densos residuales para reconstruir imágenes tridimensionales con un alto nivel de detalle. Los resultados obtenidos con el conjunto de datos BraTS demuestran que este enfoque supera a métodos tradicionales de interpolación y a redes generativas en 2D, logrando imágenes de mayor fidelidad visual y menores niveles de distorsión. Este estudio representa un avance significativo en la superresolución de imágenes médicas, especialmente en contextos donde no se dispone

de datos emparejados, lo que lo convierte en un valioso referente para el desarrollo de modelos de superresolución en resonancia magnética cerebral [36].

Capítulo 3

Desarrollo Metodológico

3.1. Obtención del Conjunto de Datos

Como se mencionó en el Objetivo Específico 1, para este proyecto se requirió un conjunto de estudios de resonancia magnética anatómicos T1, con los cuales fue posible entrenar y evaluar los modelos de Super-Resolución.

Estos estudios fueron gestionados y adquiridos desde el repositorio público IDA del *Laboratory of Neuro Imaging* de la *University of Southern California*. El *Image and Data Archive* (IDA, por sus siglas en inglés) es un repositorio global en línea, optimizado para almacenar, explorar y compartir datos e imágenes neurológicas, como MRI, PET, SPECT, EEG, entre otros [37]. Este repositorio cuenta con tecnologías flexibles de desidentificación de datos y transmisión cifrada de archivos, lo que garantiza el cumplimiento de las regulaciones sobre el tratamiento de datos y la privacidad de los pacientes [37]. Las políticas de propiedad y acceso a los datos en IDA establecen que todos los datos almacenados pertenecen exclusivamente a sus propietarios, quienes tienen el completo control sobre las decisiones relacionadas con acceso y descarga de imágenes [38]. Para acceder a los datos en IDA, es necesario completar un proceso de registro y crear una cuenta de usuario. Una vez la cuenta haya sido creada, es posible descargar datos de los estudios disponibles; sin embargo, algunos estudios específicos requieren una solicitud adicional de uso de datos, que puede realizarse a través del portal web [37].

Para el desarrollo de nuestro proyecto, decidimos trabajar con estudios MRI dentro del enfoque de investigación de la enfermedad de Parkinson, específicamente con la iniciativa *Parkinson's Progression Markers Initiative* (PPMI) [39]. En IDA se encuentran aproximadamente 2936 GB de archivos relacionados con el PPMI, que incluye datos de más de 4314 sujetos que se han cargado desde el año 2010.

Desde IDA se descargó un conjunto de datos con las siguientes características: 333 estudios MRI 3D de corte sagital, tomados con máquinas de intensidad de campo de 1.5 Tesla, donde cada imagen estaba almacenada en formato DICOM.

3.2. Exploración de Datos

En los 333 estudios MRI 3D descargados hay un total de 55858 imágenes (slides o cortes) de alta resolución en formato DICOM, de las cuales 31474 tienen un tamaño de 256x256, todas en escala de grises. De los 333 estudios, 182 corresponden a género masculino y 151 a género femenino, con rango de edad que van desde los 32 años hasta los 82 años.

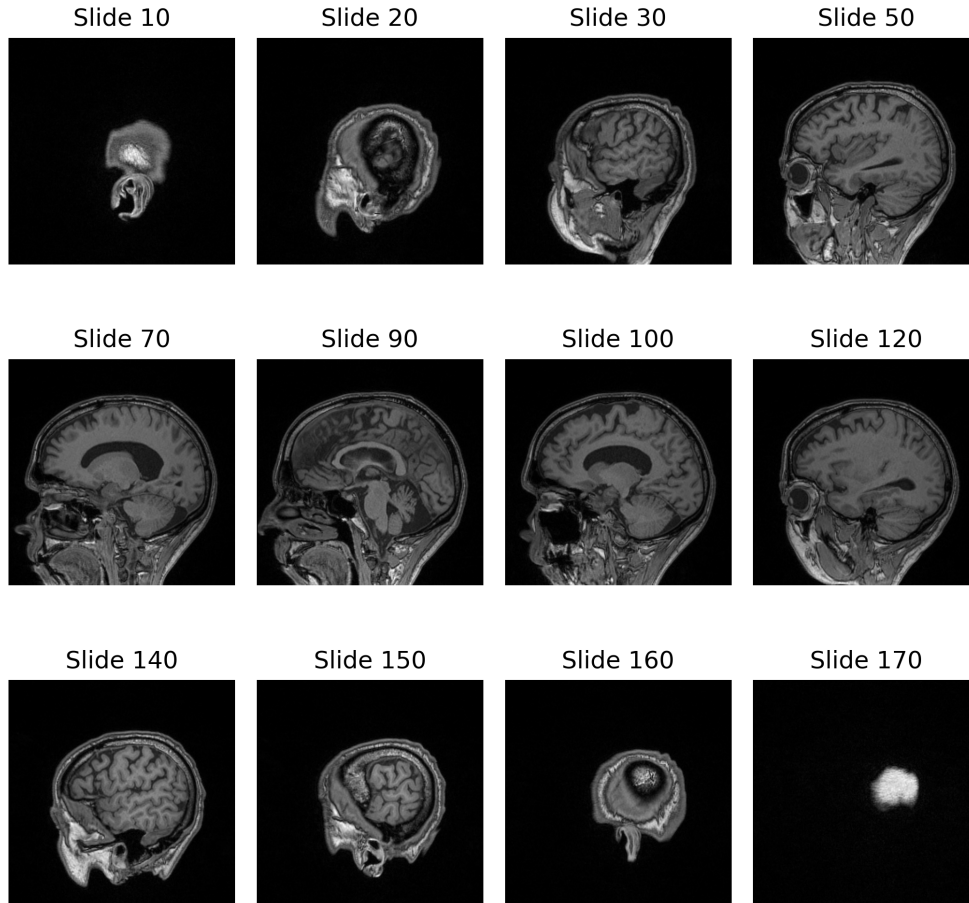


Figura 3.1: Ejemplo de estudio MRI 3D de corte sagital.

3.3. Preprocesamiento de Imágenes MRI

3.3.1. Preprocesamiento de Imágenes MRI 3D

Se inicia con la carga de los estudios en formato DICOM, cada uno compuesto por múltiples cortes en el plano sagital. Estos cortes se ensamblan en volúmenes tridimensionales de tamaño $176 \times 256 \times 256$ píxeles. Para asegurar una correcta secuencia espacial, los archivos DICOM se ordenan de acuerdo con el atributo InstanceNumber. Se prosigue con la normalización de valores de intensidad de cada volumen al rango $[0,1]$ utilizando una normalización min-max, lo que permite homogenizar las entradas para su uso en redes neuronales.

Para simular datos degradados, se crearon versiones de baja resolución de cada volumen original. Este proceso se aplicó corte por corte, utilizando un escalado hacia abajo al 50% mediante interpolación por área (INTER AREA), seguido de un re-muestreo al tamaño original mediante interpolación bilineal (INTER LINEAR). Gracias a esto, se preservan las dimensiones espaciales del volumen mientras se introduce una pérdida controlada de calidad.

Finalmente, los volúmenes se dividieron en conjuntos de entrenamiento (90%) y validación (10%). Debido a que cargar los volúmenes completos superaba la capacidad de la VRAM disponible, se usó una estrategia basada en parches tridimensionales. Cada volumen fue subdividido en bloques de tamaño $64 \times 64 \times 64$ con un stride 32, permitiendo obtener múltiples ejemplos por estudio sin exceder la memoria. Los pares de parches de baja y alta resolución fueron organizados como muestras de entrenamiento y validación, asegurando

compatibilidad con arquitecturas convolucionales 3D.

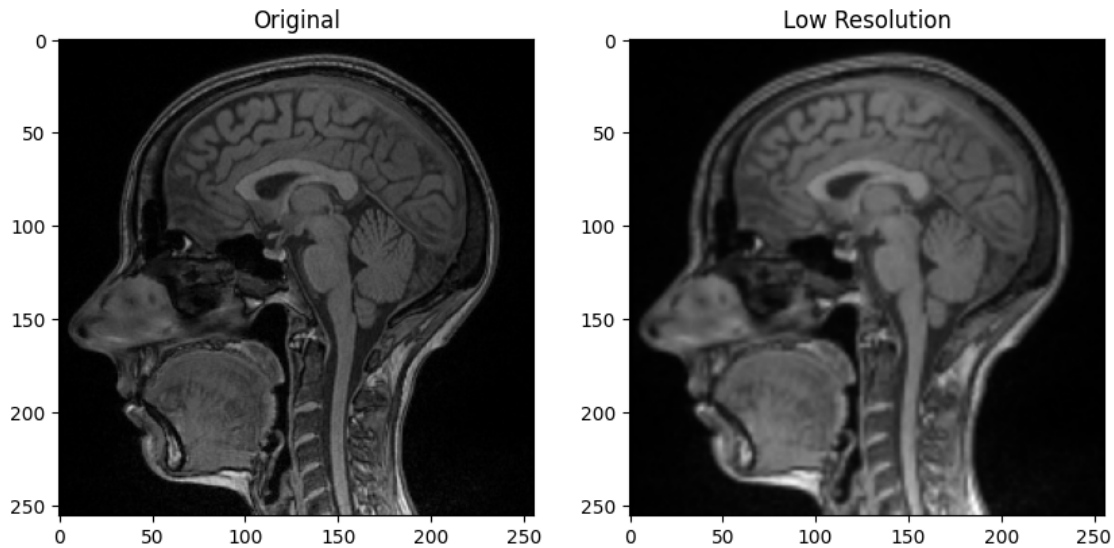


Figura 3.2: Ejemplo 1 de MRI con resolución original y resolución reducida.

3.3.2. Preprocesamiento de Imágenes MRI 2D

El preprocesamiento inicia explorando el directorio para identificar archivos DICOM, que se filtran y ordenan según el atributo InstanceNumber para mantener la secuencia espacial correcta. Solo se cargan imágenes con dimensiones de 256x256 píxeles. Luego, se normalizan los valores de píxel, originalmente en un rango de 0 a 255, convirtiéndolos a tipo float con `astype(np.float32)` y dividiéndolos por 255 para obtener valores entre 0 y 1. Esta normalización mejora la estabilidad numérica y la eficiencia en algoritmos de aprendizaje automático. Se utilizan las librerías `pydicom` para leer imágenes y `NumPy` para manipular los datos numéricos.

Una vez cargadas las imágenes, se realiza un proceso de reducción de resolución controlado a través de un factor de escala (0.5 en este caso). Cada imagen se submuestra al 50% de su tamaño original utilizando interpolación por área (`INTER AREA`) para preservar la calidad visual al reducir la resolución. Posteriormente, estas imágenes reducidas se remuestrean al tamaño original (256x256 píxeles) mediante interpolación bilineal (`INTER LINEAR`), simulando imágenes de menor calidad, pero manteniendo las dimensiones originales. Este procedimiento permite generar un conjunto de datos de baja resolución que puede ser utilizado para tareas como super-resolución o reconstrucción de imágenes.

Las imágenes de alta y baja resolución se dividen en conjuntos de entrenamiento (80%), validación (10%) y prueba (10%) para garantizar una adecuada evaluación del modelo. Cada conjunto se reestructura añadiendo una dimensión adicional, transformando las imágenes en tensores de forma (256, 256, 1), lo cual es necesario para su compatibilidad con redes convolucionales. Este preprocesamiento incluye la preparación de bloques de *downsampling* y *upsampling*, que serán utilizados en arquitecturas de redes neuronales para reducir y aumentar la resolución de las imágenes, respectivamente. Estas técnicas aseguran que los datos estén listos para ser utilizados en tareas de aprendizaje profundo relacionadas con imágenes médicas.

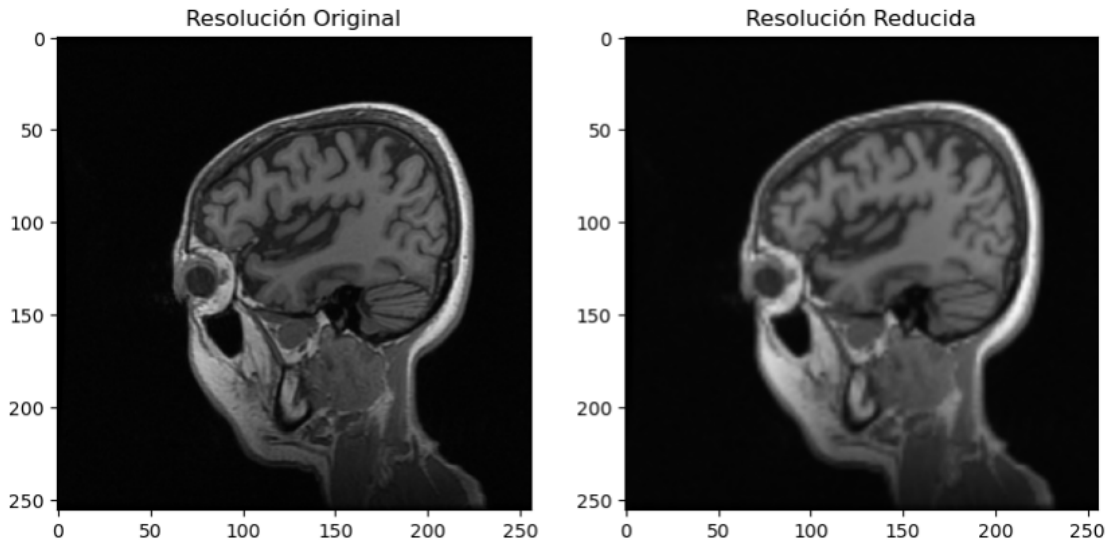


Figura 3.3: Ejemplo 2 de MRI con resolución original y resolución reducida.

3.4. Selección de Hiperparámetros

La selección de hiperparámetros es una etapa fundamental en el desarrollo de modelos de aprendizaje automático, dado su impacto directo en la capacidad de generalización y en el rendimiento final del modelo. Existen diversas técnicas sistemáticas para optimizar estos valores, tales como *random search*, *grid search* y enfoques más avanzados como la *optimización bayesiana*, que pueden implementarse mediante herramientas como *Optuna*, *Keras Tuner*, *Ray Tune* o *Hyperopt*.

En este trabajo, debido a las limitaciones computacionales y al alto costo temporal de los entrenamientos, se decidió no aplicar estrategias automatizadas de optimización de hiperparámetros. En el caso de los modelos tridimensionales, cada experimento implicaba tiempos de entrenamiento que oscilaron entre 4 y 8 horas, lo que llevó a priorizar la eficiencia en el uso de GPU mediante el ajuste manual de la complejidad de cada arquitectura. Para los modelos bidimensionales, aunque algunos entrenamientos fueron más rápidos (entre 30 minutos y 3 horas), otros casos llegaron a demorar hasta 10 horas, lo que también limitó la posibilidad de realizar múltiples pruebas sistemáticas.

Ante este escenario, se adoptó una estrategia empírica basada en prueba y error, orientada por la experiencia previa y las recomendaciones encontradas en la literatura especializada. Durante el entrenamiento, se monitorearon métricas clave como la función de pérdida (por ejemplo, L1 o MSE, según la arquitectura), además del PSNR y SSIM calculados, lo cual permitió realizar ajustes manuales sobre hiperparámetros sensibles como la tasa de aprendizaje, la regularización o la profundidad del modelo. Este enfoque permitió contar con un criterio de comparación objetivo y consistente entre modelos, pese a no haber empleado un proceso automatizado de optimización de hiperparámetros.

3.5. Implementación de Modelos

3.5.1. Modelos 3D

Para las arquitecturas tridimensionales, se utilizó un equipo de cómputo conformado por un procesador AMD Ryzen 5 5600X, una tarjeta gráfica NVIDIA GeForce RTX 4060 con 8 GB de VRAM, y 16 GB de memoria RAM. El desarrollo se realizó en Python, utilizando el framework PyTorch, aprovechando su com-

patibilidad con CUDA para ejecutar los entrenamientos en la GPU. Como entorno de trabajo se empleó Jupyter Notebook, lo cual facilitó la organización del código y el análisis iterativo de resultados. Además, se utilizaron algunas librerías complementarias como lo son OpenCV, pydicom, scikit-image y scikit-learn para el manejo, procesamiento y evaluación de las imágenes médicas.

Se utilizaron, en todos los modelos 3D, un total 30 estudios MRI en el plano sagital para el proceso de entrenamiento y validación.

A continuación, se realiza un repaso de los parámetros e hiperparámetros de cada arquitectura 3D.

UNET 3D

Arquitectura: Está compuesto por codificador (encoder), cuello de botella (*bottleneck*) y decodificador (decoder). El codificador tiene dos capas convolucionales 3D (Conv3d) con filtros de $3 \times 3 \times 3$ y activación ReLU, que transforman el volumen de entrada de un canal a 32 canales, seguidas por una operación de MaxPool3D que reduce la resolución espacial. El cuello de botella incorpora dos capas Conv3D con 64 filtros, permitiendo una representación más abstracta del volumen comprimido. El decodificador inicia con una capa ConvTranspose3d que realiza un *upsampling*, seguida de una capa Conv3D con 32 filtros para refinar características, y una última Conv3D con kernel $1 \times 1 \times 1$ que reduce los canales a uno, generando la salida. Se emplea ReLU en todas las capas, excepto en la última, que no incluye activación para permitir una salida continua.

Compilación: Se eligió la función de pérdida L1 (MAE), por su capacidad para conservar detalles finos. El optimizador utilizado fue AdamW, con una tasa de aprendizaje inicial de 0.0001 y un factor de regularización L2 (*weight decay*) de 0.00001 para prevenir sobreajuste. Además, se incorporó un *scheduler* que reduce la tasa de aprendizaje cuando la pérdida en validación se estanca, con un factor de reducción de 0.8 y una paciencia de 10 épocas. Se usaron PSNR y SSIM como métricas de evaluación.

Entrenamiento: El modelo se entrenó durante 50 épocas, utilizando un tamaño de *batch* de 4 volúmenes tanto para el conjunto de entrenamiento como para el de validación.

EDSR 3D

Arquitectura: El modelo se compone de tres bloques principales: entrada (*head*), cuerpo (*body*) y salida (*tail*). El bloque de entrada contiene una capa convolucional tridimensional (Conv3d) con un kernel de $3 \times 3 \times 3$, que extrae características iniciales del volumen de entrada. La activación ReLU no se aplica en esta etapa, sino en los bloques residuales del cuerpo. El cuerpo está compuesto por una serie de bloques residuales 3D, cada uno con dos capas Conv3d (con kernel $3 \times 3 \times 3$), una activación ReLU después de la primera convolución y una conexión residual (*skip connection*) que facilita el flujo de gradientes y estabiliza el entrenamiento. Se utilizaron cuatro bloques residuales, todos operando con una profundidad fija de 32 canales. Finalmente, el bloque de salida utiliza una capa Conv3d con kernel $3 \times 3 \times 3$ que proyecta los mapas de características a una única salida volumétrica. No se aplica función de activación en la última capa para permitir una salida continua.

Compilación: Se usó L1 (MAE) como función de pérdida, debido a su utilidad en la preservación de estructuras detalladas, lo cual es especialmente importante en imágenes médicas. Como optimizador se empleó AdamW, con una tasa de aprendizaje inicial de 0.0001 y una regularización L2 (*weight decay*) de 0.00001, con el objetivo de reducir el sobreajuste. Se añadió un scheduler para reducir la tasa de aprendizaje en caso de que la pérdida no mejorara, con un factor de reducción de 0.9 y una paciencia de 10 épocas.

Entrenamiento: El modelo se entrenó durante 50 épocas, utilizando un tamaño de *batch* de 4 volúmenes tanto para el conjunto de entrenamiento como para el de validación.

cGAN 3D

Arquitectura: El modelo está compuesto por dos subredes: un generador y un discriminador. El generador toma como entrada volúmenes de baja resolución (LR) y produce su correspondiente versión en super-resolución (SR). Su arquitectura inicia con una capa Conv3d con kernel $3 \times 3 \times 3$ y 32 canales de salida, seguida por una activación Leaky ReLU. A continuación, se incluyen tres bloques residuales, cada uno conformado por dos capas Conv3d con kernel $3 \times 3 \times 3$. En cada bloque, la primera convolución es seguida por una normalización por lotes (BatchNorm3d) y activación Leaky ReLU, mientras que la segunda convolución se aplica directamente. Se emplean conexiones residuales para facilitar la retropropagación del gradiente. La capa final del generador es una convolución 3D adicional con activación sigmoide, que proyecta el resultado a un único canal, formando el volumen SR.

El discriminador recibe como entrada el volumen de baja resolución (LR) concatenado con una imagen de alta resolución (real HR o generada SR), y predice la probabilidad de que esta última sea auténtica. Su arquitectura consiste en tres capas Conv3d con kernel $3 \times 3 \times 3$ y 32, 64 y 128 canales, respectivamente, cada una seguida de una activación Leaky ReLU. Luego, se emplea una capa de pooling adaptativo que reduce el volumen a una representación de tamaño fijo, la cual es pasada a una capa totalmente conectada con activación sigmoide que produce una salida escalar.

Compilación: Se utilizó una función de pérdida compuesta que combina el error adversarial con una pérdida MSE ponderada, multiplicando el término MSE por un factor de 3. Esto busca equilibrar la generación de imágenes realistas con la preservación de detalles estructurales. La función adversarial consistió en una pérdida MSE entre la salida del discriminador y el valor esperado (1 para imágenes reales, 0 para falsas). Tanto el generador como el discriminador se optimizaron con AdamW, utilizando una tasa de aprendizaje de 0.0001 y parámetros beta de (0,0, 0,9), con el fin de favorecer la estabilidad del entrenamiento.

Entrenamiento: El entrenamiento se realizó durante 50 épocas, utilizando un tamaño de *batch* de 4 volúmenes, tanto para el conjunto de entrenamiento como para el de validación. En cada época, el generador y el discriminador fueron actualizados alternadamente con base en parches 3D extraídos aleatoriamente de los estudios MRI.

SRCNN 3D

Arquitectura: El modelo está compuesto por una arquitectura secuencial de cinco capas Conv3d, con funciones de activación ReLU aplicadas después de las primeras cuatro capas. En la entrada, se utiliza una capa Conv3d con kernel $9 \times 9 \times 9$ y 64 canales, lo que permite una captación más amplia del contexto espacial desde el volumen de entrada. Posteriormente, se usan tres capas Conv3d intermedias con kernel $3 \times 3 \times 3$, reduciendo progresivamente el número de canales a 64, 32 y 16, respectivamente, logrando una compresión efectiva de las características. En la capa de salida se emplea un kernel $5 \times 5 \times 5$ para proyectar los mapas de características a una única imagen volumétrica de salida. No se aplica función de activación en la última capa, lo que permite que el modelo genere salidas continuas.

Compilación: Se utilizó la función de pérdida MSE, para minimizar las diferencias voxel a voxel entre el volumen reconstruido y el de referencia. El optimizador fue Adam, con una tasa de aprendizaje inicial de 0.0001 y una regularización L2 (*weight decay*) de 0.00001, para mitigar el riesgo de sobreajuste. También, se implementó un scheduler con un factor de reducción de 0.5 en caso de que la pérdida no mejore durante 5 épocas consecutivas, y así favorecer una convergencia más estable.

Entrenamiento: El modelo se entrenó durante 50 épocas, con un tamaño de batch de 4 volúmenes, tanto para el conjunto de entrenamiento como para el de validación.

3.5.2. Modelos 2D

Para las arquitecturas Autoencoder, DRCN, SRGAN, UNET y VDSR, el entrenamiento se realizó en un Mac Studio M2 Max con 64 GB de RAM, CPU de 12 núcleos (8 de rendimiento y 4 de eficiencia), GPU de 30 núcleos y Neural Engine de 16 núcleos. El conjunto de datos utilizado consistió en 31.474 imágenes

MRI de 256x256, de estas, se usaron 25.179 imágenes para entrenamiento, 3.147 imágenes para validación y 3.148 imágenes para pruebas (80 %, 10 %, 10 %). El software corresponde a Visual Studio Code para MacOS, Python versión 3.10.15.

Los modelos SRResNet, SRDenseNet y SR3 fueron implementados y entrenados en Google Colab utilizando una suscripción Colab Pro, la cual proporciona un acelerador de hardware GPU A100 con 40 GB de memoria GPU, 83.5 GB de RAM del sistema y un entorno de ejecución Python 3.

Las librerías utilizadas son:

- **os:** Manejo de operaciones del sistema operativo, como navegación y manipulación de archivos y directorios.
- **pydicom:** Lectura y escritura de archivos DICOM, usados para el manejo de los archivos MRI.
- **numpy:** Manipulación eficiente de arreglos y operaciones matemáticas sobre datos numéricos.
- **matplotlib.pyplot:** Visualización de datos mediante gráficos y representaciones visuales.
- **cv2 (OpenCV):** Procesamiento y análisis de imágenes y videos.
- **tensorflow:** Framework para construir y entrenar modelos de aprendizaje automático y redes neuronales.
- **tensorflow.keras (layers, Model):** Componentes para crear y estructurar modelos de redes neuronales dentro de TensorFlow.
- **pandas:** Manipulación y análisis de datos tabulares, como hojas de cálculo y bases de datos.
- **math:** Funciones matemáticas básicas y avanzadas.

A continuación, se hace la descripción de las arquitecturas e hiperparámetros de cada modelo.

Autoencoder 2D

Arquitectura: La arquitectura utilizada consta de 8 capas. Comienza con una capa convolucional (Conv2D) que aplica 32 filtros de 3x3 a la imagen de entrada de baja resolución, extrayendo características iniciales. La siguiente capa, MaxPool2D, reduce la dimensionalidad espacial a la mitad (128x128) para enfocar las características más importantes. Posteriormente, una capa de UpSampling2D duplica la resolución espacial (256x256) mediante interpolación bilineal, preparando la imagen para las siguientes capas convolucionales que refinarán los detalles. Le siguen 4 capas convolucionales, todas con 64 filtros de 3x3, que aprenden representaciones más complejas de las características de alta resolución. Finalmente, la capa convolucional de salida utiliza un único filtro de 3x3 para generar la imagen de salida de alta resolución con un solo canal, aplicando una función sigmoide para normalizar los valores de los píxeles.

Compilación: incluyen la función de pérdida MAE, optimizador Adam con una tasa de aprendizaje de 0.0001, encargado de ajustar los pesos para minimizar dicha función de pérdida. Además, se emplean como métricas SSIM y PSNR para evaluar el desempeño del modelo.

Entrenamiento: 30 iteraciones completas sobre todo el conjunto de datos para entrenar el modelo (*epochs*) y un *batch size* de 32 imágenes procesadas por lote en cada paso de entrenamiento.

VDSR 2D

Arquitectura: La extracción inicial de características se realiza mediante una capa convolucional con 64 filtros de tamaño 3x3 y activación ReLU, encargada de captar las características básicas de la imagen. A continuación, se aplican 20 capas convolucionales recursivas, cada una con 64 filtros de 3x3 y activación ReLU, que permiten al modelo aprender representaciones profundas y complejas para mejorar la calidad

de la imagen. Finalmente, la reconstrucción de la imagen de alta resolución se lleva a cabo con una capa convolucional final que utiliza un filtro de 3x3 y activación lineal para generar la salida reconstruida.

Compilación: Incluyen la función de pérdida MAE, el optimizador es Adam con una tasa de aprendizaje de 0.0001. Además, se emplean como métricas SSIM y PSNR para evaluar el desempeño del modelo.

Entrenamiento: 30 iteraciones completas sobre todo el conjunto de datos para entrenar el modelo (*epochs*) y un *batch size* de 32 imágenes procesadas por lote en cada paso de entrenamiento.

SRGAN 2D

Arquitectura: La arquitectura consta de dos modelos principales: un generador y un discriminador. El generador recibe imágenes de baja resolución (256x256, escala de grises) y produce imágenes de alta resolución. Comienza con una capa convolucional inicial con 64 filtros de 3x3 y activación ReLU para extraer características básicas. Luego, aplica 20 capas recursivas convolucionales (*num_recursive_layers=20*), todas con 64 filtros de 3x3 y activación ReLU, que permiten aprender características profundas y complejas. Finalmente, una capa convolucional de salida con un filtro de 3x3 y activación lineal genera la imagen reconstruida. El discriminador es una red convolucional que clasifica imágenes reales y generadas. Está compuesto por varias capas convolucionales con filtros crecientes (64, 128, 256, 512), cada una seguida de una activación LeakyReLU ($\alpha=0.2$) y con *stride* de 2 para reducción espacial. La salida se aplanada y pasa a una capa densa con activación sigmoide.

Compilación: Incluyen la función de pérdida MAE, el optimizador es Adam con una tasa de aprendizaje de 0.0002. Además, se emplean como métricas SSIM y PSNR para evaluar el desempeño del modelo.

Entrenamiento: 30 iteraciones completas sobre todo el conjunto de datos para entrenar el modelo (*epochs*) y un *batch size* de 32 imágenes procesadas por lote en cada paso de entrenamiento.

DRCN 2D

Arquitectura: La arquitectura utilizada consta de una capa de entrada seguida por una capa convolucional inicial que extrae características con 64 filtros de 3x3 y activación ReLU. Luego, se aplican 4 capas convolucionales recursivas (*num_recursive_layers=4*), todas con 64 filtros de 3x3 y activación ReLU, que permiten aprender representaciones profundas y complejas de la imagen. Finalmente, una capa convolucional de salida con un único filtro de 3x3 y activación lineal genera la imagen reconstruida en escala de grises.

Compilación: Incluyen la función de pérdida MAE, el optimizador es Adam con una tasa de aprendizaje de 0.001. Además, se emplean como métricas SSIM y PSNR para evaluar el desempeño del modelo.

Entrenamiento: 30 iteraciones completas (*epochs*) y un *batch size* de 32.

UNET 2D

Arquitectura: La arquitectura utilizada consta de tres bloques. En el primero (codificación), se aplican dos capas convolucionales con 32 filtros de 3x3 y activación ReLU, seguidas de una capa de MaxPool2D 2x2 para reducir la resolución espacial. Luego, se repite un bloque similar con 64 filtros y otra capa de MaxPool2D para una mayor reducción. En el segundo (*bottleneck*), se utilizan dos capas convolucionales con 128 filtros de 3x3 y activación ReLU, que hacen la representación más comprimida y abstracta de la imagen. Por último, en el bloque de decodificación se realizan operaciones de *upsampling* para aumentar la resolución espacial, concatenando las características correspondientes del bloque de codificación (conexiones de salto) para recuperar detalles finos. Se aplican dos capas convolucionales con 64 filtros después del primer *upsampling* y dos capas con 32 filtros tras el segundo. La capa de salida es convolucional con un filtro de 1x1 y activación sigmoide.

Compilación: Incluyen la función de pérdida MAE, el optimizador es Adam con una tasa de aprendizaje de 0.001. Además, se emplean como métricas SSIM y PSNR para evaluar el desempeño del modelo.

Entrenamiento: 20 iteraciones completas (*epochs*) y un *batch size* de 32.

SRResNet 2D

Datos: Se utilizaron 25 179 imágenes para entrenamiento, 3147 imágenes para validación y 3148 imágenes para pruebas. Los conjuntos de entrenamiento, validación y prueba están compuestos por pares de imágenes de alta y baja resolución, de tamaño 256x256x1. Se utilizó un *Data Generator* para generar lotes de datos de entrenamiento y validación con *batch size* igual a 32.

Arquitectura: El modelo inicia con cinco capas convolucionales para la etapa de *downsampling*, que reducen la imagen de entrada de 256x256x1 a 8x8x512. Las dos primeras capas tienen 128 filtros, la tercera capa tiene 256 filtros y las capas cuarta y quinta tienen 512 filtros.

En la etapa de *upsampling*, se implementan cinco capas convolucionales transpuestas que escalan las imágenes de 8x8x512 a 256x256x3. La primera capa convolucional transpuesta tiene 512 filtros, la segunda tiene 256 filtros, las capas tercera y cuarta tienen 128 filtros, y la quinta capa tiene 3 filtros.

Las conexiones residuales (*skip connections*) enlazan las capas convolucionales con las capas de convolución transpuesta: la salida de la primera capa de convolución transpuesta se concatena con la salida de la cuarta capa de convolución; la segunda capa transpuesta se concatena con la tercera convolucional; la tercera capa transpuesta se concatena con la segunda capa de convolución; la cuarta capa transpuesta se concatena con la primera capa convolucional; y la quinta capa transpuesta se concatena con la imagen de entrada.

Finalmente, se añade una capa de convolución con un filtro y un kernel de tamaño 2x2 para la reconstrucción de la imagen de salida de tamaño 256x256x1.

Todas las capas convolucionales y convolucionales transpuestas utilizan un kernel de tamaño 3x3 (excepto por la capa de reconstrucción) y emplean la función de activación ReLU. El modelo cuenta con 9 595 284 parámetros entrenables.

Compilación: El modelo se compila utilizando el optimizador Adam, con una tasa de aprendizaje de 0,0001 y un valor de β_1 igual a 0,5. La función de pérdida empleada es el *Mean Absolute Error*, y las métricas usadas para la evaluación del desempeño son el PSNR y el SSIM.

Entrenamiento: El modelo se entrenó en 50 épocas utilizando el *Data Generator* que genera lotes de imágenes de entrada en baja resolución y las imágenes objetivo en alta resolución.

SRDenseNet 2D

Datos: Se utilizaron 25 179 imágenes para entrenamiento, 3147 imágenes para validación y 3148 imágenes para pruebas. El conjunto de datos está compuesto por pares de imágenes de baja resolución de tamaño 128x128x1 y de alta resolución (imágenes objetivo) de tamaño 256x256x1. Para entrenar este modelo se implementó un *Data Generator* encargado de generar lotes de datos de entrenamiento y validación de tamaño 64.

Arquitectura: La entrada del modelo tiene una forma de 128x128x1 seguida de una capa convolucional con 16 filtros con kernel de tamaño 3x3 y función de activación ReLU.

Posteriormente, se incorporan ocho bloques densos, en los cuales cada entrada tiene la concatenación de la salida del bloque anterior con la salida de la primera capa convolucional. Cada bloque denso está compuesto por ocho capas convolucionales. En estos bloques, cada capa convolucional recibe como entrada la concatenación de las salidas de todas las capas anteriores. Todas las capas tienen 16 filtros con kernel de tamaño 3x3, emplean la función de activación ReLU y un regularizador L2 con λ igual a 0,01.

Luego de los bloques densos viene una etapa *bottleneck* con una capa convolucional con 256 filtros, kernel de tamaño 1x1 y función de activación ReLU, que lleva el mapa de características a un tamaño de 128x128x256. En la etapa de deconvolución se utiliza una capa de convolución transpuesta con 256 filtros, kernel de tamaño 3x3 y función de activación ReLU. Esta capa escala el mapa de características a un tamaño de 256x256x256. Finalmente, en la etapa de reconstrucción se usa una capa convolucional con 1 filtro y con kernel de tamaño

3x3 para generar la imagen de salida de alta resolución de tamaño 256x256x1. Así, el modelo cuenta con 1 294 241 parámetros entrenables.

Compilación: El modelo utiliza el optimizador Adam con una tasa de aprendizaje de 0,0001 y un valor de β_1 igual a 0,5. La función de pérdida fue el *Mean Absolute Error*, y las métricas usadas para la evaluación del desempeño son el PSNR y el SSIM.

Entrenamiento: El modelo se entrenó en 30 épocas utilizando el *Data Generator* con *batch size* igual a 64, el cual provee las imágenes de entrada en baja resolución y las imágenes objetivo en alta resolución.

SR3 2D

Datos: Debido al alto consumo de recursos computacionales y de memoria que requiere este modelo durante el proceso de entrenamiento (se entrena completamente tantas veces como pasos de difusión se definan), se utilizaron únicamente 800 imágenes para el entrenamiento y 100 imágenes para las pruebas. El conjunto de datos utilizado para entrenamiento e inferencia está compuesto por pares de imágenes de baja resolución de tamaño 256x256x1 y de alta resolución de tamaño 256x256x1.

Arquitectura: El modelo implementado tiene 9 438 467 parámetros entrenables. Este modelo tiene una arquitectura U-Net dividida en 6 etapas y construida usando bloques convolucionales:

- *Bloque convolucional:* A este bloque ingresa la concatenación de la señal de las imágenes (imagen ruidosa y la imagen en baja resolución), junto con el parámetro temporal (*time step*). La concatenación de las imágenes pasa por una capa convolucional de 64 filtros con kernel de tamaño 3x3 y activación ReLU. El parámetro temporal pasa por una capa densa de 64 neuronas, seguida de una capa de activación ReLU y una capa de redimensionamiento (*Reshape*) de forma (1,1,64). La salida de la primera capa convolucional es multiplicada con el parámetro temporal redimensionado. También se incluye una segunda capa convolucional con 64 filtros y kernel de tamaño 3x3, a la cual ingresa la concatenación de las imágenes ruidosas y de baja resolución. La salida de la segunda capa convolucional se suma al resultado de la multiplicación entre la salida de la primera capa convolucional y el parámetro temporal. El resultado de esta suma pasa por una capa de normalización y una capa de activación ReLU. El bloque convolucional se visualiza en la Figura 3.4.

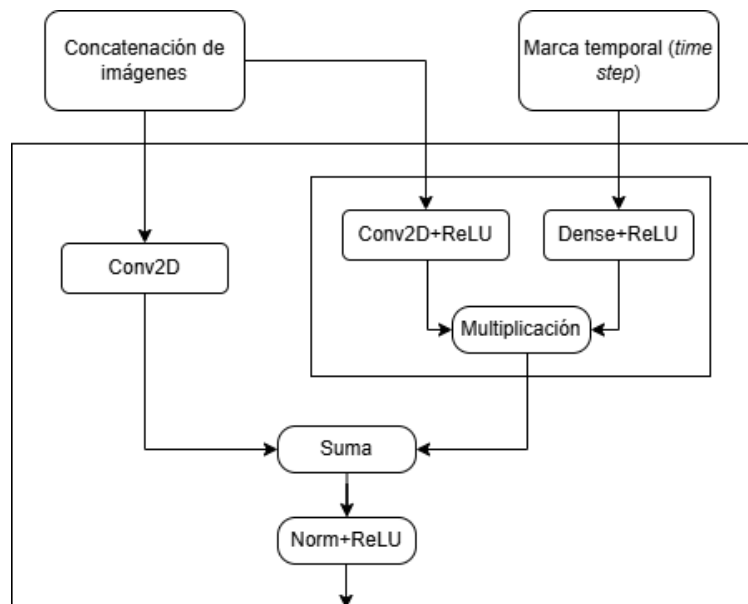


Figura 3.4: Bloque convolucional en SR3.

■ *Etapas de la arquitectura:*

1. Entradas del modelo: Al modelo entra la imagen ruidosa de tamaño $256 \times 256 \times 1$ y la imagen en baja resolución de tamaño $256 \times 256 \times 1$. Estas dos imágenes son concatenadas en una sola variable.
2. Entrada de la marca de tiempo (*time steps*): Se cuenta con una entrada que es un vector unidimensional, el cual pasa por una capa densa de 64 neuronas, una capa de normalización y una capa de activación ReLU.
3. *Downsampling*: En esta etapa se utilizan cuatro bloques convolucionales sucesivos, los cuales reciben como entrada las señales de las imágenes y la marca temporal. Luego de los tres primeros bloques se aplica una capa MaxPool2D con *pool size* de 2×2 . En este proceso se lleva la señal de las imágenes de $256 \times 256 \times 2$ a $32 \times 32 \times 64$.
4. Perceptrón multicapa (MLP): En esta etapa primero se aplana la señal de las imágenes para luego concatenarla con la marca temporal. Luego se emplea una capa densa con 64 neuronas, seguida de una capa de normalización y una capa de activación ReLU. Finalmente, la señal pasa por una capa densa de 65,536 neuronas, seguida por una capa de normalización, una capa de activación ReLU y una capa de redimensionamiento con forma $(32, 32, 64)$.
5. *Upsampling*: Para aumentar la resolución espacial de los mapas de características de las imágenes, se utilizan cuatro bloques convolucionales que incorporan conexiones residuales (*skip connections*) y capas UpSampling2D de tamaño 2×2 . Después de los tres primeros bloques convolucionales se utilizan las capas UpSampling2D. Las conexiones residuales se establecen de la siguiente manera: al primer bloque convolucional ingresa la concatenación de la salida del MLP con la salida del último bloque convolucional de *downsampling*. Al segundo bloque convolucional entra la concatenación de la salida del primer bloque convolucional con la salida del tercer bloque de *downsampling*. El tercer bloque recibe como entrada la concatenación de la salida del segundo bloque convolucional con la salida del segundo bloque de *downsampling*. Y en el cuarto bloque convolucional entra la concatenación de la salida del tercer bloque con la salida del primer bloque convolucional de *downsampling*. Este proceso entrega un mapa de características de tamaño $256 \times 256 \times 64$.
6. Salida: En esta última etapa se concatena la imagen ruidosa de entrada, la imagen de baja resolución de la entrada y el mapa de características entregado por el proceso de *upsampling*. Finalmente, la señal pasa por una capa convolucional con un único filtro y con kernel de tamaño 1×1 , entregando un tensor resultado de tamaño $256 \times 256 \times 1$.

Compilación: El modelo utiliza el optimizador Adam con una tasa de aprendizaje de 0,001. La función de pérdida empleada es el *Mean Squared Error*.

Entrenamiento: Para el proceso de entrenamiento se definieron 500 *time steps* (marcas temporales), y para cada marca temporal se realiza un entrenamiento completo del modelo. Durante el entrenamiento se guardan los modelos entrenados en diferentes marcas de tiempo; en este caso se guardó el modelo en las marcas de tiempo 400, 300, 200, 100, 50, 10 y 1. Esto se hace debido a que para el proceso de inferencia se utilizará el modelo correspondiente en el tramo en el cual se esté realizando el proceso iterativo de reconstrucción: el modelo predecirá el ruido en el *time step* 350 usando el modelo entrenado y almacenado en la marca temporal 300.

Durante el entrenamiento es en donde se realiza el proceso de difusión hacia adelante, en el cual en cada marca temporal se irá agregando ruido a la imagen hasta degradarla completamente. La Figura 3.5 muestra cómo se visualiza ese proceso con imágenes MRI, en la cual se observa la degradación progresiva a lo largo de diferentes marcas temporales. Se inicia con el paso cero, en donde se tiene la imagen original en alta resolución, y se avanza por las marcas temporales 10 y 50, hasta llegar al paso 499, donde la imagen se ha degradado casi por completo en ruido.

Lo primero que se hace es agregar ruido al conjunto completo de imágenes en alta resolución, obteniendo así tanto las imágenes ruidosas como el mapa de ruido agregado. Luego se entrena el modelo, al cual se le proporcionan como datos de entrada las imágenes ruidosas (obtenidas en el momento anterior a partir de las imágenes de alta resolución), las imágenes en baja resolución y la marca temporal, mientras que el objetivo a predecir es el ruido agregado. Y este proceso se repite para las 500 marcas temporales definidas.

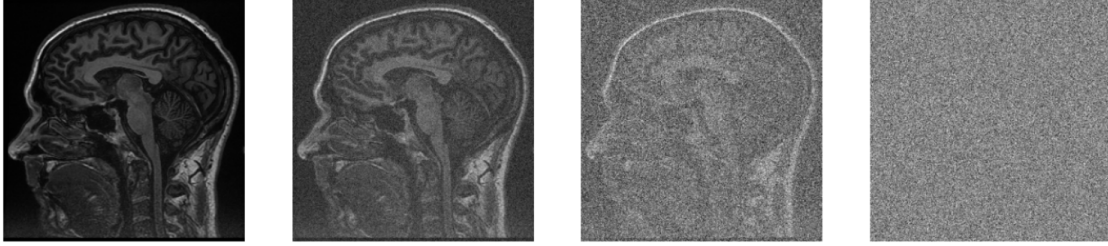


Figura 3.5: Degradación progresiva de una imagen MRI.

El modelo se entrenó para cada *time step* en 10 épocas con *batch size* igual a 64.

Eliminación de ruido: En un proceso iterativo de refinamiento, que comienza en el paso 500, en donde se utiliza el modelo entrenado correspondiente a cada tramo específico de las marcas temporales, se proporciona inicialmente al modelo una imagen completamente ruidosa, la imagen en baja resolución (la cual se desea mejorar) y la marca temporal, con el fin de que el modelo infiera el ruido que debe ser eliminado. En cada iteración, el modelo predice el ruido a eliminar de forma progresiva, desde el paso 500 hasta el paso 1, hasta lograr la reconstrucción completa de la imagen en alta resolución.

Capítulo 4

Resultados y Discusión

4.1. Modelos 3D

Antes de iniciar con los resultados obtenidos en cada arquitectura 3D, se requiere aclarar que, con el fin de evaluar visualmente la capacidad de reconstrucción de imágenes de alta resolución a partir de entradas de baja resolución, se realizó una prueba con un volumen completo de resonancia magnética que no había sido utilizado durante el entrenamiento ni la validación. En cada arquitectura se utilizó un volumen diferente para esta evaluación, asegurando así una prueba justa e independiente del aprendizaje previo del modelo. Estos volúmenes se procesaron generando una versión de baja resolución mediante reducción espacial y posterior interpolación, y luego se dividieron en parches tridimensionales de tamaño $64 \times 64 \times 64$. Cada parche fue pasado a través del modelo correspondiente para obtener su versión reconstruida, tras lo cual se reensamblaron los parches procesados para reconstruir el volumen completo. Finalmente, se comparó visualmente la reconstrucción con la imagen original de alta resolución y la entrada de baja resolución, mostrando cortes representativos del volumen en el plano sagital. Esta metodología permitió observar cualitativamente el desempeño de cada arquitectura más allá de las métricas cuantitativas.

Además del análisis cualitativo, se llevó a cabo una evaluación cuantitativa del desempeño de cada arquitectura empleando cinco estudios completos de resonancia magnética que no fueron utilizados en las etapas de entrenamiento ni validación. Estos estudios permitieron medir la capacidad general de generalización de cada modelo ante nuevos datos. A cada arquitectura se le suministraron los mismos cinco volúmenes completos, procesados previamente para generar su versión de baja resolución, con el fin de garantizar condiciones de prueba equitativas. Luego de reconstruir cada volumen tridimensional a partir de sus respectivos parches procesados por el modelo, se calcularon métricas estándar de calidad de imagen como PSNR, SSIM y MSE, comparando el volumen reconstruido con su correspondiente volumen original de alta resolución.

4.1.1. UNET 3D

El entrenamiento tardó 4 horas y 13 minutos en ejecutarse. Se observa como durante las 50 épocas de entrenamiento, el modelo mejora progresivamente en la reconstrucción de los volúmenes de imagen. Se alcanzaron unos valores de pérdida de 0,0091, PSNR de 37,85 dB y SSIM de 0,9668. En validación las métricas alcanzadas también fueron consistentes, con una pérdida de 0,0101, un PSNR de 36,99 dB y un SSIM de 0,9632.

En la Figura 4.1 se aprecia una tendencia clara de convergencia. La curva de pérdida disminuye de forma constante en entrenamiento y validación, reflejando una buena capacidad de generalización del modelo. En cuanto a las curvas de PSNR y SSIM muestran un incremento progresivo a lo largo de las épocas, alcanzando valores altos sin signos de sobreajuste.

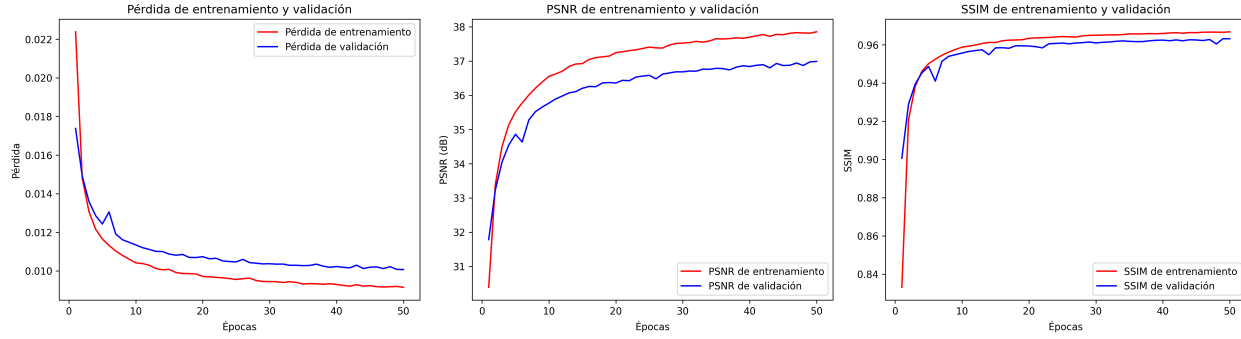


Figura 4.1: Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo UNET 3D.

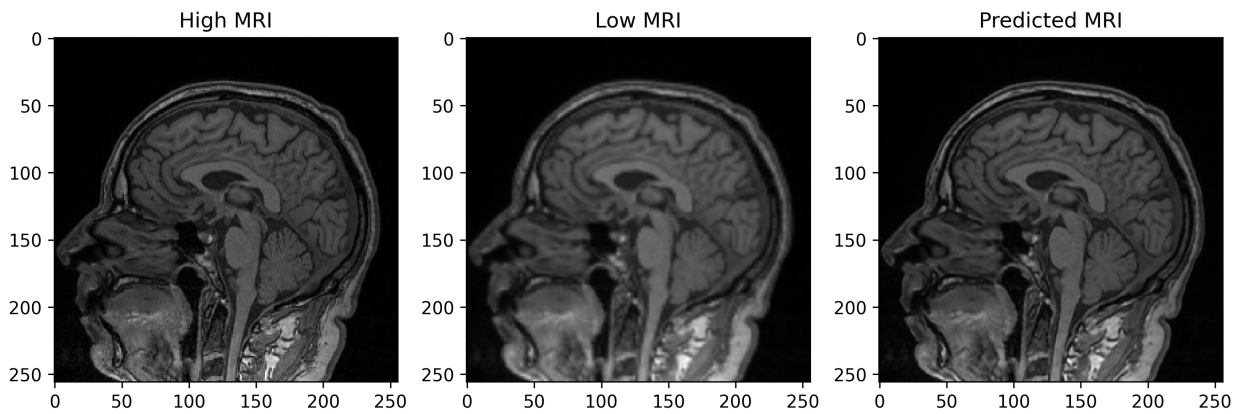


Figura 4.2: Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo UNET 3D.

La Figura 4.2 muestra un corte representativo de un estudio completo de resonancia magnética que fue procesado por el modelo. En ella se compara la imagen de entrada de baja resolución, la imagen reconstruida por el modelo y la imagen original de alta resolución. Se observa que el modelo logra recuperar detalles anatómicos relevantes, como la definición de los contornos corticales y de estructuras internas, mejorando notablemente la nitidez con respecto a la imagen degradada. Aunque se evidencian ligeras diferencias en regiones con texturas finas, la salida generada se aproxima visualmente a la imagen original, lo que refleja una buena capacidad de reconstrucción. Esta apreciación visual se ve respaldada por los valores obtenidos para todo el volumen reconstruido, con un PSNR de 35,89 dB y un SSIM de 0,9579, lo que indica una alta similitud estructural y una recuperación precisa de la información espacial.

Estudio MRI	PSNR (dB)	SSIM	MSE
1	37,33	0,9648	0,000185
2	34,46	0,9571	0,000358
3	30,51	0,9232	0,000889
4	37,86	0,9671	0,000164
5	35,89	0,9579	0,000257

Cuadro 4.1: Análisis comparativo *Gold Standard* con estudios MRI de prueba para el modelo UNET 3D.

En el Cuadro 4.1 se observa como esta arquitectura mostro un desempeño solido y consistente en los

cinco estudios evaluados. Sus valores de PSNR oscilan entre 30,51 dB y 37,86 dB, y presentan una media notablemente alta, lo que indica una buena fidelidad en la reconstrucción volumétrica respecto al original. El índice SSIM se mantiene elevado ($\geq 0,9232$), lo que sugiere que la estructura interna de las imágenes fue conservada de manera efectiva. Los valores de MSE son relativamente bajos, reforzando la precisión global del modelo. Aunque en el tercer estudio se observa una caída en el PSNR y un aumento en el error cuadrático medio, el desempeño general refleja una arquitectura robusta, con una buena capacidad para generalizar sobre volúmenes no vistos previamente.

4.1.2. EDSR 3D

El entrenamiento tuvo una duración de 8 horas y 24 minutos. Durante las 50 épocas, las métricas mejoraron progresivamente hasta estabilizarse. Se obtuvieron valores de pérdida de 0,0085, PSNR de 38,61 dB y SSIM de 0,9704. En la fase de validación se alcanzaron resultados de 0,0095 en pérdida, 37,62 dB en PSNR y 0,9664 en SSIM.

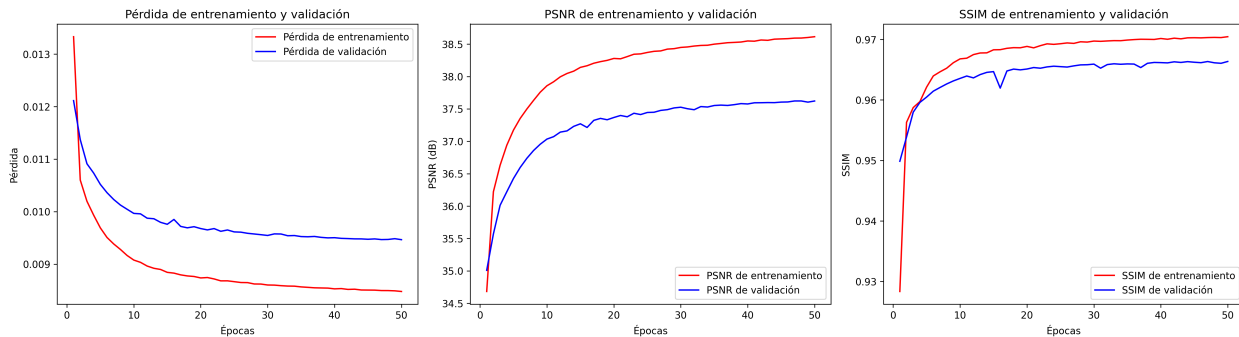


Figura 4.3: Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo EDSR 3D.

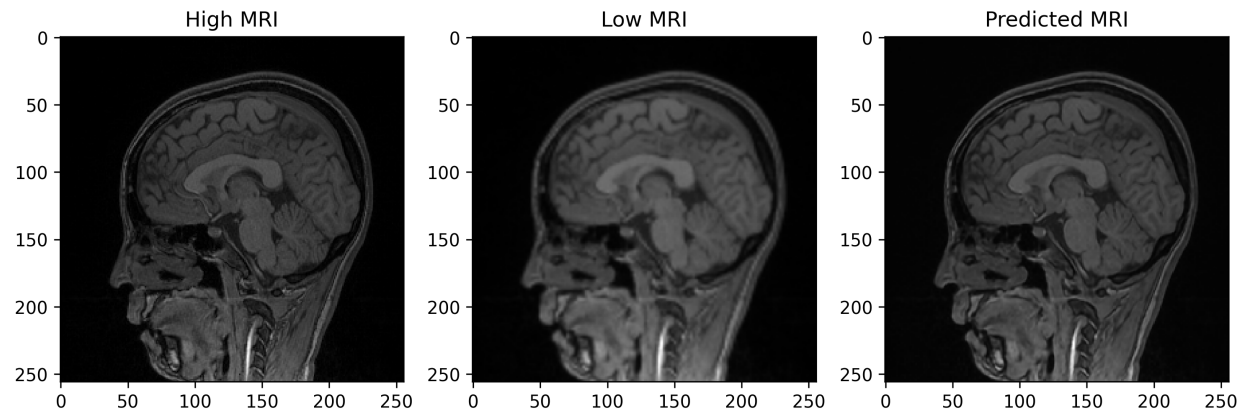


Figura 4.4: Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo EDSR 3D.

Gráficamente, en la Figura 4.3 se muestra la evolución de las métricas. Se observa una disminución constante de la pérdida en el conjunto de entrenamiento y el de validación, lo que indica un aprendizaje constante sin signos importantes de sobreajuste. De igual manera, el PSNR y el SSIM aumentan sostenidamente, hasta alcanzar una meseta en las últimas épocas. Aunque los valores de validación son un poco inferiores a los de

entrenamiento, dicha diferencia es pequeña y sugiere que el modelo logra una buena capacidad de generalización.

La Figura 4.4 presenta un corte representativo de un estudio MRI completo, procesado por el modelo. Aquí se comparan la imagen de baja resolución, la imagen generada por el modelo y la imagen original de alta resolución. A simple vista, se aprecia una mejora considerable en la calidad visual de la imagen reconstruida en comparación con la imagen de entrada. La versión generada muestra contornos más definidos y una mayor nitidez general, lo que facilita la identificación de las estructuras anatómicas presentes. Si bien persisten diferencias sutiles respecto a la imagen original, el resultado producido por el modelo se aproxima de forma convincente, evidenciando su capacidad para recuperar detalles importantes y mejorar la percepción visual del volumen original. Esta mejora se ve respaldada por los resultados obtenidos para todo el volumen reconstruido, con un PSNR de 38,26 dB y un SSIM de 0,9697, lo que refleja una alta fidelidad en la reconstrucción espacial y estructural.

Estudio MRI	PSNR (dB)	SSIM	MSE
1	38,03	0,9678	0,000157
2	34,75	0,9601	0,000335
3	30,64	0,9266	0,000863
4	38,26	0,9697	0,000149
5	36,12	0,9607	0,000244

Cuadro 4.2: Análisis comparativo *Gold Standard* con estudios MRI de prueba para el modelo EDSR 3D.

Como se puede observar en el Cuadro 4.2 la EDSR 3D fue la arquitectura que logró los puntajes más altos de PSNR entre todos los modelos evaluados, alcanzando hasta 38,26 dB, con un rendimiento particularmente fuerte en los estudios 1 y 4. Los valores de SSIM también se mantuvieron por encima de 0,92 en todos los casos, lo que evidencia una conservación estructural sobresaliente. Esta arquitectura también logró algunos de los valores más bajos de MSE, indicando una menor discrepancia promedio entre los voxels reconstruidos y los originales.

4.1.3. cGAN 3D

El entrenamiento tuvo un tiempo de ejecución de 5 horas y 3 minutos. A diferencia de modelos anteriores, esta arquitectura generativa adversarial presentó un comportamiento más inestable en las métricas, caracterizado por oscilaciones propias de la dinámica entre el generador y el discriminador. No se observó una convergencia clara o estabilización sostenida, aunque se lograron picos de alto rendimiento. En el conjunto de entrenamiento, se alcanzaron valores máximos de 33,28 dB en PSNR y 0,9411 en SSIM, mientras que en validación se obtuvieron hasta 32,64 dB en PSNR y 0,9253 en SSIM. En cuanto a las pérdidas, se observaron fluctuaciones en los valores del generador y del discriminador a lo largo del entrenamiento; en los mejores momentos, la pérdida del generador descendió hasta 0,012 y la del discriminador hasta 0,251.

En la Figura 4.5 se aprecia el comportamiento mencionado anteriormente. Se observan oscilaciones marcadas en la pérdida del generador, mientras que la pérdida del discriminador tiende a disminuir progresivamente, con algunos repuntes ocasionales. Esto sugiere un juego dinámico entre ambos componentes a lo largo del entrenamiento. En cuanto a las métricas de PSNR y SSIM, se evidencia una mayor variabilidad en el conjunto de validación, con picos y caídas abruptas, en contraste con un comportamiento algo más estable en el conjunto de entrenamiento.

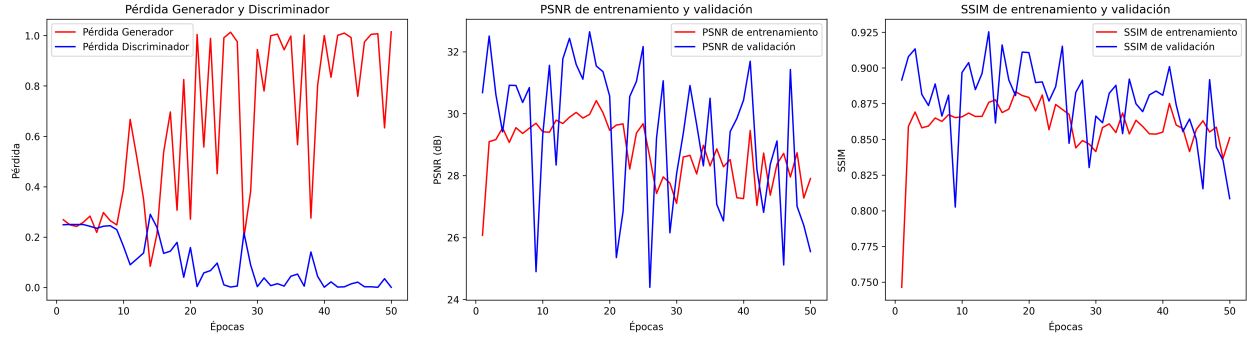


Figura 4.5: Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo cGAN 3D.

Estudio MRI	PSNR (dB)	SSIM	MSE
1	28,15	0,8341	0,001532
2	27,85	0,8307	0,001639
3	25,72	0,7685	0,002682
4	29,25	0,8501	0,001189
5	28,93	0,8400	0,001280

Cuadro 4.3: Análisis comparativo *Gold Standard* con estudios MRI de prueba para el modelo cGAN 3D.

En contraste con los demás modelos, cGAN 3D presentó métricas notablemente más bajas en los cinco estudios evaluados. En el Cuadro 4.3 Los valores de PSNR no superan los 29,25 dB, y el SSIM se encuentra en un rango significativamente inferior (entre 0,7685 y 0,8501), lo cual indica una menor capacidad para preservar las características estructurales del volumen original. Además, el MSE en este modelo es considerablemente más alto, lo que revela mayores errores en la reconstrucción voxel a voxel. Si bien las GANs tienen el potencial de generar resultados visualmente realistas, estos resultados reflejan que esta arquitectura aún requiere ajustes adicionales para alcanzar un rendimiento comparable con los modelos deterministas, al menos en términos de métricas cuantitativas.

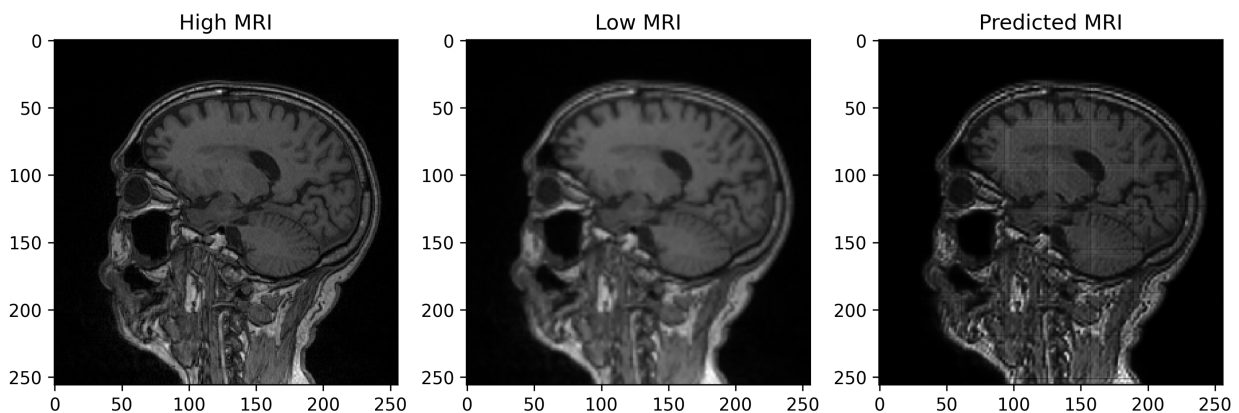


Figura 4.6: Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo cGAN 3D.

La Figura 4.6, presenta un corte representativo de un estudio completo de resonancia magnética tras ser procesado por el modelo. Se comparan la imagen de entrada de baja resolución, la imagen generada

por el modelo y la imagen original de alta resolución. La imagen reconstruida presenta una mejora en la definición de bordes y estructuras internas en comparación con la imagen degradada. Sin embargo, persisten diferencias notables frente a la imagen original, incluyendo artefactos propios de modelos generativos, como patrones cuadrículados y regiones con texturas artificiales. Estas observaciones se reflejan en los resultados cuantitativos obtenidos para todo el volumen, con un PSNR de 27,85 dB y un SSIM de 0,8307, valores que indican una mejora respecto a la entrada degradada, aunque con menor fidelidad que otros modelos evaluados.

4.1.4. SRCNN 3D

El entrenamiento tardó en finalizar un total de 7 horas y 33 minutos. A lo largo de las 50 épocas, se evidenció una mejora progresiva en las métricas de desempeño. El modelo alcanzó en entrenamiento un PSNR de 38,12 dB, un SSIM de 0,9679 y una pérdida de 0,0003. En el conjunto de validación, los resultados también fueron positivos, con una pérdida de 0,0003, un PSNR de 37,24 dB y un SSIM de 0,9641.

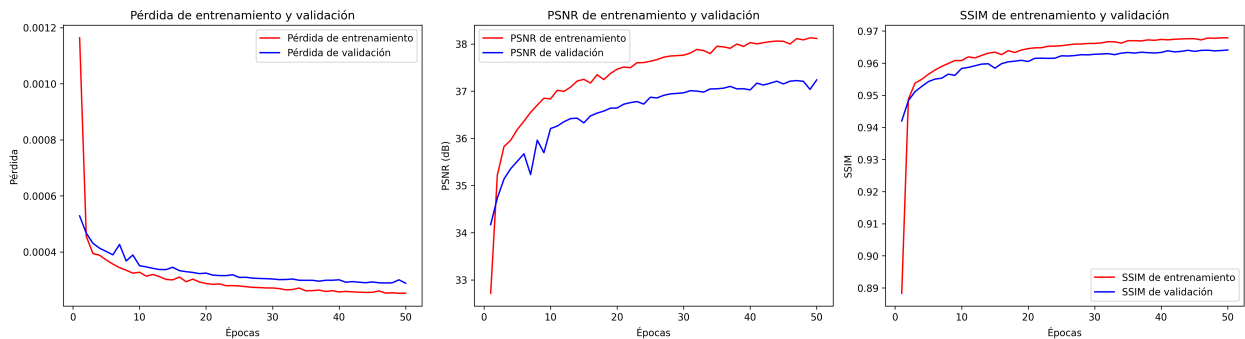


Figura 4.7: Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo SRCNN 3D.

En la Figura 4.7 se aprecia una evolución positiva y consistente del desempeño del modelo. Se observa una disminución progresiva de la pérdida en los conjuntos de entrenamiento y validación, lo que indica un aprendizaje constante sin señales claras de sobreajuste. Asimismo, las métricas PSNR y SSIM muestran un aumento sostenido a medida que avanzan las épocas, estabilizándose hacia el final del entrenamiento. Si bien, los valores alcanzados en el conjunto de validación son ligeramente inferiores a los del entrenamiento, la diferencia es reducida y sugiere que el modelo logra generalizar adecuadamente a nuevos datos.

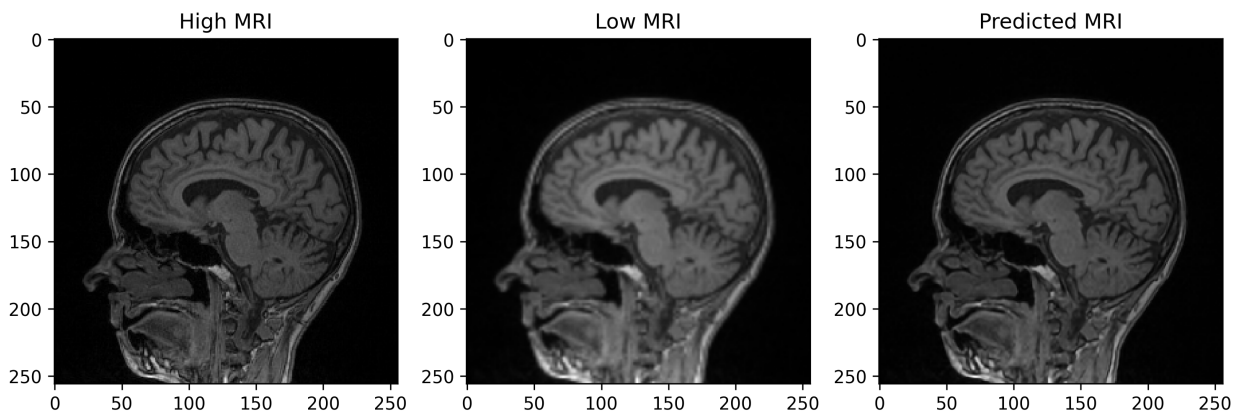


Figura 4.8: Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo SRCNN 3D.

La Figura 4.8, ilustra un corte representativo extraído de un estudio completo de resonancia magnética procesado por el modelo. En esta figura se comparan la imagen de entrada de baja resolución, la imagen generada por el modelo y la imagen original de alta resolución. La imagen reconstruida muestra una mejora visual frente a la imagen de entrada, evidenciada por una mayor nitidez y definición en los contornos, así como por una recuperación parcial de los detalles estructurales. Si bien se mantienen algunas diferencias sutiles con respecto a la imagen original, la salida generada logra representar de manera más precisa los aspectos espaciales y anatómicos del volumen. Esta mejora se sustenta en los resultados cuantitativos obtenidos para todo el volumen reconstruido, con un PSNR de 37,62 dB y un SSIM de 0,9658, lo que refleja una alta calidad en la reconstrucción y preservación de la información estructural.

Estudio MRI	PSNR (dB)	SSIM	MSE
1	37,62	0,9658	0,000173
2	34,60	0,9582	0,000347
3	30,56	0,9242	0,000878
4	38,02	0,9678	0,000158
5	35,99	0,9588	0,000252

Cuadro 4.4: Análisis comparativo *Gold Standard* con estudios MRI de prueba para el modelo SRCNN 3D.

En el Cuadro 4.4 se puede apreciar como la SRCNN 3D ofreció un rendimiento competitivo, logrando valores de PSNR muy cercanos a los de UNet 3D y EDSR 3D en varios estudios. Destaca el estudio 4, con un PSNR de 38,02 dB y un MSE de apenas 0,000 158, que se sitúan entre los mejores resultados obtenidos en toda la evaluación. El índice SSIM también se mantuvo alto y estable, superando el 0,924 en todos los casos. Esto indica que, a pesar de su simplicidad relativa frente a modelos más complejos, SRCNN 3D fue capaz de reconstruir con alta calidad los volúmenes no vistos. Su eficiencia y buen rendimiento la convierten en una opción viable para escenarios donde se requiere un balance entre precisión y bajo costo computacional.

4.2. Modelos 2D

4.2.1. Autoencoder 2D

El entrenamiento tardó en ejecutarse un total de 2 horas y 43 minutos. La pérdida (*loss*) fue de 0,42, SSIM alcanzó un valor de 0,66 y un PSNR de 17,23 dB, lo que refleja una calidad moderada en la reconstrucción de las imágenes. Esto indica que, aunque el autoencoder captura la estructura global de las imágenes, presenta limitaciones en la reconstrucción de detalles finos y texturas biológicas.

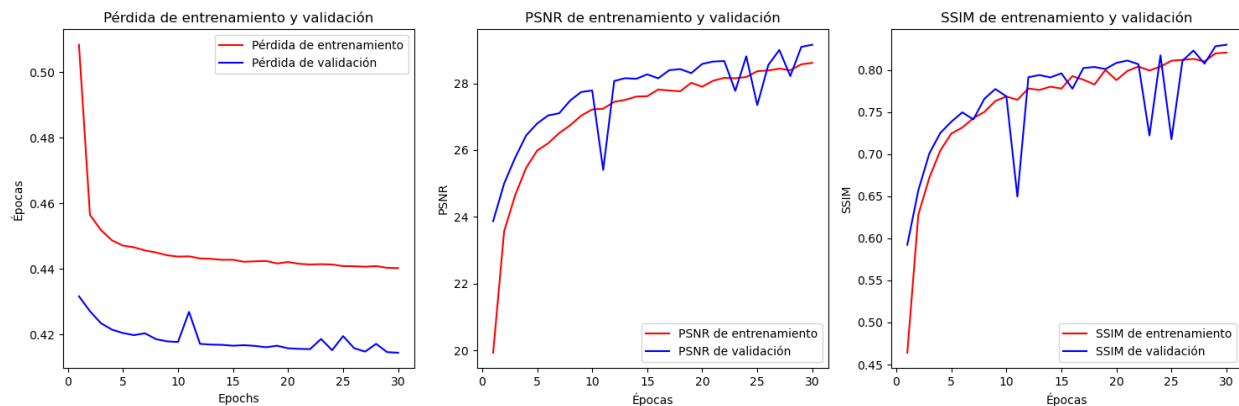


Figura 4.9: Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo Autoencoder 2D.

La Figura 4.9 muestra estabilización temprana de las curvas (alrededor de la época 15), sugiriendo que el modelo alcanzó su máxima capacidad de aprendizaje con la arquitectura actual, particularmente en regiones anatómicas complejas.

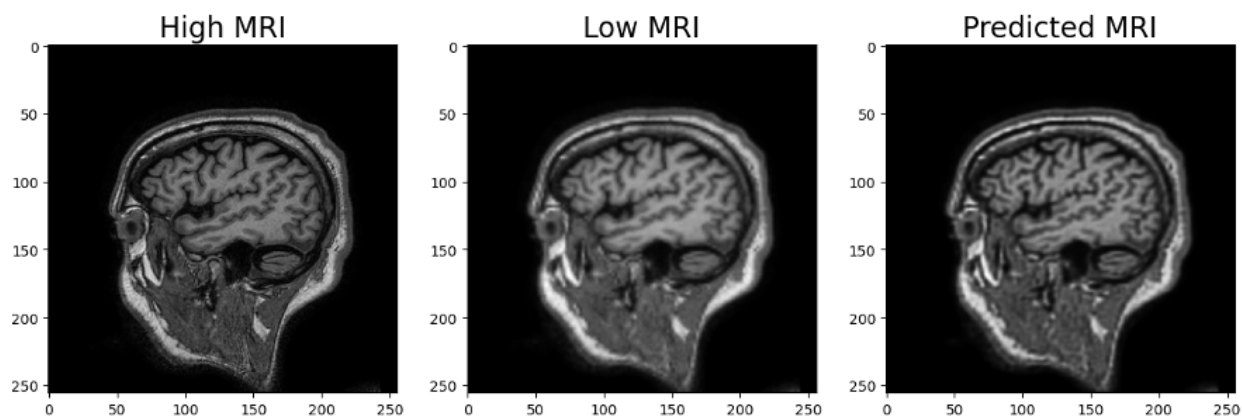


Figura 4.10: Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo Autoencoder 2D.

Las comparaciones visuales presentadas en la Figura 4.10 demuestran que las imágenes reconstruidas conservan la integridad estructural general de las originales. Además, se observa una mejora notable en la definición de los bordes y una recuperación de detalles estructurales; ciertos elementos que aparecían difuminados en las imágenes de baja resolución se visualizan con mayor nitidez tras la reconstrucción.

Slide	PSNR (dB)	SSIM	MSE
3105	51,5293	0,5869	0,4572
2302	78,7735	0,9747	0,0009
24	86,2222	0,9828	0,0002
285	78,6514	0,9808	0,0009
2326	78,7901	0,9763	0,0009
2770	83,9159	0,9869	0,0003
2385	59,5338	0,6962	0,0724
2663	75,4681	0,9682	0,0018
1689	78,1954	0,9642	0,0010
192	53,0538	0,6560	0,3219

Cuadro 4.5: Análisis comparativo *Gold Standard* con imágenes MRI de prueba para el modelo Autoencoder 2D.

Como se observa en el Cuadro 4.5, los resultados de la arquitectura autoencoder muestran un rendimiento variable en las imágenes evaluadas. Los valores de PSNR oscilan entre aproximadamente 51,5 y 86,2 dB, indicando que mientras algunas imágenes se reconstruyen con alta calidad y poco ruido (valores altos de PSNR), otras presentan mayor distorsión. Los valores de SSIM varían significativamente, desde 0,5869 hasta 0,9869, lo que sugiere que la preservación de la estructura y detalles visuales es inconsistente: algunas imágenes mantienen bien la calidad visual, mientras que otras pierden detalles importantes. Además, los valores de MSE reflejan esta variabilidad, con errores medios cuadrados que van desde valores muy bajos (0,0002) hasta valores altos (0,4572 y 0,3219), indicando que en ciertos casos el autoencoder no logra reconstruir con precisión la imagen original. En conjunto, estos datos sugieren que el autoencoder puede ser efectivo para ciertas imágenes, pero su desempeño no es tan robusto ni uniforme como otras arquitecturas.

4.2.2. VDSR 2D

La implementación de la arquitectura VDSR para la super-resolución concluyó su fase de entrenamiento en 10 horas y 59 minutos, logrando indicadores de rendimiento robustos.

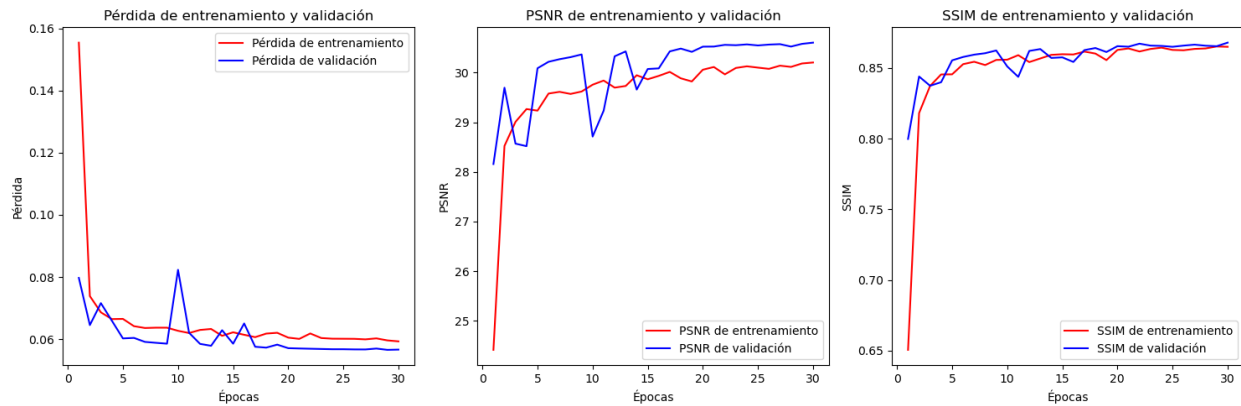


Figura 4.11: Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo VDSR 2D.

Se obtuvo un valor final de pérdida de 0,056, lo que evidencia una adecuada convergencia del modelo durante el entrenamiento, minimizando el error entre las imágenes predichas y las de referencia. Este buen ajuste se ve refrendado por un SSIM de 0,8663 y PSNR de 30,59 dB. Estas métricas cuantitativas indican una alta calidad en la reconstrucción de las imágenes, con una excelente preservación de las estructuras anatómicas y una significativa reducción de ruido y artefactos en comparación con las imágenes de entrada de baja resolución.

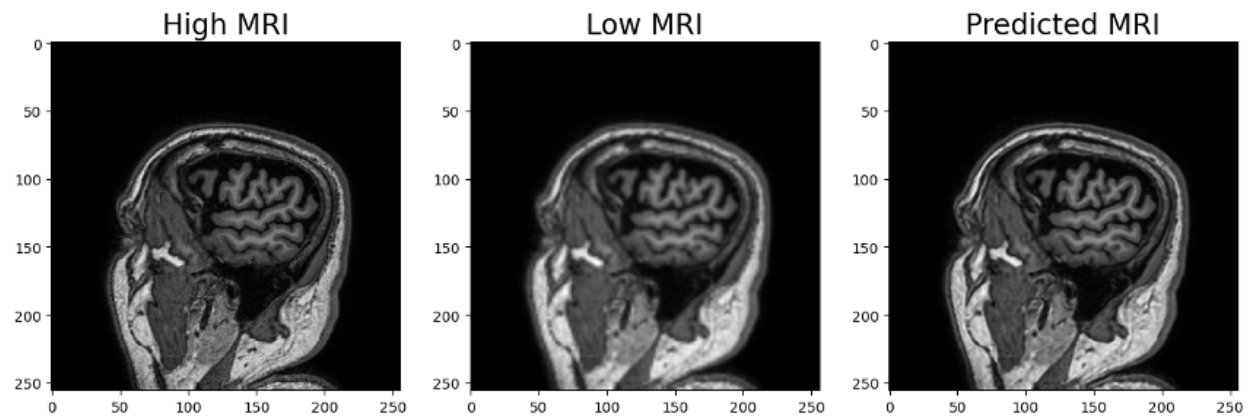


Figura 4.12: Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo VDSR 2D.

Visualmente, los resultados de la arquitectura VDSR demuestran su eficacia en la mejora de la calidad de las imágenes MRI, como se puede apreciar en la Figura 4.12. La imagen de la derecha (*Predicted MRI*) es la salida generada por esta VDSR, la cual exhibe una notable recuperación de detalles y una claridad superior en comparación con la imagen de baja resolución, acercándose considerablemente a la calidad de la imagen original de alta resolución. Esto subraya la capacidad del modelo para aprender y aplicar las transformaciones necesarias para la superresolución de manera efectiva.

Como se observa en el Cuadro 4.6, los resultados de la arquitectura VDSR muestran un desempeño sólido y consistente en la mayoría de las imágenes, con valores de PSNR que oscilan entre aproximadamente 60,6

Slide	PSNR (dB)	SSIM	MSE
3096	64,9938	0,9892	0,0206
2687	77,8380	0,9836	0,0011
1288	60,6374	0,9921	0,0561
2616	76,9056	0,9755	0,0013
2612	77,1266	0,9757	0,0013
1822	77,2316	0,9783	0,0012
3002	68,0621	0,9942	0,0102
2040	76,8563	0,9799	0,0013
3125	69,9652	0,9944	0,0066
1086	85,4723	0,9908	0,0002

Cuadro 4.6: Análisis comparativo *Gold Standard* con imágenes MRI de prueba para el modelo VDSR 2D.

y 85,5 dB, indicando una alta calidad de reconstrucción en términos de relación señal-ruido. Los valores de SSIM son muy elevados, todos por encima de 0,97, lo que refleja una excelente preservación de la estructura y detalles visuales en las imágenes reconstruidas. Además, los valores de MSE son muy bajos (todos menores a 0,06), lo que confirma que el error medio cuadrático entre las imágenes originales y reconstruidas es mínimo. En conjunto, estos indicadores sugieren que VDSR es eficaz para mejorar la resolución manteniendo una fidelidad visual alta en las imágenes procesadas.

4.2.3. SRGAN 2D

La implementación de la arquitectura SRGAN para la superresolución completó su ciclo de entrenamiento en 10 horas y 41 minutos, alcanzando métricas de rendimiento sólidas. Se registró un valor final de pérdida de 0,0578, lo que indica una buena convergencia del modelo generativo y del discriminador hacia un equilibrio durante el proceso de entrenamiento adversarial. Este ajuste se complementa con SSIM de 0,8597 y PSNR de 30,31 dB. Estos indicadores cuantitativos sugieren una notable calidad en la reconstrucción de las imágenes, con una adecuada preservación de las características estructurales.

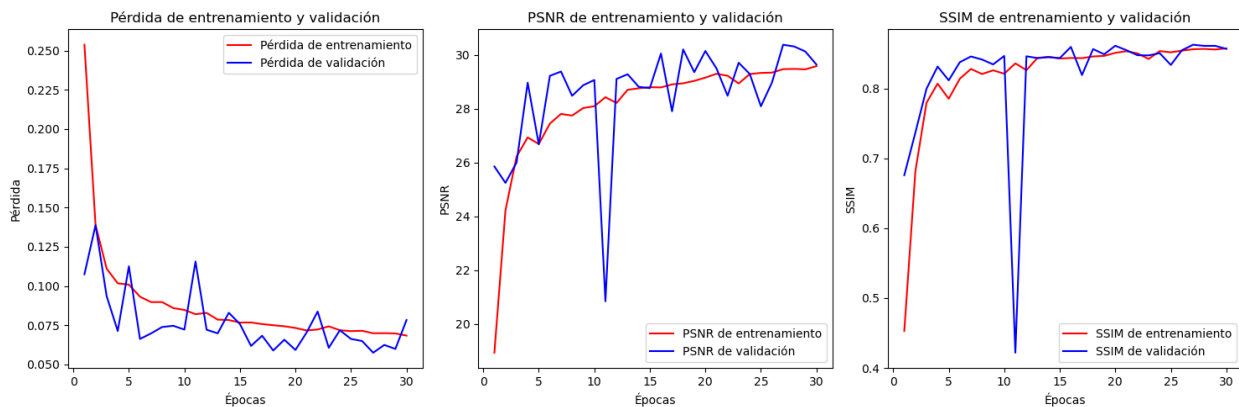


Figura 4.13: Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo SRGAN 2D.

Desde una perspectiva visual, los resultados generados por el modelo SRGAN evidencian su capacidad para realzar la calidad de las imágenes MRI, como se ilustra en la Figura 4.14. La imagen a la derecha (*Predicted MRI*) es la resultante de la aplicación del modelo SRGAN. Esta última demuestra una mejora apreciable en la nitidez y en la definición de los detalles anatómicos en comparación con la imagen de baja resolución, logrando una reconstrucción que, si bien puede introducir algunas texturas propias de las GANs, se aproxima a la calidad de la imagen original. Las curvas de entrenamiento y validación, como las mostradas

en Figura 4.13, muestran que la arquitectura se estabiliza a partir de la época 15.

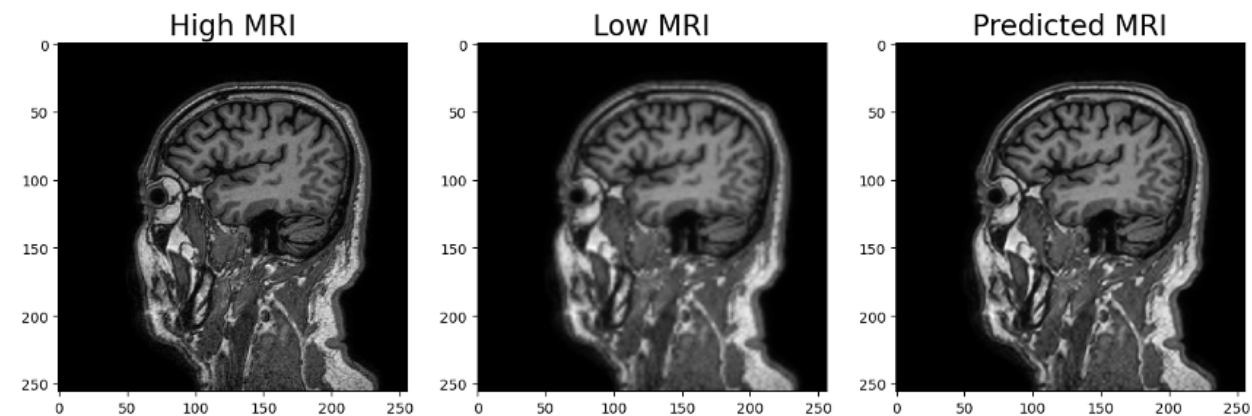


Figura 4.14: Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo SRGAN 2D.

Slide	PSNR (dB)	SSIM	MSE
0	80,27	0,9857	0,0128
1	58,44	0,9869	0,1726
2	63,98	0,9846	0,0935
3	94,61	0,9870	0,0009
4	79,67	0,9757	0,0153

Cuadro 4.7: Análisis comparativo *Gold Standard* con imágenes MRI de prueba para el modelo SRGAN 2D.

Los resultados del SRGAN (Cuadro 4.7), muestran un desempeño muy bueno en la mayoría de las imágenes evaluadas, con valores de PSNR altos (entre 58,44 y 94,61), lo que indica una reconstrucción con baja distorsión. Los valores de SSIM están muy cercanos a 1 (de 0,9757 a 0,9870), reflejando una excelente preservación de la estructura y detalles visuales. Además, los valores de MSE son bajos en general, evidenciando un error absoluto medio pequeño, salvo en un caso particular (índice 1) que presenta un MSE más alto pero mantiene un SSIM elevado, sugiriendo que, aunque hay mayor diferencia numérica, la calidad perceptual sigue siendo buena. En conjunto, estos resultados indican que el SRGAN logra reconstrucciones de alta calidad y precisión en la mayoría de las imágenes.

4.2.4. DRCN 2D

La implementación de la arquitectura DRCN para la tarea de superresolución se destacó por su eficiencia, completando la fase de entrenamiento en tan solo 3 horas y 18 minutos. Al finalizar, se obtuvieron métricas de rendimiento que demuestran su capacidad: un valor de pérdida de 0,0580, indicativo de una buena minimización del error. El SSIM alcanzó 0,8646 y el PSNR se situó en 30,43 dB. Estos valores cuantitativos reflejan una reconstrucción de alta calidad, con una notable preservación de la integridad estructural de las imágenes y una buena supresión de ruido, lo cual es fundamental para la utilidad diagnóstica de las MRI.

En la evaluación visual de los resultados, como se observa en la Figura 4.15, el modelo DRCN muestra su efectividad en la mejora de la resolución de las imágenes MRI. Se puede apreciar que la imagen predicha exhibe una mejora sustancial en la nitidez y en la definición de los detalles finos en comparación con la entrada de baja resolución, acercándose a la calidad de la imagen original. Las gráficas de evolución de las métricas durante el entrenamiento y la validación, como las presentadas en la Figura 4.16, muestran que el entrenamiento se estabiliza a partir de la época 15.

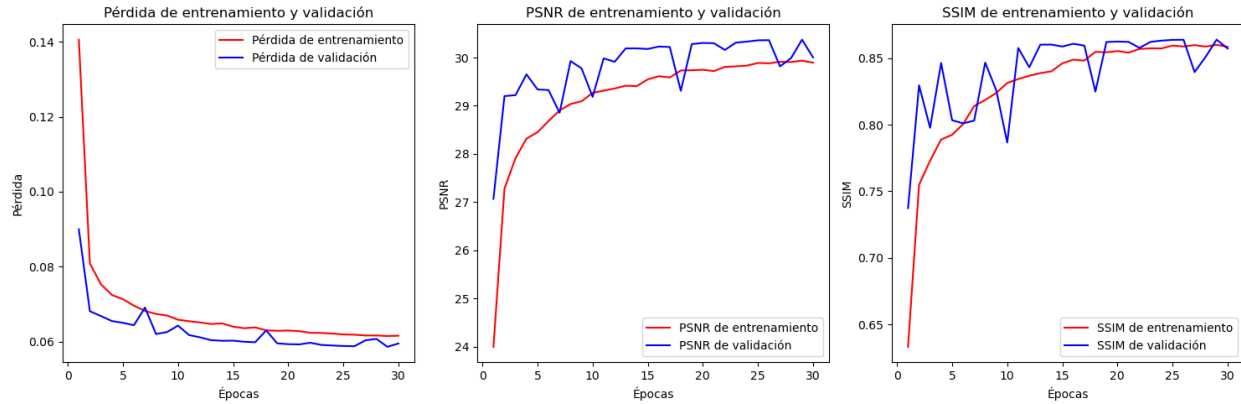


Figura 4.15: Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo DRCN 2D.

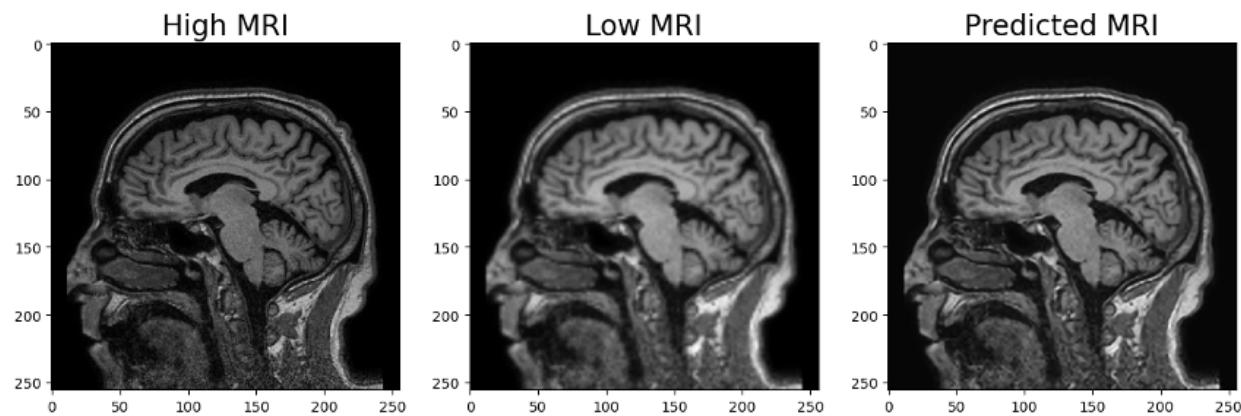


Figura 4.16: Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo DRCN 2D.

Slide	PSNR (dB)	SSIM	MSE
3017	64,6869	0,9900	0,0206
3059	65,1176	0,9864	0,0200
2385	73,0024	0,9917	0,0033
169	65,5327	0,9878	0,0182
227	75,8358	0,9935	0,0017
1402	68,5166	0,9967	0,0091
2881	82,5271	0,9756	0,0004
1118	64,0752	0,9948	0,0254
466	63,5617	0,9854	0,0286
1420	64,2426	0,9939	0,0245

Cuadro 4.8: Análisis comparativo *Gold Standard* con imágenes MRI de prueba para el modelo DRCN 2D.

En el Cuadro 4.8, los 10 ejemplos presentados muestran valores de PSNR superiores a 63 dB, lo que indica que el modelo es capaz de reconstruir imágenes visualmente muy similares a las originales en alta resolución. Además, todos los valores de SSIM están muy cercanos a 1, demostrando que el modelo preserva con alta fidelidad la similitud estructural en la predicción de la imagen. Asimismo, los valores de MSE son muy bajos, reflejando una mínima pérdida de información durante la reconstrucción de las imágenes. Estos

resultados muestran que el modelo utilizado es altamente efectivo para la super-resolución de imágenes, logrando reconstruir imágenes de alta calidad a partir de versiones de baja resolución.

4.2.5. UNET 2D

La implementación de la arquitectura U-Net para la super-resolución de imágenes de resonancia magnética (MRI) requirió un tiempo de entrenamiento de un día, 0 horas y 28 minutos. Al concluir esta fase, se obtuvo un valor de pérdida (*loss*) de 0,4111. En cuanto a las métricas de calidad de imagen, el SSIM alcanzó 0,8682, mientras que el PSNR se situó en 30,68 dB. El análisis de las curvas de entrenamiento y validación, visibles en la Figura 4.17, sugiere que el modelo comienza a mostrar una convergencia estable en su desempeño en el conjunto de validación alrededor de la época 15, momento en el cual la pérdida de validación alcanza un mínimo y las métricas de PSNR y SSIM tienden a estabilizarse. Estos indicadores cuantitativos reflejan una capacidad adecuada del modelo U-Net para reconstruir las imágenes, manteniendo una buena similitud estructural y una calidad de señal aceptable.

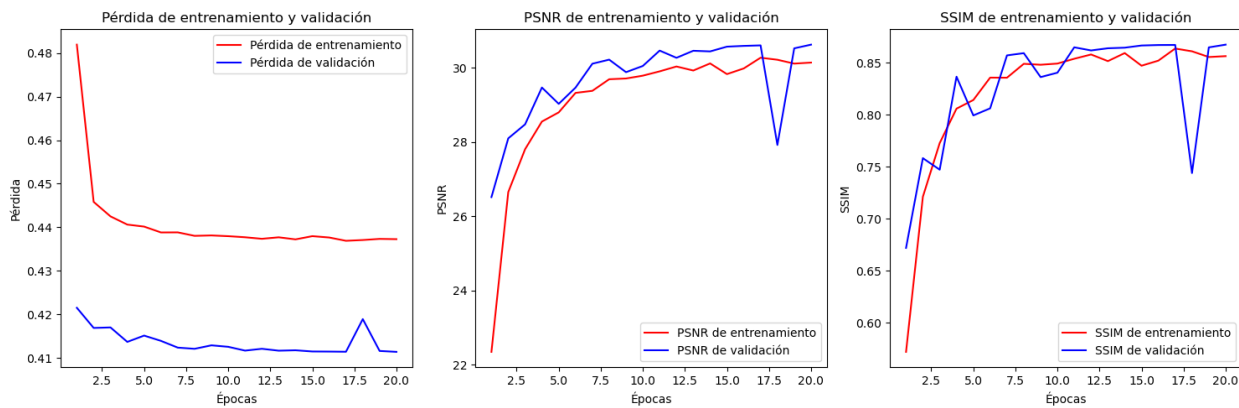


Figura 4.17: Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo UNET 2D.

Visualmente, el desempeño del modelo U-Net en la mejora de la resolución se puede evaluar en la Figura 4.18. La imagen predicha muestra una mejora en la claridad y en la definición de las estructuras anatómicas en comparación con la imagen de baja resolución, demostrando la efectividad del modelo para la tarea de superresolución y su potencial para generar imágenes con detalles restaurados, aproximándose a la calidad de la imagen original.

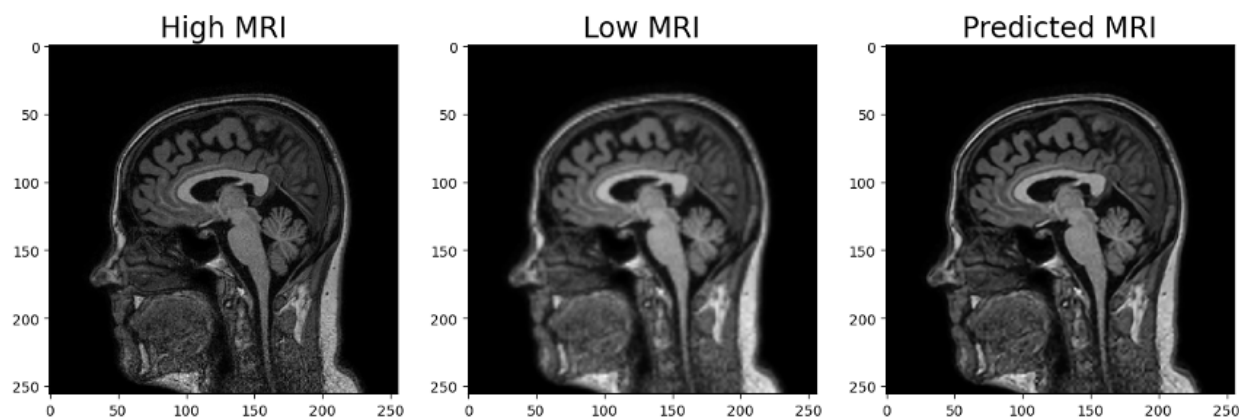


Figura 4.18: Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo UNET 2D.

Slide	PSNR (dB)	SSIM	MSE
505	48,7452	0,4757	0,8681
2470	50,1957	0,5564	0,6216
695	57,7557	0,7686	0,1090
1768	94,1125	0,9859	0,0000
320	80,6911	0,9810	0,0006
510	48,8657	0,4744	0,8443
2831	84,0027	0,9817	0,0003
1206	41,5154	0,1462	4,5871
226	65,1318	0,8976	0,0199
2319	80,5356	0,9826	0,0006

Cuadro 4.9: Análisis comparativo *Gold Standard* con imágenes MRI de prueba para el modelo UNET 2D.

El Cuadro 4.9, muestra que la arquitectura UNET muestra un rendimiento variable: en varios casos logra reconstrucciones de alta calidad con valores elevados de PSNR (hasta 94 dB), SSIM cercanos a 1 y errores MSE mínimos, indicando gran fidelidad a las imágenes originales; sin embargo, en otras imágenes presenta resultados pobres con PSNR bajos (alrededor de 41-50 dB), SSIM muy bajos y MSE altos, reflejando pérdidas significativas y menor precisión, lo que evidencia que su efectividad depende de las características específicas de cada imagen.

4.2.6. SRResNet 2D

El entrenamiento del modelo tomó 22 minutos, alcanzado con los datos de validación un valor de pérdida (MAE) de 0,001 01, un PSNR de 61,58 dB y un SSIM de 0,9954.

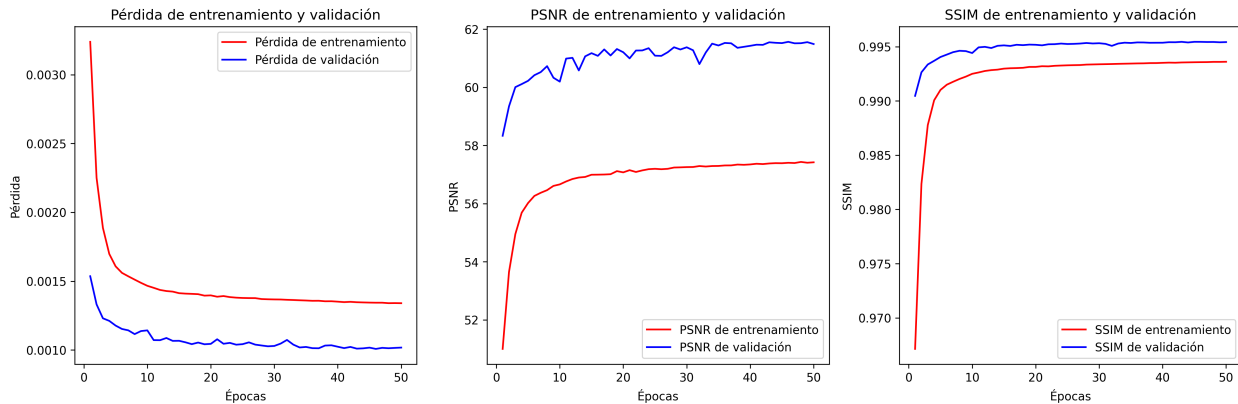


Figura 4.19: Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo SRResNet 2D.

En la Figura 4.19, el gráfico de pérdida de entrenamiento y validación (izquierda) muestra que el modelo ha aprendido progresivamente de manera regular, mostrando un mejor comportamiento con los datos de validación que de entrenamiento, lo que indica que no hay señales de sobreajuste. El gráfico de PSNR (centro) indica que el modelo alcanza valores altos en las primeras épocas y luego entra en fase de refinamiento; al igual que el gráfico de pérdida, se tienen mejores resultados para los datos de validación. La gráfica de SSIM (derecha) evidencia un incremento acelerado del desempeño del modelo para luego estabilizarse, comportamiento que se presenta tanto para los datos de entrenamiento como para los de validación. Los resultados observados en la Figura 4.19 sugieren que el modelo tiene una buena capacidad de generalización en la reconstrucción de imágenes.

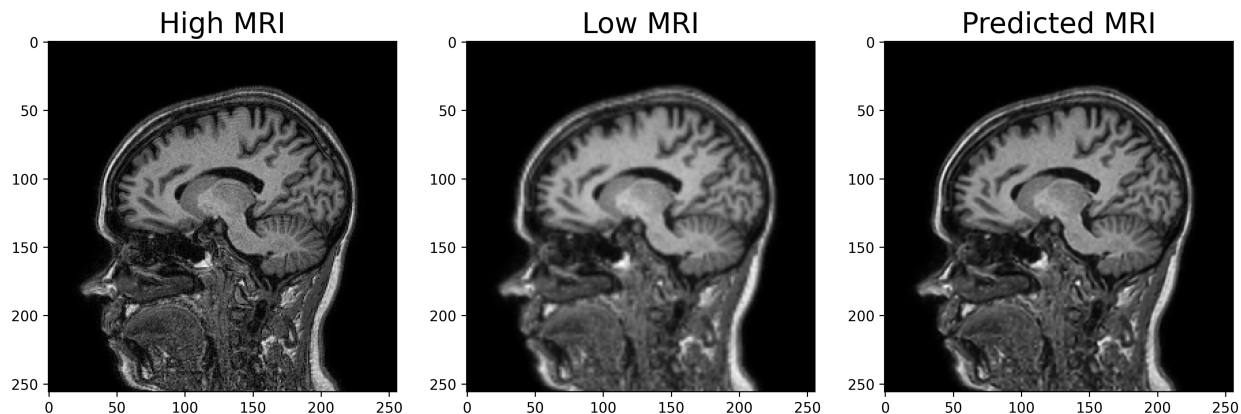


Figura 4.20: Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo SRResNet 2D.

Al calcular el PSNR, el SSIM y el MSE (*Mean Squared Error*) entre la imagen generada por el modelo (Figura 4.20, derecha) y la imagen de referencia en alta resolución (Figura 4.20, izquierda), se obtuvieron los siguientes valores: PSNR de 63,77 dB, SSIM de 0,9994 y MSE de $4,2 \times 10^{-7}$. Estos valores sugieren que el modelo realizó una reconstrucción de alta fidelidad, tanto estructural como perceptual, y ofreció un buen rendimiento computacional con un alto volumen de imágenes de entrenamiento.

Slide	PSNR (dB)	SSIM	MSE
1	64,55	0,9995	$3,51 \times 10^{-7}$
2	64,23	0,9995	$3,78 \times 10^{-7}$
3	63,67	0,9994	$4,29 \times 10^{-7}$
4	63,77	0,9994	$4,20 \times 10^{-7}$
5	63,82	0,9994	$4,14 \times 10^{-7}$
6	63,76	0,9994	$4,20 \times 10^{-7}$
7	63,32	0,9994	$4,66 \times 10^{-7}$
8	64,15	0,9995	$3,84 \times 10^{-7}$
9	64,05	0,9995	$3,94 \times 10^{-7}$
10	65,90	0,9996	$2,57 \times 10^{-7}$

Cuadro 4.10: Análisis comparativo *Gold Standard* con imágenes MRI de prueba para el modelo SRResNet 2D.

El Cuadro 4.10 presenta un análisis comparativo entre las imágenes generadas por el modelo y las imágenes referencia de alta resolución *Gold Standard*, utilizando el conjunto de datos de prueba (desconocido por el modelo). En este análisis se evaluaron las métricas PSNR, SSIM y MSE sobre 10 imágenes de prueba seleccionadas aleatoriamente.

En el Cuadro 4.10, los 10 ejemplos presentan valores de PSNR superiores a 63 dB, lo que indica que el modelo es capaz de reconstruir imágenes visualmente muy similares a las originales en alta resolución. Todos los valores de SSIM son muy cercanos a 1, lo que demuestra que el modelo preserva con alta fidelidad la similitud estructural en la predicción de la imagen. Asimismo, los valores tan bajos de MSE reflejan una mínima pérdida de información durante la reconstrucción.

4.2.7. SRDenseNet 2D

El modelo SRDenseNet fue entrenado durante una hora y 37 minutos. Con los datos de validación alcanzó un valor de pérdida (MAE) de 0,0012, un PSNR de 59,41 dB y un SSIM de 0,9941.

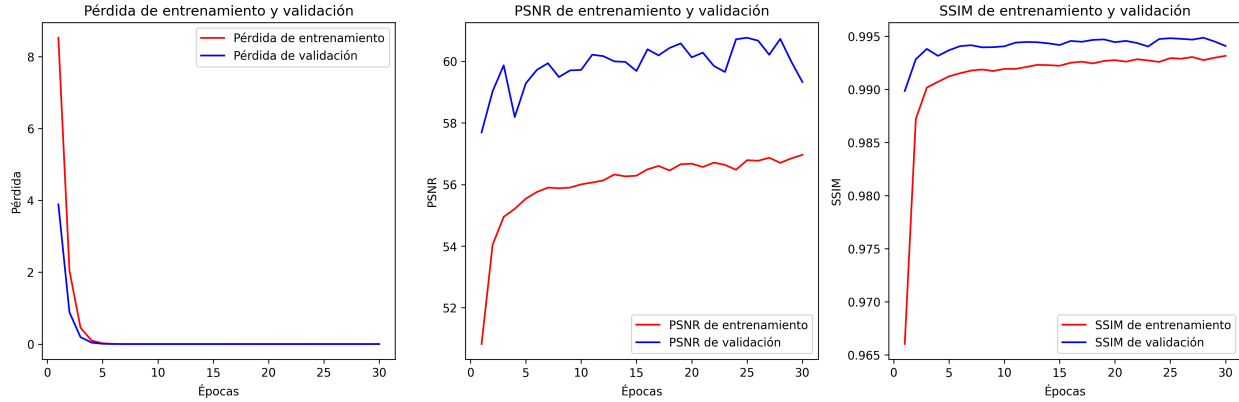


Figura 4.21: Gráficos de Pérdida, PSNR y SSIM para datos de entrenamiento y validación del modelo SRDenseNet 2D.

El gráfico de pérdida (izquierda), ilustrado en la Figura 4.21, muestra una caída brusca y temprana en los valores de MAE, llegando prácticamente a cero. Este comportamiento podría significar inestabilidad en el modelo y un posible sobreajuste temprano. La gráfica de PSNR (centro) presenta una mayor variabilidad para los datos de validación, aunque en general las dos líneas (validación y entrenamiento) muestran incrementos durante todas las épocas. Por último, la gráfica del SSIM (derecha) también muestra un crecimiento temprano en los valores, así como cierta oscilación en la curva correspondiente a los datos de validación. Las gráficas de las métricas PSNR y SSIM muestran mejores resultados en los datos de validación que en los de entrenamiento, esto debido a la aplicación de una regularización fuerte (en este caso L2). De igual forma, se descarta la presencia de sobreajuste y se concluye que el modelo presenta una rápida convergencia, por lo que podría entrenarse en un menor número de épocas.

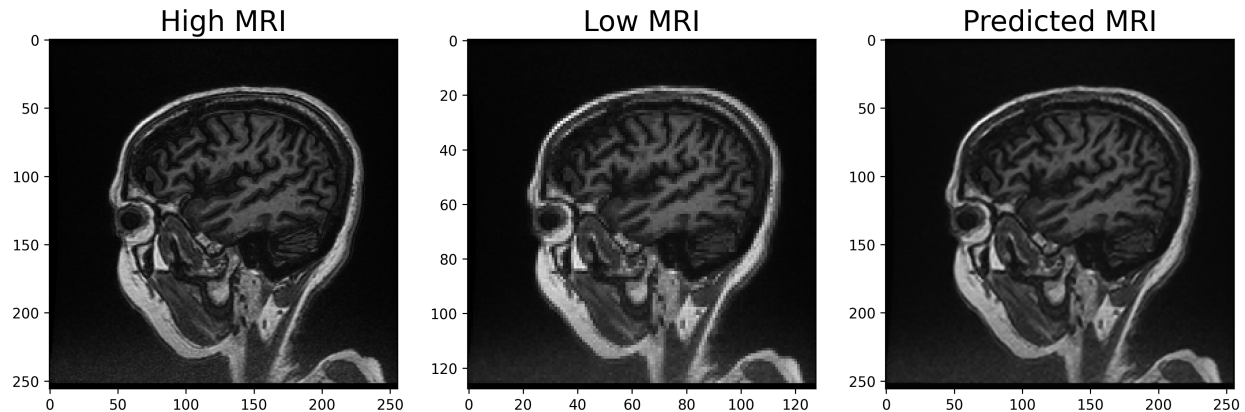


Figura 4.22: Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo SRDenseNet 2D.

La Figura 4.22 muestra la imagen generada por el modelo (derecha) a partir de una imagen de baja resolución (centro). Comparando la imagen generada por el modelo SRDenseNet con la imagen de alta resolución (izquierda), se obtienen las siguientes métricas: PSNR de 43,49 dB, SSIM de 0,9787 y MSE de $4,47 \times 10^{-5}$. Estos valores, junto con la visualización en la Figura 4.22, confirman que la reconstrucción de la imagen mantiene la fidelidad estructural con muy bajo nivel de distorsión.

Slide	PSNR (dB)	SSIM	MSE
1	42,31	0,9738	$5,9 \times 10^{-5}$
2	42,16	0,9715	$6,1 \times 10^{-5}$
3	42,25	0,9717	$5,9 \times 10^{-5}$
4	43,09	0,9726	$4,9 \times 10^{-5}$
5	43,01	0,9728	$5,0 \times 10^{-5}$
6	42,48	0,9738	$5,6 \times 10^{-5}$
7	42,54	0,9751	$5,6 \times 10^{-5}$
8	43,49	0,9787	$4,5 \times 10^{-5}$
9	45,01	0,9837	$3,1 \times 10^{-5}$
10	48,15	0,9905	$1,5 \times 10^{-5}$

Cuadro 4.11: Análisis comparativo *Gold Standard* con imágenes MRI de prueba para el modelo SRDenseNet 2D.

El Cuadro 4.11 muestra valores consistentes en las tres métricas para las 10 imágenes del conjunto de datos. Esta consistencia indica que, en la comparación con *Gold Standard* el modelo siempre presenta buenos resultados al reconstruir la imagen. Todos los valores de PSNR superan los 40 dB, lo que demuestra una reconstrucción visual muy aceptable (visible en la imagen de la derecha de la Figura 4.22). Los buenos valores de SSIM y MSE reflejan una alta similitud estructural entre las imágenes comparadas, con una pérdida de información prácticamente nula.

4.2.8. SR3 2D

El modelo SR3 se entrenó durante 3 horas y 40 minutos. Debido a su diseño conceptual basado en modelos de difusión, SR3 utiliza únicamente datos de entrenamiento. En este caso, el proceso de difusión se repitió 500 veces durante el entrenamiento.

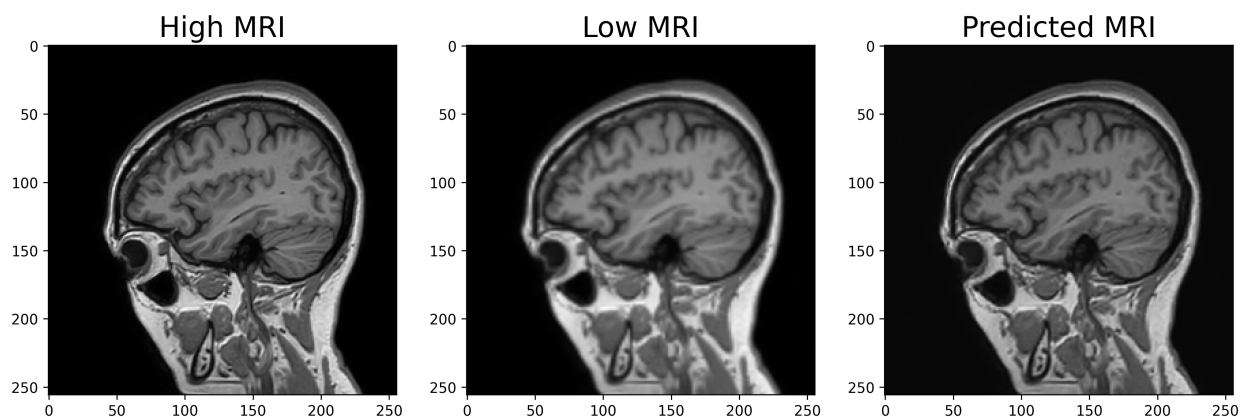


Figura 4.23: Comparación entre imagen de alta resolución, su par en baja resolución, y la imagen generada por el modelo SR3 2D.

La Figura 4.24 muestra las gráficas de pérdida de entrenamiento (MSE) en las marcas temporales 500, 400, 300, 200, 100, 50, 30, 10 y 1, durante el proceso de difusión hacia adelante. En este proceso, el modelo aprendió a predecir el ruido añadido a cada imagen de alta resolución mediante una secuencia de agregación progresiva. En general, durante todos los procesos de entrenamiento, el valor de la pérdida (MSE) inició en un nivel bajo y continuó disminuyendo hasta aproximarse a cero, como se observa en la Figura 4.24. En la mayoría de los pasos de entrenamiento hacia adelante, los valores de pérdida descienden rápidamente durante las primeras iteraciones, hasta estabilizarse a partir de la cuarta época.

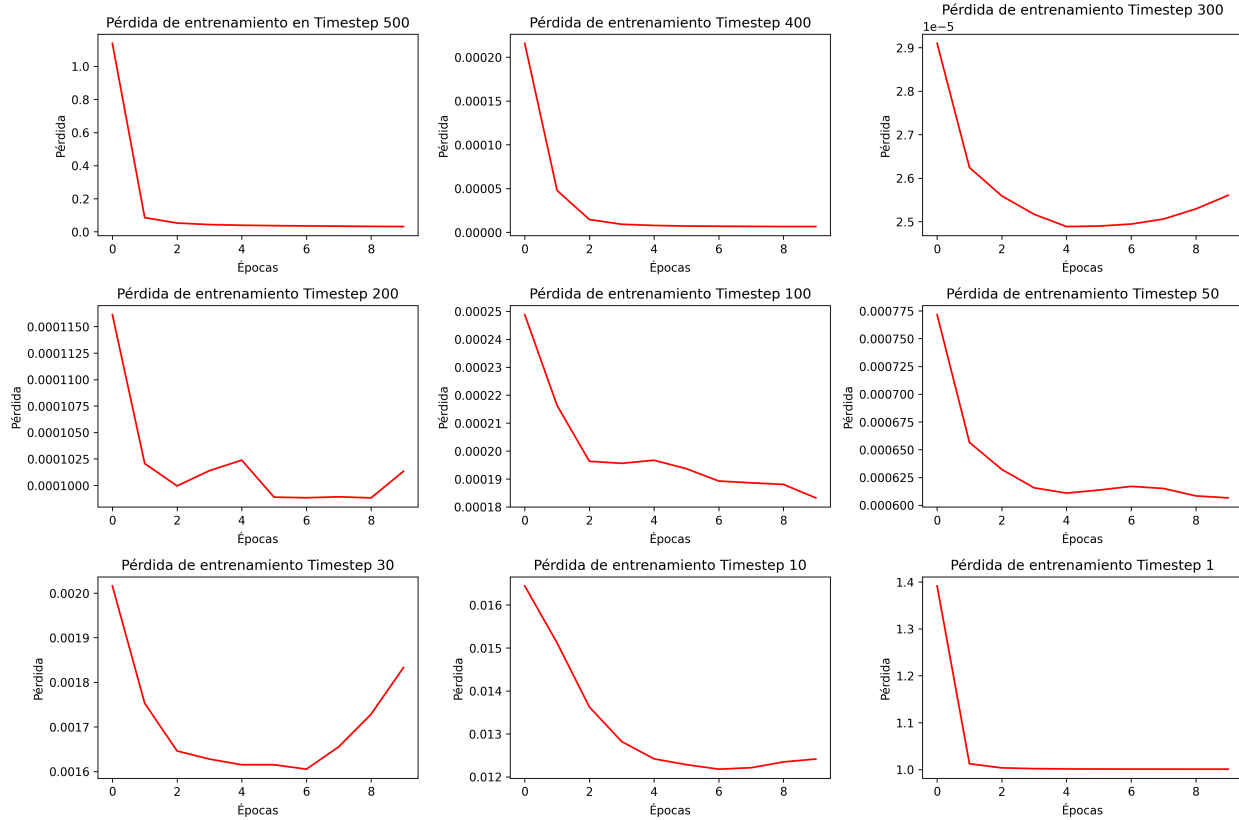


Figura 4.24: Pérdida de entrenamiento en las marcas temporales 500, 400, 300, 200, 100, 50, 30, 10 y 1 para el modelo SR3 2D.

El proceso de inferencia y reconstrucción de imágenes de alta de resolución con el modelo SR3 requerirá más recursos computacionales durante un tiempo considerable, ya que implica realizar inferencias (predicción de ruido) y eliminación de ruido (*denoising*) a lo largo de las marcas temporales en las que el modelo fue entrenado, para este caso 500 *time steps*. Para el ejemplo mostrado en la Figura 4.23, en donde se compara la imagen generada por SR3 (derecha) y la imagen de alta resolución (izquierda), se obtuvieron los siguientes resultados: 51,38 dB de PSNR, 0,9937 de SSIM y 0,000 007 de MSE.

El SR3 es un modelo que tuvo buenos resultados en las métricas de calidad de reconstrucción de imágenes, incluso siendo entrenado con un conjunto de datos pequeño, pero el costo computacional de entrenamiento, inferencia y reconstrucción es demasiado elevado.

Slide	PSNR (dB)	SSIM	MSE
1	53,31	0,9952	0,000005
2	53,36	0,9952	0,000005
3	52,94	0,9950	0,000005
4	52,59	0,9947	0,000005
5	52,29	0,9944	0,000006
6	52,14	0,9943	0,000006
7	52,07	0,9943	0,000006
8	51,84	0,9942	0,000007
9	51,44	0,9936	0,000007
10	51,38	0,9937	0,000007

Cuadro 4.12: Análisis comparativo *Gold Standard* con imágenes MRI de prueba para el modelo SR3 2D.

El Cuadro 4.12 presenta los resultados de la evaluación del modelo SR3 con imágenes de prueba. En este caso, para las 10 imágenes evaluadas, los resultados son consistentes, robustos y aceptables, con valores de PSNR superiores a 51 dB en todos los casos, un SSIM muy cercano a 1, y métricas de MSE que reflejan muy poca pérdida de información. Estos valores sugieren que el modelo mejora significativamente la resolución de las imágenes, conservando tanto la estructura como la información, como se observa en la imagen de la derecha de la Figura 4.23.

4.3. Resumen Comparativo

En este resumen se presentan los valores de Pérdida (*Loss*) y las métricas PSNR y SSIM que se alcanzaron durante el proceso de entrenamiento con los datos de validación.

4.3.1. Modelos 3D

Modelos	Estudios MRI (176x256x256)			Ejecución	Loss	SSIM	PSNR
	Entrenamiento	Prueba	Validación				
UNET	27	5	3	4h, 13m	0,0101	0,9632	36,99
EDSR	27	5	3	8h, 24m	0,0095	0,9664	37,62
cGAN	27	5	3	5h, 3m	0,0253	0,9253	32,64
SRCNN	27	5	3	7h, 33m	0,0003	0,9641	37,24

Cuadro 4.13: Resumen Comparativo Modelos 3D.

Nota: En todos los modelos, las métricas SSIM y PSNR corresponden al conjunto de validación. La métrica de pérdida (Loss) también fue obtenida sobre el conjunto de validación, excepto en el caso del modelo cGAN, donde se reporta únicamente la pérdida del generador sobre el conjunto de entrenamiento, ya que el discriminador no fue evaluado fuera del conjunto de entrenamiento.

4.3.2. Modelos 2D

Modelos	Imágenes			Ejecución	Loss	SSIM	PSNR
	Entrenamiento	Prueba	Validación				
Autoencoder	25179	3147	3147	0d, 3h, 43m	0.4158	0.6562	17.23
DRCN	25179	3147	3147	0d, 3h, 18m	0.0580	0.8646	30.43
SRGAN	25179	3147	3147	0d, 10h, 40m	0.2858	0.8592	30.32
UNET	25179	3147	3147	1d, 4h, 43m	0.4113	0.8682	30.08
VDSR	25179	3147	3147	0d, 10h, 59m	0.0564	0.8663	30.96
SRResNet	25179	3147	3147	0d, 0h, 22m	0.00101	0.9954	61.58
SRDenseNet	25179	3147	3147	0d, 1h, 37m	0,0012	0.9941	59.41
SR3	800	100	-	0d, 3h, 40m	-	-	-

Cuadro 4.14: Resumen Comparativo Modelos 2D.

Nota: En SR3 no se utilizaron datos de validación. Debido a que SR3 predice el ruido en cada iteración de refinamiento, no se calculan las métricas SSIM y PSNR durante el entrenamiento. Tampoco se agrega la información de pérdida porque el proceso de entrenamiento se realiza 500 veces.

Capítulo 5

Conclusiones

A lo largo del desarrollo del proyecto, se diseñaron, entrenaron y evaluaron diferentes arquitecturas de redes neuronales profundas, adaptadas a entornos bidimensionales y tridimensionales. Los resultados obtenidos, medidos a través de métricas como PSNR, SSIM y MSE, demostrando mejoras significativas en la calidad visual y cuantitativa de los estudios de resonancia magnética (MRI) completos y de las imágenes reconstruidas, evidenciando la capacidad de los modelos desarrollados para recuperar información espacial de alta fidelidad. En este contexto, se concluye que el objetivo del proyecto —*implementar técnicas de super-resolución 2D y 3D basadas en aprendizaje profundo para mejorar la resolución espacial en estudios de resonancia magnética (MRI), validando su desempeño mediante métricas cuantitativas comparadas con imágenes de referencia tipo Gold Standard*— se cumplió satisfactoriamente.

En cuanto a los resultados obtenidos con los modelos tridimensionales, se observó que las arquitecturas UNet 3D, SRCNN 3D y EDSR 3D lograron desempeños satisfactorios en términos de calidad de reconstrucción, con diferencias marcadas en los tiempos de ejecución. La UNet 3D destacó por su eficiencia computacional, ofreciendo buenos resultados con un menor tiempo de entrenamiento en comparación con las demás. Por su parte, la cGAN 3D fue la arquitectura que requirió más pruebas y ajustes para estabilizar su desempeño. Aunque mostró potencial, alcanzar resultados comparables a los otros modelos fue más desafiante. Se evidenció que este tipo de arquitecturas generativas requiere un mayor poder de procesamiento, no solo por la complejidad del entrenamiento conjunto entre generador y discriminador, sino también por la necesidad de realizar un ajuste fino de hiperparámetros para lograr una convergencia estable. No obstante, trabajar con cGANs permitió explorar nuevas formas de mejorar la percepción visual de las reconstrucciones, lo cual es especialmente relevante en el contexto de imágenes médicas tridimensionales.

En el dominio bidimensional, los modelos SRResNet 2D y SRDenseNet 2D emergieron como los más eficaces, alcanzando valores de PSNR excepcionalmente altos (65,9 dB y 48,15 dB, respectivamente) y SSIM muy cercanos a 1 (0,9996 y 0,9905), indicando una reconstrucción casi perfecta y una rápida convergencia. El modelo SRResNet 2D demostró ser muy eficiente y de alto alto rendimiento. Es un modelo compacto, robusto, fácilmente interpretable y altamente escalable, capaz de entrenarse rápidamente con grandes volúmenes de imágenes. Además, ofreció excelentes métricas en la reconstrucción de imágenes con datos de prueba. Las arquitecturas SRResNet 2D y SRDenseNet 2D son modelos flexibles que ofrecen amplias capacidades de exploración y mejora para tareas de super-resolución en imágenes diagnósticas, principalmente por la eficiencia demostrada y por su facilidad de implementación.

Modelos como VDSR, DRCN y SRGAN también ofrecieron un rendimiento aceptable (PSNR 30 dB, SSIM 0,86), demostrando ser opciones viables aunque con margen de mejora. Sin embargo, el Autoencoder 2D (PSNR 17,23 dB) y el modelo de difusión SR3 2D mostraron limitaciones importantes; el Autoencoder por su baja calidad de reconstrucción y el SR3 2D por su altísimo costo computacional y la necesidad de un entrenamiento extensivo, a pesar de lograr buenas métricas (PSNR > 51 dB) con datos limitados.

La exploración de diversos modelos en dos y tres dimensiones permitió analizar las fortalezas y debili-

dades de cada uno frente a una problemática real en un campo de aplicación exigente. Este análisis puede contribuir a la definición de procesos en las fases de implementación y validación médica de sistemas para la mejora de la resolución en imágenes diagnósticas.

Este trabajo resalta la relevancia de las técnicas de super-resolución basadas en aprendizaje profundo como una solución innovadora para superar las limitaciones inherentes a la adquisición de MRI. La capacidad de mejorar la resolución espacial a posteriori tiene un impacto significativo, especialmente en entornos con acceso limitado a equipos de última generación. Facilitar diagnósticos más precisos y tempranos, particularmente en enfermedades neurodegenerativas o cáncer, y reducir la necesidad de repetir estudios, no solo mejora la atención al paciente sino que también optimiza los recursos del sistema de salud. La investigación contribuye al avance de la imagenología médica y la inteligencia artificial aplicada, demostrando que arquitecturas como las basadas en bloques residuales y encoder-decoder son especialmente prometedoras.

Capítulo 6

Trabajos Futuros

1. Validación clínica y evaluación por expertos.

Dado que el alcance actual del proyecto no incluye validación clínica, médica, neurológica o de radiología, un paso fundamental para el futuro sería someter las imágenes mejoradas mediante super-resolución a una evaluación por parte de radiólogos y otros especialistas médicos. Esto permitiría determinar el impacto real de las técnicas implementadas en la precisión diagnóstica y su utilidad en la práctica clínica diaria.

2. Extensión a otras modalidades de resonancia magnética y patologías.

El proyecto se enfoca en estudios MRI anatómicos T1. Investigaciones futuras podrían explorar la aplicación de estas técnicas de super-resolución a otras secuencias de MRI (como T2, FLAIR, difusión DWI) que son cruciales para el diagnóstico de diferentes condiciones. Además, sería valioso extender el estudio a cohortes de pacientes con otras patologías específicas para evaluar la generalización y efectividad de los modelos.

3. Incorporación de nuevas arquitecturas de aprendizaje profundo y técnicas híbridas.

El campo del aprendizaje profundo evoluciona rápidamente. Trabajos futuros podrían investigar la implementación y adaptación de arquitecturas de super-resolución más recientes o emergentes que no se incluyeron en la selección inicial. Se podrían explorar enfoques híbridos que combinen las fortalezas de diferentes modelos o que integren información de múltiples fuentes o escalas.

4. Optimización para entornos con recursos limitados y despliegue clínico.

El proyecto justifica la relevancia del proyecto en contextos con equipos obsoletos o limitaciones de acceso a tecnología avanzada. Una línea de trabajo futuro podría centrarse en optimizar los modelos de super-resolución para su ejecución eficiente en hardware con capacidades computacionales limitadas, facilitando su despliegue en centros de salud con menos recursos.

Bibliografía

- [1] E. J. R. van Beek, C. Kuhl, Y. Anzai et al., «Value of MRI in medicine: More than just another test?» *J. Magn. Reson. Imaging*, vol. 49, n.º 7, e14-e25, jun. de 2019. DOI: 10.1002/jmri.26211.
- [2] G. Katti, S. A. Ara y A. Shireen, «Magnetic resonance imaging (MRI)—A review,» *International Journal of Dental Clinics*, vol. 3, n.º 1, págs. 65-70, 2011.
- [3] U.S. Food and Drug Administration. «MRI (Magnetic Resonance Imaging) - Benefits and Risks.» [En línea]. (2017), dirección: <https://www.fda.gov/radiation-emitting-products/mri-magnetic-resonance-imaging/benefits-and-risks>.
- [4] D. Moratal-Pérez, J. Millet-Roig, F. Castells, L. Martí-Bonmatí y M. E. Brummer, «Surcando el espacio- k para mejorar la imagen por resonancia magnética,» *Radiología*, vol. 46, n.º 3, págs. 133-150, 2004.
- [5] M. E. Ladd, «High-field-strength magnetic resonance: potential and limits,» *Topics in Magnetic Resonance Imaging*, vol. 18, n.º 2, págs. 139-152, 2007.
- [6] M. Roy, B. Upadhyaya, J. Rai y K. Sharma, «Resolution Enhancement of Brain MRI Images Using Deep Learning,» *Engineering Proceedings*, vol. 59, n.º 1, pág. 158, 2024.
- [7] Z. Zhong, «Advancing Spatial and Temporal Resolution of MRI for Biological and Physical Applications,» Doctoral dissertation, University of Illinois at Chicago, 2021.
- [8] OASIS Brain Project, Washington University School of Medicine, St. Louis. «Open Access Series of Imaging Studies.» [Online]. (n.d.), dirección: <https://sites.wustl.edu/oasisbrains/>.
- [9] Alzheimer’s Disease Neuroimaging Initiative (ADNI). «Alzheimer’s Disease Neuroimaging Initiative.» [Online]. (n.d.), dirección: <https://adni.loni.usc.edu/data-samples/access-data/>.
- [10] Z. Chen, K. Pawar, M. Ekanayake, C. Pain, S. Zhong y G. F. Egan, «Deep learning for image enhancement and correction in magnetic resonance imaging—state-of-the-art and challenges,» *Journal of Digital Imaging*, vol. 36, n.º 1, págs. 204-230, 2023.
- [11] D. Crosby, S. Bhatia, K. M. Brindle et al., «Early detection of cancer,» *Science*, vol. 375, n.º 6586, eaay9040, 2022. DOI: 10.1126/science.aay9040.
- [12] M. J. Oude Wolcherink, C. M. Behr, X. G. Pouwels, C. J. Doggen y H. Koffijberg, «Health economic research assessing the value of early detection of cardiovascular disease: a systematic review,» *Pharmacoeconomics*, vol. 41, n.º 10, págs. 1183-1203, 2023.
- [13] C. A. Vilarroig, *Universidad Politécnica de Valencia*, [Online], jul. de 2019. dirección: <https://riunet.upv.es/handle/10251/124500>.
- [14] Y. LeCun, Y. Bengio y G. Hinton, «Deep learning,» *Nature*, vol. 521, n.º 7553, págs. 436-444, 2015. DOI: 10.1038/nature14539.
- [15] P. P. Shinde y S. Shah, «A Review of Machine Learning and Deep Learning Applications,» en *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, IEEE, ago. de 2018, págs. 1-6. DOI: 10.1109/ICCUBEA.2018.8697857.
- [16] A. Shrestha y A. Mahmood, «Review of Deep Learning Algorithms and Architectures,» *IEEE Access*, vol. 7, págs. 53 040-53 065, 2019. DOI: 10.1109/ACCESS.2019.2912200.

- [17] J. Guerreiro, P. Tomás, N. Garcia y H. Aidos, «Super-resolution of magnetic resonance images using generative adversarial networks,» *Computerized Medical Imaging and Graphics*, vol. 108, pág. 102-280, 2023. DOI: 10.1016/j.compmedimag.2023.102280.
- [18] P. Ronneberger, O. Fischer y T. Brox, «U-Net: Convolutional Networks for Biomedical Image Segmentation,» en *Medical Image Computing and Computer-Assisted Intervention*, 2015, págs. 234-241.
- [19] B. Lim, S. Son, H. Kim, S. Nah y K. Mu Lee, «Enhanced Deep Residual Networks for Single Image Super-Resolution,» en *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, págs. 136-144.
- [20] M. Mirza y S. Osindero, *Conditional Generative Adversarial Nets*, arXiv preprint arXiv:1411.1784, 2014. dirección: <https://arxiv.org/abs/1411.1784>.
- [21] C. Dong, C. C. Loy, K. He y X. Tang, «Learning a Deep Convolutional Network for Image Super-Resolution,» en *Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part IV*, ép. Lecture Notes in Computer Science, vol. 8692, Springer International Publishing, 2014, págs. 184-199. DOI: 10.1007/978-3-319-10593-2_13.
- [22] F. G. López, A. M. M. García y R. M. Carrión, «Influencia de la selección de hiper-parámetros en el rendimiento de autoencoders para la detección de ataques en red,» en *Actas de las VIII Jornadas Nacionales de Investigación en Ciberseguridad: Vigo, 21 a 23 de junio de 2023*, Vigo, España: Universidade de Vigo, 2023, págs. 93-100.
- [23] H. Ochoa Domínguez, «Estudio comparativo de algoritmos de súper resolución de una sola imagen basados en aprendizaje profundo,» Trabajo de tesis, Instituto de Ingeniería y Tecnología, 2020.
- [24] C. Ledig, L. Theis, F. Huszár et al., «Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network,» en *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2017, págs. 4681-4690.
- [25] M. Hassan, K. Illanko y X. N. Fernando, «Single Image Super Resolution Using Deep Residual Learning,» *AI*, vol. 5, n.º 1, págs. 426-445, 2024. DOI: 10.3390/ai5010023.
- [26] K. He, X. Zhang, S. Ren y J. Sun, «Deep Residual Learning for Image Recognition,» en *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2016, págs. 770-778.
- [27] G. Huang, Z. Liu, L. Van Der Maaten y K. Q. Weinberger, «Densely Connected Convolutional Networks,» en *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2017, págs. 4700-4708.
- [28] T. Tong, G. Li, X. Liu y Q. Gao, «Image Super-Resolution Using Dense Skip Connections,» en *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, IEEE, 2017, págs. 4799-4807.
- [29] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet y M. Norouzi, «Image Super-Resolution via Iterative Refinement,» *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, n.º 4, págs. 4713-4726, 2022. DOI: 10.1109/TPAMI.2022.3209428.
- [30] J. Ho, A. Jain y P. Abbeel, «Denoising Diffusion Probabilistic Models,» en *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, págs. 6840-6851.
- [31] Y. Wang, W. Yang, X. Chen et al., «SINSR: Diffusion-Based Image Super-Resolution in a Single Step,» en *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2024, págs. 25 796-25 805.
- [32] Q. Lyu, C. You, H. Shan y G. Wang, «Super-Resolution MRI Through Deep Learning,» *arXiv preprint arXiv:1810.06776*, 2018. dirección: <https://arxiv.org/abs/1810.06776>.
- [33] A. Estler, L. Zerweck, M. Brunnée et al., «Deep Learning-Accelerated Image Reconstruction in MRI of the Orbit to Shorten Acquisition Time and Enhance Image Quality,» *Journal of Neuroimaging*, vol. 34, n.º 2, págs. 232-240, 2024. DOI: 10.1111/jon.13156.
- [34] A. Ayaz, R. Boonstoppel, C. Lorenz, J. Weese, J. Pluim y M. Breeuwer, «Effective deep-learning brain MRI super resolution using simulated training data,» *Computers in Biology and Medicine*, vol. 183, pág. 109 301, 2024. DOI: 10.1016/j.compbiomed.2024.109301.

- [35] M. Padwekar, «Super-resolution of Brain MRI Images,» en *2025 Fourth International Conference on Power, Control and Computing Technologies (ICPC2T)*, 2025, págs. 1-4. DOI: 10.1109/ICPC2T63847.2025.10958683.
- [36] H. Do, D. Helbert, P. Bourdon, M. Naudin, C. Guillevin y R. Guillevin, «MRI super-resolution using 3D cycle-consistent generative adversarial network,» en *2021 Sixth International Conference on Advances in Biomedical Engineering (ICABME)*, 2021, págs. 85-88. DOI: 10.1109/ICABME53305.2021.9604810.
- [37] Laboratory of Neuro Imaging (LONI), *IDA User Manual*, [Online]. dirección: https://ida.loni.usc.edu/explore/jsp/support/IDA_User_Manual.pdf.
- [38] K. L. Crawford, S. C. Neu y A. W. Toga, «The Image and Data Archive at the Laboratory of Neuro Imaging,» *NeuroImage*, vol. 124, págs. 1080-1083, 2016. DOI: 10.1016/j.neuroimage.2015.09.070.
- [39] Parkinson's Progression Markers Initiative, *Parkinson's Progression Markers Initiative*, <https://www.ppmi-info.org/>, [En línea].

Anexos

A continuación se presentan los enlaces a los repositorios de GitHub donde se desarrollaron los modelos de Super-resolución 2D y 3D:

- Repositorio de modelos de Super-resolución 2D:
<https://github.com/carlosarcos1/super-resolution-mri>
- Repositorio de modelos de Super-resolución 3D:
<https://github.com/GioOrtega7/Super-resolucion-MRI-3D.git>