



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Informática

Generación de conjuntos de datos para entrenamiento de
métodos de reconstrucción de imágenes de tomografía por
emisión de positrones

Trabajo Fin de Grado

Grado en Ciencia de Datos

AUTOR/A: Marin Colome, Miquel

Tutor/a: Morillas Gómez, Samuel

Cotutor/a: Rodríguez Álvarez, María José

CURSO ACADÉMICO: 2021/2022

Resum

La intenció d'aquest treball és la d'arribar a trobar un mètode eficaç per a la generació d'una base de dades sintètica i realista que permeti la reconstrucció d'imatges de tomografies per emissió de positrons (PET), com a suport a la tasca dels radiòlegs en l'estudi de les lesions neoplàsiques de mama. Per assolir aquest objectiu, el projecte s'ha centrat en l'ús de models basats en l'aprenentatge profund. L'escull principal que existeix per entrenar aquests tipus de models és l'escassetat d'imatges clíniques disponibles de diagnosi oncològica. Per fer front a aquest obstacle s'ha confiat en les xarxes neuronals generatives adversàries (GAN). Partint d'un ampli estudi previ d'investigació sobre la bibliografia al·lusiva al tema i dels articles més recents publicats sobre aquesta qüestió, es descriuen els diferents models generatius GAN, llurs components, funcionament i procés d'entrenament així com els problemes de convergència que se'n deriven d'aquest procés. L'experimentació s'ha realitzat amb tres models diferents de xarxes GAN i fent-ne ús de dues bases de mostres de característiques heterogènies: primer una amb imatges de tomografies PET i, posteriorment, una altra amb mostres de mamografies. En la creació de les diferents arquitectures s'han emprat les llibreries *Keras* i *Tensorflow* i, per a l'entrenament dels models seleccionats, s'ha fet ús d'una GPU Tesla T4 facilitada per Google Colab. Amb tot això s'han anat aconseguint progressivament imatges sintètiques d'una millor resolució. Els pesos d'aprenentatge obtinguts amb mostres de mamografies, en el model amb el millor rendiment assolit prèviament, han permès el reentrenament dels tres models GAN proposats amb les mostres PET inicials. Això ha facilitat finalment, mitjançant la tècnica de *transfer learning*, l'obtenció de les imatges amb més definició de l'anatomia de la mama.

Paraules clau: càncer de mama, tomografies PET, generació d'imatges, aprenentatge profund, xarxes GAN, *transfer learning*, intel·ligència artificial

Resumen

El propósito de este trabajo es hallar un método eficaz para la generación de una base de datos sintética y realista que permita la reconstrucción de imágenes de tomografías por emisión de positrones (PET), como apoyo a la labor de los radiólogos en el estudio de las lesiones neoplásicas de mama. Para alcanzar este objetivo, el presente proyecto se ha centrado en el empleo de modelos basados en aprendizaje profundo. La principal dificultad que existe para entrenar estos modelos es la escasez de imágenes clínicas disponibles de diagnóstico oncológico. Para hacer frente a este obstáculo se ha confiado en el empleo de las redes neuronales generativas adversarias (GAN). Partiendo de un amplio estudio previo de investigación sobre la bibliografía relativa al tema y de los más recientes artículos publicados sobre dicha cuestión, se describen los principales modelos generativos GAN, su funcionamiento, componentes y proceso de entrenamiento así como los problemas de convergencia derivados del mismo. La experimentación se ha realizado con tres modelos diferentes de redes GAN y utilizando dos bases de muestras de características heterogéneas: primero una con imágenes de tomografías PET y, posteriormente, otra con muestras de mamografías. En la creación de las diferentes arquitecturas se han empleado las librerías *Keras* y *Tensorflow* y, para el entrenamiento de los modelos seleccionados, se ha utilizado una GPU Tesla T4 facilitada por Google Colab. Con todo ello se han ido obteniendo progresivamente imágenes sintéticas de una mayor resolución. Los pesos de aprendizaje obtenidos con muestras de mamografías en el modelo con el mejor rendimiento observado previamente, han permitido reentrenar los tres modelos GAN propuestos con las muestras PET iniciales. Esto ha facilitado finalmente, mediante la técnica de *transfer learning*, la obtención de las imágenes con mayor definición de la anatomía de la mama.

Palabras clave: cáncer de mama, tomografías PET, generación de imágenes, aprendizaje profundo, redes GAN, *transfer learning*, inteligencia artificial

Abstract

The aim of this work is to find an effective method for the generation of a synthetic and realistic database that allows the reconstruction of positron emission tomography images (PET), to support the radiologists' tasks in the study of neoplastic breast lesions. To achieve this goal, this project has focused on the use of deep learning-based models. The main difficulty in training these models is the scarcity of available clinical images of oncological diagnoses. To address this obstacle, we have relied on the use of generative adversarial neural networks (GANs). Based on an extensive previous research study of the literature and the most recent articles published on the subject, we describe the main GAN generative models, their performance, components and training process, as well as the convergence problems derived from it. The experimentation has been carried out with three different models of GANs and using two sample bases with heterogeneous characteristics: firstly, a database of PET scans and, subsequently, another base with mammography samples. *Keras* and *Tensorflow* libraries were used to create the different architectures, and a Tesla T4 GPU provided by Google Colab was used to train the selected models. This has resulted in progressively higher resolution synthetic images. The learning weights obtained with mammography samples in the previously observed best-performing model, have allowed the three proposed GAN models to be retrained with the initial PET samples. This has finally made it possible, using the transfer learning technique, to obtain images with greater definition of the breast anatomy.

Key words: breast cancer, PET scan, image generation, deep learning, GAN networks, transfer learning, artificial intelligence

Índice general

Índice general	VII
Índice de figuras	IX
Índice de tablas	XII
<hr/>	
1 Introducción	1
1.1 Motivación	2
1.2 Objetivos	3
1.3 Impacto esperado	3
1.4 Estructura de la memoria	4
2 Contexto	7
2.1 Técnicas de diagnóstico por imágenes	7
2.2 Radiofármacos	9
2.3 Metabolismo de la glucosa y utilización de isótopos radiactivos	9
2.4 Conceptos básicos de física de partículas	10
2.5 Generación de imágenes PET	11
3 Estado del arte	15
3.1 Datos sintéticos	15
3.2 Métodos de generación sin entrenamiento	16
3.3 Métodos de generación con entrenamiento	16
3.3.1 Métodos estadísticos	17
3.3.1.1 Simulaciones de Monte Carlo	17
3.3.1.2 Modelos de Naive Bayes	18
3.3.2 Métodos de visión artificial	18
3.4 Crítica al estado del arte	20
4 Descripción del problema	23
4.1 Análisis del marco legal y ético	23
4.2 Solución propuesta	26
5 Redes generativas adversarias GAN. Marco teórico	29
5.1 Introducción	29
5.2 Conceptos teóricos de las redes neuronales convolucionales	30
5.2.1 Concepto de convolución	31
5.2.2 Bloque convolucional	31
5.2.3 Bloque de clasificación	33
5.3 Formulación matemática del modelo	33
5.3.1 Planteamiento	33
5.3.2 Teorema minimax	35
5.3.3 Funciones de pérdida	36
5.3.3.1 Función de pérdida del discriminador	37
5.3.3.2 Función de pérdida del generador	37

5.3.3.3	Función de pérdida del modelo	37
5.4	Proceso de entrenamiento del modelo	38
5.5	Problemas de convergencia	40
5.5.1	Colapso de modo	41
5.5.2	Desvanecimiento del gradiente	41
5.6	Variantes de GAN	43
5.6.1	Deep Convolutional GAN (DCGAN)	43
5.6.2	Wasserstein GAN (WGAN)	45
5.6.2.1	Entrenamiento de la red	50
5.6.3	Wasserstein GAN Gradient Penalty (WGAN - GP)	51
5.6.3.1	Entrenamiento de la red	52
6	Recursos utilizados	53
6.1	Herramientas tecnológicas	53
6.1.1	Software	53
6.1.2	Hardware	54
6.2	Conjuntos de datos	55
6.2.1	Imágenes del i3M	56
6.2.2	Imágenes de la University of South Florida (USF)	57
7	Experimentación y resultados	59
7.1	Procesamiento previo	59
7.2	Métricas utilizadas	60
7.3	Modelos propuestos	61
7.3.1	GAN	61
7.3.2	DCGAN	63
7.3.3	WGAN - GP	64
7.4	Experimentos realizados. Parámetros	65
7.4.1	GAN y DCGAN	66
7.4.2	WGAN - GP	66
7.5	Resultados	67
7.5.1	Experimentos con muestras i3M	67
7.5.1.1	Modelo i3M - GAN	67
7.5.1.2	Modelo i3M - DCGAN	71
7.5.1.3	Modelo i3M - WGAN - GP	74
7.5.2	Experimentos con muestras USF	76
7.5.2.1	Modelo USF - GAN	76
7.5.2.2	Modelo USF - DCGAN	79
7.5.2.3	Modelo USF - WGAN - GP	82
7.5.3	Transfer learning aplicado a las muestras i3M	83
8	Conclusiones	87
8.1	Legado	89
8.2	Relación del trabajo desarrollado con los estudios cursados	90
9	Desarrollos futuros	93
	Bibliografía	95
<hr/>		
	Apéndices	
A	Objetivos de Desarrollo Sostenible (ODS)	103
B	Distancia Wasserstein. Ejemplo	107
C	Función de pérdida W-Loss. Desarrollo matemático	111

Índice de figuras

2.1	De izquierda a derecha y de arriba a abajo, imágenes de RX, ecografía, CT, MRI, SPECT y PET. Imágenes extraídas de [9][10][11][12][13] y [14].	8
2.2	Composición química de la molécula de 2-desoxi-D-glucosa y cómo el isótopo Flúor-18 se une a ella para formar el trazador fluorodesoxiglucosa 18 F-FDG. Imagen extraída de [19].	9
2.3	Diagrama de Feynman, donde se observa el intercambio de un fotón, representado por una línea ondulada γ , generado por una aniquilación entre un electrón y un positrón. Imagen extraída de [25].	11
2.4	Emisión de fotones por aniquilación de electrones-positrones y su detección por el tomógrafo. Imagen extraída de [27].	11
2.5	Representación de una línea de coincidencia LOR. Imagen extraída de [28].	12
2.6	Representación de la técnica empleada en los tomógrafos PET. Imagen extraída de [29].	13
3.1	Arquitectura DeepPET. Imagen extraída de [40].	19
4.1	Esquema de la solución propuesta.	27
5.1	Red neuronal generativa adversaria.	30
5.2	Arquitectura de una red neuronal convolucional. Imagen extraída de [52].	31
5.3	De izquierda a derecha se pueden ver las funciones de activación sigmoide, tangente hiperbólica, Rectified Linear Unit (ReLU) y Exponential Linear Unit (ELU). Imagen extraída de [53].	32
5.4	Hiperparámetros que intervienen en una red neuronal convolucional. . .	33
5.5	Proceso de entrenamiento de una red generativa adversaria. Imagen extraída de [55].	34
5.6	Proceso de entrenamiento de una GAN. Fragmento del artículo de Goodfellow et al. Imagen extraída de [51].	38
5.7	Ruta de aprendizaje de una red neuronal estándar (a la izquierda) y ruta de aprendizaje de una red generativa adversaria (a la derecha). Imagen extraída de [58].	40
5.8	Conjunto de imágenes creadas por una GAN. Imagen extraída de [59]. . .	41
5.9	Evolución del rendimiento del discriminador y del generador. Imagen extraída de MathWorks ©.	42
5.10	Arquitectura de una red DCGAN. En la parte superior se encuentra el generador y, en la inferior, el discriminador. Imagen extraída de Medium.com.	43
5.11	Aproximación mediante WGAN de una distribución de datos generados (en naranja) a una distribución de datos reales (en azul). Imagen extraída de Medium.com.	46
5.12	Proceso de entrenamiento del crítico, empleando el método weight-clipping. Imagen extraída de [62].	50

5.13	Proceso de entrenamiento de una WGAN. Fragmento del artículo de Arjovsky et al. Imagen extraída de [61].	50
5.14	Proceso de entrenamiento de una WGAN. Fragmento del artículo de Gulrajani et al. Imagen extraída de [62].	52
6.1	Descripción de características de la GPU Tesla T4, obtenida del <i>notebook</i> de Google Colab.	55
6.2	Ejemplos de imágenes de tomografía PET procedentes del i3M.	57
6.3	Detalle de un caso particular de cribado de mama procedente de la <i>University of South Florida</i>	58
7.1	Modelo Inception-v3. Imagen extraída de Google Cloud.	60
7.2	Arquitectura del generador GAN.	62
7.3	Arquitectura del discriminador GAN.	63
7.4	Arquitectura del generador DCGAN.	63
7.5	Arquitectura del discriminador DCGAN.	64
7.6	Modelo i3M - GAN. Evolución del <i>loss</i> según optimizador aplicando <i>dropouts</i> de 0.4, 0.5 y 0.6.	68
7.7	Modelo i3M - GAN. Evolución del <i>loss</i> según <i>dropout</i> , aplicando optimizadores RMSprop, Adam y Adadelta.	68
7.8	Modelo i3M - GAN. Evolución del <i>loss</i> según <i>batch size</i> , aplicando optimizador RMSprop.	69
7.9	Modelo i3M - GAN. Evolución del <i>loss</i> según <i>learning rate</i> , aplicando optimizador RMSprop. La imagen derecha aporta la misma información que la de la izquierda pero se han eliminado las primeras 150 <i>epochs</i> por cuestiones de escala, pudiéndose así visualizar mejor los valores.	70
7.10	Modelo i3M - GAN. Imágenes de la mama obtenidas con la mejor configuración posible.	71
7.11	Modelo i3M - DCGAN. Evolución del <i>loss</i> según optimizador, aplicando tasas de abandono (<i>dropouts</i>) de 0.4, 0.5 y 0.6.	71
7.12	Modelo i3M - DCGAN. Evolución del <i>loss</i> según <i>dropout</i> , aplicando optimizadores RMSprop, Adam y Adadelta.	72
7.13	Modelo i3M - DCGAN. Evolución del <i>loss</i> según <i>batch size</i> , aplicando optimizador RMSprop. La imagen derecha aporta la misma información que la izquierda pero, en este caso, se han eliminado las primeras 150 <i>epochs</i> por cuestiones de escala, pudiéndose así visualizar mejor los valores.	72
7.14	Modelo i3M - DCGAN. Evolución del <i>loss</i> según <i>learning rate</i> , aplicando optimizador RMSprop. La imagen derecha aporta la misma información que la izquierda pero, en este caso, se han eliminado las primeras 150 <i>epochs</i> por cuestiones de escala, pudiéndose así visualizar mejor los valores.	73
7.15	Imágenes de la mama obtenidas con la mejor configuración.	74
7.16	Modelo i3M - WGAN-GP. Evolución del <i>loss</i> según <i>batch size</i> , aplicando valores λ de 0.1, 1 y 10 respectivamente.	74
7.17	Modelo i3M - WGAN-GP. Evolución del <i>loss</i> según <i>batch size</i> , aplicando valores λ de 0.1, 1 y 10 respectivamente pero donde, por cuestiones de escala, se han eliminado las primeras 150 <i>epochs</i> , pudiéndose así visualizar mejor los valores.	75

7.18	Modelo i3M - WGAN-GP. Evolución del <i>loss</i> según λ , aplicando tamaños de lote de 8 y 16. Las gráficas inferiores aportan la misma información que las superiores donde, por cuestiones de escala, se han eliminado las primeras 150 <i>epochs</i> , pudiéndose así visualizar mejor los valores.	75
7.19	Modelo i3M - WGAN-GP. Imágenes de la mama obtenidas con la mejor configuración.	76
7.20	Modelo USF - GAN. Evolución del <i>loss</i> según optimizador, aplicando tasas de abandono (<i>dropouts</i>) 0.4, 0.5 y 0.6.	77
7.21	Modelo USF - GAN. Evolución del <i>loss</i> según <i>dropout</i> , aplicando optimizadores RMSprop, Adam y Adadelta.	77
7.22	Modelo USF - GAN. Evolución del <i>loss</i> según <i>batch size</i> , aplicando optimizador RMSprop. La imagen de la derecha aporta la misma información que la de la izquierda pero en ésta, por cuestiones de escala, se han eliminado las primeras 50 <i>epochs</i> , pudiéndose así visualizar mejor los valores.	78
7.23	Modelo i3M - GAN. Evolución del <i>loss</i> según <i>batch size</i> , aplicando optimizador RMSprop.	78
7.24	Modelo USF - GAN. Imágenes de la mama obtenidas con la mejor configuración.	79
7.25	Modelo USF - DCGAN. Evolución del <i>loss</i> según optimizador, aplicando tasas de abandono (<i>dropouts</i>) 0.4, 0.5 y 0.6.	79
7.26	Modelo USF - DCGAN. Evolución del <i>loss</i> según <i>dropout</i> , aplicando optimizadores RMSprop, Adam y Adadelta.	80
7.28	Modelo USF - DCGAN. Evolución del <i>loss</i> según <i>batch size</i> , aplicando optimizador RMSprop. La imagen derecha aporta la misma información que la de la izquierda pero, en este caso, se han eliminado por cuestiones de escala las primeras 50 <i>epochs</i> , pudiéndose así visualizar mejor los valores.	81
7.29	Modelo USF - DCGAN. Imágenes de la mama obtenidas con la mejor configuración.	81
7.30	Modelo USF - WGAN-GP. Evolución del <i>loss</i> según <i>batch size</i> , aplicando valores λ de 0.1, 1 y 10. Las gráficas inferiores aportan la misma información que las superiores pero, en este caso, se han eliminado por cuestiones de escala las primeras 250 <i>epochs</i> , pudiéndose así visualizar mejor los valores.	82
7.31	Modelo USF - WGAN-GP. Evolución del <i>loss</i> según λ , aplicando tamaños de lote de 16 y 32. Las gráficas inferiores aportan la misma información que las superiores pero, en este caso, se han eliminado por cuestiones de escala las primeras 150 <i>epochs</i> , pudiéndose así visualizar mejor los valores.	82
7.32	Modelo USF - WGAN-GP. Imágenes de la mama obtenidas con la mejor configuración.	83
7.33	Imágenes de tomografías PET originales del i3M.	84
7.34	Imágenes PET generadas por el modelo DCGAN con RMSprop, tasa de abandono de 0.4 y <i>learning rate</i> de 0.0008.	85
7.35	Tomografías PET generadas después de la aplicación del método de <i>transfer learning</i> mediante los pesos de aprendizaje correspondientes a la mejor configuración obtenida con mamografías de la USF.	86
A.1	Imagen extraída de UNESCO.	103
A.2	Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).	106

B.1	Imagen extraída de Medium.com.	107
B.2	Imagen extraída de Medium.com.	108
B.3	Imagen extraída de Medium.com.	108

Índice de tablas

7.1	Parámetros utilizados en las arquitecturas GAN y DCGAN para imágenes del i3M.	66
7.2	Parámetros utilizados en las arquitecturas GAN y DCGAN para imágenes del repositorio de la <i>University of South Florida</i>	66
7.3	Parámetros utilizados en la arquitectura WGAN para imágenes del i3M.	66
7.4	Parámetros utilizados en la arquitectura WGAN para imágenes del repositorio de la <i>University of South Florida</i>	67
7.5	Modelo i3M - GAN. Valores FID según optimizador y <i>dropout</i>	68
7.6	Modelo i3M - GAN. Valores FID según <i>batch size</i>	69
7.7	Modelo i3M - GAN. Valores FID según <i>learning rate</i>	70
7.8	Modelo i3M - DCGAN. Valores FID según optimizador y <i>dropout</i>	72
7.9	Modelo i3M - DCGAN. Valores FID según <i>batch size</i>	73
7.10	Modelo i3M - DCGAN. Valores FID según <i>learning rate</i>	73
7.11	Modelo i3M - WGAN-GP. Valores FID según <i>batch size</i> y <i>lambda</i>	76
7.12	Modelo USF - GAN. Valores FID según optimizador y <i>dropout</i>	77
7.13	Modelo USF - GAN. Valores FID según <i>batch size</i>	78
7.14	Modelo USF - GAN. Valores FID según <i>learning rate</i>	78
7.15	Modelo USF - DCGAN. Valores FID según optimizador y <i>dropout</i>	80
7.16	Modelo USF - DCGAN. Valores FID según <i>batch size</i>	80
7.17	Modelo USF - DCGAN. Valores FID según <i>learning rate</i>	81
7.18	Modelo USF - WGAN-GP. Valores FID según <i>batch size</i> y <i>lambda</i>	83

CAPÍTULO 1

Introducción

El cáncer es la segunda causa principal de mortalidad en los países de la Unión Europea (UE), después de las enfermedades cardiovasculares. En el año 2020, unos 2,7 millones de personas fueron diagnosticadas en la UE con esta patología, y otros 1,3 millones de personas perdieron la vida a causa de ella [1]. Se prevé que las muertes provocadas por el cáncer aumenten en un 24 % de aquí a 2035 ¹, convirtiéndose así esta enfermedad en la principal causa de muerte en Europa.

El cáncer de mama femenino es el cáncer más comúnmente diagnosticado. Se estima que más de 355.000 mujeres en la UE han sido diagnosticadas de cáncer de mama en 2020 (el 13,3 % de todos los diagnósticos de cáncer). Según la Oficina Europea de Estadística (Eurostat), en 2016 el cáncer de mama se cobró 85.300 vidas en la UE, lo que representa un 7 % de todas las muertes por cáncer [2]. En la lucha contra esta enfermedad, la detección temprana de patologías de mama mediante exámenes de diagnóstico ha demostrado ser un potente aliado, al contribuir de manera eficaz en la disminución del porcentaje de mortalidad general.

Pero la detección del cáncer de mama es una de las tareas de diagnóstico más difíciles debido a que el tejido mamario tiene muchas características que pueden llegar a confundir al ojo humano. Para poder realizar un cribado a gran escala que sea efectivo se requiere mucho tiempo, dinero y recursos. Es obvio pues, que existe un claro potencial para que los sistemas automatizados reduzcan los costes y los tiempos de procesamiento, apoyando así la labor de los radiólogos y ayudando a reducir los errores humanos.

Parece cada vez más claro que, por los motivos apuntados anteriormente, el futuro de la detección del cáncer de mama, y de los diagnósticos en general, va a verse condicionado por las tecnologías de inteligencia artificial (IA) al acelerar éstas los tiempos de procesamiento de las imágenes clínicas. Actualmente ya existen algoritmos que pueden detectar anomalías en tomografías por emisión de positrones (PET) tan bien como un radiólogo experto, consiguiendo diagnosticar cánceres antes de que estos tumores malignos sean perceptibles por el ojo humano.

¹<https://gco.iarc.fr/tomorrow/en/>

1.1 Motivación

Si bien el trabajo de la IA resulta sorprendente, es evidente que ésta funciona mejor cuando se construye a partir de conjuntos de datos bien diseñados. Los investigadores necesitan grandes volúmenes de datos para entrenar, y luego testar, los sistemas de aprendizaje profundo desarrollados. El gran reto que se plantea es cómo obtener los datos necesarios para desarrollar los algoritmos que puedan detectar anomalías patológicas en las exploraciones PET.

La Comisión Europea ya está trabajando en este objetivo, tal y como recoge el Plan Europeo de Lucha contra el Cáncer [3] aprobado en 2021. En este plan se proponen diez *iniciativas emblemáticas* y múltiples medidas para hacer frente a todas las etapas de la enfermedad.

Concretamente, la segunda de ellas, conocida como *Iniciativa Europea de Imagen del Cáncer*, pretende desarrollar un atlas de la UE de imágenes relacionadas con esta enfermedad, como parte de la creación del Espacio Europeo de Datos Sanitarios o EHDS según las iniciales de su nombre en inglés (*European Health Data Space*).

Recientemente, en mayo de 2022, la Comisaria Europea de Salud y Seguridad Alimentaria, la chipriota Stella Kyriakides, presentó la propuesta definitiva para la creación de ese Espacio Europeo de Datos sanitarios con 2025 fecha límite para su desarrollo ².

A pesar de que esta actuación ha de favorecer en gran manera la labor de los investigadores, al poner a su disposición un mayor volumen de imágenes de lesiones neoplásicas de mama con las que entrenar sus modelos de IA, la precariedad de bases de datos de calidad continua siendo un problema acuciante. El sector sanitario es un sector «rico en datos pero pobre en información» ³. Los datos y las tecnologías disponibles son suficientes, pero se procesan y utilizan de manera ineficaz, debido al origen estático y resistente al cambio de las instituciones del sector. No sólo es necesario llevar a cabo un proceso de digitalización sino también una revisión de las estructuras organizativas del sector.

La motivación del presente trabajo sería precisamente la de poder ayudar a subsanar en parte esta carencia, contribuyendo a la generación de bases de datos simuladas y realistas, a partir de exploraciones PET que incorporen la gran variabilidad del tejido mamario de las pacientes, para su posterior utilización en el entrenamiento de métodos de reconstrucción de imágenes por parte de los investigadores. Para alcanzar esta meta, se ha confiado en técnicas de *deep learning* (DL) o aprendizaje profundo, ya que éstas permiten trabajar con múltiples tipos de datos de entrada, lo que es especialmente importante para datos sanitarios heterogéneos, como es el caso de los tejidos mamarios.

Esta motivación ha actuado como catalizador de las propias inquietudes personales del autor, desarrolladas a lo largo de la carrera, hacia el procesamiento de imágenes y el desarrollo de modelos de aprendizaje profundo. El presente proyecto ha proporcionado la oportunidad de poder profundizar en el campo de la aplicación de algoritmos de visión por computador (*Computer Vision*) que permiten el tratamiento y comprensión de imágenes mediante el uso de redes neuronales convolucionales, o CNN (por sus siglas en inglés *Convolutional Neural Network*).

²<https://gacetamedica.com/politica/espacio-europeo-de-datos-sanitarios-el-principal-pilar-de-una-ue-de-la-salud-real-robusta-y-coherente/>

³<https://www.oecd-ilibrary.org/sites/e3b23f8e-en/index.html?itemId=/content/publication/e3b23f8e-en>

Por otra parte, la posibilidad de utilización de estas técnicas de DL en el ámbito del *Medical Imaging*⁴, un campo hasta ahora desconocido para el autor, se ha traducido en un valor añadido para el impulso de su formación académica, al permitir la puesta en contacto con las infinitas variantes de arquitecturas existentes en el campo de las redes neuronales.

Finalmente, no hemos de perder de vista que el objetivo final de este proceso de generación de bases de datos es el diseño de algoritmos cada vez más eficaces en el diagnóstico precoz del cáncer de mama, que proporcionen a los profesionales de la medicina razones plausibles para confiar en la IA.

1.2 Objetivos

El objetivo principal de este trabajo es el diseño y desarrollo de un modelo que propicie la creación de una base de datos simulada y realista de tomografías de mama por emisión de positrones. Para ello se realizará previamente un análisis de las técnicas actuales más idóneas en este tipo de tareas, como son las que forman parte de los denominados modelos generativos o redes neuronales generativas adversarias (por las iniciales en inglés de *Generative Adversarial Networks*, GAN). Con dichas arquitecturas como eje central de la investigación, el propósito es llevar a cabo una serie de experimentaciones que permitan determinar cuál es el modelo óptimo para una generación satisfactoria de muestras sintéticas.

Adicionalmente, otro objetivo del proyecto y complementario al anterior, se concretaría en descubrir cómo conseguir emplear en la misma investigación otro conjunto de imágenes clínicas de naturaleza y características diferentes a las muestras PET. Estamos hablando de las mamografías y la importancia de su uso radica en el hecho de poder llegar a entrenar con ellas el modelo propuesto, en el caso de que las imágenes sintéticas resultantes carezcan de suficiente calidad.

Sobre estas muestras procedentes de mamografías se procederá a la realización de diversas experimentaciones para determinar cuál resulta ser el modelo que mejor rendimiento ofrece en este tipo de imágenes. Los pesos de este proceso de aprendizaje serán conservados cuidadosamente y, utilizando la técnica del *transfer learning*, se aplicarán a las muestras procedentes de tomografías PET para intentar conseguir así la definición de imágenes sintéticas de mayor calidad.

1.3 Impacto esperado

Idealmente, una de las aspiraciones de este trabajo sería la de poder servir de recopilación e informe-guía de las nuevas técnicas existentes en el ámbito de la generación de imágenes sintéticas de cáncer de mama. De esta manera, cualquier usuario interesado en la materia, podría disponer de una base con la que ampliar la investigación en este campo.

Existe todavía mucho camino por recorrer hasta que el *deep learning* consiga crear nuevas imágenes de una manera perfecta propiciando, gracias a ello, la generación de diagnósticos precisos. Esto lograría reducir en buena medida el actual obligado nexo

⁴Conjunto de técnicas y procesos utilizados para la creación de imágenes del cuerpo humano.

existente entre profesionales de la salud y desarrolladores de software para lograr este objetivo. Este avance redundaría en una mejora significativa en la reducción de costes de tiempo y recursos invertidos en el proceso.

Evidentemente, y desde una perspectiva bastante más amplia, lo que perseguiría en definitiva este proyecto, y los futuros que puedan beneficiarse de las conclusiones alcanzadas en el mismo, sería la ayuda para reducir la carga de trabajo de los radiólogos y médicos en general. Hasta el momento, este objetivo ha resultado ser una cuestión realmente ardua, ya que la mayor parte de las técnicas de inteligencia artificial utilizadas en el análisis de imágenes médicas se basan en el aprendizaje supervisado, lo que requiere la existencia de una gran cantidad de datos etiquetados. Desgraciadamente, el volumen de imágenes disponibles de tomografías por emisión de positrones suele ser escaso y la etiquetación de las ya disponibles es un procedimiento costoso.

El interés de este trabajo en el estudio de las redes generativas adversarias se deriva del hecho de que éstas constituyen un método no supervisado que permite simular la distribución de imágenes reales y reconstruir otras imágenes sintéticas aproximadas. Es decir, los métodos de aprendizaje profundo permitirán generar suficientes imágenes que resuelvan el problema de la escasez de datos. Es precisamente aquí, con la puesta en marcha de nuevas perspectivas en la generación de imágenes médicas por parte de los nuevos investigadores, donde a este proyecto le gustaría dejar su pequeña contribución.

En comparación con otras técnicas de análisis de imágenes clínicas, el estudio y aplicación de las GAN en este campo aún está en estado embrionario y el volumen de la literatura disponible no es lo suficientemente abundante, sobre todo en lengua española. De ahí el deseo de este Trabajo de Fin de Grado (TFG) al respecto.

Por otra parte, otro impacto esperado adicional sería también el de fomentar el interés por el estudio de nuevas métricas de evaluación de la calidad de las imágenes generadas. Aunque actualmente ya se están creando imágenes médicas sintéticas, la calidad de éstas todavía se evalúa utilizando métodos tradicionales como *accuracy* o *loss*, que no son lo suficientemente útiles para dicho objetivo.

En este proyecto se presenta una nueva métrica empleada exclusivamente en las arquitecturas GAN, como es la *Fréchet Inception Distance* (FID). A pesar de que ya existen diversas métricas para evaluar imágenes PET reconstruidas a partir de datos directos, e incluso estándares y normativas⁵ que regulan estas imágenes médicas, lo que sería realmente gratificante es conseguir suscitar el suficiente interés por el impulso de nuevos desarrollos de métricas para muestras PET sintéticas, que mejoren significativamente la calidad de las imágenes de tomografías por emisión de positrones.

1.4 Estructura de la memoria

La presente memoria está estructurada siguiendo un patrón bastante intuitivo. En primer lugar, después de una breve introducción relativa al impacto que supone para nuestra sociedad el grave problema que supone el cáncer de mama y cómo la generación de bases de datos sintéticos puede contribuir a su estudio, se expone la motivación personal y profesional que ha suscitado el suficiente interés en el autor para abordar esta problemática, con la intención de poder ofrecer una nueva aportación a la resolución de la misma.

⁵NEMA (National Electrical Manufacturers Association). <https://www.nema.org>

Para ello, se han detallado a continuación los objetivos del proyecto, que definen la estrategia seguida para enfrentarse a este reto, así como la repercusión que se espera obtener con el logro de éstos.

Antes de avanzar en el desarrollo de la exposición, se ha considerado oportuno introducir un capítulo dedicado exclusivamente a poner en contexto al lector en el ámbito de las tomografías por emisión de positrones. Se ha estimado aconsejable hacer una referencia a la descripción de qué son, cómo funcionan, para qué sirven y cómo se generan las imágenes PET utilizadas por los radiólogos para efectuar sus diagnósticos clínicos.

Una vez alcanzada una perspectiva general sobre la cuestión, ya se dispondrá de la suficiente información para considerar de manera adecuada cuál es la actual situación del estado del arte en el campo de la generación de imágenes sintéticas clínicas y, en especial, de aquellas técnicas que pretenden emplearse en este proyecto.

La exposición de estos tres primeros capítulos permite adentrarnos ya en la descripción y análisis del problema planteado en este trabajo, junto con la pertinente solución propuesta para enfrentarse al mismo.

A partir de aquí, la estructura de la memoria cobra un cariz marcadamente teórico-técnico, como requiere a nuestro juicio un trabajo de ciencia de datos. Así, se procederá a la presentación de las denominadas redes generativas adversarias, como herramienta principal de trabajo para alcanzar las metas propuestas. Esto requerirá un esfuerzo especial para exponer previamente todo el marco teórico que subyace bajo las diferentes arquitecturas GAN, convirtiéndose de esta manera, en una de las partes más importantes de este TFG. Aquí se verá cuál es el funcionamiento interno de las GAN, su proceso de entrenamiento, los problemas de convergencia que pueden surgir durante este aprendizaje, etc., además de introducir finalmente las variantes GAN concretas que se utilizarán en la experimentación posterior.

Antes de sumergirse en este proceso de experimentación será preciso hacer referencia a los recursos de *software* y *hardware* utilizados en dicho proceso, así como de las bases de imágenes utilizadas como muestras. Este último aspecto no es en absoluto una cuestión baladí ya que, como veremos, la heterogeneidad existente entre las dos diferentes bases de datos utilizadas requerirá la justificación de su uso.

Con todo este bagaje, seguidamente podremos iniciar ya el proceso de experimentación y la obtención de resultados, detallando pormenorizadamente cuáles son los que se logren alcanzar en cada uno de los modelos propuestos.

Por último, como en todo proyecto de investigación científica, los resultados de la experimentación exigirán definir cuáles son las conclusiones a las que se han llegado con ellos, detallando las fortalezas de estos resultados pero también las carencias de los mismos. Será necesario indicar muy bien los motivos y causas de dichas carencias para, finalmente, poder precisar una hoja de ruta que contemple el desarrollo de nuevas investigaciones que puedan llevarse a cabo en un futuro en base a estas conclusiones.

Adicionalmente, al final de la memoria se han incluido una serie de apéndices que contienen información sobre aquellos temas tratados en el TFG y sobre los cuales el lector puede estar interesado en profundizar, pero que no son indispensables para la correcta comprensión del desarrollo de la memoria.

CAPÍTULO 2

Contexto

En la lucha contra el cáncer de mama existen diversos procedimientos a disposición de los profesionales de la medicina para el diagnóstico de tumores malignos. De entre ellos destaca la utilización de imágenes obtenidas de las pacientes. Aún sin estar exentas de riesgos, las técnicas de diagnóstico por imagen destacan por su eficacia ya que reducen la necesidad de cirugías exploratorias innecesarias y permiten determinar patologías concretas y el alcance de las mismas.

El propósito de este capítulo es, partiendo de una descripción de los principales procedimientos empleados en el tratamiento de imágenes diagnósticas utilizadas actualmente, llegar a comprender cómo se llegan a generar las imágenes en las tomografías por emisión de positrones.

2.1 Técnicas de diagnóstico por imágenes

Las técnicas de exploración por imágenes proporcionan a los radiólogos una herramienta para poder visualizar el interior del cuerpo humano y detectar indicios de patologías, posibilitando así un diagnóstico precoz de numerosas enfermedades y facilitando la rápida adopción de terapias que permitan ampliar las posibilidades de éxito en el tratamiento de las mismas.

Estas técnicas de diagnóstico basadas en imágenes utilizan la energía procedente de diversas fuentes emisoras para realizar su trabajo [4]. Las radiaciones ionizantes [5] generadas por rayos X (RX), producen unos efectos biológicos sobre la materia viva al desprender electrones de los átomos o moléculas irradiados. Esto puede ser utilizado en radiografías [6], mamografías, densitometrías y tomografías computerizadas (CT, *Computerize Tomography*) [7]. A diferencia de éstas, la resonancia magnética (MRI, *Magnetic Resonance Imaging*) [8] trabaja sin radiación, utilizando los pulsos de radiofrecuencia que se generan al crear un campo magnético alrededor del cuerpo. Análogamente, las ecografías tampoco utilizan la radiación, sino el envío de ultrasonidos hacia la parte del cuerpo que se quiere estudiar.

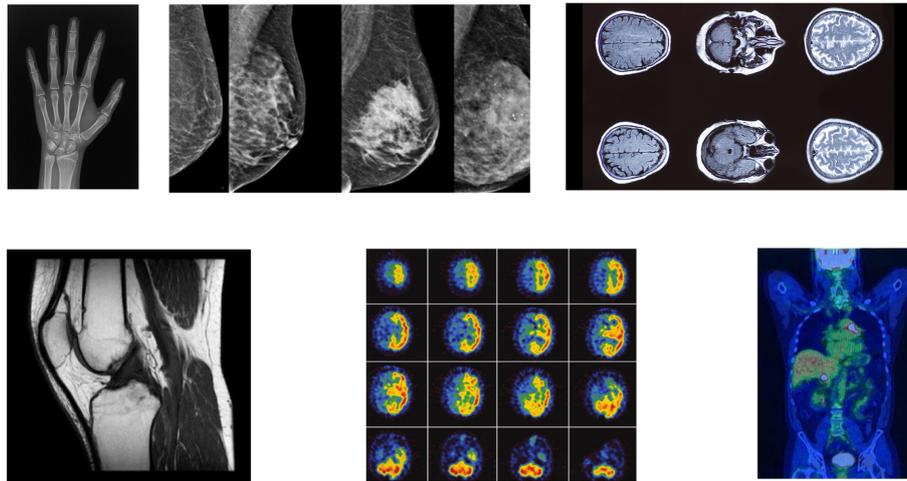


Figura 2.1: De izquierda a derecha y de arriba a abajo, imágenes de RX, ecografía, CT, MRI, SPECT y PET. Imágenes extraídas de [9][10][11][12][13] y [14].

En contraste con lo expuesto anteriormente, la medicina nuclear utiliza radiofármacos (isótopos radiactivos) que se introducen en el organismo humano y viajan a través de la zona explorada, emitiendo radiación en forma de rayos gamma que son detectados por un dispositivo especial que localiza la radiación y, posteriormente, los datos obtenidos son transferidos a una computadora para crear las imágenes del interior del cuerpo [15].

Las imágenes de medicina nuclear son pues de naturaleza distinta a las vistas anteriormente, ya que registran la radiación emitida desde el interior del cuerpo en lugar de la radiación generada por fuentes externas como los rayos X. Las dos modalidades de imágenes más comunes en medicina nuclear son la tomografía computerizada por emisión de fotón simple (SPECT, *Single-Photon Emission Computed Tomography*) y la tomografía por emisión de positrones (PET, *Positron Emission Tomography*), la cual es capaz de detectar simultáneamente dos rayos gamma γ [16], como veremos posteriormente.

En la Figura 2.1 se pueden apreciar diversos ejemplos de cada uno de los tipos de imágenes reseñados anteriormente.

Las imágenes obtenidas en las resonancias magnéticas (MRI) o en las tomografías computerizadas (CT) son de tipo *estructural*¹, es decir, sólo nos dan información sobre la estructura de los tejidos y órganos objeto de estudio. Por el contrario, las imágenes proporcionadas por la medicina nuclear son de tipo *funcional*, permitiendo obtener información sobre dónde y cómo está teniendo lugar un determinado proceso metabólico dentro del organismo. Esto permite detectar las alteraciones metabólicas previas a determinados cambios anatómicos (detectados por MRI o CT), consiguiendo así un diagnóstico precoz de la enfermedad.

La técnica PET, en la que se centra este trabajo, es de gran utilidad en el diagnóstico y seguimiento de enfermedades tumorales, en afecciones del corazón y en trastornos neurodegenerativos como la enfermedad de Alzheimer o la de Parkinson.

¹Existen resonancias magnéticas más modernas (fMRI) que son de tipo *funcional*.

2.2 Radiofármacos

La tomografía por emisión de positrones (PET) es una técnica diagnóstica basada en la detección y análisis de la distribución tridimensional que adopta, en el interior del cuerpo humano, un *radiofármaco* administrado al paciente a través de una inyección intravenosa. Esto permite medir la actividad metabólica del organismo. Según sea la parte del cuerpo que se desee estudiar, se pueden utilizar diferentes radiofármacos emisores de positrones.

Estos radiofármacos o radiosondas son el resultado de la unión de una molécula biológica normal con una sustancia radioactiva de vida media ultracorta [17]. Tienen la capacidad de acumularse en aquellas células que presentan una mayor actividad metabólica (p. ej. las células tumorales).

Al realizar una tomografía PET, lo que se obtiene es una imagen en donde se pueden visualizar esas zonas hipercaptantes. Existen diversos radiofármacos utilizados en medicina nuclear. Uno de los más utilizados es el *Flúor-18* [18F], un isótopo del flúor.

2.3 Metabolismo de la glucosa y utilización de isótopos radiactivos

Lo que hace que el Flúor-18 [18F] sea particularmente útil en medicina es su capacidad para unirse a la molécula *2-desoxi-D-glucosa* para obtener así el trazador *fluorodesoxiglucosa* (18F - FDG), una molécula similar a la glucosa. Esta característica es de suma importancia, ya que las células cancerosas son metabólicamente más activas y pueden absorber glucosa a una tasa más alta [18].

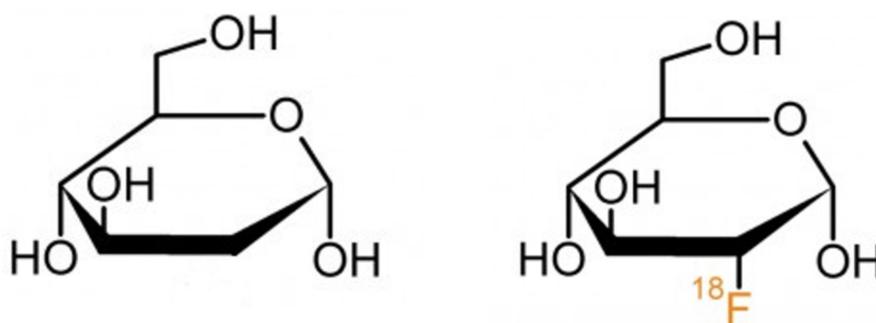


Figura 2.2: Composición química de la molécula de 2-desoxi-D-glucosa y cómo el isótopo Flúor-18 se une a ella para formar el trazador fluorodesoxiglucosa 18 F-FDG. Imagen extraída de [19].

En la figura 2.2 se ilustra cómo el isótopo radiactivo ¹⁸F se une a la molécula 2-desoxi-D-glucosa para obtener el trazador fluorodesoxiglucosa ¹⁸F - FDG.

Gracias a las emisiones de rayos gamma del trazador ¹⁸F - FDG, tenemos la posibilidad de poder detectar y evaluar el consumo de glucosa del cuerpo humano. Esto resulta ser una herramienta de vital trascendencia en la diagnosis clínica, ya que indica cuáles son las zonas del organismo con un metabolismo glucídico elevado, lo cual es un rasgo distintivo de los tejidos neoplásicos malignos.

2.4 Conceptos básicos de física de partículas

Anteriormente nos hemos referido al Flúor-18 [18F] como un *isótopo* del flúor. Es importante conocer qué es lo que subyace bajo el concepto de isótopo porque nos ayudará a entender el por qué este elemento es capaz de emitir radiación que después puede ser detectada por un tomógrafo.

Todos los átomos de un mismo elemento químico tienen en su núcleo idéntico número de protones (*número atómico*). En cambio, el número de neutrones de ese núcleo (*número másico*) puede variar de un átomo a otro y, aun así, tratarse del mismo elemento ya que no ha variado su número atómico. Sólo se trataría de versiones más pesadas o ligeras del mismo elemento químico [20].

Solamente pueden hallarse ciertas combinaciones de neutrones y protones que formen núcleos estables. Si existen demasiados o muy pocos neutrones para un número determinado de protones, el núcleo resultante no será estable y sufrirá una desintegración radiactiva [21]. El flúor, con 9 protones y 10 neutrones en su núcleo, es un elemento químico estable.

La industria farmacéutica elabora el isótopo Flúor-18 [18F] utilizando aceleradores de partículas (*ciclotrones*) mediante el bombardeo con protones de alta energía [22]. La nueva versión del elemento, producida en el laboratorio, no contiene ya 10 neutrones en su núcleo como el flúor, sino sólo 9. Este nuevo balance entre neutrones y protones hace que el isótopo generado tenga ahora un núcleo atómico inestable.

Los núcleos atómicos inestables tienden a estabilizarse espontáneamente emitiendo partículas, energía, o ambas. Este fenómeno se conoce como *radiactividad*². Esta inestabilidad es la causante de que el Flúor-18 [18F] sea radiactivo y emita energía y partículas a medida que va mutando hacia una forma más estable. Su vida media es de 110 minutos y se desintegra por emisión de positrones y radiación gamma.

Los *positrones* o antielectrones son partículas elementales subatómicas, antipartículas del electrón. Una *antipartícula* posee la misma masa que su partícula e igual carga, pero de signo contrario. Así, de la misma manera, el antiprotón es una partícula con la misma masa que el protón pero cuya carga es igual a una unidad negativa [23].

El positrón posee la misma cantidad de masa que el electrón; sin embargo, su carga es positiva e igual a (+1e), mientras que la del electrón es (-1e). No forma parte de la materia ordinaria, sino de la *antimateria*.

Las antipartículas tienen una vida muy corta, ya que cuando se encuentran con una partícula se aniquilan mutuamente, liberando energía. Esta *aniquilación* recíproca no significa su destrucción, sino una transformación que da lugar a *fotones*³ de alta energía, generando *rayos gamma* γ [24].

La tomografía por emisión de positrones detecta esta radiación y procesa la información recabada para poder obtener imágenes tridimensionales del interior del cuerpo humano.

²<https://www.ulpgc.es/sprlyupr/que-es-radiactividad>

³<https://www.sea-astronomia.es/glosario/foton>

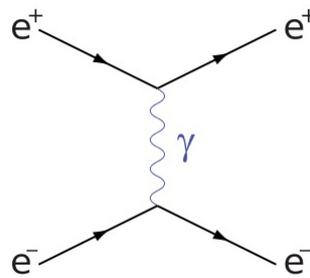


Figura 2.3: Diagrama de Feynman, donde se observa el intercambio de un fotón, representado por una línea ondulada γ , generado por una aniquilación entre un electrón y un positrón. Imagen extraída de [25].

En la Figura 2.3 se puede observar, mediante el denominado diagrama de Feynman, cómo la aniquilación entre un electrón ($-1e$) y un positrón ($+1e$) genera el intercambio de un fotón, emitiendo rayos gamma.

2.5 Generación de imágenes PET

Sabemos pues que el isótopo radiactivo ^{18}F , inyectado por vía intravenosa a la paciente, tiene una vida media de 110 minutos. Como ya hemos visto también, a medida que este radioisótopo se desintegra, emite positrones y que cada uno de estos positrones puede colisionar con un electrón, aniquilándose mutuamente.

Esta aniquilación genera la emisión de dos rayos gamma (γ) en la misma dirección pero en sentidos opuestos [26]. La radiación gamma es un tipo de radiación electromagnética, es decir, compuesta por fotones. El escáner o tomógrafo es capaz de detectar esta radiación.

La detección de los fotones gamma, emitidos por el organismo del paciente, es posible gracias a unos anillos detectores instalados en el tomógrafo. La imagen se obtiene gracias a estos escáneres.

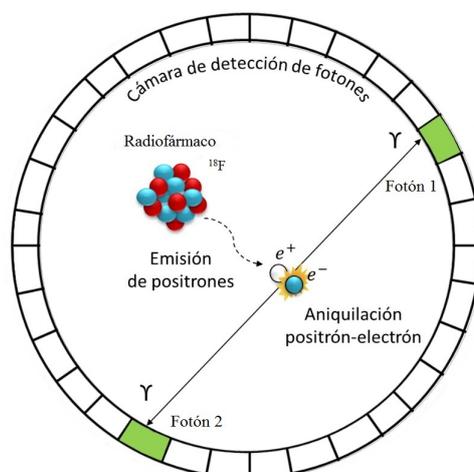


Figura 2.4: Emisión de fotones por aniquilación de electrones-positrones y su detección por el tomógrafo. Imagen extraída de [27].

En la Figura 2.4 se puede apreciar cómo se produce la generación de dos fotones debido a la aniquilación entre electrón y positrón y su detección por el anillo del tomógrafo.

Los anillos de detección del tomógrafo están constituidos por unidades independientes denominadas *bloques detectores*. Estos bloques resultan ser unos dispositivos matriciales formados por cristales de centelleo, conectados a unos fotomultiplicadores que se encargan de transformar la energía electromagnética en impulsos eléctricos⁴.

Debido a que los dos fotones derivados del proceso de colisión se desplazan en la misma dirección, pero en sentidos opuestos, el punto de aniquilación estará ubicado en una línea recta que une ambos puntos de detección. Esto propicia que ambos fotones puedan ser detectados *en coincidencia*, es decir, al mismo tiempo. La línea que une estos dos detectores recibe el nombre de *línea de coincidencia* o línea de respuesta (en inglés LOR por las iniciales de *Line Of Response*). El proceso de almacenamiento en el tomógrafo PET de los datos obtenidos, se simplifica gracias a la simetría axial de los dispositivos detectores.

La proyección de una de estas líneas de coincidencia se puede representar gracias a las coordenadas de detección obtenidas en los cristales de centelleo⁵ situados en el anillo de detección del tomógrafo $(x_1, y_1; x_2, y_2)$ [28].

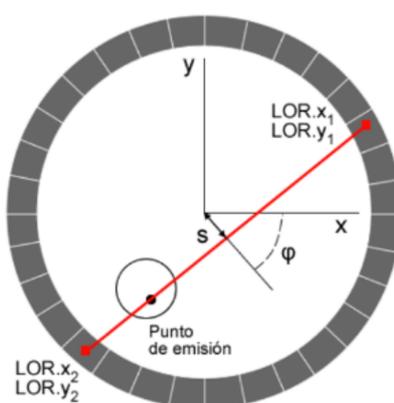


Figura 2.5: Representación de una línea de coincidencia LOR. Imagen extraída de [28].

En la figura 2.5 se puede observar una representación de una LOR, donde $f(x,y)$ es la representación bidimensional de la actividad y s es la coordenada radial.

Una vez detectados todos los sucesos, se procesan las coincidencias, se almacenan y, finalmente, se reconstruyen las imágenes en un ordenador.

⁴https://es.wikipedia.org/wiki/Tomograf%C3%ADa_por_emisi%C3%B3n_de_positrones

⁵La gran mayoría de los detectores están compuestos de cristales de centelleo, pero también existen otros tipos en los cuales el material centelleador es líquido e incluso ha habido algún intento de utilizar detectores semiconductores basados en CZT (teluro de cadmio-zinc o CdZnTe), que no centellean, sino que generan directamente pares electrón-hueco.



Figura 2.6: Representación de la técnica empleada en los tomógrafos PET. Imagen extraída de [29].

En la Figura 2.6 se ilustra un diagrama con el proceso de detección, procesamiento y reconstrucción de imágenes en un tomógrafo PET.

Existen diversos efectos físicos que pueden degradar la calidad de la imagen PET, como por ejemplo la *atenuación*⁶ que sufren los fotones al atravesar el cuerpo en estudio, la cual reduce el número de fotones detectados. Si durante el viaje a través del organismo uno de los dos fotones fuera absorbido por éste y no llegara a ser detectado, se originaría una pérdida de detección de todo el suceso. En general, el impacto de tal efecto de atenuación se traduciría en una disminución del número de sucesos detectados, de las no-uniformidades de las imágenes generadas y de la distorsión observada. Para evitarlo, es necesario llevar a cabo un preprocesamiento de los datos para intentar reducir los factores que provocan estos problemas [30].

Es necesaria pues la llegada al anillo detector de los dos fotones para la detección de eventos de positrones. La detección de coincidencia no será posible si la atenuación provoca la pérdida de cualquiera de los dos fotones. En otras palabras, la probabilidad de que ambos fotones alcancen el anillo detector del tomógrafo es la que determina el número de eventos detectados.

La señal obtenida, una vez recogida por el tomógrafo, ha de ser luego interpretada. Como en el caso de las radiografías, cada una de las imágenes obtenidas es de naturaleza bidimensional. No obstante, estas imágenes bidimensionales conseguidas desde distintos ángulos alrededor de la paciente, pueden llegar a combinarse para poder disponer así de representaciones tridimensionales. Éstas deben después ser manipuladas informáticamente para poder obtener secciones dimensionales de la mama en cualquier orientación [31].

En tomografía, la reconstrucción de imágenes de un objeto, a partir de los datos detectados, se puede plantear como la resolución de un sistema con infinitud de ecuaciones. En éstas, cada incógnita representaría un *vóxel*⁷ del volumen a reconstruir. Por otra parte, los datos de las ecuaciones vendrían dados por cada una de las líneas de coincidencia del anillo detector del tomógrafo. El empleo de diversos métodos estadísticos es el que permitiría la resolución de este sistema de ecuaciones.

⁶Desaparición progresiva de los fotones que constituyen los rayos gamma al penetrar en un medio natural. Esta disminución del número de fotones se debe a la interacción de un cierto número de ellos con los átomos que componen el medio [32].

⁷Vóxel (*volumetric pixel*). Unidad cúbica que representa la unidad mínima de un objeto tridimensional.

Los datos obtenidos por un tomógrafo PET se suelen almacenar mediante los denominados *sinogramas*. Un sinograma es un histograma bidimensional en donde se registran las coincidencias de las diversas adquisiciones mediante las coordenadas polares (r, θ) que determinan las líneas LOR. Reciben este nombre porque los máximos de este tipo de funciones se configuran en forma de función seno [33]. En cada suceso de adquisición queda registrado el número de aniquilaciones detectadas en cada una de las direcciones. Esto permite asignar a la imagen resultante un nivel de intensidad diferente para cada concentración del radiofármaco observada.

La generación de imágenes PET se realiza gracias a un *algoritmo de reconstrucción*, que permite interpretar el conjunto de eventos adquiridos de la paciente en el tomógrafo. El método de reconstrucción más utilizado es la *retroproyección filtrada* o, en inglés, *Filtered Back Projection* (FBP), debido a su bajo coste computacional. Este procedimiento, basado en el teorema de cortes de Fourier [34], es ágil y sencillo, pero adolece de una falta de empleo de información estadística.

Existen otros algoritmos que, para la reconstrucción de imágenes PET, tienen en cuenta la información estadística de los datos recogidos. Estamos hablando de los llamados *métodos iterativos* de reconstrucción de imagen. Estos algoritmos son una alternativa a FBP, pero tienen un coste computacional más alto, ya que generan imágenes de mejor calidad y definición. Dentro de estos métodos iterativos de reconstrucción de imagen se incluiría el uso del conocido algoritmo *Maximum Likelihood Expectation Maximization* (MLEM), formulado por Lawrence Shepp y Yehuda Vardi en 1982 [35].

CAPÍTULO 3

Estado del arte

A lo largo de los últimos años, la inteligencia artificial (IA) ha logrado grandes avances en el campo de la investigación sanitaria gracias al desarrollo de robustos modelos que permiten la elaboración de diagnósticos a partir del análisis de imágenes médicas. Sin embargo, para poder elaborar modelos consistentes a partir de estas imágenes se requieren cantidades masivas de datos iniciales para que los modelos de *machine learning* y *deep learning* puedan generalizar lo suficiente en el proceso de aprendizaje, es decir, evitar que el modelo simplemente se limite a memorizar las características de las imágenes introducidas (*overfitting*). Esto permitiría a los investigadores poder generar conclusiones que faciliten la toma de decisiones libres de sesgos indeseados.

Esta situación, por lo pronto, plantea un doble problema. Por una parte, la obtención de estos datos clínicos resulta un procedimiento generalmente costoso y, por otra parte, lleva aparejada una serie de impedimentos legales y éticos. El acceso a los datos sanitarios, debido a la información sensible que contienen, suele estar muy restringido debido a las diversas leyes existentes en el ámbito de la protección de datos. Estas restricciones merman las posibilidades de investigación con dichos datos [36].

3.1 Datos sintéticos

Una propuesta para solucionar esta escasez de datos puede venir de la mano del empleo de los datos generados sintéticamente. Los datos sintéticos (*synthetic data*), no son más que información generada artificialmente por un modelo matemático para replicar distribuciones de datos históricos reales. Con ellos se pueden llegar a entrenar modelos de IA cuando los conjuntos de datos reales disponibles tienen un volumen escaso o adolecen de una falta de variedad o calidad. En principio, esta alternativa resolvería la problemática ética y legal existente en el tratamiento de datos de salud, ya que preservaría la privacidad de la información sensible de los pacientes. Asimismo, con la utilización de datos sintéticos, se dispondría de una mayor variabilidad de los mismos, lo cual redundaría en una menor probabilidad de aparición de sesgos en la toma de decisiones.

La evolución tecnológica actual ha permitido la aparición de una gran variedad de procedimientos orientados a la generación de datos sintéticos, no sólo en el ámbito sanitario sino también en otros muchos sectores no relacionados con el de la salud. Ciertamente ésta no es una cuestión trivial ya que, en junio de 2021, la consultora norteamericana Gartner realizó un informe sobre datos sintéticos y predijo que para el año 2030 la mayoría de los datos empleados por la IA estarían generados artificialmente mediante modelos

estadísticos, simulaciones o técnicas parecidas. “El hecho es que no se podrán construir modelos de inteligencia artificial de alta calidad y alto valor sin datos sintéticos”, dice el informe ¹.

3.2 Métodos de generación sin entrenamiento

De entre estos métodos, podemos destacar, en primer lugar, las técnicas básicas utilizadas para tal fin en cualquier área de investigación. No sólo se emplean en el ámbito médico sino en cualquier otra área de estudio. Son técnicas que, para generar nuevos datos, no requieren del entrenamiento de los sistemas. Su implementación es sencilla y suelen ser un buen punto de partida para mejorar el rendimiento de los modelos propuestos. Existen muchas técnicas básicas para la generación de datos sintéticos, pero aquí sólo se pretende ofrecer una breve reseña de las más utilizadas actualmente. De entre ellas, destacaríamos las siguientes técnicas [37]:

- *Cropping*: esta técnica, también conocida como RICAP (por sus siglas en inglés, *Random Image Cropping and Patching Data Augmentation*) se basa en el recorte aleatorio de subimágenes o subconjuntos de características originales a las imágenes iniciales, para después construir nuevas muestras de entrenamiento.
- *Rotation*: también conocida como *Random Rotate Data Augmentation*, consiste en la rotación de una muestra de forma aleatoria en una cierta cantidad de grados, creando nuevas versiones de imágenes basadas en las originales. Esto ayuda a que el modelo aprenda cuál es el aspecto general de un objeto en lugar de tener que memorizar la forma específica en la que aparecen los objetos en los datos de entrenamiento. La rotación aleatoria puede mejorar un modelo de visión por ordenador sin necesidad de tener que recopilar y etiquetar más datos ².
- *Translation*: se basa en la creación de nuevas muestras de datos desplazando los patrones existentes en el conjunto de datos a lo largo de uno o varios ejes. La traslación sólo implica mover la imagen a lo largo de la dirección X o Y (o ambas). Esto obliga a que el modelo realice la búsqueda de objetos en todas direcciones.
- *Combination*: combinación de dos o más imágenes para obtener una nueva muestra resultante.

3.3 Métodos de generación con entrenamiento

En el epígrafe anterior se han analizado las técnicas utilizadas para obtener, de manera sencilla, “nuevas” imágenes a partir de imágenes ya existentes. Básicamente, estas técnicas se limitaban a modificar las características de la imagen de origen (posición, tamaño, etc.) para obtener nuevas muestras.

Sin embargo, si lo que realmente se pretende es obtener una imagen completamente nueva, nunca vista anteriormente, será necesario el uso de una serie de métodos que requieran un proceso de aprendizaje que permita poder definir las características de las

¹<https://la.blogs.nvidia.com/2021/07/20/que-son-los-datos-sinteticos/>

²<https://blog.roboflow.com/why-and-how-to-implement-random-rotate-data-augmentation/>

muestras de origen. Para tal fin, actualmente suelen utilizarse tanto métodos estadísticos como métodos basados en visión artificial.

3.3.1. Métodos estadísticos

3.3.1.1. Simulaciones de Monte Carlo

En el ámbito estadístico, un método muy útil que permite modelar la probabilidad de que determinados sucesos acaben ocurriendo, es el que ofrece la denominada *simulación Monte Carlo*. Esta estrategia, que se conoce también como método Monte Carlo o simulación de probabilidad múltiple, toma su nombre de la ciudad de Montecarlo (Mónaco) y de su famoso casino, al considerar su juego más famoso, la ruleta, como un claro ejemplo de mecanismo generador de sucesos aleatorios. El método Monte Carlo fue desarrollado durante la Segunda Guerra Mundial por los matemáticos John von Neumann y Stanislaw Ulam para la mejora de la toma de decisiones en condiciones de incertidumbre³.

Se trata de un procedimiento matemático utilizado para la estimación de los potenciales resultados de un determinado suceso de naturaleza incierta, soliendo emplear en procesos que no pueden predecirse fácilmente debido a la intervención de diversas variables aleatorias. Básicamente, este método se emplea para valorar los efectos del riesgo y la incertidumbre en los modelos de previsión.

La simulación Monte Carlo se distingue de otros modelos de previsión por estar orientada a la creación de modelos mediante la sustitución de una determinada distribución de probabilidad para cualquier variable con incertidumbre implícita. Su objetivo consiste en el cálculo repetido de resultados, empleando en cada ocasión un conjunto de números aleatorios situado entre el valor máximo y el mínimo de la función de la probabilidad⁴. Para llegar a completar la simulación, pueden ser necesarios miles de recálculos.

Este método ha sido de gran utilidad en el ámbito sanitario para paliar la escasez de datos clínicos disponibles a la hora de implementar una investigación científica. Un ejemplo de su uso se puede encontrar en la propuesta de Pourshahrokhi et al. [38], en donde se plantea un enfoque bayesiano para la creación de datos sintéticos sanitarios (*augmentation of healthcare data*). Para ello se utiliza el denominado *algoritmo Hamiltoniano Monte Carlo* (HMC), también conocido como *Monte Carlo híbrido*, el cual permite procesar de forma eficiente un conjunto de datos con un número de muestras disponibles muy bajo.

Este particular planteamiento de *data augmentation* permite mejorar significativamente la generalización en el proceso de aprendizaje, evitando así la aparición de sesgos en el modelo. Utilizando el HMC, un modelo puede generar muestras sintéticas a gran escala cuya distribución de probabilidad se asemeje a la distribución original de la información disponible y que sean adecuadas para el entrenamiento y la generalización de dicho modelo.

Formalmente, se puede precisar que el algoritmo HMC es un método de Monte Carlo basado en cadenas de Markov que permite la obtención de una secuencia de muestras aleatorias que se aproxima a una determinada distribución de probabilidad objetivo, para la cual realizar el muestreo directo resulta complicado. Es decir, el HMC, a partir del

³<https://programmerclick.com/article/6615816992/>

⁴<https://www.ibm.com/es-es/cloud/learn/monte-carlo-simulation>

concepto de *random walk*⁵ permite generar un conjunto de muestras sintéticas cuya distribución de probabilidad se asemeja a la distribución original, consiguiendo con ello entrenar el modelo y generalizar el proceso de aprendizaje.

Es bien conocido que los indicadores registrados de los datos de salud de un paciente pueden ser diferentes y numerosos, es decir, las diversas variables configurarían un espacio dimensional elevado. En esta situación, los modelos basados en métodos de regresión lineal o máquinas de vectores de soporte (*support vector machines*) no tendrían un rendimiento adecuado. El concepto del *random walk* permite trabajar en escenarios con espacios dimensionales elevados.

3.3.1.2. Modelos de Naive Bayes

Para intentar resolver la problemática de la escasez de datos en una investigación no sólo se utilizan métodos basados en simulaciones de Monte Carlo. Otra manera de abordar el tema es mediante la puesta en práctica de algoritmos de aprendizaje automático. Unos de los modelos más utilizados son los denominados modelos de *Naive Bayes*. Estos modelos pertenecen a un tipo especial de algoritmos de clasificación de aprendizaje automático basados en el teorema de Bayes.

Una muestra del empleo de estos modelos es la sugerida por Aviñó et al. en un artículo dedicado a la generación de datos sintéticos plausibles a partir de registros sanitarios observables [39], proponiendo la adopción de un algoritmo de aproximación para tal fin.

El anterior estudio parte de la suposición de la existencia de ciertas variables ocultas o *latentes*⁶ que representan el verdadero estado clínico de un paciente y no son observables por el investigador. Los datos observables del paciente, a los cuales sí se puede acceder, no serán más que la manifestación del verdadero estado clínico del mismo y dependen de él.

De esta manera, el procedimiento propuesto asume que los datos sintéticos sean generados por un modelo de Naive Bayes con k variables latentes, el cual está entrenado mediante un *método de momentos*⁷. El objetivo es generar muestras realistas, pero sintéticas, e indistinguibles de las observables.

3.3.2. Métodos de visión artificial

El siguiente método de reconstrucción con entrenamiento es diferente a los dos analizados en los apartados anteriores. En ellos se conseguía generar imágenes partiendo de otras ya disponibles. Aquí lo que se intentará es reconstruir imágenes a partir de datos, como en el caso del empleo de los algoritmos FBP y MLEM tratados en el epígrafe 2.5.

La visión artificial o visión por computador (del inglés, *computer vision*), es un campo de la inteligencia artificial que se centra en el análisis de imágenes y vídeos digitales con el fin de generar información procesable por un ordenador. En los últimos años ha cosechado progresos significativos y ha demostrado ser de gran aplicación en el entorno sanitario. Esta evolución ha sido posible, en gran medida, gracias al desarrollo de téc-

⁵Formulación matemática de la trayectoria final que resulta de dar una serie de paseos aleatorios.

⁶Variable que no puede observarse directamente pero que se pueden inferir mediante un método matemático a partir de otras variables que sí son observables.

⁷Método estadístico utilizado para estimar los parámetros poblacionales de un modelo a partir de un conjunto de datos.

nicas basadas en redes neuronales convolucionales o CNN (por las siglas en inglés de *convolutional neural network*).

Las CNN son útiles en diversas aplicaciones dentro del *machine learning*, sin embargo, donde ha conseguido revolucionar el estado del arte es en el campo de la visión artificial y el procesamiento de imágenes⁸. Durante mucho tiempo se ha estado intentado utilizar la IA para resolver problemas pero, con la visión convencional, se requerían algoritmos muy complejos. Esto ha cambiado radicalmente durante los últimos años con el aprendizaje profundo y las CNN, estableciéndose así un nuevo paradigma.

Los modelos de generación de datos sintéticos han ido adoptando progresivamente estos avances proporcionados por las CNN en el campo de la visión artificial. En esta línea, uno de los más significativos progresos desarrollados recientemente es la arquitectura propuesta por Häggström et al. para la reconstrucción de imágenes de tomografías por emisión de positrones [40]. Este nuevo modelo ha sido bautizado como *DeepPET*.

Aunque es cierto que una parte de DeepPET, como red neuronal que es, se encarga del entrenamiento del modelo, básicamente debe considerarse como un método de reconstrucción de imágenes a partir de datos.

Como ya se vio en el epígrafe 2.5, dedicado a la exposición del proceso de generación de imágenes PET, los datos de proyección de una tomografía (sinogramas) no pueden ser interpretados directamente, sino que deben reconstruirse previamente mediante un algoritmo de reconstrucción. Los más utilizados son, como vimos, la retroproyección filtrada y los llamados métodos iterativos de reconstrucción. Sin embargo, por lo general, las imágenes resultantes de la aplicación de estos métodos estándar adolecen de diversos desajustes que pueden provocar la aparición de distorsión en las imágenes reconstruidas.

El modelo DeepPET propuesto está formado por una red neuronal convolucional profunda de codificación-decodificación (*encoder-decoder*) que utiliza el aprendizaje supervisado para resolver directamente el problema inverso de la reconstrucción de imágenes PET, tomando como datos de entrada los procedentes de los sinogramas y generando automáticamente imágenes PET de alta calidad.

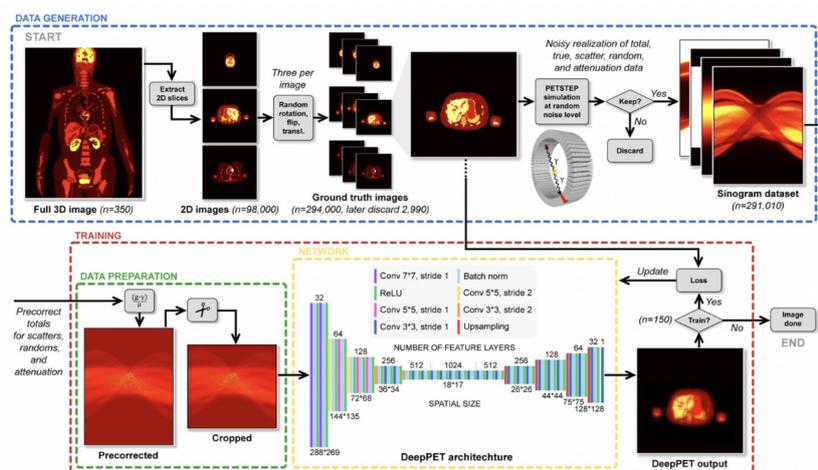


Figura 3.1: Arquitectura DeepPET. Imagen extraída de [40].

⁸<https://www.ellaberintodefalken.com>

En la Figura 3.1 puede apreciarse todo el proceso de reconstrucción de imágenes PET propuesto en el modelo DeepPET.

3.4 Crítica al estado del arte

A partir de una primera exploración de la literatura disponible en Riunet (Repositorio Institucional de la Universitat Politècnica de València), se ha podido constatar la existencia de diversos proyectos realizados en la Escola Tècnica Superior d'Enginyeria Informàtica (ETSInf) relacionados con el tratamiento y procesamiento de imágenes de diagnóstico médicas, procediendo algunas de ellas de tomografías por emisión de positrones. En dichos estudios, cada uno de ellos con un objetivo distinto, se puede valorar el proceso de aplicación de las diferentes técnicas y herramientas disponibles en el estado del arte, utilizadas para poder alcanzar las metas propuestas.

De entre todos estos proyectos, se ha considerado oportuno destacar tres de ellos, presentados como trabajos de fin de grado en Ingeniería Informática, por entender que guardan suficiente relación con algunos de los conceptos tratados en la presente memoria. El propósito es el de poder llegar a identificar y analizar las posibles coincidencias o divergencias de nuestra propuesta con dichos proyectos.

Para tal fin, según el criterio personal del autor de este texto, se procederá a señalar cuál ha sido el propósito principal de cada uno de estos trabajos, para así poderlos confrontar con los planteamientos desarrollados en nuestra memoria. Siguiendo este enfoque, se han querido poner de relieve los dos escenarios principales que se exponen en estos tres trabajos seleccionados, obviamente bajo un criterio puramente subjetivo. Así, se ha podido identificar, por un lado, una inquietud por el desarrollo de *librerías software* para el análisis de imágenes y, por otro, un interés en el *diseño de redes neuronales convolucionales*, en el campo también del análisis de imágenes clínicas.

Dentro del escenario inicial dedicado a las denominadas *software libraries*, el autor del primer trabajo, Pablo Poves Carrasco [41], propone la creación de una biblioteca software, mediante el empleo del lenguaje de programación Python, que permita llevar a cabo un análisis automatizado de la calidad de imágenes PET, para así poder evaluar la calidad de éstas. El objetivo perseguido se concreta en la voluntad de que los usuarios de esta biblioteca puedan integrar los nuevos recursos ofrecidos en ella en su propio software, aportando flexibilidad para crear nuevas librerías que compartan características con ésta.

En el segundo de los contextos mencionados anteriormente, donde se ponía el énfasis en el diseño de modelos de redes neuronales convolucionales, se ha considerado conveniente destacar otros dos proyectos de fin de grado.

En el primero de ellos, el autor Álvaro López Chilet [42] presenta un sistema que posibilita la clasificación de un conjunto de imágenes de resonancia magnética cerebral en el diagnóstico médico de la enfermedad de Alzheimer, mediante técnicas de aprendizaje profundo. Si bien es cierto que en este estudio las imágenes analizadas no proceden de tomografías PET como en nuestro caso, se ha creído aconsejable tenerlo en consideración debido a dos razones principales:

- Las imágenes de resonancia magnética utilizadas guardan cierta relación con las que se utilizan en la presente memoria y, asimismo, se ha demostrado que la combinación de estas imágenes con las obtenidas por tomografías PET “aporta una alta fiabilidad en el diagnóstico del cáncer”, según los resultados de un estudio de investigación desarrollado por especialistas de la Clínica Universidad de Navarra ⁹.
- En este trabajo seleccionado se utilizan variantes de las redes neuronales convolucionales como técnica principal en la tarea de clasificación. Estas arquitecturas son el pilar básico para el desarrollo de las redes neuronales generativas adversarias. Dichas redes, que permiten reconocer y estudiar los patrones de un conjunto de imágenes, son uno de los objetos de estudio y aplicación en el proyecto presentado en esta memoria.

El segundo TFG seleccionado, igualmente relativo al desarrollo de redes neuronales, está muy relacionado con el anterior ya que también utiliza las CNN para el diagnóstico del Alzheimer. Para tal fin, el autor del estudio, Rafael Vicente Sánchez Romero [43], sí utiliza en este caso las imágenes procedentes de tomografías por emisión de positrones así como también las obtenidas por resonancia magnética.

El nivel de estos tres trabajos aquí revisados es ciertamente excelente y sus propuestas son relevantes, ya que ponen a disposición de la comunidad científica soluciones que pueden permitir a los profesionales de la salud realizar diagnósticos precoces de enfermedades graves. Sin embargo, en todos ellos, la investigación se realiza a partir de bases de imágenes completas y perfectamente disponibles, además de conjuntos de datos ya etiquetados. En estos tres proyectos, los conjuntos de datos disponibles han sido suficientes para poder entrenar los modelos de clasificación de imágenes propuestos.

Desgraciadamente, como ya se apuntó en el epígrafe 3.1, para poder alcanzar un rendimiento elevado de los modelos de inteligencia artificial que puedan proponerse en otros proyectos de investigación, resulta indispensable poder disponer de un extenso conjunto de datos, o *dataset*, para que dichos modelos alcancen una suficiente generalización durante el proceso de aprendizaje.

En el trabajo aquí presentado se ha querido resaltar la dificultad añadida que se puede encontrar en cualquier investigación basada en imágenes PET de cáncer de mama, resultado del déficit en la disponibilidad de éstas. A ello se le ha de sumar la propia complejidad del tejido mamario, que dificulta el estudio y clasificación de estas imágenes.

Estas serían principalmente las razones que diferenciarían el punto de partida del análisis de este proyecto con el de los otros tres trabajos comparados. La escasez de conjuntos de datos lo suficientemente extensos en el campo de la investigación del cáncer de mama, que dificulta la obtención de un diagnóstico clínico acertado y precoz, ha sido justamente el catalizador del interés en el proyecto ahora presentado. Este innovador objetivo pretende poder generar conjuntos de datos sintéticos cuya verosimilitud con los datos reales ya existentes sea lo más fiable posible. Para ello, como se verá a continuación, se propone el empleo de redes neuronales generativas adversarias GAN como herramienta principal de uso.

⁹<https://www.cun.es/actualidad/noticias/pet-resonancia-cancer-prostata>

CAPÍTULO 4

Descripción del problema

Para la precoz detección de posibles lesiones neoplásicas de mama, el empleo de redes neuronales convolucionales (CNN) en la clasificación de imágenes procedentes de tomografías por emisión de positrones (PET), puede llegar a ser de inestimable ayuda a los radiólogos y oncólogos en general.

No obstante, dichos modelos de redes neuronales necesitan alimentarse de una gran cantidad de datos para poder implementar en ellos un correcto entrenamiento y permitir así ofrecer un rendimiento satisfactorio.

El problema surge cuando no existe suficiente disponibilidad de muestras para llevar a cabo un proceso de entrenamiento adecuado en los distintos modelos de aprendizaje profundo propuestos. Debido a ello, el modelo puede no ser capaz de generalizar de forma óptima y, por lo tanto, no tener la capacidad de discernir si una imagen en concreto resulta ser verdadera o falsa.

Las redes generativas adversarias (GAN) nos ofrecen una original propuesta de solución a esta problemática. Este tipo de redes sí son capaces de generar imágenes sintéticas con una distribución de probabilidad que se asemeja lo mejor posible a la distribución de probabilidad de las tomografías PET originales. En este proyecto se intentará exponer, paso a paso, todo el proceso con el cual se pretende resolver el problema de la falta de generalización de los modelos debido a la escasez de muestras disponibles para su entrenamiento.

El potencial que proporcionan las redes GAN permite que éstas puedan emplearse en diversas aplicaciones, no sólo en el campo de la salud, sino también en otros muchos ámbitos como el diseño industrial, generación de imágenes de rostros humanos, súper resolución, *text-to-image translation*¹, etc.

4.1 Análisis del marco legal y ético

El presente trabajo se ha podido llevar a cabo gracias a imágenes de mama reales obtenidas de dos fuentes principales. Por un lado, las proporcionadas por el Instituto de Instrumentación para Imagen Molecular (UPV - i3M) y, por otro lado, las imágenes de libre acceso obtenidas en la página web de la *University of South Florida*².

¹<https://towardsdatascience.com/text-to-image-a3b201b003ae>

²<http://www.eng.usf.edu/cvprg/Mammography/Database.html>

El Instituto i3M es un centro de investigación mixto creado en 2010 por la Universitat Politècnica de València (UPV) y el Consejo Superior de Investigaciones Científicas (CSIC) con sede en el Campus de Vera de la UPV. El *Medical Imaging Reconstruction Group* (MIRG), liderado por la doctora María José Rodríguez Álvarez, es uno de los grupos pertenecientes al i3M y tiene como actividad prioritaria “la investigación de nuevas técnicas de reconstrucción de imágenes para aplicaciones en el campo biomédico”³.

Ante la perspectiva de la utilización de imágenes médicas facilitadas por estos proveedores, la pregunta que todo profesional debe plantearse es la licitud del uso de dichos datos. No hemos de olvidar que las imágenes para el diagnóstico médico son consideradas como información sensible, por lo cual es primordial un tratamiento prudente y cuidadoso de este tipo de datos. Es por este motivo que los profesionales de la salud están obligados por ley a garantizar la privacidad y confidencialidad de los datos de cada paciente.

Conviene pues conocer cuándo es necesario el consentimiento del paciente y en qué supuestos es posible recurrir a otra base de legitimación para el tratamiento de datos de salud. El Reglamento General de Protección de Datos (RGPD) [44], que entró en vigor el 24 de mayo de 2016, es la normativa europea que regula la protección de las personas físicas en lo concerniente al tratamiento de sus datos personales y a la libre circulación de éstos. La propia RGPD disponía la actualización de las leyes nacionales de los países miembros a la nueva norma europea. Por dicho motivo, la Ley Orgánica de Protección de Datos de Carácter Personal (LOPD), vigente desde 1999, tuvo que actualizarse y acabó siendo sustituida por la Ley Orgánica de Protección de Datos Personales y Garantía de los Derechos Digitales (LOPDGDD) [45] el 6 de diciembre de 2018.

El Reglamento recuerda que se deberá permitir que los pacientes puedan dar su consentimiento para el tratamiento de sus datos de salud en “determinados ámbitos de investigación científica que respeten las normas éticas reconocidas para la investigación científica”⁴. Este consentimiento es, de hecho, uno de los pilares en la legitimación del tratamiento de datos personales, tal y como se establece en el art. 6.1.a) del RGPD.

Por consentimiento, el RGPD contempla “toda manifestación de voluntad libre, específica, informada e inequívoca por la que el interesado acepta, ya sea mediante una declaración o una clara acción afirmativa, el tratamiento de datos personales que le conciernen”⁵. Esto exige que dicho consentimiento se obtenga mediante un claro acto afirmativo, excluyéndose el consentimiento tácito.

Sin la existencia de este consentimiento expreso de la persona interesada, el tratamiento de estos datos de salud sólo será posible en el caso en que concurra alguna de las circunstancias enumeradas en el apartado 2 del art. 9 del RGPD y resumidas a continuación:

³<https://www.i3m-stim.i3m.upv.es/research/medical-imaging-reconstruction-group-mirg/#Members>

⁴Considerando 33 del RGPD.

⁵Art. 4.11) del RGPD.

- Por razones de interés público esencial, según lo que estipule el Derecho de la Unión o el de los Estados miembros (letra g).
- Por razones de interés público en el ámbito de la salud pública (letra i).
- Cuando sea de interés público con fines de archivo, de investigación científica o histórica o fines estadísticos (letra j).

Estas excepciones tan estrictas al tratamiento de datos sin consentimiento previo se derivan del hecho de que los datos de salud están incluidos dentro de las categorías especiales de datos personales, reguladas por el art. 9.1 RGPD, así como también por la Disposición Adicional decimoséptima de la LOPDGDD.

En lo referente a la investigación biomédica, la normativa española se sitúa en la misma línea que el RGPD, exigiendo que, para poder llevar a cabo investigaciones de esta índole, sea preciso “el cumplimiento de los derechos, la protección de la salud y la garantía de la seguridad de la sociedad, los pacientes y los profesionales involucrados en la actividad de investigación”⁶. Asimismo, según la normativa específica aplicable en este ámbito de investigación, para la realización de tareas de esta naturaleza será necesario recabar el consentimiento previo, expreso y por escrito de los pacientes, una vez éstos hayan recibido la información oportuna⁷.

En resumen, para el tratamiento de datos con fines de investigación biomédica la regla general es el consentimiento del paciente, tal y como subraya el informe 073667/2018 de la Agencia Española de Protección de Datos (AEPD)⁸.

No obstante, el art. 89 del RGPD abre la posibilidad al tratamiento de datos de salud sin la obligatoriedad de la obtención del consentimiento siempre y cuando se respeten las debidas garantías relativas a los derechos y libertades de las personas interesadas y, además, sólo en los casos que hagan referencia al tratamiento de datos “con fines de archivo en interés público, fines de investigación científica o histórica o fines estadísticos”.

Dichas garantías se concretarían en la toma de aquellas medidas que fueran necesarias para evitar una posible identificación futura de los pacientes. De entre estas medidas destacaríamos la *seudonimización* de los datos obtenidos.

Llegados a este punto es importante distinguir entre los conceptos de datos seudonimizados y datos anonimizados. Una de sus mayores diferencias reside en la normativa legal que protege a cada uno de ellos [48]. Mientras que los datos anonimizados no se encuentran bajo el ámbito de aplicación del RGPD, el conjunto de datos seudonimizados sí lo están.

La información anónima hace mención a un conjunto de datos que “no guarda relación con una persona física identificada o identificable, ni a los datos convertidos en anónimos de forma que el interesado no sea identificable, o deje de serlo”⁹. Es decir, “los datos se considerarán anonimizados en la medida que no exista una probabilidad razonable que cualquier persona pueda identificar a la persona física en el conjunto de datos”.

⁶Art. 44.c) de la Ley 16/2003, de 28 de mayo, de cohesión y calidad del Sistema Nacional de Salud [46].

⁷Arts. 4.1, 5, 13, 45 y 48.1 de la Ley 14/2007, de 3 de julio [47].

⁸<https://www.aepd.es/es/documento/2018-0046-investigacion-biomedica.pdf>

⁹Considerando 26 del RGPD.

La información seudonimizada haría referencia a un tipo de datos que ya no pueden “atribuirse a un interesado sin utilizar información adicional, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física identificada o identificable”¹⁰.

Por el contrario, en lo referente a los datos anonimizados, el RGPD sólo velará por la fiabilidad del método utilizado para anonimizar dichos datos, para evitar así cualquier posible revelación ulterior de la identidad de las personas titulares de los mismos.

Todo este armazón normativo no hace más que recoger las diversas cuestiones éticas que se plantean en el inicio de cualquier investigación científica, debatidas y discutidas en diferentes asambleas y congresos mundiales. Así, el Código de Núremberg de 1947 ya defendía que “el consentimiento voluntario del sujeto humano es absolutamente esencial”¹¹, aunque dejaba el control ético de la investigación en manos del propio científico.

En el año 1964 se produjo un gran avance en el planteamiento de estos aspectos éticos derivados de las investigaciones científicas que incorporan datos personales de pacientes. La Declaración de Helsinki, adoptada 18ª Asamblea Médica Mundial en junio 1964, planteó la propuesta de unos principios éticos para la investigación biomédica, incluidos los datos. Su objetivo era el de establecer un control ético en la investigación¹².

Sin pretender llegar a formular ningún juicio de valor al respecto, no obstante resulta razonable afirmar que, después de todo lo visto anteriormente, la legislación actual, tanto la española como la europea, recoge sobradamente esta preocupación por el respeto de los planteamientos éticos implícitos en cualquier investigación biomédica.

En el caso del presente trabajo, y a la vista del análisis normativo realizado previamente, parece plausible inferir que el uso de las imágenes disponibles para realizar este trabajo, independientemente de aspectos que desconocemos como el método de obtención en origen del consentimiento de sus titulares, ha tenido una indudable finalidad de investigación científica de acorde con lo dispuesto, tanto por el RGPD como por la LOPDGDD. Además, la solidez ética de dicha finalidad viene reforzada por el rasgo distintivo de la anonimidad de las imágenes utilizadas.

4.2 Solución propuesta

En este epígrafe se intentará aportar una visión general de la propuesta planteada en la presente memoria para llegar a alcanzar el objetivo principal del proyecto presentado.

Para la realización de este trabajo se tomó, como punto de partida, el acceso a la base de imágenes de tomografías por emisión de positrones facilitada por el Instituto de Instrumentación para Imagen Molecular (i3M) de la Universitat Politècnica de València (UPV). Aunque las características de dicha base de datos se analizarán de manera pormenorizada en el subepígrafe 6.2.1, se puede indicar ya que ésta contiene alrededor de 400 imágenes procedentes de tomografías PET de mama.

¹⁰Art. 4.5 del RGPD.

¹¹<http://www.san.gva.es/documents/151744/228971/36codigodenuremberg.pdf>

¹²<https://www.wma.net/es/polices-post/declaracion-de-helsinki-de-la-amm-principios-eticos-para-las-investigaciones-medicas-en-seres-humanos/>

El desarrollo propuesto del trabajo sigue el siguiente esquema:

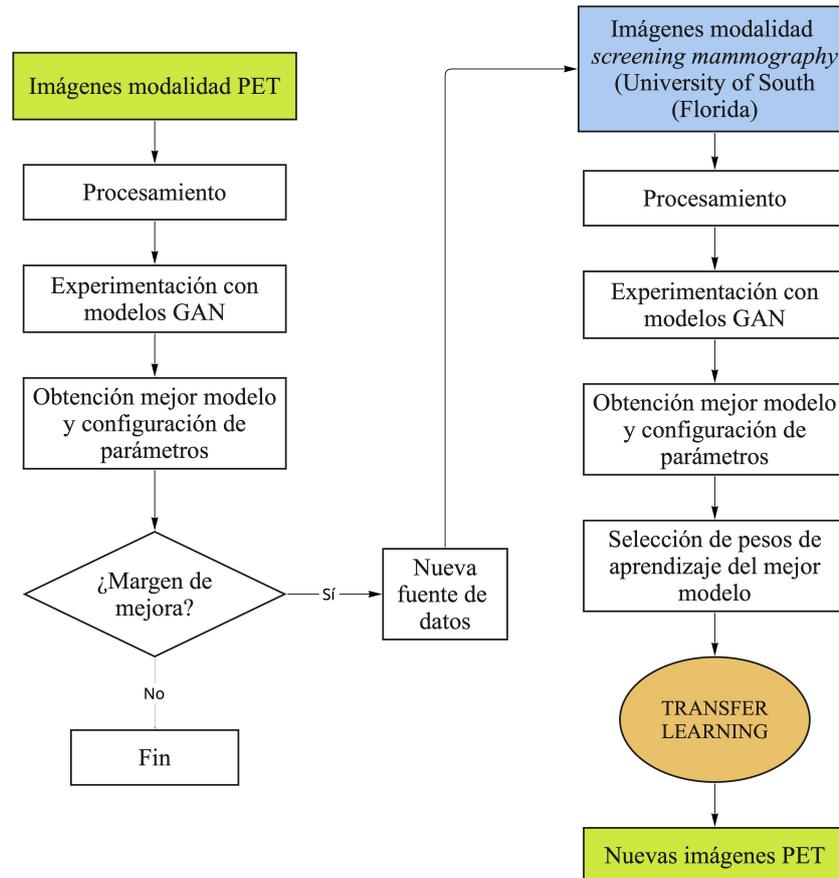


Figura 4.1: Esquema de la solución propuesta.

En primer lugar, y como etapa más importante del proceso, se llevaron a cabo un conjunto de tareas de procesamiento de las imágenes suministradas por el i3M para poder asegurar que dichas representaciones pudieran ser utilizadas correctamente como entrada o *input* de cada una de las arquitecturas de redes neuronales generativas adversarias (GAN) empleadas en el estudio.

El procesamiento previo de dichas imágenes facilitó posteriormente la realización de diversos experimentos con estas redes generativas adversarias GAN, cuyo desarrollo se expondrá con detenimiento en el capítulo 7. Estos experimentos permitieron dilucidar qué modelo utilizado resultó ser, finalmente, el que obtenía el mejor rendimiento. Lógicamente, junto con la elección del mejor modelo, se seleccionaron los parámetros óptimos correspondientes. Para ello fue necesario el empleo de la *Fréchet Inception Distance* (FID), una nueva métrica utilizada exclusivamente en tareas en las que intervienen modelos generativos.

El modelo óptimo seleccionado proporcionó una serie de imágenes de mama sintéticas creadas por la GAN. A partir de ahí, el siguiente paso supuso la observación de los detalles de mama de estas imágenes, calibrando mediante la FID si existía o no margen de mejora en la generación de nuevas imágenes.

Si en efecto, la métrica utilizada nos informaba de que sí se podía avanzar en la mejora de las imágenes sintéticas, el escenario quedaba abierto a la introducción de una nueva base de datos para el refuerzo del entrenamiento del modelo. En nuestro caso, las nuevas imágenes utilizadas procedían de un repositorio de la *University of South Florida*, del cual hablaremos detalladamente en el subepígrafe 6.2.2. Baste decir por el momento, que estas muestras presentaban el resultado de diversas pruebas de mamografías (*screening mammography*), un método de diagnóstico diferente al de las PET.

Evidentemente, la procedencia de dichas imágenes no era la misma que la utilizada en un principio, lo cual suponía el riesgo de que su empleo en nuestro modelo pudiera llegar a distorsionar los resultados obtenidos. La heterogeneidad de los datos empleados podía hacer muy difícil la correlación de manera efectiva de los mismos y la cuantificación de sus similitudes.

Bajo la supervisión de los tutores del TFG se consideró oportuno, dentro de nuestro modelo GAN, avanzar en la exploración en el campo de la síntesis *cross modality* para llegar a generar imágenes PET a partir de muestras de mamografías. Estas técnicas de síntesis todavía se encuentran en estado embrionario y, como hemos dicho, permiten “sintetizar muestras de una modalidad de imagen a partir de imágenes adquiridas con una modalidad diferente” [49]. Se espera que, en el futuro, estos nuevos procedimientos faciliten la reducción de costes de atención y mantenimiento de los sistemas de salud, así como la agilización de los diferentes procesos médicos en los que intervengan un elevado número de pacientes.

De esta manera, con las nuevas muestras de mamografías de la *University of South Florida*, se llevó a cabo el mismo proceso que el ejecutado con las imágenes PET. Es decir, se realizó la correspondiente experimentación para así obtener la mejor configuración de parámetros posible. Los pesos obtenidos en este proceso de aprendizaje fueron seleccionados y guardados. Aplicando dichos parámetros a las imágenes PET originales, mediante la técnica de *transfer learning*, se pudo llegar a comprobar si los resultados finales obtenidos mejoraban sustancialmente los iniciales, ya que este procedimiento nos permitía “transferir los conocimientos adquiridos en una o varias tareas y utilizarlos para mejorar el aprendizaje en una tarea de destino relacionada.” [50]

CAPÍTULO 5

Redes generativas adversarias GAN. Marco teórico

5.1 Introducción

A lo largo de los últimos años, la inteligencia artificial (IA) ha tenido un impacto significativo en la sociedad, siendo de vital importancia en varios sectores como la ciberseguridad, economía, educación, entre otros. Además, gracias a su capacidad de aprendizaje, la IA está llevando a cabo una transformación en el área de la salud, concretamente en el desarrollo de nuevos tratamientos médicos, diagnóstico de enfermedades, gestión y prestación de servicios sanitarios, etc. Lamentablemente, a pesar de dichos avances, la IA no ha llegado a alcanzar el máximo de su potencial debido a la poca disponibilidad de datos clínicos y de salud existentes.

En el año 2014 el ingeniero informático Ian J. Goodfellow desarrolló una original idea que permitiría afrontar esa problemática de escasez de datos. Goodfellow, que también ha trabajado como científico investigador en Google Brain y, hasta abril de 2022 ¹, como director de aprendizaje automático en Apple, desarrolló durante su etapa de doctorado en la Universidad de Montreal un concepto totalmente disruptivo que se convertiría en uno de los mayores avances en la historia de la IA: las *redes neuronales generativas adversarias* o antagonicas (GAN, por las siglas en inglés de *Generative Adversarial Nets*). [51]

En ese momento, ya se estaban empleando redes neuronales como modelos generativos plausibles de datos sintéticos. Pero aunque la IA era capaz de reconocer imágenes de manera satisfactoria, no las sabía crear de forma realista. Para que los ordenadores pudieran generar imágenes por sí solos, el método empleado hasta entonces consistía en utilizar análisis estadísticos complejos de los elementos que componían las imágenes originales, es decir, se debía informar al ordenador qué era exactamente todo aquello que aparecía en los datos de entrenamiento. A pesar de este esfuerzo, los resultados no eran demasiado alentadores: además de demostrar ser un proceso costoso, las imágenes generadas solían presentar ruido o errores.

¹<https://www.macrumors.com/2022/05/07/apple-director-of-machine-learning-resigns/>

Goodfellow tuvo la genial idea de enfrentar entre sí dos redes neuronales ². En esta novedosa arquitectura, ambas redes estarían entrenadas con el mismo conjunto de datos reales. La primera red se encargaría de generar datos sintéticos lo más realistas posible, por eso recibió el nombre de *red generativa*. La segunda red, conocida como *discriminadora*, compararía los datos sintéticos creados por la red generativa con los datos del conjunto de entrenamiento, para así intentar descubrir cuáles eran reales y cuáles no. Ambas redes competirían entre sí, de ahí el nombre de “adversarias”.

Se puede decir pues, que el objetivo principal de las GAN es crear información sintética a partir de un conjunto de datos reales y hasta el punto de que las muestras generadas sean indistinguibles de los datos reales.

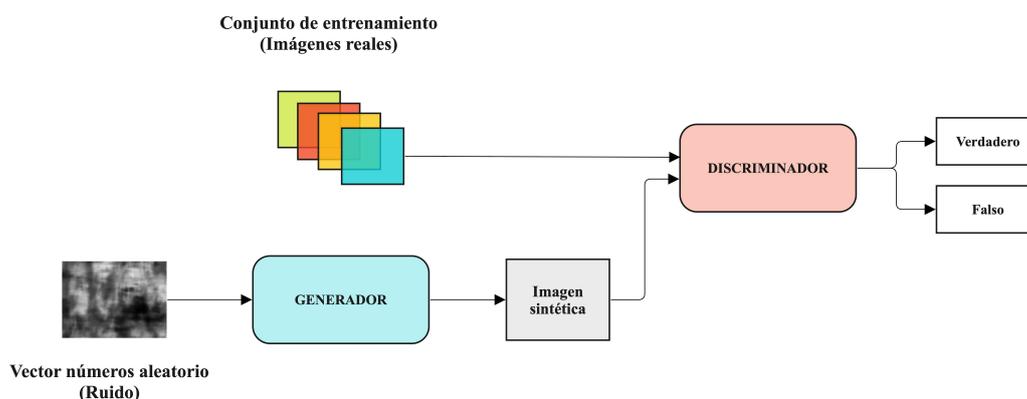


Figura 5.1: Red neuronal generativa adversaria.

En la Figura 5.1 se puede observar, a grandes rasgos, cómo funciona una red neuronal generativa adversaria, formada por dos redes neuronales antagónicas. La primera de éstas, la generadora, recibe los datos de entrada en forma de vector de valores numéricos aleatorios y se encarga de crear una nueva imagen sintética de salida. Paralelamente, existe otra red neuronal, la discriminadora, que se encarga de revisar el conjunto de imágenes producidas por la red generadora y que va aprendiendo a discernir si éstas son reales o falsas. En función de las suposiciones de la red discriminadora, la red generativa reajusta sus parámetros para crear nuevas imágenes. Y así sucesivamente hasta que la red discriminadora es incapaz de distinguir lo sintético de lo real.

5.2 Conceptos teóricos de las redes neuronales convolucionales

Debido a que la mayoría de variantes de las GAN están formadas por redes neuronales convolucionales (CNN), es importante conocer algunos aspectos fundamentales de dicha arquitectura ya que, en la evaluación de los modelos y su correspondiente experimentación, dichos conceptos serán objeto de desarrollo.

²<https://www.technologyreview.es/s/10016/el-senor-de-las-gan-el-hombre-que-dio-imaginacion-las-maquinas>

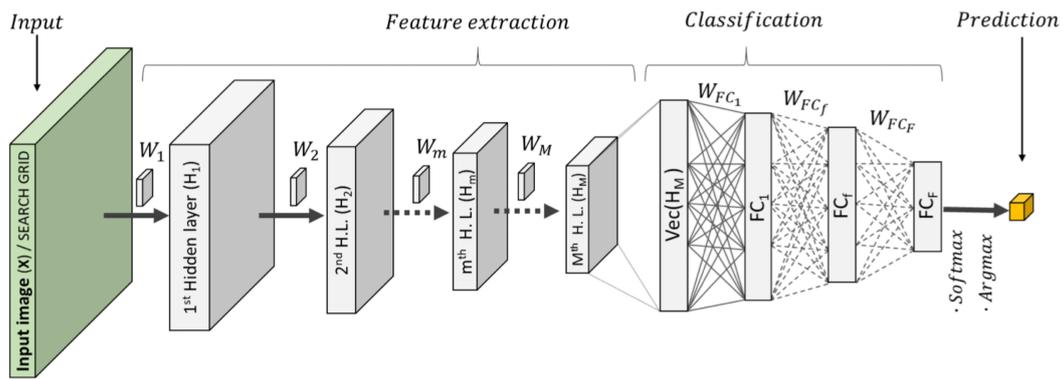


Figura 5.2: Arquitectura de una red neuronal convolucional. Imagen extraída de [52].

Como se puede observar en la Figura 5.2, una red neuronal convolucional está formada, básicamente, por dos bloques. Estos bloques reciben el nombre de bloque convolucional (*feature extraction*) y bloque de clasificación (*classification*).

Para la correcta comprensión de los dos bloques que forman una red neuronal convolucional, será de gran ayuda repasar previamente el concepto de convolución lo cual, a su vez, servirá de introducción al tema.

5.2.1. Concepto de convolución

Una convolución es una función que transforma una señal de entrada en una nueva señal de salida. Formalmente, se puede definir como el producto de dos funciones, f (*función original*) y g (*función base*), cuyo resultado es una tercera función que define la magnitud de ambas cuando éstas se superponen.

Así, suponiendo que tenemos dos funciones, $f(x)$ y $g(x)$, la convolución sería igual a una integral que representaría la magnitud en que se superpone una función g a medida que se desplaza sobre la función f . La función de convolución se define como $(f * g)(x)$ y se expresa como

$$(f * g)(x) = f(x) * g(x) = \int_{-\infty}^{+\infty} f(x') g((x - x')) dx' \quad (5.1)$$

donde x' es una variable de integración.

En el caso de redes neuronales que utilizan imágenes, una de las funciones representaría la matriz de una imagen. El filtro que se deslizará a lo largo de esa imagen estará representado por la otra función. Estos filtros no son más que matrices y más adelante profundizaremos en el empleo de los mismos.

5.2.2. Bloque convolucional

Siguiendo el esquema de la Figura 5.2, se puede decir que el bloque convolucional es el que se encarga de observar todos los detalles representativos de las imágenes que se incorporan como datos de entrada a la red. Llegados a este punto, es importante diferenciar varios conceptos importantes implicados en el funcionamiento de este bloque:

- *Capas convolucionales*: Este bloque incluye varias capas de convolución. Tienen un papel crucial en la obtención de las características principales de las imágenes, tomando a éstas como valores de entrada y asignándole diferentes pesos a ciertos elementos de las mismas según su importancia.
- *Función de activación*: Es uno de los elementos esenciales de una red neuronal. Se encarga de definir el resultado obtenido en cada unidad básica de la red (*neurona*), el cual se utiliza como valor de entrada para la siguiente neurona. Existen una gran variedad de funciones de activación, cada una empleada según el objetivo que se persiga. Las funciones más utilizadas suelen ser la *sigmoide*, la *tangente hiperbólica*, la función *rectificadora* o *Rectified Linear Unit* (ReLU) y la *unidad lineal exponencial* o *Exponential Linear Unit* (ELU). [53]
- *Pooling* o capa de submuestreo: Es la función que permite obtener una nueva representación de la imagen de entrada, reduciendo su tamaño para que pueda ser utilizada en el proceso de aprendizaje de nuevos datos. Esto optimiza los cálculos en dicho proceso, ya que reduce el número de parámetros a entrenar por la red.

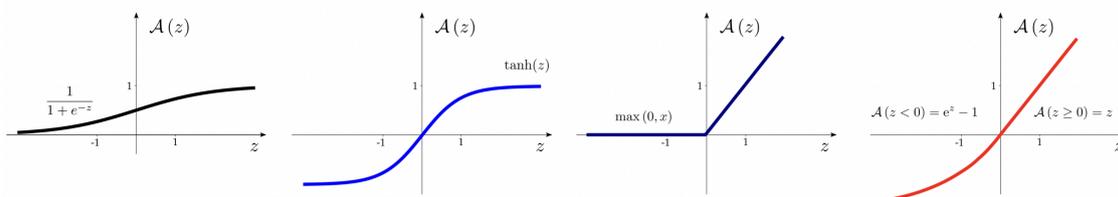


Figura 5.3: De izquierda a derecha se pueden ver las funciones de activación sigmoide, tangente hiperbólica, Rectified Linear Unit (ReLU) y Exponential Linear Unit (ELU). Imagen extraída de [53].

La capa principal en una CNN es la *capa de convolución*. Como hemos visto anteriormente, la superposición de un filtro sobre una imagen, definida como una convolución de dos funciones, tendrá como resultado una nueva imagen. Para controlar el proceso de aprendizaje, en la capa convolucional se utilizan una serie de hiperparámetros o filtros. El rendimiento de un modelo puede llegar a depender de la elección de estos hiperparámetros.

De entre estos hiperparámetros que requiere la capa convolucional, destacaríamos principalmente los cuatro siguientes [54]:

- *Kernel size* o tamaño del filtro (K). Es importante definir la anchura y altura del filtro o ventana deslizante de la convolución para poder determinar distintos patrones. El kernel permite detectar diferentes características de una imagen, así como enfocar y desenfocar dicha imagen.
- Número de filtros (F). Es el número total de ventanas deslizantes necesarias para que la red pueda aprender.
- *Stride* o salto (S). Hace referencia al número de cambios de píxeles sobre la matriz de entrada. Se define como la distancia recorrida por la ventana deslizante en cada iteración del entrenamiento.

- *Padding* (P). Después de cada convolución, la imagen sufre una reducción de su tamaño, lo cual impediría la construcción de redes más profundas. Para solucionar este problema, se lleva a cabo un proceso que consiste en la agregación de capas de ceros a las imágenes de entrada, es decir, la adición a dicha imagen de un determinado número de píxeles. La función de este parámetro es determinar si se rellena o no el borde de la matriz de la imagen de entrada cuando se realizan operaciones de convolución. No es imprescindible pero, en algunos casos, la aplicación de este hiperparámetro es conveniente.

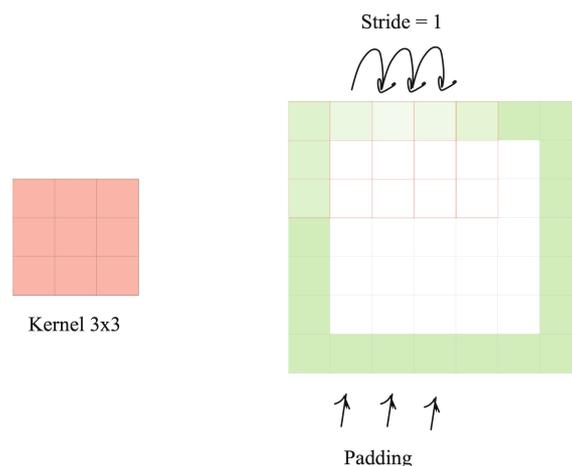


Figura 5.4: Hiperparámetros que intervienen en una red neuronal convolucional.

A modo de resumen, en la Figura 5.4 se pueden observar todos los hiperparámetros que intervienen en la convolución. A la izquierda se ha definido un *kernel 3x3* (altura = 3 y anchura = 3), el cual se aplica a la matriz (imagen). Se ha establecido que el *kernel* tenga *stride* unitario, es decir, un desplazamiento igual a una unidad en cada iteración. Finalmente, los píxeles en verde representan el padding aplicado a la matriz, en otras palabras, el número de píxeles que se añaden a la imagen.

5.2.3. Bloque de clasificación

En segundo lugar, el bloque de clasificación está formado por capas densamente conectadas entre sí (*fully connected*). Una capa densa (*dense layer*) es aquella que contiene una serie de neuronas completamente conectadas con las de la siguiente capa. Esta capa se emplea para la clasificación de las imágenes resultantes de las capas convolucionales. Un aumento en el número de neuronas provoca un aumento en el número final de parámetros a entrenar en la red neuronal.

5.3 Formulación matemática del modelo

5.3.1. Planteamiento

Una vez introducidos los conceptos básicos de las redes neuronales convolucionales y detallado el funcionamiento general de las redes GAN, es preciso conocer con más de-

tenimiento cómo trabaja internamente este último tipo de arquitecturas, profundizando en la base matemática que las sustenta.

Sea X nuestra variable aleatoria de interés, formada por un conjunto de imágenes reales $X = \{X_1, \dots, X_N\}$, las cuales son independientes entre sí y están idénticamente distribuidas y sea P_X la función de distribución de probabilidad de X .

Sea Z el vector latente formado por las variables aleatorias que alimentan el modelo. Este vector puede seguir dos distribuciones de probabilidad:

- Una distribución gaussiana: $Z \sim N(0, 1)$
- Una distribución uniforme: $Z \sim U(-1, 1)$

Como ya sabemos, en un modelo GAN intervienen dos tipos de redes neuronales, la red generadora y la red discriminadora.

Sea g_θ el generador del modelo y θ los parámetros de dicho generador, siendo éste una red neuronal representada por $g_\theta : Z \rightarrow X$. Es decir, el generador recibe como entrada una variable aleatoria gaussiana $Z \sim N(0, 1)$ que, después del entrenamiento de la red, da como resultado un nuevo conjunto de imágenes sintéticas, $g_\theta(Z)$. La distribución de probabilidad de $g_\theta(Z)$ se representa como P_θ ³.

El objetivo del modelo es generar nuevas imágenes que estén distribuidas de forma similar a las del conjunto de imágenes reales. Formalmente, este objetivo se concretaría en la elección de los parámetros θ adecuados para que la distribución de probabilidad de salida P_θ se asemeje a P_X , es decir, $P_\theta \sim P_X$.

La segunda red neuronal que interviene en el modelo es el discriminador, el cual está representado por $D_\omega : X \rightarrow [0, 1]$, donde ω son los parámetros del mencionado discriminador. Éste recibe, por un lado, las imágenes reales con distribución P_X y, por otro lado, las imágenes sintéticas provenientes del generador (con distribución P_θ). La finalidad del discriminador es asignar el valor 1 a las imágenes reales y el valor 0 a las sintéticas.

El objetivo principal del modelo sería pues conseguir que la distribución P_θ se encuentre lo más cercana posible a la distribución P_X . Gráficamente, este objetivo se podría ilustrar de la siguiente manera.

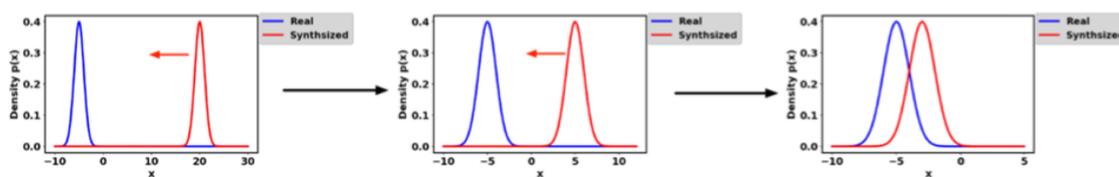


Figura 5.5: Proceso de entrenamiento de una red generativa adversaria. Imagen extraída de [55].

En la Figura 5.5 se puede observar como, a medida que se va entrenando la red, la distribución P_θ (*synthesized*) se va asemejando cada vez más a la distribución P_X (real).

³https://courses.engr.illinois.edu/ece598pv/fa2017/Lecture12_GAN_AmirT.pdf

5.3.2. Teorema minimax

Según lo expuesto anteriormente, de forma alternativa se podría argumentar también que el objetivo del discriminador sería el de clasificar correctamente las imágenes generadas sintéticas como falsas (valor 0) y las imágenes reales, como verdaderas (valor 1). Por el contrario, el objetivo del generador sería el de confundir al discriminador tanto como pueda, de modo que éste clasifique erróneamente las imágenes generadas como verdaderas.⁴

Ian Goodfellow, en su artículo, expuso el planteamiento de estos objetivos del discriminador y generador como la resolución de un juego *minimax*, donde el discriminador perseguiría maximizar su objetivo mientras que el generador buscaría conseguir lo contrario.

El *teorema minimax*, desarrollado por John von Neumann en 1928, forma parte de la Teoría de Juegos y establece un “método de decisión para minimizar la pérdida máxima esperada en juegos con adversario”⁵. Veamos a continuación como este nuevo concepto tiene su aplicación en el problema que nos ocupa.

Para entender dicho teorema, es necesario definir previamente un juego T de suma cero con dos jugadores I y II. Cada uno de estos jugadores dispone de un conjunto de estrategias puras a elegir a lo largo del juego. Asimismo, cada jugador puede asignar también una probabilidad de elección a cada una de sus estrategias puras, dando lugar a las denominadas estrategias mixtas. El conjunto de estrategias mixtas del jugador I se denota como $X = \{x_1, \dots, x_i, \dots, x_n\}$ y el del jugador II, como $Y = \{y_1, \dots, y_j, \dots, y_m\}$.

Para interpretar el resultado del juego, cuando uno o ambos jugadores emplean estrategias mixtas, se utiliza el concepto de *valor esperado*. Éste se puede definir como el resultado que obtendría el jugador I al elegir una estrategia mixta determinada x_i , suponiendo que el jugador II escogiese una estrategia mixta y_j . Todos estos valores esperados se pueden recoger en una *matriz de pagos*⁶ A o matriz de decisión.

El objetivo del jugador I será el de maximizar el valor esperado de su estrategia mixta x_i , es decir,

$$\max_X X^T AY \quad (5.2)$$

y el objetivo del jugador II será:

$$\min_Y X^T AY \quad (5.3)$$

Obviamente, dada la estrategia x_i elegida por el jugador I, lo que intentará hacer el jugador II será tomar una estrategia mixta y_j que minimice su pérdida esperada [56]. Así, el teorema minimax para juegos suma cero 2×2 , puede formularse como,

⁴<https://jaketae.github.io/study/gan-math/>

⁵<https://es.wikipedia.org/wiki/Minimax>

⁶Matriz de dimensión $n \times m$ que refleja el resultado posible de cada combinación de estrategias en un juego de suma cero entre dos jugadores, donde el jugador I dispone de n posibles movimientos y el jugador II, de m .

$$\max_X \min_Y X^T AY = \min_Y \max_X X^T AY = v \quad (5.4)$$

donde v es el resultado del juego.

Teniendo en cuenta todo lo explicado hasta el momento, la función objetivo de una red neuronal generativa adversaria (GAN) se definiría como la resolución del siguiente problema minimax,

$$\min_{\theta} \max_w \mathbb{E} [\log (D_w(X)) + \log (1 - D_w(g_{\theta}(Z)))] \quad (5.5)$$

donde,

- $D_w(X)$: evaluación que realiza el discriminador de los datos reales, es decir, la probabilidad de que una imagen X proceda de las imágenes reales y no de las imágenes creadas por el generador con probabilidad P_{θ} .
- $g_{\theta}(Z)$: conjunto de imágenes sintéticas procedentes del generador.
- $D_w(g_{\theta}(Z))$: evaluación que realiza el discriminador de los datos sintéticos, es decir, de los datos creados por el generador. En otras palabras, es la estimación que hace el discriminador de la probabilidad de que una imagen sintética sea como una real.
- \mathbb{E} : valor esperado de la ecuación.

5.3.3. Funciones de pérdida

Para la obtención de la ecuación 5.5, es necesario recurrir al concepto de *función de pérdida*. Hemos de tener en cuenta que el modelado de las redes neuronales requiere de dos atributos: la propia arquitectura del modelo y una función de pérdida que permita evaluar la calidad de dicho modelo comparando las diferencias entre las imágenes generadas y las reales.

Siguiendo el desarrollo de Goodfellow et al., la función de pérdida del modelo se puede derivar de la fórmula de pérdida de entropía cruzada binaria (*binary cross-entropy*), la cual se puede escribir como:

$$L(\hat{Y}, Y) = [Y \cdot \log \hat{Y} + (1 - Y) \cdot \log(1 - \hat{Y})] \quad (5.6)$$

donde Y representa las imágenes reales, e \hat{Y} las imágenes sintéticas.⁷

Para alcanzar la función de pérdida del modelo, será necesario previamente calcular las funciones de pérdida del discriminador y del generador.

⁷<https://towardsdatascience.com/the-math-behind-gans-generative-adversarial-networks-3828f3469d9c>

5.3.3.1. Función de pérdida del discriminador

En el proceso de entrenamiento del discriminador, se le asigna el valor $Y=1$ a los datos procedentes de P_X (imágenes reales) y un valor $\hat{Y} = D(X)$ a la evaluación de estas imágenes reales realizada por el discriminador. Sustituyendo estos valores en la función anterior, obtendríamos

$$L(D(X), 1) = \log(D(X)) \quad (5.7)$$

Por otra parte, a los datos procedentes del generador (datos sintéticos) se les asigna el valor $Y=0$ y un valor $\hat{Y} = D(G(Z))$ a la evaluación de las imágenes sintéticas realizada también por el discriminador. De igual manera, sustituyendo en 5.6, obtendríamos

$$L(D(G(Z)), 0) = \log(1 - D(G(Z))) \quad (5.8)$$

Llegados a este punto, el objetivo ahora del discriminador es clasificar el conjunto de imágenes reales y sintéticas. Para ello, las ecuaciones 5.7 y 5.8 deben maximizarse, quedando la función de pérdida final del discriminador definida como

$$L(D) = \text{máx}[\log(D(X)) + \log(1 - D(G(Z)))] \quad (5.9)$$

5.3.3.2. Función de pérdida del generador

Dado que el discriminador y el generador compiten entre sí, lo que hará este último es intentar minimizar la ecuación 5.9, resultando la función de pérdida del mismo como

$$L(G) = \text{mín}[\log(D(X)) + \log(1 - D(G(Z)))] \quad (5.10)$$

5.3.3.3. Función de pérdida del modelo

Las ecuaciones 5.9 y 5.10 se pueden fusionar para obtener,

$$L = \text{mín}_G \text{máx}_D [\log(D(X)) + \log(1 - D(G(Z)))] \quad (5.11)$$

que sería la función de pérdida general del modelo, donde L podría definirse como una función de G y D

$$V(G, D) = \text{mín}_G \text{máx}_D [\log(D(X)) + \log(1 - D(G(Z)))] \quad (5.12)$$

Se ha de tener en cuenta que la función de pérdida anterior sería válida únicamente en el caso de la existencia de un solo dato (imagen). Para contemplar el conjunto de datos completo, en la ecuación anterior debería considerarse el valor esperado de la misma. De esta manera, obtendríamos

$$\text{mín}_G \text{máx}_D V(D, G) = \mathbb{E}[\log D(X)] + \mathbb{E}[\log(1 - D(G(Z)))] \quad (5.13)$$

Aplicando los parámetros θ del generador y ω del discriminador, que optimizan ambas funciones de pérdida, la ecuación 5.13 se podría expresar como,

$$\min_{\theta} \max_{\omega} \mathbb{E} [\log (D_{\omega}(X)) + \log (1 - D_{\omega}(g_{\theta}(Z)))] \quad (5.14)$$

tal y como habíamos visto en 5.5, y utilizando la misma notación que allí.

Es importante tener en cuenta que el generador no puede influir en el término de la ecuación $\log (D_{\omega}(X))$, ya que éste no se ve afectado por el parámetro θ . En este caso, para el generador, minimizar su función de pérdida es equivalente a minimizar la expresión $\log (1 - D_{\omega}(g_{\theta}(Z)))$ ⁸.

Teniendo en cuenta esta última puntualización, el planteamiento del problema min-max se podría presentar alternativamente como,

$$\max_{\omega} \mathbb{E} [\log (D_{\omega}(X)) + \log (1 - D_{\omega}(g_{\theta}(Z)))] \quad (5.15)$$

y

$$\min_{\theta} \mathbb{E} [\log (1 - D_{\omega}(g_{\theta}(Z)))] \quad (5.16)$$

5.4 Proceso de entrenamiento del modelo

En el epígrafe anterior se ha estudiado cómo funciona una red generativa adversaria, analizando las expresiones matemáticas que intervienen en ella. Ahora bien, es muy importante que, tanto la arquitectura del generador como la del discriminador, estén bien diseñadas y, además, que el proceso de entrenamiento de ambas esté formulado correctamente ya que, sin una definición adecuada, las dos redes neuronales no tendrán el rendimiento esperado.

Algorithm 1 Minibatch stochastic gradient descent training of generative adversarial nets. The number of steps to apply to the discriminator, k , is a hyperparameter. We used $k = 1$, the least expensive option, in our experiments.

for number of training iterations **do**

for k steps **do**

- Sample minibatch of m noise samples $\{z^{(1)}, \dots, z^{(m)}\}$ from noise prior $p_g(z)$.
- Sample minibatch of m examples $\{x^{(1)}, \dots, x^{(m)}\}$ from data generating distribution $p_{\text{data}}(x)$.
- Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log D(x^{(i)}) + \log (1 - D(G(z^{(i)})))] .$$

end for

- Sample minibatch of m noise samples $\{z^{(1)}, \dots, z^{(m)}\}$ from noise prior $p_g(z)$.
- Update the generator by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log (1 - D(G(z^{(i)}))) .$$

end for

The gradient-based updates can use any standard gradient-based learning rule. We used momentum in our experiments.

Figura 5.6: Proceso de entrenamiento de una GAN. Fragmento del artículo de Goodfellow et al. Imagen extraída de [51].

⁸<https://developers.google.com/machine-learning/gan/loss>

En la Figura 5.6, se ilustra un fragmento del artículo de Goodfellow et al. donde se incluye el planteamiento formal de dicho proceso de entrenamiento formulado por el científico norteamericano. En él, se puede observar como el entrenamiento de las dos redes neuronales de la GAN se consigue gracias a la aplicación del denominado algoritmo de *descenso de gradiente*⁹ (en inglés, *stochastic gradient descent*).

En dicho artículo se aprecia cómo el generador y el discriminador se entrenan simultáneamente, pero por separado¹⁰. En él, se definen también las funciones de pérdida asociadas a cada red neuronal durante el proceso de entrenamiento, las cuales están actualizadas por un vector de gradiente estocástico. El método de descenso de gradiente se aplica tanto al discriminador como al generador, para maximizar o minimizar la función de pérdida en cada caso.

Manteniendo la notación utilizada por Goodfellow et al., ambas funciones de pérdida se podrían expresar de la siguiente manera:

- Discriminador:

$$\nabla_{\omega} \frac{1}{m} \sum_{i=1}^m \left[\log D(x^{(i)}) + \log \left(1 - D(G(z^{(i)})) \right) \right] \quad (5.17)$$

- Generador:

$$\nabla_{\theta} \frac{1}{m} \sum_{i=1}^m \log \left(1 - D(G(z^{(i)})) \right) \quad (5.18)$$

donde i representa la muestra actual que se está evaluando y (ω, θ) los parámetros del discriminador y generador, respectivamente¹¹.

Formalmente, el proceso de entrenamiento consistiría en la aplicación de un método iterativo de descenso de gradiente estocástico simultáneo en el problema de maximización – minimización. Los parámetros ω del discriminador y θ del generador, convergerán hacia sus respectivos gradientes. En cada proceso iterativo se emplea un *batch* de imágenes reales procedentes de la función de distribución objetivo $X = \{X_1, \dots, X_N\} \sim P_X$ y otro de muestras sintéticas del generador $\{g_{\theta}(Z_1), \dots, g_{\theta}(Z_N)\}$ y los parámetros ω y θ se van actualizando según una tasa η de aprendizaje.

Al ponerse en marcha el proceso de entrenamiento del discriminador, éste va clasificando las imágenes procedentes de la red generadora. De esta manera, el discriminador se autopenalizará en cada ocasión que clasifique erróneamente una imagen real como falsa o viceversa, es decir, en el caso que catalogue una imagen generada sintéticamente por el generador (falsa), como real. Esto permitirá la maximización de la función de pérdida del discriminador 5.17. El proceso de autopenalización permite que el discriminador mejore su rendimiento a la hora de evaluar si una imagen es verdadera o falsa.

⁹Método de optimización que tiene como objetivo alcanzar el valor mínimo de una función en un proceso iterativo. [57]

¹⁰<https://towardsdatascience.com/the-math-behind-gans-generative-adversarial-networks-3828f3469d9c>

¹¹Nótese que aquí se ha mantenido nuestra notación en los parámetros del discriminador y del generador, ω y θ , respectivamente. La notación utilizada por Goodfellow para los parámetros de ambas redes sería θ_d y θ_g , respectivamente.

Por otra parte, el generador se encarga de obtener una imagen de salida, dada una muestra de ruido aleatorio. Dicha imagen llega al discriminador y, tal y como se ha comentado anteriormente, se clasifica como verdadera o falsa. El discriminador calcula el error del generador en función de la clasificación de las imágenes que recibe de este último. De este modo, se recompensa al generador si consigue “engañar” al discriminador, minimizando así su función de pérdida 5.18. Obviamente, si el generador fracasa en su empeño, éste se ve penalizado.

Finalmente, debido a que se está trabajando con dos redes neuronales, ambas deben ser entrenadas simultáneamente, ya que los resultados obtenidos en una red influyen en el comportamiento final de la otra.

5.5 Problemas de convergencia

El objetivo de una red generativa adversaria es, como ya sabemos, encontrar el equilibrio. Formalmente consistiría en hallar el punto de silla (*saddle point*)¹² de un problema de minimización – maximización empleando el algoritmo de descenso de gradiente.

Aunque en la práctica este tipo de arquitecturas parecen funcionar bien, a priori se desconoce si la optimización de las mismas convergerá realmente en dicho equilibrio. Incluso si una GAN está muy cerca de este equilibrio, ésta puede simplemente girar entorno a un punto de silla sin llegar a converger en ella. O peor aún, pueden existir regiones conflictivas, donde el sistema puede atascarse, que provoquen que la GAN acabe desviándose de ese equilibrio.

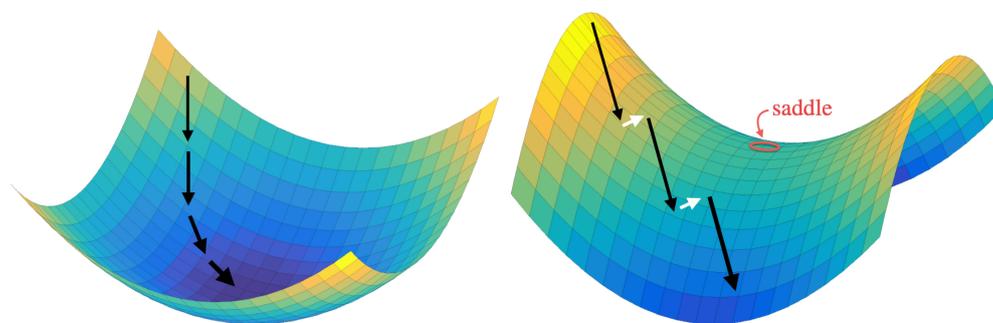


Figura 5.7: Ruta de aprendizaje de una red neuronal estándar (a la izquierda) y ruta de aprendizaje de una red generativa adversaria (a la derecha). Imagen extraída de [58].

En la Figura 5.7 se destacan las diferencias entre las rutas o procesos de aprendizaje de un red neuronal básica y de una GAN. En la representación de la derecha se pueden observar dichas regiones conflictivas en las cuales el sistema se acaba desviando.

Este sería, de forma general, el inconveniente principal que suele aparecer durante el entrenamiento de una arquitectura de estas características. Debido a ello, la dinámica de la optimización GAN debe ser estudiada para observar si este equilibrio es localmente estable.

¹²Puntos críticos donde una función tiene un máximo local en una dirección y un mínimo, también local, en otra.

Relacionado con lo anterior, pueden definirse algunos problemas más específicos de convergencia. De entre ellos se podrían destacar el *colapso de modo* y el *desvanecimiento de gradiente*.

5.5.1. Colapso de modo

El *mode collapse*, o colapso de modo en español, es uno de los grandes problemas que pueden surgir en el proceso de convergencia de una GAN. Se puede definir como el colapso parcial del generador, derivado de su incapacidad para generar muestras para todos los valores existentes en el dataset de entrenamiento dado. Es decir, el generador sólo sería capaz de generar una variedad limitada de salidas a partir de todas las muestras disponibles en la base de entrenamiento.

El generador no es lo suficientemente potente y sólo es capaz de generar un número limitado de muestras, apareciendo varios duplicados de las mismas. Esto provoca que, en este punto, la mejor estrategia que puede tomar el discriminador sea la de aprender a rechazar siempre ese número limitado de salidas.

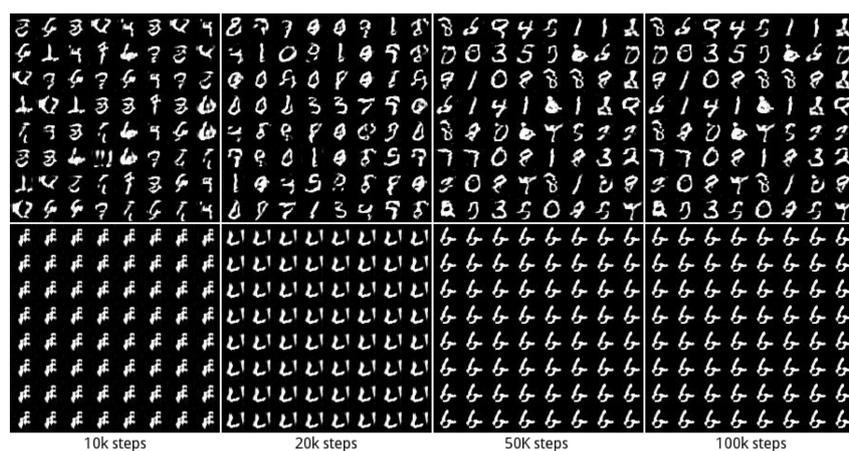


Figura 5.8: Conjunto de imágenes creadas por una GAN. Imagen extraída de [59].

En la Figura 5.8 se representa un ejemplo del problema de *mode collapse* en las GAN. La fila superior de la imagen muestra un entrenamiento sin colapso de modo, donde el generador es capaz de producir todas las muestras (dígitos) a partir de la base de entrenamiento MNIST. La fila inferior muestra un mal entrenamiento con colapso de modo, en el cual el generador sólo es capaz de crear algo parecido al dígito “6”.

El colapso de modo puede llegar a ser un problema difícil de diagnosticar y resolver. Algunas de las soluciones usualmente propuestas pueden incluir la modificación de la propia arquitectura GAN, el aumento de tamaño del conjunto de datos de entrenamiento o el empleo de un algoritmo de entrenamiento diferente.

5.5.2. Desvanecimiento del gradiente

Puede darse el caso de que una de las dos redes del modelo tenga un rendimiento excepcional y anule el aprendizaje de la otra. Así, por ejemplo, si el aprendizaje del discriminador es demasiado satisfactorio, éste conseguirá identificar fácilmente los da-

tos sintéticos de los reales. Esto puede provocar el desvanecimiento del gradiente de la función de pérdida del generador (*vanishing gradient*), provocando que el proceso de aprendizaje del mismo converja lentamente o incluso se detenga.

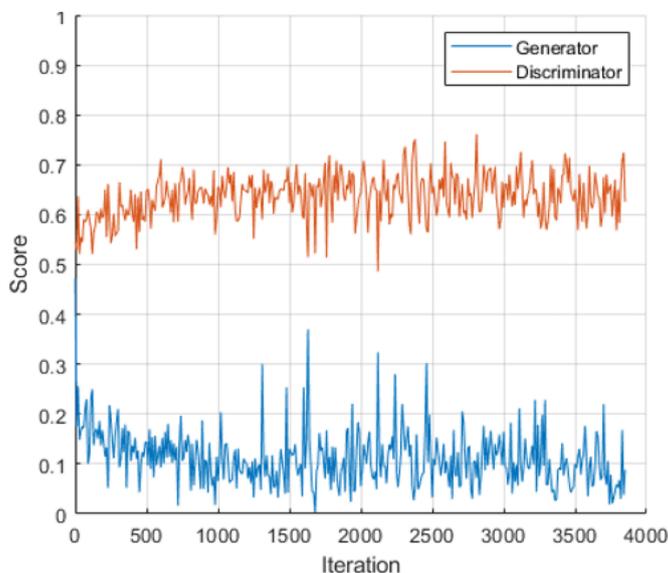


Figura 5.9: Evolución del rendimiento del discriminador y del generador. Imagen extraída de MathWorks ©.

En la Figura 5.9 se representa, mediante una gráfica, la evolución del rendimiento (*score*) del discriminador y del generador¹³. Se ve claramente como el *score* del discriminador (en naranja), es excelente. Esto se traduce en un bajo rendimiento del generador (en azul), imposibilitando de esta manera la producción de una gran variedad de muestras sintéticas.

Según Goodfellow et al., el problema generado por el *vanishing gradient* puede provocar que, dada la función de pérdida del modelo,

$$\min_G \max_D V(D, G) = \mathbb{E}[\log D(X)] + \mathbb{E}[\log(1 - D(G(Z)))] \quad (5.19)$$

el generador G no pueda tener un proceso de aprendizaje adecuado. Esto es debido a que, en las primeras iteraciones de dicho proceso, las muestras generadas por G son de baja calidad, asemejándose al ruido. En este caso, el discriminador D no tendría demasiados problemas para rechazar estas muestras, ya que serían marcadamente diferentes a los datos originales. En esta situación, el término de la ecuación anterior

$$\log(1 - D(G(Z))) \quad (5.20)$$

llegaría a saturarse¹⁴.

¹³<https://www.mathworks.com/help/deeplearning/ug/monitor-gan-training-progress-and-identify-common-failure-modes.html>

¹⁴Jesús Javier Moralo García. *Generative Adversarial Networks (GAN), una introducción* (LinkedIn).

5.6 Variantes de GAN

5.6.1. Deep Convolutional GAN (DCGAN)

Como ya se ha podido observar, uno de los principales inconvenientes de las GAN originales es la dificultad que tienen éstas para converger en un punto de equilibrio. Con la finalidad de mejorar su aprendizaje durante el proceso de entrenamiento, se han llegado a plantear diferentes propuestas. Uno de los enfoques que más ha contribuido en este aspecto ha sido el desarrollo de las denominadas *Deep Convolutional GAN (DCGAN)*, o redes GAN convolucionales profundas.

La principal idea de las DCGAN, en comparación con los modelos originales de redes generativas adversarias, es la de ir agregando sucesivas capas convolucionales de muestreo en el generador, entre el vector de entrada y la imagen de salida. De la misma manera, en el discriminador, y como en una CNN básica, también se utilizan diversas capas convolucionales para poder clasificar las imágenes sintéticas y las reales.

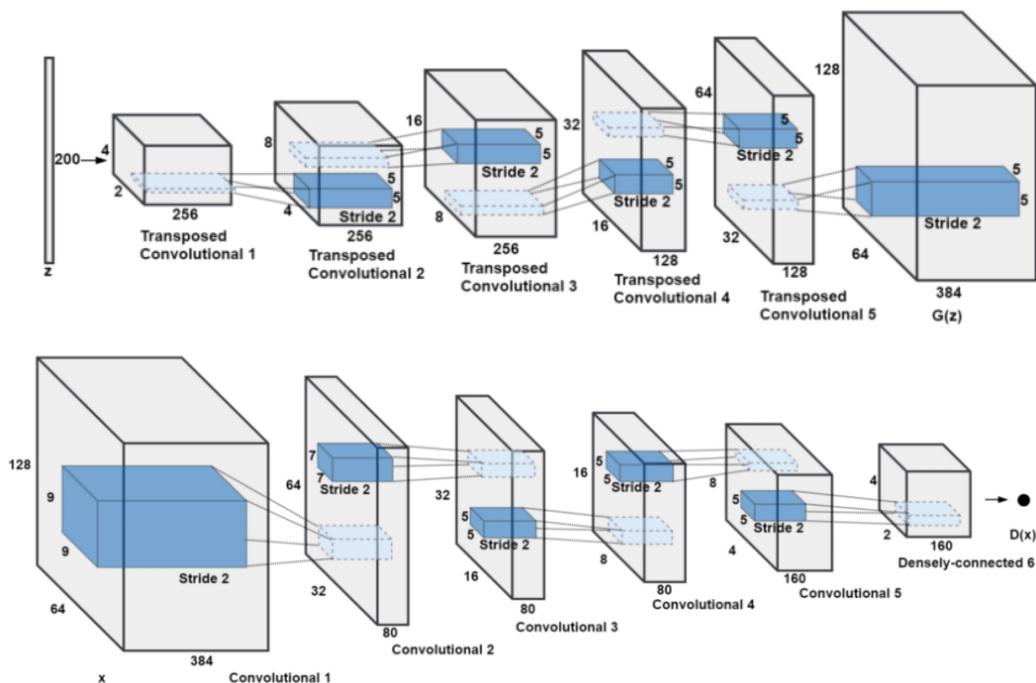


Figura 5.10: Arquitectura de una red DCGAN. En la parte superior se encuentra el generador y, en la inferior, el discriminador. Imagen extraída de Medium.com.

En la Figura 5.10¹⁵, se pueden observar ejemplos de las arquitecturas del generador y discriminador dentro de una DCGAN.

Radford et al., en un artículo publicado en 2015 sobre aprendizaje no supervisado con redes DCGAN [60], propusieron un conjunto de directrices para lograr estabilizar el entrenamiento de esta nueva arquitectura. Es necesario advertir que no existe una fundamentación matemática que avale el empleo de estas pautas. Los resultados obtenidos tras

¹⁵Himanshu Dutta. *DCGAN Under 100 Lines of Code*. <https://medium.com/swlh/dcgan-under-100-lines-of-code-fc7fe22c391>

la adopción de las mismas, que han evidenciado una mayor estabilidad en los modelos, son fruto únicamente de la evidencia empírica tras una exhaustiva experimentación.

Las directrices propuestas por Radford serían las siguientes:

- Sustitución de las capas de submuestreo (*pooling layers*) por capas de convoluciones stride (*strided convolutions*) en el discriminador y capas deconvolucionales (*fractional-strided convolutions*) en el generador.

Las capas de submuestreo o *pooling layers*, son capas que se utilizan normalmente para reducir, de una capa a la siguiente, la resolución de una imagen. En ellas no intervienen parámetros para ser entrenados, por lo que el número total de éstos no se verá incrementado.

Las convoluciones tipo *stride* (*strided convolutions*) realizan la misma función. Así, por ejemplo, en el caso del discriminador si se establece que el *stride* de una convolución sea igual a 2, se estará reduciendo la resolución de la imagen a la mitad. Pero en este caso, y a diferencia de las *pooling layers*, las convoluciones *strided* sí que determinan el tamaño de la salida y el número de operaciones, por lo que el volumen total de parámetros sí que aumentará.

En el generador se utilizan las denominadas *fractional-strided convolutions*, a veces conocidas como capas deconvolucionales. Éstas realizan el proceso inverso al descrito anteriormente, es decir, permiten aumentar la dimensión final de la imagen.

- Empleo de la normalización de lotes (*batchnorm*) tanto en el generador como en el discriminador.

La normalización por lotes (en inglés, *batch normalization* o simplemente *batchnorm*) es un método diseñado para conseguir acelerar y estabilizar las redes neuronales. Se consigue mediante un proceso de normalización de los valores que ingresan en cada capa, evitando así que dichos valores aparezcan sesgados.

Formalmente se puede exponer como la transformación a la que se someten los datos para que éstos tengan una media igual a cero y una desviación estándar unitaria. Este proceso, como ya hemos dicho, permite acelerar el entrenamiento de las GAN.

- Supresión de las capas ocultas densamente conectadas de las arquitecturas más profundas.

En el diseño de una red neuronal convolucional habitualmente suelen existir demasiados parámetros asociados implícitamente a las capas ocultas densamente conectadas. Esto provoca que dichas capas resulten menos eficientes en comparación con las capas convolucionales. Para resolver este problema, en las CNN clásicas se suele utilizar un método conocido como *Global Average Pooling*, el cual se basa en una operación de agrupación diseñada para reemplazar capas densamente conectadas. Pero se ha demostrado que este procedimiento perjudica la velocidad de convergencia.

Radford et al. proponen una solución intermedia, consistente en conectar directamente las capas convolucionales más densamente conectadas en la entrada y salida del generador y discriminador, respectivamente.

- Empleo de la función de activación ReLU en el generador y de la LeakyReLU en el discriminador.

La función de activación es un elemento esencial en el diseño de una red neuronal, ya que otorga un control total sobre el proceso de entrenamiento del modelo. Añadiendo

este tipo de funciones en las capas ocultas se consigue un aprendizaje más eficiente del modelo.

La función de activación ReLU (por las siglas en inglés de *Rectified Linear Unit*), ayuda a las redes neuronales a crear modelos de aprendizaje profundo. Básicamente, lo que hace es transformar los valores de entrada, anulando los valores negativos y dejando los positivos tal como entran.

Esta función posibilita la superación de los problemas generados por el *vanishing gradient*, permitiendo que el proceso de aprendizaje del modelo sea mejor y más rápido. Se emplea en todas las capas del generador, excepto en la capa de salida, donde se utiliza la función *tanh*¹⁶ como función de activación.

En el discriminador se emplea la función de activación *Leaky ReLU* (*Rectified Lineal Unit*), en todas las capas. Esta función transforma los valores de entrada, multiplicando los negativos por un coeficiente rectificativo y dejando los positivos tal como entran.

5.6.2. Wasserstein GAN (WGAN)

Las redes Wasserstein GAN fueron propuestas por Arjovsky et al. en un artículo publicado en el 2017 y titulado “*Wasserstein GAN*” [61]. Este tipo de redes reciben su nombre de la denominada distancia de Wasserstein, más conocida en inglés como *Earth Mover’s Distance* (EMD). Esta EMD, sobre la cual volveremos a incidir más adelante, es una medida de la distancia existente entre dos distribuciones de probabilidad y “se utiliza ampliamente en la comparación de imágenes”¹⁷.

En el artículo mencionado, las WGAN se presentan como un desarrollo de las GAN originales para tratar de encontrar una forma alternativa de entrenamiento de la red neuronal del generador. El objetivo es conseguir que la distribución de datos generados logre una mejor aproximación a la distribución de datos originales (*dataset* de entrenamiento), evitando así los problemas de convergencia o los derivados de un colapso de modo. Las WGAN son un tipo de redes que minimizan la aproximación de la distancia EMD, como explicaremos a continuación.

Como ya sabemos, en la formulación GAN original, la misión del discriminador es clasificar una imagen de entrada dada como real o falsa. En otras palabras, el discriminador entrega como salida un resultado binario (0 ó 1), es decir, nos dice si la muestra es falsa o verdadera. El inconveniente de este resultado radica en que está expresado como una puntuación y no como una medida de probabilidad, la cual permitiría conocer el grado de verosimilitud de las imágenes de entrada. Lo que propone Arjovsky es emplear un *crítico* en lugar del discriminador, para predecir la probabilidad de que las imágenes generadas sean reales o falsas.

A partir de las distribuciones de probabilidad de las imágenes reales (P_r) y de las imágenes sintéticas (P_g), la finalidad de la WGAN es conseguir que ambas distribuciones de probabilidad se acerquen entre sí lo máximo posible. Esto, obviamente, se traducirá en un mayor realismo de la salida generada. Este propósito puede ilustrarse mediante una gráfica¹⁸.

¹⁶*Tangent Hyperbolic* (Función tangente hiperbólica).

¹⁷<http://www.recsi2020.udl.cat/static/site/files/BlancoDomingo-XVI-RECSI.pdf>

¹⁸Se puede apreciar que la notación utilizada en la Figura 5.11, para las distribuciones de probabilidad de las imágenes reales y de las sintéticas, es P_{data} y P_G respectivamente.

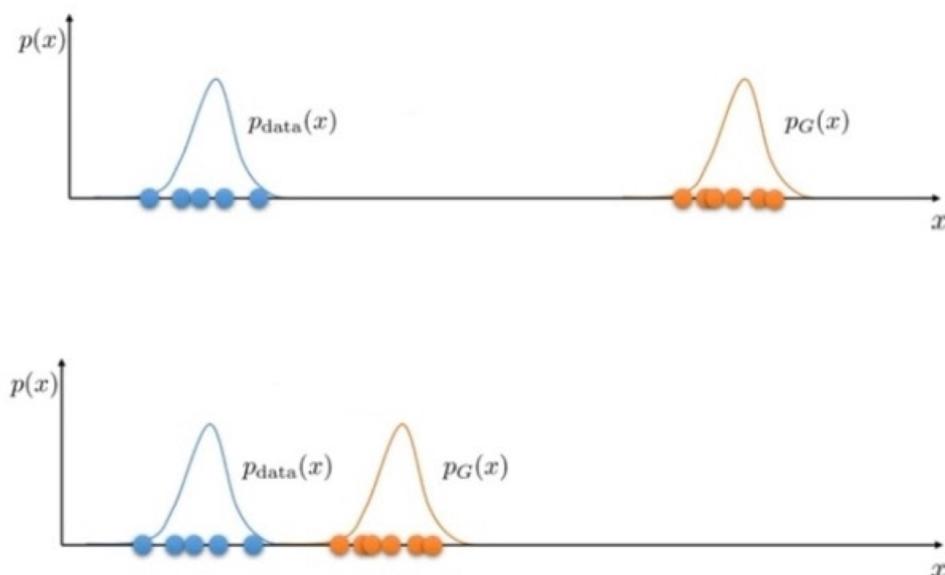


Figura 5.11: Aproximación mediante WGAN de una distribución de datos generados (en naranja) a una distribución de datos reales (en azul). Imagen extraída de Medium.com.

En la parte superior de la Figura 5.11 se puede observar como, en principio, ambas distribuciones P_{data} y P_G están muy alejadas¹⁹. El objetivo de las WGAN es que la distribución P_G se acerque lo máximo posible a la distribución P_{data} , tal y como se puede ver en la parte inferior de la misma figura. Para ello, será preciso medir previamente la distancia existente entre ambas distribuciones.

Arjovsky sugiere cuatro diferentes propuestas como medidas de cálculo de dicha distancia entre distribuciones de probabilidad:

- Distancia de variación total o *Total Variation distance (TV)* en inglés.
- Divergencia de Kullback-Leibler o *Kullback-Leibler divergence (KL)*.
- Divergencia de Jensen-Shannon o *Jensen-Shannon divergence (JS)*.
- Distancia de Wasserstein o *Earth-Mover Distance (EMD)*.

La arquitectura WGAN está basada en la última de estas propuestas, la distancia de Wasserstein (EMD). Arjovsky utiliza esta métrica porque considera que es la medida más adecuada para calcular la distancia entre dos distribuciones de probabilidad. El propósito de la WGAN es calcular primero dicha distancia y luego, a través del entrenamiento, conseguir que ésta llegue a hacerse lo más pequeña posible para que ambas distribuciones se aproximen. Esto permitirá que los datos sintéticos generados logren asemejarse lo más posible a los datos originales.

¹⁹Samuele Bolotta. *Generative modeling*. <https://medium.com/mlearning-ai/generative-modeling-7a88e847f62e>

Intuitivamente la EMD puede entenderse como el trabajo mínimo necesario para transformar una distribución en otra, donde el trabajo se define como el producto entre la “masa” de la distribución de probabilidad que debe desplazarse y la distancia final recorrida en dicho desplazamiento.

Para explicar este concepto es conveniente confiar en la ayuda proporcionada por el siguiente ejemplo ²⁰. Supongamos una situación definida por un problema de transporte de cierto producto entre ciudades, con los datos siguientes:

- Sea C un conjunto de ciudades $C = \{C_1, \dots, C_N\}$.
- Sea $d(c_x, c_y)$ la distancia entre dos ciudades cualesquiera, en este caso, c_x y c_y .
- Sea $p(c)$ y $q(c)$ la oferta y demanda de cierto producto en una ciudad c .
- Sea $\gamma(c_x, c_y)$ una función no negativa que representa la cantidad de producto que se debe transportar desde la ciudad c_x hasta la ciudad c_y .

Hay que tener en cuenta además, las siguientes premisas:

- La oferta de una ciudad $p(c)$ se define como la cantidad total de producto que sale de dicha ciudad.
- La demanda de una ciudad $q(c)$ se determina como la cantidad total de producto que finalmente entra en ella.

En estadística, las funciones de oferta y demanda (p) y (q) se pueden concretar como las distribuciones de probabilidad de las mismas sobre el conjunto de ciudades C .

Con toda esta información, podemos definir el coste C_d de transportar m unidades de ese bien [$m = \gamma(c_x, c_y)$] desde la ciudad c_x hasta la c_y , como el producto $m * d(c_x, c_y)$, es decir, la “masa” del producto transportado por la distancia entre ambas ciudades. Si formulamos esta última expresión según la notación utilizada por Arjovsky, obtendríamos:

$$C_d(\gamma) = \sum_{c_x, c_y \in C} \gamma(c_x, c_y) d(c_x, c_y) \quad (5.21)$$

La ecuación 5.21 podría escribirse también como,

$$C_d(\gamma) = \mathbb{E}_{c_x, c_y \sim \gamma} [d(c_x, c_y)] \quad (5.22)$$

expresándose en función del coste promedio esperado del transporte para todos los pares de ciudades (c_x, c_y) .

Si representamos el conjunto de todos los planes de transporte válidos por $\Pi(p, q)$, la distancia de Wasserstein puede definirse matemáticamente, para dos distribuciones p y q , como el mayor límite inferior (*ínfimo*) de cualquier plan de transporte, es decir, el coste del plan más barato. Esto puede formularse como sigue:

²⁰Ejemplo extraído de <https://www.vanillabug.com/posts/wasserstein/>. Puede encontrarse otro ejemplo alternativo en el Apéndice B.

$$W_d(p, q) = \inf_{\gamma \in \Pi(p, q)} \mathbb{E}_{c_x, c_y \sim \gamma} [d(c_x, c_y)] \quad (5.23)$$

En la ecuación 5.23, el multiplicatorio $\Pi(p, q)$ representaría el total de las distribuciones conjuntas $\gamma(c_x, c_y)$ cuyas distribuciones marginales serían iguales a p y q respectivamente. Intuitivamente, $\gamma(c_x, c_y)$ indicaría la cantidad de “masa” que debe moverse desde c_x hasta c_y , para transformar la distribución q en la distribución p . Así pues, la distancia de Wasserstein nos estaría informando de cuál es el coste de transporte óptimo que permite convertir q en p .

Hemos de tener en cuenta que la distancia es un vector del cual nos interesa conocer su longitud. Para ello, será necesario definir un operador *norma* que determine la longitud del mismo. Así, la ecuación 5.23 podría escribirse también como,

$$W(p, q) = \inf_{\gamma \in \Pi(p, q)} \mathbb{E}_{(x, y) \sim \gamma} [\|x - y\|] \quad (5.24)$$

Nótese que en la anterior ecuación se ha modificado la notación original de las distribuciones utilizada por Arjovsky. Las distribuciones p y q corresponderían a la distribución de probabilidad de imágenes reales P_r y a la distribución de imágenes sintéticas P_g , respectivamente. Así, conservando la notación original de Arjovsky, la ecuación 5.24 se podría reescribir como,

$$W(P_r, P_g) = \inf_{\gamma \in \Pi(P_r, P_g)} \mathbb{E}_{(x, y) \sim \gamma} [\|x - y\|] \quad (5.25)$$

Sin embargo, la distancia de Wasserstein definida en la ecuación 5.25 puede llegar a ser una cuestión difícilmente abordable, ya que no se puede calcular el ínfimo sobre todos $\gamma \in \Pi(P_r, P_g)$ debido a que $\gamma(x, y)$ puede tratarse de una distribución muy compleja.

Arjovsky et al. proponen una solución más manejable para el entrenamiento posterior de la red. Así, utilizando la dualidad de *Kantorovich-Rubinstein*, se puede simplificar el cálculo y formular alternativamente la ecuación 5.25 como,

$$W(P_r, P_g) = \sup_{\|f\|_L \leq 1} \mathbb{E}_{x \sim P_r} [f(x)] - \mathbb{E}_{x \sim P_g} [f(x)] \quad (5.26)$$

donde *sup* es el límite superior mínimo (*supremo*) y f es una función de *1-Lipschitz*. De esta manera, para calcular la distancia de Wasserstein, sólo necesitamos encontrar una función de *1-Lipschitz*.

La dualidad de Kantorovich-Rubinstein proporciona una forma mucho más cómoda para resolver la distancia de Wasserstein. La nueva formulación permite expresar la diferencia de probabilidades entre P_r y P_g de forma parecida a cómo lo hace una GAN básica.

La función 5.26 es conocida también como función de pérdida Wasserstein (*W-Loss*) y resuelve algunos grandes problemas a los que se enfrentan las GAN simples, como el *mode collapse* y el *vanishing gradient*. Sin embargo, ya hemos visto que el tratamiento de esta función puede llegar a ser un tema espinoso debido a la dificultad para encontrar un valor óptimo del parámetro γ .

El lector interesado en conocer el desarrollo matemático íntegro que permite llegar a esta función de pérdida W -Loss, definida en la ecuación 5.26, lo puede encontrar en el Apéndice C.

El generador intentará minimizar la función W -Loss tratando que las muestras sintéticas por él creadas se aproximen lo máximo posible a las muestras reales. Pero, por otra parte, el *crítico* quiere maximizar la misma función para poder diferenciar de la mejor manera posible las imágenes reales de las sintéticas. Es decir, lo que persigue el *crítico* es que la distancia entre ambas probabilidades sea la mayor posible.

Asimismo, para el correcto funcionamiento de la red, el *crítico* debe cumplir una condición especial: que la función sea 1 -Lipschitz (1 -L) continua, es decir, que la norma euclídea del gradiente de la función sea menor o igual a 1.

$$\|\nabla f(x)\|_2 \leq 1 \quad (5.27)$$

Una función es 1 -Lipschitz continua si su gradiente (pendiente) es menor o igual a 1. Esta condición es de suma importancia para que la función de pérdida W -Loss sea continua y diferenciable y, de esta forma, mantenga cierta estabilidad durante el entrenamiento.

La dificultad en las WGAN es hacer cumplir esta restricción de Lipschitz. Durante el entrenamiento, para satisfacer la condición 1 -L, se utilizaría la técnica del *weight-clipping*. Mediante este método, los pesos del crítico se mantienen acotados dentro de un intervalo fijo, es decir, "recortándose" (*clipping*) todos aquellos valores que se encuentren fuera del intervalo.

O sea, se establecerá un valor máximo y otro mínimo para el rango de pesos que pueden asignarse al *crítico*, recortándose a dichos valores todos aquellos que superen los mismos. Este concepto puede explicarse mejor con un ejemplo.

Supongamos que los pesos del *crítico*, en una iteración concreta de descenso de gradiente, son los siguientes:

$$[0, 1, 12, 67, -0.5, 1.7, -7.8] \quad (5.28)$$

y se desea aplicar el método del *weight-clipping* para que estos pesos se encuentren dentro del rango definido por el intervalo entre $[-1, 1]$. El vector resultante quedará como:

$$[0, 1, 1, 1, -0.5, 1, -1] \quad (5.29)$$

Se puede observar que todos aquellos valores que sobrepasan los límites establecidos han sido recortados.

Esta técnica es muy efectiva para el correcto cumplimiento de la condición 1 -Lipschitz y, además, permite que el entrenamiento de la red sea más estable. Sin embargo, esto repercute en la capacidad de aprendizaje del *crítico* ya que, el hecho de que los pesos estén acotados dentro de un rango limitado de valores, provoca que éste solamente pueda aprender funciones muy simples, limitando su capacidad para modelar funciones más complejas. Esta circunstancia se puede observar en la Figura 5.12.

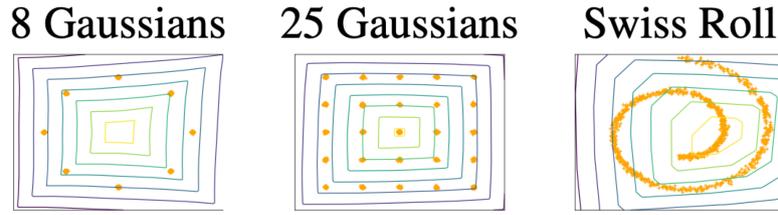


Figura 5.12: Proceso de entrenamiento del crítico, empleando el método weight-clipping. Imagen extraída de [62].

5.6.2.1. Entrenamiento de la red

El proceso de entrenamiento de una red WGAN seguiría el enfoque presentado por Arjovsky et al. en su artículo de 2017.

Algorithm 1 WGAN, our proposed algorithm. All experiments in the paper used the default values $\alpha = 0.00005$, $c = 0.01$, $m = 64$, $n_{\text{critic}} = 5$.

Require: : α , the learning rate. c , the clipping parameter. m , the batch size. n_{critic} , the number of iterations of the critic per generator iteration.

Require: : w_0 , initial critic parameters. θ_0 , initial generator's parameters.

```

1: while  $\theta$  has not converged do
2:   for  $t = 0, \dots, n_{\text{critic}}$  do
3:     Sample  $\{x^{(i)}\}_{i=1}^m \sim \mathbb{P}_r$  a batch from the real data.
4:     Sample  $\{z^{(i)}\}_{i=1}^m \sim p(z)$  a batch of prior samples.
5:      $g_w \leftarrow \nabla_w [\frac{1}{m} \sum_{i=1}^m f_w(x^{(i)}) - \frac{1}{m} \sum_{i=1}^m f_w(g_\theta(z^{(i)}))]$ 
6:      $w \leftarrow w + \alpha \cdot \text{RMSProp}(w, g_w)$ 
7:      $w \leftarrow \text{clip}(w, -c, c)$ 
8:   end for
9:   Sample  $\{z^{(i)}\}_{i=1}^m \sim p(z)$  a batch of prior samples.
10:   $g_\theta \leftarrow -\nabla_\theta \frac{1}{m} \sum_{i=1}^m f_w(g_\theta(z^{(i)}))$ 
11:   $\theta \leftarrow \theta - \alpha \cdot \text{RMSProp}(\theta, g_\theta)$ 
12: end while

```

Figura 5.13: Proceso de entrenamiento de una WGAN. Fragmento del artículo de Arjovsky et al. Imagen extraída de [61].

En la Figura 5.13 se representa un fragmento del artículo de Arjovsky donde puede apreciarse el detalle del proceso de entrenamiento de una WGAN. Inicialmente, mediante un lote de imágenes reales y otro de imágenes creadas por la red generadora, se pone en marcha el proceso de entrenamiento de la red crítica definiendo unos parámetros iniciales. A continuación, estos mismos lotes se emplean para calcular la pérdida de dicha red (la distancia de Wasserstein) y, posteriormente, se aplicaría el método *weight-clipping* para cumplir la condición *1-Lipschitz*.

Seguidamente, gracias a las muestras sintéticas creadas por el generador, el crítico informa a la red generadora sobre la distancia entre la distribución de probabilidad de las imágenes reales y la distribución de las imágenes que ésta última ha creado. Para finalizar, mediante *backpropagation*, esta distancia o error se propaga capa a capa hasta el

generador, el cual va a encargarse de actualizar sus pesos de aprendizaje para minimizar la distancia de Wasserstein.

5.6.3. Wasserstein GAN Gradient Penalty (WGAN - GP)

Como acabamos de ver, uno de los principales inconvenientes de la arquitectura WGAN original, que utiliza el método de *weight-clipping* o reducción de peso, es que no siempre funciona de forma óptima, debido a que el *crítico* solamente es capaz de aprender funciones muy simples.

Afortunadamente, existe una segunda técnica para asegurar que la función de pérdida WGAN sea *1-Lipschitz* continua y, a la vez, solucionar la problemática anterior. Se trata del método *Gradient Penalty* (GP) o penalización de gradiente, propuesto por Gulrajani et al. en un artículo titulado *Improved Training of Wasserstein GANs* [62] y publicado a finales de 2017.

La WGAN-GP, para hacer cumplir la continuidad de Lipschitz, reemplaza el recorte de pesos visto anteriormente por la introducción de una restricción en la norma del gradiente del crítico. “Esto permite un entrenamiento más estable de la red que WGAN y requiere muy pocos ajustes de hiperparámetros”²¹.

Con *Gradient Penalty* se añade un término de regularización a la función de pérdida. WGAN-GP penaliza al *crítico* cuando su norma de gradiente es mayor que 1. Pero el proceso de comprobación del *crítico* en todos los puntos posibles del espacio de características es poco práctico. Gulrajani propone obtener una muestra “intermedia” mediante una interpolación \hat{x} entre imágenes reales y generadas, utilizando un número aleatorio ϵ .

En el artículo mencionado, el *Gradient Penalty* se podría formular mediante la siguiente expresión:

$$(\|\nabla_{\hat{x}}D_w(\hat{x})\|_2 - 1)^2 \quad (5.30)$$

La interpolación \hat{x} se definiría como,

$$\hat{x} = \epsilon x + (1 - \epsilon)g(z) \quad (5.31)$$

siendo x la imagen real y $g(z)$ la imagen generada.

Es en la muestra interpolada \hat{x} en donde se quiere conseguir que el gradiente del *crítico* sea menor o igual a 1. En la siguiente ecuación, el parámetro λ es el coeficiente de regularización que se utiliza para ponderar el término de penalización del gradiente:

$$D_w(\tilde{x}) - D_w(x) + \lambda (\|\nabla_{\hat{x}}D_w(\hat{x})\|_2 - 1)^2 \quad (5.32)$$

donde,

- $D_w(\tilde{x}) - D_w(x)$ permite que la WGAN – GP sea menos susceptible de sufrir problemas como el *mode collapse* y el *vanishing gradient*.
- $\lambda (\|\nabla_{\hat{x}}D_w(\hat{x})\|_2 - 1)^2$ ayuda a que la red crítica sea *1-Lipschitz* continua.

²¹<https://towardsdatascience.com/demystified-wasserstein-gan-with-gradient-penalty-ba5e9b905ead>

5.6.3.1. Entrenamiento de la red

Esta nueva arquitectura permite un nuevo proceso de entrenamiento de la WGAN.

Algorithm 1 WGAN with gradient penalty. We use default values of $\lambda = 10$, $n_{\text{critic}} = 5$, $\alpha = 0.0001$, $\beta_1 = 0$, $\beta_2 = 0.9$.

Require: The gradient penalty coefficient λ , the number of critic iterations per generator iteration n_{critic} , the batch size m , Adam hyperparameters α, β_1, β_2 .

Require: initial critic parameters w_0 , initial generator parameters θ_0 .

```

1: while  $\theta$  has not converged do
2:   for  $t = 1, \dots, n_{\text{critic}}$  do
3:     for  $i = 1, \dots, m$  do
4:       Sample real data  $\mathbf{x} \sim \mathbb{P}_r$ , latent variable  $\mathbf{z} \sim p(\mathbf{z})$ , a random number  $\epsilon \sim U[0, 1]$ .
5:        $\tilde{\mathbf{x}} \leftarrow G_\theta(\mathbf{z})$ 
6:        $\hat{\mathbf{x}} \leftarrow \epsilon \mathbf{x} + (1 - \epsilon) \tilde{\mathbf{x}}$ 
7:        $L^{(i)} \leftarrow D_w(\tilde{\mathbf{x}}) - D_w(\mathbf{x}) + \lambda(\|\nabla_{\hat{\mathbf{x}}} D_w(\hat{\mathbf{x}})\|_2 - 1)^2$ 
8:     end for
9:      $w \leftarrow \text{Adam}(\nabla_w \frac{1}{m} \sum_{i=1}^m L^{(i)}, w, \alpha, \beta_1, \beta_2)$ 
10:   end for
11:   Sample a batch of latent variables  $\{\mathbf{z}^{(i)}\}_{i=1}^m \sim p(\mathbf{z})$ .
12:    $\theta \leftarrow \text{Adam}(\nabla_\theta \frac{1}{m} \sum_{i=1}^m -D_w(G_\theta(\mathbf{z})), \theta, \alpha, \beta_1, \beta_2)$ 
13: end while

```

Figura 5.14: Proceso de entrenamiento de una WGAN. Fragmento del artículo de Gulrajani et al. Imagen extraída de [62].

En la Figura 5.14, se puede observar el detalle del proceso de entrenamiento de esta nueva arquitectura mejorada. Es importante subrayar que aquí no se puede utilizar *batch normalization* en el crítico. Tal y como menciona Gulrajani en su artículo, “la normalización por lotes cambia la forma del problema de discriminación, pasando de un mapeo de una sola entrada a una sola salida a un mapeo de un lote completo de entradas a un lote de salidas”. Debido a que se está trabajando con la dualidad de Kantorovich-Rubinshtein, el gradiente de cada par (imagen real, imagen sintética) será único y no un *batch* de muestras, como ocurre en otras arquitecturas.

CAPÍTULO 6

Recursos utilizados

6.1 Herramientas tecnológicas

Para la confección del presente trabajo se han aplicado y puesto en práctica diferentes conocimientos informáticos adquiridos a lo largo de todo el Grado en Ciencia de Datos. Se ha considerado adecuado presentar a continuación una breve reseña con el detalle de los recursos computacionales más utilizados en la elaboración de esta memoria, tanto a nivel de software como de hardware.

6.1.1. Software

Como ya se ha expuesto anteriormente, este proyecto está enfocado en el estudio de diferentes imágenes procedentes de tomografías PET y mamografías. Existen diversas herramientas de software para poder extraer, ver y analizar conjuntos de datos volumétricos de imágenes médicas. Para tal fin, en el trabajo se han utilizado visores DICOM de código abierto. Estos visores DICOM (por sus siglas en inglés, *Digital Imaging and Communication In Medicine*), son muy útiles para la visualización y análisis de imágenes médicas.

DICOM es un estándar de transmisión de imágenes clínicas. Las aplicaciones más comunes de este estándar son la visualización, manipulación, almacenamiento y transmisión de información de imágenes médicas. Este protocolo incluye la definición del formato del fichero DICOM y de los protocolos de comunicación de red basados en TCP/IP, los cuales permiten intercambiar información entre dos entidades que tengan capacidad para enviar y recibir información en dicho formato DICOM ¹.

Las dos herramientas de software libre utilizadas en el presente documento, compatibles con este estándar DICOM, son dos visores de código abierto conocidos como Miele-LXIV DICOM Viewer ² (*Medical Imaging ELEments*) y AMIDE ³ (*Amide's a Medical Imaging Data Examiner*), ambos ejecutables en sistemas operativos Mac OS X.

En otro orden de cosas, cabe mencionar que, tanto para el preproceso de las imágenes como para el desarrollo de las arquitecturas de aprendizaje profundo, el lenguaje de programación utilizado ha sido Python ⁴. Este lenguaje cuenta con una vasta comunidad de

¹<https://ca.wikipedia.org/wiki/DICOM>

²<https://dicom.3utilities.com/viewer.php>

³<http://amide.sourceforge.net>

⁴<https://www.python.org/>

usuarios, la cual ha desarrollado una amplia variedad de librerías que permiten afrontar cualquier problemática que surja en el marco de la programación. De entre estas librerías, las más utilizadas han sido las siguientes:

- Numpy: biblioteca de funciones matemáticas de alto nivel utilizada para el procesamiento multidimensional de vectores (*arrays*) y matrices ⁵.
- Matplotlib: biblioteca para la creación de gráficos a partir de los datos de entrada contenidos en listas o arrays introducidos en lenguaje Python ⁶.
- Pillow: biblioteca que permite el procesamiento de imágenes en formatos de archivo diferentes ⁷.
- OpenCV: librería de programación de código abierto dirigida a la visión por computador utilizada en el análisis de imágenes y vídeos en tiempo real ⁸.
- Keras / Tensorflow: Ambas librerías son de código abierto. Por una parte, *Tensorflow* ⁹ es una de las librerías más utilizadas para la creación de modelos de machine learning y deep learning. Fue diseñada por el equipo *Google Brain* de Google y está escrita en lenguaje C++. Por otra parte, *Keras* ¹⁰, escrita en lenguaje Python, es una interfaz API que se ejecuta sobre Tensorflow y aporta un nivel de abstracción de alto nivel de complejidad respecto a la misma, permitiendo la automatización de las tareas realizadas por dicha librería.

6.1.2. Hardware

Las unidades de procesamiento gráfico, o GPU (del inglés *Graphics Processing Unit*), son microprocesadores dedicados al procesamiento de gráficos y utilizados como apoyo del procesador principal del ordenador, o CPU (*Central Processing Unit*). Permiten acelerar considerablemente los procesos de cálculo al liberar de carga de trabajo a la CPU y posibilitando que ésta pueda dedicarse a otras tareas. Las GPU, en su concepción teórica, tienen una arquitectura similar a las CPU, pero son mucho más rápidas.

Son una parte esencial de la infraestructura moderna para la inteligencia artificial ya que, para aprovechar sus ventajas, se necesita una gran capacidad de cálculo. Las GPU han sido especialmente diseñadas y optimizadas para tareas de aprendizaje profundo.

En este proyecto se ha utilizado la GPU Tesla T4, proporcionada por *Google Colab* ¹¹. Este servicio cloud permite el uso gratuito de las GPU y TPU (*tensor processing unit*) ¹² de Google.

⁵<https://numpy.org>

⁶<https://matplotlib.org>

⁷<https://pillow.readthedocs.io/en/stable/>

⁸<https://opencv.org>

⁹<https://www.tensorflow.org>

¹⁰<https://keras.io>

¹¹<https://colab.research.google.com/?hl=es>

¹²Unidad de Procesamiento Tensorial. Circuito integrado desarrollado por Google para el aprendizaje automático.

La GPU Tesla T4 es una unidad de procesamiento gráfico basada en la arquitectura Turing y dirigida a la “aceleración de inferencia del modelo de aprendizaje profundo”¹³. *Turing* es la denominación en clave de esta GPU desarrollada por la compañía californiana Nvidia, especializada en el desarrollo de unidades de procesamiento gráfico. Lleva el nombre del destacado matemático e informático Alan Turing¹⁴.

```

+-----+
| NVIDIA-SMI 460.32.03   Driver Version: 460.32.03   CUDA Version: 11.2   |
+-----+-----+-----+-----+-----+-----+
| GPU  Name          Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf    Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
|=====  

|  0   Tesla T4             Off   | 00000000:00:04:0 Off  |          0          |
| N/A   59C    P8     10W /  70W |  0MiB / 15109MiB |          0%      Default |
|=====  

+-----+-----+-----+-----+-----+-----+
| Processes:                                     |
|  GPU   GI   CI          PID  Type   Process name                      GPU Memory |
|  ID   ID   ID                                 |             Usage                      |
+-----+-----+-----+-----+-----+-----+
| No running processes found                    |
+-----+

```

Figura 6.1: Descripción de características de la GPU Tesla T4, obtenida del *notebook* de Google Colab.

6.2 Conjuntos de datos

Este proyecto se ha podido llevar a cabo gracias a las imágenes históricas reales de mama facilitadas por dos fuentes principales. Por un lado, las proporcionadas por el *Instituto de Instrumentación para Imagen Molecular, i3M (UPV - CSIC)*¹⁵ y, por otro lado, las imágenes de libre acceso obtenidas en la página web de la *University of South Florida*¹⁶.

Uno de los primeros problemas a los que se enfrentó este trabajo fue, obviamente, la cuestión de la clasificación automática de las imágenes de las cuales se disponía. Era necesario poder contar con diversas alternativas que permitieran solucionar el reto propuesto de manera eficaz.

La utilización de redes neuronales, sustentadas en librerías como las analizadas anteriormente, está siendo en la actualidad de gran ayuda en problemáticas de este tipo, proporcionando unos excelentes resultados. El principal objetivo de estas redes neuronales es poder obtener estimaciones de alta precisión a partir de bases o conjuntos formados por un gran número de datos (*datasets*).

¹³https://colab.research.google.com/github/d2l-ai/d2l-tvm-colab/blob/master/chapter_gpu_schedules/arch.ipynb#scrollTo=YfUox2_VbXSI

¹⁴[https://en.wikipedia.org/wiki/Turing_\(microarchitecture\)](https://en.wikipedia.org/wiki/Turing_(microarchitecture))

¹⁵Centro de investigación mixto creado en 2010 por la Universidad Politécnica de Valencia (UPV) y el Consejo Superior de Investigaciones Científicas (CSIC) con sede en el Campus de Vera de la UPV.

¹⁶<http://www.eng.usf.edu/cvprg/Mammography/Database.html>

6.2.1. Imágenes del i3M

Este conjunto de imágenes, procedentes del i3M y facilitadas por los tutores del presente trabajo, recibe el nombre de MAMMI por las iniciales en inglés del Proyecto Europeo *Mammography with Molecular Imaging* (Mamografía Mediante Imagen Molecular). Este conjunto de mamografías MAMMI-PET obtenidas mediante imagen molecular PET, surgió como un proyecto dentro de la Unión Europea ¹⁷ con el objetivo de mejorar la detección en el cáncer de mama y su Coordinador fue el Dr. José M. Benlloch ¹⁸, actual Director del i3M.

Como ya hemos visto en el capítulo 2, donde se expone el actual contexto tecnológico en el cual se desarrollan este tipo de imágenes, la tomografía por emisión de positrones es una técnica que permite visualizar la actividad metabólica del organismo. Es decir, proporciona imágenes tridimensionales de la actividad celular de la mama para detectar posibles lesiones neoplásicas.

La gran mayoría de las actuales mamografías requieren la compresión de la mama analizada entre las dos placas de los mamógrafos convencionales. Esto, además de ser incómodo para la paciente, genera una distorsión de la anatomía del seno y supone una dificultad añadida para determinar con exactitud la localización y tamaño del tumor. La nueva técnica incorporada en el MAMMI-PET elimina el aplastamiento de la mama al tumbar a las pacientes en decúbito prono, de tal manera que las mamas quedan suspendidas y situadas dentro del anillo detector del tomógrafo, sin necesidad de comprimir el tejido mamario.

Este sistema, diseñado por investigadores del i3M y desarrollado por la empresa valenciana Oncovision, supone una mejora significativa en el “diagnóstico precoz de cáncer de mama, pudiendo llegar a detectar la patología antes y con mayor seguridad que los mejores equipos de diagnóstico utilizados hoy en día” ¹⁹.

La alta resolución de MAMMI (1.6 mm) permite la detección de tumores muy pequeños, incluso en mujeres con mamas densas, lo cual es vital para el diagnóstico del cáncer de mama en estadios tempranos de la enfermedad [63].

El material proporcionado por el Dr. Samuel Morillas y la Dra. María José Rodríguez está organizado en dos carpetas principales: SR (*Single Ring*) y DR (*Double Ring*). Como ya se ha podido ver en el epígrafe 2.5, dedicado a la generación de imágenes PET, las señales de una tomografía por emisión de positrones se obtienen mediante unos anillos detectores instalados en el escáner, de ahí la denominación de *ring* en los nombres de cada una de las carpetas.

Siguiendo las indicaciones de ambos tutores, en el proyecto se han utilizado principalmente imágenes procedentes de la primera carpeta (SR), ya que las contenidas en la segunda de ellas (DR) corresponden a casos especiales que no entrarían dentro del objeto de estudio de este trabajo.

La carpeta SR está organizada en 178 subcarpetas, que contienen alrededor de unos 400 archivos de imagen DICOM. Asociado a cada archivo de imagen DICOM existe un fichero con la extensión *.lm que proporciona información complementaria. Los ficheros LM son archivos de datos de Microsoft Windows que contienen información estadística utilizada para mostrar texto predictivo en aplicaciones de Windows y Microsoft.

¹⁷Proyecto FP6-2005-LIFESCIHEALTH-7-037555, 2007-2010

¹⁸<https://www.i3m.upv.es/view.php/DIM/Members/benlloch>

¹⁹https://www.upv.es/entidades/I3M/noticia_776567c.html

Este tipo de archivos no se encuentran categorizados, ya que están asociados con varias aplicaciones y ficheros del sistema. De hecho, pertenecen a los denominados “archivos misceláneos” (*misc files*). La mayoría de las veces, este tipo de archivos son en realidad metadatos, es decir, son «datos sobre datos» y nos proporcionan información adicional sobre los datos contenidos en un documento.

En nuestro caso, estos archivos LM contienen distintas variables que se utilizan en el proceso de reconstrucción de las imágenes PET obtenidas por el escáner del tomógrafo. Las variables que más se suelen emplear en dicho proceso de reconstrucción de imágenes son las siguientes:

- Instante exacto en el tiempo en el que se produce la aniquilación mútua entre electrones y positrones.
- Cuantificación de la energía emitida cuando se produce dicha aniquilación.
- Proyección de la línea de respuesta LOR, definida a partir de las coordenadas de detección de los impactos de los fotones en cada uno de los cristales de centelleo del anillo detector del tomógrafo, es decir, (x_1, y_1) en un cristal y (x_2, y_2) en el opuesto.
- Identificador de la pareja de bloques de cristales detectores implicados en el impacto.

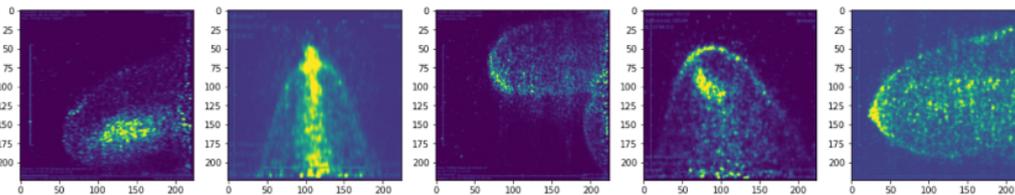


Figura 6.2: Ejemplos de imágenes de tomografía PET procedentes del i3M.

6.2.2. Imágenes de la University of South Florida (USF)

Las imágenes de mamografías de la *University of South Florida* (USF) utilizadas en este trabajo, son de libre acceso y están disponibles en la página web de dicha universidad ²⁰. Son el resultado de un proyecto de la USF para la creación de una base de datos digital de mamografías de cribado (DDSM por las siglas en inglés de *Digital Database for Screening Mammography*), en colaboración junto con el Hospital General de Massachusetts y los laboratorios *Sandia National Laboratories*, dependientes de la Administración Nacional de Seguridad Nuclear de los Estados Unidos.

El objetivo principal de esta base de datos es la de facilitar la investigación en el desarrollo de algoritmos de ayuda en el cribado de mamografías y, de forma secundaria, el soporte a la enseñanza y la formación.

El DDSM contiene mamografías obtenidas en el Hospital General de Massachusetts (*Massachusetts General Hospital*), la Facultad de Medicina de la Universidad Wake Forest (*Wake Forest University*), el Hospital del Sagrado Corazón (*Sacred Heart Hospital and ISMD*,

²⁰<http://www.eng.usf.edu/cvprg/Mammography/Database.html>

Incorporated) y la Facultad de Medicina de la Universidad Washington de Sant Louis (*Washington University in St. Louis*).

La base de datos se completó en otoño de 1999 y todos los casos procedían de mamografías realizadas entre octubre de 1988 y febrero de 1999. Contiene 2620 estudios diferentes de mamografías de cribado. Cada estudio incluye dos imágenes de cada mama, junto con alguna información asociada de la paciente (edad, índice de densidad mamaria ACR²¹, descripción de las anomalías, índice de sutileza de las anomalías²²) e información de la imagen (escáner, resolución espacial, etc.)²³.

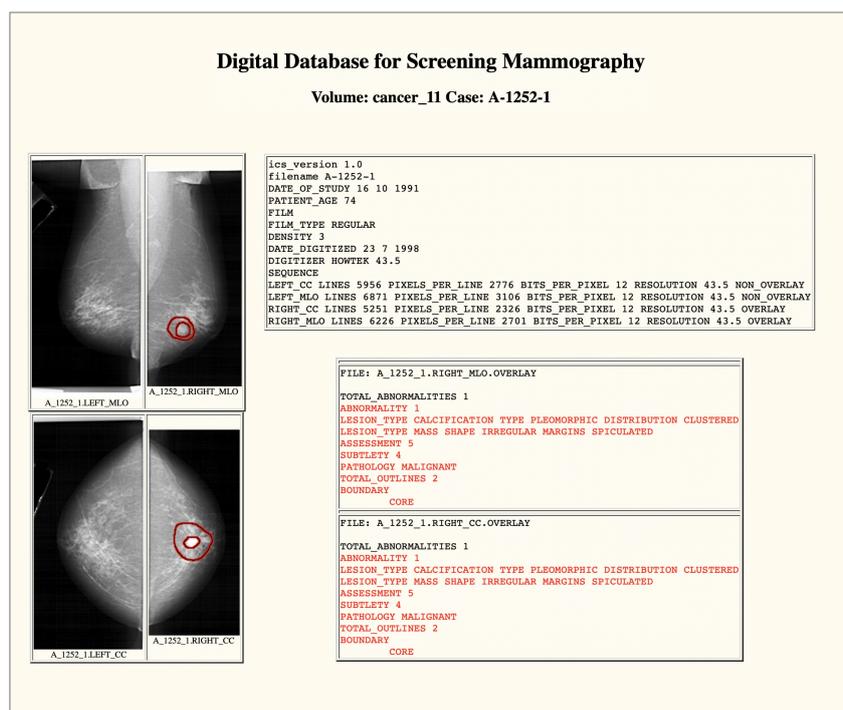


Figura 6.3: Detalle de un caso particular de cribado de mama procedente de la *University of South Florida*.

En la Figura 6.3 se muestra el caso 1252 del DDSM, donde se puede apreciar el detalle de la información asociada a las imágenes de esta base de datos [64].

²¹La sigla ACR corresponde a la clasificación realizado por el Colegio Americano de Radiología para indicar que las mamas son densas, es decir que están compuestas predominantemente por tejido glandular y de sostén (tejido fibroglandular). <https://www.diagnosticorojas.com.ar/blog/mamografia/que-significa-acr-c/>

²²*Subtlety Score*. Una calificación subjetiva de la visibilidad de una lesión en una imagen médica, que va de no visible a evidente.

²³<https://www.imaginis.com/pro-teaching-files/university-of-south-florida-digital-database-for-screening-mammography?r>

CAPÍTULO 7

Experimentación y resultados

En este capítulo, una vez detalladas en profundidad el tipo de redes neuronales que se emplearán en el estudio, así como los *datasets* de imágenes seleccionados, se van a llevar a cabo una serie de experimentaciones sobre las distintas arquitecturas GAN ya propuestas, para así poder valorar cuál de ellas es la que permite generar las imágenes de mama más realistas posibles.

Inicialmente, se expondrá el procesamiento previo aplicado a las imágenes empleadas en el entrenamiento de los distintos modelos. Asimismo, se explicarán brevemente las métricas empleadas para la evaluación y comparación de las distintas redes construidas. Seguidamente se procederá a definir los distintos modelos propuestos, haciendo hincapié en las capas que forman cada una de estas arquitecturas. Para finalizar, se realizará una descripción de los distintos experimentos realizados, junto con los resultados obtenidos en cada uno de ellos.

7.1 Procesamiento previo

No se ha llevado a cabo un procesamiento exhaustivo de las imágenes disponibles ya que las herramientas de software libre utilizadas son compatibles con el estándar DICOM y estas imágenes pueden visualizarse perfectamente. Se ha optado por realizar una estandarización de las muestras, es decir, conseguir que todas ellas tengan el mismo formato, resolución y escala de colores ¹.

Las modificaciones efectuadas han sido las siguientes:

- Transformación de formato *.gif* a *.jpg*. Empleada principalmente en las mamografías procedentes del repositorio de la *University of South Florida*, ya que éstas se encuentran disponibles en formato *.gif*.
- Se ha decidido escoger una resolución media (224x224x1), no muy elevada, ya que la generación de imágenes de alta resolución requiere grandes cantidades de memoria GPU debido a su mayor coste computacional ².
- Transformación a escala de grises.

¹Estas transformaciones se pueden consultar en el repositorio de GitHub.

²<https://developer.nvidia.com/blog/generating-and-editing-high-resolution-synthetic-images-with-gans/>

- Por motivos de eficiencia en el entrenamiento de las distintas arquitecturas, se han unido todas las imágenes en una única matriz y ésta se ha convertido en un *NumPy Binary File (.npy)*³. Se trata de un formato de fichero binario estándar que permite almacenar archivos individuales en forma de matriz NumPy en disco.

7.2 Métricas utilizadas

Como bien es sabido, las métricas más utilizadas en cualquier algoritmo de aprendizaje automático y aprendizaje profundo son la precisión (*accuracy*) y la pérdida (*loss*). Éstas permiten evaluar el rendimiento general del modelo, aunque en ciertos casos no suelen ser medidas muy robustas. En el ámbito de las GAN sí conviene tener como referencia una métrica robusta que nos dé información acerca del grado de similitud de una imagen sintética generada respecto a la original.

Es por este motivo por el cual se ha utilizado en este trabajo la *Fréchet Inception Distance* (FID), muy útil para estimar la calidad de las muestras generadas por las GAN. Cuanto menor sea el valor de esta métrica, más realistas serán las imágenes creadas. Su función principal es la de calcular la distancia entre las distribuciones de probabilidad de las muestras reales y sintéticas. Para ello, es necesario integrar el conjunto de estas distribuciones en un vector de características, el cual se utilizará en el cálculo de la media y la varianza de los datos reales y sintéticos, medidas que permitirán estudiar la similitud entre estos dos grupos de imágenes. [65]

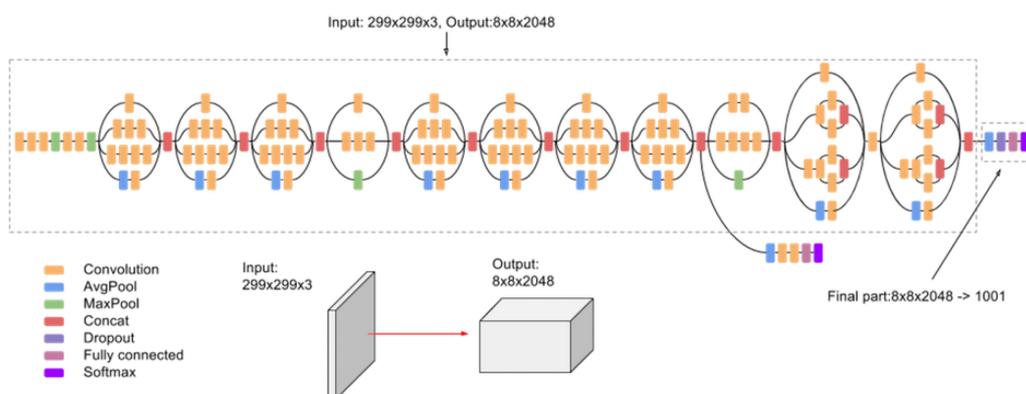


Figura 7.1: Modelo Inception-v3. Imagen extraída de Google Cloud.

Esta métrica recibe el nombre de *Inception* debido a que se encarga de obtener ese vector de características de una capa intermedia de la red denominada *Inception-v3* y desarrollada por Google. Con ella se obtiene una representación de alto nivel de la imagen. Como puede observarse en la Figura 7.1, se trata de un modelo basado en redes neuronales convolucionales y empleado, principalmente, en tareas de clasificación de imágenes.

Tal y como se enuncia en la propia publicación de Google, “el modelo en sí se compone de bloques de construcción simétricos y asimétricos, incluyendo convoluciones, *pooling* promedio, *pooling* máximo, concatenaciones, *dropouts* y capas totalmente conectadas. La normalización por lotes se utiliza ampliamente en todo el modelo y se aplica a las en-

³<https://numpy.org/devdocs/reference/generated/numpy.lib.format.html>

tradas de activación. Las pérdidas se calculan con Softmax”⁴. La última capa de *pooling* es la que se utiliza como vector de características para calcular la medida correspondiente.

Definido ya el origen de dicho vector, la métrica FID se puede formular de la siguiente manera:

$$\text{FID} = \|\mu_r - \mu_g\|^2 + \text{Tr} \left(\Sigma_r + \Sigma_g - 2 (\Sigma_r \Sigma_g)^{1/2} \right) \quad (7.1)$$

donde,

- μ_r y μ_g son las medias de las distribuciones de probabilidad de datos reales y sintéticos, respectivamente.
- Σ_r y Σ_g son las covarianzas de las distribuciones de probabilidad de los datos reales y sintéticos, respectivamente.
- Tr es la traza de la matriz, es decir, la suma de los elementos de la diagonal principal.

Para la validación de los resultados obtenidos de nuestra experimentación, el uso de la métrica FID ha resultado fundamental, aunque también se ha recurrido al empleo de la función de pérdida *loss*.

7.3 Modelos propuestos

Una vez conocido cómo se ha realizado el procesamiento de las imágenes y qué métricas se emplearán en la validación de los resultados, es preciso desarrollar la parte más importante de este proyecto, la creación de los diferentes modelos a utilizar. En el capítulo 5 se expuso el marco teórico de las diferentes redes neuronales generativas adversarias empleadas en este trabajo (GAN, DCGAN y WGAN-GP), y se profundizó también en los mecanismos de entrenamiento de las mismas.

En los tres siguientes subepígrafes, se detallan los elementos significativos que forman las arquitecturas de cada una de estas redes neuronales.

7.3.1. GAN

El primer modelo planteado se fundamenta en una arquitectura muy básica, la cual se suele utilizar, no sólo en tareas como la que nos ocupa, sino también en otros cometidos en donde se emplea el aprendizaje profundo o, en términos más generales, la inteligencia artificial.

Como ya hemos visto, un modelo GAN está compuesto por dos redes neuronales, generador y discriminador, las características de las cuales son:

⁴<https://cloud.google.com/tpu/docs/inception-v3-advanced>

Generador

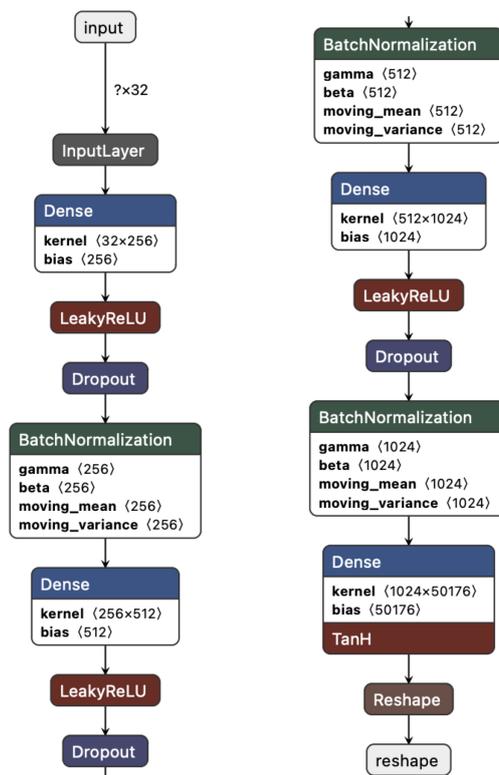


Figura 7.2: Arquitectura del generador GAN.

En la Figura 7.2 se observa la topología de la arquitectura del generador GAN empleado en el primer modelo. Como puede apreciarse, se han ido añadiendo una serie de capas densamente conectadas las cuales permitirán la generación de imágenes con la mayor resolución posible, es decir, que contengan el mayor número de detalles posible.

Por otra parte se han añadido, además, bloques *batch normalization* para evitar el sobreentrenamiento de la red. Relacionado con ello, se ha procedido también a la aplicación de una regularización *dropout*⁵, que suele ser muy útil cuando no estamos trabajando con *datasets* lo suficientemente extensos.

Finalmente, debido al hecho de encontrarnos en el generador, el último paso a realizar es la ejecución de un *reshape* en la última capa densa, para así quedarnos con el tamaño de salida que deseamos, en este caso, 224x224x1.

⁵Técnica de regularización que se encarga de eliminar una serie de neuronas de forma aleatoria.

Discriminador

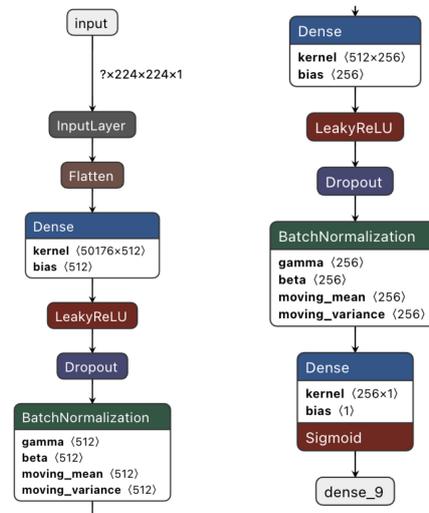


Figura 7.3: Arquitectura del discriminador GAN.

En la Figura 7.3 se puede observar la arquitectura del discriminador, cuyos elementos son los mismos que intervienen en el generador aunque el objetivo, en este caso, no es el de generar imágenes sintéticas, sino decidir si éstas son reales o falsas.

7.3.2. DCGAN

El segundo modelo aplicado en el proceso de experimentación ha sido una *Deep Convolutional GAN*, la cual presenta una serie de mejoras respecto a la red tratada anteriormente. Para su correcta implementación, se han seguido las directrices presentadas por Radford et al. en su artículo publicado en 2015 [60].

Generador

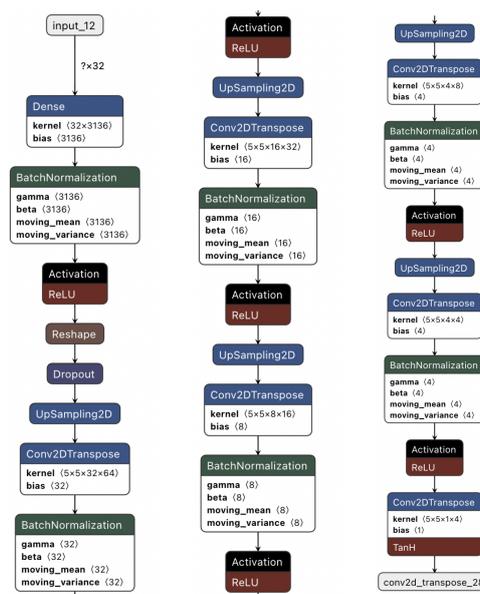


Figura 7.4: Arquitectura del generador DCGAN.

En la Figura 7.4 se muestra la arquitectura de un generador en un modelo DCGAN. Aquí se han añadido capas deconvolucionales (*Conv2DTranspose*) y de submuestreo (*Up-Sampling2D*) las cuales van a permitir aumentar la dimensión final de la imagen.

En los bloques *Conv2DTranspose* se ha aplicado un *kernel size* con valor 5. No hay ninguna razón matemática por la cual se haya escogido dicha configuración. Después de estudiar el estado del arte actual y la experiencia de otros investigadores, se ha podido observar que los valores más comunes de dicho parámetro son 3 y 5.

De la misma forma que en el modelo anterior, se ha aplicado *batch normalization* después de cada capa deconvolucional. Finalmente, a diferencia de la GAN básica, se ha escogido la función de activación *ReLU* para cada una de las capas, a excepción de la última, que debe contener la tangente hiperbólica *tanh*.

Discriminador

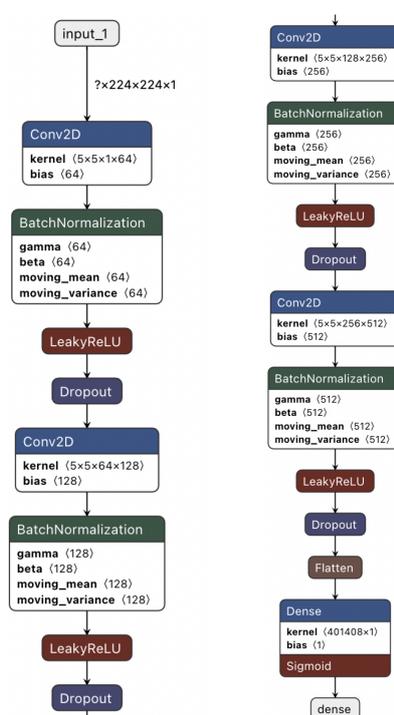


Figura 7.5: Arquitectura del discriminador DCGAN.

La estructura del discriminador, la cual se puede observar en la Figura 7.5, es muy similar a la del generador, aunque existen algunas diferencias en las capas y funciones de activación. En este caso, Radford recomienda emplear bloques convolucionales (*Conv2D*) y funciones de activación *LeakyReLU* en cada una de éstas.

7.3.3. WGAN - GP

El último modelo aplicado en la experimentación es una Wasserstein GAN con *penalización de gradiente* (WGAN-GP), la cual utiliza una función de pérdida distinta a las dos anteriores redes, tal y como se ha estudiado en el subepígrafe 5.6.3.

Se ha optado por no adjuntar la topología del generador y discriminador de este modelo puesto que dicha red está basada en la misma arquitectura que la presentada ante-

riormente, a excepción de las capas de *batch normalization*, las cuales no pueden utilizarse aquí ya que este tipo de normalización trabaja con *batches*. No obstante, debido a la dualidad de Kantorovich-Rubinstein, el gradiente de cada uno de los pares es único en este modelo.

7.4 Experimentos realizados. Parámetros

Para la correcta validación de los modelos GAN propuestos se han realizado distintos experimentos, teniendo en cuenta los distintos parámetros que intervienen en este tipo de arquitecturas: optimizadores, tasas de abandono (*dropout*), tamaños de lote (*batch size*) y tasas de aprendizaje (*learning rate*) y, en el caso de las arquitecturas WGAN-GP, el *gradient penalty loss weight* (λ).

Dichos parámetros varían en función del tipo de red que se está evaluando. Sin embargo, en este caso, éstos no varían entre las arquitecturas GAN y DCGAN ya que ambas comparten muchas similitudes, excepto en lo referente a las capas convolucionales: las redes DCGAN contienen capas convolucionales y las GAN básicas, no.

En este epígrafe se ha intentado realizar una comparación “justa” de todos los modelos propuestos ya que en cada uno de ellos se aplican los mismos parámetros, a excepción del tamaño de lote que varía en función del conjunto de imágenes que se esté empleando. Sin embargo, tal y como se verá en el subepígrafe 7.4.2, el dilatado periodo de entrenamiento del modelo WGAN-GP no ha permitido poder valorar el rendimiento del mismo utilizando todos sus parámetros y, por ello, en dicho modelo sólo se han empleado el tamaño de lote y el parámetro λ .

Sería plausible afirmar pues que, si se hubiese podido aplicar una mayor cantidad de valores a los distintos parámetros presentados, la experimentación hubiera podido ofrecer un mayor valor que el ya conseguido, es decir, se hubiese podido llegar a obtener un número de conclusiones más considerable.

La limitación existente en la disponibilidad de recursos computacionales ha originado que los tiempos de entrenamiento de los distintos modelos, en cada ejecución realizada, resultaran ser muy extensos. Concretamente, en el caso de una red GAN básica, el tiempo de ejecución de cada experimento suponía aproximadamente unos treinta minutos. Este periodo se alargaba hasta los cincuenta minutos en el caso del modelo DCGAN y alrededor de dos horas en las arquitecturas WGAN-GP.

Este es el motivo por el cual no se han considerado más alternativas de experimentación de las ya propuestas. En el capítulo 9, relativo a los posibles desarrollos futuros que pudiesen derivarse de este trabajo, nos referiremos otra vez a esta posibilidad de ampliar el número de parámetros en la experimentación para obtener mejores resultados.

7.4.1. GAN y DCGAN

En las Tablas 7.1 y 7.2 se detallan todos los parámetros que se han utilizado en las arquitecturas GAN y DCGAN elegidas. La principal diferencia radica en el tamaño de lote o *batch size*, debido a la diferencia en el número de imágenes que existen en un repositorio u otro.

Tabla 7.1: Parámetros utilizados en las arquitecturas GAN y DCGAN para imágenes del i3M.

Parámetros	Valores
Optimizador	Adadelata/ Adam / RMSprop
Dropout	0.4/0.5/0.6
Batch Size	8/16/32
Learning Rate	0.0008/0.0004/0.001

Tabla 7.2: Parámetros utilizados en las arquitecturas GAN y DCGAN para imágenes del repositorio de la *University of South Florida*.

Parámetros	Valores
Optimizador	Adadelata/ Adam / RMSprop
Dropout	0.4/0.5/0.6
Batch Size	16/32/64
Learning Rate	0.0008/0.0004/0.001

7.4.2. WGAN - GP

De la misma forma, en las Tablas 7.3 y 7.4 se detallan todos los parámetros utilizados en la WGAN con penalización de gradiente. En esta red se han realizado menos experimentos debido a que los tiempos de ejecución son mucho más extensos en comparación con el resto de las arquitecturas.

Como en el caso anterior, el tamaño del lote varía en función del tamaño de la base de imágenes utilizada. En esta ocasión no se ha realizado una experimentación para comparar el rendimiento de las distintas tasas de abandono, debido a la escasez de recursos computacionales.

Por último conviene destacar que, en esta red, siguiendo las recomendaciones de Gulrajani et al. [62] en el proceso de entrenamiento, solamente se ha utilizado el optimizador *Adam*.

Tabla 7.3: Parámetros utilizados en la arquitectura WGAN para imágenes del i3M.

Parámetros	Valores
Batch Size	8/16
Gradient Penalty loss weight	0.1/1/10

Tabla 7.4: Parámetros utilizados en la arquitectura WGAN para imágenes del repositorio de la *University of South Florida*.

Parámetros	Valores
Batch Size	16/32
Gradient Penalty loss weight	0.1/1/10

7.5 Resultados

Analizados ya el procesamiento de las imágenes, las métricas de evaluación, las topologías de las distintas redes y sus correspondientes experimentos, finalmente se pueden presentar los resultados obtenidos de los diferentes experimentos realizados.

Tal y como se ha explicado en el epígrafe 4.2, donde se hace referencia a la solución propuesta en el proyecto, el proceso de experimentación se inicia con los ensayos realizados sobre el conjunto de tomografías MAMMI-PET proporcionado por el i3M. En el caso de existir margen de mejora en la representación de las imágenes, la experimentación se amplía con el repositorio de mamografías de la *University of South Florida (USF)*.

Por último es conveniente indicar que, debido a la importante diferencia en el número de muestras entre una base de imágenes y otra, el número de *epochs* será distinto en cada una de éstas, entrenando los modelos un total de 4000 iteraciones en las tomografías y 2000 en las mamografías.

La estructura de este proceso de experimentación sigue el siguiente patrón:

- Se selecciona inicialmente el repositorio de imágenes proporcionado por el i3M y, posteriormente, el de la USF.
- Cada uno de estos *datasets* de imágenes es empleado sucesivamente en los modelos GAN, DCGAN y WGAN-GP.
- Para cada modelo anterior se realizan varios experimentos: tres, en los casos de GAN y DCGAN, y un solo experimento en las WGAN-GP, ya que aquí el tiempo de cómputo resulta ser mucho más extenso y, por ello, la experimentación sufre una restricción.

7.5.1. Experimentos con muestras i3M

7.5.1.1. Modelo i3M - GAN

Experimento 1: Optimizadores y dropouts

El primer experimento se enfocó en la evaluación del rendimiento de distintos optimizadores y tasas de abandono para las muestras.

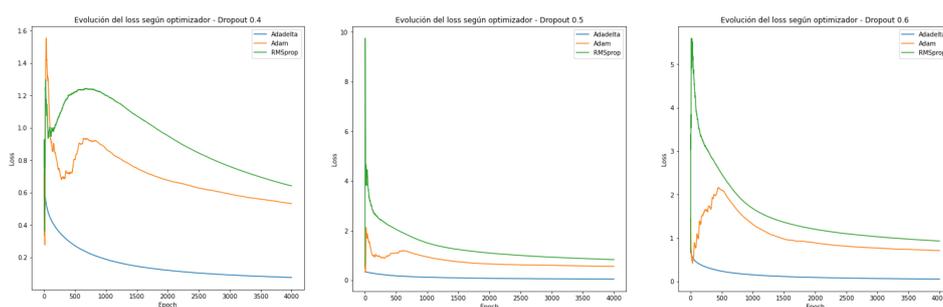


Figura 7.6: Modelo i3M - GAN. Evolución del *loss* según optimizador aplicando *dropouts* de 0.4, 0.5 y 0.6.

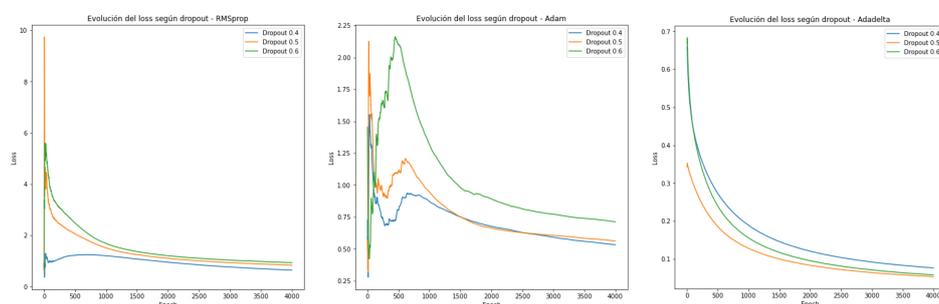


Figura 7.7: Modelo i3M - GAN. Evolución del *loss* según *dropout*, aplicando optimizadores RMSprop, Adam y Adadelta.

En la Figura 7.6 se muestra la evolución del *loss* de cada uno de los optimizadores después de 4000 iteraciones (*epochs*).

Se puede observar que, iteración tras iteración, *RMSprop* es el que reporta un mayor *loss* en todas las situaciones. Al estar midiendo el error del discriminador, se debe tener en cuenta que éste está intentando maximizar la distancia para, de esta forma, saber distinguir mejor entre las imágenes originales y las sintéticas. Teniendo esto en cuenta, *RMSprop* resultaría ser el que tiene un mayor rendimiento, tanto si se aplican *dropouts* de 0.4, 0.5 ó 0.6.

Por otra parte, en la Figura 7.7 también se muestra la evolución del error pero, en este caso, el de cada uno de los *dropouts* con los tres optimizadores que ya sabemos. Si se aplica una tasa de abandono de 0.6, a excepción de *Adadelta*, se obtiene el error más elevado.

Para corroborar estas dos últimas conclusiones, se procedería a calcular el FID para, de esta forma, saber cómo de realistas son las imágenes creadas por esta red GAN.

Tabla 7.5: Modelo i3M - GAN. Valores FID según optimizador y *dropout*.

FID	$D = 0.4$	$D = 0.5$	$D = 0.6$
RMSprop	373.394	356.103	346.530
Adam	350.327	301.263	320.038
Adadelta	478.088	444.362	419.815

Finalmente, en la Tabla 7.5 se pueden comprobar los resultados de dicha métrica en cada una de las ejecuciones. Se puede observar que *Adam - Dropout 0.5* es la configuración que reporta un menor valor, concretamente 301.263.

A modo de conclusión, podemos observar que el *loss* puede dar una aproximación del rendimiento de nuestro modelo. Sin embargo, para confirmar los resultados obtenidos, es conveniente utilizar una medida adecuada para la tarea que se está tratando, en este caso, medir el desempeño de redes generativas a través de la comparación de imágenes.

Experimento 2: Batch size

En este segundo experimento se procede a estudiar el rendimiento de las GAN básicas comparando distintos tamaños de lote (*batch size*). En este caso, para evitar un tamaño de experimentación demasiado extenso, se utilizó únicamente el optimizador *RMSprop*, ya que éste es el más empleado en el entrenamiento de las redes neuronales.

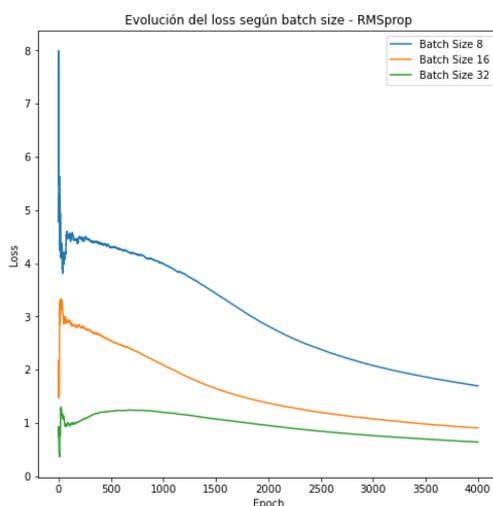


Figura 7.8: Modelo i3M - GAN. Evolución del *loss* según *batch size*, aplicando optimizador *RMSprop*.

En la Figura 7.8 se observa la evolución del error de cada uno de los *batch sizes*. Se puede apreciar que, utilizando un tamaño de lote con valor igual a 8, se obtiene el error de discriminador más elevado, concretamente 1.69 en la última *epoch*. Sin embargo, según la métrica FID, utilizando un tamaño de lote 16 se obtiene un mejor resultado (323.803).

Tabla 7.6: Modelo i3M - GAN. Valores FID según *batch size*.

FID	<i>B.S</i> = 8	<i>B.S</i> = 16	<i>B.S</i> = 32
<i>RMSprop</i>	329.973	323.803	373.394

Experimento 3: Learning rate

En este último experimento se evalúa el rendimiento de distintas tasas de aprendizaje (*learning rate*). Cabe decir que todas las configuraciones anteriores se han ejecutado con un *learning rate* de 0.0008. De forma análoga que en el experimento anterior, únicamente se ha utilizado el optimizador *RMSprop*.

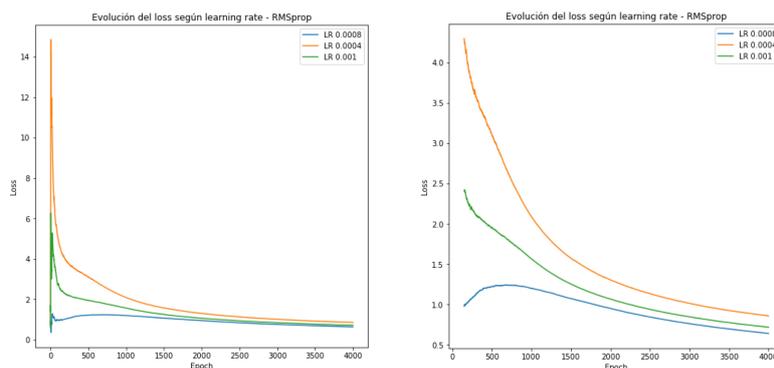


Figura 7.9: Modelo i3M - GAN. Evolución del *loss* según *learning rate*, aplicando optimizador *RMSprop*. La imagen derecha aporta la misma información que la de la izquierda pero se han eliminado las primeras 150 *epochs* por cuestiones de escala, pudiéndose así visualizar mejor los valores.

La evolución del error utilizando dicho parámetro se puede apreciar en la Figura 7.9. A priori, aunque la escala de valores es muy extensa en las primeras iteraciones, se puede observar que el rendimiento de las tres configuraciones es muy similar.

En la Tabla 7.7 se presentan los valores FID de cada tasa de aprendizaje. A la vista de los mismos, se elige finalmente un *learning rate* de 0.001 (FID=330.344) ya que es el que ofrece el mejor resultado.

Tabla 7.7: Modelo i3M - GAN. Valores FID según *learning rate*.

FID	<i>L.R</i> = 0.0008	<i>L.R</i> = 0.0004	<i>L.R</i> = 0.001
<i>RMSprop</i>	373.394	371.663	330.344

Imágenes finales obtenidas con la mejor configuración en el modelo i3M - GAN

En base a lo expuesto anteriormente podemos decir que, mediante una arquitectura GAN únicamente con capas densamente conectadas, la configuración *Adam* con una tasa de abandono 0.5 genera las representaciones MAMMI-PET más realistas.

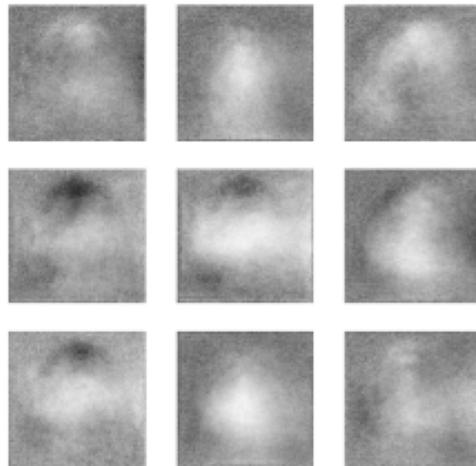


Figura 7.10: Modelo i3M - GAN. Imágenes de la mama obtenidas con la mejor configuración posible.

Sin embargo, si observamos las imágenes representadas en la Figura 7.10, el resultado de éstas deja mucho que desear. Debido a que estamos trabajando con capas densamente conectadas, solamente se puede llegar a apreciar el contorno de la mama.

En el siguiente apartado se llevarán a cabo los mismos experimentos realizados anteriormente pero, en este caso, empleando una *Deep Convolutional GAN*. La combinación de capas convolucionales y *fully connected* que nos aporta dicha arquitectura nos permitirá obtener, no sólo el contorno de la mama, sino también detalles más concretos de su anatomía.

7.5.1.2. Modelo i3M - DCGAN

Experimento 1: Optimizadores y dropouts

Anteriormente hemos identificado que, para una GAN sencilla, la configuración *Adam* y tasa de abandono con valor 0.5 era la que generaba representaciones más realistas. Si nos fijamos en la Figura 7.11 y 7.12, los resultados no son tan evidentes como entonces, ya que los errores son muy similares iteración tras iteración.

Los resultados más nítidos se obtienen cuando se emplea un *dropout* con valor 0.4, con el optimizador *Adadelta* reportando el error del discriminador más elevado (0.4793 en la última *epoch*), aunque está bastante emparejado con *RMSprop* (0.3732 en la última *epoch*).

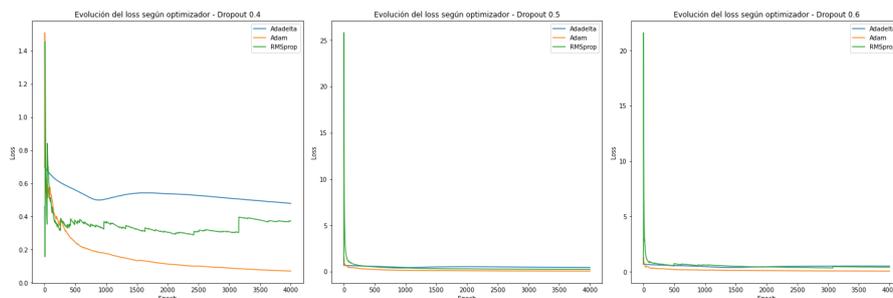


Figura 7.11: Modelo i3M - DCGAN. Evolución del *loss* según optimizador, aplicando tasas de abandono (*dropouts*) de 0.4, 0.5 y 0.6.

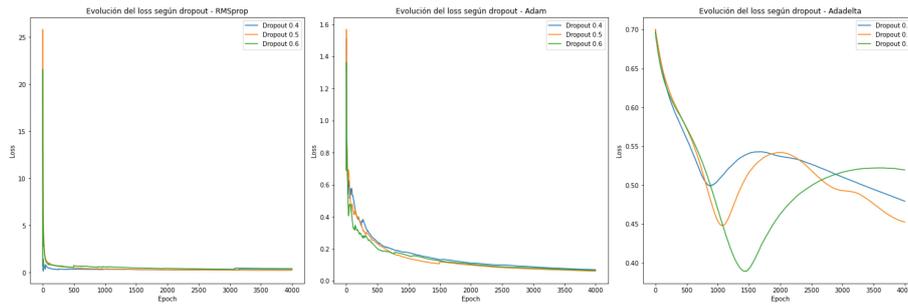


Figura 7.12: Modelo i3M - DCGAN. Evolución del *loss* según *dropout*, aplicando optimizadores RMSprop, Adam y Adadelta.

Tabla 7.8: Modelo i3M - DCGAN. Valores FID según optimizador y *dropout*.

FID	$D = 0.4$	$D = 0.5$	$D = 0.6$
RMSprop	219.420	345.329	294.231
Adam	354.287	345.195	336.373
Adadelta	301.414	365.519	226.043

A continuación, se procede al empleo de la métrica FID para la confirmación de nuestros resultados, los cuales se pueden apreciar en la Tabla 7.8. A simple vista, los valores son muy inferiores a los que se han obtenido en el modelo anterior.

Finalmente, con un FID de 219.420, se puede concluir que *RMSprop* con *dropout* 0.4 es la mejor combinación de parámetros en este primer experimento.

Experimento 2: Batch size

Si nos fijamos en la Figura 7.13, volvemos a observar que la evolución del *loss* en las tres situaciones es prácticamente igual, con unos valores de 0.4205, 0.3958 y 0.3732, respectivamente. Asimismo, la métrica FID confirma esa similitud, a excepción del tamaño de lote 8. La tercera configuración sería la óptima, aunque nos podríamos quedar con cualquiera de las dos últimas.

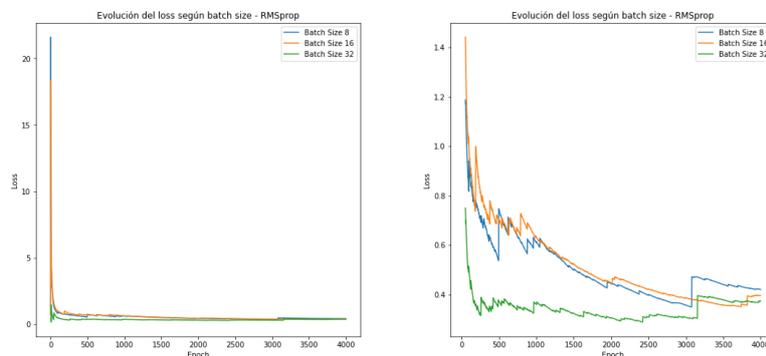


Figura 7.13: Modelo i3M - DCGAN. Evolución del *loss* según *batch size*, aplicando optimizador RMSprop. La imagen derecha aporta la misma información que la izquierda pero, en este caso, se han eliminado las primeras 150 *epochs* por cuestiones de escala, pudiéndose así visualizar mejor los valores.

Tabla 7.9: Modelo i3M - DCGAN. Valores FID según *batch size*.

FID	$B.S = 8$	$B.S = 16$	$B.S = 32$
RMSprop	289.900	224.051	219.420

Experimento 3: Learning rate

Para terminar con la experimentación de este segundo modelo i3M - DCGAN, en la Figura 7.14 podemos observar la evolución del error utilizando distintas tasas de aprendizaje. De forma semejante al anterior experimento, los valores obtenidos son muy similares en las tres configuraciones (0.3732, 0.2389 y 0.4128, respectivamente).

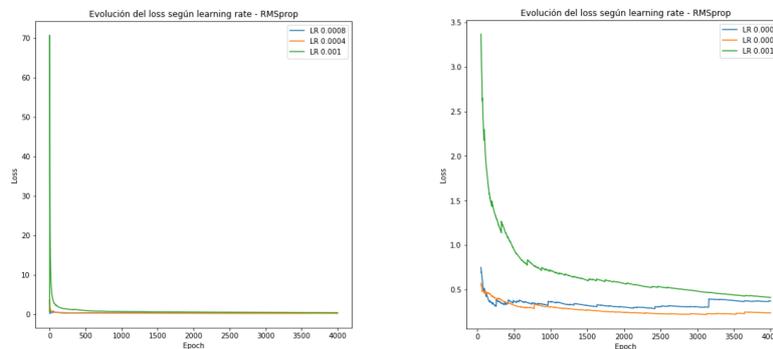


Figura 7.14: Modelo i3M - DCGAN. Evolución del *loss* según *learning rate*, aplicando optimizador RMSprop. La imagen derecha aporta la misma información que la izquierda pero, en este caso, se han eliminado las primeras 150 *epochs* por cuestiones de escala, pudiéndose así visualizar mejor los valores.

Tabla 7.10: Modelo i3M - DCGAN. Valores FID según *learning rate*.

FID	$L.R = 0.0008$	$L.R = 0.0004$	$L.R = 0.001$
RMSprop	219.420	234.777	276.022

Empleando el FID apreciamos que, utilizando una tasa de aprendizaje de 0.0008, se obtienen los mejores resultados de este último experimento.

Imágenes finales obtenidas con la mejor configuración en el modelo i3M - DCGAN

Teniendo en cuenta todos los valores FID previos, se puede concluir que en este segundo modelo, empleando la configuración *RMSprop* con tasa de abandono 0.4, *learning rate* de 0.0008 y *batch size* 32 se obtienen las imágenes de mama más realistas, tal y como se puede apreciar en la Figura 7.15. En este caso, no sólo obtenemos el contorno, sino que ya se pueden apreciar también más detalles de la mama.

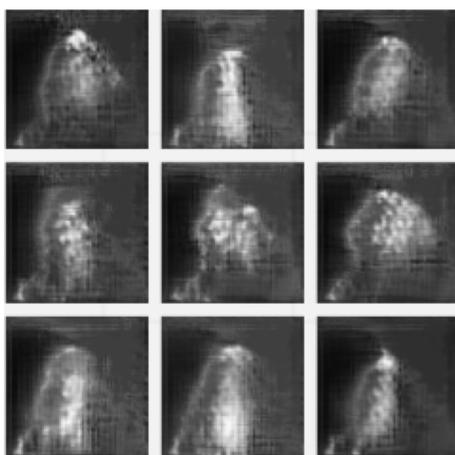


Figura 7.15: Imágenes de la mama obtenidas con la mejor configuración.

7.5.1.3. Modelo i3M - WGAN - GP

Por último, para completar las experimentaciones con las imágenes MAMMI-PET del i3M, nos vamos a centrar en una red que mejora las dos anteriores arquitecturas. Nos estamos refiriendo a la WGAN con penalización de gradiente. Tal y como hemos explicado en el subepígrafe 5.6.3, relativo al marco teórico de este modelo, en esta red se utiliza una nueva función de pérdida, la cual permite estabilizar el proceso de entrenamiento.

En este último modelo únicamente se realizará un solo tipo de experimentación, debido a que los tiempos de ejecución son muy extensos y no se ha dispuesto de suficiente capacidad de cómputo como para abordarla de manera exhaustiva.

Experimento: Batch size y gradient penalty loss weight

En dicho experimento intervienen dos parámetros: tamaño de lote o *batch size* (empleado ya anteriormente), y un nuevo término denominado *gradient penalty loss weight* (λ) que, tal y como se ha explicado en 5.6.3, se encarga de ponderar el término de penalización del gradiente.

En el proceso de entrenamiento propuesto por Gulrajani [62] se suele emplear $\lambda = 10$ como valor predeterminado. En este caso, se han utilizado los valores 0.1, 1 y 10.

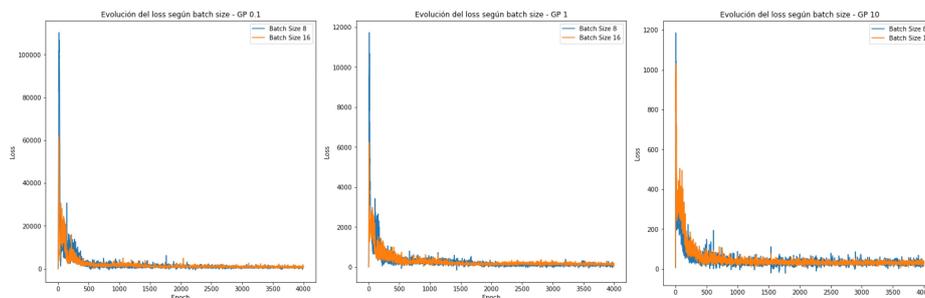


Figura 7.16: Modelo i3M - WGAN-GP. Evolución del *loss* según *batch size*, aplicando valores λ de 0.1, 1 y 10 respectivamente.

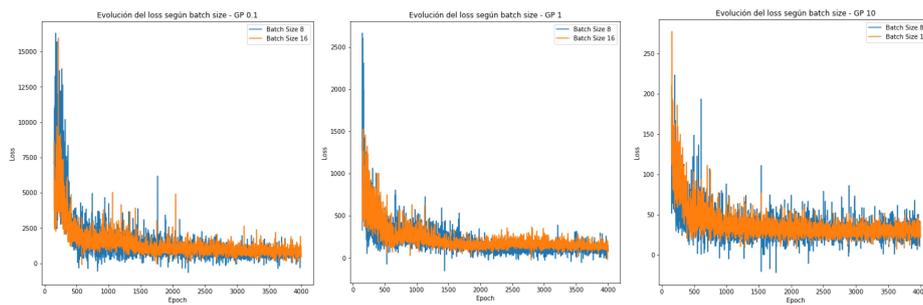


Figura 7.17: Modelo i3M - WGAN-GP. Evolución del *loss* según *batch size*, aplicando valores λ de 0.1, 1 y 10 respectivamente pero donde, por cuestiones de escala, se han eliminado las primeras 150 *epochs*, pudiéndose así visualizar mejor los valores.

En las Figuras 7.16 y 7.17 podemos observar la evolución del error según el tamaño de lote y aplicando el parámetro lambda.

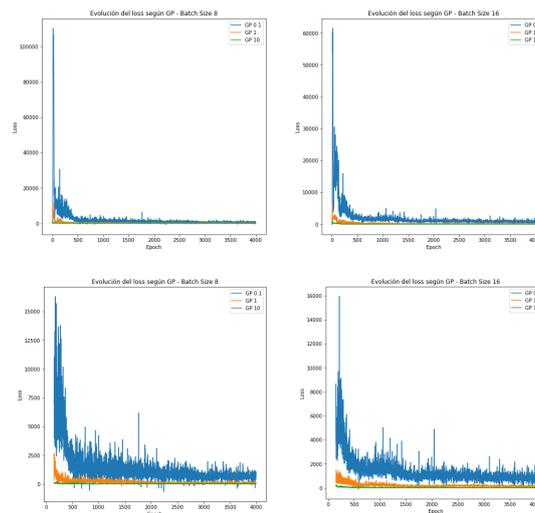


Figura 7.18: Modelo i3M - WGAN-GP. Evolución del *loss* según λ , aplicando tamaños de lote de 8 y 16. Las gráficas inferiores aportan la misma información que las superiores donde, por cuestiones de escala, se han eliminado las primeras 150 *epochs*, pudiéndose así visualizar mejor los valores.

Por el contrario, la Figura 7.18 refleja el rendimiento de ese parámetro λ , aplicando ahora distintos tamaños de lote.

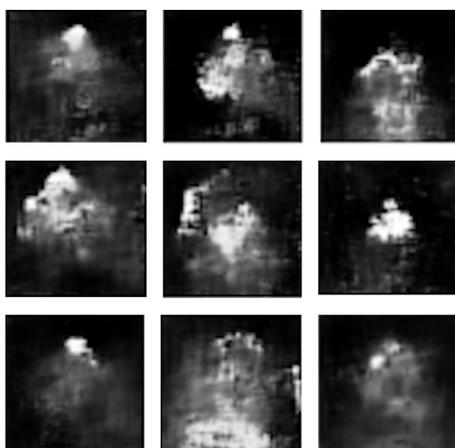
La principal diferencia respecto a las anteriores experimentaciones es que ahora no existe un descenso del error tan suavizado sino que, *epoch* tras *epoch*, se forman distintos picos.

Por otra parte, si nos fijamos en la escala de valores de cada una de las gráficas anteriores vemos que, utilizando un $\lambda = 0.1$ se obtiene un error mucho más elevado. Sin embargo, no está tan claro qué tamaño de lote debería utilizarse ya que el *loss* es muy parecido, tanto utilizando uno de 8 como otro de 16.

Tabla 7.11: Modelo i3M - WGAN-GP. Valores FID según *batch size* y *lambda*.

FID	$\lambda = 0.1$	$\lambda = 1$	$\lambda = 10$
B.S = 8	274.927	252.796	259.189
B.S = 16	256.898	239.813	264.724

La métrica FID, cuyos valores figuran en la Tabla 7.11, nos indica que con el empleo de un *batch size* 16 y un *gradient loss penalty weight* igual a 1, se generan las imágenes más realistas de este experimento, las cuales se pueden apreciar en la Figura 7.19. Se detectan algunos detalles de la mama pero éstos podrían estar mejor definidos. De hecho, las representaciones obtenidas mediante la DCGAN son mejores que éstas, ya que el valor de la métrica FID es menor.

**Figura 7.19:** Modelo i3M - WGAN-GP. Imágenes de la mama obtenidas con la mejor configuración.

7.5.2. Experimentos con muestras USF

7.5.2.1. Modelo USF - GAN

En el subepígrafe anterior hemos podido observar cómo, utilizando una DCGAN, se podía obtener una buena representación de la mama (FID = 219.420, la mejor configuración), aunque ésta podría contener detalles más concretos.

De ahí que, tal y como se ha comentado en el epígrafe 4.2 relativo a la solución planteada, se proponga utilizar ahora otro conjunto de imágenes más extenso (*University of South Florida*). De forma análoga al subepígrafe 7.5.1, se llevarán a cabo una serie de ensayos y se guardarán los pesos resultantes de la mejor configuración obtenida para aplicar posteriormente *transfer learning* a las muestras originales de MAMMI-PET.

Experimento 1: Optimizadores y dropouts

El optimizador *Adam* junto con *dropout* 0.5 era la combinación de parámetros que reportaba un mayor rendimiento en las tomografías PET. En el caso de las mamografías USF el resultado obtenido es el mismo, como puede apreciarse tanto en las Figuras 7.20 y 7.21 como también en la Tabla 7.12.

En principio, aunque con una escala de valores distinta, el descenso del error de cada una de las configuraciones es muy similar al obtenido en los experimentos con las muestras del otro repositorio (i3M). Finalmente, empleando el optimizador *Adam* y *dropout* 0.5, se consigue el menor valor FID.

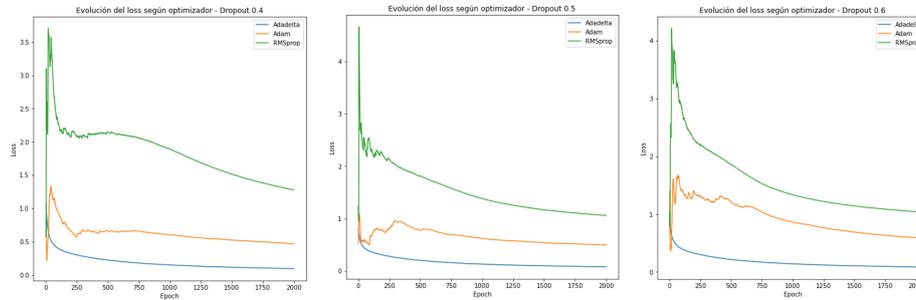


Figura 7.20: Modelo USF - GAN. Evolución del *loss* según optimizador, aplicando tasas de abandono (*dropouts*) 0.4, 0.5 y 0.6.

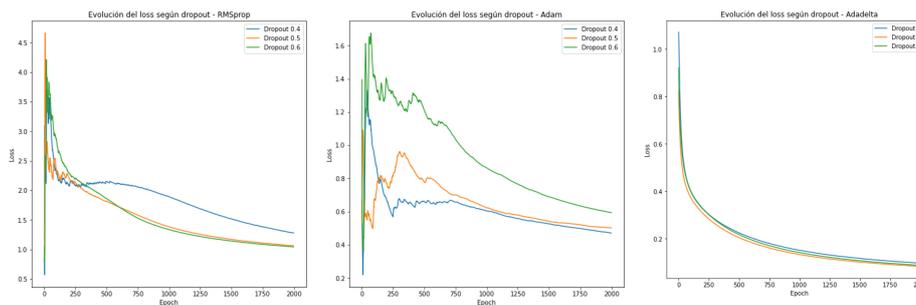


Figura 7.21: Modelo USF - GAN. Evolución del *loss* según *dropout*, aplicando optimizadores RMSprop, Adam y Adadelta.

Tabla 7.12: Modelo USF - GAN. Valores FID según optimizador y *dropout*.

FID	$D = 0.4$	$D = 0.5$	$D = 0.6$
RMSprop	371.248	388.168	374.613
Adam	449.061	363.068	366.376
Adadelta	558.790	534.553	514.508

Experimento 2: Batch size

Basándonos en las gráficas representadas en la Figura 7.22, se observa que en este segundo experimento los resultados han sido también muy similares a los obtenidos en los realizados con las imágenes PET. Se puede apreciar que el error del discriminador más elevado se obtiene con el menor *batch size*. Sin embargo, fijándonos en la tabla de valores FID, sería una cuestión complicada la elección de una solución final, ya que las dos últimas tienen una distancia muy parecida aunque, en este caso, nos quedaríamos con un tamaño de lote de 64 imágenes.

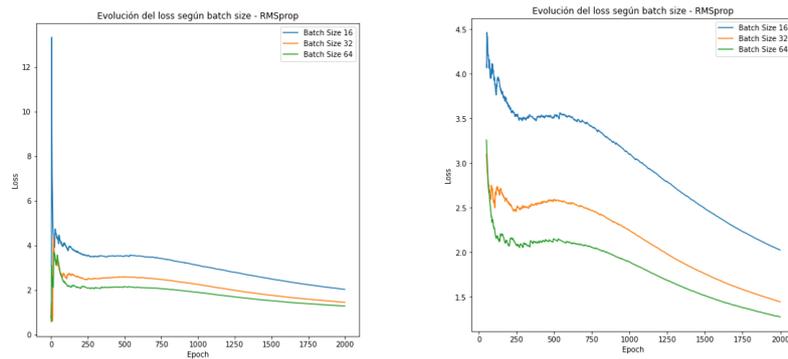


Figura 7.22: Modelo USF - GAN. Evolución del *loss* según *batch size*, aplicando optimizador RMSprop. La imagen de la derecha aporta la misma información que la de la izquierda pero en ésta, por cuestiones de escala, se han eliminado las primeras 50 *epochs*, pudiéndose así visualizar mejor los valores.

Tabla 7.13: Modelo USF - GAN. Valores FID según *batch size*.

FID	<i>B.S</i> = 16	<i>B.S</i> = 32	<i>B.S</i> = 64
RMSprop	400.530	372.429	371.248

Experimento 3: Learning rate

En la Figura 7.23 se puede comprobar que el descenso del error es prácticamente igual al obtenido en el mismo experimento con las imágenes PET. A partir de los datos reflejados en la Tabla 7.14 puede afirmarse que, con el empleo de la tasa de aprendizaje LR más pequeña, se obtiene el mejor resultado en este último experimento.

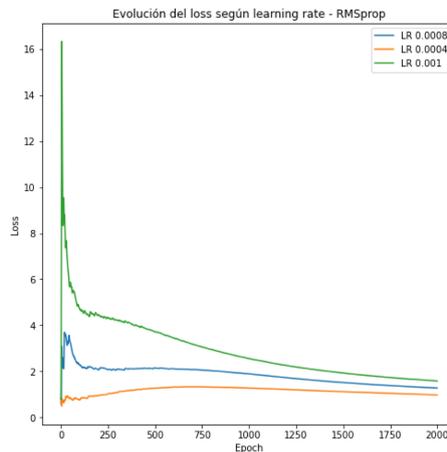


Figura 7.23: Modelo i3M - GAN. Evolución del *loss* según *batch size*, aplicando optimizador RMSprop.

Tabla 7.14: Modelo USF - GAN. Valores FID según *learning rate*.

FID	<i>L.R</i> = 0.0008	<i>L.R</i> = 0.0004	<i>L.R</i> = 0.001
RMSprop	371,248	408.380	411.708

Imágenes obtenidas con la mejor configuración

Teniendo en cuenta todos los resultados anteriores, concluimos que la mejor combinación de parámetros en el modelo GAN es el optimizador *Adam* junto con *dropout* de 0.5 (FID = 363.068), la misma que se obtuvo utilizando los datos clínicos PET, a excepción del tamaño de lote ya que, en este caso, se ha empleado un valor de 64 en la mayoría de los experimentos.

En la Figura 7.24 podemos observar las imágenes de mama generadas por la GAN básica en donde, al igual que en las tomografías PET, solamente se puede apreciar el contorno del seno y no otros detalles más concretos.

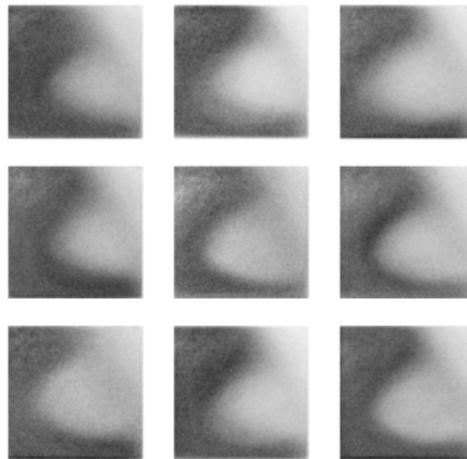


Figura 7.24: Modelo USF - GAN. Imágenes de la mama obtenidas con la mejor configuración.

7.5.2.2. Modelo USF - DCGAN

Experimento 1: Optimizadores y dropouts

Fijándonos en las Figuras 7.25 y 7.26 se puede apreciar que el error mínimo se obtiene utilizando el optimizador *Adam*. En cambio, la tarea de escoger entre una tasa de abandono de 0.4 ó 0.6 es complicada debido a que su *loss* final es 0.0993 y 0.0818, respectivamente. Sin embargo, utilizando FID, finalmente podemos inclinarnos por *Adam* con *dropout* 0.4 como configuración, ya que reporta un mejor resultado.

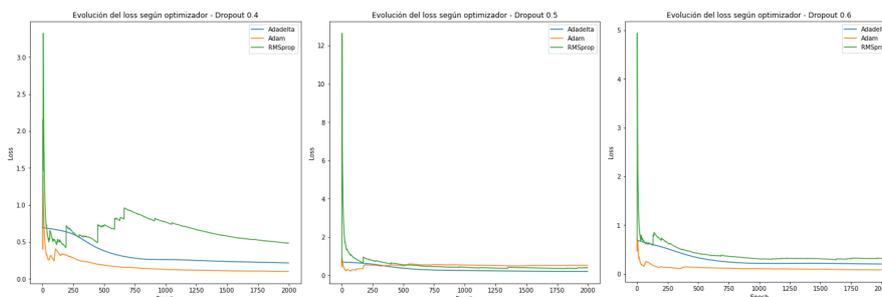


Figura 7.25: Modelo USF - DCGAN. Evolución del *loss* según optimizador, aplicando tasas de abandono (*dropouts*) 0.4, 0.5 y 0.6.

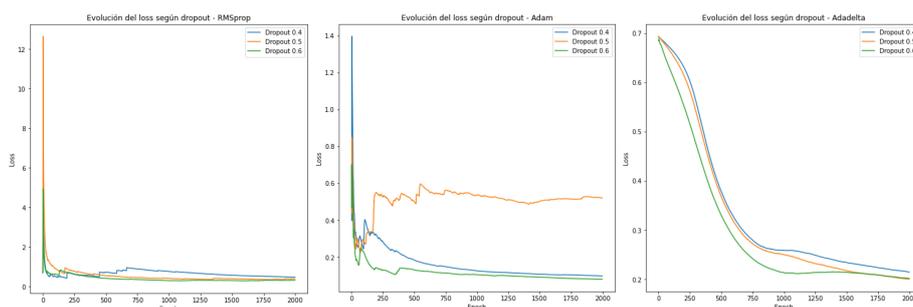


Figura 7.26: Modelo USF - DCGAN. Evolución del *loss* según *dropout*, aplicando optimizadores RMSprop, Adam y Adadelta.

Tabla 7.15: Modelo USF - DCGAN. Valores FID según optimizador y *dropout*.

FID	$D = 0.4$	$D = 0.5$	$D = 0.6$
RMSprop	261.559	344.163	319.923
Adam	243.168	372.165	271.761
Adadelta	374.205	348.309	451.659

Experimento 2: Batch size

En este caso se puede comprobar que los errores en las tres configuraciones son casi idénticos, a pesar de que el tamaño de lote de 64 muestras reporta el mayor error. Por su parte, la métrica FID nos clarifica que la segunda opción (*batch size 32*) es la óptima para este segundo experimento, cuya distancia es muy similar a la que se ha obtenido en el ensayo anterior.

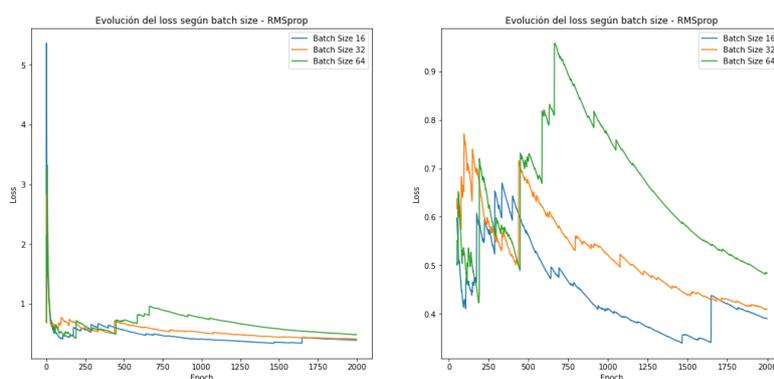


Figura 7.27: Modelo USF - DCGAN. Evolución del *loss* según *batch size*, aplicando optimizador RMSprop. La imagen derecha aporta la misma información que la de la izquierda pero, en este caso, se han eliminado por cuestiones de escala las primeras 50 *epochs*, pudiéndose así visualizar mejor los valores.

Tabla 7.16: Modelo USF - DCGAN. Valores FID según *batch size*.

FID	$B.S = 16$	$B.S = 32$	$B.S = 64$
RMSprop	270.083	245.455	261.559

Experimento 3: Learning rate

Para finalizar los experimentos con la versión convolucional de la GAN observamos que utilizando tasas de aprendizaje de 0.0008 y de 0.0004, se obtiene un error muy similar, en concreto, 0.4840 y 0.2566, respectivamente. Ahora bien, con un valor FID de 232.726 se logra el mejor resultado de toda esta experimentación.

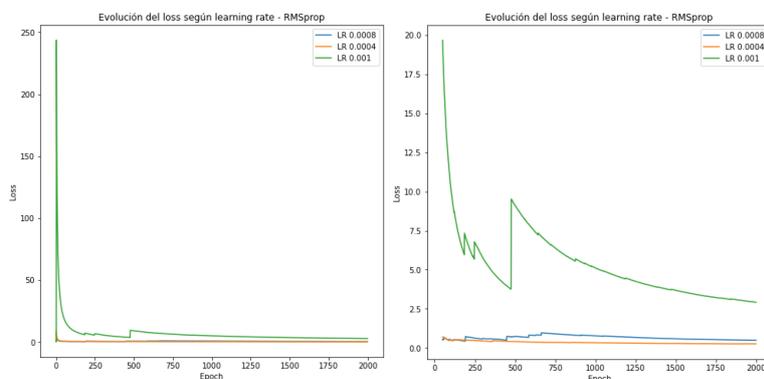


Figura 7.28: Modelo USF - DCGAN. Evolución del *loss* según *batch size*, aplicando optimizador RMSprop. La imagen derecha aporta la misma información que la de la izquierda pero, en este caso, se han eliminado por cuestiones de escala las primeras 50 *epochs*, pudiéndose así visualizar mejor los valores.

Tabla 7.17: Modelo USF - DCGAN. Valores FID según *learning rate*.

FID	$L.R = 0.0008$	$L.R = 0.0004$	$L.R = 0.001$
RMSprop	261.559	232.726	242.926

Imágenes obtenidas con la mejor configuración

Tal y como se ha explicado ya anteriormente, los parámetros óptimos a utilizar son *RMSprop*, *dropout* de 0.4 y *learning rate* de 0.0004 (FID = 232.726). Las imágenes de mama generadas por dicha configuración se pueden apreciar en la Figura 7.29.

Podemos ver que los resultados mejoran sustancialmente respecto a los ensayos anteriores, utilizando una GAN sólo con capas densamente conectadas. La diferencia de distancia FID es muy significativa, siendo una de 363.068 y otra de 232.726, lo que da lugar a que se puedan apreciar mayores detalles de la mama.

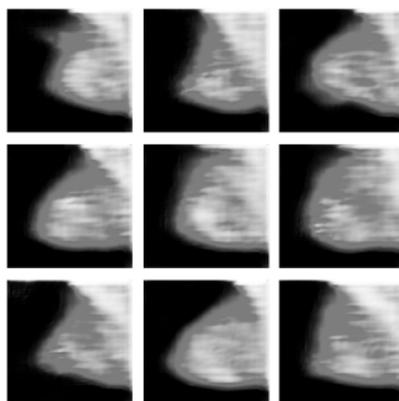


Figura 7.29: Modelo USF - DCGAN. Imágenes de la mama obtenidas con la mejor configuración.

7.5.2.3. Modelo USF - WGAN - GP

Todavía restaría efectuar un experimento más con las muestras USF. Para ello emplearemos la red neuronal Wasserstein GAN. Como ya hemos comentado anteriormente, únicamente se procederá a ejecutar un solo tipo de experimentación con este modelo debido a la escasez de recursos computacionales disponibles. En dicho experimento intervendrán los parámetros *batch size* y λ , siendo éste último el que nos informa del peso que se le otorga a la penalización de gradiente correspondiente.

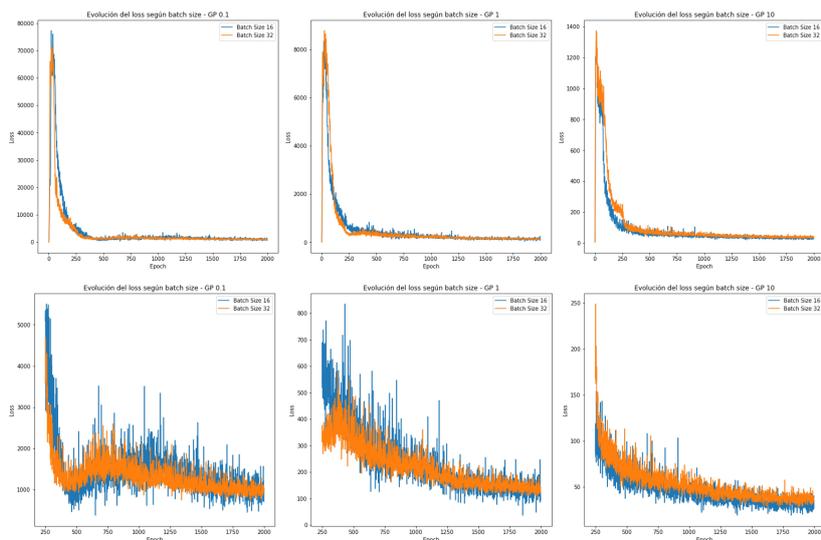


Figura 7.30: Modelo USF - WGAN-GP. Evolución del *loss* según *batch size*, aplicando valores λ de 0.1, 1 y 10. Las gráficas inferiores aportan la misma información que las superiores pero, en este caso, se han eliminado por cuestiones de escala las primeras 250 *epochs*, pudiéndose así visualizar mejor los valores.

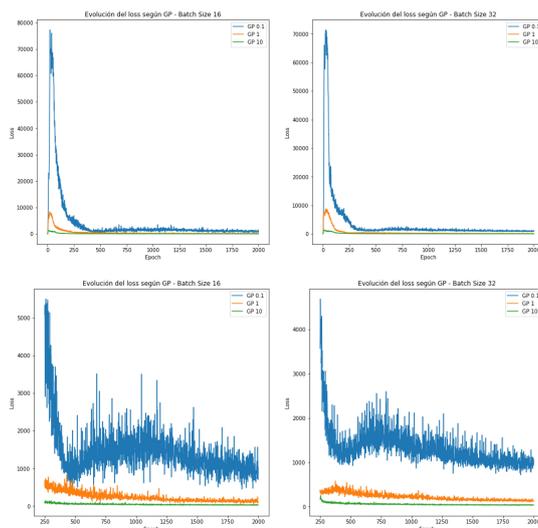


Figura 7.31: Modelo USF - WGAN-GP. Evolución del *loss* según λ , aplicando tamaños de lote de 16 y 32. Las gráficas inferiores aportan la misma información que las superiores pero, en este caso, se han eliminado por cuestiones de escala las primeras 150 *epochs*, pudiéndose así visualizar mejor los valores.

Analizando las Figuras 7.30 y 7.31, se puede apreciar la evolución del error en función del tamaño de lote y el parámetro λ . Se observa el mismo fenómeno que se producía en el ensayo realizado con las tomografías PET, es decir, el descenso del *loss* no es suavizado sino que existen distintos picos *epoch* tras *epoch*.

De igual manera que en dicho experimento con muestras PET, el error más elevado se obtiene para $\lambda = 1$. Por otra parte, la decisión de qué tamaño de lote escoger resulta difícil de tomar, ya que sus respectivos errores son muy parejos. Mediante el uso de la métrica FID, cuyos valores figuran en la Tabla 7.18, se obtiene el mismo resultado que en el caso de empleo de imágenes PET. Dicho resultado consiste en escoger un *batch size* de 16 y un λ unitario.

Tabla 7.18: Modelo USF - WGAN-GP. Valores FID según *batch size* y *lambda*.

FID	$\lambda = 0.1$	$\lambda = 1$	$\lambda = 10$
B.S = 16	233.590	223.749	230.097
B.S = 32	251.335	227.034	227.619

Las imágenes creadas por esta red WGAN-GP se pueden examinar en la Figura 7.32. Son las muestras más realistas creadas con este repositorio, y en ellas se puede observar correctamente la anatomía de la mama.

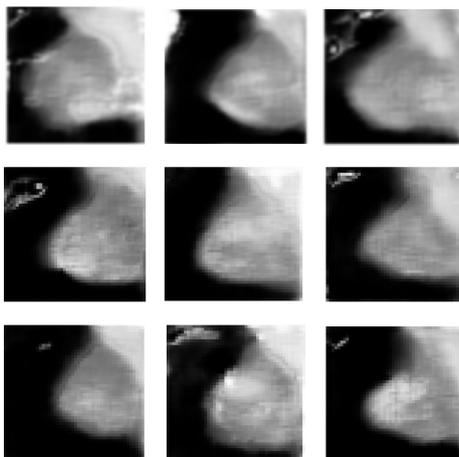


Figura 7.32: Modelo USF - WGAN-GP. Imágenes de la mama obtenidas con la mejor configuración.

7.5.3. Transfer learning aplicado a las muestras i3M

De toda la experimentación realizada anteriormente, tanto con imágenes PET como con mamografías, podemos concluir que los dos modelos que han tenido un mejor rendimiento han sido los siguientes:

- En las muestras del i3M ha resultado ser el modelo propuesto por la red DCGAN con optimizador RMSprop, *dropout* con valor 0.4, *learning rate* de 0.0008 y *batch size* de 32.
- Con las imágenes de la USF, el modelo que ha funcionado mejor ha sido el planteado por la red WGAN-GP, con un tamaño de lote de 16 y una λ de 1.

La exposición de toda la serie de experimentos anteriores, realizados con muestras de mamografías, no debe hacernos olvidar que el objetivo principal de nuestro proyecto continúa siendo el de obtener resultados tangibles en el empleo de imágenes de tomografías PET. En este contexto, tal como se propuso en el epígrafe 4.2, la aplicación de la técnica *transfer learning* a nuestras imágenes PET puede suponer una gran mejora en la generación de nuevas muestras. Esta aplicación se concretaría en la implementación, en el repositorio i3M, de los pesos de aprendizaje de la mejor configuración obtenida en los experimentos con mamografías de la USF, es decir, un tamaño de lote=16 y una $\lambda = 1$.

Los resultados finales, obtenidos a través de la aplicación de este último modelo, se pueden apreciar en las próximas tres figuras. En la Figura 7.33 se representan las tomografías PET originales del i3M, punto de partida de toda la experimentación.

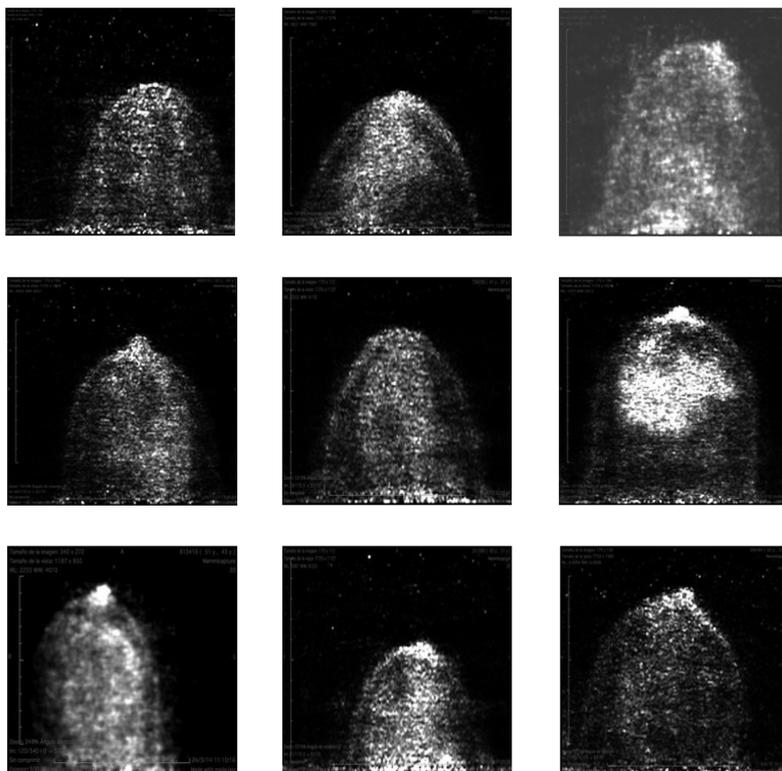


Figura 7.33: Imágenes de tomografías PET originales del i3M.

En la Figura 7.34 se pueden observar las imágenes PET creadas con el modelo que ha proporcionado el mejor rendimiento, red DCGAN, y con las muestras del *dataset* del i3M.

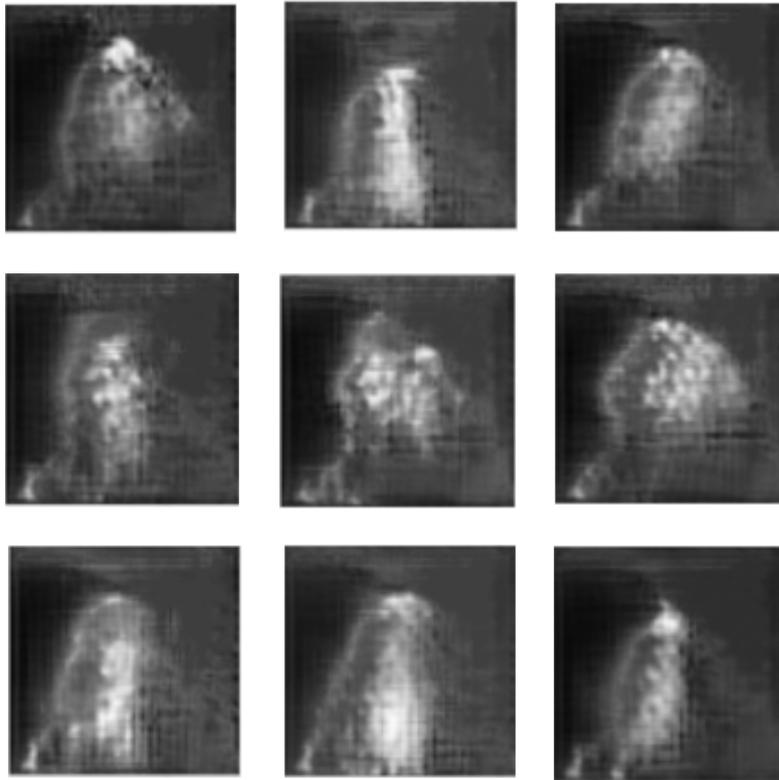


Figura 7.34: Imágenes PET generadas por el modelo DCGAN con RMSprop, tasa de abandono de 0.4 y *learning rate* de 0.0008.

Finalmente, en la Figura 7.35 se aprecian las muestras PET generadas después de haber aplicado el método de *transfer learning*, con los pesos de aprendizaje obtenidos en la mejor configuración resultante de la red WGAN-GP y con las muestras de mamografías de la USF.

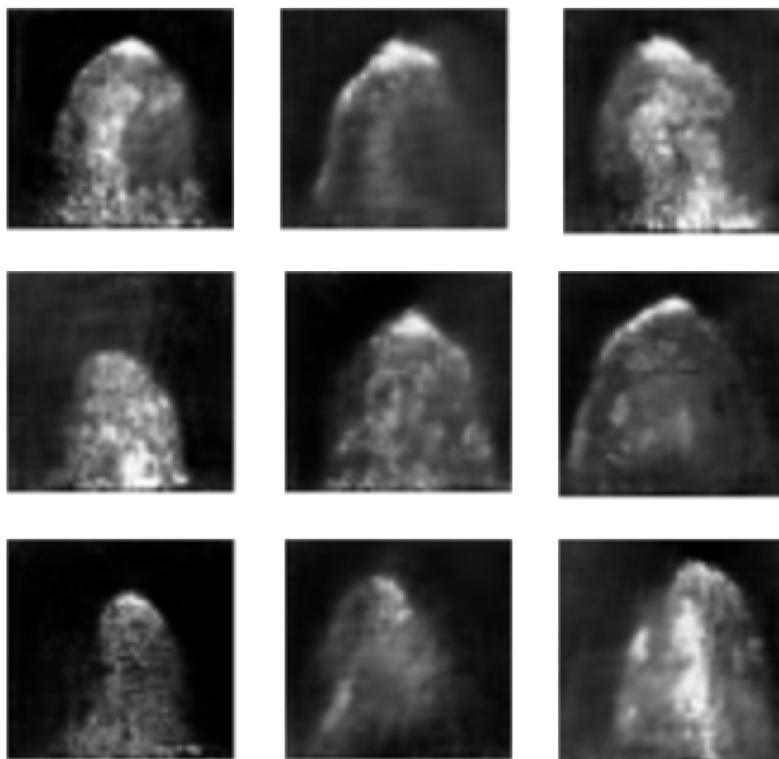


Figura 7.35: Tomografías PET generadas después de la aplicación del método de *transfer learning* mediante los pesos de aprendizaje correspondientes a la mejor configuración obtenida con mamografías de la USF.

Asimismo, mediante la aplicación del método de *transfer learning*, se ha conseguido un valor FID de 183.973, suponiendo esto una mejora respecto a los resultados obtenidos hasta el momento. Vemos, por tanto, que los modelos GAN son capaces de generar de forma satisfactoria imágenes clínicas que se asemejan lo suficiente a las originales.

Para finalizar podría afirmarse que, en general, el hecho de utilizar en una misma investigación imágenes pertenecientes a modalidades diferentes, puede suponer una alternativa válida para resolver el problema de la escasez de muestras y, así, poder llegar a entrenar satisfactoriamente modelos basados en redes neuronales generativas adversarias con un adecuado número de *inputs*. Se puede concluir pues que, las mejoras logradas a través de *transfer learning*, abren en un futuro la puerta al estudio de técnicas más robustas para el empleo de muestras con diferentes características intrínsecas en una misma investigación. Una de ellas sería la conocida como *cross modality*, ya mencionada en el epígrafe relativo a la solución propuesta en el presente proyecto.

CAPÍTULO 8

Conclusiones

En este capítulo, a modo de síntesis, se intentarán exponer detalladamente las principales conclusiones a las cuales se ha podido llegar después la preparación y desarrollo del presente Trabajo de Fin de Grado (TFG).

Estas conclusiones permitirán precisar posteriormente, en el epígrafe 8.1, todos aquellos conocimientos y técnicas que se han tratado en el proyecto y que podrían llegar a transmitirse en forma de legado a todas aquellas personas interesadas en el desarrollo de soluciones para problemáticas similares a las tratadas en esta memoria.

Finalmente, en el epígrafe 8.2, concluiremos el capítulo haciendo una profunda revisión acerca de las asignaturas y competencias transversales cuyo dominio ha resultado fundamental, a nuestro juicio, para poder llevar a cabo este TFG.

Como bien sabemos, el objetivo principal de este trabajo era poder llegar a crear una base de datos simulada y realista de tomografías de mama por emisión de positrones que sirviera de ayuda en un futuro a radiólogos y oncólogos en la reconstrucción de imágenes PET para la detección precoz de lesiones neoplásicas.

Para ello, se consideró oportuno concretar este propósito en tres objetivos bien diferenciados:

- Estudio y selección de las arquitecturas que nos pudieran proporcionar una generación de muestras sintéticas realistas y de calidad. Para ello nos centramos en la redes neuronales generativas adversarias (GAN).
- Experimentación detallada con cada uno de los modelos analizados para obtener los resultados deseados.
- Conseguir aumentar la calidad de las muestras obtenidas mediante una nueva experimentación y la aplicación de técnicas de *transfer learning*.

El primer objetivo descrito anteriormente se pudo alcanzar después de haber realizado un análisis exhaustivo de los modelos generativos existentes actualmente y que consideramos que eran candidatos para adaptarse a nuestros fines. Para ello, como en todo proyecto de ciencia de datos, se empezó a trabajar con una arquitectura sencilla para, a partir de ésta, construir otras más robustas que permitieran mejorar los resultados obtenidos. Este objetivo se pudo alcanzar no sin antes abordar el estudio de la base matemática en la que se fundamentaba cada uno de estos modelos.

En el apartado de experimentación con muestras del i3M, las tres redes generativas utilizadas, detalladas en el capítulo 5, fueron proporcionando diferentes resultados. Al principio solamente se pudieron obtener resultados en donde sólo era apreciable el contorno de la mama (GAN básica con capas densamente conectadas). No obstante, añadiendo capas convolucionales a nuestro modelo de experimentación (DCGAN), se pudieron llegar a extraer muestras con un detalle más preciso.

Si bien con el modelo DCGAN el rendimiento obtenido se podía considerar plausiblemente correcto, aparecieron ciertos problemas de convergencia durante la última fase de entrenamiento del modelo. Esta eventualidad hizo que, a nuestro juicio, se considerara oportuno llevar a cabo una tercera fase de experimentación con un nuevo modelo. Para ello se propuso una nueva red, la WGAN con penalización de gradiente, cuyo proceso de entrenamiento propició una mayor estabilidad. En términos globales, los resultados obtenidos fueron lo suficientemente prometedores para poder aseverar que el segundo de nuestros objetivos se había alcanzado también de manera satisfactoria.

A pesar de ello, se consideró que aún se podía llegar un poco más allá y obtener una mejor calidad en las muestras sintéticas generadas. A tal efecto se contempló volver a realizar toda la experimentación anterior pero esta vez con muestras de mamografías procedentes del repositorio de libre acceso facilitado por la *University of South Florida* (USF). Esto nos proporcionó un conjunto de datos de mayor volumen que, lógicamente, permitiría una mejor generalización de los modelos durante el proceso de entrenamiento.

Las características de las muestras de este nuevo repositorio (*screening mammography*) eran diferentes a las de las tomografías PET iniciales del i3M, lo cual llegó a plantear el dilema de si los buenos resultados obtenidos con ellas, en el nuevo proceso de experimentación, podrían llegar a quedar distorsionados debido a la heterogeneidad de las muestras utilizadas.

Sin embargo, la técnica del *transfer learning* permitió exportar de manera satisfactoria los resultados obtenidos con muestras de la USF a los del i3M. Este proceso dio como resultado la generación de unas muestras sintéticas de una calidad superior, lo cual permite confirmar el cumplimiento de nuestro último objetivo propuesto.

No obstante, como se comentará posteriormente en el capítulo 9, queda abierta la puerta al desarrollo de nuevas experimentaciones con imágenes clínicas de naturaleza heterogénea mediante el empleo de técnicas más robustas especialmente diseñadas para ello. El acceso a esta nueva investigación ha de venir, sin lugar a dudas, de la mano del estudio y aplicación de técnicas de síntesis o *cross modality*.

Es evidente que todo este proceso de investigación y experimentación no hubiera podido llevarse a cabo sin un dominio de las diversas tecnologías que permitían dar solución a los retos planteados. Entre ellas cabe destacar el uso de *Keras* y *Tensorflow* como herramientas imprescindibles para desarrollar el diseño de las arquitecturas GAN propuestas.

El conocimiento de ambas librerías ya era conocido antes de iniciar el TFG, pero su empleo continuo y sistemático a lo largo de todo el proyecto ha permitido mejorar sustancialmente el conocimiento previo que se tenía sobre ellas. Concretamente nos estamos refiriendo al aprendizaje de nuevas funcionalidades que han aportado estas librerías, como la creación de las distintas capas que conforman estas redes generativas, de las funciones de activación, el cálculo de gradientes (*keras.backend.gradients*) y otros conceptos como el de *batch normalization*, *dropout*, etc.

El desarrollo de arquitecturas de vanguardia, como las redes neuronales generativas adversarias, no ha sido objeto de estudio a lo largo del Grado en Ciencia de Datos, por lo que ha sido necesario un esfuerzo adicional para poderlas llegar a entender y aplicar en nuestra experimentación. Pero, sin la capacitación conseguida en el Grado sobre el concepto de redes convolucionales, hubiera sido realmente difícil haber podido llegar a buen puerto.

La realización del presente proyecto ha abierto el camino al autor de esta memoria hacia un nuevo campo de investigación tan amplio y en constante evolución como el que representan los modelos generativos. Éstos, de bien seguro, tienen un gran potencial, no sólo en medicina, sino también en otros muchos sectores. Según Ian Goodfellow, creador de las GAN, éstas podrían usarse “para diseñar chips más rápidos, autos más eficientes, edificios que sean más resistentes a los terremotos, o que sean más económicos de construir. Toda esta tecnología es aún incipiente”.

Para finalizar, es justo reconocer que una de las mayores dificultades encontradas a lo largo de la elaboración de este TFG, ha sido la comprensión de toda la base matemática que subyace bajo estas arquitecturas GAN, en concreto algunos conceptos empleados como el de la dualidad de *Kantorovich-Rubinstein* en las redes WGAN – GP.

8.1 Legado

A lo largo de todo este Trabajo de Fin de Grado ya se ha apuntado en diversas ocasiones que la principal aportación que pretende hacer la investigación llevada a cabo, es la generación de una base de datos sintéticos que permita entrenar modelos de inteligencia artificial capaces de detectar tumores de mama. Obviamente, los beneficiarios últimos de los resultados obtenidos han de ser los profesionales de la medicina que se dedican a dicha labor, concretamente radiólogos y oncólogos. El desarrollo de estos nuevos modelos de entrenamiento han de favorecer progresivamente la aparición de métodos más precisos en la detección precoz del cáncer de mama, generando así unos menores costes de tiempo y de recursos.

La elaboración de este proyecto ha llevado también consigo la creación de un repositorio de Github con todos aquellos elementos utilizados en el desarrollo del trabajo. Este conjunto de archivos se ha creado con la intención adicional de que, todas aquellas personas interesadas en este tipo de investigación, puedan replicar la experimentación llevada a cabo para obtener sus propios resultados o profundizar en su estudio si así lo desean.

En este repositorio ¹ se pueden encontrar los siguientes elementos:

- Notebook *.ipynb con el procesamiento previo llevado a cabo.
- Notebook *.ipynb con la creación de las distintas arquitecturas propuestas.
- Notebook *.ipynb con las gráficas de evolución del *loss* en cada uno de los experimentos.
- Notebook *.ipynb con el cálculo de la métrica FID para cada uno de los experimentos.
- Imágenes de tomografías PET obtenidas mediante las distintas arquitecturas propuestas, además de los resultados conseguidos con la técnica de *transfer learning*.

¹https://github.com/miquelmarin43/test_tfg_repo

El empleo de todos estos elementos permitirá, a todo aquel que lo desee, continuar experimentando con las arquitecturas propuestas o tenerlas como punto de referencia para futuros desarrollos.

Debido a la rápida evolución que han tenido las redes generativas adversarias desde 2014, el autor de esta memoria es consciente de que los detalles compartidos en este repositorio pueden quedar obsoletos en un futuro próximo. Sin embargo, esperamos que puedan servir de guía para las personas que estén interesadas en la búsqueda de métodos de mejora del rendimiento de estas redes, disponiendo así de un punto de partida.

8.2 Relación del trabajo desarrollado con los estudios cursados

Durante estos últimos cuatro años, la estructura del plan de estudios del Grado en Ciencia de Datos de la UPV nos ha brindado la oportunidad de disfrutar de un aprendizaje activo, continuo y práctico, proporcionando una visión multidisciplinar de la ciencia de datos. Sin duda, el carácter de la formación recibida ha dotado, a los estudiantes de esta carrera, de las habilidades de ingeniería necesarias y del suficiente razonamiento crítico para poder gestionar problemas complejos reales.

Gracias a dicha preparación, han sido diversos los conocimientos y competencias alcanzados durante esta etapa, los cuales han permitido disfrutar de la capacitación necesaria para elaborar el presente Trabajo de Fin de Grado (TFG).

No resulta tarea sencilla sintetizar en pocas líneas solamente algunas de las técnicas y métodos utilizados en este proyecto ya que, en mayor o menor medida, todo el acervo de conocimientos adquiridos ha contribuido a su realización. No obstante, estructurando el empleo de dichas tecnologías en cuatro grandes áreas, se han podido llegar a concretar aquellas disciplinas más utilizadas en la confección de esta memoria.

En primer lugar, es importante destacar que el estudio de las redes neuronales artificiales ha supuesto un aspecto crucial en la elaboración del TFG. Fue necesario entender perfectamente el proceso previo de desarrollo y entrenamiento de las redes propuestas en este trabajo, para poder estructurar sobre él todo el análisis posterior.

Para esta tarea resultaron de inestimable ayuda todos los conocimientos aprendidos en asignaturas como *técnicas escalables en aprendizaje automático* y *análisis de imágenes y vídeos*, sin descartar tampoco todos los fundamentos proporcionados en *modelos descriptivos y predictivos*.

Como habrá podido observarse a lo largo del trabajo, el tratamiento de las redes neuronales adversarias (GAN) ha resultado ser uno de los principales ejes sobre el que ha basculado el mismo. Aunque este tipo concreto de redes no haya sido objeto de estudio específico durante la carrera, sí que ha resultado de gran utilidad el conocer previamente las bases de las redes neuronales convolucionales, esenciales para la creación de las arquitecturas GAN. Las asignaturas mencionadas anteriormente han cubierto bien la misión de proporcionar una correcta comprensión de dichos principios.

Por otra parte, es indudable que no se hubiera logrado ninguno de los objetivos propuestos sin una oportuna puesta en práctica de las tecnologías utilizadas en el proyecto. En este sentido, además de las materias detalladas anteriormente, ha resultado indispensable una adecuada aplicación de todos los conocimientos adquiridos en disciplinas como *programación y algorítmica*. No quisiéramos olvidarnos tampoco de una asignatura

de gran utilidad en la creación de las diversas librerías de Python utilizadas en el TFG. Nos referimos, obviamente, a la asignatura de *visualización*.

En otro orden de cosas, pero al mismo nivel de importancia que el punto anterior, cabe destacar también, en las tareas de procesamiento de las imágenes utilizadas en el trabajo, la relevante contribución de los conocimientos aprendidos en la asignatura de *análisis de imágenes y videos*, de gran utilidad en el empleo de las librerías OpenCV y Pillow.

Finalmente, es necesario destacar en especial todas las *asignaturas de cariz matemático-estadístico* estudiadas durante el Grado, ya que han proporcionado una excelente base conceptual para comprender los diferentes desarrollos teóricos que aparecen en el transcurso de la exposición. No en vano, la palabra "matemática" deriva del griego *máthēma*, "conocimiento".

Este conjunto de materias han permitido alcanzar una de las más importantes competencias transversales propuestas al inicio de la carrera. Estamos haciendo alusión, obviamente, al *análisis y resolución de problemas*. Parece plausible pues, no poner en tela de juicio el hecho de que, disponer de la capacidad de analizar y resolver problemas de manera eficiente, sea una de las competencias más significativas a conseguir al término de unos estudios universitarios y, más aún, si éstos son de índole científico-tecnológica. Consideramos que este trabajo ha supuesto un reto en lo relativo a la identificación, planteamiento y propuesta de resolución de un problema real y ciertamente complejo. Esto hubiera sido difícil de abordar sin contar previamente con dicha competencia.

En la misma línea, tal como se ha comentado al principio de este epígrafe, el carácter activo y práctico de la formación recibida permitió disponer, desde un buen principio, de una precisa metodología que pudo utilizarse a lo largo de todo el TFG, aplicando así los conocimientos teóricos recibidos. Gran parte de las disciplinas detalladas anteriormente facilitaron el acceso a la competencia de *aplicación y pensamiento práctico*. Esto permitió llevar a cabo un diseño funcional de la planificación concreta del procedimiento a adoptar, para poder alcanzar así el objetivo propuesto en el proyecto. Una meta sin un plan es sólo un deseo.

Si bien la planificación era fundamental, lógicamente, era necesario también disponer del tiempo suficiente para ejecutarla. Siendo el plazo disponible para realizar nuestro trabajo un valor dado e inamovible, la *planificación y gestión del tiempo* de manera apropiada se convertía en una condición imprescindible para llegar a buen puerto. El dominio de esta competencia transversal ayudó a gestionar y programar las actuaciones necesarias para lograr los retos definidos dentro del plazo marcado.

Como hemos explicado anteriormente, el planteamiento de este TFG requería el estudio de cuestiones no abordadas a lo largo de la carrera. Esto significaba todo un desafío, ya que no se disponía aún de los conocimientos técnicos específicos sobre dichas cuestiones. A pesar de ello, el haber podido alcanzar durante el Grado la competencia en *aprendizaje permanente*, habilitó el empleo de la formación recibida, no sólo en la adquisición y asimilación de los nuevos conocimientos requeridos, sino también en su consecución de una manera completamente autónoma.

Evidentemente este proceso de adquisición de nuevos conocimientos precisaba de un esfuerzo por llegar a descubrir los fundamentos en los cuales estaban basados todos ellos. Es decir, era necesario disponer de un sutil *razonamiento crítico* para poder discernir cuáles, de todos los novedosos conceptos descubiertos, eran los que precisaba nuestro proyecto. Esto sólo se podía conseguir si previamente existía una clara voluntad por co-

nocer el propósito para el cual habían sido desarrollados cada uno de ellos y valorar después si estos propósitos se adaptaban a nuestros intereses.

Para concluir, no quisiéramos dejar de mencionar algunas competencias transversales adicionales adquiridas durante la carrera, las cuales también han tenido su importancia en la elaboración de este TFG. Así, el *conocimiento de problemas contemporáneos* ha permitido abordar un tema tan actual, y en constante desarrollo, como el procesamiento de imágenes médicas mediante el uso de inteligencia artificial. Finalmente, como puede comprobarse en el epígrafe dedicado al marco ético y legal, en el diseño del trabajo se tuvo especial cuidado en que éste se englobara en un contexto que respetase la *responsabilidad ética y profesional* que se requiere en cualquier proyecto de investigación.

CAPÍTULO 9

Desarrollos futuros

A partir de las conclusiones expuestas en el capítulo 8, se pueden derivar algunas líneas de desarrollo que podrían impulsarse en una futura investigación en el campo de las redes neuronales generativas adversarias (GAN).

Es obvio que, para la realización de un Trabajo de Fin de Grado, el tiempo disponible para la materialización del mismo constituye un valor dado fijo e invariable. Por tal motivo, en la planificación del proyecto, se decidió limitar el proceso de experimentación a tres de estas arquitecturas GAN.

No obstante, existe una gran variabilidad de tipos de redes GAN que merecen ser estudiadas con detenimiento y profundizar en su desarrollo para poder obtener así cada vez mejores resultados en el objetivo fijado de la generación de imágenes sintéticas de alta resolución que permitan apreciar mejor la anatomía de la mama.

En esta línea, dos posibles propuestas de investigación podrían venir definidas por el estudio y desarrollo de las denominadas *Progressive growing GAN* (PGGAN) y las *Super-resolution GAN* (SRGAN). Tal propósito exigiría la utilización de tarjetas gráficas adecuadas que permitan reducir los elevados tiempos de computación que requieren dichas arquitecturas.

Otra vía de actividad podría abrirse con la ampliación del número de parámetros a utilizar en la experimentación y que, en nuestro proyecto, se han tenido que descartar por razones de tiempo. Estaríamos hablando de los siguientes:

- Empleo de otras técnicas de normalización de capas, tanto en el generador como en el discriminador, como *InstanceNorm*, *GroupNorm*, etc.
- Utilización de otras funciones de activación, como *ELU*, *SELU* y *GELU* (similares a *ReLU*).
- En el caso de optimizadores, aplicación de variantes de *RMSprop*, como por ejemplo *SC-RMSprop*.
- Uso de un mayor número de tasas de aprendizaje (*learning rate*) y abandono (*dropouts*).
- Empleo de más valores λ en el entrenamiento de las WGAN – GP.
- Consideración de la posibilidad de aplicar *data augmentation* como paso previo al entrenamiento de las redes.

Por otra parte, tal y como se ha mencionado en el epígrafe 4.2, la ampliación de la investigación en la aplicación de técnicas como *cross modality* mejoraría muchísimo la calidad de las imágenes sintéticas generadas por los modelos GAN. Esto posibilitaría el empleo, como en nuestro caso, de datos con características heterogéneas para suplir posibles carencias en el número de muestras disponibles para el entrenamiento de los modelos generativos propuestos. Queda mucho trabajo futuro por hacer, ya que se tiene que estudiar en qué condiciones exactas se puede aplicar el *cross modality* y qué tipo de imágenes se pueden incluir en este tipo de entrenamiento.

Consideramos que una tarea que debería tenerse muy en cuenta en el futuro sería la investigación en los modelos conocidos como *diffusion models* o modelos de difusión. En los últimos años estos modelos están cobrando una importancia cada vez mayor en el campo de la generación de imágenes sintéticas, ya que son capaces de conseguir unos mejores resultados FID y una mayor estabilidad que los proporcionados por las redes neuronales disponibles en el actual estado del arte [66].

Bibliografía

- [1] European Commission. *2020 Cancer incidence and mortality in EU-27 countries* [en línea]. Bruselas, 2020 [consulta: 12/05/2022]. Disponible en https://joint-research-centre.ec.europa.eu/jrc-news/2020-cancer-incidence-and-mortality-eu-27-countries-2020-07-22_en
- [2] Eurostat. *World Cancer Day: 1 in 4 deaths caused by cancer* [en línea]. Luxemburgo, 2022 [consulta: 12/05/2022]. Disponible en <https://ec.europa.eu/eurostat/en/web/products-eurostat-news/-/edn-20200204-1>
- [3] European Commission. *Europe's Beating Cancer Plan. Flagship 2* [en línea]. Bruselas, 2021 [consulta: 12/05/2022]. Disponible en https://ec.europa.eu/health/system/files/2022-02/eu_cancer-plan_en_0.pdf
- [4] BRANT, William E.; HELMS, Clyde A. Diagnostic Imaging Methods. *Fundamentals of Diagnostic Radiology*. 4th ed. Lippincott Williams & Wilkins, Philadelphia, Pensilvania, EE.UU., 2012, pp. 2-25. ISBN 9781608319114.
- [5] World Health Organization. *Ionizing radiation, health effects and protective measures* [en línea]. 29 April 2016. Last reviewed March 2022 [consulta: 04/05/2022]. Disponible en <https://www.who.int/news-room/fact-sheets/detail/ionizing-radiation-health-effects-and-protective-measures>
- [6] CARROLL, Quinn B. An Introduction to Radiographic Science. *Radiography in the Digital Age*. 1st ed. Charles C Thomas, Springfield, Illinois, EE.UU., 2011, pp. 3-15. ISBN 9780398086466
- [7] Mayo Clinic. *CT Scan* [en línea]. Archived from the original on 15 October 2016. Retrieved 20 October 2016, © Mayo Foundation for Medical Education and Research [consulta: 04/05/2022]. Disponible en <https://www.mayoclinic.org/tests-procedures/ct-scan/about/pac-20393675>.
- [8] LÓPEZ-MORANCHEL, Ignacio; SÁNCHEZ-GONZÁLEZ, Lidia. Principios físicos de la resonancia magnética. *Fundamentos de la técnica de imagen por resonancia magnética*. Síntesis, Madrid, España, 2020, pp. 33-45. ISBN 9788491714330.
- [9] Prinz. *Radiografía: ¿Qué es y para qué sirve?* [en línea]. Redactado por Dr. Raúl Espino y revisado por Dr. Sergio Robles [consulta: 04/05/2022]. Disponible en <https://prix.com/salud/radiografia-que-es-y-para-que-sirve/>

- [10] Clínica Pedrosa. Mastopatía fibroquística. *Blog de Ginecología y Obstetricia* [en línea]. 2018 [consulta: 04/05/2022]. Disponible en <https://cpedrosa.com/mastopatia-fibroquistica/>
- [11] BROOKSHIRE, Bethany. *Scientists Say: CT scan* [en línea]. Science News for Students, Washington, D.C, EE.UU, 2018 [consulta: 04/05/2022] Disponible en <https://www.sciencenewsforstudents.org/article/scientists-say-ct-scan>
- [12] NEGOITIA-NEGOITIA, Adina; CAPILLA-AMPUDIA, Jesús María; GAMOVILLEGAS, Paula [et al.]. *Valoración con Resonancia Magnética 1,5 y 3T de rodilla postquirúrgica*. Sociedad Española de Radiología Médica (SERAM), Madrid, España, 2018 [consulta: 04/05/2022]. Disponible en <https://piper.espacio-seram.com/index.php/seram/article/view/2793/1363>
- [13] Vegaspect Nükleer Tip Merkezi. *SPECT* [en línea]. Esmirna, Turquía. 2021 [consulta: 04/05/2022]. Disponible en <https://vegaspect.com/spect/>
- [14] Radiological Society of North America, Inc. (RSNA). *PET / CT* [en línea]. Oak Brook, Illinois, EE.UU., 2022, © RadiologyInfo.org. [consulta: 04/05/2022]. Disponible en <https://www.radiologyinfo.org/es/gallery/image/233>
- [15] Radiological Society of North America, Inc. (RSNA). *General Nuclear Medicine* [en línea]. © RadiologyInfo.org., Oak Brook, Illinois, EE.UU., 2010, Reviewed 04-20-2018 [consulta: 05/05/2022]. Disponible en <https://www.mcrads.com/pdfs/genuclear.pdf>.
- [16] NÚÑEZ, Margarita. *Fundamentos de SPECT* [en línea]. Escuela Universitaria de Tecnología Médica UdelaR. Comité de Tecnólogos de ALASBIMN (Asociación Latinoamericana de Sociedades de Biología y Medicina Nuclear), Montevideo, Uruguay, 2008 [consulta: 05/05/2022]. Disponible en http://www.alasbimn.net/comites/tecnologos/material/Fundamentos_SPECT.pdf
- [17] Clínica Universidad de Navarra. *Diagnóstico por radiofármacos* [en línea]. © CUN, Pamplona, España, 2022 [consulta: 14/05/2022]. Disponible en <https://www.cun.es/enfermedades-tratamientos/pruebas-diagnosticas/diagnostico-radiofarmacos>
- [18] PISANO-CASTRO, María. *Radiofármacos PET en España. Indicaciones diagnósticas y comercialización* [en línea]. Trabajo Final de Grado, Facultad de Farmacia, Universidad Complutense, Madrid, España, 2017. Disponible en <https://eprints.ucm.es/id/eprint/56518/1/MARIA%20PISANO%20CASTRO.pdf>
- [19] *PET - CT y cáncer de mama* [en línea]. Presentación del Máster Internacional de Especialización en Mastología, Universidad Internacional Menéndez Pelayo, Madrid, España, 2021. Disponible en https://www.master-mastologia.com/wp-content/uploads/2021/01/PRESENTACION_PET-TAC-y-cancer-de-mama_Luis-Dominguez-Gadea.pdf
- [20] Sociedad Española de Astronomía. *Isótopo* [en línea]. Facultad de Física, Universidad de Barcelona, España [consulta 04/05/2022]. Disponible en <https://www.sea-astronomia.es/glosario/isotopo>

- [21] Radiation Dosimetry. *What is Stable Nuclei?* [en línea]. © Radiation Dosimetry, 2022 [consulta 04/05/2022]. Disponible en <https://www.radiation-dosimetry.org/what-is-stable-nuclei-unstable-nuclei-definition/>
- [22] LÓPEZ-DURÁN, Fred Alonso; DRUCKER-COLÍN, René. La radioquímica de la [18F]-FDG: la primera experiencia en México. En *Tip Revista Especializada en Ciencias Químico-Biológicas*. Universidad Nacional Autónoma de México, Distrito Federal, México, 2004, vol. 7, núm. 2, pp. 76-82. ISSN 1405-888X.
- [23] GATO-RIVERA, Beatriz. La antimateria. En *Revista Española de Física*. Real Sociedad Española de Física, Madrid, España, 2020, vol. 34, núm. 3, pp. 1-6. arXiv:2205.08348v1. Disponible en <https://arxiv.org/ftp/arxiv/papers/2205/2205.08348.pdf>
- [24] KAISER, David. La física y los diagramas de Feynman. En *Investigación y Ciencia*. Barcelona, España, 2005, núm. 348, pp. 74-83. Disponible en <https://www.investigacionyciencia.es/revistas/investigacion-y-ciencia/dixido-de-carbono-bajo-tierra-404/la-fsica-y-los-diagramas-de-feynman-4508>
- [25] Physics. *What is the physical interpretation of the two tree-level Feynman diagrams for $ee \rightarrow ee$ scattering?* [en línea] Stack Exchange Inc., Nueva York, EE.UU., 2022, [consulta: 05/05/2022]. Disponible en <https://physics.stackexchange.com/questions/696979/what-is-the-physical-interpretation-of-the-two-tree-level-feynman-diagrams-for>
- [26] LOPEZ-DURAN, Fred Alonso; ZAMORA-ROMO, Efraín; ALONSO-MORALES, José Luis [et al.]. Tomografía por emisión de positrones: los nuevos paradigmas. En *Tip Revista Especializada en Ciencias Químico-Biológicas*. Universidad Nacional Autónoma de México, Distrito Federal, México, 2007, vol. 10, núm. 1, pp. 26-35. ISSN 1405-888X.
- [27] MARTÍNEZ-ZALDÍVAR, Nananci. *Modelos matemáticos de clasificación automática de pacientes de Parkinson usando análisis de datos adquiridos mediante tomografía por emisión de positrones*. Tesis de Licenciatura en Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, Distrito Federal, México, 2019. Disponible en https://www.researchgate.net/figure/Figura-2-Camara-PET-y-aniquilacion-positron-electron_fig2_335777641
- [28] CAÑADAS-CASTRO, Mario. *Aspectos metodológicos para la evaluación de sistemas de tomografía por emisión de positrones empleando técnicas Montecarlo, protocolos estandarizados y diferentes trazadores*. Tesis Doctoral, Departamento de Ingeniería Electrónica, Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid, España, 2009. Disponible en https://oa.upm.es/14654/1/MARIO_CANADAS_CASTRO.pdf
- [29] MOLINET-DRONDA, Francisco. *Análisis in-vivo mediante tomografía por emisión de positrones (PET) del metabolismo cerebral inducido por la depleción dopaminérgica nigroestriada en modelos animales de enfermedad de Parkinson*. Memoria de Doctorado, Facultad de Medicina, Universidad de Navarra, Pamplona, España, 2016. Disponible en https://www.researchgate.net/figure/Figura-9-Representacion-de-la-tecnica-empleada-en-los-tomografos-PET_fig7_304571795

- [30] VICENTE-TORRICO, Esther. *Evaluación de las características y rendimiento de escáneres PET de pequeños animales* [en línea]. Trabajo de Investigación de Doctorado, Departamento de Física Atómica, Molecular y Nuclear, Facultad de Ciencias Físicas. Universidad Complutense, Madrid, España, 2012. Disponible en http://nuclear.fis.ucm.es/research/thesis/dea_esther.pdf
- [31] NÚÑEZ, Margarita. *Tomografía por emisión de positrones (PET): Fundamentos* [en línea]. Escuela Universitaria de Tecnología Médica UdelaR. Comité de Tecnólogos de ALASBIMN (Asociación Latinoamericana de Sociedades de Biología y Medicina Nuclear), Montevideo, Uruguay, 2008 [consulta: 06/05/2022]. Disponible en https://www.alasbimn.net/comites/tecnologos/material/Fundamentos_PET.pdf
- [32] Consejo de Seguridad Nuclear. *Curso de PR para dirigir instalaciones de Rayos X con fines de diagnóstico médico (IRD-OP-GR-T02)* [en línea]. © CSN, Madrid, España, 2009 [consulta 08/05/2022]. Disponible en https://csn.ciemat.es/MDCSN/recursos/ficheros_md/133100241_2411200913036.pdf
- [33] LÓPEZ-HERRAIZ, Joaquín. *Técnicas avanzadas de reconstrucción de imagen nuclear PET, X-CT y SPECT* [en línea]. Memoria del Trabajo del Máster de Física Biomédica, Departamento de Física Atómica, Molecular y Nuclear, Facultad de Ciencias Físicas, Universidad Complutense, Madrid, España, 2008. Disponible en http://nuclear.fis.ucm.es/research/thesis/MASTER_JOAQUIN_LOPEZ_REV.PDF
- [34] KONTAXAKIS, George.; VAQUERO-LÓPEZ, Juan José; SANTOS, Andrés. Reconstrucción de imagen en tomografía por emisión de positrones. En *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales*. Madrid, España, 2002, vol. 96, núms. 1-2, pp. 45-57. ISSN 1137-2141.
- [35] SHEPP, Lawrence A.; VARDI, Yehuda. Maximum likelihood reconstruction for emission tomography. En *IEEE Transactions on Medical Imaging*, Buffalo, Nueva York, EE. UU., 1982, vol.1, núm. 2, pp.113-122. ISSN 1558-254X.
- [36] BHANOT, Karan; QI, Miao; ERICKSON, John S. [et al.]. The Problem of Fairness in Synthetic Healthcare Data. En *Entropy*. Multidisciplinary Digital Publishing Institute (MDPI), Basilea, Suiza, 2021, vol. 23, núm. 9, p. 1165. ISSN 1099-4300. Disponible en <https://www.mdpi.com/1099-4300/23/9/1165>
- [37] CHLAP, Phillip; MIN, Hang; VANDENBERG, Nym [et al.]. A review of medical image data augmentation techniques for deep learning applications. En *Journal of Medical Imaging and Radiation Oncology*. The Royal Australian and New Zealand College of Radiologists, 2021, vol. 65, núm. 5, pp. 545-563. ISSN 1754-9477.
- [38] POURHAHROKHI, Narges; KOUCHAKI, Samaneh; KOBER, Kord M. [et al.]. A Hamiltonian Monte Carlo Model for Imputation and Augmentation of Healthcare Data. *arXiv preprint arXiv:2103.02349*, 2021.
- [39] AVIÑÓ, Laura; RUFFINI, Matteo; GAVALDÀ, Ricard. Generating synthetic but plausible healthcare record datasets. *arXiv preprint arXiv:1807.01514*, 2018.

- [40] HÄGGSTRÖM, Ida; SCHMIDTLEIN, C.Ros; CAMPANELLA, Gabriele [et al.]. DeepPET: A deep encoder-decoder network for directly solving the PET reconstruction inverse problem. En *Medical image analysis*. Elsevier B.V., 2019, vol. 54, p. 253-262. ISSN 1361-8415.
- [41] POVES-CARRASCO, Pablo. *Desarrollo de una librería software para el análisis y validación de imágenes 2D/3D PET basado en el protocolo NEMA*. Trabajo Final de Grado en Ingeniería Informática, Escola Tècnica Superior d'Enginyeria Informàtica (ETSInf), Universitat Politècnica de València, España, 2022.
- [42] LÓPEZ-CHILET, Álvaro. *Clasificación de imágenes de resonancia magnética cerebral mediante redes neuronales para el diagnóstico médico*. Trabajo Final de Grado en Ingeniería Informática, Escola Tècnica Superior d'Enginyeria Informàtica (ETSInf), Universitat Politècnica de València, España, 2019.
- [43] SÁNCHEZ-ROMERO, Rafael Vicente. *Análisis de imágenes médicas para dar soporte al diagnóstico de Alzheimer*. Trabajo Final de Grado en Ingeniería Informática, Escola Tècnica Superior d'Enginyeria Informàtica (ETSInf), Universitat Politècnica de València, España, 2019.
- [44] Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 (Reglamento General de Protección de Datos), *Diario Oficial de la Unión Europea*, núm. 119, de 4 de mayo de 2016, páginas de L 119/1 a L 119/88. ISSN 1977-0685.
- [45] Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (LOPDGDD), *Boletín Oficial del Estado*, núm. 294, sección I, de 6 de diciembre de 2018, páginas de 119788 a 119857. ISSN 0212-033X.
- [46] Ley 16/2003, de 28 de mayo, de cohesión y calidad del Sistema Nacional de Salud, *Boletín Oficial del Estado*, núm. 128, de 29 de mayo de 2003, páginas 20567 a 20588.
- [47] Ley 14/2007, de 3 de julio, de Investigación biomédica. *Boletín Oficial del Estado*, núm. 159, de 4 de julio de 2007, páginas de 28826 a 28848.
- [48] Agencia Española de Protección de Datos (AEPD). *Anonimización y seudonimización* [en línea]. Madrid, España, 2021 [consulta 22/05/2022]. Disponible en <https://www.aepd.es/es/prensa-y-comunicacion/blog/anonimizacion-y-seudonimizacion#:~:text=Es%20decir%2C%20los%20datos%20se, en%20el%20conjunto%20de%20datos>
- [49] FARD, Azin Shokraei; REUTENS, David C.; VEGH, Viktor. CNNs and GANs in MRI-based cross-modality medical image estimation. *arXiv preprint arXiv:2106.02198*, 2021.
- [50] TORREY, Lisa; Shavlik, Jude; Walker, Trevor; et al. Transfer Learning via Advice Taking. En *Advances in Machine Learning I*. Springer, Berlin, Heidelberg, 2010. pp. 147-170. ISBN 9783642051760.
- [51] GOODFELLOW, Ian; POUGET-ABADIE, Jean; MIRZA, Mehdi [et al.]. Generative Adversarial Nets. En *Advances in Neural Information Processing Systems 27 (NIPS 2014)*. 28th Annual Conference on Neural Information Processing Systems, 2014. Curran Associates, Inc. Nueva York, EE.UU., 2015, vol. 1, pp. 2672-2680. ISBN 9781510800410.

- [52] AVALOS, Sebastian; ORTIZ, Julian M. Geological modeling using a recursive convolutional neural networks approach. *arXiv preprint arXiv:1904.12190*, 2019.
- [53] MASI, Filippo; STEFANO, Ioannis; VANNUCCI, Paolo [et al.]. Thermodynamics-based Artificial Neural Networks for constitutive modeling. En *Journal of the Mechanics and Physics of Solids*, 2021, vol. 147, p. 104277. Disponible en https://www.sciencedirect.com/science/article/pii/S0022509620304841?casa_token=BjK47mza92wAAAAA:g-ySUx_Wpq7CnWNeSAMVcByjQJQZHURZVMaPHMkSFGuTPU2q5JxspN5-3c32hWgRTkKLDE1uxw
- [54] KALYANI, G.; JANAKIRAMAIAH, B. Deep learning-based detection and classification of adenocarcinoma cell nuclei. En *Trends in Deep Learning Methodologies, Hybrid Computational Intelligence for Pattern Analysis*, 2021, pp. 241-264. ISBN 9780128222263.
- [55] WANG, Zhengwei; SHE, Qi; WARD, Tomas E. Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy. En *ACM Computing Surveys (CSUR)*. Nueva York, EE.UU., 2021, vol. 54, núm. 2, art. 37, pp. 1-38. ISSN 0360-0300.
- [56] FERNÁNDEZ-ISASI, Isabel. *Juegos de suma cero. Teoremas de Von Neumann, Sion y Knieser-Fan* [en línea]. Trabajo Final de Grado, Facultad de Ciencias, Universidad de Valladolid, España, 2016. [consulta: 11/05/2022] Disponible en: <https://uvadoc.uva.es/handle/10324/19057>
- [57] SALIMANS, Tim; GOODFELLOW, Ian; WOJCIECH, Zaremba [et al.]. Improved Techniques for Training GANs. En *Advances in Neural Information Processing Systems 29 (NIPS 2016). 30th Annual Conference on Neural Information Processing Systems, 2016*. Curran Associates, Inc. Nueva York, EE.UU., 2017, vol. 1, pp. 2234-2243. ISBN 9781510838819.
- [58] YADAV, Abhay; SHAH, Sohil; XU, Zheng [et al.]. Stabilizing Adversarial Nets With Prediction Methods. En *6th International Conference on Learning Representations (ICLR)*. Vancouver, Columbia Británica, Canada, 2018. arXiv preprint arXiv:1705.07364.
- [59] METZ, Luke; POOLE, Ben; PFAU, David [et al.]. Unrolled Generative Adversarial Networks. En *5th International Conference on Learning Representations (ICLR)*. Toulon, Francia, 2017. arXiv preprint arXiv:1611.02163.
- [60] RADFORD, Alec; METZ, Luke; CHINTALA, Soumith. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015
- [61] ARJOVSKY, Martin; CHINTALA, Soumith; BOTTOU, Léon. Wasserstein Generative Adversarial Networks. En *Proceedings of the 34th International Conference on Machine Learning*. Proceedings of Machine Learning Research (PMLR), 2017, vol. 70, pp. 214-223. ISSN 2640-3498.
- [62] GULRAJANI, Ishaan; AHMED, Faruk; ARJOVSKY, Martin [et al.]. Improved Training of Wasserstein GANs. En *Advances in Neural Information Processing Systems 30 (NIPS 2017). 31th Annual Conference on Neural Information Processing Systems, 2017*. Curran Associates, Inc. Nueva York, EE.UU., 2018, vol. 1, pp. 5768-5778. ISBN 9781510860964.

-
- [63] O'CONNOR, Michael; RHODES, Deborah; HRUSKA, Carrie. Molecular breast imaging. En *Expert Review of Anticancer Therapy*. Taylor Francis Online, Londres, Reino Unido, 2009, vol. 9, núm. 8, pp. 1073-1080. e-ISSN 1744-8328.
- [64] HEATH, Mike; BOWYER, Kevin; KOPANS, Daniel [et al.]. The digital database for screening mammography. En *Proceedings of the Fifth International Workshop on Digital Mammography*, Madison, Wisconsin, EE.UU., 2001, pp. 212-218. ISBN 1930524005
- [65] NUNN, Eric J.; KHADIVI, Pejman; SAMAVI, Shadrokh. Compound Frechet Inception Distance for Quality Assessment of GAN Created Images. *arXiv preprint arXiv:2106.08575*, 2021.
- [66] DHARIWAL, Prafulla; NICHOL, Alexander. Diffusion models beat gans on image synthesis. En *Advances in Neural Information Processing Systems*, 2021, vol. 34, pp. 8780-8794. ISBN 9781713845393.

APÉNDICE A

Objetivos de Desarrollo Sostenible (ODS)

Consideraciones sobre los nexos de este Trabajo de Fin de Grado (TFG) con los Objetivos de Desarrollo Sostenible promulgados por la ONU.

Es de dominio público que la Agenda 2030 sobre el desarrollo sostenible, aprobada por la ONU en 2015, es una oportunidad para que los países y sus sociedades emprendan un nuevo camino con el que mejorar la vida de las personas.

La Agenda cuenta con 17 objetivos de desarrollo sostenible (ODS), que incluyen desde la eliminación de la pobreza hasta el combate al cambio climático, la educación, la igualdad de la mujer, la defensa del medio ambiente o el diseño de nuestras ciudades.

Estos objetivos, son:



Figura A.1: Imagen extraída de UNESCO.

La presente memoria pretende evidenciar la gran aportación que la inteligencia artificial (IA) puede hacer en el entrenamiento de métodos de reconstrucción de imágenes de tomografías por emisión de positrones (PET). Pero es obvio que las ventajas de la utilización de la IA no se circunscriben únicamente a este ámbito.

Es por ello, que a continuación se intentará mostrar cómo la IA puede ser un facilitador de los objetivos de desarrollo sostenible y cómo el desarrollo de las tecnologías basadas en la IA podría convertirse en un valor para que nuestro mundo sea más justo y equitativo. Estamos plenamente convencidos de que la IA debe dar su apoyo para lograr cada uno de los 17 ODS.

A pesar de la amplitud de aplicaciones de la IA en cada uno de los 17 objetivos de Naciones Unidas, nos limitaremos a reseñar aquí cuáles son, según el criterio del autor de este texto, los objetivos ODS más vinculados con el propósito de este TFG.

Cada uno de estos ODS es intrínsecamente complejo, debido al amplio número de facetas que los componen y al hecho de que muchos de ellos interaccionen entre sí. A pesar de ello, se ha intentado sintetizar cuatro vínculos de estos objetivos de Naciones Unidas con la finalidad del presente trabajo.

- En primer lugar, interesa especialmente destacar que el estudio del proceso de reconstrucción de imágenes PET aquí analizado, entronca directamente con el *Objetivo #3* de Naciones Unidas, el cual trata de cómo "favorecer el acceso de la población a una vida más sana y promover el bienestar de todas las personas, sea cual sea su edad".

El empleo de la IA, de los algoritmos de aprendizaje profundo y las redes neuronales, junto con la disposición y utilización libre de los datos, han de facilitar una mejor predicción y diagnóstico de las enfermedades. La disponibilidad de estos diagnósticos, gracias a la IA, cada vez será mayor entre los profesionales de la medicina, con una menor generación de costes y, por tanto, con mayor accesibilidad para cualquier persona o grupo poblacional.

No obstante, para que esto sea posible consideramos que son indispensables dos premisas fundamentales: por una parte, la unificación y armonización de los flujos de datos médicos disponibles y, por otra parte, la formación digital de los facultativos, para que así puedan utilizar convenientemente esta nueva tecnología a su disposición.

- Esta última reflexión, sobre la capacitación del personal sanitario, estaría estrechamente vinculada con otro de los ODS de la ONU, concretamente con el *Objetivo #4*, al "garantizar una educación inclusiva y equitativa de calidad y la promoción de oportunidades de aprendizaje permanente para todos".

El mero hecho de que, para la utilización en medicina de las nuevas técnicas de IA, sea condición imprescindible la formación digital de los profesionales sanitarios, se traduce automáticamente en un hito hacia la eliminación de la llamada *brecha digital*.

Una educación digital de calidad, no sólo ha de posibilitar el acceso a mejores y mayores oportunidades de aprendizaje para todas las personas, sino que también que los métodos y estilos de aprendizaje se liberen de los impedimentos de la presencialidad física. En este aspecto concreto, la IA puede cubrir un hueco significativo.

- En otro orden de consideraciones, es primordial no olvidar que una de las más importantes aplicaciones del entrenamiento de métodos de reconstrucción de imágenes y, por ende, de las tomografías por emisión de positrones, es la detección temprana de cánceres de mama.

Creemos firmemente que el diagnóstico precoz de las lesiones neoplásicas gracias a la ayuda de la IA, enlazaría directamente con el *Objetivo #5* de Naciones Unidas, que no es otro que la adquisición por parte de las mujeres de un empoderamiento sobre su propia patología. De esta manera, las mujeres podrán acrecentar su protagonismo

y control sobre la evolución de la enfermedad. Esto cobra especial relevancia en estas difíciles situaciones y se traduce en una mayor probabilidad de supervivencia de las mujeres afectadas.

- Por último, consideramos indispensable destacar, como ya se ha apuntado anteriormente, que la IA puede conseguir que la detección del cáncer de mama sea más accesible.

Actualmente, las altas tasas de mortalidad de las pacientes con cáncer de mama a menudo se deben a una detección tardía, especialmente en las zonas rurales, donde puede ser difícil acceder a exámenes de detección confiables y asequibles.

La mamografía de detección es un método ampliamente utilizado, pero su uso en países de escasos recursos es limitado debido al coste del equipo y la habilidad experta requerida para la interpretación de las mamografías.

En entornos con recursos limitados, las soluciones innovadoras asistidas por IA para detectar precozmente el cáncer de mama podrían ser el camino a seguir. Esto, obviamente, ayudaría a alcanzar el *Objetivo #10*, al contribuir a "la reducción de la desigualdad en y entre los países".

En la siguiente página se adjunta el Anexo facilitado por la ETSinf relativo al grado de relación de este trabajo con los Objetivos de Desarrollo Sostenible (ODS) de Naciones Unidas.



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



ANEXO

OBJETIVOS DE DESARROLLO SOSTENIBLE

Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).

Objetivos de Desarrollo Sostenibles	Alto	Medio	Bajo	No Procede
ODS 1. Fin de la pobreza.				X
ODS 2. Hambre cero.				X
ODS 3. Salud y bienestar.	X			
ODS 4. Educación de calidad.	X			
ODS 5. Igualdad de género.	X			
ODS 6. Agua limpia y saneamiento.				X
ODS 7. Energía asequible y no contaminante.				X
ODS 8. Trabajo decente y crecimiento económico.		X		
ODS 9. Industria, innovación e infraestructuras.		X		
ODS 10. Reducción de las desigualdades.	X			
ODS 11. Ciudades y comunidades sostenibles.			X	
ODS 12. Producción y consumo responsables.				X
ODS 13. Acción por el clima.				X
ODS 14. Vida submarina.				X
ODS 15. Vida de ecosistemas terrestres.				X
ODS 16. Paz, justicia e instituciones sólidas.				X
ODS 17. Alianzas para lograr objetivos.			X	



Escola Tècnica
Superior d'Enginyeria
Informàtica

ETS Enginyeria Informàtica
Camí de Vera, s/n. 46022. València
T +34 963 877 210
F +34 963 877 219
etsinf@upvnet.upv.es - www.inf.upv.es



Figura A.2: Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).

APÉNDICE B

Distancia Wasserstein. Ejemplo

Para comprender el concepto que subyace bajo la ecuación de las WGAN propuesta por Arjovsky et al. en el subepígrafe 5.6.2, el arquitecto de software Jonathan Hui publicó en 2018 un modelo en la plataforma blogging *Medium.com*¹.

Consideremos una situación donde se plantea el traslado de cajas de un lugar a otro, como en la Figura B.1. Inicialmente existen 6 cajas y queremos moverlas desde la izquierda a las posiciones de la derecha marcadas por cuadrados punteados. En un primer movimiento, trasladamos la caja #1 desde la ubicación 1 a la 7. El coste de dicho traslado será igual al peso de la caja por la distancia que tendrá que recorrer. Si, para simplificar, suponemos que el peso de las cajas es unitario, el coste total del traslado de la caja #1 será igual a 6, que son el número de posiciones que ha de recorrer desde la ubicación 1 hasta la 7.

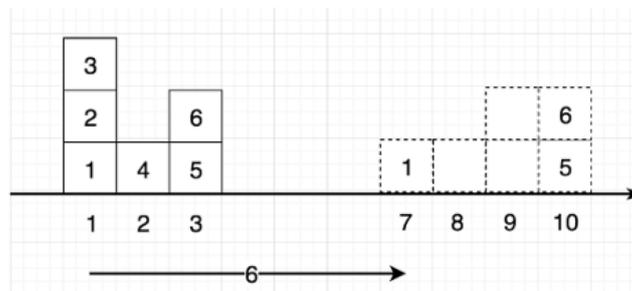


Figura B.1: Imagen extraída de Medium.com.

Siguiendo este mismo mecanismo, en la Figura B.2 se presentan dos planes de traslado γ diferentes. Las tablas γ_1 e γ_2 de la derecha de la imagen ilustran cómo se mueven las cajas en cada uno de dichos planes.

Por ejemplo, en el primer plan γ_1 , se trasladan 2 cajas desde la ubicación 1 a la ubicación 10 y, por lo tanto, a la entrada $\gamma(1,10)$ le corresponde un valor igual a 2. A la entrada $\gamma(1,7)$ le correspondería un valor de 1, ya que sólo una caja es trasladada desde la posición 1 a la 7. De igual manera iríamos completando sucesivamente las tablas γ_1 e γ_2 .

¹<https://jonathan-hui.medium.com/gan-wasserstein-gan-wgan-gp-6a1a2aa1b490>

Puede observarse que, en este ejemplo, el coste total de transporte en cada uno de los planes es igual 42.

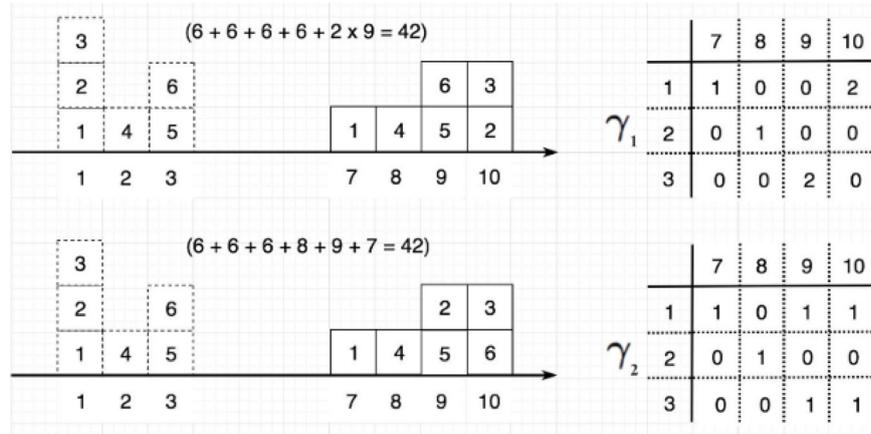


Figura B.2: Imagen extraída de Medium.com.

Sin embargo, esto no sucede siempre así. Puede darse el caso en donde se planteen planes de transporte que tengan costes diferentes. Por ejemplo, en el primer plan de traslado propuesto en la Figura siguiente, el coste total de transporte de trasladar las cajas #1 y #2 es igual a dos. Por el contrario, en la Figura B.3, el coste de transporte de trasladar las mismas cajas aquí es igual a 6.

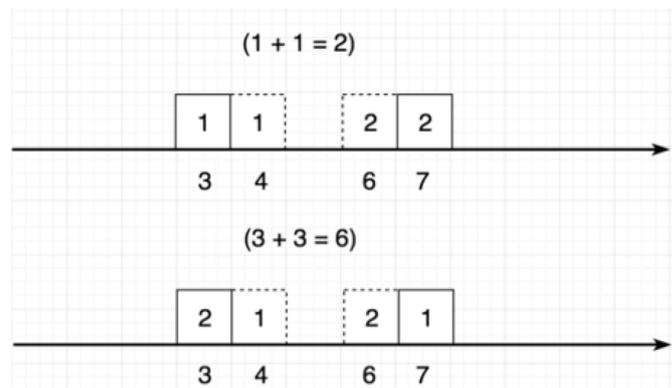


Figura B.3: Imagen extraída de Medium.com.

Éste es precisamente el concepto que utiliza la distancia de Wasserstein o EMD. En el último ejemplo ambos planes tienen un coste diferente, siendo el coste mínimo (EMD) igual a dos.

Si en lugar de cajas nos referimos a distribuciones de probabilidad, la distancia de Wasserstein sería igual al coste mínimo del transporte de “masa” en la conversión de una determinada distribución de datos q a una distribución de datos p .

Como hemos visto, la distancia de Wasserstein, para una distribución de datos reales P_r y una distribución de datos generados P_g , se define matemáticamente como el mayor límite inferior (*ínfimo*) para cualquier plan de transporte (es decir, el coste del plan más barato):

$$W(P_r, P_g) = \inf_{\gamma \in \Pi(P_r, P_g)} \mathbb{E}_{(x,y) \sim \gamma} [\|x - y\|] \quad (\text{B.1})$$

En la ecuación B.1, $\Pi(P_r, P_g)$ representa el total de las distribuciones conjuntas $\gamma(x, y)$, cuyas distribuciones marginales son iguales a P_r y P_g respectivamente.

Intuitivamente, $\gamma(x, y)$ indicaría la cantidad de "masa" que debe moverse de x a y para transformar la distribución P_r en la distribución P_g . La distancia de Wasserstein se podría entender como el coste del plan de transporte óptimo.

APÉNDICE C

Función de pérdida W -Loss. Desarrollo matemático

En el subepígrafe 5.6.2 llegamos a definir la distancia de Wasserstein como

$$W(P_r, P_g) = \inf_{\gamma \in \Pi(P_r, P_g)} \mathbb{E}_{(x,y) \sim \gamma} [\|x - y\|] \quad (\text{C.1})$$

Desafortunadamente, como ya vimos entonces, este tipo de optimización no es práctico. En muchas aplicaciones, como por ejemplo en el caso de las imágenes, la variable de entrada puede tener fácilmente miles de dimensiones. Obtener una mera aproximación de γ sería virtualmente imposible.

Debido pues a la dificultad para calcular el ínfimo de todos los $\gamma \in \Pi(P_r, P_g)$, Arjovsky et al. en su artículo propusieron utilizar la dualidad de *Kantorovich-Rubinstein* para simplificar el cálculo y así poder formular alternativamente la ecuación C.1 como,

$$W(P_r, P_\theta) = \sup_{\|f\|_L \leq 1} \mathbb{E}_{x \sim P_r} [f(x)] - \mathbb{E}_{x \sim P_\theta} [f(x)] \quad (\text{C.2})$$

donde *sup* era el límite superior mínimo (*supremo*) y f una función de *1-Lipschitz*. Esta reformulación aportaba la ventaja de que, para calcular la distancia de Wasserstein, tan sólo era preciso encontrar una función de este tipo.

En este apéndice se intentará explicar pormenorizadamente cómo se obtiene la ecuación C.2 a partir de la ecuación C.1 mediante la dualidad de Kantorovich-Rubinstein. Todo el desarrollo que se presentará a continuación fue publicado en 2017 en el blog personal de GitHub de Vincent Herrmann ¹, actualmente estudiante de doctorado en aprendizaje automático en el IDSIA (*Istituto Dalle Molle di studi sull'intelligenza artificiale*) de Lugano, Suiza.

Supongamos que nuestras distribuciones de probabilidad p_r y p_θ son continuas, es decir, tienen rangos de valores infinitos.

En este caso, el conjunto de todas las distribuciones conjuntas igual a $\pi(p_r, p_\theta)$, siendo las distribuciones marginales iguales a p_r y p_θ . Entonces la distancia de Wasserstein C.1 se podría definir como:

¹Vincent Herrmann. *Wasserstein GAN and the Kantorovich-Rubinstein Duality*. <https://vincentherrmann.github.io/blog/wasserstein/>

$$W(p_r, p_\theta) = \inf_{\gamma \in \pi} \iint_{x,y} \|x - y\| \gamma(x, y) dx dy = \inf_{\gamma \in \pi} \mathbb{E}_{x,y \sim \gamma} [\|x - y\|] \quad (\text{C.3})$$

Si añadimos los términos adecuados, podemos eliminar todas las restricciones en las distribuciones $(x, y) \sim \gamma$. Esto se consigue añadiendo una optimización adicional sobre la función $f : x \mapsto k \in \mathbb{R}$, lo cual descartaría como posible solución a todo $\gamma \notin \pi$:

$$\begin{aligned} W(p_r, p_\theta) &= \inf_{\gamma \in \pi} \mathbb{E}_{x,y \sim \gamma} [\|x - y\|] \\ &= \inf_{\gamma} \mathbb{E}_{x,y \sim \gamma} [\|x - y\| + \underbrace{\sup_f \mathbb{E}_{s \sim p_r} [f(s)] - \mathbb{E}_{t \sim p_\theta} [f(t)] - (f(x) - f(y))}_{f}] \\ &= \begin{cases} 0, & \text{si } \gamma \in \pi \\ +\infty & \text{en los demás casos} \end{cases} \\ &= \inf_{\gamma} \sup_f \mathbb{E}_{x,y \sim \gamma} [\|x - y\| + \mathbb{E}_{s \sim p_r} [f(s)] - \mathbb{E}_{t \sim p_\theta} [f(t)] - (f(x) - f(y))] \end{aligned}$$

Ahora tenemos una optimización binivel, donde un problema se encuentra inmerso dentro de otro. Esto significa que tomaremos la solución óptima de la optimización interna (\sup_f), para cada valor de la optimización exterior (\inf_{γ}), y de todas estas posibles soluciones elegiremos la mejor para la optimización externa.

Para continuar, tenemos que hacer uso del *principio minimax*, que dice que en ciertos casos podemos invertir el orden de *inf* y *sup* sin que la solución varíe. No obstante primero tendremos que demostrar que nuestro problema pertenezca realmente a uno de estos casos.

Consideremos una función $g : A, B \mapsto \mathbb{R}$ y definamos el ínfimo y supremo de dicha función como:

- $g(\hat{a}, \hat{b}) = \inf_{a \in A} \sup_{b \in B} g(a, b)$: “el menor de todos los máximos”.
- $g(\hat{a}', \hat{b}') = \sup_{b \in B} \inf_{a \in A} g(a, b)$: “el mayor de todos los mínimos”.

La argumentación de que $g(\hat{a}, \hat{b}) \geq g(\hat{a}', \hat{b}')$ es simple: cualquier $g(\hat{a}, \hat{b})$ es admitido automáticamente como candidato al ínfimo de $\sup_{b \in B} \inf_{a \in A} g(a, b)$, pero no a la inversa.

Esto equivale a decir que $\sup_b g(a', b) \geq g(a', b') \geq \inf_a g(a, b') \forall a', b'$, lo cual es cierto por definición.

Para $g(\hat{a}, \hat{b}) > g(\hat{a}', \hat{b}')$ al menos una de estas afirmaciones debe ser verdadera:

- $g(\hat{a} + t, \hat{b}) < g(\hat{a}, \hat{b})$ para todo $t \neq 0$. Esto sólo será posible si $\sup_{b \in B} g(a, b)$ es no convexo en a porque $g(\hat{a}, \hat{b})$ ya es un ínfimo para \hat{a} .
- $g(\hat{a}', \hat{b}' + t) > g(\hat{a}', \hat{b}')$ para todo $t \neq 0$. Esto sólo será posible si $\inf_{a \in A} g(a, b)$ es no cóncavo en b porque $g(\hat{a}', \hat{b}')$ ya es un supremo para \hat{b}' .

Esto significa, por supuesto, que si $\sup_{b \in B} g(a, b)$ es convexo e $\inf_{a \in A} g(a, b)$ es cóncavo, entonces se puede aplicar el teorema minimax, resultando:

$$g(\hat{a}, \hat{b}) = g(\hat{a}', \hat{b}') \quad (\text{C.4})$$

En nuestro caso, ya se puede ver a simple vista que la condición de convexidad se cumple.

Entonces,

$$\begin{aligned} & \sup_f \inf_{\gamma} \mathbb{E}_{x,y \sim \gamma} [\|x - y\| + \mathbb{E}_{s \sim p_r} [f(s)] - \mathbb{E}_{t \sim p_\theta} [f(t)] - (f(x) - f(y))] \\ = & \sup_f \mathbb{E}_{s \sim p_r} [f(s)] - \mathbb{E}_{t \sim p_\theta} [f(t)] + \underbrace{\inf_{\gamma} \mathbb{E}_{x,y \sim \gamma} [\|x - y\| - (f(x) - f(y))]}_{=} \\ & = \begin{cases} 0, & \text{si } \|f\|_L \leq 1 \\ -\infty & \text{en los demás casos} \end{cases} \end{aligned}$$

Se observa pues que el ínfimo es cóncavo, tal y como se requiere. Como todas las funciones f de Lipschitz continuas generan la misma solución óptima para (\inf_{γ}) , y sólo ellas son soluciones factibles de (\sup_f) , podemos convertir esta condición en una restricción. Con ello, obtenemos la forma dual de la distancia de Wasserstein:

$$\begin{aligned} W(p_r, p_\theta) &= \sup_f \mathbb{E}_{s \sim p_r} [f(s)] - \mathbb{E}_{t \sim p_\theta} [f(t)] + \inf_{\gamma} \mathbb{E}_{x,y \sim \gamma} [\|x - y\| - (f(x) - f(y))] \\ &= \sup_{\|f\|_L \leq 1} \mathbb{E}_{s \sim p_r} [f(s)] - \mathbb{E}_{t \sim p_\theta} [f(t)] \end{aligned}$$

Este es nuestro caso particular de la dualidad de Kantorovich-Rubinstein que, como puede observarse, coincide con la función de pérdida W -Loss 5.26 que vimos en el subepígrafe 5.6.2 dedicado a las redes Wasserstein GAN.

En realidad, sería válida para otras métricas además de la métrica euclidiana que hemos utilizado. Pero la función f es adecuada para ser aproximada por una red neuronal, y esta versión tiene la ventaja de que la continuidad de Lipschitz se puede conseguir simplemente fijando los pesos.