



UNIVERSITAT OBERTA DE CATALUNYA (UOC)

MASTER'S DEGREE IN DATA SCIENCE

MASTER'S THESIS

AREA: MEDICINE (TFM-MED)

**A GAN approach to synthetic PET imaging generation for
breast cancer diagnosis**

Author: Javier Cantero Lorenzo

Supervisor: Dr. Héctor Espinós Morató

Professor: Dr. Jordi Casas Roma

Barcelona, June 6, 2021

Copyright

This work is licensed under Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Spain License.



The GitHub code repository¹ of this Master's Thesis is licensed under MIT license.

¹<https://github.com/jcantlord/PET-SinoGAN>

FICHA DEL TRABAJO FINAL

Título del trabajo:	A GAN approach to synthetic PET imaging generation for breast cancer diagnosis
Nombre del autor:	Javier Cantero Lorenzo
Nombre del colaborador/a docente:	Dr. Héctor Espinós Morató
Nombre del PRA:	Dr. Jordi Casas Roma
Fecha de entrega (mm/aaaa):	06/2021
Titulación o programa:	Máster en Ciencia de Datos
Área del Trabajo Final:	Área Medicina (TFM-MED)
Idioma del trabajo:	English
Keywords	Breast Cancer, Sinogram, GAN, PET

To all the real fighters,



Acknowledgements

In the first place, I would like to acknowledge and give my full gratitude to my supervisors, Dr. Héctor Espinós Morató and David Cascales Picó, for all their guidance and support during the completion of this Master's thesis. I have had at all times a continuous mentoring with great professionalism, being able to solve my doubts and incorporate into the workflow all the ideas and advices provided. Once this document is submitted, I look forward to publishing with them all the findings gathered within this document and further investigate it together.

This thanks is extended to all the members of the Medical Imaging Reconstruction Group (MIRG) of the I3M Group of Detectors for Molecular Imaging research institute (UPV-CSIC) who welcomed me from the very beginning and allowed me to be part of one of their most important projects.

It would not be honest if I did not mention my gratitude to Dr. Claudio Mirasso and Dr. Miguel Cornelles from Institute for Cross-Disciplinary Physics and Complex Systems (UIB-CSIC), who supervised my Bachelor's Thesis and guided me in my first contact with Artificial Intelligence.

I could not have reached the point where I am right now if it had not been for the constant support I have received from my parents and family, who have been by my side on the road to achieving my dreams since the day I was born. Thank you for giving me light even in the darkest moments.

It is impossible to mention all those people who in one way or another have influenced my decisions and have allowed me to reach this point. I would especially like to thank my coaches from True Love Crossfit, my students from Aula Balear, the NotiFriends group, Marc, Marina, Tolo, Delegado, Cati, Jaume, Biel, and in general all the people who have accompanied me during the completion of this thesis.

As a final note, I would like to thank the entire Python, Data Science, and AI community for the democratization of knowledge and the tools that have made this work possible. In particular I have to name Andrew Ng and his team thanks to whom I was able to have free access to knowledge from the best universities regarding Deep Learning and generative models.

Abstract

Positron Emission Tomography (PET) is a non-invasive medical diagnostic test widely used in the early detection of breast cancer with significant advantages over other simpler techniques such as mammograms, sonograms, or biopsies. There are several factors that make it difficult to generate a large enough dataset to train AI-based PET reconstruction models, such as the controlled use of radioactive tracers or the sensitive nature of the images.

In recent years, different approaches have been explored with the aim of solving this bottleneck, including the simulation of PET tests using anthropomorphic phantoms or classical data augmentation procedures. Despite being able to solve part of the problem, these techniques are still simplified simulations with a low variability (introducing biases), in addition to an expensive consumption of computational resources.

In this work we propose DCGAN and WGAN generative networks for the unsupervised generation of synthetic data as a starting point in the training of PET reconstructors based on Artificial Intelligence. The nature of this networks allows to obtain a statistically accurate approximation to the probability distribution of the real patients dataset.

The end result of this work aims to lay the foundation for a standard workflow for synthetic data generation in those domains with a significant bottleneck in the creation of a large dataset, thus overcoming one of the main stumbling blocks in the creation of computer-aided diagnosis (CAD) in the analysis of medical images.

Keywords: Breast Cancer, PET, Deep Learning, GAN, Sinograms, Synthetic data.

Resumen

La tomografía por emisión de positrones (PET) es una prueba de diagnóstico médico no invasiva muy utilizada en la detección precoz del cáncer de mama, con importantes ventajas sobre otras pruebas más sencillas como son las mamografías, los sonogramas, o las biopsias. En la actualidad existen varios factores que dificultan la generación de un conjunto de datos lo suficientemente grande como para entrenar modelos de reconstrucción de PET basados en IA, como son el uso controlado de trazadores radiactivos o la naturaleza sensible de las imágenes.

En los últimos años se han explorado diferentes enfoques con el objetivo de resolver este cuello de botella, incluyendo la simulación de pruebas PET utilizando fantasmas antropomórficos o procedimientos clásicos de aumento de datos. A pesar de poder resolver parte del problema, estas técnicas no dejan de ser simulaciones simplificadas con una baja variabilidad (introduciendo sesgos), además de un costoso consumo de recursos computacionales.

En este trabajo se propone el uso de las redes generativas DCGAN y WGAN para la generación no supervisada de datos sintéticos como punto de partida en el entrenamiento de reconstructores PET basados en Inteligencia Artificial. La naturaleza de estas redes permite obtener una aproximación estadísticamente precisa a la distribución de probabilidad del conjunto de datos de las pacientes reales.

El resultado final de este trabajo pretende sentar las bases de un estándar para la generación de datos sintéticos en aquellos dominios con un importante cuello de botella en la creación de un gran conjunto de datos, superándose de esta manera uno de los principales obstáculos a la hora de diseñar diagnósticos asistidos por ordenador (DAO) en análisis de imagen médica.

Palabras Clave: Cáncer de mama, PET, Aprendizaje profundo, GAN, Sinogramas, Datos sintéticos.

Contents

Abstract	ix
Contents	xiii
List of Figures	xv
List of Tables	xvii
1 Introduction	1
1.1 Background and motivation	1
1.1.1 Framework	2
1.2 Objectives	2
1.3 Personal motivation	3
1.4 Methodology	4
1.5 Planning	6
2 State of the art	7
2.1 Traditional PET reconstruction methods	7
2.1.1 Analytical methods	8
2.1.2 Iterative methods	9
2.2 Deep Learning based reconstruction	11
2.2.1 Data augmentation	12
2.2.1.1 Basic image manipulations	12
2.2.1.2 RandAugment	14
2.2.1.3 Generative models	14
3 Methodology	19
3.1 Workflow overview	19
3.2 Basic principles of PET medical imaging	20
3.2.1 Nuclear interactions	20

3.2.2	Line of response (LOR)	20
3.2.3	Sinograms	22
3.3	Databases	24
3.3.1	ACRIN-6688 Database	24
3.3.2	ONCOVISION-MAMMI Database	25
3.4	Preprocessing	27
3.5	Network architectures	27
3.5.1	Deep Convolutional GAN (DCGAN)	27
3.5.2	Wasserstein GAN (WGAN)	30
3.5.2.1	Weight Clipping	31
3.5.2.2	Gradient penalty	31
3.6	Evaluation metrics	32
3.6.1	Manual evaluation	33
3.6.1.1	Visual Turing Test	34
3.6.2	Quantitative metrics	34
3.6.2.1	Inception Score	34
3.6.2.2	Fréchet Inception Distance (FID)	36
4	Experiments and Results	39
4.1	Qualitative evaluation	39
4.1.1	Visual Turing Test	39
4.1.2	Quality vs training	40
4.2	Quantitative evaluation	41
4.2.1	DCGAN training stability	41
4.2.2	Fréchet Inception Distance	42
4.3	Sinogram reconstruction	43
5	Conclusions and Future Work	45
5.1	Conclusions	45
5.2	Future work	46
Bibliografía		47

List of Figures

1.1	The Guerrilla Analytics workflow	5
1.2	Gantt chart with project planning	6
2.1	Line of response in PET data acquisition.	7
2.2	Fourier Slice Theorem	8
2.3	Flow chart of iterative image reconstruction scheme.	9
2.4	Taxonomy of common data augmentation techniques.	13
2.5	Affine and pixel-level transformations applied on breast PET scans to perform data augmentation.	13
2.6	Example images augmented by RandAugment (N=2, M=9,17,28).	15
2.7	Generative adversarial network.	15
2.8	Synthetic brain PET images generated from MRI images for two typical subjects in comparison to their corresponding real images.	16
2.9	Synthetic PET images generated by (d) LB-GAN, (e) CT-GAN, (f) M-GAN in comparison to (c) Real PET.	17
2.10	Real and synthetic brain PET images for a normal patient using DCGAN.	18
3.1	Workflow for the augmentation of sinogram datasets by the unsupervised generation of synthetic images.	19
3.2	Feynman diagram of the β^+ decay.	20
3.3	Representation of two photons impacting simultaneously on the detector ring, so that a LOR is established between them.	21

3.4	Types of coincident events. From left to right: True coincidence, random (accidental) coincidence and scattered coincidence.	21
3.5	Pet scanner.	22
3.6	Representation of the mapping between the data captured by the PET scanner in real space to Radon space for a source point.	23
3.7	Breast PET 2D slice and corresponding sinogram.	23
3.8	Sample of chest PET images captured by ACRIN 6688.	25
3.9	Images of the MAMMI detector.	25
3.10	Sample of breast PET images captured by the MAMMI detector.	26
3.11	Proposed DCGAN generator network.	28
3.12	Proposed DCGAN discriminator network.	29
3.13	KL divergence as a function of the conditional and marginal probability distributions of the generated database	36
4.1	Evolution of synthesized sinograms as a function of the number of training epochs. Real column shown for fidelity comparison.	40
4.2	Evolution of generator and discriminator loss during DCGAN network training on ACRIN-6688 database.	41
4.3	Evolution of generator and discriminator loss during DCGAN network training on ONCOVISION-MAMMI database.	42
4.4	Evolution of FID score during DCGAN network training on ACRIN-6688. . . .	42
4.5	Evolution of FID score during DCGAN network training on ONCOVISION-MAMMI	43
4.6	Evolution of FID score during WGAN-GP training on ONCOVISION-MAMMI.	43
4.7	Synthetic sinogram and reconstructed PET image for DCGAN on ACRIN-6688.	44
4.8	Synthetic sinogram and reconstructed PET image for DCGAN on ONCOVISION-MAMMI.	44
4.9	Synthetic sinogram and reconstructed PET image for WGAN-GP on ONCOVISION-MAMMI.	44

List of Tables

2.1 Statistical PET image reconstruction methods.	10
4.1 Visual Turing Test results for each studied model-database pair and evaluator. R-R: Real image recognized as real image; R-G: Real image recognized as generated image; G-R: Generated image recognized as real image; G-G: Generated image recognized as generated image.	40

Chapter 1

Introduction

1.1 Background and motivation

According to the International Agency for Research on Cancer [1], breast cancer is the most common type of non-skin cancer and leading cause of mortality among women in the European Union (EU-27) as of 2020. It is estimated that 1 in 11 women in the EU-27 will develop breast cancer before the age of 74 [2].

Unlike other types of cancer, early detection of breast cancer can lead to a better prognosis and a less radical treatment, therefore contributing to a considerable reduction of long-term mortality [3, 4]. Nevertheless, some studies suggest that the continued and abusive use of diagnostic procedures (overdiagnosis) may introduce certain adverse effects due to the nature of the techniques involved [5].

Public education and awareness is critical to promote the early detection of cancer [6]. Some of the most common breast cancer diagnostic procedures include manual self-palpation, biopsies, and various medical imaging techniques among which mammography stands out [7].

Among all diagnostic procedures, self-palpation stands for a simple and approximate test, and biopsies are highly invasive. Medical imaging makes it possible to study the morphology, structure, metabolism and functions of the breast by means of minimally invasive procedures and with an optimization of diagnostic resources [8]. Some of the most commonly used medical imaging techniques for the study of breast cancer are X-ray mammography [9], ultrasound [10], magnetic resonance imaging (MRI) [11], positron emission tomography (PET) [12] and positron emission tomography/computed tomography (PET/CT) [13, 14].

Positron emission tomography (PET) has clear advantages over other imaging techniques in that it makes it possible to visualize the body's metabolic processes, and thus the presence of metastases. The patient is initially administered a radiopharmaceutical combined with glucose,

generally 18-fluorodeoxyglucose (FDG) [15], which is naturally distributed throughout the body. This radiotracer will be distributed in large proportion to the cancerous areas, which have a high glucose consumption. Using detectors placed radially in reference to the patient, the photons resulting from the annihilation of the positrons emitted by the radioisotope nucleus and the electrons in the tissues will be detected. Next, from this data, reconstruction models will provide the final medical image [16].

Artificial intelligence has proven to be a great working methodology with incredible applications in the medical field [17, 18]. One of the direct applications of this technology is the reconstruction of PET images from raw data [19, 20]. To train these models a large volume of training data is required, which is the main bottleneck of these procedures. Due to the use of radioactive tracers and the sensitive nature of the images, there is not a large enough database of breast PET data to train efficiently the reconstruction models. Some solutions proposed for the generation of synthetic data are anthropomorphic phantoms [21] and PET simulation softwares. However, these procedures generate results with low variability, in addition to having high computational resources requirements.

Within Deep Learning, GANs are a framework for estimating a generative model from adversarial processes [22]. Using this type of networks, real patient breast PET data will be tried to be synthesized in order to performe advanced data augmentation processes, thus generating a large and diverse volume of training data for future AI-based PET reconstructors.

1.1.1 Framework

This Master thesis is framed under the project “*Desarrollo de software para un sistema PET de cristal continuo aplicado al cáncer de mama (DEEPBREAST)*” funded by Ministry of Science, Innovation and Universities with reference number PID2019-107790RB-C22, whose principal investigator (PI) is María José Rodríguez-Álvarez. The project is developed by the *Medical Imaging Reconstruction Group (MIRG)* of the I3M Group of Detectors for Molecular Imaging research institute (Universitat Politècnica de València - Consejo Superior de Investigaciones Científicas (CSIC)).

1.2 Objectives

In this section the main and secondary objectives needed to achieve the goals defined by the studied problem will be established.

The **main objective** of this work is to implement a GAN-based methodology for the generation of synthetic data to serve as a training dataset for AI-based PET reconstruction models.

The **secondary objectives** of this work establish the understanding of the domain of knowledge studied and the techniques used. Each of these objectives will be considered met when a key performance indicator (KPI) is met.

- Design and implementation of different GAN architectures (DCGAN and WGAN-GP).
KPI: Functional scripts capable of generating results.
- Documentation of the technical details of the different medical diagnostic imaging tests.
KPI: Comparison of pros and cons of PET and its influence on breast cancer diagnosis.
- Study of the influence of hyperparameters on the generative architectures.
KPI: Documentation in original papers and tests with variations of hyperparameter values.
- Comparison of the generated synthetic data with ground truth.
KPI: Generation of synthetic data and calculation of quantitative metrics.
- Research and implementation of models for reconstruction of sinograms to PET image.
KPI: Functional scripts for reconstruction of real and PET-generated sinograms.

1.3 Personal motivation

As far as I can remember, I have always been intrigued by what is surrounding us: finding patterns in nature, establishing cause-effect relationships, aiming for the optimal solution to logic problems, or sharing my trivial discoveries with anyone who happened to be nearby at that particular instant.

Even though the early interest and focus on the scientific environment, and achieving a degree in Physics, I found in Data Science an interdisciplinary field that allowed me to combine my interests in programming, solid background in mathematics, and a wide range of fields of knowledge that I was willing to explore.

My relationship with Data Science and Artificial Intelligence has always been linked to the AI4Good and AI4All movements. Their main motivation is to democratize AI through the education of society, in addition to the development of projects and applications with social impact. These technologies promise more than ever to bring the advances of modern society closer to the most disadvantaged regions around the world.

In August 2019 I joined Saturdays.AI, a non-profit association-school whose values and actions are symbiotic with the motivations mentioned above. Leading the local section, I believe I was very lucky to be enrolled in the development of different AI projects in the fields of mental health, hospital emergency management and the achievement of the Sustainable Development Goals.

Over time Saturdays.AI has become one of the reference communities in Artificial Intelligence in Spanish-speaking countries.

I personally consider the field of medicine to be one of the most promising data-driven future scenarios. Its potential is so great that some curricular modifications have already been proposed in the training of residents to include a specific module on Artificial Intelligence and Data Science.

By using Deep Learning, and in particular Generative Adversarial Networks, I intend to contribute in the development of new techniques that may favor the early detection of breast cancer, an essential condition to maximize the survival rate of this disease.

1.4 Methodology

This section defines a data mining methodology appropriate to the problem to be solved that establishes the different guidelines and procedures to be followed.

The main characteristic that defines this project is its **academic research nature** at the frontiers of the state-of-the-art.

In addition, the project proposes the development and training of its own artificial intelligence architectures, which makes it a **software development** project.

Finally, the project will aim to define a standard framework for data augmentation in medical imaging environments with a **low volume of training data**.

Because we are at the frontier of state-of-the-art, we will be working in a very dynamic environment with constant variation in the techniques to be used and susceptible to the discovery of more appropriate procedures. In addition, we must be open to the appearance or generation of new data at any point in the project. Finally, due to the availability (or not) of the supercomputers that will run the models, we have to be prepared to work with dynamic resources.

For all these reasons, **Guerrilla Analytics** [23] is considered to be the most appropriate data mining methodology for this project.

This model is suitable for projects with highly dynamic data, highly dynamic requirements and highly dynamic resourcing, where both time and tools are limited.

The basic Guerrilla Analytics workflow shown in Figure 1.1 is based on the same structure as most data mining models: Domain understanding, Data understanding, Data preparation, Modeling, Evaluation, and Deployment. However, what characterizes this workflow is that it includes possible disruptions that may cause an abrupt jump between phases.

This workflow makes this methodology highly iterative with strong data dynamics. However,

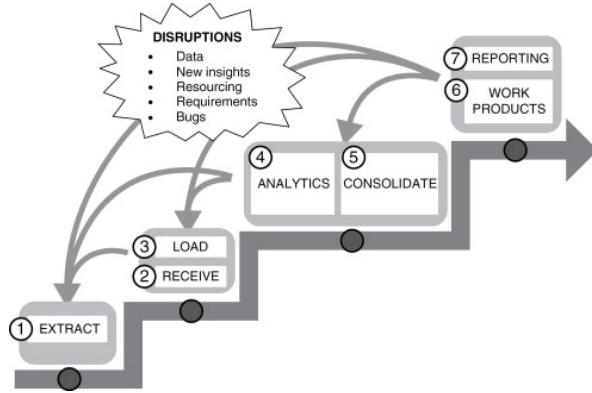


Figure 1.1: The Guerrilla Analytics workflow [Source: Ridge, 2014 [24]]

despite their dynamism, the analyses performed must be evaluable and reproducible.

In relation to this work, each of the steps is developed as follows:

- **Extraction.** Processes related to the acquisition and preprocessing of PET images to obtain the initial sinograms.
- **Receive/Load.** Loading of the sinograms on the networks and testing of the models.
- **Analytics/Consolidate.** Workflow management processes. Analysis of generated synthetic data and comparison with ground truth.
- **Work Products/Reporting.** Publication of the results, workflow, and dataset generated in public repositories and scientific research journals.

The fundamental principles of **Guerrilla Analytics** state that:

1. Space is cheap, confusion is expensive.
2. Prefer simple, visual project structures over heavily documented and project-specific rules.
3. Prefer automation with program code over manual graphical methods.
4. Maintain a link between data on the file system, data in the analytics environment, and data in work products.
5. Version control changes to data and program code.
6. Consolidate team knowledge in version-controlled builds.
7. Prefer analytics code that runs from start to finish.

In order to meet these premises, **GitHub** will be used as version control tool.

1.5 Planning

This section presents a schedule of the different tasks that will be performed throughout the development of this work. The estimated deadlines for the completion of each task and the simultaneity of subtasks should be taken into account for an optimal time management.

To visually represent this information, a **Gantt chart** [25] will be used; a type of visualization widely used in the context of project management.

The temporality will be divided with weekly granularity into 18 weeks, starting on 17 February 2021 and ending on 25 June 2021.

The main tasks will correspond to the partial deliveries of the memory defined by the university planning. The secondary tasks will be specific to the work, and will admit some minor modifications according to the changing flow of events.

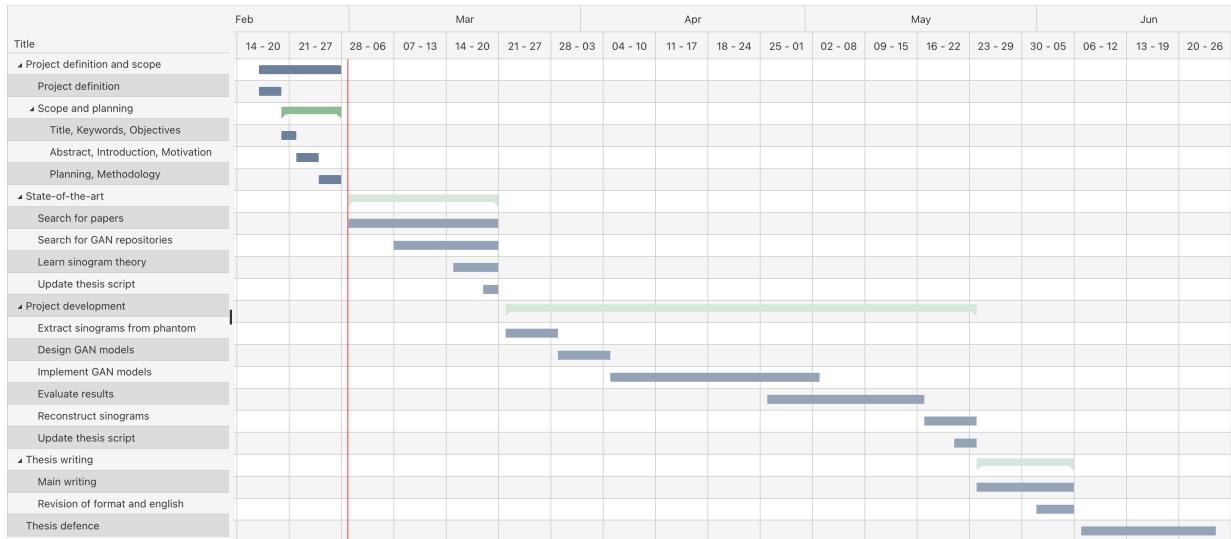


Figure 1.2: Gantt chart with project planning

Chapter 2

State of the art

Academic research related to AI-based technologies in the context of medical imaging is on trend [26]. Currently it can be found from the most formal and theoretical proposals that open the field into the unknown [27], to practical applications integrated in real clinical workflows [28].

Medical image reconstruction is one of the most fundamental and important components of medical imaging, whose major objective is to acquire high-quality medical images for clinical usage at the minimal cost and risk to the patients.

2.1 Traditional PET reconstruction methods

As it will be extensively discussed in Section 3.2, in the process of performing a PET scan, a radiotracer is injected into the patient. After different nuclear processes, two annihilation photons are generated and emitted in opposite directions. These photons will be detected by opposing detectors allowing a *coincidence* event to be registered

The imaginary line that connects these two detectors is called the **line of response** (LOR) (see Figure 2.1). The event is known to have taken place somewhere along this LOR.

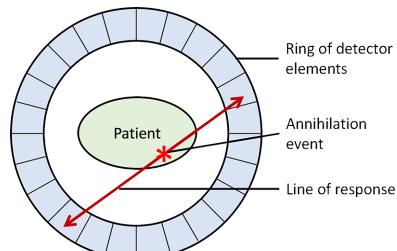


Figure 2.1: Line of response in PET data acquisition. [Source: Radiology Café].

From the set of all detections, different methods are used to reconstruct an image of the radio-tracer. These methods are classified into two groups: analytical methods and iterative methods.

2.1.1 Analytical methods

Analytical reconstruction is based on analytical inversion from the relationship between a function and its line integrals. The most used method is filtered back projection (FBP) [29].

The theoretical foundation behind FBP methods is known as the Fourier Slice Theorem [30, 31]:

Theorem 1 *The Fourier transform of the projection in the direction θ of the function f , is the Fourier transform of f on the straight line passing through the origin and having direction θ :* $\hat{P}_\theta(\omega) = \hat{f}(\omega\theta)$.

From Theorem 1, represented in Figure 2.2, it is shown that by having the projections of an image it is possible to determine the original image by performing a two-dimensional inverse Fourier transform.

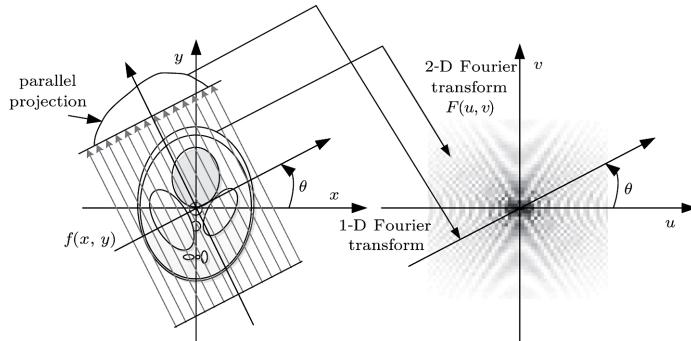


Figure 2.2: The Fourier Slice Theorem provides a relationship between the one-dimensional Fourier transform of a projection and the two-dimensional Fourier transform of the object. [Source: Jin, 2016 [32]].

The main advantages of FBP are that it is a formally consistent model and has a high computational efficiency, being suitable for applications where the number of counts is high. However, the quality of the reconstructed PET images is limited by the approximations implicit in the linear integral model on which it is based.

FBP reconstruction fails to model the statistical noise inherent to PET acquisitions. This noise is caused by different factors such as the delay of the detectors, the dead time, or the pile-up phenomenon of the pulses (when several pulses coincide and produce a single pulse that distorts the energy information and contributes to the loss of valid events).

Attenuation and dispersion correction methods exist to compensate for the factors that skew the data from its linear integral model, but they reduce its computational efficiency [33, 34].

2.1.2 Iterative methods

Iterative reconstruction methods consider the Poisson statistical characteristics of the positron emission within the source, which simplifies the problem to the standard statistical problem of estimation with incomplete data. This incompleteness in the data corresponds to the impossibility of spatially locating the annihilation events in which positrons were produced after radioactive decay.

Unlike analytical methods, whose nature makes them highly dependent on the geometry of the detector, statistical reconstruction methods can be easily adapted to non-standard geometries.

This set of statistical algorithms are mainly iterative in nature, in which, starting from a certain initial estimate, projection and back projection operators are applied to update the image of the previous step, according to a certain optimization algorithm, and in accordance with the acquired data and the system model. This workflow is presented in Figure 2.3.

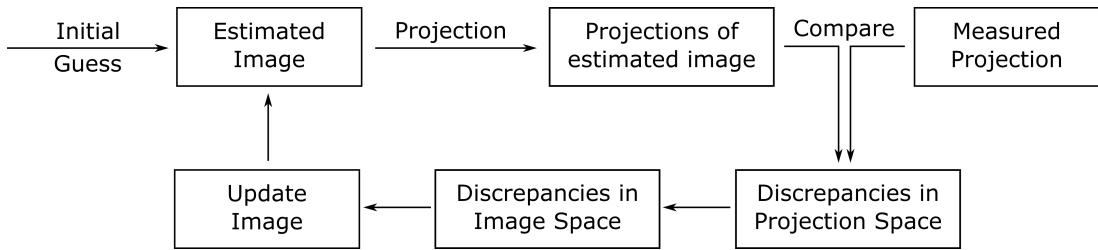


Figure 2.3: Flow chart of iterative image reconstruction scheme. [Source: Zeng, 2001 [35]].

One of the best known iterative reconstruction methods is expectation maximization (EM) [36, 37]. The name of the technique comes from the fact that in each iteration there is an expectation calculation step followed by a maximum likelihood step, using this reconstruction to make such estimates [38].

The EM algorithm is applied as an iterative technique to compute maximum likelihood estimates (MLE) from the activity density parameters [39]. Thus, the acquired data are considered as samples of a set of random variables with probability density functions related to the object distribution according to a mathematical model of the data acquisition process. In this way it is possible to calculate, for any initial density of the actual distribution, the probability that it has produced the observed data.

The EM algorithm for emission tomography provides a physically accurate model, since it allows the direct incorporation of different physical phenomena such as attenuation correction, inclusion of scatter and randoms, positron range effects, time-of-flight information [40], or information on the nature of the specific radiopharmaceutical being used, among others.

Expectation maximization algorithms have been found to produce superior results compared to

convolution-retroprojection methods [41], reducing artifacts and reducing error in the assessment of radioisotope metabolism, especially in areas of low activity [42].

On the other hand, this algorithm has an important disadvantage: its slow convergence rate; a common feature in most algebraic and iterative techniques. In addition, after a large number of iterations, noise reappears in the image; a phenomenon known as the checkerboard effect [43].

Different authors have proposed solutions to the slow convergence problem, leading to different accelerated variations of the EM algorithm, which are presented below:

The **ISRA** (image space reconstruction algorithm) reconstructs the data with a variable point spread function (PSF) [44]. This algorithm allows to reduce the required storage space by reversing the order of comparison between the projection and back projection steps.

The **WLS** (weighted least squares) algorithm is a maximization method that aims to improve the convergence rate of the EM algorithm [45].

The **SAGE** (space-alternating generalized expectation maximization) algorithm proposes to use the backprojection of image estimation to compute correction factors for updating a subset of pixels during iteration [46]. EM method is accelerated using statistical considerations.

The **OSEM** (ordered subsets EM) algorithm is an incremental gradient maximization method [47]. This set of algorithms also known as blockwise iterative divide the objective function into a sum of subfunctions, each dependent on a subset of the data. However, this method does not converge to a maximum likelihood solution, except for the case of noise-free data.

The **E-COSEM** (enhanced complete-data ordered subsets EM) algorithm is an incremental approximation of the EM method that follows the strategy of updating only one subjective function at each iteration [48]. This algorithm converges more slowly than the others in the first iterations, so it is common to start reconstructing with another algorithm and switch to E-COSEM in the last iterations.

The different iterative PET reconstruction methods are summarized in Table 2.1.

Algorithm	Class	Maximizer	Converges?	Reference
MLEM	EM	ML	Yes	(Shepp & Vardi, 1982) [37]
ISRA	Quadratic	WLS	No	(Daube, 1986) [44]
SAGE	Hidden variables	ML	Yes	(Fessler & Hero, 1994) [46]
OSEM	OS	ML	Yes	(Hudson & Larkin, 1994) [47]
E-COSEM	OS, Incremental	ML	Yes	(Hsiao et al., 2004) [48]

Table 2.1: Statistical PET image reconstruction methods. EM: expectation-maximization; OS: ordered subset, WLS: weighted least squares; ML: maximum likelihood.

As reviewed in this section, traditional PET reconstruction methods present significant drawbacks: a high dependence on the detector geometry in analytical models [49] and a low convergence speed in iterative models. In addition, the resulting images present problems and require postprocessing, such as scattering correction, normalization, or motion correction [50].

Generalist methodologies have been proposed in terms of geometry. However, these methods still require to input the manual definition of the geometry by the user; a task that requires domain expertise and consumes a lot of time [51].

2.2 Deep Learning based reconstruction

Since 2012 with the great revolution in computer vision with the introduction of AlexNet [52] and convolutional networks (CNN) [53], current state of the art for medical imaging reconstruction has been mainly focused on Deep Learning (DL) approaches [54, 55].

Traditional medical image reconstruction models are designed primarily based on human domain knowledge and a collection of hypotheses about the images to be reconstructed. The considerable increase in data volume and the improvements in the available computational resources have led to the emergence of Deep Learning-based models that allow data-driven medical imaging reconstruction, reducing the need for human intervention to a minimum [56].

The ability of DL models to extract image features from a hierarchical architecture [57] raises the possibility of using this technology for the design of reconstructors that do not require a priori knowledge of the detector geometry, or even the region over which the test has been performed.

Compared to traditional reconstruction models, with a solid mathematical basis that provides highly interpretable procedures, AI-based reconstruction models mostly work as a blackbox [55]. In order to bring these technology closer to the medical community and to gain the confidence of patients, several efforts have been made to improve the interpretability of the decisions made by these models [58].

Another major issue that arises when using AI-based techniques in medicine is the shortage of pathology data needed to train the model to perform a certain task [59]. This fact constitutes the main bottleneck when using medical image reconstruction models based on Deep Learning [60], particularly in the case of breast PET imaging.

Traditionally PET medical imaging has been used in the skull and chest regions, as these areas have a large bone structure that eases the overlap between the anatomical and metabolic structure. The use of PET with FDG is based on the principle of elevated glucose metabolism in malignant tumors. In humans, the brain accounts for about 2% of the body weight, but consumes about 20% of glucose-derived energy making it the main consumer of glucose [61].

In contrast to the above, the mammary region is composed mainly of soft tissue, with a much lower concentration of glucose. This results in a lower uptake of the administered radiopharmaceutical and therefore a considerable decrease in the number of decay counts. Due to the aforementioned complications, the volume of breast PET imaging databases is considerably lower than that found from other regions, such as the chest or the skull.

Because PET is the only medical imaging procedure that in addition to anatomical information provides metabolic information [62], efforts are being devoted to the construction of dedicated breast PET detectors [63].

Other solutions explored propose synthetically increasing the number of available samples. This procedure will constitute the cornerstone of the present work.

2.2.1 Data augmentation

Computer vision is one of the most prominent branches of Deep Learning. Following the emergence of the ResNet-152 architecture [64], for the first time in history an AI system managed to beat human performance in the ImageNet-2012 Challenge [65] on image classification [66].

Nevertheless, the networks used in computer vision are heavily reliant on big volumes of data to avoid overfitting; a phenomenon that occurs when a network learn a function with very high variance such as to perfectly model the training data [67]. However, many application domains do not have access to these data volume, which is the particular case of medical imaging [68].

Data augmentation is a solution to the problem of limited data, in which a collection of techniques are used to enhance the size and quality of training datasets. The augmented dataset will better represent the distribution of possible points, thus minimizing the distance between the training set, the validation set, and the test set [69]. Artificial inflation of the training set can be performed by either data warping or oversampling.

The taxonomy of the most common data augmentation techniques is shown in Figure 2.4.

2.2.1.1 Basic image manipulations

Data augmentation may be performed even with the simplest geometric transformations, which establish the basis for more complex techniques. Some of these transformations are [70]:

- **Flipping.** Horizontal axis flipping is much more common than vertical axis flipping.
- **Rotation.** Rotation of the image right or left on an axis between 1° and 359° .
- **Cropping.** Reduces the image size to the patch size and produces a translation effect.

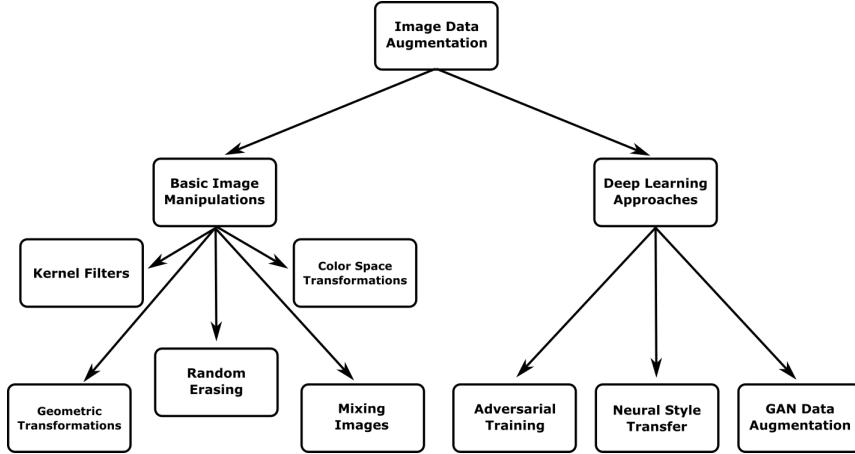


Figure 2.4: Taxonomy of common data augmentation techniques. [Source: Own elaboration based on (Shorten, 2019) [69]].

- **Translation.** Shift image position preserving size to avoid positional bias in the data.
- **Shear.** Shifts one part of the image like a parallelogram.
- **Noise injection.** Injection of a matrix of random values usually drawn from a Gaussian distribution.
- **Color space transformations.** Loop through the images and decrease or increase each RGB pixel values by a constant value. This can be useful to fix dark or light images.

The application of all these data augmentation techniques to increase the size of breast PET images datasets is shown in Figure 2.5.

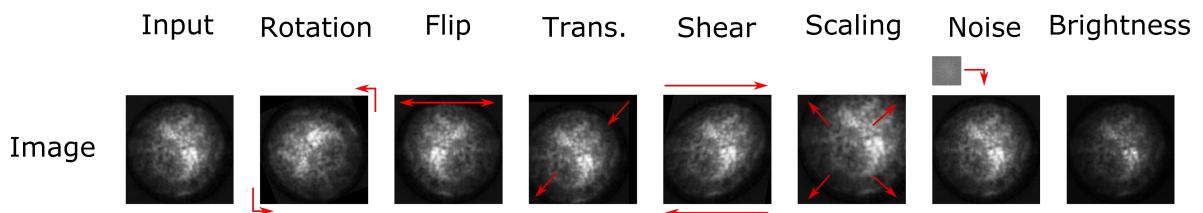


Figure 2.5: Affine and pixel-level transformations applied on breast PET scans to perform data augmentation. [Source: Own elaboration based on [70]].

When working with data augmentation, the security of the procedure must always be considered. Thus, depending on the domain of application, some transformations could result in the loss of label preservation. For example, when working with thoracic radiographs, applying flip transformations could violate the natural organs location, leading to confusion about the side on which the heart is originally located (generating false dextrocardia) [71].

Sinograms (described in more detail in Section 3.2.3) also present some limitations in the type of valid data augmentation procedures available. For example, a horizontal translation leads to

a rotation in the reconstructed image, since this axis represents the angle of the LOR. Moreover, each of the axes has a certain semantic association, and therefore a rotation or shear of the sinogram would have consequences in the reconstructed image.

In conclusion, in sinograms the concrete position of a pixel in the image has meaning, and therefore a geometric manipulation of it could lead to erroneous data. This fact highlights the importance of having synthetic data generation systems to increase the volume of data.

2.2.1.2 RandAugment

Data augmentation has led to some of the most novel advances in the field of image classification and object detection [72]. One obstacle to the adoption of these techniques on a large scale is the increased computational cost and training complexity associated with a separate search phase. It has been observed that the improvement associated with the data augmentation process is dependent on the model used and the size of the dataset. The regularization associated with each of these characteristics is not adjustable by traditional automated data augmentation techniques.

RandAugment [73] is presented in the literature as one of the best state-of-the-art data augmentation policies. Its potential lies in its simplicity and its ability to generate random input noise that will not be learnable by the models, generating an effective data augmentation.

This methodology takes an image and two integer inputs, N and M. N will indicate the number of random transformations (selected out of the thirteen available) that will be applied on the image, and M indicates the magnitude of these transformations. Thus, RandAugment is able to generate drastically different images that will result in the improvement of the learners.

For example, if we consider the case $N=2$, we can generate $13^2 = 169$ images for a certain value of M. Figure 2.6 shows some of the results of using RandAugment for $N=2$.

RandAugment can be used across different tasks and datasets, improving all previous automated augmentation approaches tested on CIFAR-10/100 and ImageNet with a significantly reduced search space (10^{30} times smaller than the next bigger data augmentation method).

2.2.1.3 Generative models

Generative models can be used to generate synthetic data that preserve the characteristics of the real data distribution [74, 75]. In particular, such models are trained to transform a random noise (and a possible class conditioning label) into a features matrix that resembles the real image. They model the conditional probability distribution between classes and features; $P(X|Y)$.

The two most common types of generative models in the literature are Variational AutoEncoders (VAE) [76] and Generative Adversarial Networks (GAN) [22].

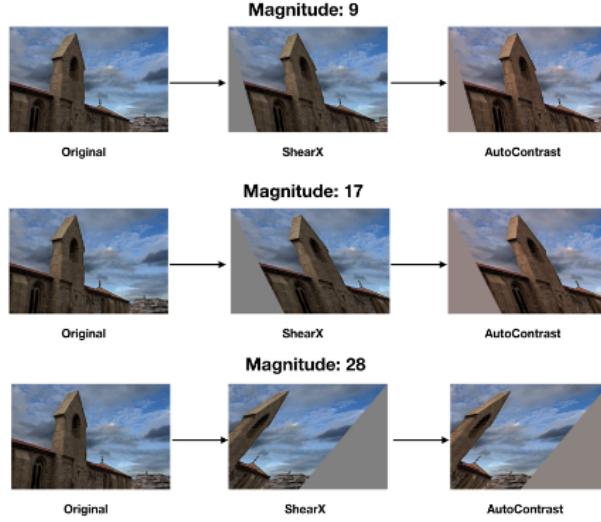


Figure 2.6: Example images augmented by RandAugment ($N=2$, $M=9,17,28$). As the distortion magnitude increases, the strength of the augmentation increases. [Source: Cubuk, 2019 [73]].

Generative Adversarial Networks (GAN) are constituted by the adversarial interaction between two independent neural networks [77, 78]:

- The **generator**, G tries to estimate the distribution that the training data came from in order to generate fake samples from the estimated distribution. Thus, $G(\mathbf{z})$ represents the generator function, which maps a latent space vector sampled from a standard normal distribution, \mathbf{z} , to data-space, \mathbf{x} .
- The **discriminator**, D , tries to discern between synthetic and real examples [79]. Thus, this network will output the probability that the generated sample came from the training data (real) rather than the generated data (fake).

The basic workflow and architecture of a GAN is shown in Figure 2.7.

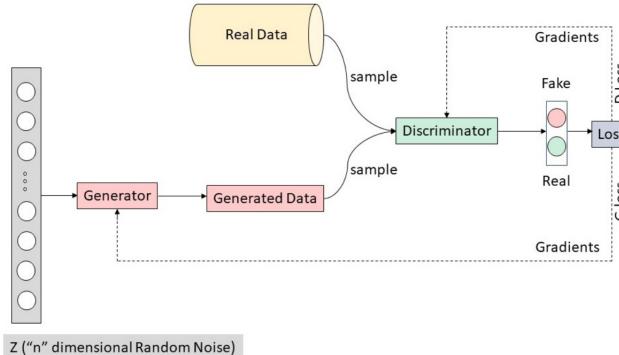


Figure 2.7: Generative adversarial network. [Source: Sagar, 2020 [78]].

The competitive nature of each network's objective will result in continuous improvement in the performance of both networks until a Nash equilibrium is reached.

During the training process, the generative network and the discriminative network will be trained simultaneously and alternately in what is known as the mini-max game until a Nash equilibrium is reached. While the discriminator tries to maximize the loss function (maximize the probability of correctly classifying real and synthetic images), the generator tries to minimize it (reduce the distance between the real distribution and the generator distribution). This behavior is shown in equation 2.1.

$$\min_G \max_D V_{\text{GAN}}(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))] \quad (2.1)$$

The optimal solution to this problem is the point at which the generator distribution equals the real distribution. However, the theory of convergence of GANs is still under study [80].

Different techniques have been proposed for the task of data augmentation using generative adversarial networks. Despite the good results shown by GANs in other fields of study, their direct application in PET imaging is not trivial. The suitability of each technique will depend primarily on the information available in accordance with its nature.

It is common in medical imaging studies to opt for a multi-modal strategy [81], where different modalities are used given the underlying correlation between them. However, in practice, it is not always possible to perform all the tests, which results in a serious problem of missing data [82]. This problem is intensified in the case of PET imaging, as some patients may refuse to undergo this test due to its high cost and exposure to radioactive substances.

Image-to-image translation learns the mapping rules between a source image and a target image. Conditional GANs (cGAN) [83] are commonly used to perform this task. CycleGan [84] is one of the best known cGANs, as it allows the training phase to be performed without the need to use paired images. These architectures have been used to transform MRI and CT images (which are easier to obtain) to PET images, allowing to augment the training datasets used by more complex models [85, 86], as shown in Figure 2.8.

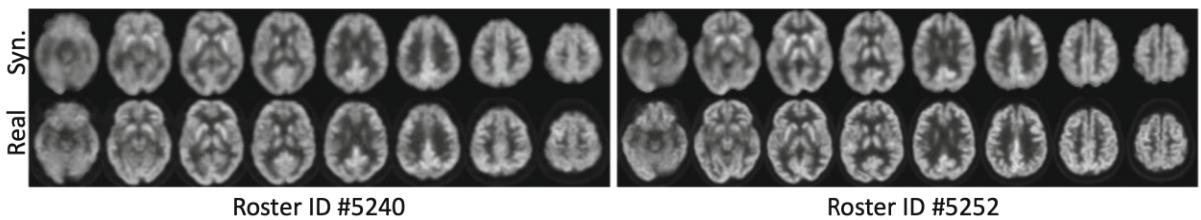


Figure 2.8: Synthetic brain PET images generated from MRI images for two typical subjects in comparison to their corresponding real images. [Source: Pan, 2018 [85]].

In (Bi, 2017) [19], the author identifies the main problems in performing different GAN-based PET synthesis processes, using mainly PET lung images. PET-to-PET is hardly able to create new variations of the input images, Label-to-PET may present lack of spatial consistency and finally CT-to-PET may not be able to synthesize the uptake regions.

The author proposes to use of a multi-channel generative adversarial network (M-GAN) architecture that integrates both label information and reference CT images in order to synthesize the high uptake and the anatomical background.

The evaluation of the results is performed using the *mean absolute error* (MAE) and *peak signal-to-noise-ratio* (PSNR) metrics and with respect to single-channel variants: LB-GAN (using labels), and CT-GAN (using CTs). The article shows how this proposed architecture achieves the best results in both metrics with respect to the other architectures.

In Figure 2.9 the comparison of the synthetic PET images generated from each of the architectures with reference to the real PET image is shown.

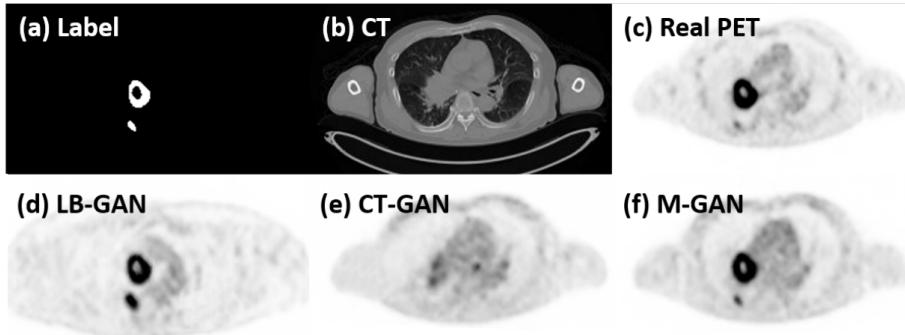


Figure 2.9: Synthetic PET images generated by (d) LB-GAN, (e) CT-GAN, (f) M-GAN in comparison to (c) Real PET. [Source: Bi, 2017 [19]].

Despite the good results obtained with these techniques, their use is conditioned to the availability of medical images of different modalities, which will not always be satisfied. In our particular case, the main objective is to obtain a large training dataset, regardless of the fact that these images may correspond to real patients.

Deep convolutional GANs (DCGANs) [87] are an improved variant of GANs where both the generator and the discriminator consist of deep convolutional networks. A large number of studies from different fields [88, 89] have traditionally opted for the use of these architectures for the data augmentation task, including the field of medical imaging [90, 78, 91].

The result of using DCGAN to generate synthetic brain PET images is shown in Figure 2.10.

Metrics to evaluate the quality-diversity trade-off in the field of GAN are domain-specific and are currently under study. In (Islam, 2020) [92], the author proposes for the task of generating new synthetic PET images of the brain from real images the use of the quantitative metrics *peak*

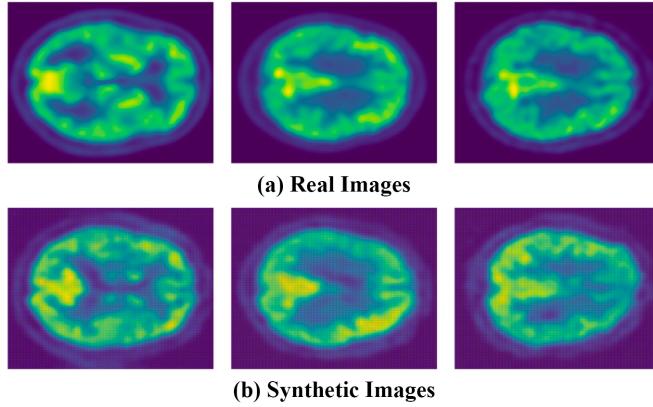


Figure 2.10: Real and synthetic brain PET images using DCGAN. [Source: Islam, 2020 [92]].

signal to noise ratio (PSNR) [93], and *structural similarity index* (SSIM) [94]. PSNR is used to measure the ratio between the maximum possible intensity value and the mean squared error of the synthetic and the real image, while SSIM finds the similarities within pixels of two images; if the pixels in the two images line up and have similar pixel density values.

The data augmentation methods described so far are not mutually exclusive. Some studies [95, 96] seem to show that data augmentation performed from the combination of synthetic data generation and the use of geometric transformations would allow a greater increase in the volume of the dataset while maintaining an improvement in the performance of the models.

Note that all the literature discussed in this section have mainly studied data augmentation techniques on brain and lungs PET imaging. At the time of writing, no studies of the application of these techniques on breast have been identified. This is probably due to the inherent complexity of using PET methodology in the breast, mainly due to a lower uptake of glucose (and therefore of the radiotracer) and the lack of hard tissue as opposed to the skull region.

This work aims to be a first approach to the use of GAN-based data augmentation in breast PET imaging. For this reason, we will rely mainly on internal evaluation metrics instead of making a comparison with other studies.

Chapter 3

Methodology

3.1 Workflow overview

The methodology proposed in this work (see Figure 3.1) covers the entire pipeline from the collection of the raw data in the clinical setting to the generation of an augmented dataset with synthetic data.

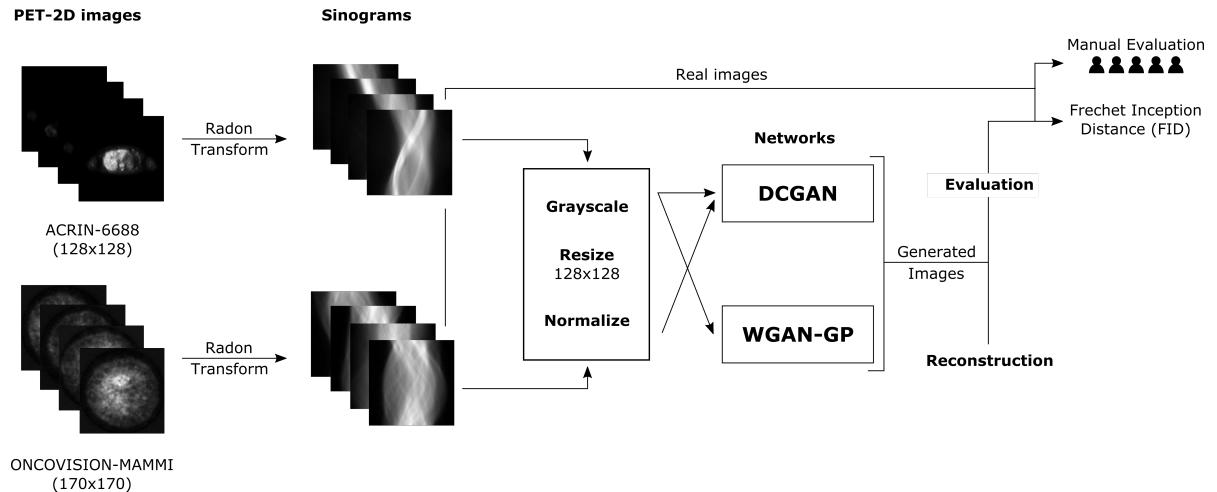


Figure 3.1: Pipeline for the augmentation of sinogram datasets by the unsupervised generation of synthetic images.

The designed protocol is not exclusive to clinical scenarios, and aims to establish the basis for the augmentation of datasets in those domains with a bottleneck in real image collection capacity.

Each of the represented phases will be described in greater depth in this chapter, exploring from the physics fundamentals of PET medical imaging to the technical details of the generative models used and the various techniques for evaluating the quality of the synthesized results.

3.2 Basic principles of PET medical imaging

Positron emission tomography (PET) is a sophisticated functional medical imaging technique capable of producing *in vivo* images of functional processes of the body. It is commonly used in patients with brain or heart conditions and cancer.

3.2.1 Nuclear interactions

Due to the nuclear nature of PET, a small amount of radioactive substance (radiopharmaceutical) with the ability to emit positrons is introduced to the patient to assist in the study of the affected tissue. These substances have a deficit of neutrons in their nucleus, reaching stability by means of a β^+ decay (see Figure 3.2) by which a proton is converted into a neutron, emitting a positron and a neutrino in the process.

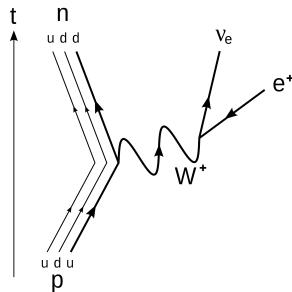


Figure 3.2: Feynman diagram of the β^+ decay.

The loss of kinetic energy by the positron when interacting with the medium leads to its annihilation with an electron. The laws of conservation of mass and energy explain the generation of two annihilation photons with the same energy (around 511 keV) traveling in opposite directions.

An event detected in a PET scanner is called a *coincidence* and consists of the interaction of a pair of annihilation photons in two crystals opposite within a time window of the order of a few nanoseconds (typically between 5 and 10 ns), since its detection must be "electronically" simultaneous to ensure they come from the same annihilation. Behind the treatment of the acquired signals it is possible to know the positions of all gamma rays incident on the crystals.

3.2.2 Line of response (LOR)

The positions corresponding to a certain coincidence are joined by an imaginary line along which is known that both photons traveled in opposite directions until reaching the detectors, the exact starting position being unknown. That is, at what point positron-electron annihilation occurred along the line. These lines where the events are recorded are called response lines or LORs (lines

of response) and constitute the basis for the reconstruction of PET images of which the position of the annihilation events is deduced (see Figure 3.3).

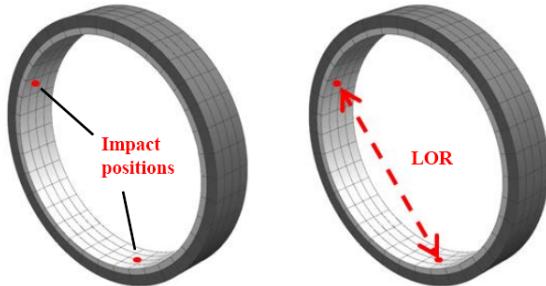


Figure 3.3: Representation of two photons impacting simultaneously on the detector ring, so that a LOR is established between them.

Coincidence events do not always represent real coincidences. Due to accidental (or random) phenomena, two photons not originating from the same annihilation event may be detected in the same LOR. Likewise, two photons originating from the same annihilation event may be detected in different LORs (Compton scattering). These effects are shown in Figure 3.4.

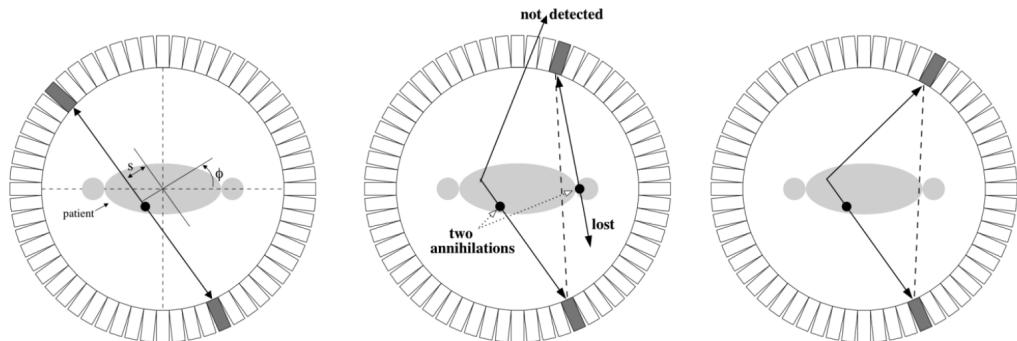


Figure 3.4: Types of coincident events. From left to right: True coincidence, random (accidental) coincidence and scattered coincidence. [Source: The Physics of PET/CT scanners, Schmitz; University of Washington].

PET data acquisition is not a straightforward process. Interactions in the patient attenuates the emitted photons, the detection efficiency of the detectors is highly dependent on the materials and configuration used, and random and scatter events interfere with the true events.

For clinically useful PET imaging all these effects must be corrected. Some of the most important processing procedures are attenuation correction, scatter correction, and normalization.

3.2.3 Sinograms

A **tomographic image** is defined as a 2D representation of the structures within a specific plane of a three-dimensional object. This image can help to discern the distribution of activity in various cross sections of the patient and provide a better quantification of the activity contained in a given point of the body.

One of the fundamental problems in conventional nuclear medicine is that the images obtained are 2D projections of 3D dose distributions. This fact results in overlapping images of structures at different depths. A possible solution is to use tomographic techniques by obtaining projections of the object at different angles, and then using mathematical algorithms (detailed in Section 2.1) to reconstruct the images from the data obtained in these projections.

In PET scanning the patient is introduced into a tubular scanner (see Figure 3.5) where the detectors are (usually) arranged in a stationary ring around the patient. N equispaced projections between 0° and 180° of the area of interest will be collected. Note that this procedure will allow us to obtain the number of events for each LOR.



Figure 3.5: Pet scanner [Source: Nova Global Healthcare]

The raw data obtained by the collection of detectors that constitute the PET scanner are processed and stored in the form of a sinogram. A **sinogram** is a grayscale display whose rows represent the projection profile for each of the angles, and the successive rows, from top to bottom, represent the projections for the different angles.

The shade of gray at each pixel in the sinogram corresponds to the counts for a particular LOR. Note that the semantic importance of every pixel in the image will impose great restrictions on classical data augmentation techniques.

To understand the formation of the sinogram a stationary reference system with respect to the detector is introduced. This coordinate system is defined as (r,s) , where the r coordinate will be orthogonal to the detector plane and the s coordinate parallel to it.

Assuming that the patient is in its own (x,y) coordinate system, and the detector is rotated an angle ϕ with respect to it, the transformation equations are described in equation 3.1

$$\begin{aligned}s &= x \cos \phi + y \sin \phi \\ r &= y \cos \phi - x \sin \phi\end{aligned}\quad (3.1)$$

The set of all projection profiles collected on the 2D matrix $p(s, \phi)$ represents the Radon transform of the real PET image. This display is often called sinogram because of the sinusoidal shape that arises from a point source in the real space, as represented in Figure 3.6.

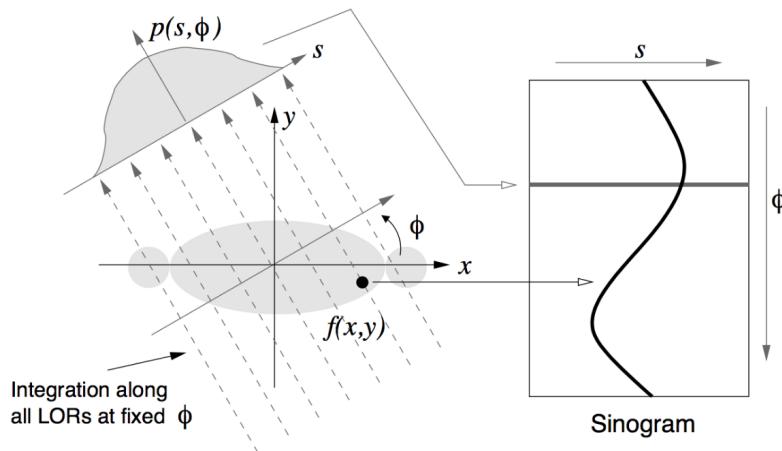


Figure 3.6: Representation of the mapping between the data captured by the PET scanner from real space to Radon space for a point source [Source: Lecture on Medical Physics, Milan Zvolský].

One of the real PET images that will be used in this work and its corresponding sinogram (Radon transform) is shown in Figure 3.7. Note that the raw information collected by the detector will be stored directly as a sinogram, being its reconstruction the final objective.

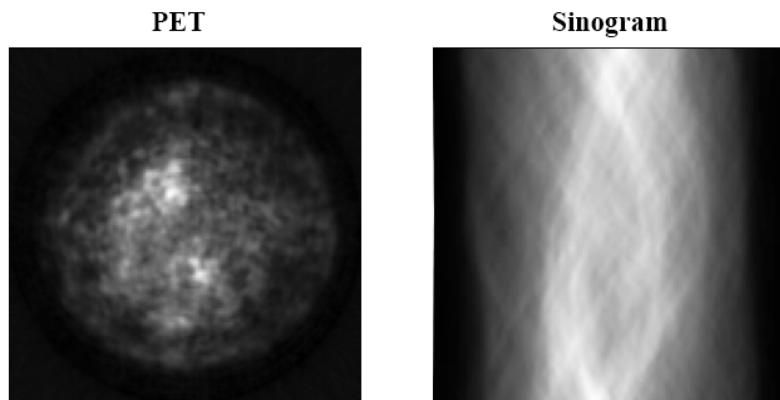


Figure 3.7: Breast PET 2D slice and corresponding sinogram.

3.3 Databases

In this work two databases will be used: one representing a quasi-ideal case in which the images have different types of corrections applied (study of the viability of the models) and the other representing a case with non-corrected images (study of the scalability to the clinical scenario).

3.3.1 ACRIN-6688 Database

The ACRIN-6688 is a database which principal objective was to correlate changes measured by ^{18}F -FLT PET imaging, a measure of cellular proliferation, in the primary tumor early during NAC (neo-adjuvant chemotherapy) with pCR (pathologic complete response) in locally advanced breast cancer patients [97, 98]. The trial also examined both pre-therapy and post-therapy association of ^{18}F -FLT uptake with the tissue proliferative marker Ki- 67 to compare ^{18}F -FLT PET/CT against an accepted reference standard for cellular proliferation.

The study participants were patients with pathologically confirmed breast cancer and determined to be a candidate for primary systemic (neoadjuvant) therapy and for whom surgical resection of residual primary tumor following completion of neoadjuvant therapy is planned.

The trial, which originated at Virginia Commonwealth University - VCU, required both FDG (fluorodeoxyglucose) and FLT (fluoro-3-deoxythymidine) imaging on separate days. The protocols were different for each tracer:

- **FDG protocol:** Following injection of FDG, a one hour uptake period occurred prior to whole body (WB) multiple field-of-view (FOV) imaging. This imaging session will have three series consisting of a CT and two emission images for FDG (an attenuation-corrected image or PET AC, and a non-attenuation corrected image series or PET NAC).
- **FLT protocol:** ACRIN 6688 began by requiring dynamic FLT imaging followed by WB scanning. Concurrent with injection, a 45 frame dynamic imaging sequence was acquired over 1 FOV for one hour. The WB scan occurred within 10 minutes of the completion of the dynamic study. Thus, there are six series associated with this imaging session, namely three single FOV series (PET AC dynamic, the CT, the summed 30-60 min PET AC) and three WB series (CT, PET AC and PET NAC).

An example of the 128×128 pixel sized images contained in the dataset is shown in Figure 3.8. A total of 42980 images from this database were used.

The images represent slices of corrected chest PET/CT and will constitute the quasi-ideal case in which the images have different types of corrections applied.

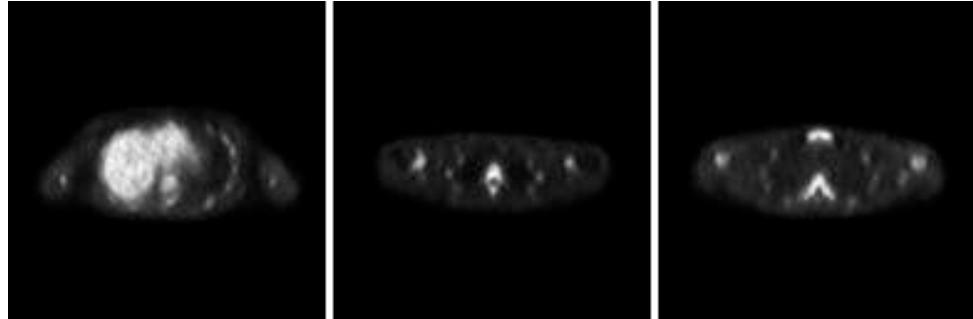


Figure 3.8: Sample of chest PET images captured by ACRIN 6688.

3.3.2 ONCOVISION-MAMMI Database

The MAMMI prototype is the result of an European project whose main members were the Spanish National Research Council (CSIC), the National Cancer Institute Antoni van Leeuwenhoek (NKI), the Technical University of Munich (TUM), the Karolinska University of Stockholm and the Spanish company Oncovision. The main objective of the MAMMI project was the investigation of early stage breast cancer through the development of a dedicated PET prototype dedicated to the exploration of this organ [99].

In order to obtain the images, the breast is placed inside a PET detector ring close to the organ, with a much larger solid angle between the breast and the detector than in whole-body systems. This allows an increase in detection sensitivity and therefore an increase in image quality. Figure 3.9 shows the equipment configuration and patient positioning for image acquisition.

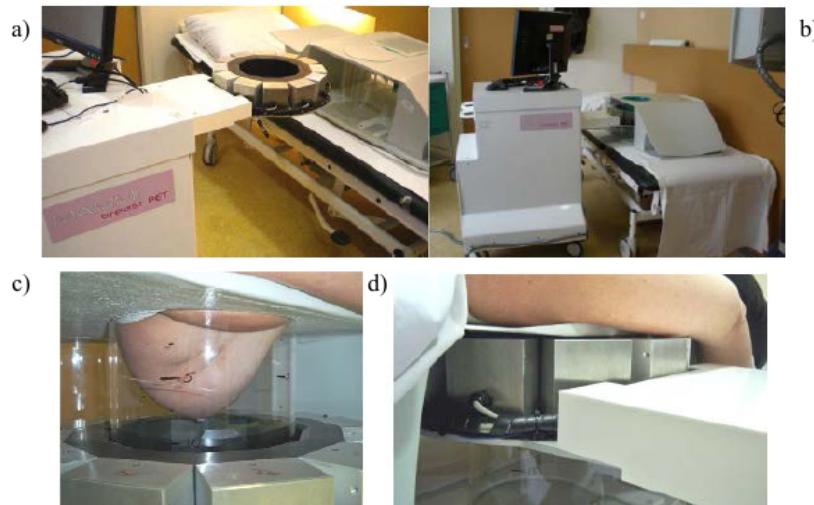


Figure 3.9: Images of the MAMMI detector. a) Detector ring. b) Housing in which the electronics are placed and specially designed stretcher to position the patient. c) Position of the breast during a scan. d) Positioning of the patient during the scan. [Source: Moliner, 2014 [99]].

The MAMMI detector ring has a transaxial FOV of 170 mm, which covers 98% and 95% of patients respectively, depending on the study considered. The amount of radiopharmaceutical to be injected into the patient in a MAMMI scan is around 2 mCi, so that the doses delivered to patients are considerably reduced and the scan can be repeated if necessary.

The ring consists of twelve detector modules connected to the acquisition electronics via the lifting arm. Each has a monolithic LYSO scintillator crystal coupled by optical grease (Rhodorsil Pâte 7) to a Hamamatsu Photonics H8500 position sensitive photomultiplier tube (PSPMT). The use of monolithic crystals versus the most common pixelated crystals means a reduction in equipment cost, as each detector pixel does not have to be cut, polished and glued to each of the gamma detectors.

The crystals have a truncated pyramid shape so that they fit together completely, reducing the dead areas in the detection. The crystal is 10 mm thick, its front detection area is (40×40) mm 2 and the PSPMT is coupled to an area of (50×50) mm 2 . All surfaces, except the one coupled to the PSPMT which is polished, are roughened and painted black to avoid unwanted internal reflections that degrade the spatial resolution and increase the effect of image compression.

MAMMI works in 3D acquisition, so that coincidences are allowed between a detector and its opposing modules, resulting in a transaxial FOV of 170 mm. The size of the detector module restricts the axial FOV to 40 mm per frame, but by means of the detector ring elevator, an axial FOV of up to 170 mm, since it is able to move in 34 mm steps, overlapping 6 mm in sequential acquisitions. After the reconstruction process these frames are joined together to form the whole breast image. The window of energies in the acquisition typically used is 350 to 650 keV.

An example of the 170×170 pixel sized images contained in the dataset is shown in Figure 3.10. A total of 37450 images from this database were used.

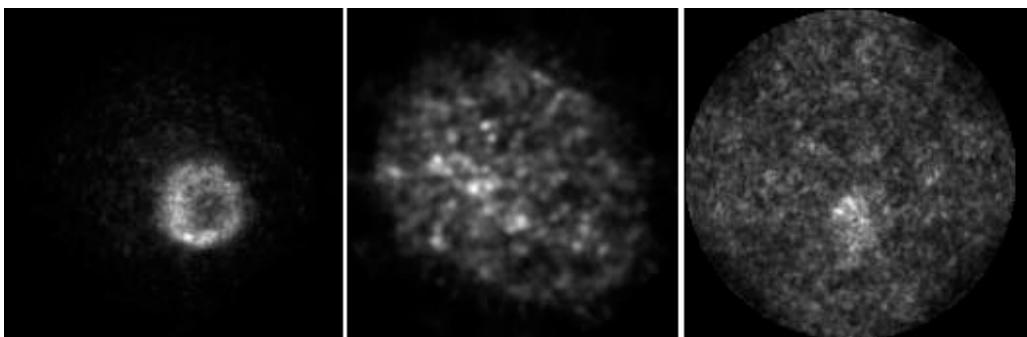


Figure 3.10: Sample of breast PET images captured by the MAMMI detector.

Note the variability in the available images, which reflects both the size of the breast and the different intensities of radiotracer absorption.

3.4 Preprocessing

The datasets shown so far consist mainly of raw PET data in binary list-mode format, while for the correct functioning of the networks it is necessary to work directly with the corresponding sinogram images.

For this it will be necessary to preprocess the raw data, for which the following workflow will be followed:

1. First, the volumetric PET image is taken and divided into different bidimensional slices.
2. Using the scikit-learn library, the Radon transform operation is applied on the matrices that characterize each of the measurements. In this way it obtains the respective sinograms, which are the necessary inputs for our GAN.
3. The values of the new matrices are normalized in the range [0, 255], which will allow to obtain a Grayscale representation of the measurement.
4. The images are resized to the input size of the network (128×128 in this case).

Note that the values of the generated images will be in the range [-1, 1] due to the use of the hyperbolic tangent activation function. These values will be rescaled to the range [0, 255] for correct representation and reconstruction.

3.5 Network architectures

Since the first GAN models, a large number of modifications have been proposed addressing the limitations of previous models and improving the quality and diversity of the generated outputs.

In the field of unconditioned image generation, the main networks that stand out historically for their high computational performance and ease of implementation and training are Deep Convolutional GAN (DCGAN) and Wasserstein GAN (WGAN).

A basic knowledge of GAN related concepts (reviewed in Section [2.2.1.3](#)) will be assumed throughout this section.

3.5.1 Deep Convolutional GAN (DCGAN)

Deep Convolutional GAN proposes building the generator and discriminator networks of the GAN structure based on convolutional neural networks (CNN). Thus, it is possible to combine the good performance of CNNs in supervised tasks with the unsupervised nature of GANs.

One of the main limitations of the vanilla GAN architecture is its high instability and the sporadic generation of low-fidelity results. For this reason, DCGAN introduces some fundamental modifications that allow both to improve the convergence speed of training and to improve the quality and diversity of the generated synthetic results:

- The filter size is **doubled** at each convolutional layer.
- All pooling layers are replaced by **strided convolutions** (discriminator) and **fractional-strided convolutions** (generator).
- **Batch Normalization** is used at each layer (except for the input layer) as it reduces the covariance shift.
- **LeakyReLU** is used as the main activation function for the discriminator as it helps to avoid the vanishing gradient effect except for the output layer, which uses sigmoid.
- **ReLU** is used as the main activation function for the generator except for the output layer, which uses tanh.

The original DCGAN paper proposes an architecture designed to generate images of size 64×64 with 3 color channels (RGB). Due to the characteristics of the PET image being worked on, both the input image and the output image are required to be of size 128×128 with 1 color channel (Grayscale). The addition of an extra layer in both the generator and the discriminator has been proposed to introduce these modifications.

The generator and discriminator models proposed for the DCGAN structure are shown in Figure 3.11 and Figure 3.12 respectively.

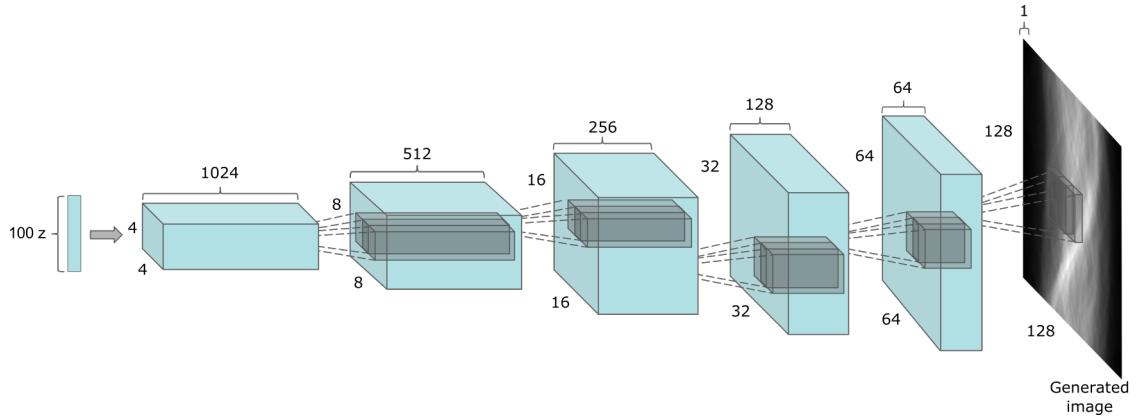


Figure 3.11: Proposed DCGAN generator network. [Source: Own elaboration based on [87]].

In the generator network, each of the intermediate layers is composed of a combination of transposed convolution, Batch Normalization and LeakyReLU layers. A kernel size of $k = 4$,

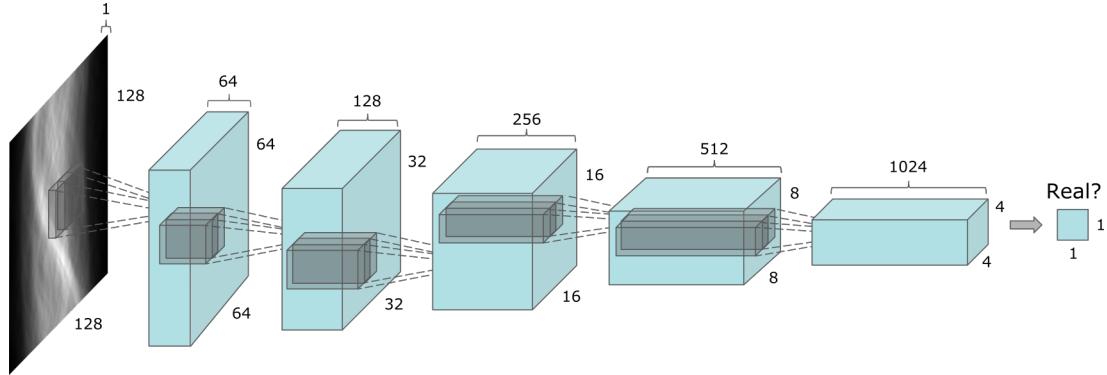


Figure 3.12: Proposed DCGAN discriminator network. [Source: Own elaboration based on [87]].

stride of $s = 2$, and padding of $p = 1$ are used. This configuration leads to a doubling of the image size in each layer while the number of filters is halved. In the last layer the number of filters is adjusted so that the generated image has the desired number of channels.

All network configurations are done in accordance with the original DCGAN paper [87]:

- The **weights** are randomly initialized from a normal distribution $\mathcal{N}(\mu = 0, \sigma = 0.02)$.
- The generator and the discriminator will have individual optimizers, being in both cases the **Adam optimizer** with momentum $\beta_1 = 0.5$ learning rate $l_{rG} = 0.0002$ for the generator and $l_{rD} = 0.00005$ for the discriminator.
- The loss function used is **Binary Cross-Entropy**. The result will be bounded between 0 and 1 and will indicate the degree of confidence that the generated image corresponds to the set of real images.

DCGAN training will occur intermittently between the generator and the discriminator. The training should be given in such a way that both improve their performance at the same rate, as a discriminator that is too good will not provide useful information for the improvement of the generator.

GANs are very sensitive to hyperparameters due to their low stability. Their value will be determined by the recommendations of the original paper, obtained by empirical review. A batch size of 128, an input size of the noise vector in the generator of 100, and 120 epochs will be used. Note that especially in the context of GANs a higher number of training epochs is not directly related to better synthetic results. For this reason the whole training process will be qualitatively monitored by manual inspection.

Despite representing a major step forward in the field of synthetic models, DCGAN suffers from some limitations directly related to the nature of the constituent techniques:

- **Checkerboard pattern** appears in the synthetic images as a consequence of the use of transposed convolutions.
- **Mode collapse** is a GAN phenomenon whereby the GAN is unable to generate images of different classes (lacks diversity) despite having been trained to do so. This occurs mainly when the generator identifies that the generation of a particular class fools the discriminator much more easily, thus reinforcing this behavior. Because there is no incentive for the generator to switch things up, the entire system will over-optimize on that one output.
- **Gradient vanishing** is a problem that arises from the limitations of the cost function used. As the discriminator confidence value approaches 0 or 1, the computed gradients approach zero and as a result, the generator is not able to obtain value information to improve its performance.

3.5.2 Wasserstein GAN (WGAN)

In the previous section, the main problems derived from the DCGAN model have been identified, which are mainly a consequence of the loss function (BCE).

(Arjovsky, 2017 [100]) proposes using Earth's Moving Distance metric (EMD), which measures the distance between two distributions. In particular, it measures the degree of difference between two distributions by estimating the effort it would require to construct the generated data distribution from the real data distribution.

W-Loss is an approximation to EMD with some advantages over BCE, whose expression is shown in equation 3.2.

$$W_{\text{loss}} = \min_g \max_c \mathbb{E}(c(x)) - \mathbb{E}(c(g(z))) \quad (3.2)$$

Using W-Loss, the output value of the discriminator is not limited to the range [0, 1], so for consistency of nomenclature, it is renamed as the *critic* in this context. This behavior of the loss function makes W-Loss much less prone to the vanishing gradients and mode collapse problems experienced by BCE.

While the generator will try to reduce the distance between distributions, the critic will try to increase it, re-encountering an adversarial training until a Nash equilibrium point is reached.

The EMD approximation process from W-Loss leads to the Kantorovich-Rubinstein duality [101], which is able to reformulate this distance as a solution to the maximization over 1-Lipschitz functions, turning the Wasserstein GAN optimization problem into a saddle-point problem [102].

Thus, in order to work with the Wasserstein Loss, the critic must be 1-L continuous, which means

that the norm of the gradient must be at most 1 for each point,

$$\|\nabla f(x)\|_2 \leq 1 \quad (3.3)$$

To enforce 1-L continuity there exist mainly two methods: weight clipping and gradient penalty. Each will represent an independent case study in this work.

3.5.2.1 Weight Clipping

With weight clipping, the weights of the critic are forced to take values between a fixed interval.

Once the weights have been updated after the gradient descent step, all weights whose value is outside the desired range are brought to the limits. If a weight is too low or too high, its value will be brought to the minimum or maximum of the allowed range, respectively.

This way of enforcing 1-L continuity could overly limit the critic's ability to learn or find an optimal state. In the opposite case, considering a wide clipping interval may not be sufficient to ensure 1-L continuity. Thus, a new hyperparameter controlling the clipping range would be added to the hyperparameter set.

3.5.2.2 Gradient penalty

With gradient penalty, a regularization term is added to the loss function which penalizes the critic when it's gradient norm is higher than one,

$$\min_g \max_c \mathbb{E}(c(x)) - \mathbb{E}(c(g(z))) + \lambda \phi_{\text{reg}} \quad (3.4)$$

where λ is a hyperparameter that weights the importance of the regularization term relative to the cost function. In the implementation studied in this work, a value of $\lambda = 10$ has been used.

Due to technical limitations, the gradient penalty estimation consists of taking one image from the set of real images and one image from the set of fake images, and performing an interpolation between them to obtain an intermediate image. The weight of each of these images is controlled by a random parameter ϵ , thus generating an interpolated image \hat{x} according to equation 3.5.

$$\hat{x} = \epsilon x + (1 - \epsilon) g(z) \quad (3.5)$$

The intermediate image is evaluated by the critic and the gradient of the critic's predictions on

it is obtained. The norm of this gradient is then determined, and an additional factor evaluates the quadratic distance between the unit and the excess value. Equation 3.6 shows the final expression for the gradient penalty regularization factor,

$$\phi_{\text{reg}} = (||\nabla c(\hat{x})||_2 - 1)^2 \quad (3.6)$$

The expression for the loss function used for training with W-Loss is shown in equation 3.7,

$$\min_g \max_c \mathbb{E}(c(x)) - \mathbb{E}(c(g(z))) + \lambda \mathbb{E} (||\nabla c(\hat{x})||_2 - 1)^2 \quad (3.7)$$

where the first part approximates the EMD (making the model less prone to mode collapse and vanishing gradient phenomena) and the second part is a regularization term that makes the loss function continuous and differentiable.

Studies have shown that although both are good solutions, gradient penalty usually provides better results [103]. For this reason, WGAN-GP will be the main case study within the Wasserstein GANs family.

As in DCGAN, the network architecture must be adapted to be able to work with the 128×128 sized images used in this work. The final architecture is equivalent to the one presented in Section 3.5.1 and figures 3.11 and 3.12.

An additional hyperparameter is added to control the number of training iterations of the critic versus each training iteration of the generator. Thus, the rate of improvement of each network is balanced. In this case we set 5 iterations of the critic for each iteration of the generator.

Finally, based on the recommendations proposed in the original papers, the value of the Adam optimizer moment parameter will be set to $\beta_1 = 0$ and Instance Normalization will be used instead of Batch Normalization.

3.6 Evaluation metrics

Evaluating GAN performance is not a trivial task [104]. Despite the great theoretical progress in the field, there is no consensus on the best metrics capable of capturing the strengths and weaknesses of GANs in order to perform a fair comparison between models.

In contrast to most Deep Learning models, which use a loss function to be trained until convergence, the adversarial nature of GANs implies the search for a Nash equilibrium [105] between the generator and the discriminator, stripping the model of an objective loss function.

Because of these characteristics, GAN models must be evaluated using the quality of the generated synthetic images. With their increasing use of numerous real-life applications, different quantitative and qualitative metrics have been proposed, each with its pros and cons [106].

It is possible to establish some essential characteristics [107] that an effective GAN metric should possess:

1. It should favour models that create distinguishable generated samples from real ones.
2. It should be sensitive to overfitting of the model.
3. It should be able to control with disentangled latent spaces as well as space continuity.
4. It should have well-defined boundary values.
5. It must be sensitive to image distortions and transformations.
6. It must agree with human perceptual judgments and human rankings of models.
7. It must have a low computational complexity.

For the aim of this work, two evaluation procedures have been chosen: a qualitative evaluation (Visual Turing Test), and a quantitative evaluation (Fréchet Inception Distance).

3.6.1 Manual evaluation

One of the simplest, most direct, and intuitive ways to evaluate GAN performance is to take advantage of the great visual identification capacity of the human brain.

By generating a batch of synthetic images, human evaluators try to estimate the quality and diversity of the images in relation to the target domain. This human capability is the main motivation behind some technologies such as Amazon Mechanical Turk [108].

Traditionally, this type of evaluations serve as a qualitative tool to monitor the training process of generative networks, being ad-hoc strategies and lacking standardization and validation.

Some approaches such as the Human eYe Perceptual Evaluation (HYPE) [109] propose to standardize the human evaluation workflow so that it is (1) grounded in psychophysics research in perception, (2) reliable across different sets of randomly sampled outputs from a model, (3) able to produce separable model performances, and (4) efficient in cost and time.

In particular, this solution evaluates people's ability to estimate the veracity of images both under and without temporal perceptual thresholds.

Due to the simplicity of the solution, numerous problems emerge:

- It is subjective, including biases of the reviewer about the model, its configuration, and the project objective.
- It requires domain expertise about what is real and what is not.
- It is limited to the number of images that can be reviewed in a reasonable time.

As (Algahtni, 2018 [107]) highlights: "*Evaluating the quality of generated images with human vision is expensive and cumbersome, biased, [...] difficult to reproduce, and does not fully reflect the capacity of models*".

The manual evaluation of the datasets used in this work requires prior familiarity with sinograms or, failing that, specific prior training. In addition, the evaluation resources available are limited.

3.6.1.1 Visual Turing Test

A Visual Turing Test was conducted in which a group of five people tried to evaluate the veracity of a batch of images for each dataset.

The group was composed of people with no previous domain expertise who were initially exposed to 20 real images. A total of 60 images where tested for each dataset by selecting 30 real and 30 generated images. The proportion of real and generated images was unknown to the evaluators. All reviewers were tested independently on the same sets of test images, with no communication between them.

In order to make the results of the Visual Turing Test reliable, high-quality generated images were mixed with low-quality real images for comparison and resized all images to 64×64 .

3.6.2 Quantitative metrics

There are numerous quantitative metrics that, focusing on different characteristics of the generated image, are able to calculate specific numerical scores used to summarize the outputs of generative models.

Two quantitative metrics frequently used for the evaluation of generated images are the Inception Score (IS) [110] and the Fréchet Inception Distance (FID) [111].

3.6.2.1 Inception Score

Inception score was a widely used quantitative metric used to evaluate the results of generative models. This metric introduced the idea of using pre-trained classifiers to evaluate GANs, although its problems led to its almost complete replacement by FID.

This section will provide an introduction to its mechanisms due to its historical relevance and its role in establishing the theoretical background for FID.

This score simultaneously evaluates two characteristics of the generated image distribution:

1. Its **diversity**, meaning that the model is able to generate images belonging to the different classes for which it has been trained.
2. Its **fidelity**, meaning that the model is able to generate realistic images in relation to the ground truth for the represented class.

To understand Inception Score we must understand the probability distributions involved:

- The **conditional probability** $p(y|x)$ is related to quality, and indicates the degree of predictability of the represented class given an image. Ideally we are looking for high predictability (low entropy).
- The **marginal probability** $p(y)$ is related to the diversity, and indicates the class distribution for the set of generated images. Ideally we are looking for high diversity (high entropy), which is related to a uniform data distribution.

The InceptionV3 classifier [112], which gives its name to the metric, is used to obtain each of these probability distributions. The details of this classifier are not relevant to the theoretical scope of this indicator.

To ensure the good performance of the GAN the two distributions should be as dissimilar as possible. To evaluate the dissimilarity between two probability distributions the Kullback-Leibler (KL) divergence will be used, defined as the entropy difference between both distributions:

$$D_{KL}(p(y|x) \parallel p(y)) = p(y|x) \log \left(\frac{p(y|x)}{p(y)} \right) \quad (3.8)$$

Figure 3.13 shows the qualitative performance of the KL divergence.

To obtain the final Inception Score in a range of values that is easier to work with, we take the exponential of the mean of the KL divergences for all images,

$$I_S = \exp(\mathbb{E}_{x \sim P_g} D_{KL}(p(y|x) \parallel p(y))) \quad (3.9)$$

If both diversity and fidelity are true, the score will be high. If either or both are false, the score will be low.

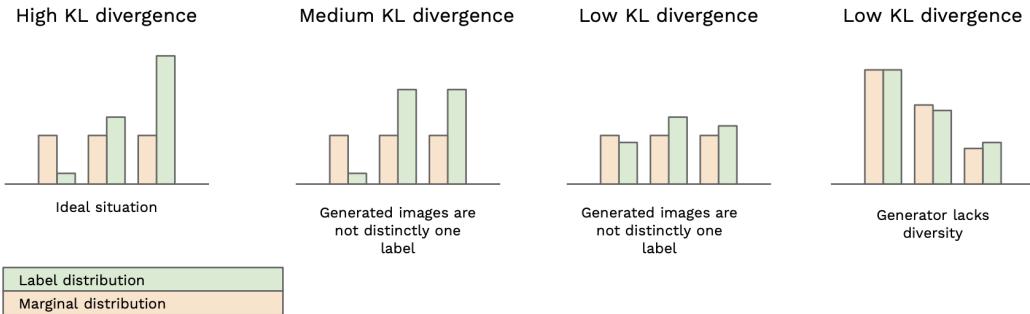


Figure 3.13: KL divergence as a function of the conditional and marginal probability distributions of the generated database. [Source: A simple explanation of the Inception Score; David Mack].

Despite its past dominance in the field of GAN evaluation, Inception Score has significant limitations that have led to it increasingly falling into disuse:

- The score is limited to InceptionV3 training data (commonly ILSVRC2014) limitations.
- The metric only looks at generated images, without comparing to real images.
- If the generator only generates one image per class it can score highly (there is no measure of intraclass diversity).
- If the generator overfits and replicates training data it can score highly.

3.6.2.2 Fréchet Inception Distance (FID)

Fréchet Inception Distance (FID) is a quantitative evaluation metric inspired by Inception Score that captures the similarity between generated and real images. It has been shown to be more consistent with human evaluation in assessing the realism and variation of the generated samples.

The main advantage of FID over IS is its evaluation based on the statistics of a collection of synthetic images compared to the statistics of a collection of real images from the target domain.

This metric recovers InceptionV3 classifier with some modifications with respect to IS. Instead of the probability distribution of the results, FID extracts the 2048-dimensional activations from the pool3 layer for real and generated samples respectively. This layer is located near the final nodes, being able to identify composite images but at a lower level of complexity, thus having a higher generalization capacity.

The distribution for each of these activations is modeled as a multidimensional Gaussian distribution of the form $\mathcal{N}(\mu_g, \Sigma_g)$ for the synthetic images and $\mathcal{N}(\mu_r, \Sigma_r)$ for real images where μ represents the mean and Σ the covariance matrix.

The distance between these two distributions is determined using the Fréchet distance (a.k.a Wasserstein-2 distance),

$$FID(x, g) = \|\mu_r - \mu_g\|_2^2 + \text{Tr} \left(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{\frac{1}{2}} \right) \quad (3.10)$$

The lower the value of this metric, the greater the similarity between the distribution of real images and generated images, indicating a better result from the generative model. Note that the minimum value of this metric is zero (full equivalence between distributions).

FID manages to overcome most of the limitations of IS, such as taking into account intraclass diversity or introducing the information of real images in the evaluation. For this reason, FID has become the state of the art evaluation metric for generative model evaluation.

Chapter 4

Experiments and Results

The assessment of the results will rely on two different methodologies: a qualitative and a quantitative analysis. DCGAN network performance will be evaluated against both training databases. DCGAN and WGAN networks performance will be compared on the ONCOVISION-MAMMI database.

4.1 Qualitative evaluation

4.1.1 Visual Turing Test

The results of this work are visual in nature. Thus, one of the main evaluation strategies will be to rely on the visual analysis capability of the human brain.

Following the strategy described in Section 3.6.1.1, five independent non-expert evaluators tried to distinguish between real sinogram images and synthetic sinogram images generated by our models. The results are summarized in Table 4.1.1.

The identification accuracy of real and generated images is approximately homogeneous for all evaluators. In general, real sinograms are correctly classified as such, but there are some cases in which evaluators confuse them with generated sinograms. This effect is accentuated in the ONCOVISION-MAMMI database study, where sinograms present noise and therefore lack of sharpness can lead to further confusion.

In the case of synthetic images, a high percentage of images is able to be confused by the evaluators. The high statistical significance with respect to the random classification baseline highlights the ability of the models to generate realistic sinogram images.

Note that using the same generative model, the quality of the realistic images is qualitatively

superior for the quasi-ideal dataset. Similarly, WGAN-GP produces better results than DCGAN for the same training dataset. This last difference is not statistically significant and does not allow us to conclude a superiority of one network over the other.

		Eval. 1	Eval. 2	Eval. 3	Eval. 4	Eval. 5	Average
ACRIN (DCGAN)	R-R	86.7 %	76.7%	83.3 %	96.7 %	86.7 %	86 %
	R-G	13.3 %	23.3 %	16.7 %	3.3 %	13.3 %	14 %
	G-R	73.3 %	73.3 %	70 %	83.3 %	80 %	76 %
	G-G	26.7 %	26.7 %	30 %	16.7 %	20 %	24 %
MAMMI (DCGAN)	R-R	76.7 %	83.3 %	70 %	86.7 %	83.3 %	80 %
	R-G	23.3 %	16.7 %	30 %	13.3 %	16.7 %	20 %
	G-R	56.7 %	70 %	80 %	73.3 %	70 %	70 %
	G-G	43.4 %	30 %	20 %	26.7 %	30 %	30 %
MAMMI (WGAN-GP)	R-R	80 %	83.3 %	76.7 %	83.3 %	93.3 %	83.3 %
	R-G	20 %	16.7 %	23.3 %	16.7 %	6.7 %	16.7 %
	G-R	63.3 %	66.7 %	80 %	83.3 %	73.3 %	73.3 %
	G-G	36.7 %	33.3 %	20 %	16.7 %	26.7 %	26.7 %

Table 4.1: Visual Turing Test results for each studied model-database pair and evaluator. R-R: Real image recognized as real image; R-G: Real image recognized as generated image; G-R: Generated image recognized as real image; G-G: Generated image recognized as generated image.

4.1.2 Quality vs training

It is logical to hypothesize that as the network is trained for a greater number of epochs, the generated images will be increasingly indistinguishable from real images. Figure 4.1 characterizes the evolution of the synthesized images at different steps of generative network training.

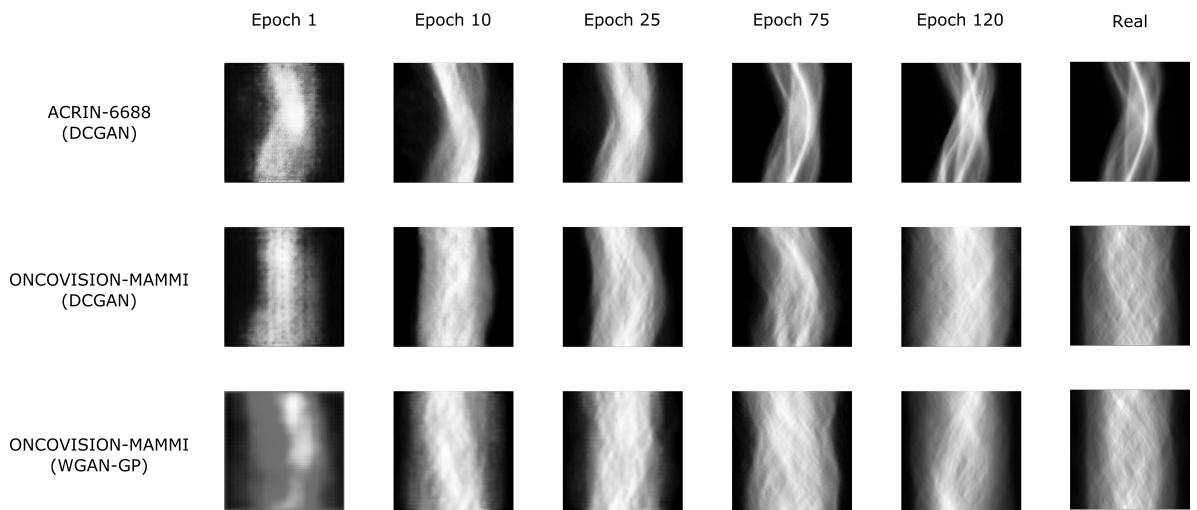


Figure 4.1: Evolution of synthesized sinograms as a function of the number of training epochs. Real column shown for fidelity comparison.

It can be observed that in the initial epochs the model focuses on the low frequencies, imitating the natural geometry of the training sinograms. As the training progresses, the different sinusoidal lines that make up the sinogram take their characteristic shape, the model being able to generate images more and more similar to the real ones. Note how from epoch 75 onwards there is no significant improvement in the quality of the images generated.

The posterior reconstruction of the generated sinograms will allow us to verify whether besides looking like real representations, the real anatomical structure studied is conserved.

4.2 Quantitative evaluation

4.2.1 DCGAN training stability

In DGANs the loss function has no meaning per se, since generator and discriminator are in an adversarial process. Since the ultimate goal of the two networks is to reach a Nash equilibrium, the evolution of the loss function will be a measure of the stability of the training.

The evolution of the generator and discriminator loss function for each of the DCGAN cases studied is shown in Figures 4.2 and 4.3.

In both cases the loss function has a high variability due to the inherent instability of GAN training. Note that the loss of the generator for the training with the first database (quasi-ideal) reaches a constant trend while for the training with the second database (non-corrected) it has an increasing trend. This would indicate a greater difficulty on the part of the generator to mimic the distribution of real uncorrected breast sinogram images.

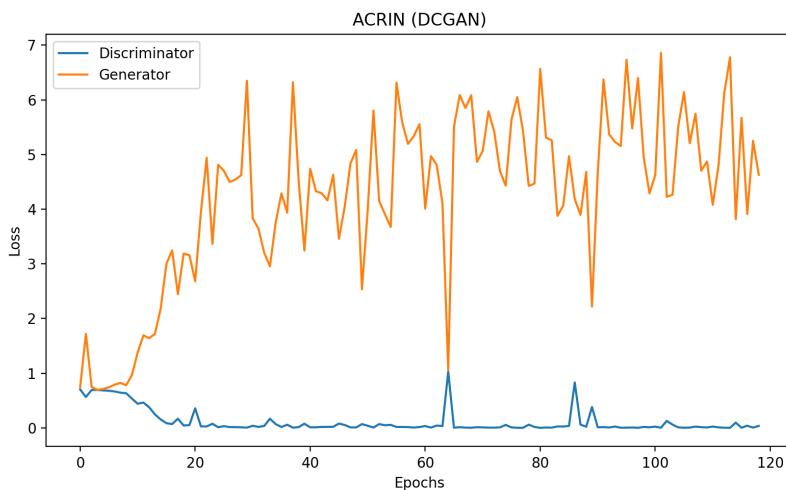


Figure 4.2: Evolution of generator and discriminator loss during DCGAN network training on ACRIN-6688 database.

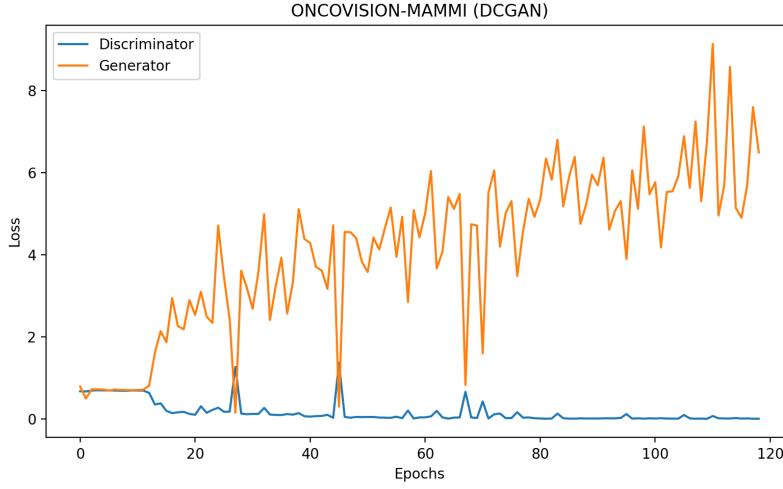


Figure 4.3: Evolution of generator and discriminator loss during DCGAN network training on ONCOVISION-MAMMI database.

4.2.2 Fréchet Inception Distance

Fréchet Inception Distance (FID) is a metric that evaluates the fidelity and diversity for a synthetic set of images. A (relative) lower value of this metric will be associated with a better overall quality of the synthetic data.

Nevertheless, the abstract nature of FID does not allow a comparison of the evolution of its value for two independent experiments to determine that one was better than the other.

The evolution of the FID values for each database-network pair have been determined using the official repository [pytorch-fid](#) [113], and are shown in Figures 4.4, 4.5, and 4.6.

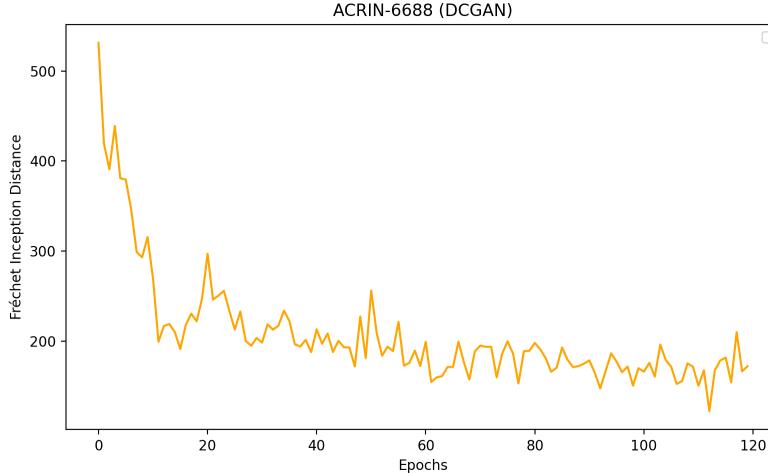


Figure 4.4: Evolution of FID score during DCGAN network training on ACRIN-6688.

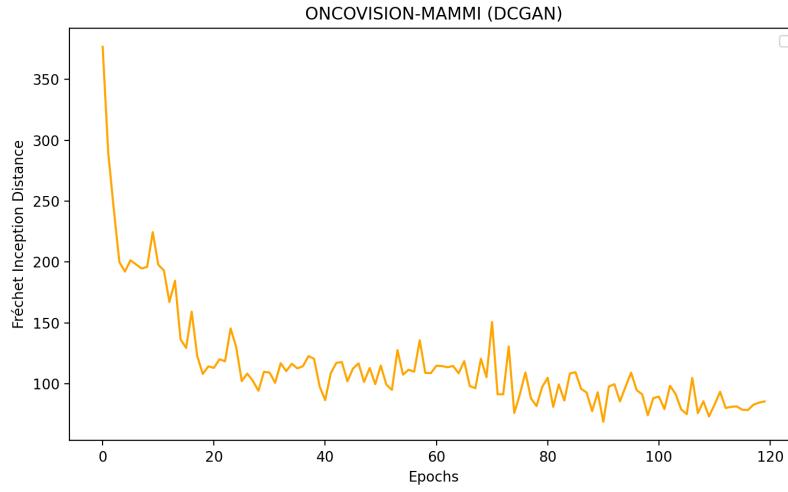


Figure 4.5: Evolution of FID score during DCGAN network training on ONCOVISION-MAMMI

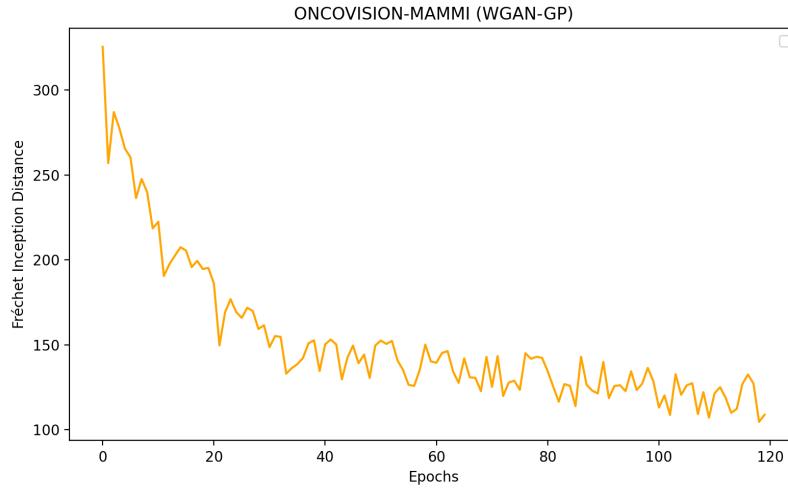


Figure 4.6: Evolution of FID score during WGAN-GP training on ONCOVISION-MAMMI.

The FID score value decreases similarly for all three networks, indicating that they all improve in generating images with higher fidelity and diversity.

4.3 Sinogram reconstruction

In this section a reconstruction for each experiment is performed in order to verify that the generated sinogram retains its semantic representation. The FBP analytical method has been used as a first approximation, since more advanced reconstruction techniques are beyond the scope of this work. The reconstructed sinograms are shown in Figures 4.7, 4.8, 4.9.

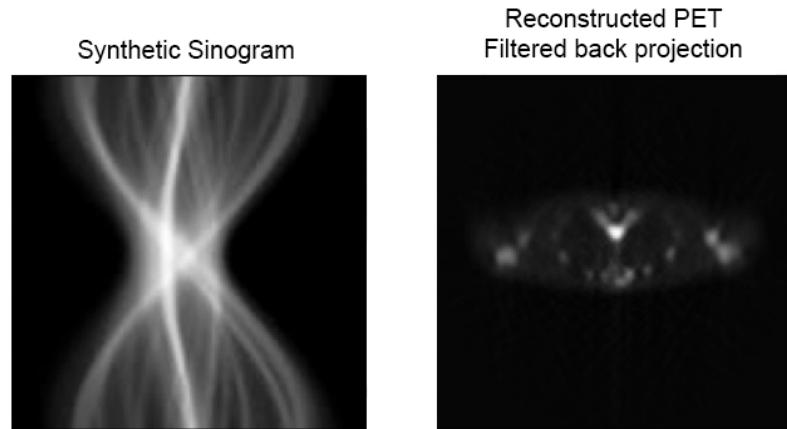


Figure 4.7: Synthetic sinogram and reconstructed PET image for DCGAN on ACRIN-6688.

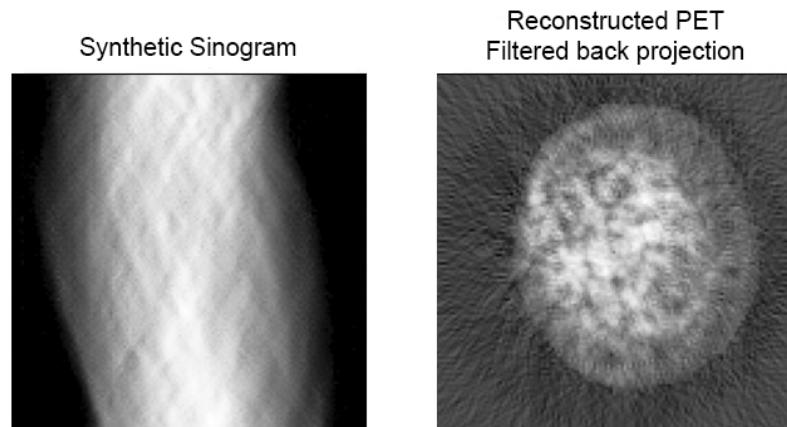


Figure 4.8: Synthetic sinogram and reconstructed PET image for DCGAN on ONCOVISION-MAMMI.

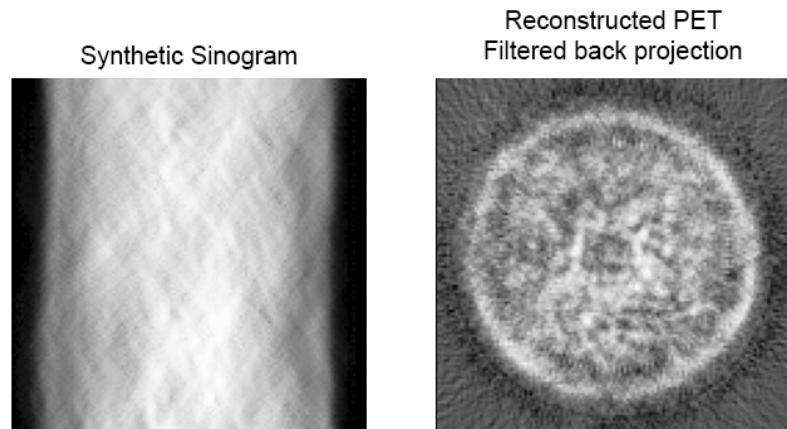


Figure 4.9: Synthetic sinogram and reconstructed PET image for WGAN-GP on ONCOVISION-MAMMI.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

There are domains in which obtaining real images for training Artificial Intelligence models is a bottleneck either because of their scarcity or because of the resources and time barriers. One of these scenarios is PET medical imaging, where the radioactive nature of the techniques used and the sensitive nature of the images limits their acquisition, especially in some regions such as the breast.

Several techniques have been proposed for the synthetic generation of images to artificially increase the volume of these databases. Classical data augmentation techniques are based on the geometrical modification of the images of the real set. In domains where the position of each pixel has an important semantic load (such as sinograms), these procedures are not valid. A possible alternative is to use AI-based generative models.

In this work, DCGAN and WGAN (with gradient penalty) networks have been explored using two case studies: a corrected (quasi-ideal) dataset and an uncorrected dataset. The latter represents the breast PET set and constitutes the main case study of this work.

Evaluating the fidelity and diversity of synthetic image sets is not a simple task. The task is made more difficult when the same metric must be able to serve as a point of comparison between two experiments.

The visual tests performed in this work show that these architectures have the ability to generate sinogram images that confound human evaluators. Furthermore, once reconstructed, it is found that they continue to represent the anatomical body under study.

The lower the amount of noise present in the real images of the training set, the faster the DCGAN network learns to generate high fidelity images.

While the techniques used by WGAN-GP avoid the mode collapse problem and result in more stable training, the results obtained are not significantly better than those produced by DCGAN. These results could be in line with a study conducted by Google Brain in which they found no evidence that other versions of GANs, including WGAN, consistently performed better than the original GAN [104].

In addition, the time and resources required for WGAN-GP training were significantly higher than those required for DCGAN (approximately 10 times more).

In conclusion, GANs (in particular DCGAN) have proven to be a useful tool for the unsupervised generation of synthetic images for a low training volume. The ability to work in the sinogram space without losing semantic load shows the potential of the proposed methodology in the clinical setting, specifically to work in complicated situations such as the use of PET medical imaging in breast.

5.2 Future work

This Master's thesis is part of a much larger ongoing research project (see Section 1.1.1). Because of its context, the scope of this work has been limited for temporal reasons and it is hoped to be able to continue this research in greater depth in the near future.

Working at the frontier of the state of the art with a very limited time range some difficulties have been unavoidable:

- The original breast PET imaging data requested by the *I3M Group of Detectors for Molecular Imaging research institute* were not available in time for this work.
- As of 2021, there are no large databases of breast PET images, as its application in this part of the body presents considerable difficulties. The steps of requesting a PET diagnostic test, patient citation, and anonymization are very slow processes, further aggravated by the nature and sensitivity of the images.
- The ONCOVISION-MAMMI images are not corrected and present some of the problems discussed in Section 3.2.2. This noise is transferred to the sinograms and amplified both in the generation of new sinograms and in their subsequent reconstruction. Further development of this specific detector is expected to yield cleaner images.
- The literary review seems to indicate that for the moment there is no notable work on unsupervised PET imaging generation. For this reason, we do not have a direct framework for comparison, and can only rely on results obtained for other types of medical imaging.

This work intends to study and establish the foundations of synthetic image generation in areas with difficulty in collecting real images. It is expected that by the time the above points become available, their integration with the rest of the defined workflow will be a trivial task.

As reviewed in Section 2.1, the reconstruction of sinogram to PET image is not a trivial process. The reconstructed PET images presented in Section 4.3 have been obtained from the analytical application of the inverse Radon transform on the sinogram matrix (analytical methods). By using specialized PET image reconstruction softwares such as *Customizable and Advanced Software for Tomographic Reconstruction* (CASToR) [114], much better reconstructed PET images could be obtained from the synthesized sinograms. The understanding of this type of tools was beyond the scope of this work, where the reconstruction process was only performed to carry out a qualitative check of the semantic representation of the generated sinograms.

Generative networks are very sensitive to hyperparameters: epochs, learning rate, number and sized of architecture layers, ... For this reason, it is expected that a much more detailed study of hyperparameters according to the input dataset could provide better results. The long training times, the lack of comparative evaluation metrics, and the need to define a global framework have suggested the default use of the hyperparameters proposed in the original papers, within a small margin of maneuver.

Throughout this study we have worked mainly with two-dimensional PET images (obtained as slices of a three-dimensional PET image), which translates into two-dimensional sinogram images. In a real clinical scenario, the raw data are volumetric data, and therefore the final objective will be to translate the studied procedures to a higher dimension, where some additional difficulties arise. This procedure constitutes the next step within the global project in which this thesis is framed.

The unsupervised methods proposed in this work assume that the network has been trained by a database of sinograms corresponding to the type of sinograms to be generated. However, by using conditional GANs, a new conditional input vector could allow to customize the image to be synthesized. Thus, different variables such as the size of the breast, the density of the breast tissue, the presence of tumors, or the presence of microcalcifications could be imposed.

The implementation of these techniques would allow greater control over the model, both at the level of use and explainability. The equivalences between each feature and its representation in the form of a sinogram may be studied. Nevertheless, for this type of methodology to be feasible, it would require not only a much larger volume of available images but also a supervised multilabel structure indicating all the features contained in each image.

The wide range of possibilities for the improvement of this work highlights its importance within the frontiers of research. The objectives of this Master's thesis within its context are considered to have been achieved.

Bibliography

- [1] Urania Dafni, Zoi Tsourtis, and Ioannis Alatsathianos. Breast cancer statistics in the european union: Incidence and survival across european countries, December 2019.
- [2] Rebecca L. Siegel, Kimberly D. Miller, and Ahmedin Jemal. Cancer statistics, 2019. *CA: A Cancer Journal for Clinicians*, 69(1):7–34, January 2019.
- [3] O M Koriech. Breast cancer and early detection. *Journal of family & community medicine*, January 1996.
- [4] Arn Migowski. A detecção precoce do câncer de mama e a interpretação dos resultados de estudos de sobrevida. *Ciencia e Saude Coletiva*, 20(4):1309–1310, 2015.
- [5] J. A. Merino Bonilla, M. Torres Tabanera, and L. H. Ros Mendoza. Breast cancer in the 21st century: from early detection to new therapies. *Radiologia*, 59(5):368–379, September 2017.
- [6] Ophira Ginsburg, Cheng-Har Yip, Ari Brooks, Anna Cabanes, Maira Caleffi, Jorge Antonio Dunstan Yataco, Bishal Gyawali, Valerie McCormack, Myrna McLaughlin de Anderson, Ravi Mehrotra, Alejandro Mohar, Raul Murillo, Lydia E. Pace, Electra D. Paskett, Anya Romanoff, Anne F. Rositch, John R. Scheel, Miriam Schneidman, Karla Unger-Saldaña, Verna Vanderpuye, Tsu-Yin Wu, Safina Yuma, Allison Dvaladze, Catherine Duggan, and Benjamin O. Anderson. Breast cancer early detection: A phased approach to implementation. *Cancer*, 126(S10):2379–2393, May 2020.
- [7] Lulu Wang. Early diagnosis of breast cancer. *Sensors (Switzerland)*, 17(7), July 2017.
- [8] Zeinab Lima, Mohammad Reza Ebadi, Ghazaleh Amjad, and Ladan Younesi. Application of imaging technologies in breast cancer detection: A review article, March 2019.
- [9] James V. Fiorica. Breast Cancer Screening, Mammography, and Other Modalities. *Clinical Obstetrics and Gynecology*, 59(4):688–709, December 2016.
- [10] Rongrong Guo, Guolan Lu, Binjie Qin, and Baowei Fei. Ultrasound Imaging Technologies for Breast Cancer Detection and Management: A Review, January 2018.

- [11] Nehmat Houssami, Robin M. Turner, and Monica Morrow. Meta-analysis of pre-operative magnetic resonance imaging (MRI) and surgical treatment for breast cancer, September 2017.
- [12] Mahdiyeh Shamsi and Jalil Pirayesh Islamian. Breast cancer: Early diagnosis and effective treatment by drug delivery tracing, January 2017.
- [13] Naima Kaabouch. A Survey of Medical Imaging Techniques Used for Breast Cancer Detection. Technical report, University of North Dakota, 2013.
- [14] Eric L. Rosen, William B. Eubank, and David A. Mankoff. FDG PET, PET/CT, and breast cancer imaging, October 2007.
- [15] David Groheux, Alexandre Cochet, Olivier Humbert, Jean Louis Alberini, Elif Hindié, and David Mankoff. 18F-FDG PET/CT for staging and restaging of breast cancer. *Journal of Nuclear Medicine*, 57:17S–26S, February 2016.
- [16] Abi Berger. Positron emission tomography, June 2003.
- [17] Kipp W. Johnson, Jessica Torres Soto, Benjamin S. Glicksberg, Khader Shameer, Riccardo Miotto, Mohsin Ali, Euan Ashley, and Joel T. Dudley. Artificial Intelligence in Cardiology, June 2018.
- [18] Ahmed Hosny, Chintan Parmar, John Quackenbush, Lawrence H. Schwartz, and Hugo J.W.L. Aerts. Artificial intelligence in radiology, August 2018.
- [19] Lei Bi, Jinman Kim, Ashnil Kumar, Dagan Feng, and Michael Fulham. Synthesis of Positron Emission Tomography (PET) Images via Multi-channel Generative Adversarial Networks (GANs). *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10555 LNCS:43–51, July 2017.
- [20] Wanting Li, Haiyan Liu, Feng Cheng, Yanhua Li, Sijin Li, and Jiangwei Yan. Artificial intelligence applications for oncological positron emission tomography imaging. *European Journal of Radiology*, 134:109448, 2021.
- [21] Jun Dang, Benjamin Frisch, Philippe Lasaygues, Dachun Zhang, Stefaan Tavernier, Nicolas Felix, Paul Lecoq, Etienne Auffray, Joao Varela, Serge Mensah, and Mingxi Wan. Development of an anthropomorphic breast phantom for combined PET, B-mode ultrasound and elastographic imaging. *IEEE Transactions on Nuclear Science*, 58(3 PART 1):660–667, June 2011.
- [22] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, October 2014.

- [23] Enda Ridge. Guerrilla Analytics - 1st Edition, 2015.
- [24] Enda Ridge. *Guerrilla Analytics: A Practical Approach to Working with Data*. Morgan Kaufmann, Amsterdam, 2014.
- [25] Ivan Valeryevich Evdokimov, Roman Yurievich Tsarev, Tatiana Nikolaevna Yamskikh, and Alexander Nikolaevich Pupkov. Using PERT and Gantt charts for planning software projects on the basis of distributed digital ecosystems. In *Journal of Physics: Conference Series*, volume 1074, page 12127. Institute of Physics Publishing, August 2018.
- [26] Louise I T Lee, Senthooran Kanthasamy, Radha S Ayyalaraju, and Rakesh Ganatra. The Current State of Artificial Intelligence in Medical Imaging and Nuclear Medicine. *BJR—Open*, 1(1):20190037, July 2019.
- [27] Curtis P. Langlotz, Bibb Allen, Bradley J. Erickson, Jayashree Kalpathy-Cramer, Keith Bigelow, Tessa S. Cook, Adam E. Flanders, Matthew P. Lungren, David S. Mendelson, Jeffrey D. Rudie, Ge Wang, and Krishna Kandarpa. A roadmap for foundational research on artificial intelligence in medical imaging: From the 2018 NIH/RSNA/ACR/The Academy workshop. *Radiology*, 291(3):781–791, April 2019.
- [28] I. Lavdas, B. Glocker, D. Rueckert, S. A. Taylor, E. O. Aboagye, and A. G. Rockall. Machine learning in whole-body MRI: experiences and challenges from an applied study using multicentre data, May 2019.
- [29] Marwa Al and Mohammed Al-Hayani. The use of filtered back projection algorithm for reconstruction of tomographic image. pages 151–156, 01 2014.
- [30] F. Natterer. *The Mathematics of Computerized Tomography*. Society for Industrial and Applied Mathematics, January 2001.
- [31] Wim van Drongelen. Fourier Transform Applications. In *Signal Processing for Neuroscientists*, pages 107–126. Elsevier, January 2007.
- [32] Zhao Jin, Han Ming Zhang, Bin Yan, Lei Li, Lin Yuan Wang, and Ai Long Cai. Novel Fourier-based iterative reconstruction for sparse fan projection using alternating direction total variation minimization. *Chinese Physics B*, 25(3):038701, January 2016.
- [33] E. Derenzo Stephen. Mathematical removal of positron range blurring in high resolution tomography. *IEEE Transactions on Nuclear Science*, 33(1):565–569, 1986.
- [34] Zhengrong Liang. Detector Response Restoration in Image Reconstruction of High Resolution Positron Emission Tomography. *IEEE Transactions on Medical Imaging*, 13(2):314–321, 1994.
- [35] G. L. Zeng. Image reconstruction - A tutorial. *Computerized Medical Imaging and Graphics*, 25(2):97–103, March 2001.

- [36] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.
- [37] L. A. Shepp and Y. Vardi. Maximum Likelihood Reconstruction for Emission Tomography. *IEEE Transactions on Medical Imaging*, 1(2):113–122, 1982.
- [38] Alan F. Karr. *Point processes and their statistical inference, second edition, revised and expanded*. CRC Press, January 2017.
- [39] Kenneth Lange and Richard Carson. Em reconstruction algorithms for emission and transmission tomography. *Journal of computer assisted tomography*, 8:306–316, 05 1984.
- [40] Chin Tu Chen and Charles E. Metz. A simplified em reconstruction algorithm for TOF-PET. *IEEE Transactions on Nuclear Science*, 32(1):885–888, 1985.
- [41] L. A. Shepp, Y. Vardi, J. B. Ra, S. K. Hilal, and Z. H. Cho. Maximum likelihood pet with real data. *IEEE Transactions on Nuclear Science*, 31(2):910–913, 1984.
- [42] J. Llacer, E. Veklerov, L. R. Baxter, S. T. Grafton, L. K. Griffeth, R. A. Hawkins, C. K. Hoh, J. C. Mazziotta, E. J. Hoffman, and C. E. Metz. Results of a clinical receiver operating characteristic study comparing filtered backprojection and maximum likelihood estimator images in FDG PET studies. *Journal of Nuclear Medicine*, 34(7):1198–1203, 1993.
- [43] Emanuel Levitan and Gabor T. Herman. A Maximum A Posteriori Probability Expectation Maximization Algorithm for Image Reconstruction in Emission Tomography. *IEEE Transactions on Medical Imaging*, 6(3):185–192, 1987.
- [44] Margaret E. Daube-Witherspoon and Gerd Muehllehner. An Iterative Image Space Reconstruction Algorithm Suitable for Volume ECT. *IEEE Transactions on Medical Imaging*, 5(2):61–66, 1986.
- [45] John M.M. Anderson, B. A. Mair, Murali Rao, and Chen Hsien Wu. Weighted least-squares reconstruction methods for positron emission tomography. *IEEE Transactions on Medical Imaging*, 16(2):159–165, 1997.
- [46] Jeffrey A. Fessler and Alfred O. Hero. Space-Alternating Generalized Expectation-Maximization Algorithm. *IEEE Transactions on Signal Processing*, 42(10):2664–2677, 1994.
- [47] H. Malcolm Hudson and Richard S. Larkin. Accelerated Image Reconstruction Using Ordered Subsets of Projection Data. *IEEE Transactions on Medical Imaging*, 13(4):601–609, 1994.

- [48] Ing Tsung Hsiao, Anand Rangarajan, Parmeshwar Khurd, and Gene Gindi. An accelerated convergent ordered subsets algorithm for emission tomography. *Physics in Medicine and Biology*, 49(11):2145–2156, June 2004.
- [49] Parisa Khateri, Jannis Fischer, Werner Lustermann, Charalampos Tsoumpas, and Günther Dissertori. Implementation of cylindrical PET scanners with block detector geometry in STIR. *EJNMMI Physics*, 6(1):15, December 2019.
- [50] Ciprian Catana. Motion correction options in PET/MRI, May 2015.
- [51] Aswin John Mathews, Ke Li, Sergey Komarov, Qiang Wang, Bosky Ravindranath, Joseph A. O’Sullivan, and Yuan Chuan Tai. A generalized reconstruction framework for unconventional PET systems. *Medical physics*, 42(8):4591–4609, August 2015.
- [52] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. Technical Report 6, 2012.
- [53] Keiron O’Shea and Ryan Nash. An Introduction to Convolutional Neural Networks. November 2015.
- [54] Alexander Selvikvåg Lundervold and Arvid Lundervold. An overview of deep learning in medical imaging focusing on MRI, May 2019.
- [55] Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C. Hansen. On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proceedings of the National Academy of Sciences of the United States of America*, 117(48):30088–30095, December 2020.
- [56] Haimiao Zhang and Bin Dong. A Review on Deep Learning in Medical Image Reconstruction. *Journal of the Operations Research Society of China*, 8(2):311–340, June 2019.
- [57] Sarat Kumar Sarvepalli. Deep learning in neural networks: The science behind an artificial brain. October 2015.
- [58] Amitojdeep Singh, Sourya Sengupta, and Vasudevan Lakshminarayanan. Explainable deep learning models in medical image analysis. Technical report, 2020.
- [59] Asaf Bar-El, Dana Cohen, Noa Cahan, and Hayit Greenspan. Improved CycleGAN with application to COVID-19 classification. In Bennett A. Landman and Ivana Isgum, editors, *Medical Imaging 2021: Image Processing*, volume 11596, page 33. SPIE, February 2021.
- [60] Martin J. Willemink, Wojciech A. Koszek, Cailin Hardell, Jie Wu, Dominik Fleischmann, Hugh Harvey, Les R. Folio, Ronald M. Summers, Daniel L. Rubin, and Matthew P. Lungren. Preparing medical imaging data for machine learning, 2020.

- [61] Philipp Mergenthaler, Ute Lindauer, Gerald A. Dienel, and Andreas Meisel. Sugar for the brain: The role of glucose in physiological and pathological brain function, October 2013.
- [62] Kiran Kumar Solingapuram Sai, Zuzana Zachar, Paul M. Bingham, and Akiva Mintz. Metabolic PET imaging in oncology. *American Journal of Roentgenology*, 209(2):270–276, August 2017.
- [63] B. B. Koolen, W. V. Vogel, M. J.T.F.D. Vrancken Peeters, C. E. Loo, E. J.Th Rutgers, and R. A. Valdés Olmos. Molecular imaging in breast cancer: From whole-body PET/CT to dedicated breast PET, 2012.
- [64] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2016-December, pages 770–778. IEEE Computer Society, December 2016.
- [65] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.
- [66] Md. Zahangir Alom, Tarek Taha, Chris Yakopcic, Stefan Westberg, Paheding Sidike, Mst Nasrin, Mahmudul Hasan, Brian Essen, Abdul Awwal, and Vijayan Asari. A state-of-the-art survey on deep learning theory and architectures. 8:292, 03 2019.
- [67] Cherry Khosla and Baljit Singh Saini. Enhancing Performance of Deep Learning Models with different Data Augmentation Techniques: A Survey. In *Proceedings of International Conference on Intelligent Engineering and Management, ICIEM 2020*, pages 79–85. Institute of Electrical and Electronics Engineers Inc., June 2020.
- [68] Félix Renard, Soulaimane Guedria, Noel De Palma, and Nicolas Vuillerme. Variability and reproducibility in deep learning for medical image segmentation. *Scientific Reports*, 10(1):1–16, December 2020.
- [69] Connor Shorten and Taghi M. Khoshgoftaar. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1):60, December 2019.
- [70] Jakub Nalepa, Michal Marcinkiewicz, and Michal Kawulok. Data Augmentation for Brain-Tumor Segmentation: A Review. *Frontiers in Computational Neuroscience*, 13:83, December 2019.
- [71] Pranav Rajpurkar. Deep Learning For Medical Image Interpretation. Technical report.
- [72] Ekin D Cubuk, Barret Zoph, Vijay Vasudevan, and Quoc V Le Google Brain. AutoAugment: Learning Augmentation Strategies from Data. Technical report.

- [73] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. RandAugment: Practical automated data augmentation with a reduced search space. Technical report, 2019.
- [74] Javier Jorge, Jesús Vieco, Roberto Paredes, Joan Andreu Sanchez, and José Miguel Benedí. Empirical Evaluation of Variational Autoencoders for Data Augmentation. 2018.
- [75] Veit Sandfort, Ke Yan, Perry J. Pickhardt, and Ronald M. Summers. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Scientific Reports*, 9(1):1–9, December 2019.
- [76] Diederik P Kingma Google, Max Welling, and Boston Delft. An Introduction to Variational Autoencoders. *Foundations and Trends R in Machine Learning*, xx, No. xx:1–18, 2019.
- [77] Ling Huang, Anthony D Joseph, Blaine Nelson, Benjamin I P Rubinstein, and J D Tygar. *Adversarial Machine Learning*. 2011.
- [78] A Preprint Sagar and Kora Venu. Evaluation of Deep Convolutional Generative Adversarial Networks for data augmentation of chest X-ray images. Technical report, 2020.
- [79] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications. Technical Report 8, 2015.
- [80] S. A. Barnett. Convergence problems with generative adversarial networks (GANs), June 2018.
- [81] Xi Cheng, Li Zhang, and Yefeng Zheng. Deep similarity learning for multimodal medical images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, 6(3):248–252, May 2018.
- [82] Mingxia Liu, Jun Zhang, Pew Thian Yap, and Dinggang Shen. View-aligned hypergraph learning for Alzheimer’s disease diagnosis with incomplete multi-modality data. *Medical Image Analysis*, 36:123–134, February 2017.
- [83] Mehdi Mirza and Simon Osindero. Conditional Generative Adversarial Nets. Technical report, 2014.
- [84] Jun Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2017-October, pages 2242–2251. Institute of Electrical and Electronics Engineers Inc., December 2017.
- [85] Yongsheng Pan, Mingxia Liu, Chunfeng Lian, Tao Zhou, Yong Xia, and Dinggang Shen. Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks

- for Alzheimer’s disease diagnosis. In *Lecture Notes in Computer Science (including sub-series Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11072 LNCS, pages 455–463. Springer Verlag, 2018.
- [86] Kui Zhao, Long Zhou, Size Gao, Xiaozhuang Wang, Yaofa Wang, Xin Zhao, Huatao Wang, Kanfeng Liu, Yunqi Zhu, and Hongwei Ye. Study of low-dose PET image recovery using supervised learning with CycleGAN. *PLOS ONE*, 15(9):e0238455, September 2020.
 - [87] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. Technical report.
 - [88] Qiufeng Wu, Yiping Chen, and Jun Meng. Dcgan-based data augmentation for tomato leaf disease identification. *IEEE Access*, 8:98716–98728, 2020.
 - [89] Aissam Jadli, Mustapha Hain, Adil Chergui, and Abderrahman Jaize. DCGAN-based data augmentation for document classification. In *2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science, ICECOCS 2020*. Institute of Electrical and Electronics Engineers Inc., December 2020.
 - [90] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification. Technical report, 2018.
 - [91] Manisha Saini and Seba Susan. Deep transfer with minority data augmentation for imbalanced breast cancer dataset. *Applied Soft Computing Journal*, 97:106759, December 2020.
 - [92] Jyoti Islam and Yanqing Zhang. GAN-based synthetic brain PET image generation. *Brain Informatics*, 7(1):3, December 2020.
 - [93] Jari Korhonen and Junyong You. Peak signal-to-noise ratio revisited: Is simple beautiful? In *2012 4th International Workshop on Quality of Multimedia Experience, QoMEX 2012*, pages 37–38, 2012.
 - [94] Jim Nilsson and Tomas Akenine-Möller. Understanding SSIM. *arXiv*, June 2020.
 - [95] Lloyd A. Courtenay and Diego González-Aguilera. Geometric morphometric data augmentation using generative computational learning algorithms. *Applied Sciences (Switzerland)*, 10(24):1–25, December 2020.
 - [96] Jason Wang and Luis Perez. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. Technical report.
 - [97] Lale Kostakoglu, Fenghai Duan, Michael O. Idowu, Paul R. Jolles, Harry D. Bear, Mark Muzi, Jean Cormack, John P. Muzi, Daniel A. Pryma, Jennifer M. Specht, Linda

- Hovanessian-Larsen, John Miliziano, Sharon Mallett, Anthony F. Shields, and David A. Mankoff. A phase ii study of 3'-deoxy-3'-18f-fluorothymidine pet in the assessment of early response of breast cancer to neoadjuvant chemotherapy: Results from acrin 6688. *Journal of Nuclear Medicine*, 56(11):1681–1689, 2015.
- [98] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, Lawrence Tarbox, and Fred Prior. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *Journal of Digital Imaging*, 26(6):1045–1057, 2013.
- [99] Laura Moliner Martínez. Reconstrucción on-line con algoritmos modo lista para equipos pet de cristal continuo. 2014.
- [100] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan, 2017.
- [101] D.A. Edwards. On the kantorovich–rubinstein theorem. *Expositiones Mathematicae*, 29(4):387–398, 2011.
- [102] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 214–223. PMLR, 06–11 Aug 2017.
- [103] Ishaan Gulrajani, Faruk Ahmed, Martín Arjovsky, Vincent Dumoulin, and Aaron C. Courville. Improved training of wasserstein gans. *CoRR*, abs/1704.00028, 2017.
- [104] Mario Lucic, Karol Kurach, Marcin Michalski, Sylvain Gelly, and Olivier Bousquet. Are gans created equal? a large-scale study, 2018.
- [105] Nicola Giocoli. Nash equilibrium. *History of Political Economy*, 36, 12 2004.
- [106] Ali Borji. Pros and cons of GAN evaluation measures. *CoRR*, abs/1802.03446, 2018.
- [107] Hamed Alqahtani, Manolya Kavakli-Thorne, and Dr. Gulshan Kumar Ahuja. An analysis of evaluation metrics of gans. 07 2019.
- [108] Kevin Crowston. Amazon mechanical turk: A research tool for organizations and information systems scholars. In Anol Bhattacherjee and Brian Fitzgerald, editors, *Shaping the Future of ICT Research. Methods and Approaches*, pages 210–221, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.
- [109] Sharon Zhou, Mitchell L. Gordon, Ranjay Krishna, Austin Narcomey, Durim Morina, and Michael S. Bernstein. HYPE: human eye perceptual evaluation of generative models. *CoRR*, abs/1904.01121, 2019.

- [110] Tim Salimans, Ian J. Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *CoRR*, abs/1606.03498, 2016.
- [111] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6629–6640. Curran Associates Inc., 2017.
- [112] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, and Jonathon Shlens. Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567, 2015.
- [113] Maximilian Seitzer. pytorch-fid: FID Score for PyTorch. <https://github.com/mseitzer/pytorch-fid>, August 2020. Version 0.1.1.
- [114] Thibaut Merlin, Simon Stute, Didier Benoit, Julien Bert, Thomas Carlier, Claude Comtat, Marina Filipovic, Frédéric Lamare, and Dimitris VISVIKIS. CASToR: a generic data organization and processing code framework for multi-modal and multi-dimensional tomographic reconstruction. *Physics in Medicine and Biology*, 63(18):185005, September 2018.