

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Deep Generative Models for 3D Breast MRI Shapes and Textures

Maria de Fátima da Silva Barros

DISSERTATION



Mestrado Integrado em Bioengenharia

Supervisor: Helder Filipe Pinto de Oliveira, PhD

Co-supervisor: João Pedro Fonseca Teixeira, MSc

July 27, 2021

Deep Generative Models for 3D Breast MRI Shapes and Textures

Maria de Fátima da Silva Barros

Mestrado Integrado em Bioengenharia

Approved in oral examination by the committee:

Chair: Luís Filipe Teixeira, PhD

External Examiner: Francesco Renna, PhD

Supervisor: Hélder Filipe Pinto de Oliveira, PhD

July 27, 2021

Resumo

Atualmente, o cancro da mama é o tipo de cancro mais comum em mulheres. As técnicas de diagnóstico dependem fortemente da imagiologia da mama, como Mamografia e Imagem por Ressonância Magnética. Na maioria dos casos, o tratamento envolve cirurgia para remoção do tumor, que pode ser conservadora e remover apenas as áreas afetadas ou remover toda a mama. Mesmo com um tratamento conservativo, ocorrem deformações na área que podem afetar a qualidade de vida das pacientes. Ferramentas tridimensionais (3D) que combinem modalidades de dados, estimando resultados de cirurgia ou mesmo com sistemas integrados que ajudam no diagnóstico, podem ajudar tanto os médicos a tratar o cancro, como os pacientes a compreender que tecidos vão ser removidos durante a cirurgia e sobretudo como o procedimento vai afetar o seu aspeto físico. Mas para desenvolver estas ferramentas são necessárias bases de dados relevantes, tanto para a criação de modelos 3D de mama como para os investigadores de visão computacional compararem estratégias e usarem um conjunto de dados confiável que atenda às expectativas da pesquisa. Considerando isto, o objetivo desta dissertação é criar modelos generativos, baseados em arquiteturas de Aprendizagem Profunda, de forma a representar dados 3D de mama, tanto de Ressonância Magnética como de formas de superfície.

Para as Imagens por Ressonância Magnética de mama, uma variedade de modelos para reconstrução e geração de volumes de dados foram explorados. As estratégias envolveram métodos que visam sintetizar volumes de ressonância magnética na resolução original, e métodos baseados em *patches*, onde os volumes inteiros são divididos em pequenos volumes que são computacionalmente mais eficientes de treinar, e que resultam num aumento do número de instâncias de entrada. Com os dados disponíveis, não foi possível desenvolver um método baseado em volumes que obtivesse imagens com a qualidade necessária para uso clínico. Nos métodos baseados em *patches*, foi desenvolvida uma arquitetura 3D X-Net que mostrou resultados promissores ao completar *patches* de Ressonância Magnética usando anotações de estruturas relevantes. Contudo, processo de treino pode ser melhorado de forma a capturar mais precisamente algumas componentes de alta-frequência presentes nas Imagens por Ressonância Magnética de mama.

Para a modelação de formas de mama, diferentes modelos baseados na arquitetura PointNet foram desenvolvidos. Para aprendizagem de representações, um *autoencoder* (AE) foi desenhado e, para a síntese de dados, foi desenvolvido um *variational autoencoder* (VAE). O AE pode representar formas de mama, criar novas formas a partir da interpolação linear de vetores latentes e realizar tarefas de conclusão de formas. Este modelo também foi usado para a transformação da pose de braços para ereta, mostrando a capacidade de aprender características relevantes da pose. O VAE foi capaz de sintetizar formas além das presentes nos dados de treino. O uso de aprendizagem por transferência de conhecimento de uma base de dados maior mostrou-se benéfica para o VAE, enquanto o AE consegue aprender representações usando treino direto nos dados de mama. Além disso, o modelo VAE também pode ser usado para criar formas através de manipulação vetores latentes criados do zero e, em seguida, decodificando-os. Um VAE condicional foi testado para

gerar formas de acordo com uma classe de rotação da forma, indicando que modelos condicionais poderiam ser usados com classes mais relevantes, como volume ou formato da mama.

As metodologias apresentadas nesta dissertação representam avanços na exploração e avaliação de modelos generativos para a síntese de formas e texturas de ressonância magnética de mama, promovendo o desenvolvimento de ferramentas automáticas para dados 3D de mama.

Abstract

Currently, Breast Cancer is the most common type of cancer in women. Diagnostic techniques rely heavily on breast imagiology, such as Mammography and Magnetic Resonance Imaging (MRI). In most cases, the treatment involves surgery for breast cancer removal, which can either be conservative and remove only the affected areas or remove the entire breast. Even with conservative treatment, deformations occur in the area that can affect the patients' quality of life. Three-dimensional (3D) tools combining data modalities, estimating surgery outcomes, or even with integrated systems that aid the diagnosis, can help both physicians treating cancer, and patients comprehend what tissues are being removed during surgery and how the procedure is going to affect their physical aspect. However, developing these tools demands relevant data, either for creating 3D models of the breasts or for computer vision researchers compare their strategies and use a reliable dataset that fulfil research expectations. This dissertation aims to create generative models, based on Deep Learning (DL) architectures, that allow the representation of 3D breast data, both MRI and surface shapes.

In breast MRI, a variety of models for reconstructing and generating volumes of data were explored. The methodologies involved methods that aimed to synthesize volumes of MRI in their original resolution, and patch-based methods, where the MRI scans are divided into small volumes, computationally more efficient to train, and that result in an increase of the number of input instances. With the available data, it was not possible to develop a method based on volumes that would obtain images with the necessary quality for clinical use. For the patch-based methods, a 3D X-Net architecture was developed, showing promising results on completing MRI patches taking advantage of annotations of relevant structures. However, the training process can be improved to capture some high-frequency components present in breast MRI more accurately.

For breast shape modelling, different models based on the PointNet were developed. An autoencoder (AE) model for learning representations was designed as well as a variational autoencoder (VAE) for the synthesis of data. The AE could represent breast shapes, create new shapes from linear interpolation of latent vectors, and perform shape completion tasks. This model was also used for prone to upright pose transformations, showing the capacity to learn relevant pose features. The VAE was able to synthesize shapes beyond the training data. Transfer learning from a bigger dataset showed to be beneficial for the VAE model, whilst the AE model can learn representations using direct train on breast data. Furthermore, the VAE model can also be used to create shapes by crafting latent vectors and then decoding them. A conditional VAE was tested to generate shapes according to a label of rotation of the shape, indicating that conditional models could be used with more relevant labels, such as breast volume or shape.

The methodologies presented in this dissertation represent advances in the exploitation and evaluation of generative models for synthesizing breast MRI shapes and textures, ultimately promoting the development of automatic tools for 3D breast data.

Agradecimentos

Em primeiro lugar quero agradecer ao Prof. Hélder Oliveira por ter possibilitado o alargamento a minha educação de uma forma que nunca pensei ser possível durante os meus anos na faculdade. Ao João e à Sílvia, que me ensinaram muito mais do que talvez imaginam, por todo o tempo que investiram em mim, e por estarem sempre dispostos a ouvir a minha opinião e ideias. Tenho a certeza que a esta dissertação não seria a mesma sem vocês. Sei também que estes dois anos me elevaram como estudante, profissional e colega de equipa e por isso, obrigada. Ainda a todo o grupo que mesmo à distância esteve disponível para dar ideias, ajudar em todas as questões, e trazer novas perspetivas.

Agradeço a todos os amigos que me acompanharam nos últimos cinco anos. Àqueles que insistiam para eu trabalhar mais um pouco porque acreditavam em mim e em todas as minhas capacidades. Mas também aos que insistiam para deixar o trabalho de parte e desfrutar de todos os lazares da vida. Só com todos vocês aprendi a atingir o equilíbrio certo.

Finalmente, quero agradecer à minha família, que me apoiou desde sempre a seguir o meu caminho. Ao Zé, que me consegue acalmar e motivar a ser uma pessoa melhor, desde o primeiro dia. Às minhas irmãs, por me aturarem e mimarem desde sempre, o meu percurso académico foi muito mais fácil depois de vos ter como exemplo. E por fim aos meus pais, que mesmo sem terem tido esta oportunidade na altura deles, trabalharam para que as filhas a tivessem. Tudo isto é graças a vós.

Fátima Barros

“Hard work doesn’t guarantee success, but improves its chances.”

B. J. Gupta

Contents

1	Introduction	1
1.1	Motivation	2
1.2	Goals and Contributions	3
1.3	Document Structure	4
2	Clinical Context	5
2.1	Anatomy and Physiology of the Breast	5
2.2	Breast Cancer	6
2.2.1	Statistics	6
2.2.2	Treatments	7
2.3	Breast Imagiology	9
2.3.1	Radiological Exams	9
2.3.2	Surface Information	11
2.3.3	Available datasets	13
2.4	Summary	14
3	Deep Generative Modelling	15
3.1	Overview	15
3.1.1	Autoencoder based models	16
3.1.2	Generative Adversarial Network based models	20
3.1.3	Evaluation of generative models	26
3.2	Applications to Radiological data	31
3.3	Summary	34
4	3D Data Representations	35
4.1	Acquisition methods of 3D data	35
4.2	Representation of 3D data	36
4.2.1	Point clouds	37
4.3	Deep Learning and 3D data	38
4.3.1	PointNet	39
4.3.2	Generative Modelling and Point Clouds	41
4.4	Quantifying differences between point clouds	42
4.4.1	Earth Mover's Distance	42
4.4.2	Closest point distances	43
4.4.3	Jensen-Shannon Divergence (JSD)	43
4.5	Summary	44

5	Synthesis of Breast MRI	45
5.1	Data Preparation	45
5.2	Recurrent networks Breast MRI reconstruction	47
5.2.1	Predicting the next MRI slice	48
5.2.2	Memory-based network for 3D breast MRI reconstruction	50
5.3	Auto-encoding GAN for Breast MRI Generation	59
5.4	3D X-Net for Breast MRI Reconstruction	60
5.4.1	Generator Training on (64, 64, 8) patches	63
5.4.2	Generator and Adversarial Training on (64, 64, 4) patches	64
5.5	Summary	69
6	Breast Shape Representations	71
6.1	Learn Breast Surface from a Small Dataset	71
6.1.1	Datasets	72
6.1.2	Sensitivity of Distance Metrics in PCLs	73
6.1.3	Evaluation	74
6.2	Breast Representations with PointNet based Autoencoder	75
6.3	Modelling Breast Pose Transformation	80
6.4	Breast Synthesis with PointNet based Variational Autoencoder	82
6.5	Conditional Generation of Breast Shapes	89
6.6	Cross Validation and Discussion	90
6.7	Summary	92
7	Conclusions	93
7.1	Contributions and Final Remarks	93
7.2	Future Work	94
A	Visual Results	97
B	Tables	101
	References	103

List of Figures

2.1	Female breast anatomy.	5
2.2	Statistics from Globocan 2020.	7
2.3	Breast Cancer Surgery.	8
2.4	Different categories of breast density estimated by radiologists.	10
2.5	Multifocal carcinoma detected on MRI and negative results from the mammography.	11
2.6	Visualization of breast surface data and breast point cloud from two different points of view.	12
2.7	TCIA breast cancer results.	14
3.1	Convolutional Autoencoder architecture.	17
3.2	Variational Autoencoder Architecture.	18
3.3	Encodings of MNIST data from different AE based models.	19
3.4	GANs basic framework.	20
3.5	DCGAN generator architecture.	22
3.6	Conditional GAN for class-conditional image synthesis.	23
3.7	InfoGAN with a discriminator that estimates class label.	23
3.8	Coarse-to-fine generator of $pix2pix_{HD}$ architecture.	24
3.9	CycleGAN training cycle.	26
3.10	MRNGAN end-to-end framework.	33
3.11	Auto-encoding GAN architecture.	33
4.1	Representations of 3D data.	37
4.2	Stanford Bunny point cloud.	37
4.3	Breast point cloud from top view.	37
4.4	PointNet vanilla.	40
4.5	PointNet Architecture.	40
4.6	Architecture of the energy function of the generative PointNet.	42
5.1	Schematization of the creation of the different Breast MRI subsets from the original dataset.	46
5.2	Framework for reconstructing breast MRI using recurrent networks.	47
5.3	Representation of the $pix2pix$ network for prediction of the next MRI slice.	48
5.4	Visual performance of the best $pix2pix$ models.	49
5.5	Sequence of original and generated slices with $pix2pix$ model.	50
5.6	Representation of the Recurrent Autoencoder network for 3D MRI reconstruction.	51
5.7	Sequence of slices generated with the Recurrent Autoencoder.	51
5.8	Diagram of ConvLSTM Cell.	52
5.9	Architectures of One-to-Many ConvLSTM Base.	53
5.10	Visual performance of One-to-Many ConvLSTM Base models.	53

5.11	Architectures of One-to-Many ConvLSTM Chain.	54
5.12	Visual performance of One-to-Many ConvLSTM Chain models.	55
5.13	Architecture of Many-to-One ConvLSTM.	56
5.14	Visual performance of Many-to-One ConvLSTM model on Volumes Padded Subset.	57
5.15	Sequence of slices generated with the Many-To-One ConvLSTM model.	57
5.16	Architecture of Many-to-One ConvLSTM Autoencoder.	58
5.17	Architecture of Auto-encoding GAN (α -WGAN).	59
5.18	Visual performance of Auto-encoding GAN (α -WGAN) during train.	60
5.19	Architecture of the 3D X-Net for breast MRI reconstruction.	61
5.20	Generator and Adversarial Training using the 3D X-Net architecture.	62
5.21	Architecture of the 3D PatchGAN for adversarial training of 3D X-Net architecture.	62
5.22	Schematization of the incomplete (64, 64, 8) patches used with X-Net architecture.	63
5.23	Schematization of the incomplete (64, 64, 4) patches used with X-Net architecture.	64
5.24	Reconstruction of entire MRI process from 3D patches.	66
5.25	Visual performance of 3D X-Net on reconstructed slices with GenT-M2 model.	66
5.26	Visual performance of 3D X-Net on close-up of reconstructed slices from validation scan with GenT-M2 and AdvT-Annot-M1 models.	68
6.1	SHREC14 object used for experimenting metric values.	73
6.2	PointNet based autoencoder diagram.	75
6.3	Reconstructions of unseen shapes from 3D Breast in the validation dataset for the best models.	78
6.4	Linear interpolation of AE-CD latent spaces with breast shapes.	78
6.5	Shape completion: robustness of AE models to noise, holes and missing data.	79
6.6	Reconstructions of unseen shapes from 3D Breast for pose transformation in the validation set.	81
6.7	PointNet based Variational Autoencoder diagram.	82
6.8	Synthesis of breast shapes.	86
6.9	Visualization of latent space experiments.	87
6.10	Synthetic breast shapes created from discriminative features.	87
6.11	Linear interpolation between latent vectors in β -VAE-EMD ($LS=256$) TL1 model.	88
6.12	Arithmetic operations in the latent space of β -VAE-EMD ($LS=256$) TL1 model.	88
6.13	PointNet based Conditional Variational Autoencoder diagram.	89
6.14	Conditional synthesis of breast shapes, with rotation.	90
A.1	Visual performance of 3D X-Net reconstructed slices with GT-M2 model and (64, 64, 4) patches - additional visual results on training scan.	97
A.2	Visual performance of 3D X-Net reconstructed slices with GT-M2 model and (64, 64, 4) patches - additional visual results on validation scan.	98
A.3	Reconstruction of unseen shapes from ShapeNet dataset in the validation set.	98
A.4	Reconstruction of unseen shapes from SHREC14 dataset.	99
A.5	Reconstruction of unseen shapes from 3D Breast dataset in the validation set.	99
A.6	Linear interpolation between objects from ShapeNet dataset.	100
A.7	Synthesis of shapes with β -VAE-EMD ($LS=256$).	100
A.8	Conditional synthesis of objects in ShapeNet model.	100

List of Tables

3.1	Relevant generative modelling applications to radiological data.	32
4.1	Main strategies used for modelling 3D data with deep learning techniques.	38
5.1	Numerical performance of the <i>pix2pix</i> models for generating the next MRI slice.	49
5.2	Experiments performed with Many-to-One ConvLSTM model on Volumes Resized Subset.	56
5.3	Numerical performance of 3D X-Net Generator Training on (64, 64, 8) patches.	63
5.4	Methods with different hyperparameters used in 3D X-Net.	64
5.5	Numerical performance of scan reconstructions with 3D X-Net Generator and Adversarial Training on (64, 64, 4) patches.	65
5.6	Numerical performance of 3D X-Net Generator and Adversarial Training on (64, 64, 4) patches.	67
5.7	Numerical performance of reconstructions with 3D X-Net Generator and Adversarial Training for each label.	69
6.1	Summary of the point cloud datasets used.	73
6.2	Sensitivity of CD and EMD in PCLs from PCL from SHREC14 dataset with modifications.	74
6.3	Description of different experiments performed with autoencoder model.	76
6.4	Numerical performance of PointNet Autoencoder models.	77
6.5	Numerical performance of prone to upright pose transformation with PointNet Autoencoder models.	81
6.6	Description of different experiments performed with Variational Autoencoder model.	83
6.7	Numerical performance of PointNet Variational Autoencoder models.	84
6.8	Quantitative results of β -VAE-EMD (LS=256) on the 3D Breast validation set (4 instances).	85
6.9	Numerical performance of PointNet Autoencoder and Variational Autoencoder models in cross-validation methods.	91
6.10	Quantitative results of the β -VAE model trained with cross-validation.	91
B.1	Numerical performance of PointNet Autoencoder models.	101
B.2	Numerical performance of PointNet Autoencoder models with data augmentation.	102
B.3	Numerical performance of PointNet Variational Autoencoder models.	102

Abbreviations

BC	Breast Cancer
MRI	Magnetic Resonance Imaging
BCS	Breast-Conserving Surgery
3D	Three Dimensional
ML	Machine Learning
DL	Deep Learning
CAD	Computer-aided Diagnosis
CT	Computerized Tomography
PET	Positron Emission Tomography
2D	Two Dimensional
PCL	Point Cloud
DDSM	Digital Database for Screening Mammography
BI-RADS	Breast Imaging-Reporting and Data System
ACR	American Cancer Society
MIAS	Mammographic Image Analysis Society
TCIA	The Cancer Imaging Archive
AE	Autoencoder
GAN	Generative Adversarial Network
CAE	Convolutional Autoencoder
VAE	Variational Autoencoder
KLD	Kullback-Leiber Divergence
CVAE	Conditional Variational Autoencoder
AAE	Adversarial Autoencoder
CNN	Convolutional Neural Network
COV	Coverage
MMD	Minimum Matching Distance
IS	Inception Score
FID	Fréchet Inception Distance
PSNR	Peak Signal-to-Noise Ratio
MSE	Mean Squared Error
SSIM	Structural Similarity Index
VTT	Visual Turing Test
MLP	Multi Layer Perceptron
EMD	Earth Mover’s Distance
CD	Chamfer Distance
JSD	Jensen-Shannon Divergence
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit

Chapter 1

Introduction

Cancer usually refers to the abnormal proliferation of cells and their unusual modification. The cancer cells aggregate in what are formally called tumors. Breast Cancer (BC) appears within the breast tissue or in the ducts that associate the nipples to the glands of the breasts. Although BC has a low mortality rate, it is the most frequent cancer amongst women, impacting the lives of around 2.3 million women per year, according to the World Health Organization [95]. BC incidence rises worldwide, although developing countries have more prevalence of diagnosed cases.

Diagnosing BC starts with self-examination of the breast and armpit area since the most common symptoms are palpable lumps that are mainly painless [83]. Awareness of BC is raised to women since a very young age from various campaigns to promote the execution of the self-palpable exam [63]. Performing the self-examination exam helps to catch the disease in its early stage, where the tumors are small, and treatment can be more successful [63]. In the absence of self-examination or non-detection of the lumps, BC can go undetected and spread to other regions of the body, such as the lymph nodes present under the arms, especially due to the asymptomatic nature of the disease [83]. Given that, regular imagiology and screening of the chest area in mature adult women are essential for the successful treatment of BC. Depending on the country, regular monitoring of BC can start around the age of 40, where screening programs offer a routine screening of the breast.

Clinical breast examination includes an X-ray of the breast, which is known as the mammography [83, 82]. According to the history of the patient, the density of the breast, and the risk of developing BC, a thoracic Magnetic Resonance (MRI) or an Ultrasound can also be performed to obtain more discriminating results [83, 126].

Treatments include chemotherapy, radiotherapy, surgery, among others [4]. Surgical options include the removal of the entire breast affected, known as a mastectomy, or just a portion containing the tumor, which is called Breast-Conserving Surgery (BCS). The mastectomy used to be the choice for most women undergoing BC removal, as they considered it a safer option concerning the recurrence of the disease. Improvements in the technology and imagiology resources lead to a more precise localization of the tumors, which makes the removal of the entire breast unnecessary sometimes, making BCS a viable option for more and more patients nowadays. BCS is a conservative treatment, and for that, it always requires radiotherapy. The outcome may differ

from expected, due to, for instance, the lack of experience of the surgeon, complications in the preservation of the tissues surrounding the tumor, or retraction of tissues after radiotherapy [4]. These can often disturb the mental health of the patients. The distortion of the breasts can cause patients to develop depression, anxiety, and unpleasant thoughts about their bodies [32]. Every day, physicians make efforts for including the patient in the decision-making process, so that the patient is aware of the results of the BCS.

Automatic tools gathering information from breast data can reduce the time of BC detection and improve the relation with the patient offering visual estimates of the breast after a BCS, for instance [21, 43]. Studies show women are more willing to deal with the outcome of surgery when they are involved in the decision-making process [134]. However, current methods lack satisfactory accuracy for clinical usage, enabling the possibility for improvement [23, 64].

1.1 Motivation

Applications and tools with 3D models of the breast that combine radiological and surface data with information of both interior and exterior parts of the breast are important for facing BC and helping to treat the disease [154]. These tools can be helpful during surgical planning or even in education, providing the user the ability for synthesizing and morphing the data in an anatomical plausible way.

Biomechanical studies that help, for instance, predict deformations from BCS can integrate these tools. However, biomechanical methods often require lots of time and have high computational costs, making them impractical for rapid clinical usage [154].

Nowadays, Machine (ML) and Deep Learning (DL) techniques offer more and more precise results because of the advances in technology [67]. These techniques often integrate Computer-Aided Diagnostics (CAD) systems that help with segmentation and classification of tumors, for instance [74]. These CAD systems help physicians to decide more efficiently and can integrate frameworks to help the interaction between physicians and patients.

To create these tools, datasets that represent populations reasonably are necessary since ML and DL techniques usually need large amounts of data and variety within this data for achieving comparable performance [74]. In medical imaging this is a challenging task, due to the difficulties in gathering data from hospitals, which implies the acceptance from ethical committees. Furthermore, acquiring ground truth data, such as annotation of structures for segmentation tasks or establishing known relationships between points of interest in different modalities of data for registration, or any other specifications needed for the research is expensive and time-consuming. Given the privacy of the data, comparison of techniques between researchers is nearly impossible, unyielding improvements in the area.

Taking all of this into consideration, the creation of synthetic yet realistic breast imagery datasets is indispensable for promoting the collaboration between researchers, to ultimately, develop 3D breast tools with relevant clinical usage. The synthetic data should be designed to model specific

characteristics of interest for multi-purpose applications, such as the validation of new computer vision methodologies, like segmentation and multimodal registration.

DL models can learn data distributions to create semi-synthetic datasets from prior data. While discriminative models, such as convolutional or recurrent neural networks, achieve state-of-the-art results in a variety of computer vision tasks, they do not efficiently learn data distributions. To do so, generative models aim to learn these distributions and have been further developed over the past few years [136]. There are only a few articles that refer to the use of generative modelling in medical data or non-rigid point clouds settings. A thorough search of the relevant literature yielded no other work on how to learn relevant representations of 3D breast data, enhancing our motivation.

1.2 Goals and Contributions

This dissertation focuses on the research of DL strategies to create 3D representations of the breasts, in a plausible and realistic way. Generative DL models are explored to synthesize radiological MRI information of the breast, while geometric DL strategies that consumed point clouds directly are used to learn breast surface distributions, enabling breast shape synthesis and predictive modelling.

The major contributions to the scientific community are:

- Research and deployment of DL methodologies for synthesizing volumetric breast MRI, circumventing the challenges of high variability of structures between different individuals, and promoting 3D continuity between slices.
 - Test state-of-the-art generative DL architectures for 3D MRI reconstruction with breast data.
 - Development of novelty strategies for dealing with breast MRI data by dividing the volumes into smaller patches.
- Research and deployment of geometric DL methodologies for the synthesis of surface point clouds of non-rigid objects, including breast anterior surfaces.
 - Learn breast shape representations and probability distributions from a limited number of exemplar data, testing the advantages of techniques for improving the performance of the models, such as transfer learning.
 - Predictive shape modelling of breast pose transformation from prone to upright (successful mimicking of biomechanical simulations).
 - Prospective study on the intuitiveness of breast representations from generative models.

This dissertation was developed within the scope of the LeGeM (Learning Representations and Generative Models for 3D Breast Data) project (UIDB/50014/2020), a financed seed project at INESC TEC, that the author integrated. The results reported in this dissertation directly result from

the author's work, and the methodologies developed for 3D point clouds of the breast are currently under submission for a scientific journal:

- Bessa, S., Barros, F, Teixeira, J. F., Cardoso, J. S. & Oliveira, H. P. (2021). *Representation and Generative Learning for 3D Breast Shape Modelling*. IEEE Access (in submission).

1.3 Document Structure

The remaining document is structured as follows:

- Chapter 2, Clinical Context - refers to the anatomy of the breast followed by a review of breast cancer and statistics of the disease around the world. Aspects such as the imagiology of the breast will be treated with a brief explanation behind the techniques used. Also, the currently available datasets will be addressed.
- Chapter 3, Deep Generative Modelling - gives a detailed review of state-of-the-art methods for generating synthetic data through Deep Learning. The focus is on Autoencoder and Generative Adversarial Networks based models and how to evaluate them. The current strategies concerning the generation of radiological data are also discussed.
- Chapter 4, 3D Data Representations - main methods for acquiring 3D data are discussed as well as possible representations of this type of data. Furthermore, a review of Deep Learning techniques for dealing with 3D data is also presented, focusing on methods of interest.
- Chapter 6, Generating Breast Surface Scans - describes the methodologies implemented for the generation of data associated with the surface of the breast.
- Chapter 5 Generating Thoracic MRI - presents the main methods tested for the generation of volumetric thoracic MRI.
- Chapter 7, Conclusions - where some final considerations are proposed, along with the main contributions of the dissertation for this research topic.

Chapter 2

Clinical Context

This Chapter addresses the fundamentals in the clinical background of the illness of interest, BC. First, a review of the main aspects of the anatomy and physiology of the breast is presented, followed by an overview of BC itself and a description of the imaging techniques used for diagnosing and characterizing the disease. The main treatments available will also be discussed, centering on breast surgery and its planning. Finally, the publicly available databases for breast imagery are presented.

2.1 Anatomy and Physiology of the Breast

The breasts are present in the chest area (see Figure 2.1), with their base lying over the *pectoralis major* muscle, somewhere between the 2nd and 6th ribs. The Cooper's ligaments, anchored to the dermis, connect the base of the breast to the *pectoralis major* fascia. Breasts are medially contoured by the sternum, superiorly by the clavicles, laterally by the *latissimus dorsi*, and inferiorly by *rectus abdominis* muscle.

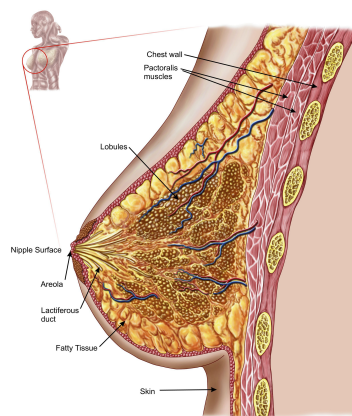


Figure 2.1: Female breast anatomy. Adapted from [143].

The size and shape of the breast vary tremendously from woman to woman. Besides this, all breasts have the same components. These can be divided into exterior and interior.

Exterior parts include the skin which contains the areola, the nipple, and the Montgomery glands. The moderate dermal layer of skin covering the chest is irrigated by arterial perforators both laterally and medially through a robust dermal vascular network. The nipple protrudes outward from the center of the areola and has several openings for allowing the flow of breast milk. The areola is a circular or oval-shaped area at the center of the breast that has a darker color than the skin surrounding it. The small raised, bumpy glands that can appear on the outside of the nipple and areola are the Montgomery glands. These produce a secretion that cleanses and moisturizes the nipple and areola area.

Interior parts include glandular and fatty tissues, Cooper's ligaments, nerves, and lactiferous ducts. Glandular tissue (lobules) produces breast milk whereas fatty tissue determines the size of the breast. The ratio between fatty and glandular tissues depends mostly on the age and hormonal production of each woman. In menopause, this ratio is increased due to the lower estrogen production. These tissues are embedded with fibrous ligaments - Cooper's ligaments -, that allow their suspension, connecting the base of the breast to the muscle underneath it, as previously explained. These ligaments tend to relax with weight and age, leading to the ptosis of the breast. Lastly, lactiferous ducts transport breast milk throughout the nipple.

2.2 Breast Cancer

BC is the most prevalent cancer amongst the female population. All the screening methods and awareness raised for this type of cancer make it a disease with a low mortality rate when compared to other forms of cancer.

The type of BC is determined by the specific cells in the breast that are affected by the disease. Ductal cancer occurs when cancer starts at the milk ducts and lobular cancer when it starts in the milk-producing glands. Ductal cancer is the most common type of cancer. Most breast cancers are carcinomas: tumors that start in the epithelial cells that line organs and tissues throughout the body. Carcinomas in the breast are designated adenocarcinomas. BC type can also depend on whether it spread yet or not: *in situ* BC has not grown into the rest of the breast tissue while invasive BC describes any cancer that has already invaded the surrounding breast. Other types of BC affect other parts of the breast, these are much less common and usually need other types of treatment [4].

2.2.1 Statistics

Globocan is an online database that provides estimates of incidence and mortality within 185 countries and 36 types of cancer, and for all cancer sites combined. In Globocan 2020¹ (most recent data up to today), amongst all new cancer cases during 2020, 11.7% of them corresponded to BC. This outputs 2.3 million newly diagnosed breast cancer cases worldwide. The statistics also show that 24.5% of new cancer cases in females are BC, making it the most frequent cancer for women. Amongst all cancer-related deaths, 6.9% are related to BC, while 15.9% of female

¹<https://gco.iarc.fr/today/data/factsheets/populations/900-world-fact-sheets.pdf>

deaths related to cancer are caused by this disease, being also the cancer site with higher mortality for women. More insight into incidence and mortality in the female population can be inspected in Figure 2.2. Incidence versus mortality numbers shows that the effort given to awareness and screening programs can decrease mortality rates. The incidence of BC in men is less than 1%, so in this work only considers breast cancer in women.

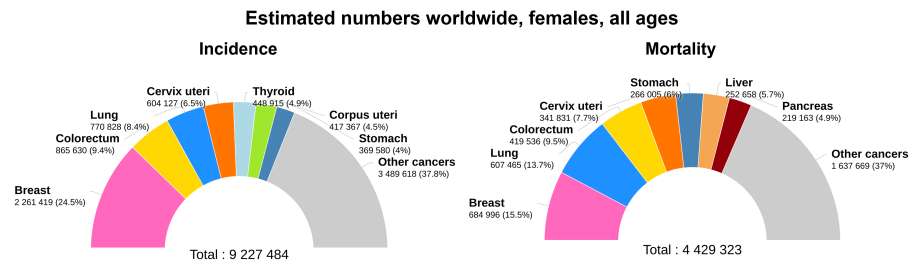


Figure 2.2: Statistics from Globocan 2020². Estimated numbers worldwide for females of all ages for incidence of cancer (on the left) and mortality (on the right).

In Portugal, with a female population of over 5 million women, it is estimated that about 7000 new cases of BC occur each year (19 cases per day), leading to the death of 5 women each day from this illness. Of all these cases, only 5-10% emerge from possible hereditary or genetic reasons, that need earlier treatment and tracking when comparing to other women [73].

2.2.2 Treatments

Breast cancer treatments depend strongly on the stage of the disease and also on other factors, such as the size of the tumor when comparing to the size of the breast itself, the results from the pathological tests, the age of the patients (and for women, whether they have gone through menopause or not), the general health of the patient and family history [4].

After considering all the aspects referred to above, there are different treatment options, and the treatment plans often include a combination of them after evaluating the specific problem of the patient [4].

Surgery involves the removal of the tumor and surrounding margins. Surgery can also find out if cancer has spread (sentinel lymph node biopsy or axillary lymph node dissection), restore the breast shape after BC removal (breast reconstruction), or relieve the symptoms of advanced cancer.

Although surgery is part of 90% of BC treatments, there are other options available for treating besides surgery or combined with it. For instance, radiotherapy is often used post-surgery to destroy any remaining mutated cells by using high-voltage rays. Another viable option is hormone therapy if tumor cells positively stimulated by hormones (estrogen or progesterone). Immunotherapy can also stimulate the immune system of the patient for recognizing and destroying cancer cells more efficiently. Before surgery, chemotherapy can be used to destroy cancer cells, and sometimes it may destroy all BC, so surgery may only be needed to remove the dead tissues. At last, targeted

²<https://gco.iarc.fr/today/online-analysis-pie?v=2020>

therapy is used to attack specific cancer cells without damaging normal cells and sometimes these targeted drugs work even when chemotherapy drugs do not.

Surgery and radiotherapy are considered local treatments since they only treat the tumor without affecting the rest of the body. The other treatments are systemic, as they can reach cancer cells in the whole body with cancer-fighting agents [4].

2.2.2.1 Breast Cancer Removal Surgery

Removing the tumor through breast surgery is very common in BC. About 80% of women with BC need to go through surgery during treatment [85]. Besides the immediate tumor removal, the surgery may give information about whether or not cancer has spread to the surrounding lymph nodes. Also, it helps to reshape the breast and can relieve symptoms in case of advanced forms of the disease [4].



(a) Lumpectomy. Retrieved from [30].

(b) Mastectomy. Retrieved from [113].

Figure 2.3: Breast Cancer Surgery.

There are types of breast cancer removal surgery: Breast-Conserving Surgery and Mastectomy. In Breast-Conserving Surgery (BCS), also known as lumpectomy or partial mastectomy, only the tumor and surrounding tissues are removed (see Figure 2.3a). The location of the tumor determines the amount of tissue that needs to be removed. This surgery is considered the first treatment option for patients with early-stage breast cancer. BCS is recommended if the size of the tumor is small (less than 5 cm of diameter or 20% of volume of the breast) when in comparison to the size of the breast, the tumor is placed in a suitable position for the procedure and there are only cancer cells in one region of the breast [30]. In the case of a mastectomy (Figure 2.3b), all the breast is removed, usually including all breast tissues, skin, nipple, areola, and sometimes surrounding area tissues, such as some of the chest wall. This procedure is recommended on different occasions, such as when a large lump is present in smaller breasts, the tumor is placed in the center of the breast, cancer appears in more than one region, or the patient has had radiotherapy in the past [113].

The breast surgery adequate for treating BC depends on the breast cancer features and the patient features, including their medical history. Despite this, the patient can and should also intervene in these decisions, so they can be aware of the outcomes of performing either BCS or a total mastectomy [4].

Although BCS does not remove the whole breast, performing this surgery for tumor extraction always comes with radiation treatments after the surgery, even for early-stage cancer. When performing a mastectomy in these conditions, it is less likely the patient will need further radiation. BCS also comes with the possibility that the outcome shape of the breast is too modified and different, especially when a large portion of the breast needs to be removed, being the major reason for going with mastectomy. Although studies have proven that there is no influence of the type of BC removal surgery and the recurrence of the disease, some women tend to go for a mastectomy out of fear or due to their family history [4].

Undergoing BC is, for most women, a traumatic experience that comes with a lot of negative emotions, such as depression and anxiety [32], symptoms associated with a higher mortality rate for breast cancer carriers [17]. These symptoms can be linked to concerns of recurrence of the disease and passing it through the next generations of their family. Also, some women report decreased libido due to loss of sensations in the area affected by the BC, and alteration of femininity [2]. The body image of women is impaired after going through breast surgery, especially for women who had gone through a mastectomy without performing breast reconstructing surgery posteriorly.

2.3 Breast Imagiology

Accessing, diagnosing, and evaluating breast pathologies often comes through the analysis of several imaging techniques. The mortality rate for breast cancer decreased by about 40% between 1980 and 2020 [94]. Many studies link the execution of screening programs to this decrease in the mortality rate [8]. Imagiology of the breast is required after self or clinical breast examination when results do not show conclusiveness. Breast cancer prevention plans also demand imaging analysis, such as Portugal's Program for Tracking Breast Cancer [73], where they do mammography as prevention in women from ages 50-69. Other conditions or family history may be other reasons for doing breast imaging routinely.

2.3.1 Radiological Exams

The most common way to address the screening of the breast is through radiological examination. Visualizing breast tissues and surrounding chest structures allows the detection of possible anomalies.

Mammography The mammogram is the most common radiological exam used to diagnose breast cancer [82]. It uses low energy X-rays (around $30kVp$) showing off four different densities (air, fat, water and, metal or calcium), it can further distinguish between different water-density tissues through differences in the thickness and compressibility and by whether the tissues contain calcium or fatty density, being able to detect unpalpable tumors. A lot of countries invite women to perform this exam after they reach a certain age.

An important feature retrieved from the mammogram is the density of the breasts. The attenuation of the X-rays happens mainly due to the present tissues in the breast (fatty and glandular tissues). Different combinations of these result in different breast densities, as shown in Figure 2.4.

The fibroglandular tissue is radiographically dense, including the stroma and parenchyma. Women with high mammographic densities (high proportion of radiographically-opaque stroma and parenchyma) have an increased risk of developing breast cancer when compared with women whose breasts are composed mainly of fatty tissues [96]. This feature has also an important role in pose transformation or postoperative outcomes from breast surgery.

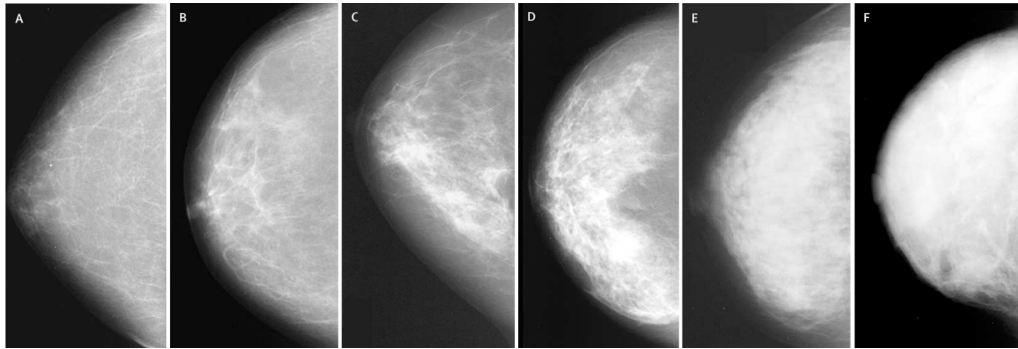


Figure 2.4: Different categories of breast density (in percentage) estimated by radiologists. A=0. B<=10%. C<=25%. D<=50%. E<=75%. F<=75%. Retrieved from [107].

Although this technique is the standard for breast care screening worldwide, the sensitivity of the image can decrease in young women, women with dense breast tissue, and women that carry BRCA (BReast CAncer gene) mutation in their genetical information [58, 98, 12]. In these, other techniques should be used to meticulously inspect results.

Magnetic Resonance Imaging MRI is an important tool when it comes to the diagnosis and evaluation of diseases. When an atomic nucleus has an odd number of protons or neutrons, it possesses a magnetic moment. This magnetic moment describes the strength and direction of a microscopic magnetic field surrounding the nucleus. When a strong and constant external magnetic field (like an imaging magnet) is nearby, the nuclear moments align themselves with this magnetic field, which leads to a measurable macroscopic magnetic moment [108]. The interaction of this magnetic moment with the external field causes each spinning nucleus to change the orientation of its rotation axis. This phenomenon is known as precess and occurs at a characteristic resonance frequency for each nucleus, which is proportional to the strength of the external field and can be calculated through the Larmor equation [108]. The images are then obtained through the density of the protons and relaxation times (i.e., the rate at which the magnetization returns to its equilibrium). Tissues where hydrogen atoms (the most abundant atom in the human body) exist in a significantly higher amount than others, for instance, tissues rich in lipid content (e.g. breasts) will display brighter than the ones where hydrogen atoms appear in lower concentrations (e.g. bone).

Breast MRI is one of the main methods used nowadays for diagnosing breast diseases, conjointly with mammography, ultrasound, and image-guided needle biopsy. To diagnose or exclude cancer, intravenous administration of gadolinium-containing contrast material is needed. Without it, the MRI cannot inform clinically besides the evaluation of the integrity of breast implants. When breast cancer is detected, MRI can be used to determine the tumor extent, since its performance for that

purpose is higher than mammography or ultrasound, as shown in Figure 2.5. It can also be used in the preoperative process since it can detect many additional enhancing lesions unseen within the other techniques. MRI does detect an increased number of false positives when compared to other breast screening techniques. The overall sensitivity of MRI for breast cancer is about 90%. This makes MRI a strong diagnostic tool for the breasts. It is recommended for women with high risks for breast cancer (due to family history) and genetic mutations, such as the BRCA genes [77].

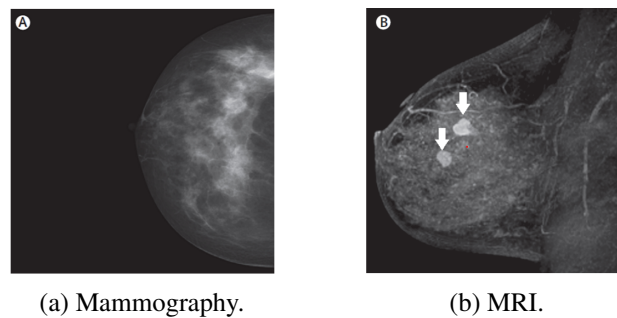


Figure 2.5: Multifocal carcinoma detected on MRI and negative results from the mammography. Retrieved from [107].

Others screening options Besides the mammography and MRI, there are many other BC screening options, such as Ultrasounds, Computerized Tomography (CT), and Positron Emission Tomography (PET). Ultrasounds can distinguish amongst different types of water-density tissues [127] and its waves do not impact the human body with harmful radiation, so they offer a proper alternative for BC diagnosis in younger women [111], or women with higher breast density, when used along with mammography [56, 58]. CT is used in the monitoring process of cancer and uses iodinated contrast media for studying the presence of axillary lymph nodes [126]. PET is mainly used for accessing metastatic disease, since it uses a radioactive substance that when injected into the human body, moves to places where cells have high active rates, enhancing cancerous cells. It is normally not used in the first stage of breast cancer since it does not accurately detect tumors smaller than 5-10 mm [83]. CT and PET scans can be combined for evaluating metastatic cancer.

2.3.2 Surface Information

Surface data of patients can be obtained by two-dimensional (2D) photographs or 3D information [91]. With 2D surface data, physicians compare measurements marked on photographs for assessment of breast aesthetic outcome after BCS, by evaluating the asymmetries between treated and non-treated breasts [20]. Since these methods do not consider the third dimension, they lack the global appearance of the aesthetics results [20, 91].

On the other hand, 3D data of patients can be used as a tool for objective cosmetic evaluation. With it, breasts can be observed from different points of view, estimate the volume ratio before and after surgery (volume/volume deficit), and plan future surgeries. This type of data can be used in simple volumetric analysis or more intricate programs for computing quantitative measurements

and simulating postoperative outcomes through software [91]. Taking this into consideration, the development of these tools could help physicians to better explain the procedures to the patients, showing the patients where the tumor is or what tissues will be removed, improving their communication on what is the most suitable surgical option, and also help on the surgery planning itself [92]. Besides, 3D models of preoperative and postoperative breasts can be used in medical school for helping future physicians to understand how the breasts behave to BCS since this kind of knowledge is mainly acquired by experience [91].

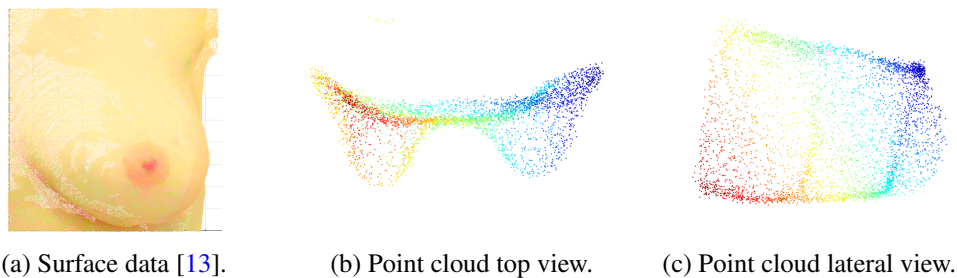


Figure 2.6: Visualization of breast surface data and breast point cloud from two different points of view.

However, to develop breast surgery tools a complete 3D model of the breast is necessary. In Figure 2.6, a surface model of the breast as well as a point cloud (PCL) of the surface are presented. In [92], a review of the main available technology for BCS planning is presented, describing the main challenges for obtaining these models.

Over the past years, different strategies have been developed for obtaining 3D breast models. 3D laser scanning images are obtained through the principle of triangulation: a laser beam projected onto the patient's torso, while the reflected rays are collected through a detector that is sensitive to the orientation of these rays [90]. 3D photographic images can be obtained through the perception of the objects from multiple viewpoints [90], creating multiple camera stereo systems to capture the surface of the breast, such as in [48], that used 4 pods and 8 cameras. However, most systems for 3D surface imaging are heavy, bulky, and expensive, limiting their usage to just one room or a single hospital. With that in mind, newer versions of equipment are being developed with more user-friendly features (more portable and cost-effective). One of these is the Microsoft Kinect device used in Xbox360 and Windows gaming consoles. It generates colored 3D models through a pseudo-structured light scanning approach, calculating the distance to the objects in the field of view using a 3D depth camera. The free software Kinect for Windows concedes the acquisition of depth maps and consequently the point cloud of the object. In [141], a study was conducted on breast surface imaging using this technology and results show acceptable agreement between manual measurements and Euclidean distances between anatomical landmarks with the Kinect-based system. The system used two calibrated Kinects (minimum for acquiring a complete PCL) and the data from it – RGB and depth cameras – acquired the models within 2 seconds.

2.3.3 Available datasets

Image analysis requires the existence of relevant image databases, and, in many clinical applications, there is also a need for ground truth annotations given by specialists. Most publicly available datasets for breast cancer are from mammography exams since it is the most used diagnostic imagery tool for this disease. Others may include combinations of mammograms and MRIs but exist in a smaller scope.

Concerning 3D surface information of the breast, the research conducted could not find any available datasets. This happens mainly because 3D models are mostly acquired in private studies with strategies that are not made public, jeopardizing comparison. There are also no traces of available datasets combining surface information with other radiological exams, enhancing the motivation for creating multimodality datasets, and sharing them with other researchers.

In the case of mammograms, one of the most used databases is DDSM (Digital Database for Screening Mammography) [47], MIAS (Mammographic Image Analysis Society) [130], which is the oldest mammogram database, BC-DR (Breast Cancer – Digital Repository) [72] and the INbreast [81]. BC-DR and INbreast are only available upon request. All of these contain some form of additional information besides the images such as density and lesion annotations and classification, but MIAS annotations are not categorized according to the Breast Imaging-Reporting and Data System (BI-RADS) categorization system, from the American Cancer Society (ACR).

Besides these, **The Cancer Imaging Archive (TCIA)**³ also provides a variety of databases for breast cancer. Although the TCIA collections include all sorts of cancer-related datasets, fast filtering over them, provides a series of databases for breast cancer, a screenshot of these can be found in Figure 2.7.

After further investigation of the available datasets on breast MRI (11 of 19 entries – one of them corresponding to mouse and not human data), most of them do not have ground truth annotation data regarding most structures present in the scans. For instance, the collection *Duke-Breast-Cancer-MRI* offers data from 922 patients with invasive breast cancer, but only contains annotations for the lesions (tumors). Another large collection *ACRIN-Contralateral-Breast-MR (ACRIN 6667)* only offers the DICOM data (the scans itself) and clinical data is only available upon request. A smaller collection, *QIN-BREAST-02*, contains scans of females diagnosed with invasive breast cancer prior and after treatment, but the data is limited (only 13 patients) and it is only available upon request. Taking this into consideration, there are no relevant datasets made available for studying and develop research with ML or DL algorithms, as most of them lack the specifications needed for these methodologies to achieve comparable performance.

³<https://www.cancerimagingarchive.net/>

⁴<https://www.cancerimagingarchive.net/collections/>

Show 125 entries

Filter table: breast cancer

Collection	Cancer Type	Location	Species	Subjects	Image Types	Supporting Data	Access	Status	Updated
Duke-Breast-Cancer-MRI	Breast Cancer	Breast	Human	922	MR	Clinical, Genomics, Image Analyses	Public	Complete	2021-04-06
CMMB	Breast Cancer	Breast	Human	1775	MG	Clinical	Public	Complete	2021-04-06
ACRIN-Contralateral-Breast-MR (ACRIN 6667)	Breast Cancer	Breast	Human	984	MR, CR	Clinical	Limited	Complete	2021-03-05
CPTAC-BRCA	Breast Cancer	Breast	Human	134	Pathology	Clinical, Genomics, Proteomics	Public	Ongoing	2021-02-02
Breast Cancer Screening-DBT	Breast Cancer	Breast	Human	985	DBT	Clinical	Public	Ongoing	2021-01-15
ACRIN-FLT-Breast (ACRIN 6688)	Breast Cancer	Breast	Human	83	PET, CT, OT	Clinical	Public	Complete	2020-01-24
QIN-Breast	Breast Cancer	Breast	Human	67	MR, PT, CT	Clinical	Public	Complete	2020-01-10
Post-NAT-BRCA	Breast Cancer	Breast	Human	64	Pathology		Public	Complete	2019-10-17
SLN-Breast	Breast Cancer	Breast	Human	78	Pathology		Public	Complete	2019-08-06
QIN-BREAST-02	Breast Cancer	Breast	Human	13	MR	Clinical	Limited	Complete	2019-07-10
VICTRE	Breast Cancer	Breast	Human	2994	MG		Public	Complete	2019-03-08
CBIS-DDSM	Breast Cancer	Breast	Human	1566	MG	Image Analyses	Public	Complete	2017-09-27
ISPY1 (ACRIN 6657)	Breast Cancer	Breast	Human	222	MR, SEG	Clinical, Image Analyses	Public	Complete	2016-08-31
Breast-MRI-NACT-Pilot	Breast Cancer	Breast	Human	64	MR, SEG	Clinical, Image Analyses	Public	Complete	2016-01-26
Mouse-Mammary	Breast Cancer	Abdomen	Mouse	32	MR		Public	Complete	2015-03-30
TCGA-BRCA	Breast Cancer	Breast	Human	139	MR, MG, Pathology	Clinical, Genomics, Image Analyses	Public	Complete	2014-12-30
QIN Breast DCE-MRI	Breast Cancer	Breast	Human	10	MR, KO	Clinical, Image Analyses	Public	Complete	2014-07-31
Breast Diagnosis	Breast Cancer	Breast	Human	88	MR, PT, CT, MG	Clinical, Image Analyses	Public	Complete	2011-11-09
RIDER Breast MRI	Breast Cancer	Breast	Human	5	MR		Public	Complete	2011-11-08

Showing 1 to 19 of 19 entries (filtered from 142 total entries)

Figure 2.7: TCIA breast cancer results⁴. Accessed on April 23, 2020.

2.4 Summary

In this Chapter, some anatomical aspects concerning the breasts were covered. Most relevant screening methods for detecting breast pathologies were described, revealing their differences, and demonstrating their completeness when combined in some clinical cases. It is important to recognize that breasts are anisotropic tissues that can undergo extreme deformations, which makes them hard to model in computer vision.

Breast cancer is the most common cancer in females, and it can be described by the cells it affects, the appearing site, or even its spreading. Diagnosis and treatments vary according to a variety of physical and background factors. Nevertheless, breast surgery is a very common treatment for BC. The decision-making process on whether to perform a total mastectomy of BCS should be done with the consciousness of the patient on possible outcomes.

Breast databases available nowadays were presented and denoted that in most cases these do not comprise multi-modality images. Also, there are no 3D surface information databases for the breasts, neither a combination of the latter with radiological imagery, which is helpful during surgery planning. Besides, most MRI collections do not provide ground truth annotations, which are required for most automatic methods to perform adequately.

Chapter 3

Deep Generative Modelling

The generation of semi-synthetic datasets that combine breast MRI and surface information is essential for the creation of medical tools for BC surgery planning. These tools can integrate CAD systems that help to diagnose BC, offer visual estimates of the surgery outcome, providing image and shape retrieval for the current patient's data, represent the tissues to be removed during surgery, being valuable for both the physicians and the patients. Beyond this, the augmentation of these types of data can validate some ML and DL techniques that aim to predict surgical outcomes for breast cancer treatments, such as [154]. To understand how these datasets can be synthesized, the relevant state-of-the-art techniques used for generating data need to be explored. Generative Models are based on DL architectures and aim to learn distributions of data so that they can generate synthetic data. In this Chapter, the main architectures of interest are explored, mainly based on Autoencoders (AE) or Generative Adversarial Networks (GAN). An overview of some relevant state-of-the-art techniques used in radiological data will also be presented.

3.1 Overview

Generative modelling can be considered as an unsupervised learning task in Deep Learning. Generative models have multiple descriptions and can be used for a wide variety of tasks. Ian Goodfellow describes a generative model as any model that takes a training set, consisting of samples drawn from a prior distribution, and learns to represent an estimate of that distribution, resulting in either a probability distribution or samples from it [39]. They have the capability of generating data not only for estimating density, being valuable in:

- Generating synthetic but realistic samples.
- Completing missing parts of data, since they give an idea about data manifold, which is called manifold learning [51].
- Manipulation of original samples based on predefined features – latent codes. These techniques can be used for manipulating pose, image parts, objects, such as in [36], [18],

and [103], but also to image-to-image translation tasks (where images are mapped from one domain to another), such as in [52], [70] or [153].

- Training a single model for multi-model output [46].
- Obtaining better models through the acquisition of more training data from the same distribution – data augmentation, as in [61], [76], and [148].
- Generation of super-resolution images, enhancing data quality [114], [149], [62].

Generative Models can also be used for data augmentation. Data augmentation comprises the expansion of the original dataset with artificial samples. Traditional techniques for data augmentation include the addition of random noise to the data or the application of geometric or other transformations (rotation, flip, etc.). Yet these are domain-agnostic transformations since they do not consider the nature, characteristics, and features of the data, producing only synthetic samples that come from the original distribution [76]. Augmenting datasets with meaningful and diverse samples is a more challenging task. To do so, Generative Models need to produce samples that belong to the original data distribution, but also differ from every sample in it.

Generative models have many categories, and this section will focus on the ones that are considered the most significant for the purposes of this work, which are Autoencoder (AE) and Generative Adversarial Network (GAN) based models. Some of the evaluation methods for generative models will also be described.

3.1.1 Autoencoder based models

Autoencoders (AE) can be defined as unsupervised Deep Learning models where the input is the same as the output, so they are trained to reconstruct their input data [9].

These types of models are usually composed of an Encoder E , a bottleneck region, and a Decoder D . The Encoder compresses the high-dimensional (n) input into a low-dimensional (p) latent representation z , in the bottleneck region. The latter goes through the Decoder that reconstructs the z into the output of dimension n , acting as a generator [136]. To do so, the functions $E : \mathbb{R}^n \rightarrow \mathbb{R}^p$ and $D : \mathbb{R}^p \rightarrow \mathbb{R}^n$ need to be learned to satisfy the following equation:

$$\arg \min_{E,D} \mathbb{E}[\mathcal{L}(x, D \circ E(x))] \quad (3.1)$$

where $\mathbb{E}$ is the expectation over the distribution of x , and $\mathcal{L}$ is a reconstruction loss function, measuring the differences between the input of the network and the output [9].

In their most popular form, E and D are neural networks and for imagery, convolutional neural networks - **Convolutional Autoencoders (CAE)**, as represented in Figure 3.1. In CAEs, the traditional fully connected layers from the AEs are replaced by convolutional and pooling layers for the encoding to extract useful features. For the decoding, upsampling layers are used in addition to the ones previously mentioned, for reconstructing images [54].

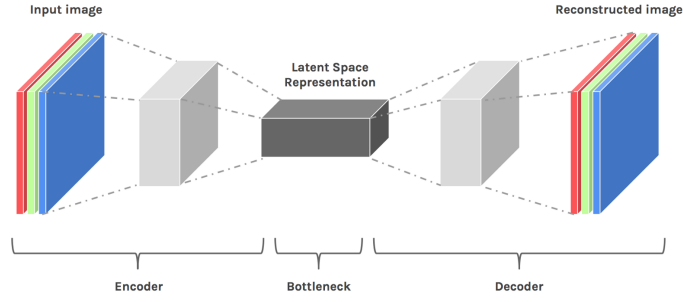


Figure 3.1: Convolutional Autoencoder architecture. Retrieved from [54].

3.1.1.1 Variational Autoencoders

The major limitation of AEs to the generation of data is the unstructured and unorganized bottleneck region, since the encoder outputs a single value for each encoding dimension, producing a non-linear latent space, that does not allow interpolation. In other terms, some points of the latent space in an autoencoder without explicit regularization will decode into meaningless representations. To overcome this issue, **Variational Autoencoders (VAE)** [57] describe the observations in the latent space in a probabilistic manner, so instead of outputting a single scalar value for each attribute in the latent space, the encoder unveils a probability distribution (commonly a Gaussian-shaped one). VAEs differ from other types of regularized AEs such as Sparse, Denoising, and Contractive autoencoders due to their statistical motivations. Let us consider the data x and latent variables z in VAE. These are modeled by the prior probability distributions $p(x)$ and $p(z)$, respectively. The model aims to infer values for the latent variables z that represent well the observed data, that is, $p(z|x)$, which can be computed by Bayes rule:

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)} \quad (3.2)$$

The denominator in Equation 3.2 can be calculated through Equation 3.3. By computing the integral in Equation 3.3 all possible configurations of latent variables are evaluated - for each candidate z , it is assessed if it can generate x , which isn't practical.

$$p(x) = \int p(x|z)p(z)dz \quad (3.3)$$

To overcome this, $p(z|x)$ can be approximated with $q_{\lambda}(z|x)$ distribution, that will be used as a proxy of $p(z|x)$ to produce predictions about future data. The λ refers to the parameters of the distribution, in a Gaussian distribution $\lambda = (\mu, \sigma^2)$. Evaluating the efficiency of the distribution to approximate $p(z|x)$ can be done through the Kullback-Leiber Divergence ($\mathcal{D}_{KL}$). Given so, the λ

parameters of the q distribution are obtained for minimizing the $\mathcal{D}_{KL}$, through Equation 3.4.

$$\begin{aligned} \arg \min_{\lambda} \mathcal{D}_{KL}(q_{\lambda}(z|x) || p(z|x)) &= \\ = \arg \min_{\lambda} \mathbb{E}_q[\log q_{\lambda}(z|x)] - \mathbb{E}_q[\log p(x, z)] + \log p(x) & \quad (3.4) \\ = \arg \min_{\lambda} \mathbb{E}_q[\log p(x, z)] - \mathbb{E}_q[\log q_{\lambda}(z|x)] + \log p(x) \end{aligned}$$

The minimization proposed in the Equation 3.4 is unmanageable since it again implies the calculation of $p(x)$. The $\mathcal{D}_{KL}$ is always greater or equal to zero and $\log p(x)$ does not depend on q , so minimizing the expression above is equivalent to maximizing the expression below (Equation 3.5), which is known as evidence lower bound (*ELBO*).

$$ELBO = \mathbb{E}_q[\log p(x, z)] - \mathbb{E}_q[\log q_{\lambda}(z|x)] \quad (3.5)$$

To conclude, in VAE, the encoder will take x and learn the optimal distribution $q_{\lambda}(z|x)$, returning the parameters that describe the distributions for the attributes of the latent spaces. Then the decoder learns $q_{\lambda}(x|z)$ in order to output the reconstructed object, $D \circ E(x) = \hat{x}$ from the distributions of the latent space, as seen in Figure 3.2.

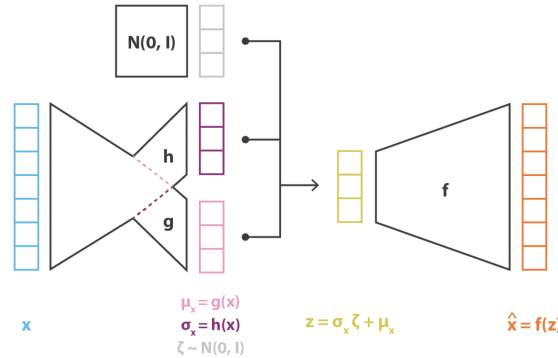


Figure 3.2: Variational Autoencoder Architecture. The x stands for the input data, mean vector is represented as μ and covariance with σ , which are calculated through independent functions $g(x)$ and $h(x)$, respectively. The z is the latent representation and $\hat{x}$ refers to the reconstruction. Retrieved from [106].

In VAE models, usually, there are two terms for the loss function:

- The reconstruction term, which is the same as in the AE model, measuring the resemblance between the input and the output of the network, $\mathcal{L}(x, \hat{x})$.
- The regularization term, which regularizes the latent space, *ELBO*.

Combined, they result in the following Equation:

$$\arg \min_{\lambda} \mathcal{L}(x, \hat{x}) + ELBO \quad (3.6)$$

To demonstrate the differences between AE and VAE in the latent space, in Figure 3.3a shows the resulting latent space distribution of an autoencoder model trained with MNIST dataset (handwritten digits), where it is visible that some parts of the latent space are totally unoccupied, leading to meaningless representations through the decoding. In Figure 3.3b, a VAE model trained with only the $\mathcal{D}_{KL}$ loss term results in a dense latent space, but randomly distributed, so it is impossible to decode meaningful representations since the model did not learn any meaning from the data. Lastly, in Figure 3.3c, the combination of the reconstruction and regularization losses results in a latent space that maintains the similarity of nearby encodings on a local scale through clustering and very densely packed when compared to the one in Figure 3.3a.

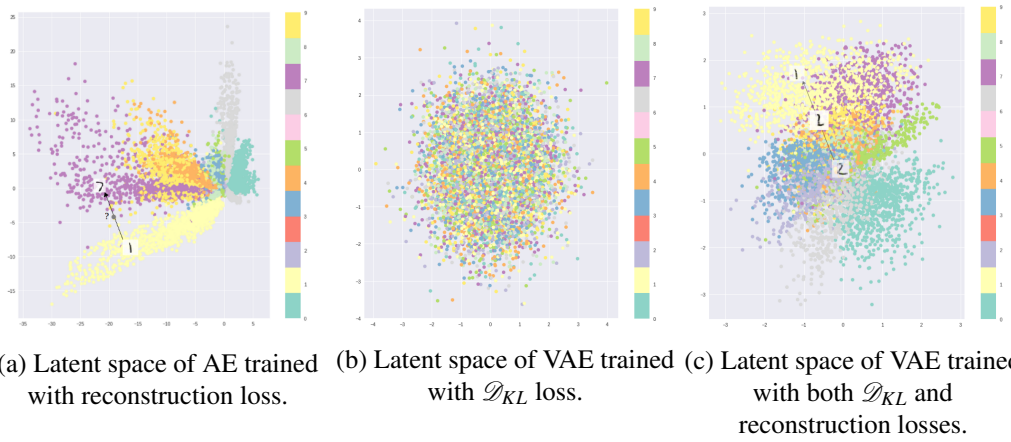


Figure 3.3: Encodings of MNIST data from different AE based models. Retrieved from [118].

3.1.1.2 Conditional Variational Autoencoders

Despite VAE models create an organized latent space, they do not allow to generate specific data because the encoder models the latent variable directly based on the non-latent variable. Similarly, this phenomenon occurs on the decoder since it models the non-latent variable on the latent one. The **Conditional Variational Autoencoder (CVAE)** is an extension of VAE that tries to overcome this issue by adding the label of the corresponding input in the encoder and the decoder. This label can be anything, if it is categorically distributed, this can be accomplished by representing the label as a one-hot vector of the chosen category [125].

To accomplish this, the *ELBO* calculation (Equation 3.5) suffers some alteration, as represented in Equation 3.6.

$$ELBO_{CVAE} = \mathbb{E}_q[\log p(x|y, z)] - \mathbb{E}_q[\log q_\lambda(z|x, y)] \quad (3.7)$$

Although VAE models may offer a reasonable alternative for generating data, they do not always work, since the integral of the $\mathcal{D}_{KL}$ only corresponds to an analytic solution for some classes of distributions or the fact that the representation of the latent space with a discrete variable does not allow the back-propagation, leading to the addition of reparameterization tricks in training. For defeating these problems, the **Adversarial Autoencoders (AAE)** [75] were created. These

include an AE as a generative network, and another network – the discriminator - for the adversarial training, as in GANs, which will be explained in the following section. The second network learns to distinguish between samples generated from the AE and user-defined prior distribution. They are trained in two separate phases: in the first phase, only the AE is trained to learn how to reconstruct the samples; then, a regularization phase trains both the AE and the discriminator to properly distinguish between the output of the generator and the original data.

3.1.2 Generative Adversarial Network based models

Generative Adversarial Networks are used for semi-supervised and unsupervised learning. They have gained popularity amongst the computer vision community for data generation tasks, and the main idea of the network is to generate data from scratch [24].

A GAN consists of a pair of networks competing: the generator G and the discriminator D . These networks set up a min-max game, where G creates samples that are intended to come from the same distribution as the training data, but without ever accessing this data, and D examines the generator's output and tries to determine whether it is real or fake [39]. The competition established between the two network models leads them to increasingly improve each other: as the training progresses, the generator learns how to create more and more realistic data, while the discriminator enhances its ability to classify the data as real or fake. Figure 3.4 schematizes this process.

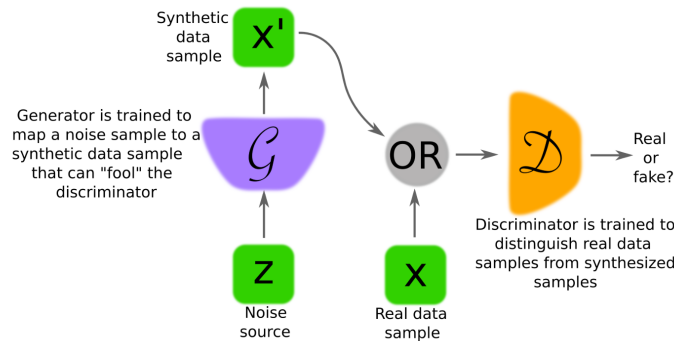


Figure 3.4: GANs basic framework. Retrieved from [24].

Since the generator does not directly access the real data samples, its only learning mechanism is through the interaction it has with the discriminator. The latter has access to both real data and the synthetic data generated by G , which also concedes the error signal of the network, through the ground truth of knowing if the sample is real or fake. The same error signal is then passed to the generator, enabling the creation of more and more realistic samples as the train proceeds.

The generator and the discriminator must be differentiable functions, and, in most cases, they consist of multi-layer neural networks with convolutional and/or fully connected layers. The generator acts as a decoder from the previously explained AE models since it maps an input noise sample z from some representation space (latent space) to the space of the data, then $G : G(z) \rightarrow \mathbb{R}^{|x|}$, where $z \in \mathbb{R}^{|z|}$ is a sample from the latent space, taken from the prior $p_z(z)$, $x \in \mathbb{R}^{|x|}$ is a real sample,

and $|\cdot|$ denotes the number of dimensions. G then produces a synthetic data sample, x' , and the network can be represented as $G(z; \theta_g)$, where θ_g refers to the network's parameters. As for the discriminator, $D(x)$, represents the probability that x came from the real data distribution (p_{data}) instead of the generator distribution (p_g). So $D : D(x) \rightarrow (0, 1)$, classifying samples as either belonging in the training data – real, close to 1 – or from the generator – fake, close to 0. This mapping can be represented as $D(x; \theta_d)$, where θ_d represented to the network's parameters [40].

Concerning the training process, global optima of GANs is encountered when $p_g = p_{data}$, as demonstrated in [40]. The discriminator is trained to maximize the probability of assigning the correct label to the training samples or the synthetic ones. Simultaneously, the generator is trained to minimize $\log(1 - D(G(z)))$. They interact through a minmax game with value function $V(D, G)$:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} \log(1 - D(G(z))) \quad (3.8)$$

Understandably, Equation 3.8 alone does not provide a good learning basis for the GAN model, since, at the beginning of the training process, G is poor at creating new samples, allowing D to have high accuracy, since generated samples are different from the real ones. Given this, $\log(1 - D(G(z)))$ saturates and for overcoming that issue G can be trained to maximize $\log D(G(z))$ instead. Besides this, D cannot be optimized to completion before the optimization of G , due to computational costs and over-fitting. To surmount this problem, Unrolling GANs [79] can be used, where a local optimum of the discriminator parameters can be obtained as the fixed point of an iterative optimization procedure. By unrolling for k steps, a surrogate objective for the update of the generator is created. Through this, the discriminator is maintained near its optimal solution, pending on the slow changes that occur in G .

Although the lack of an explicit representation of $p_g(x)$ and the difficulties in training are considered the main disadvantages for using GANs, adversarial models have a statistical advantage since the generator does not learn through the real data but the gradients that come from the discriminator, which makes the model less prone to overfitting.

Despite all of this, GANs have common issues that are not trivial to solve:

- Mode collapse, where the generator only produces a limited variety of samples, even if it performs well generating these samples, it does not generalize for all the data distribution.
- Diminished gradients, which happens when the discriminator gets too successful and disables the learning capabilities of the generator.
- Lack of convergence, when the network's parameters fail to converge, even if they oscillate.
- High sensitivity to hyperparameters.

3.1.2.1 Convolutional GANs

Convolutional GANs were introduced for the usage of GANs framework for image applications since Convolutional Neural Networks (CNNs) are apt for modelling image data. One of the most

popular families of networks is Deep Convolutional GAN (DCGAN) [102], where the generator and discriminator correspond to deep convolutional networks. Fractionally-strided convolutions in G, and strided convolutions in D, which enables the network to learn their spacial downsamplings. The generator network is schemetized in Figure 3.5.

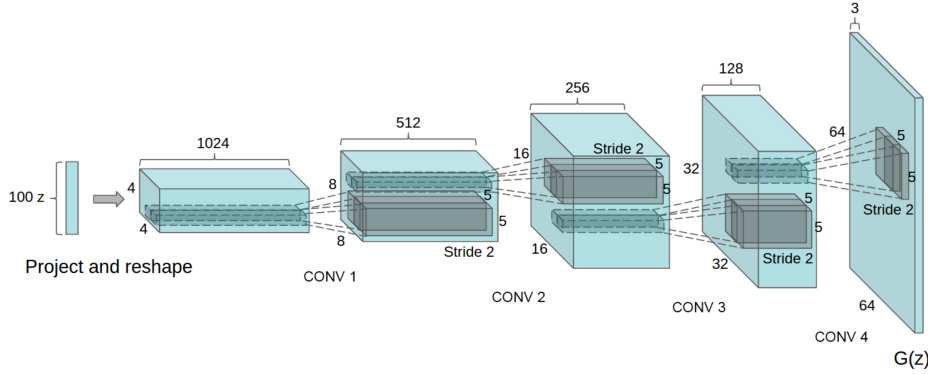


Figure 3.5: DCGAN generator architecture. Retrieved from [102].

Other features of DCGAN include the use of batch normalization, which stabilized the training, helps the gradient flow of deep models, preventing mode collapse. Also, DCGAN also pays attention to the correct choice of activation functions: generator uses ReLU (Rectified Linear Unit) in all layers except the last one, where a hyperbolic tangent is applied, and the discriminator uses Leaky ReLU in all layers since it is adequate for high-resolution images.

3.1.2.2 Conditional GANs

For a **conditional setting**, [80] transformed the GAN original framework into one where the G and D into class-conditional networks, allowing representations for multi-modal data generation, as shown in Figure 3.6. Another possibility for conditional GANs is the InfoGAN [19], present in Figure 3.7. The InfoGAN decomposes the noise source into an incompressible source and latent code, intending to discover latent factors of variation by maximizing the mutual information between the latent code and the output of the generator. The latent code is then used for discovering object classes in an unsupervised learning approach.

Considering that a conditional GAN learns a mapping from an observed image x and random noise vector z , to produce y , that results in $G : \{x, z\} \rightarrow y$. This objective can be denoted by the following equation:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x,y} [\log D(x, y)] + \mathbb{E}_{x,z} [\log (1 - D(x, G(x, y)))] \quad (3.9)$$

GANs have been used in many areas of research and new applications for adversarial training of deep networks is an active area of research. Recent frameworks for image synthesis with GANs aim to improve image quality and utility of the generation capabilities, such as LAPGAN [27].

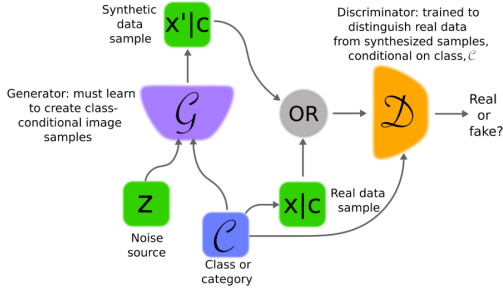


Figure 3.6: Conditional GAN [80] for class-conditional image synthesis (the discriminator performs class-conditional discrimination of real from fake images). Retrieved from [24].

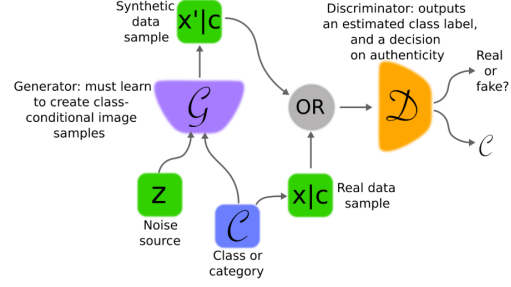


Figure 3.7: InfoGAN proposed by [19] with a discriminator that estimates class label. Retrieved from [24].

3.1.2.3 Domain translation with GANs

Concerning **image-to-image translation**, the *pix2pix* model introduced in [52] offers a conditional GAN that can serve as a general solution since it is not application-specific. The original generator is U-Net based architecture [109], that uses skip connections, which have been proven to produce results with higher quality, and the discriminator is a PatchGAN classifier, that only penalizes structures at the scale of image patches, similar to the one presented in [65].

In *pix2pix*, for the loss function of the generator, a L_1 distance term was added to measure similarity of corresponding real and synthetic images. The L_1 term has been proven to help generators synthesize more realistic images, since it performs a pixel-wise comparison of the pair, capturing low frequency information on the images, which results in less blurry images than the L_2 loss. The train of the discriminator is unchanged, but the generator is trained to both fool the discriminator and to be near the output in a L_1 sense. The PatchGAN in the discriminator is, consequently, more focused on the high frequency components. Considering the same notation as in Equation 3.9, the proposed loss function is described in Equation 3.10, where the λ controls the importance of the L_1 distance term. Despite this additional term reducing artifacts in generated images, the latter may be blurred or grayish.

$$\min_G \max_D V(D, G) = \mathbb{E}_{x,y} [\log D(x, y)] + \mathbb{E}_{x,z} [\log (1 - D(x, G(x, y)))] + \lambda \mathbb{E}_{x,y,z} [\|y - G(x, z)\|_1] \quad (3.10)$$

A variation from *pix2pix* is the *pix2pix_{HD}*, proposed in [139], which aims to produce images with higher resolution by using a coarse-to-fine generator and multi-scale discriminators. The *pix2pix_{HD}* coarse-to-fine generator is detailed in Figure 3.8. It consists of two networks: a global generator network G_1 and a local enhancer network G_2 . These consist of a convolutional front-end (G_1^F and G_2^F), followed by a sequence of residual blocks (G_1^R and G_2^R), and, finally, a convolutional back-end (G_1^B and G_2^B). Despite this, G_2 operates at a greater resolution than G_1 . These networks

are connected and have the following interactions: in the local enhancer, a semantic label map is passed as input to G_2^F , which outputs a set of feature maps with half the size of the input and, for the global generator, a two times downsampled version of the semantic label map and this is fed to G_1 as input. The feature maps resulting of G_1^B are summed pixel-wise to the set of feature maps that resulted of G_2^F , and this constitutes the input of G_2^R , ending the forward pass on G_2 .

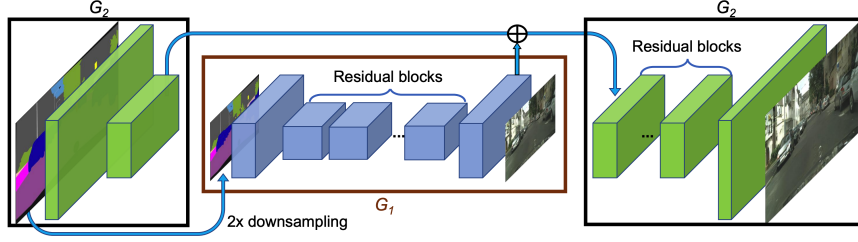


Figure 3.8: Coarse-to-fine generator of *pix2pixHD* architecture. Retrieved from [139].

The multi-scale discriminators have similar structure but operate at different resolutions, so each is trained with images at different scales, obtained by downsampling twice the original and synthetic samples. This forms a 3 level image pyramid. With this strategy, using a single discriminator with a high receptive field is avoided, which would be necessary due to the high resolution of images. The latter would require the usage of a deep architecture or large convolutional kernels, which are linked to overfitting. Also, the multi-scale discriminators have proved to avoid repetition of patterns.

The loss function described for *pix2pixHD* combines the classical GAN loss, modified to optimized the 3 discriminators, and two extra terms for evaluating feature matching and perceptual loss.

For **feature matching**, the L_1 distance is computed between the same intermediate feature maps of the discriminator on the synthesized image and the corresponding original one. For a discriminator D_k , this is computed as follow:

$$\mathcal{L}_{FM}(G, D_k) = \mathbb{E}_{x,y} \sum_{i=1}^T \frac{1}{N_i} \left[\left\| D_k^{(i)}(x, y) - D_k^{(i)}(x, G(x)) \right\|_1 \right] \quad (3.11)$$

where $D_k^{(i)}$ corresponds to the i^{th} layer feature maps of the discriminator D_k , T is the total number of layers, and N_i is the number of elements of the i^{th} layer.

The **perceptual loss** computes the L_1 distance between the same intermediate feature maps of the VGG network [123], which is obtained with the usage of the synthetic sample and the original one. It is calculated through Equation 3.12, where $F^{(i)}$ represents the i^{th} layer with M_i elements on the VGG network.

$$\mathcal{L}_{perceptual}(y, G) = \sum_{i=1}^N \frac{1}{M_i} \left[\left\| F^{(i)}(y) - F^{(i)}(G(x)) \right\|_1 \right] \quad (3.12)$$

All the terms combined result in the final loss function, present in Equation 3.13.

$$\min_G \left(\left(\max_{D_1, D_2, D_3} \sum_{k=1,2,3} \mathcal{L}_{GAN}(G, D_k) + \lambda \sum_{k=1,2,3} \mathcal{L}_{FM}(G, D_k) \right) + \lambda \mathcal{L}_{perceptual}(y, G) \right) \quad (3.13)$$

The *pix2pix_{HD}* also features a binary boundary map. In it, pixels belonging to the edges of all object instances across all classes are 1 in the semantic label. This prevents different objects of the same class to merge in the map, since semantic label maps do not differentiate objects from the same class. The instance boundary map is concatenated with a one-hot vector that represents the label map, and it's fed as input to the generator.

For performing one-to-many mapping, that is, producing various samples from the same semantic label map, in *pix2pix_{HD}*, a autoencoder network is trained to embed visual features for each object instance in the ground truth sample. This corresponds to the instance-level feature embedding, and the new samples are synthesized concatenating the output of the AE network with the semantic label map and provide it as the input of the generator. The manipulation of the visual features before the concatenation will provide different results.

The *pix2pix* and *pix2pix_{HD}* require paired data, which is often hard to obtain especially in medical imaging, where annotations must be provided by specialized professionals. To overcome this, an unsupervised GAN based architecture was created for image-to-image translation without paired data, the **CycleGAN** [152], that takes four neural networks: two generators and two discriminators: the generator G translates each image from a domain X to a domain Y and the discriminator D_Y analysis the resulting images from G by calculating the resemblance of the latter with the images from the actual Y domain; while this happens, the generator F does the opposite of G , transforming images from a domain Y to a domain X and being judged by the discriminator D_X . CycleGAN aims to learn appropriate G and F generators according to the training data available at each domain.

The CycleGAN (see Figure 3.9) uses a combination of the classical GAN loss and a **cycle consistency** term, which prevents contradictions between the mappings learned by F and G , since when translating an image from a domain X to Y . Ideally, if the same image is then transformed into domain X again, the result should be identical to the original one (and vice-versa). The cycle consistency is achieved through the computation of the L_1 distance between each image and the corresponding reconstruction for each domain:

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)} [\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{data}(y)} [\|G(F(y)) - y\|_1] \quad (3.14)$$

The cycle consistency term is controlled by λ , giving the total optimization function present in Equation 3.15.

$$(G, F, D_X, D_Y) = \arg \min_{G, F} \max_{D_X, D_Y} \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, X, Y) + \lambda \mathcal{L}_{cyc}(G, F) \quad (3.15)$$

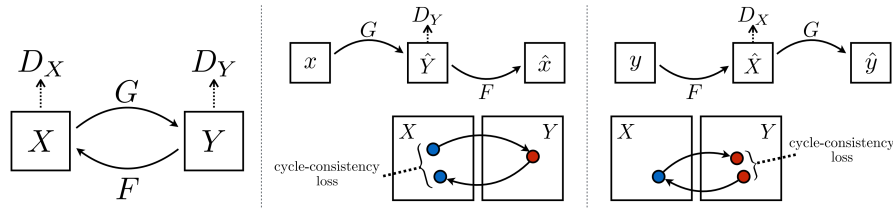


Figure 3.9: CycleGAN training cycle. On the left, the two mapping functions of the CycleGAN: $G : X \rightarrow Y$ and $F : Y \rightarrow X$. In the center, the forward cycle-consistency loss: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$. On the right, the backward cycle-consistency loss: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$. Retrieved from [152].

The models presented so far only correspond to a small fraction of current methods for generative modelling. A more precise insight of the generative models is presented in [136], where an extensive review of these models is presented as well as relationships between certain models.

3.1.3 Evaluation of generative models

Besides the fact that training a generative model is not a trivial achievement, evaluating these models is also a daunting task. Several metrics for evaluating generative models have been proposed over the years, but there is no consensus on which ones are better for measuring the strengths and limitations of the models and should be used for enabling model comparisons amongst researchers. In [10], they make an extensive critical review on this topic, and in this document only some measures of interest are described.

Measures for evaluating Generative Models are divided into two main categories: quantitative and qualitative. Both have strengths and limitations, for instance, one may think that fooling a person with the synthetically generated sample is the perfect measure for the model, but a qualitative measure for this may favor models that concentrate on limited sections of the data, such as GAN models that suffer from mode collapse. Quantitative measures are less subjective but also may not correspond to how humans perceive and judge the generated data. Complications like these and others, such as the variety of probability criteria and the lack of perceptually meaningful image similarity measures, increase the difficulty in evaluating generative models [133].

Before passing on to describing some of the existing metrics, some of the properties that a good evaluation metric should have need to be understood [10]:

- Favor models that produce high fidelity samples.
- Favor models that generate diverse samples, being sensitive to overfitting or mode collapse.
- Favor models with disentangled latent spaces and space continuity.
- Have well-defined bounds (lower, upper, and chance).

- Be sensitive to distortions and transformations of the samples, since many Generative Models use domain agnostic data augmentation techniques that do not change the semantic meaning of the samples.
- Be in harmony with human perceptual judgments or ranking models.
- Have low sample and computational complexity.

Given this, the evaluation metrics will be discussed and assessed concerning the above criteria.

3.1.3.1 Quantitative Measures

Quantitative measures can be model agnostic when they do not require a density estimation from the model and the generator network is used as a black box to sample images. On the other hand, some can demand a probability distribution from samples.

Coverage The coverage metric (COV) was firstly proposed in [135], using the probability mass of the real data “covered” by the model distribution, P_{model} , as a metric. The computation is as follows:

$$C = P_{data}(dP_{model} > t) \quad (3.16)$$

In this Equation, t is such that $P_{model}(dP_{model} > t) = 0.95$. The approximation density of P_{model} is calculated through a kernel density estimation method and the researchers claim that this metric is more interpretable than the likelihood.

In [1], the coverage is described for two sets of samples: A and B . For each sample in A , the closest neighbor in B is found and the coverage is measured as the fraction of samples from on set B that match the samples on the set A . Given so, a high coverage score indicated that most of B are roughly represented within A .

Minimum Matching Distance Since Coverage does not indicate how accurately the covered samples are represented in a set A , the Minimum Matching Distance (MMD) was created as a way to measure the fidelity of A with respect to B [1]. Given that, every sample of B is matched to the one in A with the minimum distance (appropriate to the data is being generated) and report the average of distances in the matching.

Inception Score The Inception Score (IS) is one of the most common measures for evaluating GAN models in image applications. IS was proposed in [115] and it consists of using a pre-trained Inception Net neural network [33] on the ImageNet dataset [26] for capturing desirable properties of the generated samples, such as *highly classifiable* and *diverse* when it comes to class labels.

It consists on the measurement of the average $\mathcal{D}_{KL}$ between the conditional label distribution $p(y|x)$ of samples (expected to have low entropy for easily classifiable samples, better quality), and the marginal distribution $p(y)$ obtained from all the samples (expected to have high entropy if all

classes are represented equally in the set of samples, high diversity). The IS favors low entropy of $p(y|x)$ but high entropy of $p(y)$:

$$\exp(\mathbb{E}_x[KL(p(y|x)||p(y))]) = \exp(H(y) - \mathbb{E}_x[H(p(y|x))]) \quad (3.17)$$

where $p(y|x)$ is the conditional label distribution for sample x estimated using the pretrained Inception model, $p(y)$ is the marginal distribution $p(y) \approx \frac{1}{N} \sum_{n=1}^N p(y|x_n = G(z_n))$ and $H(x)$ represents the entropy of variable x . Besides the properties stated, the IS has limitations:

- It favors models that store all training samples, being unable to detect overfitting. This is even aggravated by the fact that it does not make use of a holdout validation set.
- It is agnostic to mode collapse, since it fails to detect if the model is trapped into a bad mode, although results in [151] show the contrary.
- Since it is based on the Inception model trained in ImageNet with many objects, it may favor models that generate good objects rather than realistic samples. Also, to evaluate images from a different domain than the ones in ImageNet, IS will be low even for high-quality fake images.
- It is affected by image resolution since it is an asymmetric measure.
- It does not evaluate intra-class variability, so high IS may happen even if the model only generates one type of sample per class.

For enabling the characterization of diversity within samples in a particular category, the **Modified Inception Score** (m-IS) was proposed in [42], with the usage of a cross-entropy style score : $p(y|x_i)\log p(y|x_j)$, where x_j are samples from the same class as x_i based on the inception model's output, resulting in the following m-IS:

$$\exp(\mathbb{E}_{x_i}[\mathbb{E}_{x_j}[KL(P(y|x_i)||P(y|x_j))]]) \quad (3.18)$$

Given this, the m-IS is calculated on a per-class basis and posteriorly averaged over all classes, so it favors both intra-class sample diversity and sample quality.

Fréchet Inception Distance The Fréchet Inception Distance (FID) was introduced in [49] and it measures the Fréchet distance (also known as the Wassertein-2 distance) between two multivariate Gaussians. In Generative Models, these correspond to the real (r) and generated (g) data, taking use of the mean and covariance of these distributions. To obtain these distributions, a set of real or generated samples are embedded into a feature space, which is given by a specific layer of Inception Net (or other CNN). FID can then be calculated as it follows:

$$FID(r, g) = \|\mu_r - \mu_g\|_2^2 + Tr\left(\sum_r + \sum_g - 2(\sum_r \sum_g)^{\frac{1}{2}}\right) \quad (3.19)$$

where (μ_r, Σ_r) and (μ_g, Σ_g) are the mean and covariance of the real data and model distributions, respectively. Lower values of FID mean smaller distances between synthetic and real data distributions. FID has 0 as a lower bound and no upper bound, it has shown more robustness to noise and blurring than the IS, and it is more sensitive to mode collapse and enables the evaluation of intra-class variability. Despite this, this measure also depends on the used CNN and its capabilities for detecting relevant features. Also, when using FID, there is an assumption that Gaussian distributions correctly model the features maps extracted from the networks, besides FID only considering the first two order moments of the distributions, which may not correspond to accurate approximations.

Image Quality Measures Some researchers defend the use of measures from the image quality assessment for training and evaluating generative modelling, even though they only compare the quality of the results of the generated samples with others.

Firstly introduced for assessing image quality in [140], the **Peak Signal-to-Noise Ratio** (PSNR) between two images f and g , is calculated through Equation 3.20:

$$PSNR(f, g) = 10 \log_{10} \left(\frac{255^2}{MSE(f, g)} \right) \quad (3.20)$$

where the MSE refers to the mean squared error of two images of size $M \times N$, which is calculated through:

$$MSE(f, g) = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (f_{i,j} - g_{i,j})^2 \quad (3.21)$$

As MSE reaches zero, similarity between the two images increases, so PSNR tends to infinity. On the other hand, as MSE increases, PSNR decreases, imposing great differences amongst the images. Although simple, this metric does not assess structural similarity between images. To overcome this, the **Structural Similarity Index** (SSIM), proposed in [140], measures the similarities between images accordingly to three factors: luminance distortion (l), contrast distortion (c) and loss of structural correlation (s). In texturized or bright regions of images, image distortions are more unnoticeable, motivating the usage of contrast and luminance distortion, respectively. The structural correlation is used for evaluating object structure within spatially close pixels, which are considered to have strong interdependence. SSIM between two images f and g is calculated as demonstrated in the following Equation:

$$SSIM(f, g) = l(f, g)c(f, g)s(f, g) \quad (3.22)$$

where $l(f, g) = \frac{2\mu_f\mu_g+C_1}{\mu_f^2+\mu_g^2+C_1}$, $c(f, g) = \frac{2\sigma_f\sigma_g+C_2}{\sigma_f^2+\sigma_g^2+C_2}$, and $s(f, g) = \frac{\sigma_{fg}+C_3}{\sigma_f\sigma_g+C_3}$, where μ_f and μ_g are average intensity values, σ_f and σ_g are standard deviation values, σ_{fg} is the covariance, and C_1 , C_2 , and C_3 are constants. SSIM is bounded between -1 and 1, where 0 means no correlation between the images, 1 means equal images and -1 indicates that structure is inverted.

Semantic Segmentation Metrics Also in the field of imaging, some researchers defend the usage of semantic segmentation networks to quantitatively assess image realism. To do so, a pre-trained segmentation network is trained on real images and segmentations metrics such as DICE scores are extracted for evaluating the accuracy of the segmentation. If the generated images are realistic, it is expected that the network will be able to segment them properly. This technique has been used for image-to-image translation applications, presented in [139] or [52], for instance.

3.1.3.2 Qualitative measures

Performing visual examination of the generated samples through human analysis is one of the most common and intuitive approaches for evaluating the results of the Generative Models, such as in [115]. While a human inspection of the samples can help inspect and tune models, it has several limitations since it is expensive and cumbersome, biased towards the models that overfit, not being able to address indicators of good density model in a log-likelihood sense. Also, it is difficult to reproduce, inconsiderate of the capacity of the models, not reflecting their abilities, and unfeasible in the terms that the study needs to have a large number of human inspectors due to their high variance.

Nearest Neighbors For detecting overfitting some generated samples can be shown next to their nearest neighbors in the training set. However, while using this technique, the following factors must be remembered:

- The nearest neighbors are normally calculated through distance measures based on the Euclidean distance, which is very sensitive to minor perceptual perturbations. Given so, a large Euclidean distance may not reflect the perceptual similarities between samples [133].
- Even a model that overfits may not generate samples that are similar to the ones in the training set, so no distance metric will find a close match in the latter. Since the number of analyzed samples is much smaller than the possible samples, it is nearly impossible to detect this behavior by looking at samples from the model [133].

Rating and Preference Judgement In Rating and Preference Judgment experiments, researchers asked subjects to rate fidelity of the generated images. For instance, in [124], the authors explored whether observers choose reconstructions produced through networks perceptually-optimized rather than pixel-wise-loss-optimized. To do so, the observers received a triplet of images: an original (for reference) and two reconstructions, with SSIM- or MSE-optimized, and they were asked which of the reconstructions they preferred.

Visual Turing Test The Visual Turing Test (VTT), presented in [37] as a way for standardizing perceptual studies, presents binary questions to evaluate the ability of the systems to recognize objects and identify attributes and relationships in images. VTT has been used in medical image generation, where it was presented to radiologists to judge the quality of the images and models [22].

Inspect the Internals of the Networks Another approach for evaluating Generative Models is inspecting their internal dynamics, present in the latent space. Doing this may help to understand how and what the models learn. There are many possibilities for studying the internals of generative models, but here are some of interest to the scenarios of this work:

- **Disentangled representations.** The “disentanglement” concerns the alignment of “semantic” visual concepts to axes in the latent space. Some tests can help us evaluate the existence of semantically meaningful directions in the latent space, that is, directions on which variations lead to noticeable and predictable changes (for instance, the size of the breast). The quality of the internal representations can also be accessed through a measure of disentanglement, like the one presented in [50], that verifies whether the latent space captures true factors of variation in a simulated dataset where parameters are known by construction. Arithmetic operations performed on the latent space vectors also shows interesting results for assessing how different variables can relate in the latent representation [145].
- **Space continuity** is related to the above topic, and the aim is to explore the capability of the model for extracting details. Performing linear interpolation between two latent vectors that decode into meaningful samples is a technique used to address the latent space continuity. Considering z_1 and z_2 as the latent vectors, the result of the linear interpolation, z_3 is calculated through $z_3 = z_1 + weight \times (z_2 - z_1)$, where *weight* is a parameter bounded from 0 to 1. As the *weight* increases, the z_2 vector has more importance, since for $weight = 0$ the $z_3 = z_1$ and for $weight = 1$, $z_3 = z_2$. Selecting a variety of *weight* values shows the transformation from z_1 to z_2 and if the decoded representations of z_3 are meaningful space continuity can be assumed. Another possibility is performing spherical linear interpolation which prevents diverging from the prior distribution of the model and produces sharper samples [142].

3.2 Applications to Radiological data

Because generative models are challenging to train and tune, their application to medical data is still relatively few explored. Relevant experiments and findings are depicted in Table 3.1. Most of the works found use brain MRI images for synthesis [45], [89], [121], [137], super-resolution [116], or mode translation [93], [15]. From the analysis of Table 3.1, one conclusion is that very few works respect the depth continuity of radiological data, considering single slices as different inputs and disregarding the volumetric characteristics of the data.

Table 3.1: Relevant generative modelling applications to radiological data.

Article	Year	Purpose	Models	Evaluation
[88]	2016	Brain and pelvis MRI to CT conversion	GAN model: 3D FCN generator	MAE, PSNR
[128]	2017	3D brain MRI to X-ray projections	Cascaded Refinement Network	MAE, SSIM, PSNR
[144]	2017	2D brain MRI to MRA	Cycle GAN	MAE, PSNR
[45]	2018	2D brain MRI (T1/T2w, T1c, Flair) synthesis	DCGAN, WGAN	VTT, Accuracy
[93]	2018	Brain MRI (T1/T2w) to MRA	Steerable GAN	PSNR, DICE
[34, 35]	2018	Liver CT synthesis from annotation, liver lesion classification	DCGAN	Accuracy, Specificity, Sensitivity
[15]	2018	2D brain MRI mode conversion (T1/T2w, T1c, Flair)	Multi-input, multi-output CNN	SSIM, PSNR, DICE
[116]	2018	3D brain MRI (T1w) super-resolution	Adaptation of SRGAN for 3D	SSIM, PSNR
[89]	2018	Brain and pelvis MRI synthesis	<i>pix2pix</i>	MAE, PSNR, DICE
[5]	2018	2D PET-CE conversion, MRI motion corrections, PET denoising	MedGAN	SSIM, PSNR, MSE, VIF, UQI, LPIPS
[121]	2018	3D brain MRI (T1, T1c) synthesis from annotations	<i>pix2pix</i>	DICE
[137]	2018	2D brain MRI (T1/T2w, T1c, Flair) synthesis	PGAN, CycleGAN	SSIM, PSNR
[55]	2019	2D brain MRI to CT conversion	U-net, <i>pix2pix</i>	MAE, MSE, PSNR
[44]	2019	Accelerated MRI reconstruction	Variational Network	MSE, SSIM, NRMSE
[59]	2019	3D brain MRI generation	Autoencoding GAN	MMD, MS-SSIM
[29]	2020	2D reconstruction of multicontrast MR images	X-Net, Y-Net	normalized-MSE, SSIM
[132]	2021	2D Breast MRI generation from annotations	GANs: <i>pix2pix_{HD}</i> , CycleGAN	FID, SSIM
[84]	-	3D brain MRI generation	MRNGAN	MSE, PRNR, SSIM, IS

FCN - Fully Connected Neural Network; LPIPS - Learned Perceptual Image Patch Similarity; MAE - Mean Absolute Error; MRA - Magnetic Resonance Angiography; T1w/T2w - T1/T2 weighted; UQI - Universal Quality Index; VIF - Visual Information Fidelity.

In [84] an end-to-end framework to reconstruct 3D brain MRI through single 2D slices is presented. They called it the MRNGAN, which consists of series of networks, schematized in Figure 3.10. First, for localizing each slice in space, a CNN is used to classify to what slice belongs a certain input image. Then, a series of DCGANs are trained with images from brains without lesions to reproduce a slice of MRI in a specific location (given by the previous network), capturing the basic components shared by healthy brains. A *pix2pix* model is applied for flow generation, for understanding how brain MRI slices change temporarily, being trained to generate the next slice of MRI given the previous. Ultimately, 3D reconstruction is performed with a one-to-many recurrent neural network (RNN). By concatenating the results of the pre-trained networks (brain generalization and flow generation) given an input slice, the resulting three-dimensional array is given to the RNN. This allows the reconstruction of full MRI volumes of the brain by outputting each next slice at each time step of the recurrent network.

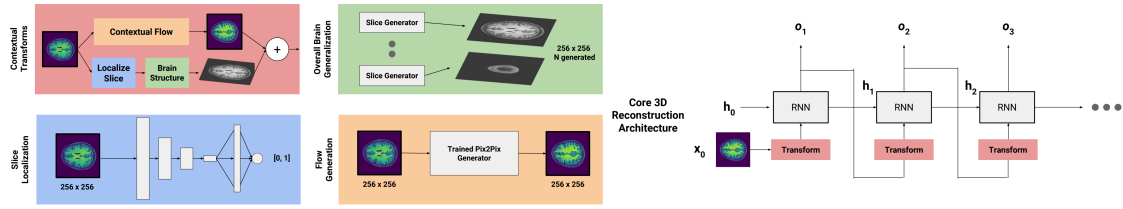


Figure 3.10: MRNGAN end-to-end framework. Retrieved from [84].

Another interesting methodology combining VAE and GAN for the generation of 3D MRI scans of the brain is the Auto-encoding GAN, presented in [59]. The architecture is detailed in Figure 3.11. The trained model can generate 3D brain MRI data from random vectors by learning the data distribution. The authors claim that Auto-encoding GAN solves both image blurriness and mode collapse problems by leveraging α -GAN that combines the advantages of VAE and GAN with an additional code discriminator network, which regularizes the latent space vectors.

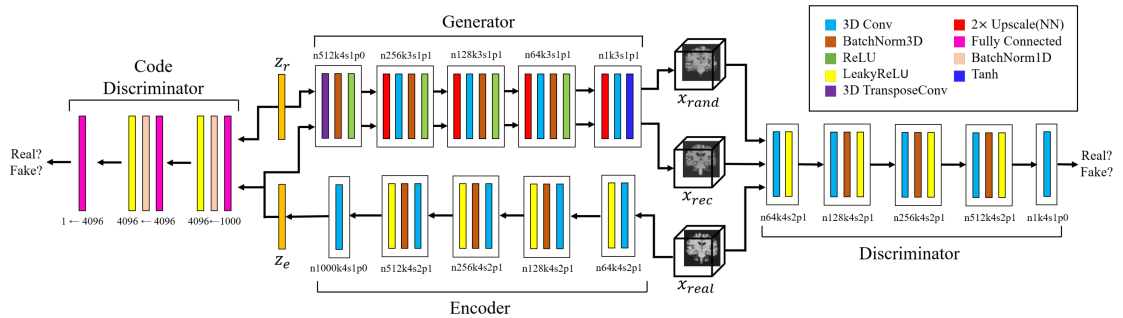


Figure 3.11: Auto-encoding GAN architecture. The number of the channels, the kernel size, the stride size, and the padding size are represented as n , k , s , and p , respectively. The generator output from random vectors z_r is represented as x_{rand} and x_{rec} is the output from encoded vectors z_e . Retrieved from [59].

For thoracic MRI, relevant work in [28] was done for generating individual slices, using *pix2pix* variations with different generator architectures (U-Net, CycleGAN, *pix2pix_{HD}*). The research conducted by the author demonstrated that these architectures can model the thoracic MRI for generating synthetic images, although some generated samples have blurred areas where important structures inside the breast are. Besides, it was also shown that evaluation metrics may provide contradicting information to the visual results, even though the latter needs a lot of expertise to be analyzed accurately. In [132], breast MRI images are synthesized from corresponding annotations and these are used to evaluate the impact on a semantic segmentation task. The authors explored variations of image-to-image translation using conditional GANs, with the best models being *pix2pix_{HD}* and Residual CycleGAN. The synthetic dataset maintained a stable segmentation score even when completely replacing the original data. No relevant research was found on generating synthetic data of 3D volumetric thoracic or breast MRI.

3.3 Summary

In this Chapter, different generative modelling strategies were addressed, focusing on frameworks of interest to this work: AE based models (VAE or CVAE) and GANs. Although they both aim to generate realistic data, they work with very different approaches. The goal of a VAE model is to compress the input into a regularized latent space, from which a sample from that space can be reconstructed into meaningful synthesized examples. On the other hand, GANs consist of two networks competing, the generator and the discriminator, where the first one tries to synthesize more and more realistic samples throughout training. The latter learns to distinguish between fake and real samples with increasing accuracy. Both AE and GAN based models are adaptive to a conditional setting. Even though VAE models are easier to train than GANs, they depend on a reconstruction loss between the original data and the output of the decoder, which is difficult to construct, so usually, they generate samples with lower quality.

A review of the main evaluation criteria for generative models was also presented, describing both quantitative and qualitative measures.

Generative modelling is a relatively new field in Deep Learning, but already some researchers have applied them to medical data, especially GAN based models, obtaining reasonable results. However, these are mainly applied to brain MRI imaging for synthesis or mode conversions. Furthermore, very few of them respect the correspondence between slices of MRI since it consists of a volume of slices per patient. Also, no work was found for translating MRI or other radiological modality to surface information.

Chapter 4

3D Data Representations

Deep Learning techniques usually require large amounts of data for successfully modeling data in the training phase. That is the most prominent fact as to why applying DL in 3D domain is not as effective as in 2D. 3D data is more scarce due to the higher difficulty in obtaining the data, as it is a more expensive and complicated process. In the past years, this has not been the case with more and more available public 3D models and affordable 3D data acquisition devices, increasing exponentially the amount of available data [3]. This combined with the fact that 3D data offers rich information about the full geometry of real-life objects, results in active research by computer vision communities for exploring this type of data with state-of-the-art DL algorithms [71]. This Chapter gives a presentation on the main methods for acquiring 3D data and the main representations for 3D data, focusing on the data of interest to this work, point clouds. Finally, a review of the leading techniques for modelling 3D data using DL is presented.

4.1 Acquisition methods of 3D data

In the medical field, using 3D representations of the surface is preferable to MRI or CT scans, since they do not overload the patients with harmful radiation [105]. However, these do not provide insight into the interior parts of the body. There are three phases for constructing 3D models: data acquisition, processing, and analysis [105].

Data acquisition and measurement techniques can be divided into two major groups: contact and non-contact [104]. In the case of contact methods, these use coordinate measuring machines and other measuring tools, such as rulers for 3D reconstruction. Non-contact methods are the most commonly used systems for the generation of 3D models and are based on light waves. Depending on the time for processing the coordinates of the objects, non-contact methods are divided into active and passive sensors. Active sensors provide the 3D coordinates used in the mesh's generation, immediately, through triangulation or time delay, as in lidar sensors, while passive sensors take time to process and get the coordinates.

After the acquisition of the data, there are mainly four methods used for object and scene modelling, using images (mostly used for basic shapes) and range (used in complex and detailed shapes) methods [104]:

- **Image-based rendering** uses input images for creating novel views of 3D environments; it either requires the knowledge of the exact position of the cameras or performing automatic stereo matching;
- **Image-based modeling** is more commonly used for geometric shapes and uses 2D image measurements for recovering the 3D object information; passive methods, for instance, the acquisition of data from multiple views is used; projective geometry or perspective camera models are used;
- **Range-based modeling** uses active sensors for capturing 3D geometric information of the surface automatically; artificial lights or pattern projection is used in these methods and the results obtained are very accurate;
- **Combining image and range-based modeling** with photogrammetry and laser scanning are used to obtain even better results, especially in large and complex architectural objects.

The number of points required to represent an object is directly proportional to their curvature. A cube can be represented by the number of vertices in it, but a complex shape as a breast needs many more points for an accurate representation [105]. Some of the methods used for obtaining 3D surface scans of the breasts have already been presented in Section 2.3.2. The data acquired is also structured so it forms a polygonal surface of the mesh. Creating more realistic 3D models involve the addition of texture of the surface, from image information.

4.2 Representation of 3D data

Different scanning devices capture raw 3D data that can differ in structure and properties. These different representations have impact on the design of the DL architectures. In [3], the possible representations for 3D data are separate in Euclidean and non-Euclidean (see Figure 4.1).

In Euclidean representations, the underlying grid-structure of the data is preserved using global parameterization and a common system of coordinates. These characteristics enable a simple transformation of the DL techniques since the convolutional operations applied are almost the same as in 2D data. Three-dimensional euclidean data representations are descriptors, projections, RGB-D data, volumetric data, and multi-view data.

For the non-Euclidean representations, also known as geometric data [11], there is no global parameterization or common system of coordinates, so they do not appear in the same format as the previous ones. Furthermore, they lack a vector space structure, implementing DL techniques not as straightforward as for Euclidean data. The main types of non-Euclidean representations for 3D data are 3D meshes, graphs, and point clouds, which are the representation of interest in this work.

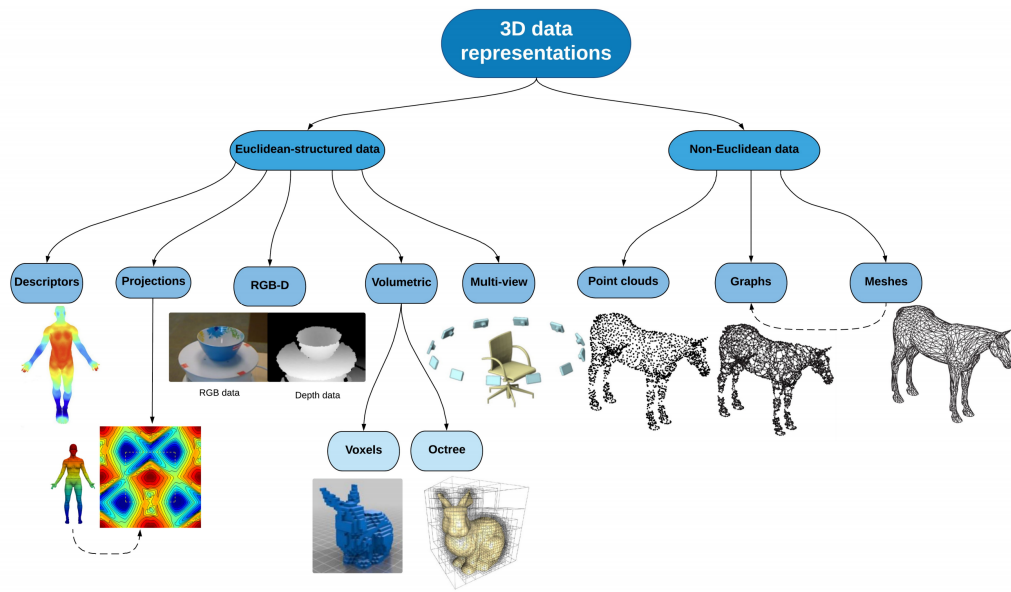


Figure 4.1: Representations of 3D data. Retrieved from [3].

4.2.1 Point clouds

Point clouds (PCLs), as the one present in Figure 4.2, consist of N unstructured 3D points that approximate the geometry of an object, normally represented by x , y , and z , but additional features such as color, material or even texture can also be added. Any combination of the points will result in the same object, making this representation permutation invariant.

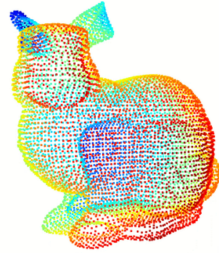
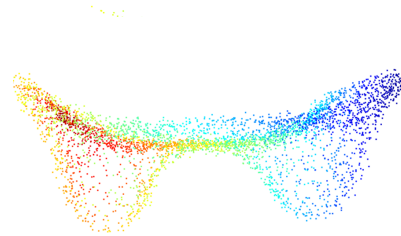
Figure 4.2: Stanford Bunny point cloud¹.

Figure 4.3: Breast point cloud from top view.

The permutation invariance makes a PCL a non-Euclidean representation of 3D data. Nevertheless, they can also be considered a collection of small Euclidean subsets. This happens since Euclidean structures can be assumed in a small scope. Given that, each subset has a global parameterization and a common system of coordinates. Also, PCLs are invariant to transformations, such as translation or rotation. Taking this into consideration, the definition of a point cloud varies according to the consideration or not of the local structure of the object.

The data can nowadays be obtained with a certain ease, with the usage of technologies such as Microsoft Kinect and structured light scanners. However, the lack of structure and absence of

¹<http://graphics.stanford.edu/data/3Dscanrep/>

connectivity information makes the data hard to process [3]. Concerning breast data, point clouds can be obtained through manual annotation or automatic segmentation from MRI scans, resulting in PCLs such as the one present in Figure 4.3, which has been down-sampled.

4.3 Deep Learning and 3D data

The multiple representations for 3D data presented above lead to different approaches for modeling and characterizing the shapes with DL techniques, also according to the target application. The most relevant methods found are present in Table 4.1.

Table 4.1: Main strategies used for modelling 3D data with deep learning techniques.

Article	Year	Input	Key aspects of the method
[146]	2015	Voxelized shapes	3D shapes represented as probability distribution of binary variables on a 3D voxel grid, using a Convolutional Deep Belief Network.
[138]	2015	Point clouds	Three-dimensional object detection through a sliding window approach, dealing with the sparsity of the data.
[129]	2015	Multi-view	Recognition of 3D shapes using a CNN to recognize their rendered views on 2D images independently.
[31]	2015	Feature vectors	Shape descriptors to craft features from the data. These features are used for classification in fully connected networks.
[41]	2015	Feature vectors	Extraction of features for each triangle face in mesh and using these features for classification with a CNN.
[100]	2016	Voxelized shapes	MO-CNN: takes multi-orientations of the 3D input, improving results over other volumetric CNNs.
[68]	2016	Distance fields	Representation of PCLs or meshes into volumetric fields, which pass through probing filters to extract features from them.
[117]	2016	Multi-view	Shape retrieval of 3D shape models rendered with different virtual cameras and further application of CNN for class predictions.
[78]	2017	Voxelized shapes	Voxelized shapes as input in a 3D CNN (VoxNet).
[99]	2017	Point clouds	PointNet consumes PCLs directly. Permutation and transformation invariant.
[101]	2017	Point clouds	PointNet++ uses the PointNet recursively on a nested partitioning of the input point set for capturization of local structures.
[1]	2018	Point clouds	AE, GAN, and Gaussian Mixture Model for reconstruction and generation.
[6]	2019	Point clouds	Transformation of a point cloud into rendered PCLs, using the original PCL and camera viewpoint angles for adversarial training with a GAN (pc2pix) for the creation of synthetic data.
[53]	2019	Point clouds	Point clouds coordinates and their second-order polynomial functions to extract features and learn shape moments for object classification.
[147]	2020	Point clouds	Energy-based generative model for unordered sets. Matching the observed examples in terms of statistical properties defined in the energy function (based on PointNet) enables generation of samples.

A few years back, it was common to transform the raw 3D onto other representations enabling an easier transition from 2D convolutional networks onto 3D applications. For instance, transforming the data into voxelized shapes and applying 3D convolutional networks lead to some accurate results as presented in [146], [100], and [78]. But this type of transformation alters the nature of the non-Euclidean data, also leading to highly voluminous data that is not necessary. The data sparsity stiffens the resolution of the volumetric representation which endorses high computational costs when performing 3D convolutions.

Methods such as [31] or [41] extract features from 3D data and apply them into a fully connected network for classification, but have no other applications than that. Another approach for dealing with 3D data is to represent it with multi-view 2D images of the object, as proposed in [129] and [117]. These show good performance in classification and retrieval but are not capable of scene segmentation or per point classification.

The major upgrade in DL for dealing with 3D point clouds came with PointNet [99], proposed in 2017, which is capable of consuming directly as input PCLs, with no other transformation needed.

4.3.1 PointNet

The framework of PointNet [99] consists on neural networks designed for scattered, unordered point data that can perform various tasks: object classification, object part segmentation, and semantic scene parsing.

The main challenges in creating the architecture thrive from the characteristics of the data earlier explained. The model needs to be invariant to the $N!$ permutations possible for the N points of the PCL, and geometric transformations, such as rotation, should not alter the classification result.

For overcoming the permutation invariance dilemma, a symmetric function needs to be applied: $f(x_1, x_2, \dots, x_N) \equiv f(x_{\pi_1}, x_{\pi_2}, \dots, x_{\pi_N})$, where $x_i \in \mathbb{R}^D$, where i corresponds to a point of a total of N points and D corresponds to the dimensions, usually 3. The trick was to construct a family of symmetric functions by using neural networks, and for that, the realization that $f(x_1, x_2, \dots, x_N) = \gamma \circ g(h(x_1), h(x_2), \dots, h(x_N))$ is symmetric if g is symmetric. So each point goes through an MLP (multi-layer perceptron) designated as h , and all the resulting MLPs enter a simple symmetric function g (max pooling demonstrated to be the best), and then the result is passed through another MLP γ for obtaining classification scores. The operations described correspond to the PointNet vanilla², schematized in Figure 4.4.

For addressing the invariance to geometric transformations impasse, the researchers propose a data-dependent transformation for automatic alignment of the data. This is accomplished with the usage of a T-Net (Transformer Network). A T-Net is applied in the input of the network, for alignment of the points, making the overall architecture rotation invariant. After that, each point is

²<http://stanford.edu/~rqi/pointnet/>

embedded in an MLP in a 64-dimensional space, where another T-Net is applied, in this case, for alignment in the embedding space.

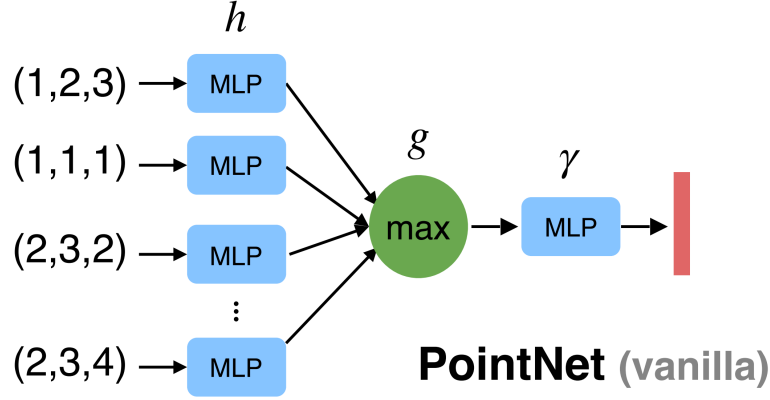


Figure 4.4: PointNet vanilla³.

Only after this the operations present in the PointNet vanilla are applied, with the embedding of each point onto another MLP of dimensional size 1024, the application of the max pooling operation resulting in a global feature vector that, for classification, is converted to the desired number of classes.

For segmentation tasks, the local embedding ($N \times 64$) is concatenated with the global feature vector (1024) for entering MLPs for point feature extraction and per point classification scores. The overall architecture of PointNet is presented in Figure 4.5.

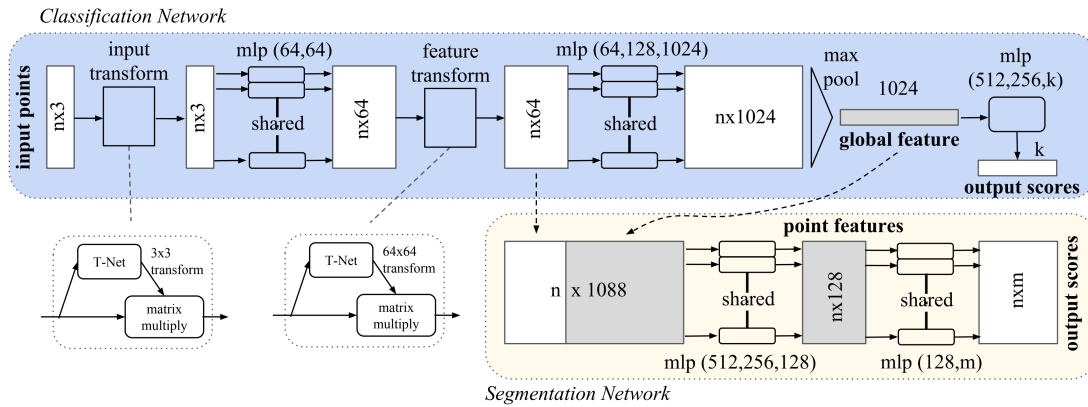


Figure 4.5: PointNet Architecture. Retrieved from [99].

PointNet shows efficiency and robustness in a variety of tasks and has been used as a model for other experiments, such as generative modeling. These models are implemented using large datasets, composed mainly of rigid objects, where deformations occur locally, like the ones available in ShapeNet [14] or ModelNet [146] datasets.

³https://web.stanford.edu/~rqi/pointnet/docs/cvpr17_pointnet_slides.pdf

The authors of PointNet also explore the effects of the bottleneck size (global feature vector) and the number of points per point cloud on the performance of the classification model. The performance increases as the number of points increases, although it saturates at around 1000 points. The size of the global feature vector is also important since an increase from 64 to 1024 results in a 2-4% performance gain. Given this, it is comprehensible that the model needs a certain size of point features to cover a 3D space for discriminating the shapes. Nevertheless, even with a lower size of the bottleneck or number of points, PointNet can still achieve a decent performance in classifying objects.

Despite all of these, PointNet does not capture local structures induced by the metric space the points live in. To overcome this, the PointNet++ [101] was created. This network applies PointNet recursively on a nested partitioning of the input point set. While in the original setting the single max-pooling operation is used to aggregate the whole point set, PointNet++ builds a hierarchical grouping of points and progressively abstracts larger and larger regions along this hierarchy. The authors of PointNet++ test the architecture in the same experiments performed in PointNet plus on non-rigid shape classification in SHREC15 [97, 69] dataset.

4.3.2 Generative Modelling and Point Clouds

As seen in Chapter 3, generative models will be the main resource for obtaining synthetic data. Despite the advances made in 3D data acquisition and processing, there is still not a lot of research on the topic of generating realistic 3D models through DL techniques.

In [1], the authors use a deep autoencoder network for reconstructing point clouds, enabling generalization. The network was inspired by PointNet, and it can learn compact representations with good reconstruction quality on unseen samples and also good classification with simple machine learning methods, such as Support Vector Machines (SVM). The encoder part of the AE is similar to PointNet vanilla up to the max-pooling operation (included). It receives a point set (2048×3) as input and has 5 1D-convolutional layers with ReLU and batch normalization. The decoder uses fully connected layers, with ReLU on the first two layers, to reconstruct the input. Besides the AE used for learning representations of the 3D PCLs, they also developed other networks for generating the samples:

- **Raw point cloud GAN (r-GAN)**, where the discriminator is like the AE previously described, but without batch normalization and with Leaky ReLUs instead of ReLUs. The output is fed into a sigmoid neuron. For the generator, it receives a Gaussian noise vector and maps it into a 2048×3 output via 5 fully connected ReLU activated layers.
- **Latent-space GAN (l-GAN)**, which does not consume raw point clouds. Instead, the input data has already passed through a pre-trained AE, trained separately for each object class. The l-GAN generator and discriminator can both operate on the bottleneck variables of the AE. After training, the code learned by the generator is converted into a PCL through the AE's decoder. The GAN networks are relatively simple when comparing to the AE, with an MLP generator of a single hidden layer and an MLP discriminator of two hidden layers.

B , with the same number of points, the EMD can be calculated as shown in Equation 4.1, where ϕ corresponds to a bijection.

$$d_{EMD}(A, B) = \min_{\phi: A \rightarrow B} \sum_{a \in A} \|a - \phi(a)\|_2 \quad (4.1)$$

4.4.2 Closest point distances

Considering two point sets, A and B , it is possible to calculate a set closest point distances, that is, for each point a in point cloud A , the closest point b in B is found.

This set of values can be used to compute Hausdorff, Euclidean and Chamfer Distances. The **Hausdorff Distance** corresponds to the maximum of set of closest point distances:

$$d_H(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (4.2)$$

The **Euclidean Distance** is calculated through the mean value of the closest points distances:

$$d_E(A, B) = \mathbb{E} \left(\min_{b \in B} \|a - b\| \right) \quad (4.3)$$

And the **Chamfer Distance (CD)** consists of the sum of the closest point distances [7]:

$$d_{Ch}(A, B) = \sum_{a \in A} \min_{b \in B} \|a - b\| \quad (4.4)$$

Metrics based on the closest point distances are asymmetric since the distances from point cloud A to B , and vice-versa, need to be calculated to properly compare both point clouds. The two corresponding measures can be summed or averaged for a global measure of each distance.

4.4.3 Jensen-Shannon Divergence (JSD)

The JSD is a measure of the distance between marginal distributions of points in different sets. It has been used for evaluating generative models with point clouds, such as in [147] or [1], along with metrics like Coverage and Minimum Matching Distance, already presented in Chapter 3. Considering a reference point set, $p_r(x)$, and a set of generated samples, $p_g(x)$, the JSD is calculated through the following:

$$JSD(p_g, p_r) = \frac{1}{2} [KL(p_r || M) + KL(p_g || M)] \quad (4.5)$$

where $M = \frac{1}{2}(p_g + p_r)$. The empirical distributions of p_g and p_r can be obtained by voxelizing the point cloud space into a $28 \times 28 \times 28$ voxel grid and counting the number of points within each voxel across all point clouds in the set.

This metric only evaluates the marginal distribution of points rather than the distribution of point clouds, consequently, a generative model that fails to capture the diversity of point clouds within a category can still generate point clouds that can obtain a satisfactory JSD score.

4.5 Summary

Over the past few years, modeling 3D data with computer vision has been an active task for researchers. 3D models can give us information about the objects that images can not, such as surface details and depth. Besides, obtaining 3D data has become easier and cheaper with the increasing availability of low-cost 3D scans. With that, DL research on different 3D representations has emerged, especially in non-Euclidean data, such as PCLs or meshes.

Breast point clouds can detail meaningful features of the surface in that region and they can be used in several medical applications, such as surgery planning tools or cosmetic and deformations evaluation.

The major techniques before PointNet involved transforming the data into other representations. This leads to dismissing some of the important information this type of data contains. PointNet is a revolution for DL with 3D data, and everyday researchers use the findings of PointNet for developing novel strategies and architectures. Possible and interesting use of PointNet is having the network function as an encoder of information for point clouds in generative models. Some of the primary measures to capture differences between point clouds were also presented, since they will be used measuring the quality of the reconstructions given by the implemented methods.

With these tools, strategies that incorporate DL with breast surface data can be explored. These strategies should represent these surfaces and, ultimately, be able to generate synthetic breast point clouds.

Chapter 5

Synthesis of Breast MRI

Most methods for synthesizing radiological data in the literature cannot generate whole volumes of data, applying generative models in slices of data, as seen in Chapter 3. After a search of the literature, no other significant research was found on synthesizing breast MRI volumes, as works typically only consider single slices. The goal of this study is to learn how to model volumes of breast MRI with DL strategies, enabling their generation, with the available data of MRI and corresponding annotations of structures. For these purposes, in this Chapter, a variety of DL models based on state-of-the-art techniques are explored. The different models are evaluated on the capability for synthesizing realistic MRI volumes, indistinguishable from the real ones.

5.1 Data Preparation

Understanding the different structures present in breast MRI is already an arduous task for human perception, so, developing DL strategies that can learn the data distribution while capturing relevant features from the data may even be more challenging. This task is even harder when the goal is to model 3D volumes of data and not just MRI slices. In this study, a variety of models is experimented, mainly divided into volume-based methods [84, 59, 116, 121] and patch-based methods [150, 60]. In volume methods, the inputs of the networks are volumes of breast MRI, either entirely (with all slices from a scan) or a part of the MRI (with the desired number of slices). For patch methods, the volumes of breast MRI are divided into smaller patches. If the small patches are created with overlap, this increases exponentially the number of instances for training, since one scan can be divided into thousands of patches. Additionally, since patches are smaller than the entire volume, these strategies are computationally more efficient. Moreover, it may be an easier task for architecture to learn relevant features from the data rather than with the entire MRI.

The breast MRI data was provided by the BCCT.plan project¹ with T1-weighted thoracic MRI exams from breast cancer patients, obtained with a Philips Ingenia 3.0T MRI scanner. Each MRI has 60 slices of gray-scale axial images, with a 3 mm thickness. The images resolution varies around 700×700 pixels (0.3-0.5 mm/pixel). The importation of the data and transformations were

¹BCCT.plan - 3D tool for planning breast cancer conservative treatment - NORTE-01-0247-FEDER-017688

done with MATLAB®. According to the demands of each method, different subsets created from the original dataset were used. This process is schematized in Figure 5.1, referring to the different operations (cropping, rescaling, zero-padding, and patch division) used for each subset. Figure 5.1 also presents the number of scans available in each subset, as well as the train and validation splits.

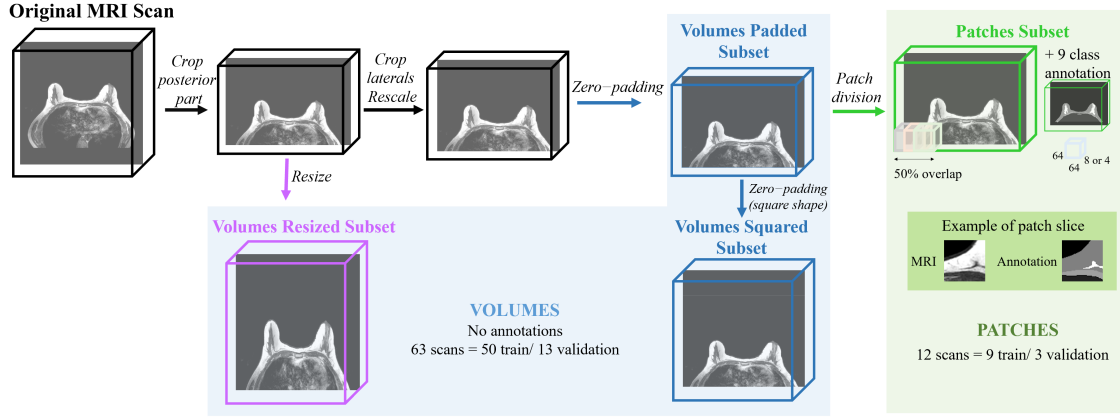


Figure 5.1: Schematization of the creation of the different Breast MRI subsets from the original dataset. Volume Subsets (only MRI) with light-blue background and Patches Subset (MRI and 9 classes annotations) with light-green background. Train and validation splits provided for each subset.

For the volume methods, three different subsets were used (Resized, Padded, and Squared), which contained 63 scans and no annotations.

In the **Volumes Resized Subset**, the most posterior part was cropped using morphological operations (example, dilation, and erosion), ellipse fitting, and active contours to estimate the armpit localization and cropping the scan. After that, the scans from this subset were resized according to the desired input size of each experiment.

Resizing the images does not preserve the anatomical characteristics of the structures, so this operation was discarded in the **Volumes Padded Subset**. After cropping the most posterior part, the maximum laterals for each scan were estimated and cropped. Then the scans were rescaled according to the smaller scan, so they all had the same width. After that, zero-padding was performed both in width and height to achieve an equal resolution between scans 511×410 .

Having a square shape is often convenient in DL architectures, so the **Volumes Squared Subset** was created from the Volumes Padded Subset by zero-padding to obtain a 512×512 resolution per slice.

For the patch methods, the **Patches Subset** was created from the Volumes Padded Subset, dividing each volume into smaller patches. This subset used annotations from 9 different structures: background, breast contour, nipple, fibroglandular tissue, clavicle, sternum, *Latissimus Dorsi* muscle, lesion, and thoracic cavity. However, these fibroglandular annotations were only available for 12 scans from the original dataset. The fibroglandular annotation is important because these tissues appear indistinctly from the fatty tissues of the breast. So, without any guidance to control these two types of tissues, they may not be well represented and learned by the models [28].

The thoracic cavity label was added to the data since, without it, the structures inside it and the background would be perceived equally by the models, threatening the learning of the different structures. Taking this into consideration, the rescale operation resulted in scans with 700×523 pixels of resolution, and the patches were created with 50% overlap and size $64 \times 64 \times 8$ or $\times 4$, according to the method.

5.2 Recurrent networks Breast MRI reconstruction

The motivation for exploring recurrent networks to reconstruct breast MRI volumes comes from the work developed in [84]. They present MRNGAN, a framework of 4 architectures used for performing 3D reconstruction of brain MRI volumes, previously explained in Section 3.2. However, the framework has to be adapted to the purposes and data specifications of this work. The adapted scheme of the framework is presented in Figure 5.2. Only two different networks may be necessary: one *pix2pix* model trained to predict the next MRI slice given the current one (Predicting the next MRI slice), and a recurrent model to perform the reconstruction of a volume given one or more slices, concatenating or not the result from the pre-trained *pix2pix* model at each input of the network (Memory-based network for 3D breast MRI reconstruction). The original framework contained two additional networks: one for slice localization, which is not needed in this case since this feature is known beforehand, and another one for generating the next slice of a "normal" brain (without lesions), which is not relevant for breast data since, unlike brain MRI, breasts structures vary too much between patients, so it is impossible to generate a generalized slice. The scheme from Figure 5.2 only represents the first perspective over the strategy based on recurrent networks.

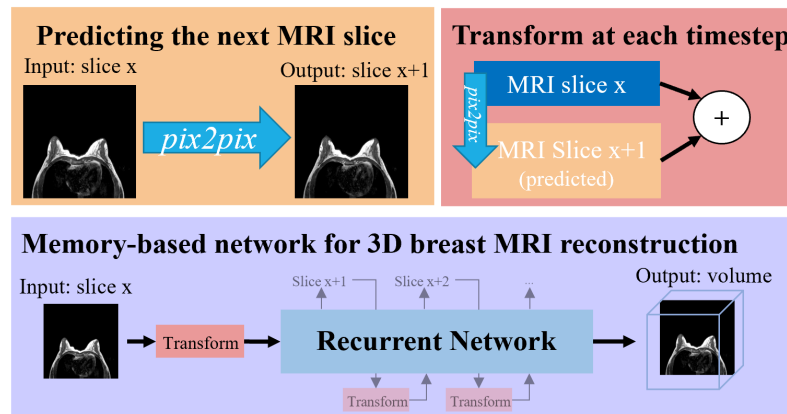


Figure 5.2: Framework for reconstructing breast MRI using recurrent networks. First, a *pix2pix* model is trained for Predicting the next MRI slice. Then, a Memory-based network is used for 3D Breast MRI reconstruction, taking advantage of the previous model and Transforming the input at each timestamp by concatenating the result from the *pix2pix* model into the recurrent network's input.

5.2.1 Predicting the next MRI slice

A *pix2pix* [52] architecture can be used for predicting the next breast MRI slice from the current one. As explained in Chapter 3, the *pix2pix* architecture is used for image-translation tasks given a condition. In this scenario, the conditional image is the previous MRI slice at timestamp x , while the generator tries to generate the next slice, at timestamp $x + 1$. The representation of the architecture and design details are present in Figure 5.3.

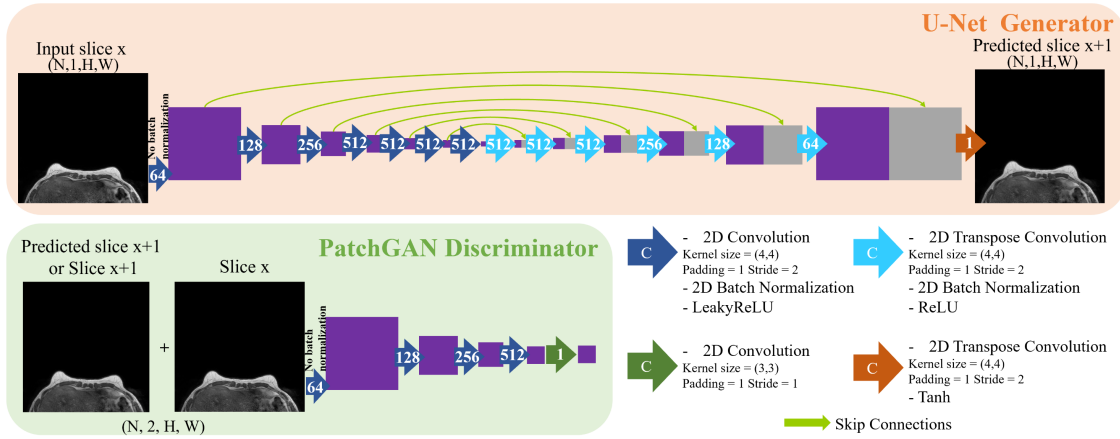


Figure 5.3: Representation of the *pix2pix* network for prediction of the next MRI slice. C is the number of filter channels in each convolutional step.

Implementation Details The dataset used for this experience was the Volumes Resized Subset (Figure 5.1). The adversarial loss used was the L2 with L1 norm for the generator, multiplied by a factor of 10. The latter is used for measuring the similarity of corresponding real and synthetic images. Adam’s optimizer was used for both networks and initialized with a learning rate of $2E-4$. Throughout half the training, the learning rate is kept constant, and on the other half, a linear decay is applied until the learning rate reaches 0. The batch size was optimized between 1 or 32 and the number of epochs between 100 or 200. Additionally, an experiment was done without the decay of the learning rate. These parameters were optimized using an input size of 256 (height and width). Posteriorly, an experiment with input size of 512, 200 epochs, batch size 16, and learning rate decay was done. Another experiment was made where each batch contained only slices from the same position, but that did not improve the results proposed by the previous methods, so it was discarded.

Reconstruction Metrics and Visual Results The evaluation of the *pix2pix* models was done through the inspection of image quality measures: the MSE (Mean Squared Error), the PSNR (Peak-to-Signal Noise Ratio), and the SSIM (Structural Similarity Index). These were calculated between the predicted slices $x + 1$ and the original ones at that timestamp, and results are presented in Table 5.1. Between models trained with images of input size 256, the overall best model was the one trained with 200 epochs with a batch size of 32 and with learning rate decay. The model trained with input size of 512 was trained in this fashion, but the batch sized needed to decrease to

16 due to high computational costs. For all models, the MSE values are relatively low, being all in the range of 10^{-3} . Although PSNR values are not as high as desired, SSIM was close to one (between 0.84-0.92). These values indicate that the model can predict the shape of the next breast MRI slice, even if with a little noise. There is a relevant discrepancy in the values from the training set and the validation set, although this was not so perceptible in the visual results.

Table 5.1: Numerical performance of the *pix2pix* models for generating the next MRI slice: mean and standard deviation are shown.

Batch Size	Epochs	Input Size	Learning Rate	Training set (2953 instances)			Validation set (777 instances)		
				MSE $\times 10^{-3} \downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	MSE $\times 10^{-3}$	PSNR $\uparrow$	SSIM $\uparrow$
32	100	256	decay	4.7 ± 2.9	23.96 ± 2.51	0.88 ± 0.03	6.1 ± 3.3	22.73 ± 2.36	0.85 ± 0.05
1	100	256	decay	5.2 ± 2.6	23.35 ± 2.15	0.88 ± 0.03	6.3 ± 3.3	22.56 ± 2.23	0.86 ± 0.05
1	200	256	decay	4.9 ± 2.4	23.48 ± 2.03	0.88 ± 0.03	7.4 ± 4.4	22.01 ± 2.48	0.86 ± 0.04
32	200	256	decay	3.0 ± 1.7	25.82 ± 2.28	0.90 ± 0.03	6.2 ± 2.9	22.61 ± 2.20	0.85 ± 0.05
32	200	256	constant	3.6 ± 1.6	24.92 ± 2.25	0.89 ± 0.02	6.8 ± 3.7	22.29 ± 2.40	0.84 ± 0.05
16	200	512	decay	1.4 ± 0.6	28.82 ± 1.93	0.92 ± 0.02	6.1 ± 2.8	22.65 ± 2.18	0.85 ± 0.05

A comparison of visual results between the best model with input size 256 and 512 is present in Figure 5.4. With the smaller input size, the input conditional image did not have enough quality in all of the structures for the model be able to produce the next slice with fidelity, as 256 pixels of resolution are not enough to represent these images. Using input size of 512 was then advantageous as the images predicted with this model have more detail. These results are in harmony with the reconstruction metrics presented previously. Nevertheless, both models were able to capture the shape of the next slice of breast MRI.

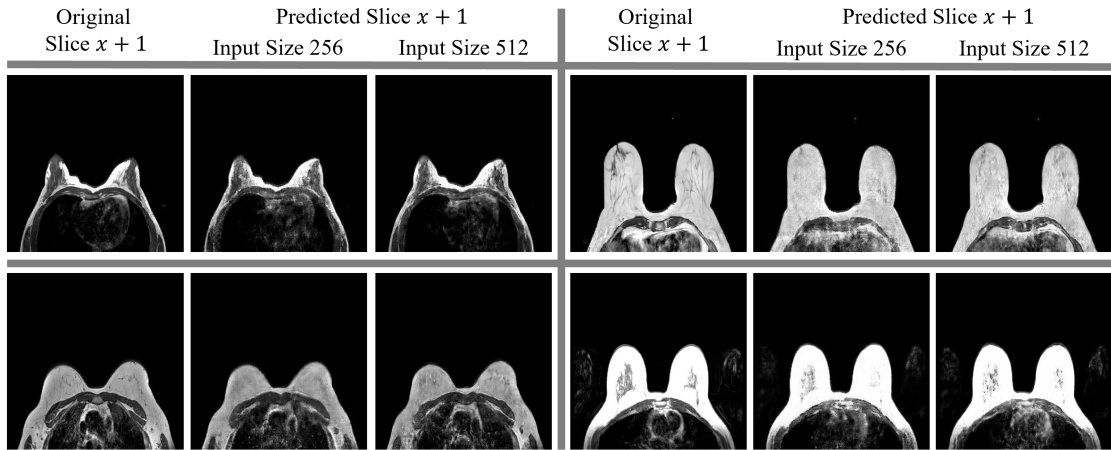
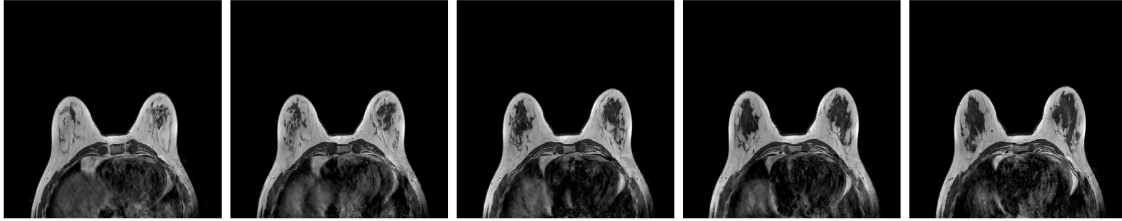


Figure 5.4: Visual performance of the best *pix2pix* models.

As discussed in [84], the *pix2pix* model could be used recursively to reconstruct the 3D volume of MRI, by using the output of the network as the input several times. The results of this operation are present in Figure 5.5, as well as the original sequence of images for comparison.

Ground truth sequence of slices from validation set



Generated images from *pix2pix* model with input size 512

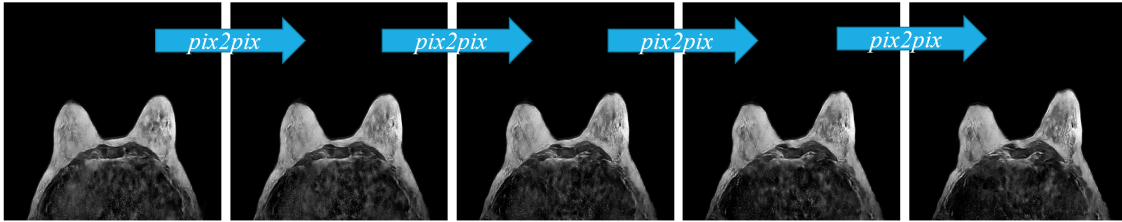


Figure 5.5: Sequence of original and generated slices with *pix2pix* model with input size 512.

It is possible to visualize degradation of image quality and a poor representation of the MRI structures as the input of the model is recursively used from the previous output, similarly to the results in [84]. The generation results are expected to improve with the use of a recurrent network.

5.2.2 Memory-based network for 3D breast MRI reconstruction

After creating a network capable of generating the next slice of breast MRI given the current one, the next step for implementing the framework present in Figure 5.2 is to develop a recurrent network capable of predicting several slices of breast MRI. In this stage, given one slice of MRI, the recurrent network should predict the following sequence of slices (one-to-many). There are two main ways for implementing a recurrent network on image data: either the data is somehow compressed and flattened so a normal Recurrent Neural Network (RNN) Cell can be applied, or using Convolutional LSTM (Long Short-Term Memory) Cells [120], which can deal with image data. Different experiments were made with these two options, although most of these could not produce high-quality results, and only visual results are shown for demonstration of the ability for synthesizing breast MRI sequences.

5.2.2.1 Recurrent Autoencoder

Most of the work found in the literature for dealing with images using recurrent networks mentioned the compression of the data with an encoder, posterior pass through a recurrent cell, and finally decoding the compressed data into its original state and size [119, 38, 66]. An architecture based on this was developed and its diagram is presented in Figure 5.6. The input slice 1 goes through an encoder with 4 blocks of 2D convolutions (kernel size 3, padding 1, stride 1), ReLU activation and max pooling, using filter channels of 16, 32, 64, and 4, respectively. Then the feature vector gets flattened so it can be consumed by a memory-based Cell. This memory-based

cell outputs a vector that is reshaped into l feature vectors and each go through that go through a decoder with 4 blocks of 2D transpose convolutions (kernel size 4, padding 1, stride 2), ReLU activation on the first 3 blocks and sigmoid on the last one and with filter channels of 64, 32, 16, and 1, respectively. For the recurrent cell itself, either an RNN, LSTM, or GRU (Gated Recurrent Unit) cell could be used.

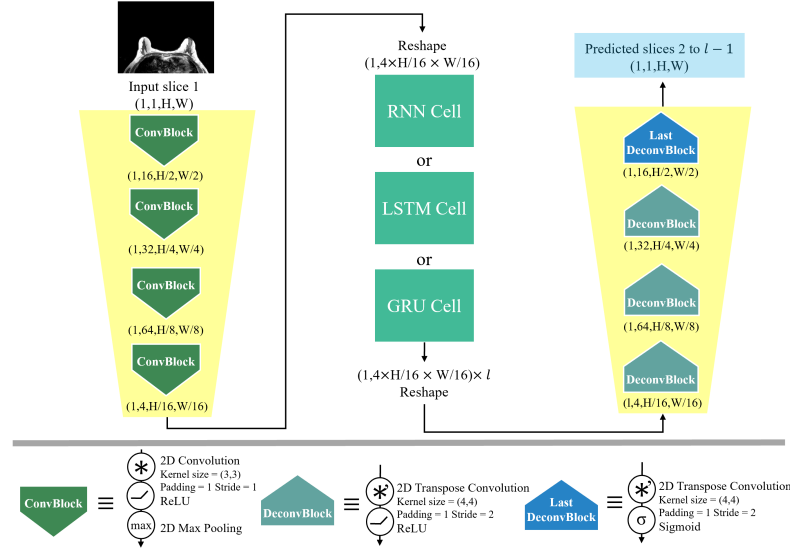


Figure 5.6: Representation of the recurrent autoencoder network for 3D MRI reconstruction.

The dataset used was a sample from the Volumes Resized Subset (Figure 5.1), only using the slices 29-35 of the scans. The autoencoder was trained before the recurrent cell, as this was proven to make the process faster. The autoencoder was trained with Adam's optimizer during 300 epochs with a learning rate of $1E-3$, with batch size 32 and input size 512 (height and width). The loss function used combined the L2 and L1 terms, with double weight for the L1. The entire architecture was trained after, with the same input size, keeping the pre-trained autoencoder on evaluation mode for 30 epochs in a total of 100 epochs. Now, each batch corresponds to a slice 29 and the following slices 30-35 ($l = 5$), so the batch size was 1, and other implementation details were similar to the autoencoder. The model was not able to reconstruct images with quality with any recurrent cell. An example of the output sequence from the model using an LSTM Cell with hidden and input sizes of 4096, and 1 layer is present in Figure 5.7. In addition to not reproducing images with enough quality, the model was not able to learn any transformations from slice to slice.

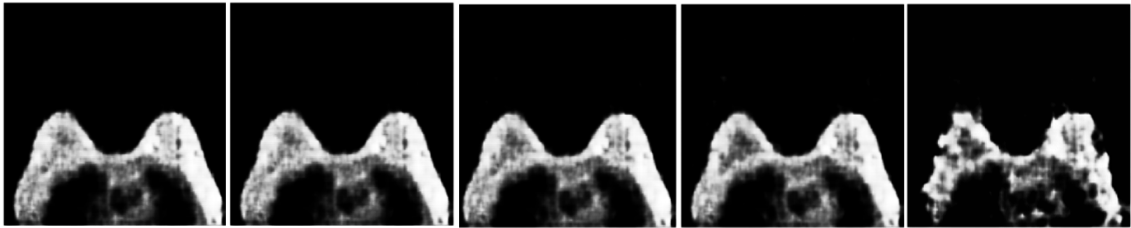


Figure 5.7: Sequence of slices generated with the Recurrent Autoencoder using LSTM Cell.

5.2.2.2 Convolutional LSTM Networks

ConvLSTM Cells [120], are a type of recurrent neural network for spatio-temporal prediction with convolutional features in both input-to-state and state-to-state transitions. This allows the network to deal with data with more than one dimension, such as images. The key functionalities of ConvLSTM Cells are described in Figure 5.8. The convolutional operators are used to determine the future state of a certain cell in the grid, by the inputs and past states of its local neighbors. This type of cells can be used to develop different architectures, in this case, for reconstructing breast MRI scans.

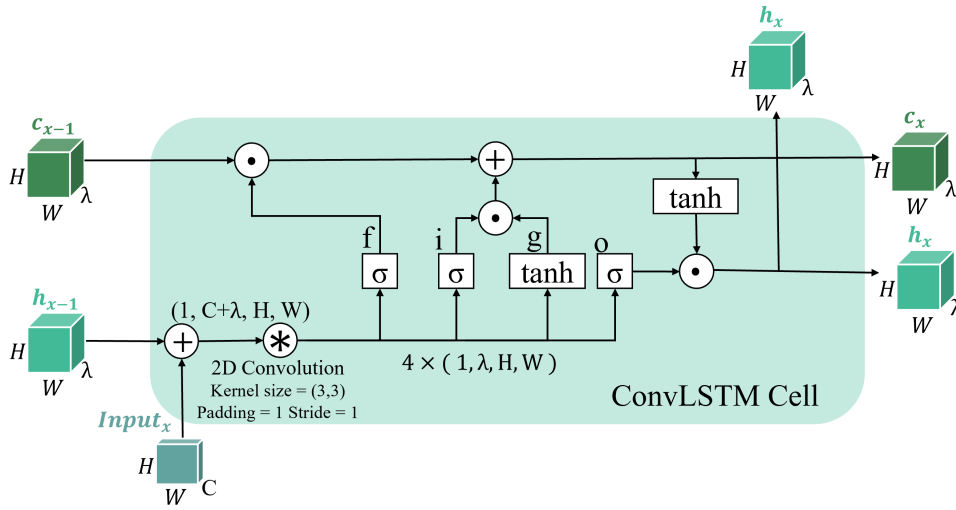


Figure 5.8: Diagram of ConvLSTM Cell. Convolutional operation denoted by (*), Hadamard product by ($\cdot$), sigmoid activation by (σ) and hyperbolic tangent activation by ($\tanh$). h is the hidden state, c is the cell state, and λ corresponds to the hidden and cell state dimension. x denotes the current timestep. Both input, hidden state and cell state have resolution H, W . Input has C channels and the hidden dimension of the cell has λ channels.

One-to-Many ConvLSTM Based on the premises of the work proposed in [84], a one-to-many ConvLSTM architecture² was developed. The goal of the model was to reconstruct a sequence of l slices of MRI given only one slice. This architecture is referred to as One-to-Many ConvLSTM Base and it is represented in Figure 5.9. This architecture is constituted by n ConvLSTM Cells vertically stacked. The input can either correspond to a MRI slice 1 and $l - 1$ black images (zeros) at each timestep (Figure 5.9a), or a MRI slice 1 followed by $l - 1$ slices predicted by the *pix2pix* model trained to generate the next slice of MRI developed in Section 5.2.1 (Figure 5.9b). Still, this model was trained so it had to reconstruct MRI slice 1, which was not necessary.

For training, the Volumes Resized Subset was used, with slices 30-35 of each scan ($l = 5$), and with input size 256. Ten ConvLSTM Cells (n) were used with hidden dimensions of sizes [4, 8, 16, 32, 64, 32, 16, 8, 4, 1], respectively. The last hidden dimension is 1 so that the output has only one channel (gray image). First experiences used the whole MRI scan, but it was computationally costly and slow. The loss function combined the L2 and L1 terms, with L1

²Inspiration: <https://github.com/spacejake/convLSTM.pytorch/blob/master/denoise-autoencoder.py>

multiplied by a factor of 10. Adam's optimizer was used with learning rate $1E-4$ during 100 epochs. The batch size used was 1 and a reset of the hidden and cell states is performed between each batch. Figure 5.10 holds the visual results from two different experiments made with this model, with and without the prediction from the previous *pix2pix* model (best model for input size 256). It is visible that using the pre-trained *pix2pix* model leads to slightly less blurry results, although the quality of all results is not favorable.

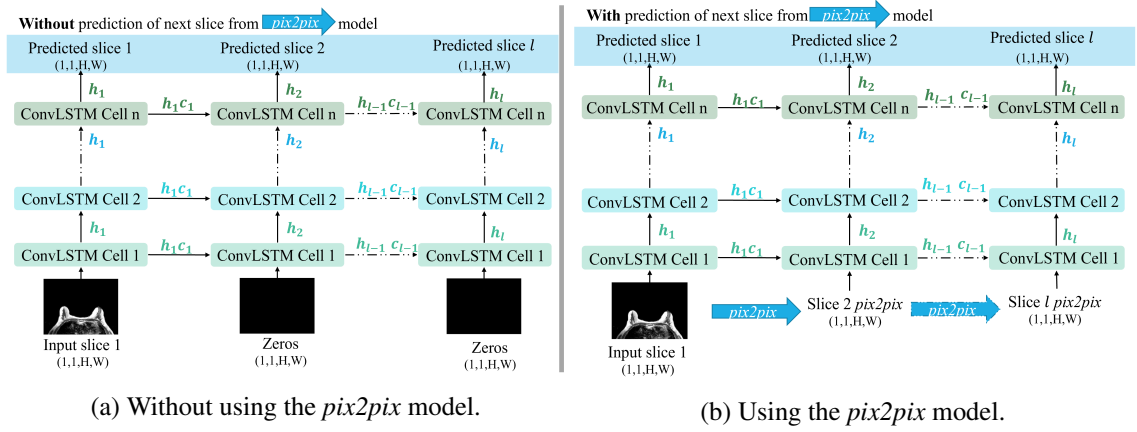
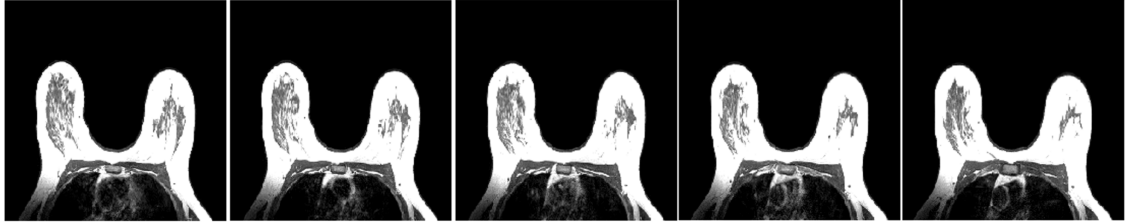


Figure 5.9: Architectures of **One-to-Many ConvLSTM Base**.

Original slices 30-35 from train set scan



Without prediction of next slice from *pix2pix* model



With prediction of next slice from *pix2pix* model



Figure 5.10: Visual performance of **One-to-Many ConvLSTM Base** models.

The One-to-Many ConvLSTM Base model was trained to reconstruct the slice from the input unnecessarily, so, in One-to-Many ConvLSTM Chain (Figure 5.11), the model was trained as a chain, feeding a prediction to the next input (from the engineering term, daisy chain). This model can also operate using the *pix2pix* model trained to generate the next slice for flow generation (Figure 5.11b) or only using the input image and the ConvLSTM model itself (Figure 5.11a).

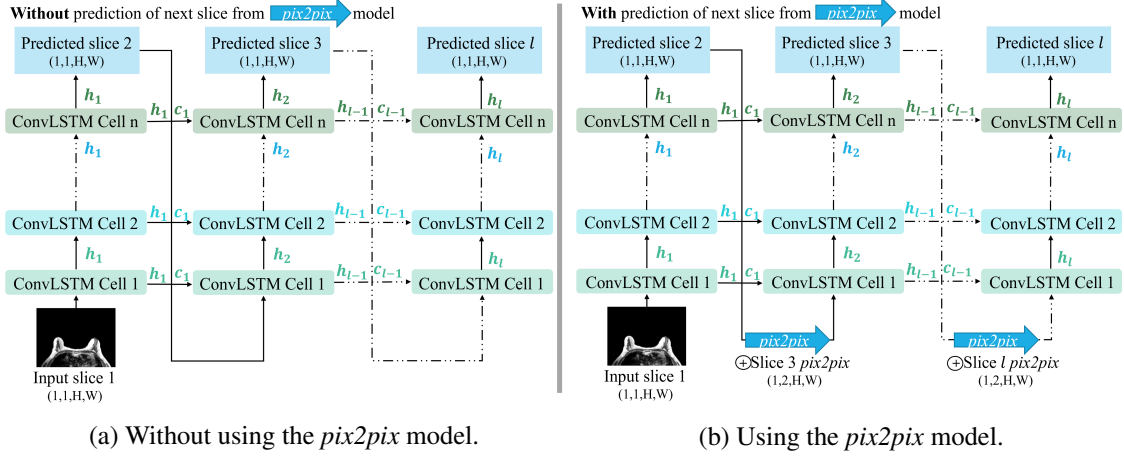


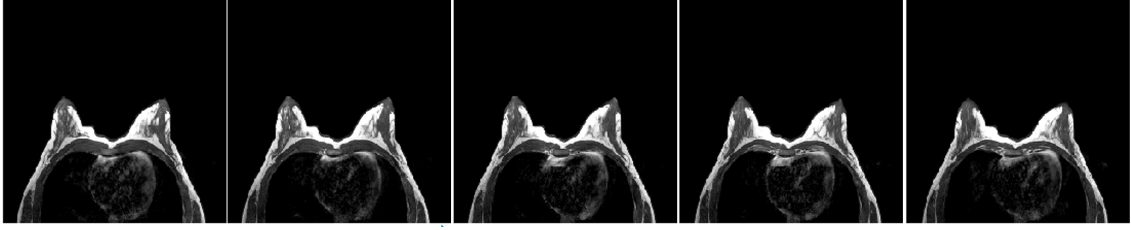
Figure 5.11: Architectures of **One-to-Many ConvLSTM Chain**.

The other implementation details were similar to the One-to-Many ConvLSTM Base, although a learning rate decayed at rate 0.7 every 5 epochs. The slice given as input was slice 29 from the MRI, while the model was trained to predict slices 30-35. Still, the results were not favorable and similar to the previous model. Given that, this experiment was discarded.

In another experiment, instead of aiming to predict only slices 30-35 from the MRI scans, all the data was used, although the sequence length of $l = 5$ slices was kept. The implementation details were also kept from the previous experiment. The intuition was that increasing the number of instances of the dataset would enable the proper learning of features. The visual results can be found in Figure 5.12b, that shows a sequence of 5 predicted slices where the input given was from the validation set. Figure 5.12a shows the original sequence. The predicted slices from the model show that the latter can predict the shape of the breast correctly, but some structures inside the breast appear too blurry to have any medical usage. For instance, the vessels and fibroglandular tissues appear indistinguishable from the fatty tissues of the breast, as this area is almost completely white.

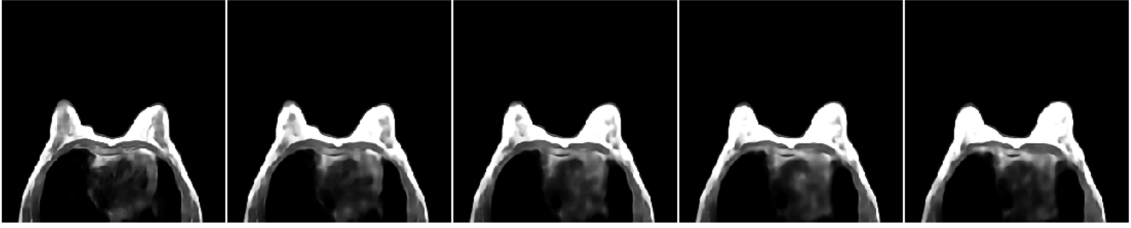
This experiment was repeated with input size 512, since using a larger image may help capture the features with more fidelity, as seen in the *pix2pix* experiment for predicting the next slice. The pre-trained *pix2pix* model used was also the one trained with input size 512. An additional trial with more two ConvLSTM Cells was performed, with hidden dimensions of sizes [4, 8, 16, 32, 64, 128, 64, 32, 16, 8, 4, 1], respectively. The results from these two experiments are shown in Figure 5.12c. The experiment with 12 cells shows some improvement in visual quality comparing to the others portrayed in Figure 5.12. However, this was not enough to portray breast MRI scans correctly.

Original sequence of 5 slices from scan on the validation set



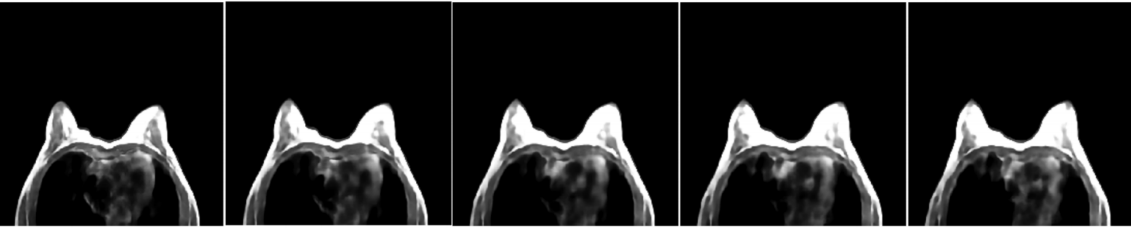
(a) Original sequence.

With prediction of next slice from *pix2pix* model
Hidden dimensions = [4, 8, 16, 32, 64, 32, 16, 8, 4, 1]

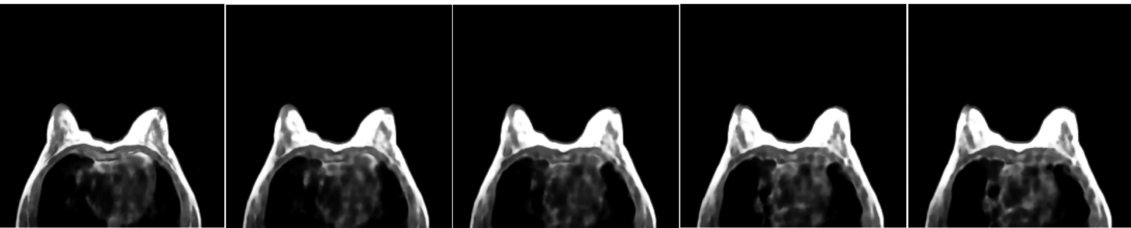


(b) Results with input size 256.

With prediction of next slice from *pix2pix* model
Hidden dimensions = [4, 8, 16, 32, 64, 32, 16, 8, 4, 1]



With prediction of next slice from *pix2pix* model
Hidden dimensions = [4, 8, 16, 32, 64, 128, 64, 32, 16, 8, 4, 1]



(c) Results with input size 512.

Figure 5.12: Visual performance of One-to-Many ConvLSTM Chain models.

Many-to-One ConvLSTM Since the results from the previous models for 3D MRI reconstruction using ConvLSTM architectures were not satisfactory, the model was modified to make the learning process easier. To do so, the previous One-to-Many ConvLSTM Chain was transformed into a Many-to-One ConvLSTM that instead of predicting a sequence of slices given one slice of MRI, aims to predict only one slice l after given a sequence of $l - 1$ slices. This way, the model can capture features from different sequential slices to reconstruct the immediate next slice. The diagram of the developed network is present in Figure 5.13. It is important to notice that no architecture including the use of the *pix2pix* pre-trained model for predicting the next slice was

used in this strategy, since this was the purpose of the Many-to-One ConvLSTM model, and it was not proven that using the *pix2pix* model improved the performance in a relevant way.

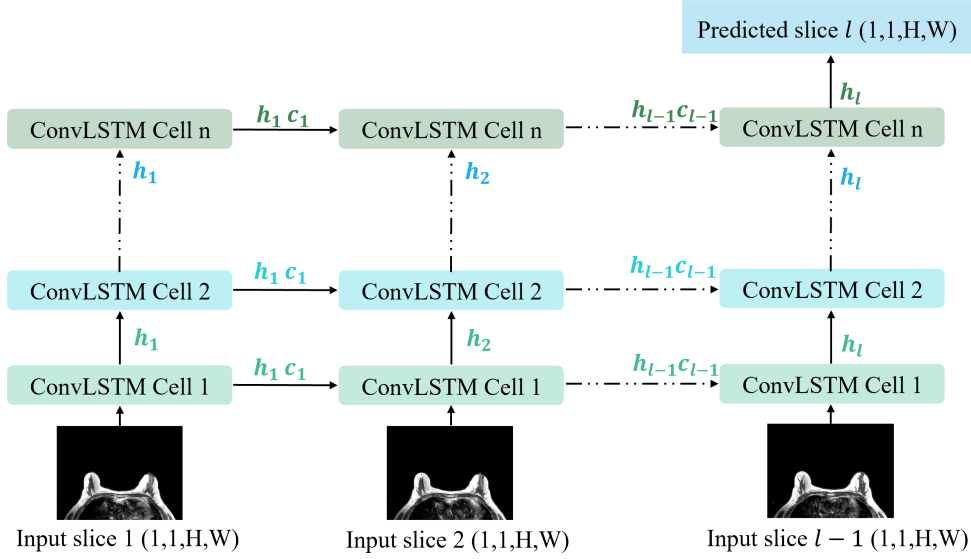


Figure 5.13: Architecture of **Many-to-One ConvLSTM**.

With this model, the influence of the number of ConvLSTM Cells n was explored as well as their hidden dimension sizes. Furthermore, the performance of the model under different loss functions and multiplying factors was explored. These experiments were made using the Volumes Resized Subset, Adam's optimizer during 100 epochs and initial learning rate of $1E-3$ with a decay rate of 0.7, every 5 epochs, and batch size 1 (resetting hidden and cell state between batches). The input size used was 512. Some brief observations over the results are present in Table 5.2.

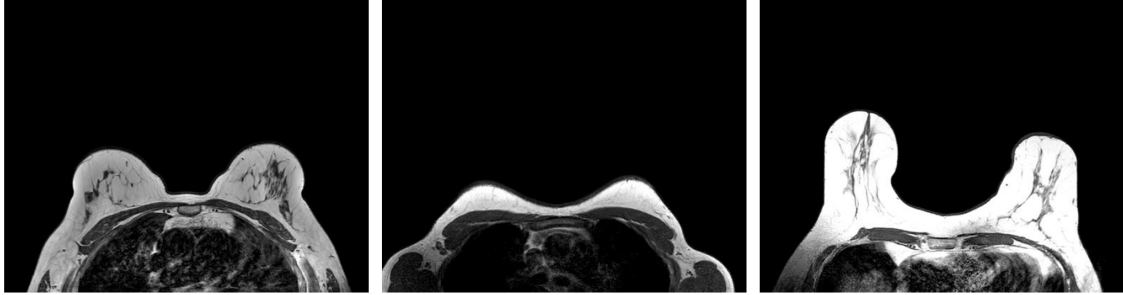
Table 5.2: Experiments performed with Many-to-One ConvLSTM model on Volumes Resized Subset.

Loss Function	n	Hidden dimension sizes	Results observations
$10 \times L1 + L2$	1	[1]	Higher quality, but very similar to slice $l-1$
	8	[8, 16, 32, 64, 32, 16, 8, 1]	Very blurry
	10	[8, 16, 32, 64, 128, 64, 32, 16, 8, 1]	Very blurry
$L1 + L2$	4	[4, 8, 4, 1]	Blurry
$L1 + MS-SSIM$	1	[1]	Higher quality, but very similar to slice $l-1$
	6	[8, 16, 32, 16, 8, 1]	Very blurry
	6	[8, 8, 8, 8, 8, 1]	Very blurry

Decreasing the number of cells and hidden dimension sizes increases the quality of the predicted slices, however, they look very similar to the last slice from the input sequence given to the model. As the number of cells and hidden dimension sizes increase, predicted slices get blurry, so it is unviable to use the models to perform reconstruction of a whole volume by using the output slice as the last input sequence slice and continuing this method for reconstructing a whole scan.

An additional experiment was done on the Volumes Padded Subset (Figure 5.1), using $n = 6$ ConvLSTM cells with hidden dimension sizes $[8, 16, 32, 16, 8, 1]$, respectively. The other implementation details were like the other experiments and the loss function used was the combination of L1 and MS-SSIM. The batch size was increased to 4, although hidden and cell states were reset between each sequence on the batch. Visual results from this model are present in Figure 5.14.

Original slices l from scan on the validation set



Predicted slices l from scan on the validation set

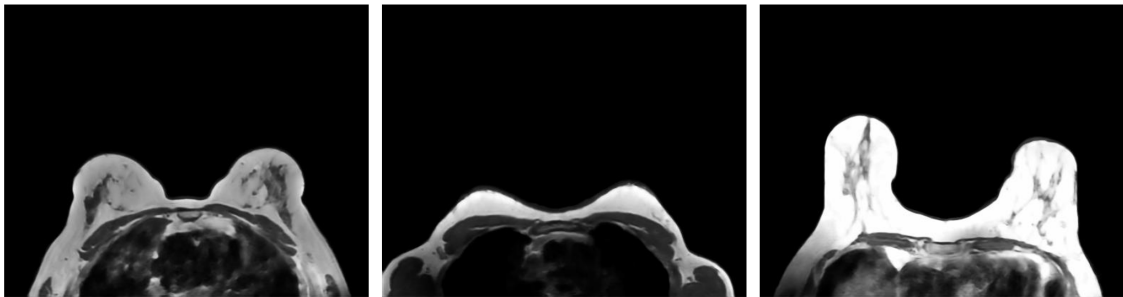


Figure 5.14: Visual performance of Many-to-One ConvLSTM model on Volumes Padded Subset.

There is still a little blurriness on the reconstructions, but much less than the other experiments. However, when trying to use this model to reconstruct a part of a volume (using the predicted slice as the last input sequence slice recursively), rapidly the results start to get more blurry and without any image quality of correct breast MRI features represented correctly (Figure 5.15).

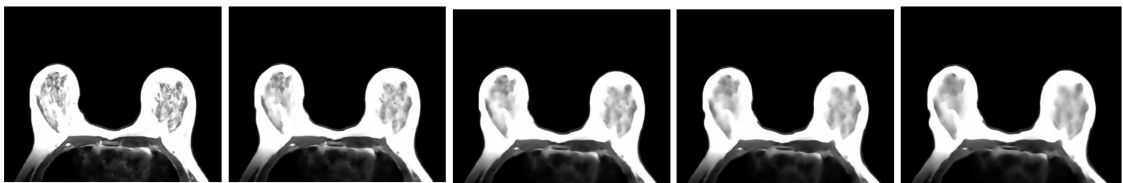


Figure 5.15: Sequence of slices generated with the **Many-To-One ConvLSTM** model. From left to right, the predicted slice is added to the input of the model while the previous first slice from the input sequence is withdrawn.

For a final experiment, the Many-to-One ConvLSTM model was updated to a Many-to-One ConvLSTM Autoencoder, which is represented in Figure 5.16. It uses an autoencoder to decrease the dimensionality of the data and only then the feature vectors go through the ConvLSTM network, which is equal to the one in Figure 5.13. The input images go through 3 downsizing blocks, with

2D convolutions (filter channels 32, kernel size 4, padding 1, stride 2), Leaky ReLU activation in the first two blocks and 3D Batch Normalization in the last two blocks. The input features go through the Many-to-One ConvLSTM network and the latter outputs a vector then decodes into the predicted slice l by an upsampling decoder with 3 blocks of 2D transpose convolutions (filter channels 32, 32, and 1, respectively, kernel size 4, padding 1, stride 2) with ReLU activation of the first two blocks and sigmoid on the last one. The rationality behind this network was to minimize the number of features needed to be learned by the model.

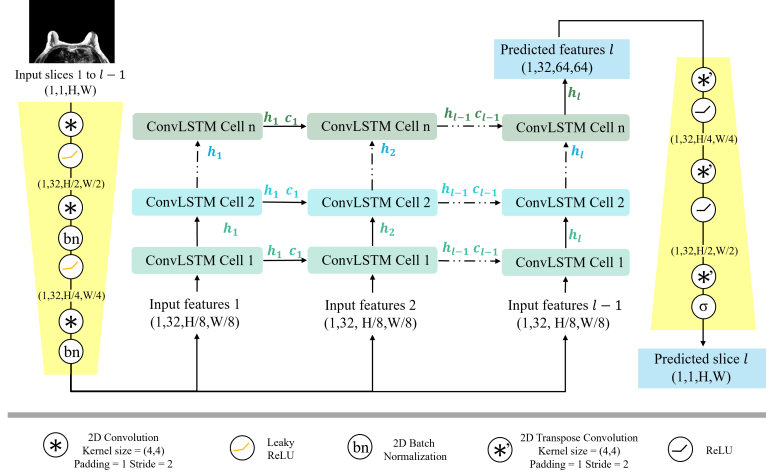


Figure 5.16: Architecture of **Many-to-One ConvLSTM Autoencoder**.

This architecture was trained on the Volumes Squared Subset, with images with size 512×512 resolution. Similar to the Recurrent Autoencoder in Section 5.2.2.1, the autoencoder was trained separately to make the learning process faster. To train this autoencoder, Adam's optimizer with constant learning of $2E-4$ was used during 100 epochs and with a batch size of 32. The loss function used between $10 \times L1 + L2$ and $L1 + MS\text{-}SSIM$ and the latter obtained better results (mean PSNR values on the validation set of 28.71 ± 4.03 versus 28.55 ± 4.12). However, while training the entire network with the ConvLSTM component using the latter trained model of the autoencoder and with hidden dimension sizes $[8, 16, 32, 64, 32]$ and loss function $10 \times L1 + L2$, the visual results are more blurry than the ones from the model without the autoencoder part, so this experiment was discarded.

Although many experiments were developed for breast MRI reconstruction using memory-based networks, such as RNNs and LSTMs, it was not possible to develop one that corresponded to the expectations of this study. The generation ability of the Many-to-One ConvLSTM models is increased when using more cells (8 or 10), but this also results in poor quality images. It is important to remember that the number of convolutions is equal to the number of cells used, so even with 10 cells, only 10 convolutions are applied to the data, which may not be enough for learning how to capture the relevant features from the breast MRI. It is arguable that a more thorough study of the number of cells and hidden dimension sizes may improve the results, but the current results show no motivation for continuing to exploring these architectures.

5.3 Auto-encoding GAN for Breast MRI Generation

One alternative for generating 3D MRI scans presented in Section 3.2 was the Auto-encoding GAN [59] (Figure 3.11). This framework uses a 3D GAN for generating data, addressing the mode collapse problems by leveraging the α -GAN, while combining a variational autoencoder (VAE) and the GAN by using an additional "code" discriminator network [110]. Additionally, the authors use the Wasserstein GAN with Gradient Penalty (WGAN-GP) loss to increase the stability of the model during training. The implementation was inspired by the original one³ although some modifications were made to the architecture, so it was able to process the available breast MRI data. The latent size was decreased from 1000 to 100, due to computational cost. Taking this into consideration, the original network was transformed in the one schemetized in Figure 5.17.

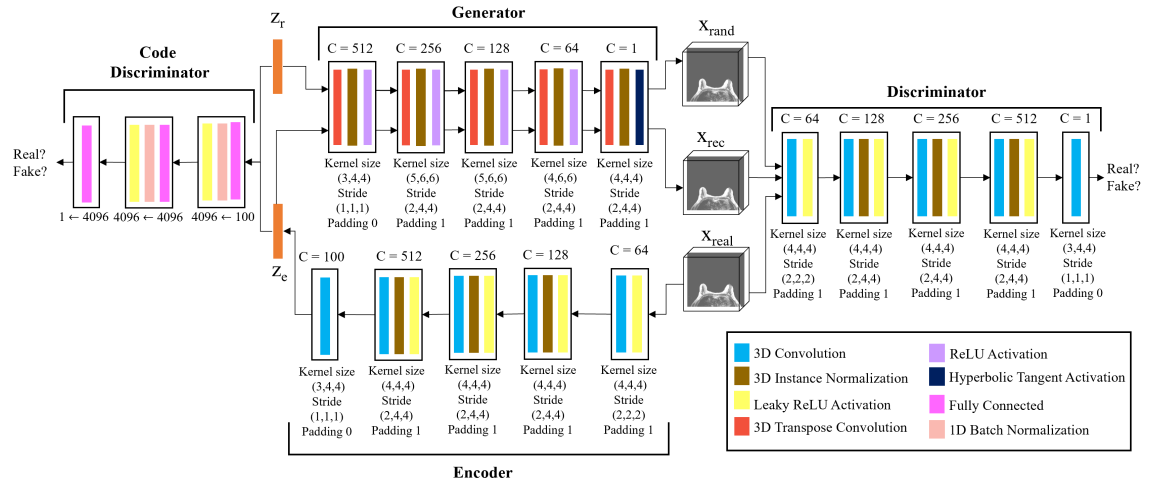


Figure 5.17: Architecture of **Auto-encoding GAN (α -WGAN)**. The Discriminator is trained to distinguish between real, reconstructed and random volumes (x_{real} , x_{rec} , and x_{rand} , respectively). The Encoder downsamples x_{real} , while the Generator upsamples the x_{rand} and x_{rec} from vectors with latent size 100, z_r and z_e , respectively. The Code Discriminator is trained to distinguish between z_r and z_e .

The model was trained using the Volumes Squared Subset (Figure 5.1) where all slices have a resolution of 512×512 . The loss function used to compare MRI scans to update the generator was the L1 term, multiplied by a factor of 10. All networks were trained using Adam's optimizer with a learning rate of $2E-4$. The model was intended to be trained throughout 200000 iterations with a batch size equal to one. However, earlier in the training process, intermediary results showed that the model was not able to learn correctly the image features (see Figure 5.18). The reconstructions are not corresponding to the input scans and have very poor image quality. In addition, the training process was time-consuming and computationally expensive, so it was not efficient for training. This may have happened due to the modifications of the network, especially in the generator, and the decrease of the latent size. Possible culprit is most likely the reduced dataset for training a network that aims to generate and entire volume of MRI.

³<https://github.com/cyclomon/3dbraingen>

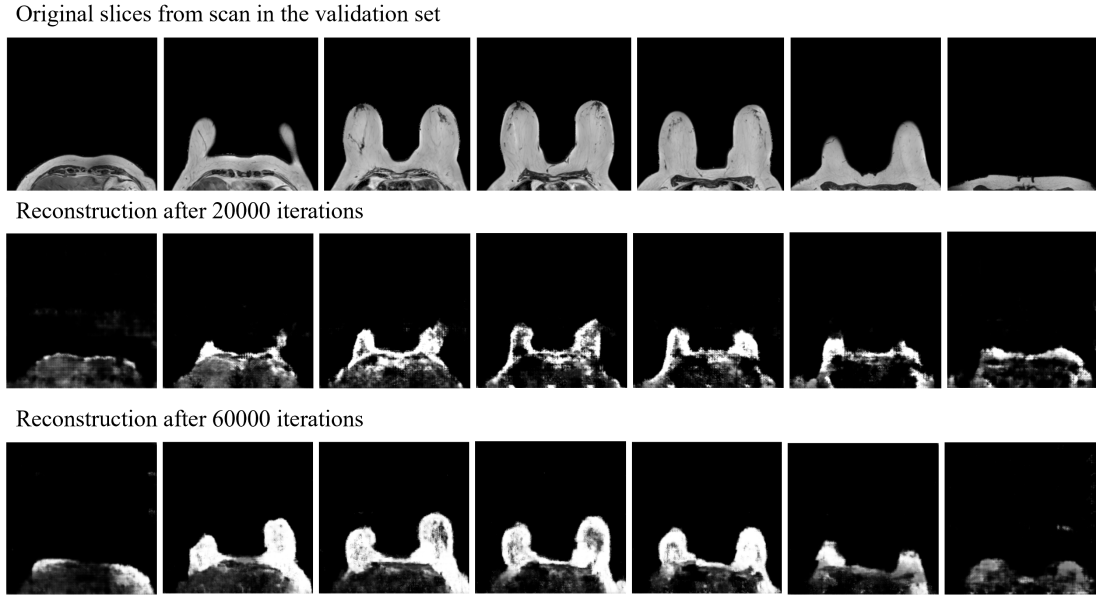


Figure 5.18: Visual performance of **Auto-encoding GAN (α -WGAN)** during train. Different slices from the same scan (slices 0, 10, 20, 30, 40, 50, 60) are shown.

5.4 3D X-Net for Breast MRI Reconstruction

As seen in the experiments described above, it is difficult to train DL architectures with MRI data, mostly because they require large amounts of data. The data itself is hard to comprehend for human trained eyes, even more for a statistical model to learn how to capture relevant features and generate new volumes. Until now, only methodologies that consumed entire MRI volumes or slice sequences and tried to reconstruct or generate new data from prior distributions were explored. These methods were not capable to perform as well as desired. Given that, in this Section, methods that use only parts of the MRI as their input are presented. To do so, the Patches Subset is used (Figure 5.1), which divides the breast MRI scans into small volumetric patches, which may be easier to model by DL models.

The learning strategy presented in this Section is inspired by the GPT models [16], where the goal is to recreate a part of the input that was not given. Here, the model should reconstruct an incomplete breast MRI, taking advantage of the complete annotation patch to learn how to predict the MRI patch completely. To summarize, the model should receive two inputs: an MRI patch with some incomplete part (all zeros), and a corresponding annotation patch. After a forward pass, the network must be able to output the MRI patch entirely reconstructed and the corresponding annotation patch.

Considering these aspects, a 3D X-Net was developed, inspired by [29], represented in Figure 5.19. This network combines the encoded parts of the two inputs to create each of the outputs. The encoding part is composed of blocks of 3D convolutions (kernel size 3, padding 1, stride 1), ReLU activations, and batch normalization. These operations are repeated twice per block. After each block, a max pooling operation is performed. On the first two blocks, the max pooling

has a kernel size of (2,2,2) and stride (2,2,2), and on the third block a kernel size of (1,2,2) and stride (1,2,2). The filter channels of each convolution are 64, 64, 128, 128, 256, and 256, respectively. Then, the latent vectors from the incomplete MRI and annotation patches are concatenated along channels and go through a 3D convolution with 256 filter channels (kernel size 3, padding 1, stride 1). The combined latent vector is copied and used for decoding the MRI and annotation patches. In the decoding, the vector goes through an interpolation operation with scale factor (1,2,2) and then, convolutional blocks similar to the ones used in the encoding part are used. Between blocks, an interpolation with scale factor of (2,2,2) is used. The last convolution has kernel size 1, stride 1, padding 0, and it is accompanied by a sigmoid operation to change the range of the output, and no batch normalization is applied. Also, the feature maps of each block on the encoding are concatenated alongside the decoding, similar to a U-Net architecture [109]. The filter channels used in the decoding are then 256, 128, 128, 64, 64, and 1, respectively.

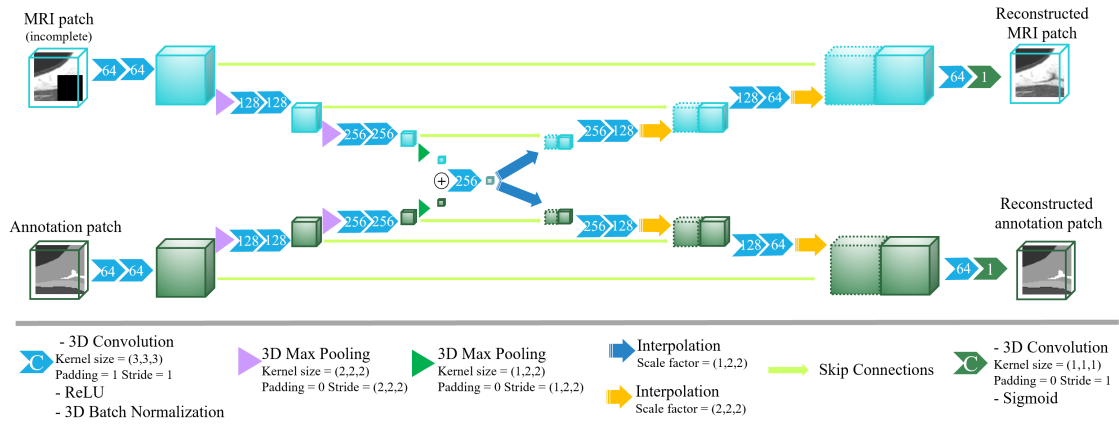


Figure 5.19: Architecture of the **3D X-Net** for breast MRI reconstruction.

This network can be trained in two fashions, represented in Figure 5.20:

- In **Generator Training (GenT)**, the 3D X-Net is trained only considering the differences between the reconstructed patches and original ones (Visual Loss).
- In **Adversarial Training (AdvT)**, by adding a 3D PatchGAN discriminator [65] it is possible to train the framework as a GAN. Here, the 3D X-Net acts as a conditional Generator, where the condition are the incomplete MRI patch and the annotation patch. Given that, the condition c given to the 3D PatchGAN discriminator can also be either one of these patches or both.

The 3D PatchGAN discriminator, represented in Figure 5.21, is trained to distinguish between the reconstructed and original MRI patches. The discriminator takes the condition c (incomplete MRI patch, annotation patch, or both) and the reconstructed or original MRI patch to output a patch used for calculating the GAN loss for the "fake" and "real" data, respectively. The inputs are concatenated and pass through a series of 3D convolutions (kernel size 3, stride 1, padding 1), ReLU activation, and 3D Batch Normalization. To downsample the patches, max pooling between

the convolutional blocks is used, with kernel size (2,2,2), stride (2,2,2), after the first convolutional block, and kernel size (1,2,2), stride (1,2,2) on the others. The last convolutional block does not have an activation function or batch normalization. The filter channels used are 64, 128, 256, 512, and 1, respectively.

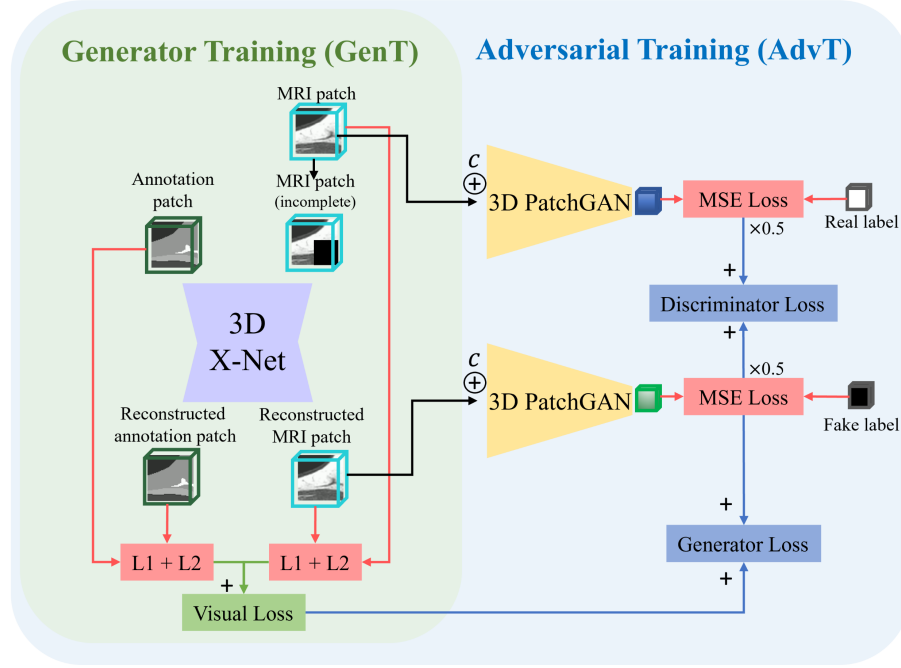


Figure 5.20: Generator and Adversarial Training using the 3D X-Net architecture. In a green background (on the left), the Generator Training setting only trains the 3D X-Net architecture with a Visual Loss between the original and reconstructed patches. In a blue background (on the right), the Adversarial Training setting uses 3D PatchGAN discriminator for a GAN setting, whereas the 3D X-Net is used as the generator. c is the condition given to the discriminator.

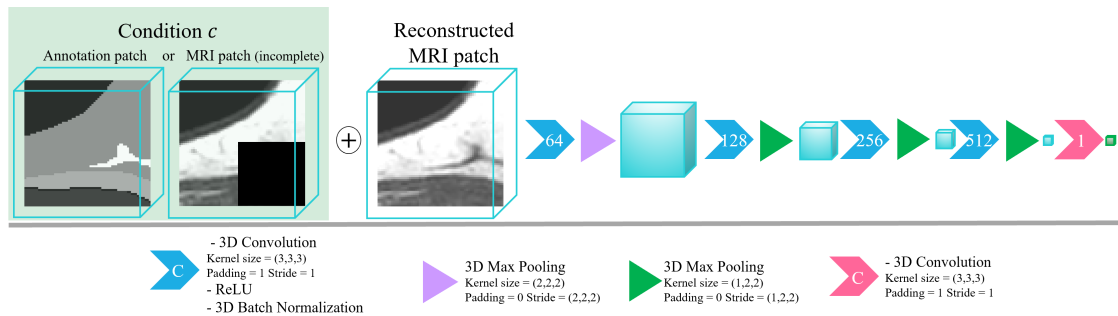


Figure 5.21: Architecture of the **3D PatchGAN** for adversarial training of 3D X-Net architecture.

In this Section, the methodologies involving the usage of these networks in these three different scenarios are presented, showing the key results of the methods which helped to evolve to the next strategies.

5.4.1 Generator Training on (64, 64, 8) patches

The first attempt to reconstructing MRI scans using the X-Net used patches with size $64 \times 64 \times 8$ where the last 4 slices on the right bottom corner (one-eighth of the patch) are incomplete, as Figure 5.22 suggests.

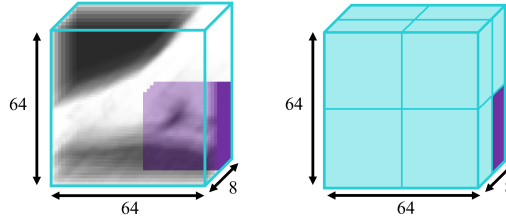


Figure 5.22: Schematization of the incomplete (64, 64, 8) patches used with X-Net architecture.

After creating all the patches from the scans available, the patches that did not contain any MRI information were withdrawn from the dataset, as they do not hold any useful information. Adam's optimizer was used with an initial learning rate of $1E-3$ with a decay rate of 0.9 every 10 epochs optimizing the combination between L1+L2 norms. Two different training cycles were completed: Equal Loss where the loss is equally calculated between the two pairs of original and reconstructed patches (MRI and annotations), and Weighted Loss where the corner to be predicted on the MRI has 8 times more weight than the other eighths of the MRI patch (loss between annotation patches was kept equal for the whole patch).

The reconstructions were evaluated in terms of MSE, PSNR, and SSIM, which are available in Table 5.3. To distinguish between these two experiments, these metrics were calculated between entire patches (MRI and Annotation) and between only the corner to be predicted in the case of MRI (MRI corner). The results from this Table show that it is advantageous to give more weight to the corner that was incomplete in the input to perform a better prediction. The reconstruction metrics of the annotation patches were also better using the weighted loss. However, the overall MRI patch reconstruction is slightly worst.

Table 5.3: Numerical performance of 3D X-Net Generator Training on (64, 64, 8) patches: mean and standard deviation distances are shown.

(a) Training set (16410 patches).

Experiment	MRI			MRI corner			Annotation		
	MSE↓	PSNR↑	SSIM↑	MSE↓	PSNR↑	SSIM↑	MSE↓	PSNR↑	SSIM↑
GenT - Equal Loss	0.05 ± 0.07	17.56 ± 7.60	0.61 ± 0.31	0.18 ± 0.28	12.92 ± 7.60	0.29 ± 0.31	0.08 ± 0.11	14.99 ± 6.11	0.53 ± 0.30
GenT - Weighted Loss	0.05 ± 0.07	16.08 ± 6.17	0.57 ± 0.31	0.15 ± 0.25	15.17 ± 9.05	0.34 ± 0.35	0.64 ± 0.10	18.14 ± 8.52	0.57 ± 0.33

(b) Validation Set (5330 patches).

Experiment	MRI			MRI corner			Annotation		
	MSE↓	PSNR↑	SSIM↑	MSE↓	PSNR↑	SSIM↑	MSE↓	PSNR↑	SSIM↑
GenT - Equal Loss	0.04 ± 0.06	17.87 ± 6.80	0.62 ± 0.32	0.14 ± 0.23	13.21 ± 7.08	0.29 ± 0.31	0.07 ± 0.10	15.21 ± 6.12	0.53 ± 0.30
GenT - Weighted Loss	0.04 ± 0.06	16.49 ± 5.44	0.59 ± 0.32	0.11 ± 0.20	15.18 ± 8.14	0.34 ± 0.33	0.06 ± 0.09	18.65 ± 8.60	0.61 ± 0.33

The reconstruction of entire MRI scans was not performed since it is time-consuming process. The intuition behind using one-eighth of the patch was that the model may need a three-dimensional neighborhood to learn the features from the patch. However, since breast MRI slices have 3 millimeters of thickness, the structures differ considerably from one to another, making the neighborhood of the third dimension not so necessary for the learning process.

5.4.2 Generator and Adversarial Training on (64, 64, 4) patches

The second round of experiments using the 3D X-Net architecture used patches with 4 slices of depth instead of 8, and one-quarter of the patch (right bottom quarter) was blacked out on the MRI patch at the input of the network (see Figure 5.23). As previously described, the neighborhood on the third dimension may not be necessary due to the data specifications, so it was discarded. Furthermore, decreasing the number of slices to reconstruct may help the model to efficiently learn the features of the patch since the reconstructed patches are smaller.

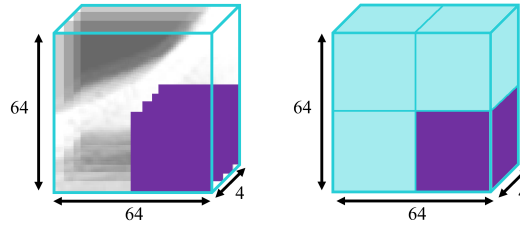


Figure 5.23: Schematization of the incomplete (64, 64, 4) patches used with X-Net architecture. On the left a volume of $64 \times 64 \times 4$ MRI slices and, on the right, a cubic representation of the incomplete patch. The incomplete part of the patch is represented in purple.

Implementation Details Weighted L1+L2 loss was used since it was beneficial in the first approach with this model. For the annotation pairs (original and reconstruction) this loss has weight 1. In the MRI patches, the corners that were given in the input had their L1+L2 loss multiplied by a factor of one quarter each, while the incomplete corner L1+L2 loss had weight 1.

Three different methods using Adam’s optimizer were used, described in Table 5.4. GenT was trained using all of these methods, while AdvT used the M1 method for both the generator and the discriminator (GenT-M1, GenT-M2, and GenT-M3). In AdvT methods, two experiments were made according to the conditions c given to the discriminator: the incomplete MRI patch (AdvT-iMRI-M1) or the annotation patch (AdvT-Annot-M1).

Table 5.4: Methods with different hyperparameters used in 3D X-Net.

Method	Initial Learning Rate	Decay Rate	Step Decay	No. of Epochs
M1	1E-3	0.9	10	100
M2	1E-3	0.9	20	200
M3	5E-3	0.9	10	100

Reconstruction Metrics and Visual Results Similar to what was done in the first experiments, the reconstruction metrics for the pairs of patches were calculated. The results are presented in Table 5.5. There is a visible increase in performance when comparing to the first approach (Table 5.3). Approximately, for GenT, all the different methods had similar results in terms of these reconstruction metrics, although GenT-M2 seems to be the one with overall better performance. M1 was used for AdvT since it takes less training time than M2 (100 epochs instead of 200) and the reconstruction metrics between GenT-M1 and GenT-M3 were better with the M1 setting. There is some difference between the training and validation sets, being more noticeable in the MRI patches. This difference is more clear in MSE than PSNR or SSIM. The SSIM reveals these models can maintain the structure of the original patches and also the incomplete input part. Regarding these metrics, AdvT with the incomplete MRI patch as condition (AdvT-iMRI-M1) was not beneficial for the MRI reconstruction, as none of the metrics for the MRI was improved with the adversarial train. However, the differences between training and validation sets seem to be smaller when using the GAN setting, indicating an improvement in generalization ability. This is also verified in the AdvT-Annot-M1 model. The PSNR values for MRI and MRI corner with the AdvT-Annot-M1 model are considerably better than any other models, which might suggest that this model reproduces MRI patches with less noise. Given that, while AdvT-Annot-M1 model produces larger MSE and SSIM errors, it has increased image quality.

Table 5.5: Numerical performance of scan reconstructions with 3D X-Net Generator and Adversarial Training on (64, 64, 4) patches: mean and standard deviation distances are shown.

(a) Training set (33745 patches).

Method	Loss	MRI			MRI corner			Annotation		
		MSE $\times 10^{-4} \downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	MSE $\times 10^{-3} \downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	MSE $\times 10^{-6} \downarrow$	PSNR $\uparrow$	SSIM $\uparrow$
GenT-M1	0.034	4.45 $\pm$ 7.61	59.15 $\pm$ 24.72	0.98 $\pm$ 0.03	1.62 $\pm$ 2.87	66.84 $\pm$ 35.68	0.92 $\pm$ 0.12	2.72 $\pm$ 5.23	59.39 $\pm$ 5.16	0.99 $\pm$ 0.00
GenT-M2	0.028	3.86 $\pm$ 6.58	59.27 $\pm$ 24.19	0.98 $\pm$ 0.03	1.45 $\pm$ 2.55	66.76 $\pm$ 6.17	0.92 $\pm$ 0.12	7.35 $\pm$ 5.10	66.26 $\pm$ 11.93	0.99 $\pm$ 0.00
GenT-M3	0.034	3.69 $\pm$ 6.29	54.46 $\pm$ 19.80	0.98 $\pm$ 0.02	1.25 $\pm$ 2.26	59.13 $\pm$ 27.92	0.93 $\pm$ 0.10	2.18 $\pm$ 5.41	65.17 $\pm$ 8.79	0.99 $\pm$ 0.00
AdvT-iMRI-M1	0.041	22.70 $\pm$ 35.60	38.13 $\pm$ 12.48	0.94 $\pm$ 0.08	5.52 $\pm$ 10.14	40.39 $\pm$ 17.13	0.85 $\pm$ 0.21	19.40 $\pm$ 63.70	67.31 $\pm$ 18.25	0.99 $\pm$ 0.00
AdvT-Annot-M1	0.051	29.60 $\pm$ 53.80	73.87 $\pm$ 46.20	0.94 $\pm$ 0.09	8.90 $\pm$ 18.58	73.52 $\pm$ 47.34	0.83 $\pm$ 0.27	37.00 $\pm$ 70.00	65.89 $\pm$ 20.64	0.99 $\pm$ 0.00

(b) Validation Set (10750 patches).

Method	Loss	MRI			MRI corner			Annotation		
		MSE $\times 10^{-3} \downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	MSE $\times 10^{-3} \downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	MSE $\times 10^{-6} \downarrow$	PSNR $\uparrow$	SSIM $\uparrow$
GenT-M1	0.070	1.20 $\pm$ 2.61	58.63 $\pm$ 26.36	0.97 $\pm$ 0.05	4.66 $\pm$ 1.03	66.80 $\pm$ 37.42	0.88 $\pm$ 0.19	2.83 $\pm$ 6.42	59.35 $\pm$ 5.18	0.99 $\pm$ 0.00
GenT-M2	0.072	1.18 $\pm$ 2.62	58.58 $\pm$ 26.07	0.97 $\pm$ 0.05	4.67 $\pm$ 10.42	66.56 $\pm$ 37.12	0.88 $\pm$ 0.19	12.71 $\pm$ 7.22	66.10 $\pm$ 12.47	0.99 $\pm$ 0.00
GenT-M3	0.068	1.26 $\pm$ 2.71	53.30 $\pm$ 22.01	0.97 $\pm$ 0.04	4.88 $\pm$ 10.70	58.17 $\pm$ 30.37	0.88 $\pm$ 0.20	2.72 $\pm$ 9.05	65.08 $\pm$ 9.00	0.99 $\pm$ 0.00
AdvT-iMRI-M1	0.052	2.35 $\pm$ 3.83	38.44 $\pm$ 12.54	0.94 $\pm$ 0.08	6.29 $\pm$ 1.17	40.84 $\pm$ 17.38	0.85 $\pm$ 0.22	33.40 $\pm$ 157.10	67.12 $\pm$ 18.76	0.99 $\pm$ 0.00
AdvT-Annot-M1	0.058	2.89 $\pm$ 5.07	75.39 $\pm$ 46.21	0.94 $\pm$ 0.09	8.87 $\pm$ 17.53	75.34 $\pm$ 47.36	0.83 $\pm$ 0.26	38.90 $\pm$ 74.40	66.12 $\pm$ 20.78	0.99 $\pm$ 0.00

For reconstructing the entire volumes of MRI in their original size, an incomplete input of this volume is needed, that is meant to be completed by the trained models afterwards. Considering this, the inputs were designed for every 4 slices, since the patches also had 4 slices in depth. Each of these inputs is then analyzed and the network starts to complete it through a sliding window of the input patch size ($64 \times 64 \times 4$). This may enable the generation of volumes if incomplete parts of MRI scans are available. Only the incomplete part of each patch (bottom right corner) is used for the reconstruction. This was a time-consuming process, so it was only performed for one scan

in the training set and one scan in the validation set. For a better understanding of this process, Figure 5.24 details the inputs and some of the steps in the reconstruction, as well as the final result for one of the slices, for both MRI and annotation. At last, Figure 5.25 shows some reconstructed slices using the GenT-M2 model. More visual results can be found in Figures A.1 and A.2.

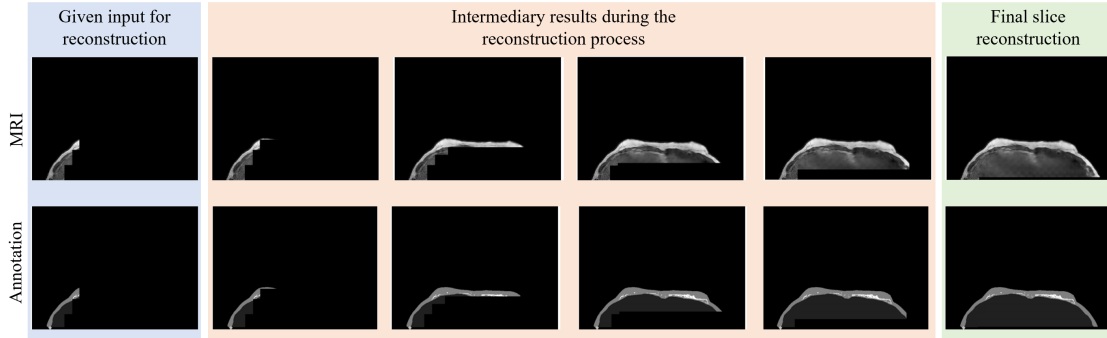


Figure 5.24: Reconstruction of entire MRI process from 3D patches.

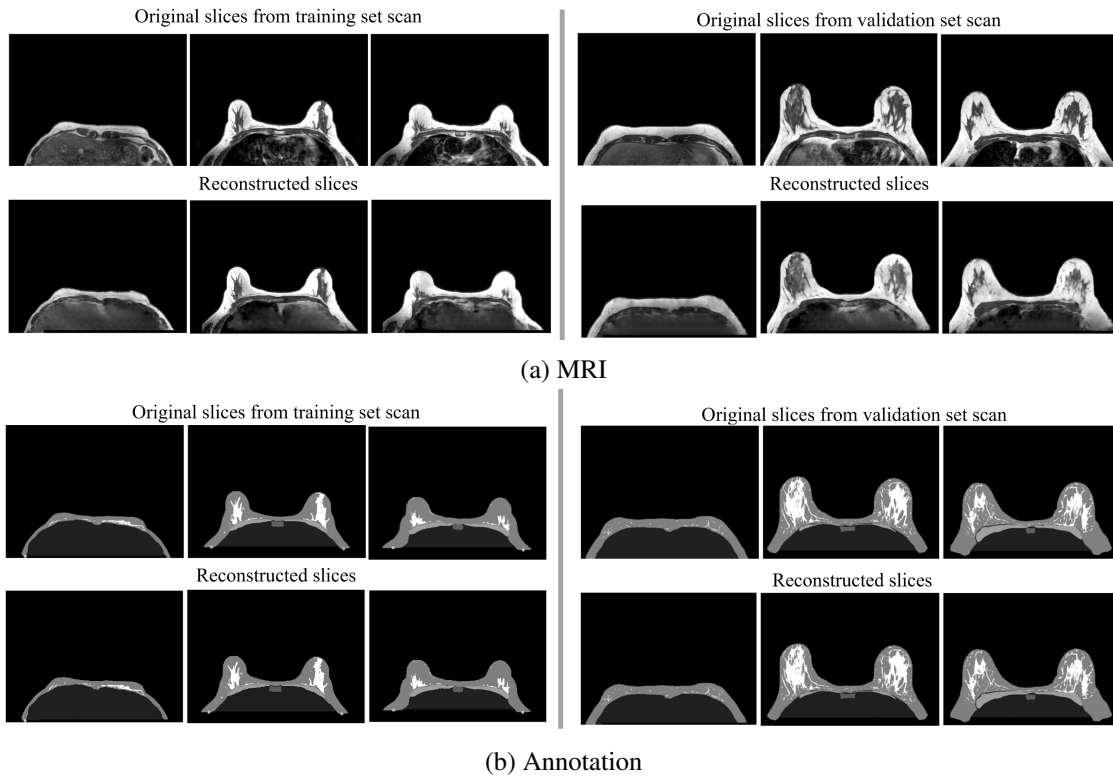


Figure 5.25: Visual performance of 3D X-Net on reconstructed slices with GenT-M2 model.

After completing the reconstruction, the reconstruction metrics the entire scans are computed. Besides MSE, PSNR, and SSIM, FID and DICE (annotation) scores were also calculated. These results are present in Table 5.6. The metrics show results with worse quality than in a patch-wise perspective. Overall, the PSNR is significantly lower (values between 18-20 instead of 60), while the SSIM and MSE were maintained similar to the patch-wise perspective. The FID in the MRI shows that the quality of the images is not yet favorable, surrounding the value of 250 for the GenT models. In the case of annotation, the FID value is low (1.20-1.57 in the training scan and 2.46-2.89 in the validation scan). This is expected since the annotation is given completely on the input, so the model only needs to reconstruct it. This was supported by all the other metrics.

Table 5.6: Numerical performance of 3D X-Net Generator and Adversarial Training on (64, 64, 4) patches: mean and standard deviation distances are shown.

(a) Reconstruction Metrics on scan from the training set.

Method	MRI				Annotation				
	MSE $\times 10^{-3}$ ↓	PSNR↑	SSIM↑	FID↓	MSE $\times 10^{-5}$ ↓	PSNR↑	SSIM↑	FID↓	DICE↑
GenT-M1	9.74 $\pm$ 1.19	20.19 $\pm$ 0.85	0.87 $\pm$ 0.01	246.93	1.40 $\pm$ 3.89	56.59 $\pm$ 6.36	0.99 $\pm$ 0.00	1.20	0.994
GenT-M2	9.26 $\pm$ 2.10	20.44 $\pm$ 0.99	0.86 $\pm$ 0.01	263.41	2.21 $\pm$ 3.87	51.68 $\pm$ 6.75	0.99 $\pm$ 0.00	1.20	0.997
GenT-M3	8.40 $\pm$ 2.01	20.89 $\pm$ 1.08	0.87 $\pm$ 0.02	266.24	1.40 $\pm$ 3.89	56.58 $\pm$ 6.43	0.99 $\pm$ 0.00	1.57	0.995
AdvT-iMRI-M1	14.40 $\pm$ 3.30	18.53 $\pm$ 1.03	0.84 $\pm$ 0.01	364.92	2.20 $\pm$ 3.80	49.36 $\pm$ 3.96	0.99 $\pm$ 0.00	2.06	0.974
AdvT-Annot-M1	13.00 $\pm$ 4.10	19.09 $\pm$ 1.37	0.84 $\pm$ 0.02	366.29	3.00 $\pm$ 3.70	46.57 $\pm$ 2.71	0.99 $\pm$ 0.00	2.45	0.965

(b) Reconstruction Metrics on scan from the validation set.

Method	MRI				Annotation				
	MSE $\times 10^{-2}$ ↓	PSNR↑	SSIM↑	FID↓	MSE $\times 10^{-5}$ ↓	PSNR↑	SSIM↑	FID↓	DICE↑
GenT-M1	1.52 $\pm$ 0.33	18.29 $\pm$ 1.08	0.84 $\pm$ 0.02	244.15	7.02 $\pm$ 12.23	43.35 $\pm$ 8.71	0.99 $\pm$ 0.00	2.50	0.994
GenT-M2	1.39 $\pm$ 0.32	18.70 $\pm$ 1.18	0.84 $\pm$ 0.03	245.93	7.00 $\pm$ 12.22	49.42 $\pm$ 8.78	0.99 $\pm$ 0.00	2.46	0.998
GenT-M3	1.49 $\pm$ 0.31	18.38 $\pm$ 1.10	0.84 $\pm$ 0.03	254.61	9.50 $\pm$ 11.3	43.15 $\pm$ 5.66	0.99 $\pm$ 0.00	2.79	0.995
AdvT-iMRI-M1	2.07 $\pm$ 0.42	16.95 $\pm$ 1.09	0.81 $\pm$ 0.02	345.48	8.90 $\pm$ 11.7	43.35 $\pm$ 4.74	0.99 $\pm$ 0.00	2.89	0.984
AdvT-Annot-M1	1.70 $\pm$ 3.90	17.95 $\pm$ 1.31	0.82 $\pm$ 0.02	318.67	9.80 $\pm$ 11.80	42.19 $\pm$ 3.92	0.99 $\pm$ 0.00	3.68	0.962

Contrarily to the patch-wise metrics, the AdvT-Annot-M1 model did not improve the PSNR values when comparing to the GenT models. However, the slices reconstructed by the model trained with the GAN setting (AdvT-Annot-M1) are much sharper than the ones from the GenT-M2 model. Figure 5.26 shows a close-up from a slice from the validation set, focusing on the right breast, along with the corresponding annotation and reconstructions from GenT-M2 and AdvT-Annot-M1 models.

The reconstructions using the AdvT-Annot-M1 model have much more detail inside the breast than the GenT-M2 model, which has more blurry reconstructed slices. The AdvT-Annot-M1 model produces reconstructed slices where the patch borders are visible (square shapes). The usage of a larger neighborhood than the patch-level for either creating a reconstruction loss or incorporating the architecture of the model could improve the training process, both for decreasing the blurriness on GenT models or decrease the squares phenomenon on AdvT models.

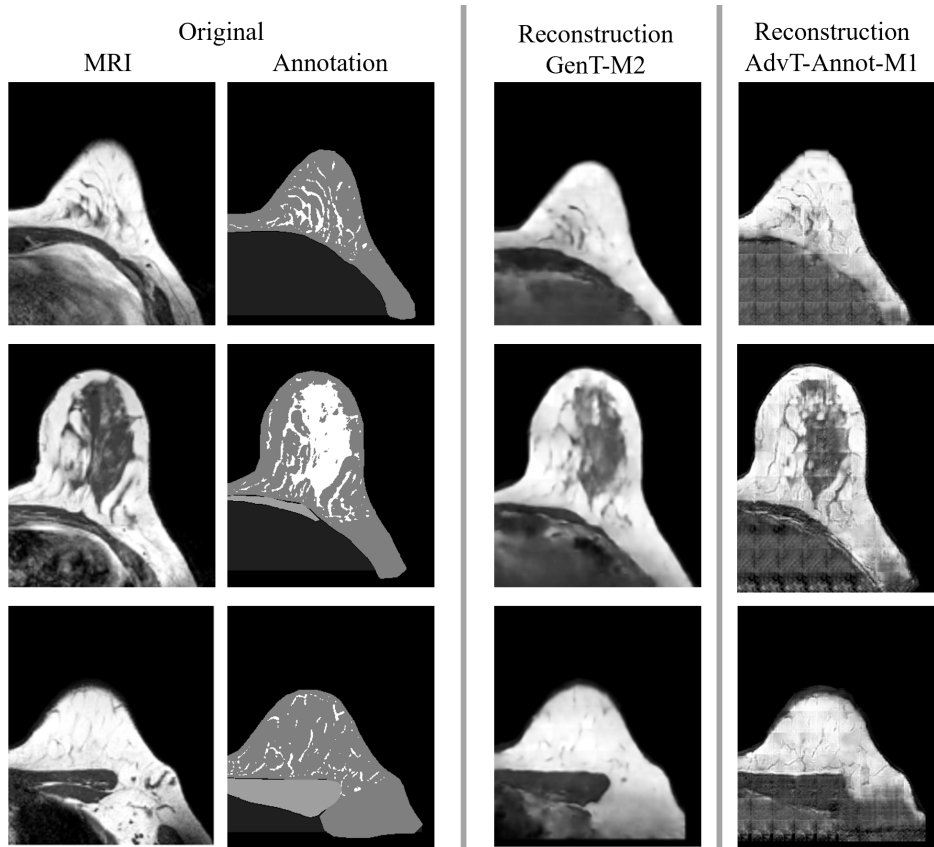


Figure 5.26: Visual performance of 3D X-Net on close-up of reconstructed slices from validation scan with GenT-M2 and AdvT-Annot-M1 models.

Although AdvT-Annot-M1 model seems to produce sharper images in the thoracic cavity, the texture produced by this model is much noisier than in the GenT-M2 model, which is expected since the structures inside this label are represented as a single class. This can negatively affect the metrics values, as they are calculated for the entire slice and not for each label. To prove or refute this hypothesis, for each label, MRI masks were created from the annotations, for both the original scans and reconstructions. Table 5.7 shows the MSE, PSNR, SSIM, and FID values between each original MRI label and the reconstruction for the GenT-M2 and AdvT-Annot-M1 models.

Most structures have similar values for both models, but in the thoracic cavity, the FID value is significantly higher on the AdvT-Annot-M1 than the GenT-M2, which is congruous with the noisier textures on this structure with the AdvT model seen in Figure 5.26. However, neither the breast contour nor the fibroglandular tissue masks have better metric results with the AdvT model than the GenT alone. Moreover, the metrics used for evaluating the image quality of the breast MRI may not evaluate the sharpness of the image within these structures.

Table 5.7: Numerical performance of reconstructions with 3D X-Net Generator and Adversarial Training for each label in validation scan: mean and standard deviation distances are shown.

Method	Thoracic Cavity					#	Latissimus Dorsi			
	#	MSE $\times 10^{-3}$ ↓	PSNR↑	SSIM↑	FID↓		MSE $\times 10^{-4}$ ↓	PSNR↑	SSIM↑	FID↓
GenT-M2	52	5.87 $\pm$ 2.35	22.77 $\pm$ 1.92	0.93 $\pm$ 0.02	148.09	38	7.43 $\pm$ 4.00	31.95 $\pm$ 2.52	0.99 $\pm$ 0.01	54.72
AdvT-Annot-M1		6.05 $\pm$ 2.11	22.49 $\pm$ 1.75	0.92 $\pm$ 0.02	223.58		7.77 $\pm$ 3.38	31.59 $\pm$ 2.29	0.99 $\pm$ 0.01	49.71
Method	Clavicle					#	Nipple			
	#	MSE $\times 10^{-4}$ ↓	PSNR↑	SSIM↑	FID↓		MSE $\times 10^{-5}$ ↓	PSNR↑	SSIM↑	FID↓
GenT-M2	4	1.05 $\pm$ 0.62	40.38 $\pm$ 2.60	0.99 $\pm$ 0.00	40.12	9	1.23 $\pm$ 1.49	53.58 $\pm$ 7.87	0.99 $\pm$ 0.00	5.12
AdvT-Annot-M1		1.74 $\pm$ 0.51	37.73 $\pm$ 1.24	0.99 $\pm$ 0.00	37.09		4.30 $\pm$ 4.00	47.28 $\pm$ 7.24	0.99 $\pm$ 0.00	6.20
Method	Sternum					#	Fibroglandular Tissue			
	#	MSE $\times 10^{-5}$ ↓	PSNR↑	SSIM↑	FID↓		MSE $\times 10^{-3}$ ↓	PSNR↑	SSIM↑	FID↓
GenT-M2	46	5.60 $\pm$ 3.60	43.41 $\pm$ 2.93	0.99 $\pm$ 0.00	14.73	52	1.21 $\pm$ 0.88	31.94 $\pm$ 6.48	0.99 $\pm$ 0.01	18.74
AdvT-Annot-M1		6.05 $\pm$ 2.11	42.55 $\pm$ 3.22	0.99 $\pm$ 0.00	17.81		1.31 $\pm$ 1.05	31.69 $\pm$ 6.25	0.99 $\pm$ 0.01	19.95
Method	Breast Contour					#				
	#	MSE $\times 10^{-3}$ ↓	PSNR↑	SSIM↑	FID↓					
GenT-M2	52	2.85 $\pm$ 0.94	25.69 $\pm$ 1.47	0.98 $\pm$ 0.01	31.75					
AdvT-Annot-M1		4.40 $\pm$ 2.50	24.25 $\pm$ 2.48	0.97 $\pm$ 0.01	36.59					

corresponds to the number of slices that have certain label.

As future development, for GenT models the learning strategies need to increase their focus on the components of higher frequencies of the MRI since these are the ones that are most affected in the reconstruction. The visual analysis of the reconstructions also supported this. One approach is to use a reconstruction loss on the original scan resolution to help the training process. In the case of AdvT models, the results from the AdvT-Annot-M1 show that using the annotation as the condition for the PatchGAN discriminator increases the performance of the model (patch-level PSNR values) and the sharpness of the breast MRI volumes reconstructions (acutance over label masks). Using a wider neighborhood for a reconstruction loss may help the model not generate such noticeable square shapes when reconstructing the slices, but it may increase training time and computational costs. The reconstruction algorithm needs to be optimized, so it computes the volume reconstructions faster, and uses over one prediction per incomplete part, by using a larger overlap.

5.5 Summary

In this chapter, different models were presented to explore the reconstruction and generation of breast MRI scans, either using the entire volumes or smaller patches. The first approach using volumes explored recurrent networks, as they use memory to learn data features from past timestamps to preview the next ones. A Recurrent Autoencoder was used, ConvLSTMs, and even ConvLSTM Autoencoders, but neither of the explored architectures could model the data with the detail needed so it has medical usability. Another volume method developed was the Auto-encoding GAN, but the available data was not large enough to enable the learning of the features, neither for reconstruction nor synthesis, leading to poor quality images with no usage.

At last, patch-based methods were explored, by dividing the MRI scans into small patches, created from the original MRI scans. Performing patch division of the data is helpful in two ways: it divides the data into smaller inputs that are easier to learn and model, and also increases the number of instances for training since with only one scan it is possible to create thousands of small volume

patches with overlap. A 3D X-Net architecture that consumes both MRI and annotation patches as inputs was developed. Pairs of incomplete MRI patches and complete annotation patches were used to train this architecture. In this way, the reconstruction of the MRI scan is also performed by using an incomplete input. This architecture reveals itself as the most promising one since the reconstruction results were the best amongst every method, allowing the reconstruction of scans with both depth continuity and reasonable image quality. Nevertheless, the high-frequency features of the images seem to not be fully captured by the GenT models. Using AdvT on the 3D X-Net with a PatchGAN discriminator showed that the reconstructions are less noisier, but the patches shape (squares) is more noticeable. As future work, reconstruction losses on the original resolution or a wider neighborhood than the patch size can be applied to the 3D X-Net architecture to test the impact on the high-frequency content of the visual results for the GenT models, and on the visual perception of the patches shapes, for the AdvT models.

Chapter 6

Breast Shape Representations

Using Deep Learning to model 3D shapes is not as straightforward as modelling images, as discovered in Chapter 4. This is especially challenging when dealing with non-Euclidean data representations, such as PCLs. Besides, most of the state-of-the-art methods use large datasets of rigid objects, such as ShapeNet [14] or ModelNet [146]. PointNet based architectures have been thriving in PCLs modelling, as it was the first architecture developed fully capable of dealing with this data without any transformation. In this study, the goal is to learn how to represent, synthesize and model non-rigid breast surface shapes, with a small dataset. To do so, in this Chapter, PointNet based models are explored, motivated by the work presented in [1], and other shape analysis networks that use PointNet as a baseline¹. The developed models are then tested for the creation of breast representations and the generation of new breast shapes. Additionally, representative models are also explored for pose transformation. These experiments will confirm or refute the hypothesis that PointNet based architectures can model small non-rigid datasets of breast shapes.

6.1 Learn Breast Surface from a Small Dataset

There is a shortage of public medical data with the particularities needed for research, as covered in Chapter 2. This makes the progress in these areas of research more difficult. Furthermore, models based on ML and DL often require large amounts of data to stabilize and achieve high performance.

Using transfer learning from larger datasets could be helpful when modelling small datasets with DL techniques [86, 122]. Datasets containing non-rigid objects are substantially smaller than the ones with rigid shapes. So, this study will compare the performance of the models with and without transfer learning from previous training cycles.

For reconstructing shapes, an AE architecture was designed and a VAE for generating shapes. Both the architecture and training strategies are simpler on the AE when comparing to the VAE. Given that, the AE was developed and stabilized first and from that, the more complex model for data synthesis was refined (VAE).

¹<https://github.com/NUAAXQ/awesome-point-cloud-analysis-2021> (last accessed on May 5, 2021) includes an up-to-date list of DL PCL analysis works.

All models were trained with PCLs of breasts in the prone position (prone PCL at the input and output). However, the representative model (AE) was tested for prone (input) to upright (output) pose transformations. Since breasts are very deformable organs, the different poses used for acquiring scans or during surgery result in significant alterations of the aesthetics of the breast. During surgery, the patient lies in the supine position. To evaluate the surgery outcome, the upright position is used as it is a more natural position for the patient. In MRI scans, the prone position is used for obtaining the data. Given that, at any point during the surgery, prone to upright transformations must happen, either directly, or from prone to supine and then upright. Taking this into consideration, prone to upright transformation is exceptionally important in breast surgery planning. These transformations are usually performed by biomechanical models, which are time-consuming and computationally expensive. Given that, this study attempts to perform prone to upright transformation using an AE model, which gives almost instant results after training.

6.1.1 Datasets

Three different PCL datasets were used in this study: 3D Breast, SHREC14 [97], and ShapeNet [14].

3D Breast is a private in-house dataset provided by INESC TEC. It holds a collection of breast shapes that were manually annotated or automatically segmented from MRI exams [131]. The PCLs used are closed and correspond to the prone position of the patient. For pose transformation tasks, a subset of shapes of prone and simulated upright positions was used. The upright shapes were obtained by biomechanical simulations [154]. The complete dataset contains 49 shapes, however, a part of these only was available later on, so only 26 shapes were used in the first experiments for stabilizing models while the 49 shapes dataset was used for cross validation methods posteriorly on the best performing models.

ShapeNet [14] is a large repository of shapes of rigid objects, such as chairs, lamps, beds. The PointNet authors and other researchers on the area have used this dataset in state-of-the-art methods [99, 1]. It contains over 55 categories, although it was only used with single-class (chair) or multi-class with 4 categories (chair, table, airplane, and lamp).

SHREC14 [97] corresponds to a human objects dataset. Humans have deformable tissues in their surface, so their shapes correspond to non-rigid objects. This dataset contains real data of scanned human participants and synthetic data of computer-generated objects. This work focuses on the real part of the dataset. SHREC14 was used as an intermediate non-rigid dataset to explore the training of the developed models before breast shapes. Although it is significantly smaller than ShapeNet, it contains considerably more objects than the breast shapes available in this study. The dataset contains data of males and females. Train and validation sets use male shapes, whereas female shapes are used as test set.

Table 6.1 provides additional information on these datasets regarding their key characteristics, while enhancing the differences in the number of objects available. The split of each dataset in the train, validation, and test (when applicable) sets is also presented in Table 6.1. Before training, all

PCLs are normalized into the unit sphere and randomly down-sampled to the desired number of points (N).

Table 6.1: Summary of the point cloud datasets used.

Dataset	Scope	Units	Type	Classes	No. of Objects	Train/Val/Test
Shapenet (single-class)	Public	Normalized	Rigid	Chair	3054	2658 / 396 / -
Shapenet (multi-class)	Public	Normalized	Rigid	Chair, Lamp, Table, Airplane	10593	9569 / 1518 / -
SHREC14	Public	Meters	Non-rigid	Male (M) Female (F)	400	140M / 60 M / 200 F
3D Breast	Private	Millimeters	Non-rigid	Prone	26	22 / 4 / -
				Prone	49	30 / 10 / 9
				Prone, Upright	22	18 / 4 / -

6.1.2 Sensitivity of Distance Metrics in PCLs

DL frameworks rely on loss functions to enable the learning of data features. For PCLs, most researchers use Chamfer or Earth Mover’s Distance Metrics as a loss function (described in Section 4.4). Although some researchers consider EMD a better choice than CD for some reconstruction tasks [87], exploring both metrics is a common practice. To understand the magnitude of the values from these metrics, an object from SHREC14 with different perturbations was used to compute CD and EMD between the original and the PCL with some disruption. The original PCL and some modifications can be found in Figure 6.1.

The experiments included the influence of the number of points in the object, rotation, translation, rotation and translation, noise (introducing random values for some points in the object), holes (Figure 6.1b), and partial objects (Figure 6.1c).

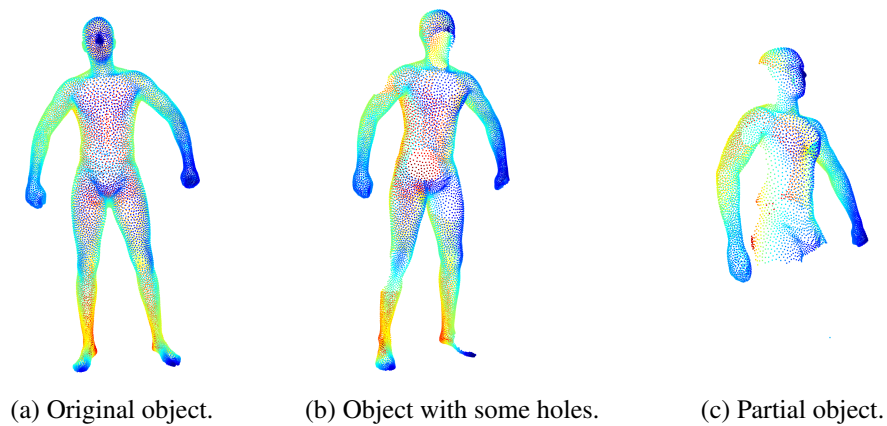


Figure 6.1: SHREC14 object used for experimenting metric values.

The result values can be found in Table 6.2. Translations have a considerable impact on both CD and EMD values (CD=8.10/EMD=25768.27), so translated objects cannot be considered similar using these distances. This is expected as they both consider the actual values of the positions from

the points in the PCL. Rotations had a higher impact on EMD than CD (CD=0.03/EMD=193.86), similar to noise perturbation values (CD=0.02/EMD=95.91). The object with holes (Figure 6.1b) did not have a great impact on the metrics (CD=0/EMD=0.98). In the partial object (Figure 6.1c), the EMD was significantly larger than CD (CD=0.05/EMD=4873.29). This might happen since CD only takes into consideration the closest point from one point set to another, while the EMD relies on the distribution of these points. With 4096 points per PCL, the noise perturbation increased the distances when comparing to 2048 points (CD=0.03/EMD=101.3 versus CD=0.02/EMD=95.91). So using higher N may lead to higher values of both CD and EMD.

Table 6.2: Sensitivity of CD and EMD in PCLs from PCL from SHREC14 dataset with modifications. Values are in meters for the Chamfer Distance.

PCL 1 (N)	PCL 2 (N)	CD	CD	EMD
		PCL 1 $\rightarrow$ PCL 2	PCL 2 $\rightarrow$ PCL 1	
Original (2048)	Original (2048)	0	0	0.03
Original (2048)	Original (4096)	0	0	0.03
Original (2048)	Rotated (2048)	0.03 ± 0.03	0.03 ± 0.03	193.86
Original (2048)	Translated (2048)	8.10 ± 1.41	9.02 ± 1.45	25768.27
Original (2048)	Rotated and Translated (2048)	8.06 ± 1.30	9.04 ± 1.42	25679.23
Original (2048)	Original w/300 points noise (2048)	0	0.02 ± 0.10	95.91
Original (4096)	Original w/300 points noise (4096)	0	0.03 ± 0.10	101.3
Original (2048)	With holes	0	0	0.98
Original (2048)	Partial	0.05 ± 0.10	0	4873.29

6.1.3 Evaluation

The reconstructions from the trained models were evaluated in terms of Euclidean and Hausdorff distances (see Section 4.4). These between the original (*Ori*) and reconstructed (*Rec*) PCLs, in the original units present in Table 6.1 (millimeters for 3D Breast). To evaluate the internal aspects of the model, linear interpolation between latent vectors was performed (procedure described in Section 3.1.3.2) and the decoded representations were visually inspected.

Reconstruction metrics were also computed for the VAE model. To evaluate the capacity for synthesizing data, quantitative measures were calculated with the generated data. More specifically, the Jensen-Shannon Divergence (JSD), Coverage (COV), and Minimum Matching Distance (MMD) [1]. To compute these, each generator delivers a synthetic set of shapes three times bigger than the number of shapes in the comparative set (validation or test) and the metrics are computed between these. This process is repeated three times and the average values are registered. The latent space of a VAE model was also evaluated by looking into the values of some latent vectors and performing arithmetic operations and linear interpolation between those, while visually evaluating the decoded representations.

6.2 Breast Representations with PointNet based Autoencoder

For learning breast shape representations (in prone to prone positions), an autoencoder was constructed with an encoder based on PointNet vanilla, previously explained in Chapter 4. Figure 6.2 shows the diagram of this network.

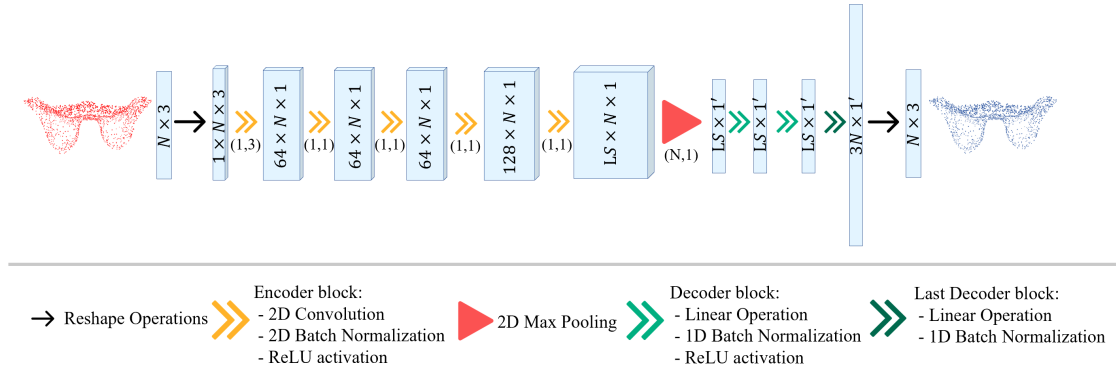


Figure 6.2: PointNet based autoencoder diagram. N is the number of points of the PCL and LS is the chosen latent space size. Kernel sizes inside parentheses.

The network receives a raw PCL data with N points in a three-dimensional vector (coordinates), and goes through a 5 block encoder, encoding the data into a latent space vector with a chosen size LS . Each block contains a 2D convolution operation, 2D batch normalization and ReLU activation. The filter sizes used were 64, 64, 128, and LS , which is converted to the global vector through a 2D max pooling operation with a kernel of size $(N,1)$. Then the decoder is composed of fully-connected blocks, with a linear operation, batch normalization and ReLU activation (except for the last layer). The output sizes of each layer were LS , LS and $N \times 3$, respectively. The result was then reshaped to the original three dimensions.

Implementation details In Table 6.3 details the experiments performed with the autoencoder model. Most scenarios used the literature references for the latent size ($LS = 1024$) and for the loss function - CD, which is the most common since it is faster to calculate. Firstly, two models were trained using the ShapeNet dataset: a single-class (Chair) and a multi-class (Chair, Table, Airplane, Lamp). The single-class models are interesting to explore because the 3D Breast dataset is itself a single-class set. Theoretically, models fitted on multi-class must learn discriminative features between such classes, whilst models trained only in one class can optimize features that distinguish intra-class objects. Hence, single-class models may be a better alternative for finetuning on a small dataset of non-rigid breast shapes. To test this, the SHREC14 dataset was trained directly and from transfer learning from the models pre-trained ShapeNet. Based on the results from SHREC14 experiments, the train of the 3D Breast dataset was also conditioned with direct train or transfer learning from single-class ShapeNet (TL1). Additionally, another experiment included the transfer

learning from models pre-trained both on ShapeNet and then SHREC14 (TL2), to determine if this added tuning step would be profitable.

All training experiments used Adam’s optimizer for training. Also, the initial learning rate was 0.001 with a decay rate of 0.7, varying only in the step of decay. For ShapeNet models, CD loss was minimized throughout 500 epochs and with a step of decay of 80 epochs, and the batch size used was 32. The non-rigid dataset SHREC14 was trained for 1000 epochs, with step decay of 100 epochs, and the batch size was decreased to 8. CD was also used here as the reconstruction loss. Finally, for the 3D Breast, a variety of models was tested since this is the data of interest of this work. The models were trained during 10000 epochs with a step of decay of 1000 epochs and the batch size used was decreased to 4. The experiments differ in the type of training (DT - Direct Train or TL - Transfer Learning), but also on the type of loss function (CD or EMD) and size of the latent space ($LS = 256$ or $LS = 1024$), enabling the comparison of different loss functions in the performance of the model and also the size of the latent space needed to correctly represent the data.

Table 6.3: Description of different experiments performed with autoencoder model.

Type of Data	Dataset	No. of Objects	Train	LS	Loss
Rigid	ShapeNet (single-class)	3054	DT	1024	CD
	ShapeNet (multi-class)	10593			
Non-Rigid	SHREC14	400	DT	1024	CD
			TL ShapeNet (single-class)		
			TL ShapeNet (multi-class)		
Non-Rigid	3D Breast	26	DT	256	EMD
			DT	1024	CD
			TL ShapeNet (single-class)		CD
			TL ShapeNet (single-class) and SHREC14		

Reconstruction Metrics For all datasets, the assessment of the results was done by the analysis of the learning curves and also by the computation of distance metrics between the original (*Ori*) PCL and respective reconstruction (*Rec*). For ShapeNet and SHREC14 experiments, the results are available in Table B.1. In the case of SHREC14, it is important to mention that the model trained with transfer learning from ShapeNet single-class outperformed the one directly trained in this dataset and the one with transfer learning from ShapeNet multi-class (Table B.1a). That conditioned the study of 3D Breast dataset only on the TL1 setting.

For 3D Breast dataset, results are shown in Table 6.4. Overall, the autoencoder proposed can reconstruct data of rigid and non-rigid 3D objects, as the presented distance values are relatively low, indicating the preservation of the original shapes throughout the model.

Table 6.4: Numerical performance of PointNet Autoencoder models: mean and standard deviation distances are shown. Distances in millimeters.

(a) Training Set (22 patients).

Train	Method		Loss	Euclidean↓		Hausdorff↓	
	LS	Loss		Rec → Ori	Ori → Rec	Rec → Ori	Ori → Rec
DT	256	EMD	48.5	8.58 ± 3.48	5.96 ± 0.98	108.02 ± 100.52	19.54 ± 4.46
		CD	0.19	5.01 ± 0.56	4.58 ± 0.41	18.88 ± 3.46	14.90 ± 2.05
DT	1024	CD	0.16	5.40 ± 1.20	4.78 ± 0.76	22.19 ± 12.05	15.21 ± 3.41
TL1			0.15	4.55 ± 0.38	4.07 ± 0.31	17.07 ± 2.66	12.82 ± 2.20
TL2			0.15	19.97 ± 25.31	6.78 ± 1.91	270.33 ± 295.00	22.95 ± 6.53

(b) Validation Set (4 patients).

Train	Method		Loss	Euclidean↓		Hausdorff↓	
	LS	Loss		Rec → Ori	Ori → Rec	Rec → Ori	Ori → Rec
DT	256	EMD	132	13.88 ± 9.64	7.78 ± 1.18	85.41 ± 73.78	26.34 ± 6.48
		CD	0.41	6.87 ± 1.42	6.75 ± 0.92	24.18 ± 3.70	24.02 ± 4.33
DT	1024	CD	0.61	9.49 ± 2.48	8.09 ± 0.75	42.59 ± 23.41	25.88 ± 2.70
TL1			0.44	7.48 ± 2.30	7.19 ± 1.56	25.99 ± 7.89	25.23 ± 7.22
TL2			0.39	7.18 ± 1.42	6.67 ± 0.89	24.33 ± 3.87	24.20 ± 8.86

TL1: transfer learning from model previously trained on single-class ShapeNet.

TL2: transfer learning from model previously trained on single-class ShapeNet and SHREC14.

The model pre-trained on the single-class ShapeNet (AE-CD ($LS=1024$) TL1) holds the best results (loss and distances) on the training set. However, the AE-CD ($LS=256$) DT model, trained directly on the 3D Breast shapes, performs better on the validation set. Besides, the difference between train and validation loss is smaller in the DT model, implying better generalization ability.

The AE-CD ($LS=256$) DT models also perform better than the one trained in the same conditions, but with higher latent space ($LS=1024$). Although most literature methods use $LS=1024$, those also use large multi-class datasets. Consequently, the model needs to learn how to capture both inter-class variability and intra-class differences. In the case of breast shapes, the differences between patients are more subtle than the differences amongst ShapeNet chairs, for instance. On top of that, 3D Breast dataset only uses 22 shapes for training. Taking all of these into consideration, using a smaller LS can be useful for capturing breast representations.

The use of EMD as a loss function did not show any benefit to the performance of the model. All distances are higher when comparing the metrics between the AE-CD and AE-EMD (both with $LS=256$). This difference is significantly higher in the Hausdorff distances, as many points in the reconstructed shapes are disperse, justifying the higher value of this metric. Using transfer learning from both ShapeNet and SHREC14 (TL2) did not show any gain in performance, when comparing to AE-CD ($LS=256$) DT or AE-CD ($LS=1024$) TL1.

Visual Results In Figure 6.3, reconstruction of breast shapes from the validation set are shown in comparison with the original ones. The results here presented correspond to the AE-CD ($LS=256$) DT and AE-CD ($LS=1024$) TL1 models. Complementary visual results can be found in Figures A.3-A.5. These visual results are congruent with the reconstruction metrics: both models can reconstruct breast shapes. The TL1 reconstructs a noisier PCL in the larger breast but its reconstruction of the small breast is "cleaner". These results also support the reconstruction metrics present in Table 6.4: the shape of the breast is maintained, leading to small Euclidean distances, but some points appear scattered from the object, increasing Hausdorff distances.

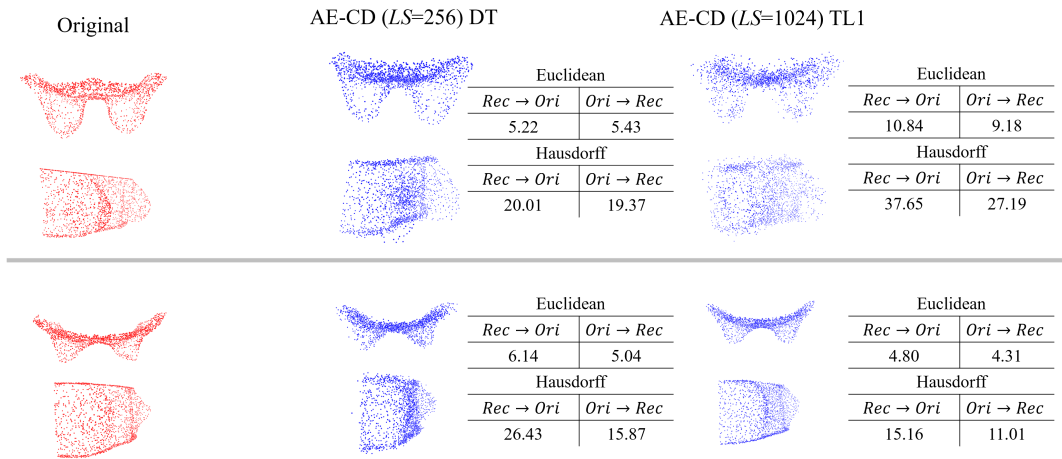


Figure 6.3: Reconstructions of unseen shapes from 3D Breast in the validation dataset for the best models. Ground truth objects are shown in red, reconstructions are shown in blue. Large breast on top, small breast on bottom. Euclidean and Hausdorff distances shown in millimeters for each shape.

Latent Space Continuity The latent space continuity was evaluated by linear interpolation of latent vectors. The corresponding reconstructions on the Shapenet single-class and multi-class models are present in Figures A.6a and A.6b, respectively. In 3D Breast, this operation was done from a smaller-sized breast to a bigger one (Figure 6.4).

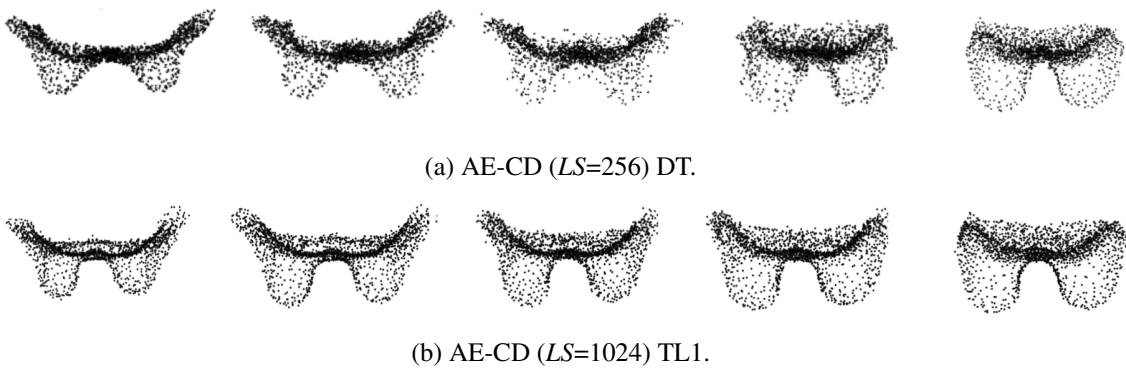


Figure 6.4: Linear interpolation of AE-CD latent spaces with breast shapes. Leftmost and rightmost shapes represent reconstructions from the dataset of smaller and larger breasts, respectively.

From left to right, it is possible to see the smooth transition from a more "pointed" breast to a more rounded one. Also, there is an increase of volume along with each decoded representation. This allows us to create new synthetic PCLs with different shapes and sizes from the originals, even though the AE is not a generative model. However, with the AE-CD ($LS=256$) DT (Figure 6.4a), some intermediary shapes are noisier. It is not possible to ensure the exact reason for that, but it may be because of the noisy reconstruction of the small breasts mentioned before.

Shape Completion The best overall models trained in 3D Breast shapes were used to test the robustness against objects with some perturbations. An object from the validation set was selected and passed through these models to visually evaluate the reconstructions. More specifically, transformations included removal of small parts (holes), cropping the most anterior part, or removal of one breast. The reconstructions of these objects can be found in Figure 6.5 for the model trained directly in the breast data AE-CD ($LS=256$) DT and the one re-trained from ShapeNet single-class AE-CD ($LS=1024$) TL1.

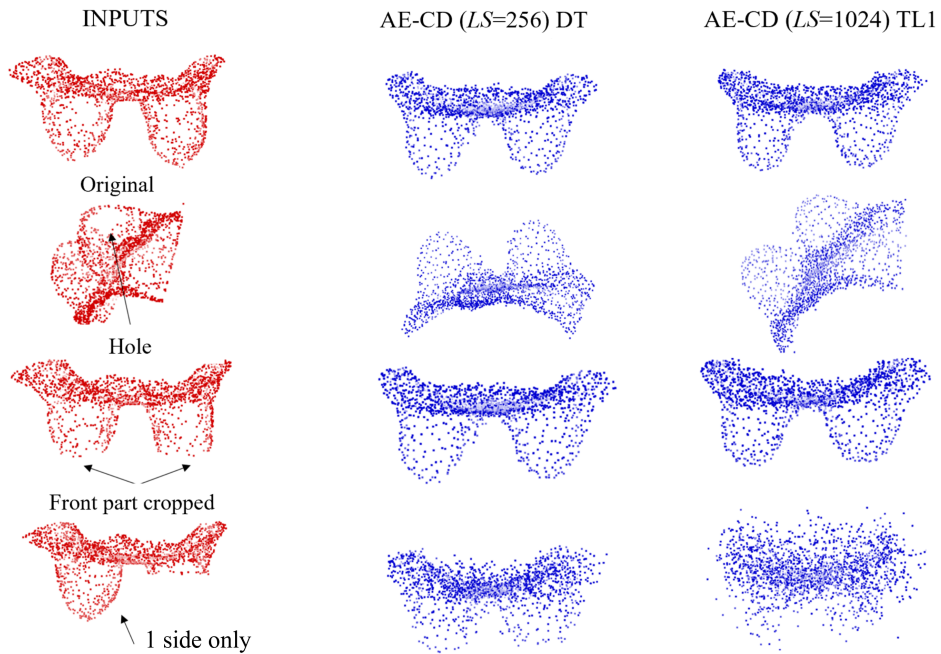


Figure 6.5: Shape completion: robustness of AE model to noise, holes and missing data. Multiple views of the breast shapes are provided; input objects are shown in red, reconstructions are shown in blue.

Both models were able to reconstruct complete breast shapes despite the small holes or cropped anterior parts. However, for the object with only one breast, the DT model reconstructed a less noisier shape than the TL1 model. It is important to remember that these two models also differ on the LS (256 and 1024). As previously explained, the smaller latent vector may be able to capture the differences between breast shapes better than the bigger one. Using transfer learning on a model with a smaller latent size could be beneficial, but the training process of the bigger dataset (ShapeNet, in this case) may be harmed.

6.3 Modelling Breast Pose Transformation

To perform prone to upright pose transformation, a set of 22 patients with data in prone and simulated upright positions was used. This set of data is used to train an AE, with the architecture presented in Figure 6.2. This way, the model was trained to capture features in the prone position (input) during the encoding that are meaningful to decode an upright pose (output).

Implementation Details The experiments on pose transformation with the AE model were tested with direct training (DT) or with transfer learning after ShapeNet single-class (TL1). Furthermore, the model was tested in other situations: transfer learning after 3D Breast prone to prone (TLPP), and after ShapeNet single-class and 3D Breast prone to prone (TL3). The DT and TLPP models were trained with $LS=256$, since the results from prone to prone demonstrated that this latent size was favorable. On the other hand, TL1 and TL3 were trained with $LS=1024$. The other implementation details were like the autoencoder models with 3D Breast shapes but without any learning rate decay. CD was used as the loss function.

Reconstruction Metrics Table 6.5 shows the losses and reconstruction metrics for the experiments on pose transformation for both train and validation sets. The best models in the training set were TLPP and TL1. Nevertheless, TL1 outperforms TLPP on the validation set. TLPP may be overfitting on the training set, which was also validated by the learning curves (data not shown). Concerning the validation set, TL1 and DT have very similar results, but the DT model shows smaller differences between train and validation sets, suggesting a higher capacity for generalization. This is also congruent with the previous experiments: while transfer learning may seem to help decreasing loss and reconstruction metrics in the training set, the generalization of the model is worse than with DT. The experiment made on TL3 did not show any benefit for the reconstruction metrics. Nevertheless, all the models trained have reasonable performance on prone to upright transformations of the breast.

Visual Results Overall, the models could transform the prone shapes into upright, preserving the general shape of the breast. Figure 6.6 shows the results on pose transformation for the AE-CD ($LS=256$) DT model, with two breast shapes from the validation set. The reconstruction of the smaller breast shows more noise, which comes along with the previous results from the AE in prone to prone. This problem may be solved with a more careful down-sampling of points instead of using a random set of N points. This operation often results in higher point density around the torso than the breast itself, which can prejudice the learning process. Also, using the same points for down-sampling on both prone and upright might improve results.

Table 6.5: Numerical performance of prone to upright pose transformation with PointNet Autoencoder models: mean and standard deviation distances are shown. Distances in millimeters.

(a) Training Set (18 patients).

Train	Method		Loss	Euclidean↓		Hausdorff↓	
	LS	Loss		Rec → Ori	Ori → Rec	Rec → Ori	Ori → Rec
DT	256	CD	0.18	5.32 ± 0.69	5.34 ± 0.68	18.38 ± 4.48	15.54 ± 3.27
TLPP			0.17	4.97 ± 0.65	4.67 ± 0.74	16.46 ± 3.64	12.87 ± 2.63
TL1	1024	CD	0.14	4.92 ± 0.65	4.62 ± 0.56	16.65 ± 3.80	12.31 ± 2.97
TL3			0.13	5.23 ± 0.99	4.74 ± 0.78	18.45 ± 9.00	12.90 ± 2.84

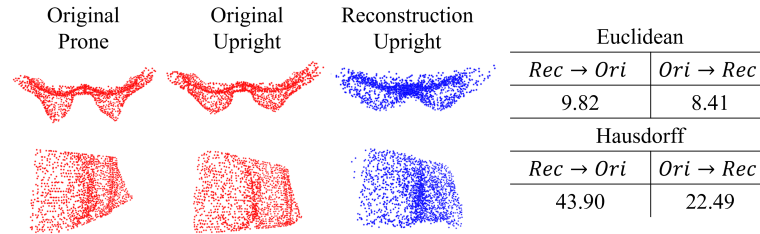
(b) Validation Set (4 patients).

Train	Method		Loss	Euclidean↓		Hausdorff↓	
	LS	Loss		Rec → Ori	Ori → Rec	Rec → Ori	Ori → Rec
DT	256	CD	0.44	7.56 ± 2.00	7.04 ± 1.29	30.10 ± 11.68	21.97 ± 5.93
TLPP			0.48	8.04 ± 2.97	7.13 ± 1.68	31.08 ± 19.32	23.00 ± 8.34
TL1	1024	CD	0.48	7.69 ± 1.74	7.05 ± 1.42	29.22 ± 7.64	25.70 ± 8.80
TL3			0.58	8.69 ± 2.31	7.45 ± 1.36	36.61 ± 15.64	25.92 ± 8.48

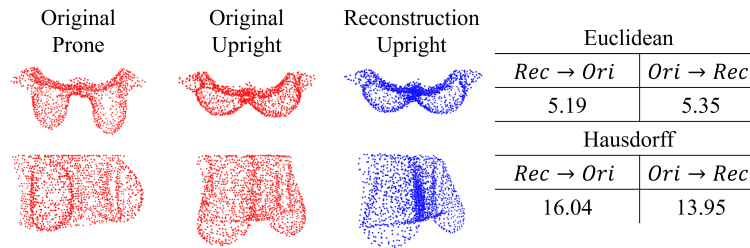
TLPP: transfer learning from model previously trained on 3D Breast prone to prone.

TL1: transfer learning from model previously trained on single-class ShapeNet.

TL3: transfer learning from model previously trained on single-class and 3D Breast prone to prone.



(a) Smaller breast.



(b) Larger breast.

Figure 6.6: Reconstructions of unseen shapes from 3D Breast for pose transformation in the validation dataset, using AE-CD (LS=256) DT model. Ground truth objects are shown in red, reconstructions are shown in blue. Euclidean and Hausdorff distances shown in millimeters for each shape.

6.4 Breast Synthesis with PointNet based Variational Autoencoder

Taking advantage of the previously developed autoencoder, for the generation of shapes the model was upgraded into a β -VAE model, maintaining the structures of both the encoder and decoder parts of the AE. The latter aims to learn a probabilistic latent space distribution while reaching a disentangled space, as explained in Chapter 3. The model uses an encoder that is permutation invariant and a decoder of fully connected layers, as represented in Figure 6.7.

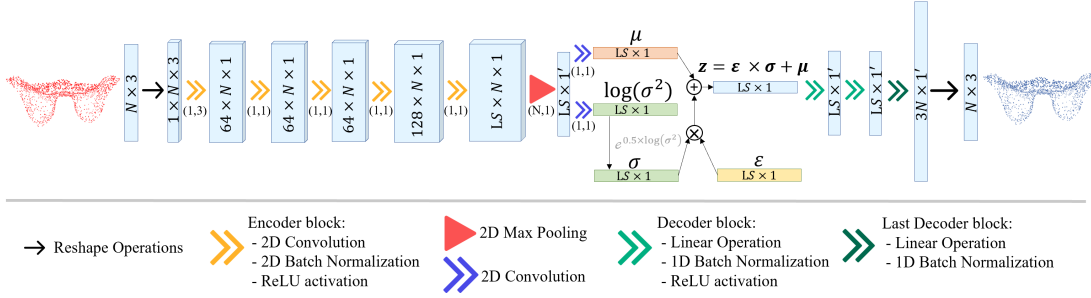


Figure 6.7: PointNet based Variational Autoencoder diagram. N is the number of points of the PCL and LS is the chosen latent space size. Kernel sizes inside parentheses.

After computing the global feature vector, mean (μ) and the variance (σ^2) layers are computed using a 2D convolution from the global feature vector for each of these layers, resulting in layers of size LS after removing dimensions of size 1. The reparameterization samples the latent vector z from the latent distribution. This is done by computing $z = \mu + \epsilon \times \sigma$, where ϵ is a random vector of size LS . The latent vector is then decoded to a $N \times 3$ matrix corresponding to the point coordinates.

To train this VAE model, a composed loss function was used: a reconstruction loss for comparing the input PCL (x) with the corresponding reconstruction ($\hat{x}$) - $\mathcal{L}(x, \hat{x})$ - and a regularization term to enforce a normal Gaussian distribution, $N(0,1)$ - the Kullback-Leibler Divergence ($\mathcal{D}_{KL}$), calculated through Equation 6.1.

$$\mathcal{D}_{KL} = \frac{\sum_{i=1}^I \left(\frac{1}{2} \sum_{ls=1}^{LS} \left(1 + \log(\sigma_{ls}^2) - \mu_{ls}^2 - \sigma_{ls}^2 \right) \right)}{I} \quad (6.1)$$

where I corresponds to the batch size. Additionally, and hyper-parameter β controls the weight of $\mathcal{D}_{KL}$ in the loss function, optimizing the balance between the reconstruction and latent space organization abilities. Using the β , the VAE model is transformed into a β -VAE, which is the same as VAE when $\beta = 1$. Given this, the final loss function is computed through:

$$Loss = \mathcal{L}(x, \hat{x}) + \beta \times \mathcal{D}_{KL}(N(\mu, \sigma), N(0, 1)) \quad (6.2)$$

Implementation details The β -VAE models were trained in ShapeNet single-class, SHREC14, and 3D Breast datasets, according to the details present on Table 6.6. The ShapeNet multi-class was not trained since the reconstruction metrics for this model proven that this setting was not

beneficial in the AE experiments (see Table B.1a). Adam’s optimizer was used in all experiments with a constant learning rate of 1E-4. In terms of the number of points, N , 2048 or 4096 points were used. The higher number of points (4096) was used in hopes that it could be helpful for the upsampling of the PCL, as it may help with the overall density of points across the shape. At first, $LS=1024$ with CD loss was used, but it was only possible to converge the models using $LS=256$ and EMD. This potentiated the experiments with these parameters on the AE posteriorly.

For ShapeNet models, the training was conducted during 1000 epochs and a batch size of 32 PCLs. The β parameters was adjusted from $\{1, 2, 4, 6\}$ and optimized to 4 and 6 for 2048 and 4096 points, respectively. The same β values were used in SHREC14 models, but the number of epochs was increased to 50000.

Training in the 3D Breast shapes dataset was not as straightforward, given the lower number of objects used in training (22). To overcome this, data augmentation was performed on the input by adding jitter to the position of the points, with Gaussian noise with mean 0 and standard deviation 0.02 and randomly rotating the shapes in the xy plane. Experiments using only one of these operations demonstrated that the performance of the model improved when using them together. Without data augmentation, the outputs from the network contained holes in the most anterior part of the breast. Although data augmentation was not necessary for training the AE models, an additional experiment with this operation was done on the AE but did not show any improvement in the results (reconstruction metrics in Table B.2).

Table 6.6: Description of different experiments performed with Variational Autoencoder model.

Type of Data	Dataset	No. of Objects	Train	LS	Loss
Rigid	ShapeNet (single-class)	3054	DT	256	CD EMD
Non-Rigid	SHREC14	400	TL ShapeNet (single-class)	256	EMD
Non-Rigid	3D Breast (data augmentation)	26	DT TL ShapeNet (single-class) TL ShapeNet (single-class) and SHREC14	256	EMD

Equally to the autoencoder, the β -VAE in 3D Breast dataset was trained directly in the dataset (DT), with transfer learning from ShapeNet single-class (TL1) and ShapeNet single-class and SHREC14 (TL2). While for ShapeNet and SHREC14 the optimal β values were of 4 and 6 (for 2048 and 4096 points, respectively), in this dataset, this hyperparameter needed to be fine-tuned (TL experiments the β used on the previous dataset was not changed). Remember that the β parameter is used to promote a higher regularization of the latent space and $\mathcal{D}_{KL}$ is also dependent on the batch size. In 3D Breast, the higher batch size possible is 22 (as there are only 22 shapes in training), which is smaller than the 32 used for ShapeNet and SHREC14. Consequently, the model needed less regularization. Given that, the β was optimized with additional 0.1 and 0.5 values. Similarly with the other datasets, when using a larger number of points, the EMD loss increases, increasing the necessity for a higher regularization weight. With higher values of β , regularization was such that the output from the model was always the same independently of the input given. The training was done through 30000 epochs for these experiments.

Reconstruction Metrics Reconstruction metrics and loss values in the last epoch are presented in Table 6.7 for 3D Breast dataset. Complementary results for the ShapeNet and SHREC14 datasets can be consulted on Table B.3.

Table 6.7: Numerical performance of PointNet Variational Autoencoder models: mean and standard deviation distances are shown. Distances in millimeters.

(a) Training Set (22 patients).								
Method			Loss		Euclidean↓		Hausdorff↓	
Train	N	β	EMD	KLD	$Rec \rightarrow Ori$	$Ori \rightarrow Rec$	$Rec \rightarrow Ori$	$Ori \rightarrow Rec$
DT	2048	0.5	75	4.5	7.94 ± 1.22	8.33 ± 1.55	25.40 ± 4.76	15.21 ± 3.41
TL1			54	5	6.22 ± 0.76	5.88 ± 0.64	20.85 ± 21.74	12.82 ± 2.20
TL2			50	4.5	6.32 ± 1.05	5.70 ± 0.98	24.08 ± 8.75	14.33 ± 2.60
DT	4096	1	132	5	6.41 ± 1.31	6.70 ± 1.46	24.55 ± 6.68	29.93 ± 6.24
TL1			60	4.3	6.85 ± 1.21	6.42 ± 1.69	31.11 ± 12.53	24.96 ± 8.59
(b) Validation Set (4 patients).								
Method			Loss		Euclidean↓		Hausdorff↓	
Train	N	β	EMD	KLD	$Rec \rightarrow Ori$	$Ori \rightarrow Rec$	$Rec \rightarrow Ori$	$Ori \rightarrow Rec$
DT	2048	0.5	90	4.1	8.13 ± 1.43	9.17 ± 1.23	35.05 ± 12.04	33.02 ± 4.00
TL1			100	4.2	8.45 ± 1.28	8.81 ± 1.06	26.99 ± 2.14	30.17 ± 3.71
TL2			112	3.9	9.97 ± 1.29	9.75 ± 1.85	38.83 ± 6.68	29.56 ± 5.80
DT	4096	1	175	4.7	7.09 ± 0.94	7.51 ± 0.74	29.88 ± 7.07	32.78 ± 6.13
TL1			196	4.5	9.44 ± 1.05	8.22 ± 1.07	36.24 ± 11.36	31.71 ± 2.32

TL1: transfer learning from model previously trained on single-class ShapeNet.

TL2: transfer learning from model previously trained on single-class ShapeNet and SHREC14.

The best overall β -VAE model in both train and validation sets regarding these reconstruction metrics was the β -VAE-EMD ($LS=256$) TL1, the model pre-trained from ShapeNet single-class with $N=2048$ points. Most of the reconstruction metrics are smaller in the validation set when using this model. However, the DT model presents results that are not much different from the TL1 and with smaller differences between train and validation sets.

Concerning the models trained in 4096 points, the DT model marginally outperformed TL1. Given that the TL2 model with 2048 points performed worse than DT and TL1 in the validation set, this setting was not tested for 4096 points.

Overall, these reconstruction metrics show results that are slightly worse than the ones presented for the AE models with CD loss in Table 6.4. This is expected since the VAE needs to learn the distribution of the data besides learning how to reconstruct input shapes.

Quantitative Evaluation Although evaluating the reconstruction of shapes gives us an idea of the capability of the model for creating PCLs, the goal of the β -VAE is to be able to generate shapes from random vectors. Considering that, the models trained in 3D Breast datasets were tested using JSD, MMD, and COV metrics between newly synthesized shapes and reconstructions from the validation set. To do so, each generator delivers a synthetic set of shapes (12) three times bigger than the number of shapes in the validation set (4) and the metrics are computed between these.

This process is repeated three times and the average values are registered. In addition, the results for a sampling-based memorization baseline are also presented for comparison: this model memorizes a random subset of the original training set (with the same size as the generated sets). The MMD and COV were computed using CD and EMD distance metrics. Small JSD and MMD and large COV indicates that unseen shapes from the validation set can be generated by the model. Results of these quantitative metrics are present in Table 6.8.

Table 6.8: Quantitative results of β -VAE-EMD ($LS=256$) on the 3D Breast validation set (4 instances). COV is shown in percentage. Metrics computed in the normalized unit sphere.

Method						
Train	N	JSD $\downarrow$	MMD-CD $\downarrow$	MMD-EMD $\downarrow$	COV-CD $\uparrow$	COV-EMD $\uparrow$
DT	2048	0.1590	0.0020	19.77	50.00	58.33
TL1		0.1389	0.0024	35.52	66.67	66.67
Baseline		0.2209	0.0058	88.83	75.00	75.00
DT	4096	0.1380	0.0015	48.73	41.67	41.67
TL1		0.1182	0.0025	85.20	83.33	91.67
Baseline		0.1874	0.0048	149.51	75.00	75.00

Synthetic shapes are $3\times$ the population of the validation.

Average results from the three repetitions.

The β -VAE-EMD ($LS=256$) TL1 models have lower JSD and higher COV values between synthesized and validation shapes, irrespective of the distance metric used in COV. Although the lower MMD values are obtained from the DT models, the other results indicate that using transfer learning may be beneficial for learning organized latent distributions. Concerning the influence of the number of points, using $N=4096$ seems to be beneficial in all quantitative metrics presented besides MMD-EMD, which was worse than when using only 2048 points.

As seen in Section 6.1 and also by the loss values in Table 6.7, the EMD values tend to increase with a higher number of points, as the effort to transform a set of points into another increases. This is also supported by the fact that besides models trained with 4096 points had higher EMD loss values, their Euclidean distances were smaller. The same phenomenon occurs in the baseline model.

Visual Results The visual analysis of the synthesized shapes can also be used to evaluate the β -VAE models. In Figure 6.8, some generated breast shapes from the β -VAE-EMD ($LS=256$) TL1 models, with 2048 or 4096 points, are presented.

The generated shapes are indistinguishable from real ones. Furthermore, the decoded PCLs are "clean" and do not contain points out of the breast shape, indicating that the object is being well represented. Hence, the VAE model proposed can generate plausible shapes so it can integrate medical tools to compare data between patients' data, for instance. Figure A.7 holds visual results for the models trained in ShapeNet and SHREC14.

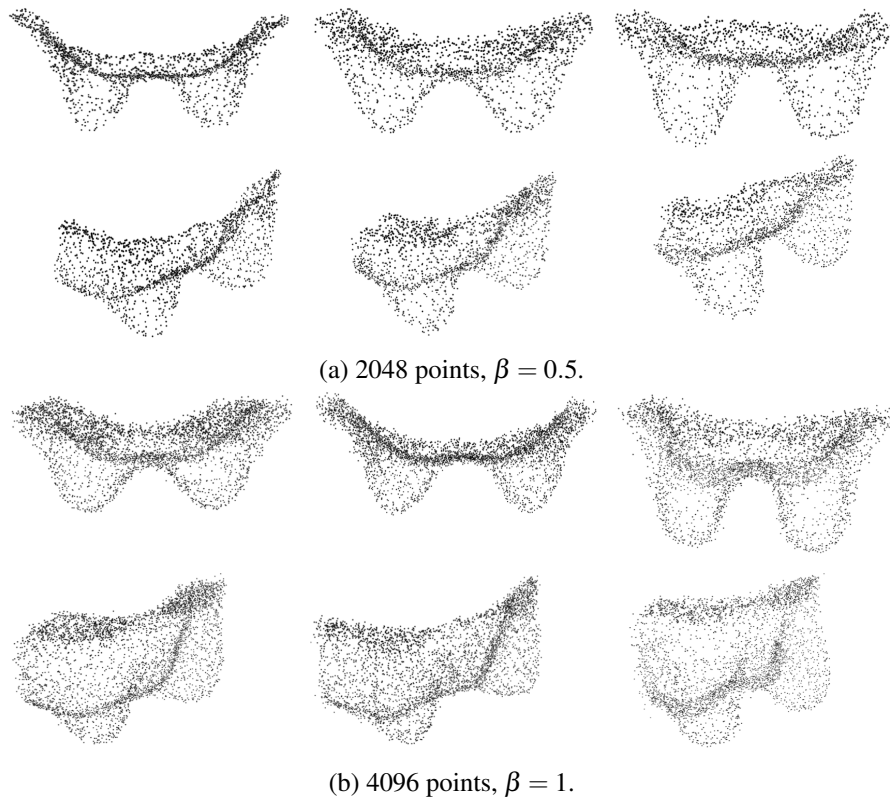


Figure 6.8: Synthesis of breast shapes with β -VAE-EMD ($LS=256$) TL1.

Visualization of the latent space To evaluate the latent space of β -VAE models, latent vectors created by β -VAE-EMD ($LS=256$) TL1 $N=2048$ model from different patients (ID17, ID22, and ID10) were used. ID17 and ID22 have larger breasts than ID10, while ID22 also has a bigger separation between the breasts. The plot of these latent vectors is present in Figure 6.9a.

In this plot, there are three distinct peaks/valleys: features 55, 77, and 101. Using the latent vector from ID17, the values of these features were altered and decoded to the corresponding shapes (Figures 6.9b, 6.9c, and 6.9d). Positive values of feature 55 seem to increase the compression of the breasts medially and laterally, while negative values result in an uncompressed shape. Comparing the latent vectors on feature 55, larger breasts have positive values, while the smaller breasts have negative ones, which is expected since larger breasts tend to be more compressed.

Concerning feature 77, it has negative values for ID17 (larger breast), but positive values for both ID22 and ID10. Looking at the three breast shapes, ID22 and ID10 have larger torso depth than ID10, so this feature may relate to that. Figure 6.9c supports this theory, increased feature 77 values elongates the breasts and increases torso depth.

At last, feature 101 seems to influence the separation of the breasts and the torso width. Negative values of this feature lead to deformations of the breast shapes to a more lateral space, with consequent medial separation of the organs. Besides, the values of this feature for ID10 are positive (less separated breasts) and for ID22 are negative (highly separated breasts).

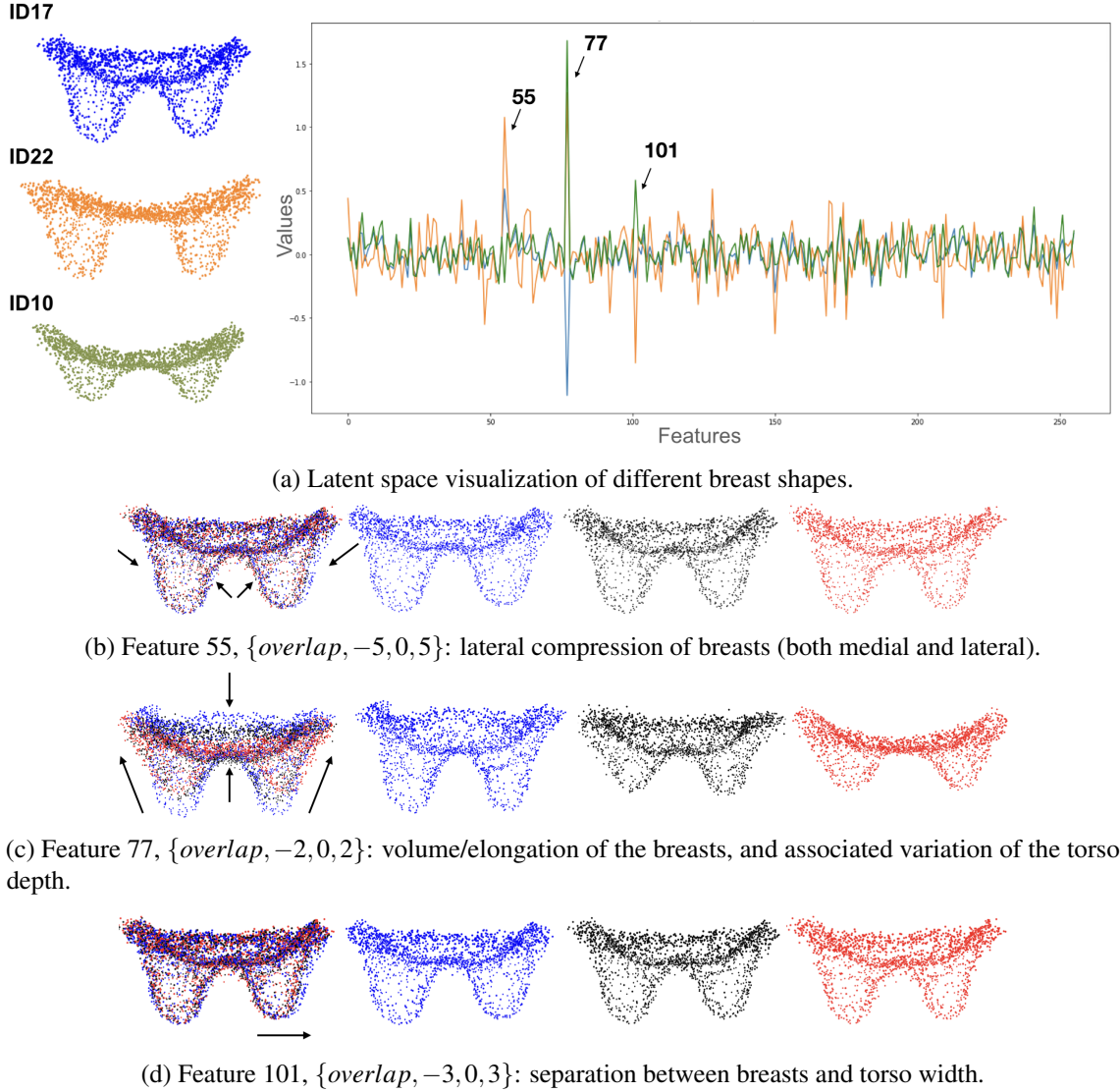


Figure 6.9: Visualization of latent space experiments on β -VAE-EMD ($LS=256$) TL1.

Another experiment was done by creating a latent vector from scratch: all features are set to zero (approximate mean value of encoding vectors) and the values of features 55, 77, and 101 were manipulated from negative to positive values. The results of some decoded representations are present in Figure 6.10. Given that the decoded representations are meaningful, the knowledge of the features can be used to create shapes according to desired purposes.



Figure 6.10: Synthetic breast shapes created from discriminative features from β -VAE-EMD ($LS=256$) TL1. From left to right, the features 55, 77, 101 are equal to $\{-5, -2, -3\}$, $\{0, 0, 0\}$ and $\{5, 2, 3\}$, while other features are set to zero.

Linear interpolation between the latent vectors of ID10 and ID17, and ID10 and ID22 was performed. Results are shown in Figure 6.11. In both cases, a smooth transition from a smaller breast to a bigger one is visible, while increasing the space between breasts in the case of ID10 to ID 22.

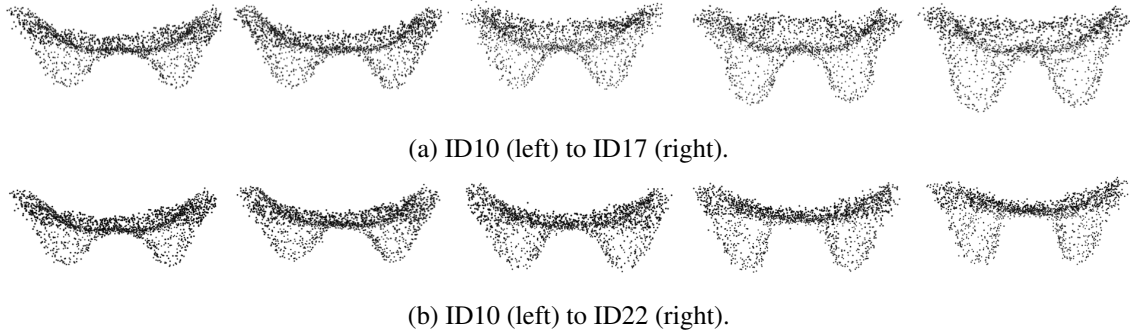


Figure 6.11: Linear interpolation between latent vectors in β -VAE-EMD ($LS=256$) TL1 model. Leftmost and rightmost shapes represent reconstructions from the dataset of smaller (ID10) and larger (ID17 and ID22) breasts, respectively.

Arithmetic operations between latent space vectors are also a form of creating new shapes and evaluate the representation of the data from the model. To prove so, additions and multiplications were performed over the same shapes (ID10, ID17, and ID22), as showed in Figure 6.12. While adding two latent vectors results in a shape with a different overall appearance than the ones used to create it, when multiplying shapes, the shape of a larger breast is more preserved. Although breast shapes do not have characteristics that could be added, such as adding an arm to a chair, arithmetic operations result in visible changes. This means that the β -VAE-EMD ($LS=256$) TL1 model can create new realistic breast shapes from others, either by interpolation or other arithmetic operations.

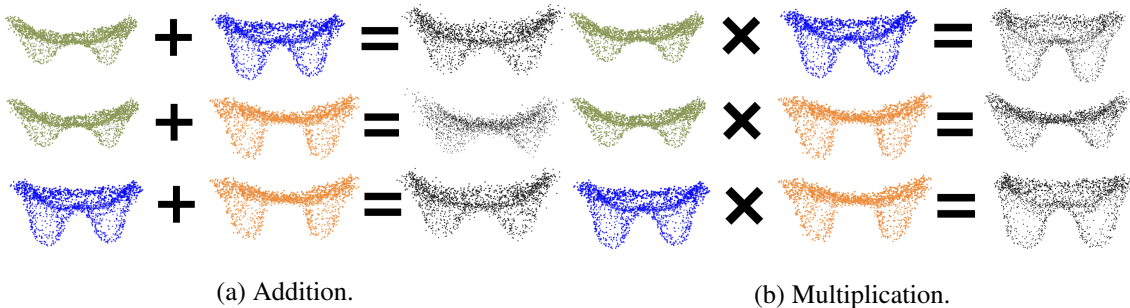


Figure 6.12: Arithmetic operations in the latent space of β -VAE-EMD ($LS=256$) TL1 model. Legend: green - ID10, blue - ID17, orange - ID22, black - result.

6.5 Conditional Generation of Breast Shapes

As seen in Chapter 3, conditioning a VAE model enables the generation of data based on a user-defined non-latent variable. To accomplish a CVAE, a one-hot encoded label was passed to the β -VAE model (both on the encoder and decoder) during the training cycle, as the architecture in Figure 6.13 suggests. This way, new shapes can be synthesized according to this label after training.

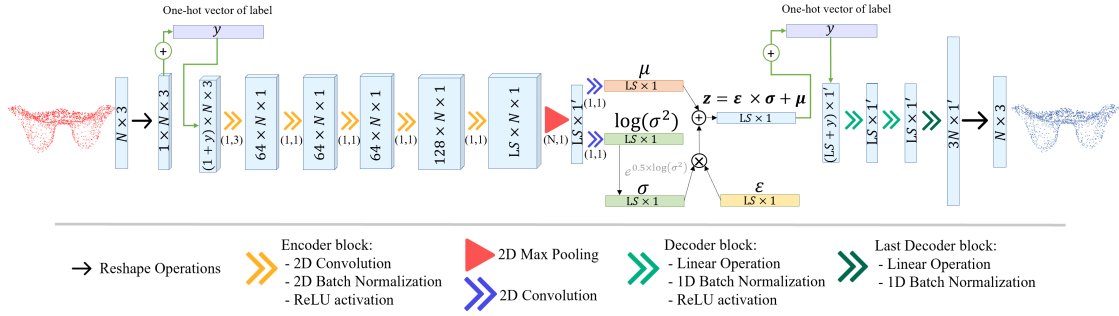


Figure 6.13: PointNet based Conditional Variational Autoencoder diagram. N is the number of points of the PCL, LS is the chosen latent space size and y is the number of possible labels. Kernel sizes inside parentheses.

At first, a CVAE on ShapeNet multi-class was trained to guarantee the correct functioning of the architecture. The implementation details were the same as the optimal β -VAE and the generation was made according to the respective class (chair, lamps, airplanes, and tables). Visual results can be consulted in Figure A.8.

In the case of breast shapes, this is particularly useful as it can enable the generation of shapes according to physical characteristics such as volume, torso width, breast shape. This type of tool could be used in education or surgery planning, enabling the rapid creation of shapes according to certain patient's criteria. A β -VAE can be used to synthesize breast with plausible and common shapes and appearances whilst a CVAE can be used to approximate a specific patient's data when this is not available. However, the available data at the time was not big or varying enough to perform these kinds of experiments, since it would result in a small set of shapes for each category. Given that, conditional generation of breast shapes with labels according to the rotation applied to the original breast shape was explored. Dividing an xy plane in four parts: label 0 for angles from $-\pi/4$ to $\pi/4$, label 1 for angles from $\pi/4$ to $3\pi/4$, label 2 for angles from $3\pi/4$ to $-\pi/4$, and label 3 for angles between $-\pi/4$ to $-\pi/4$. The model was directly trained in the 3D Breast dataset, with remaining implementation details equal to β -VAE-EMD ($LS=256$) DT model ($N=2048$ and $\beta = 0.5$).

The synthetic shapes conditioned by the created label are shown in Figure 6.14. The model was capable to generate breast shapes with rotation according to the label given. Taking this into consideration, the exploitation of CVAE models with large datasets categorized according to features such as breast shapes and volumes is a promising topic for future research.

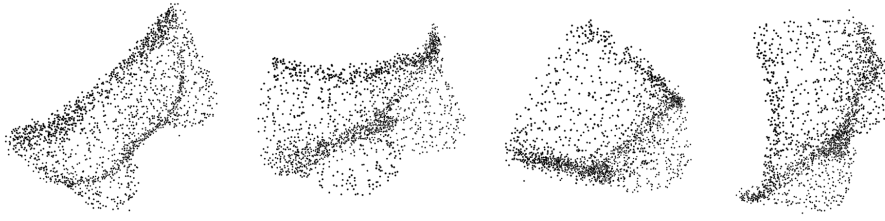


Figure 6.14: Conditional synthesis of breast shapes, with rotation. Results produced by β -CVAE-EMD ($LS=256$), direct training.

6.6 Cross Validation and Discussion

After stabilizing the autoencoder and VAE presented so far in this chapter, the best overall models were re-trained with a 4-fold cross validation technique, with new data that became available posteriorly (total of 49 shapes instead of 26). Performing cross-validation gives a better estimate of the generalization capacity of the models. For the AE and β -VAE 49 objects were available for this task. The data was divided into folds of 30/10 train/validation and 9 shapes were kept as a test set. For the pose transformation, the number of objects available was the same as previously, so an 18/4 train/validation split was made in 5-fold cross validation, without a test set.

Recalling the conclusions from the autoencoder models in Section 6.2: AE-CD ($LS=256$) DT outperformed AE-CD ($LS=1024$) TL1 both in reconstruction metrics and shape completion tasks. To understand the benefit or downside of using transfer learning, both models were tested with cross validation. TL2 option was discarded since it did not show any advantage to the model. In the case of the β -VAE models from Section 6.4, transfer learning was advantageous, so β -VAE-EMD ($LS=256$) TL1 $N=2048$ was used for cross validation.

Table 6.9 presents the reconstruction metrics for cross validation methods. For the AE-CD ($LS=256$) DT, the test set results (6.67/6.28 and 22.56/19.24 mm, $Rec \rightarrow Ori/Ori \rightarrow Rec$ mean Euclidean and Hausdorff distances, respectively) are marginally better than the results on the folded validation sets (7.34/6.94 and 25.94/19.24 mm), which shows once again that DT models can generalize. In the other autoencoder model, AE-CD ($LS=1024$) TL1, the same tendency applies to mean Euclidean and Hausdorff distances of 6.72/6.25 and 25.89/19.55 mm, respectively, on the test set, and 7.58/6.78 and 27.19/21.48 mm on the validation sets. In this experiment, the differences between the validation and test sets are smaller than the ones between train and validation in Table 6.4, showing improved generalization. Comparing the two AE models trained with cross validation, the use of transfer learning from a larger dataset cannot be entirely discredited. However, using direct train has proved enough to represent breast shapes, whilst being easier to train. Besides, the DT model has shown better results in shape completion tasks presented previously.

In the case of pose transformation from prone to upright, the results with the cross validation are more aligned with the ones reported in Table 6.5 for AE-CD ($LS=256$) DT model. This is expected since the shapes used were the same. The cross validation method displayed reconstruction metrics of 5.88/5.56 and 20.87/16.40 mm, $Rec \rightarrow Ori/Ori \rightarrow Rec$ mean Euclidean and Hausdorff distances, respectively. These are similar to those previously present in the training set (5.32/5.34

and 18.28/15.54 *mm*) while being also better than the validation set (7.56/7.04 and 30.10/21.97 *mm*). The results validate the idea that Deep Learning strategies can be used to model pose transformation techniques, being a faster alternative to biomechanical models.

Table 6.9: Numerical performance of PointNet Autoencoder and Variational Autoencoder models in cross-validation methods ($N=2048$): mean and standard deviation distances are shown.

(a) Cross Validation on the Training Set (40 patients).

Model	Method			Euclidean↓		Hausdorff↓	
	Train	LS	Loss	$Rec \rightarrow Ori$	$Ori \rightarrow Rec$	$Rec \rightarrow Ori$	$Ori \rightarrow Rec$
AE	DT	256	CD	7.34 ± 1.01	6.94 ± 0.93	25.94 ± 6.76	22.18 ± 3.74
	TL1	1024	CD	7.58 ± 1.12	6.78 ± 1.01	27.19 ± 6.38	21.48 ± 3.54
β -VAE	TL1	256	EMD	8.23 ± 1.41	8.96 ± 1.50	26.99 ± 4.86	31.61 ± 6.36
Pose AE*	DT	256	CD	5.88 ± 1.21	5.56 ± 1.03	20.87 ± 7.94	16.40 ± 4.15

*5-fold (18 train, 4 validation, without replacement)

(b) Test Set (9 patients).

Model	Method			Euclidean↓		Hausdorff↓	
	Train	LS	Loss	$Rec \rightarrow Ori$	$Ori \rightarrow Rec$	$Rec \rightarrow Ori$	$Ori \rightarrow Rec$
AE	DT	256	CD	6.67 ± 1.02	6.28 ± 1.07	22.56 ± 5.38	19.24 ± 4.46
	TL1	1024	CD	6.72 ± 0.74	6.25 ± 0.86	25.89 ± 8.89	19.55 ± 5.57
β -VAE	TL1	256	EMD	7.76 ± 0.52	8.46 ± 0.62	25.79 ± 3.54	31.14 ± 7.94

At last, for the β -VAE-EMD ($LS=256$) TL1 model, the cross validation reconstruction metrics ($8.23/8.96$ and $26.99/31.61$ *mm*, $Rec \rightarrow Ori/Ori \rightarrow Rec$ mean Euclidean and Hausdorff distances, respectively) are somewhat similar to those presented in the 26 objects dataset in the validation set in Table 6.7 ($8.45/8.81$ and $26.99/30.17$ *mm*), but with improved results in the test set ($7.76/8.46$ and $25.79/31.14$ *mm*), probably because of the larger number of objects used now for training. Similar to the work presented in Table 6.8, quantitative measures were also computed to evaluate the β -VAE model. The JSD, MMD, and COV were calculated comparing the 9 elements from the test set and $9 \times 3 = 27$ generated elements. Table 6.10 holds the results. Overall, the results with cross validation are better. This is expected since training with a larger dataset can help to capture more relevant features of the population. Taking all of this into consideration, the β -VAE models can be used to synthesize, transform, and represent breast shapes according to the desired application, constituting a possible model for medical usage.

Table 6.10: Quantitative results of the β -VAE trained with cross-validation. Metrics on the test set (9 patients) and in the normalized unit sphere.

Method	JSD↓	MMD-CD↓	MMD-EMD↓	COV-CD↑	COV-EMD↑
β -VAE-EMD ($LS=256$) TL1	0.1241	0.0019	22.43	88.89	74.07
Baseline	0.1434	0.0053	72.89	85.19	85.19

Synthetic shapes are 3× the population of the test.

Average results from the three repetitions.

6.7 Summary

This chapter guided us through the exploration of PointNet based models to represent and generate breast shapes using a small dataset. The PointNet based autoencoder can successfully model breast PCLs with reasonable reconstruction errors, create new shapes via interpolation on the latent space, and also perform shape completion tasks. Prone to upright pose transformations is also possible with this autoencoder. These tasks are common in surgical planning or aesthetical outcome evaluation, and are mainly done through complex biomechanical models, so Deep Learning models could be of use. The VAE models can synthesize new realistic shapes that are identical to real ones. The usage of transfer learning from a bigger dataset, in this case, ShapeNet single-class, was proved beneficial, while in the autoencoder direct train is enough for learning representations. The latent space of this model showed that it directly linked some specific features to the shape of the breast, which can be used for creating new shapes. However, it is not yet possible to create shapes with controlled characteristics with this model. Finally, the use of conditional VAE was explored to control the rotation of the breast. It was possible to generate shapes according to the label given, which indicates that conditional models could be used with more relevant labels, such as volume or type of shape.

The presented methods indicate that PCLs can be characterized using a small dataset in a Deep Learning manner using state-of-the-art techniques. Furthermore, it was showed that PointNet based architectures can be used for modelling non-rigid data, even in the most basic settings.

Chapter 7

Conclusions

Breast cancer is the most common form of cancer in females, with more than 2.3 million diagnoses worldwide during 2020. However, mortality rates have been decreasing over the past decades, mainly due to screening programs that help to diagnose the disease in its early stages. Treatments vary from each patient, but in 90% of cases, surgery is part of the treatment, either for removal of cancer conserving the rest of the breast (BCS) or for removal of the totality of the breast (mastectomy). The outcome from breast surgery may not always be the expected, for both physicians and patients, and the breast deformations after BC treatment can negatively affect the quality of life of the patient. Considering this, it is very important for patients to understand how the procedures may affect them. To do so, it is important to develop automatic tools gathering 3D data from the patients' breasts. These can show the areas to be removed during surgery and present visual estimates of the outcome, improving the relationship between physicians and patients. Besides helping in surgery planning, these tools can also be used in education, providing synthetic data created upon request with certain specifications, before and after surgeries. This is helpful since medical students are not always provided with resources to understand how tissues differ from patients and only collect this knowledge upon experience. To develop these tools, relevant datasets of 3D breast models that represent the population are necessary, so that both the 3D tools and CAD systems can be created using relevant data. The acquisition of medical data and annotations from specialists is a long and complicated process, that involves the approval of ethical committees, and it is usually not available off the shelf. This makes the comparison of strategies inconceivable putting a brake on the progress of research in this area. Taking all of this into consideration, the goal of this dissertation was to produce generative models and meaningful representations of 3D breast data, both MRI and breast surface shapes (PCLs), using DL methodologies.

7.1 Contributions and Final Remarks

For volumetric breast MRI, this work presented a variety of methods to learn meaningful structures present on volumes of breast MRI, either in their original resolution or at a patch level. The attempts made on the original resolution of MRI showed that the developed models (Recurrent

Autoencoder, ConvLSTMs, and Auto-encoding GAN) were not capable to capture relevant features from the data to generate or reconstruct MRI scans effectively. Moreover, the models could capture the shapes of the breasts but the quality of the slices was not enough so that they could be used medically. The patch-based methods (3D X-Net architectures) hold better results concerning image quality in the reconstruction of MRI scans. The patch division is beneficial since it divides the data into smaller volumes, which can be trained in more simpler DL architectures and that are computationally less expensive for graphical units to process. Also, using patches results in an increased number of instances for training. However, the high-frequency components of the MRI could not be captured with the detail desired.

Concerning breast shapes, the architectures developed showed that models based on PointNet Vanilla can capture relevant features from breast PCLs, even in a small dataset. The AE model could reconstruct breast shapes, create new shapes via linear interpolation of latent vectors, and even succeed on shape completion tasks. It was also possible to use this model for prone to upright pose transformation, which are sometimes necessary during surgery due to the different positions data is acquired from. Usually these transformations are done with biomechanical models that are time-consuming, while DL models are fast in inference. The VAE model can generate meaningful shapes indistinguishable from the real ones. Also, it is also possible to generate new shapes from crafted VAE latent vectors, given the prior knowledge of specific features on other latent vectors from real data. Using transfer learning on the VAE model was beneficial, while on the AE, direct training was enough to learn representations. Additionally, a CVAE was tested for generating shapes according to a label of rotation of the breast, showing that this type of architecture can be used for the controlled generation of shapes.

These two modules are a step forward in the development of synthetic data for breast MRI and shapes. The methodologies and results contributed to the LeGeM project at INESC TEC for the development of data augmentation methods in 3D breast data.

7.2 Future Work

While the developed methods can integrate medical 3D breast tools by creating surface scans using the VAE models or by reconstructing incomplete inputs of breast MRI using the 3D X-Net trained models, there are still improvements and next steps for the development of synthetic breast data for these tools:

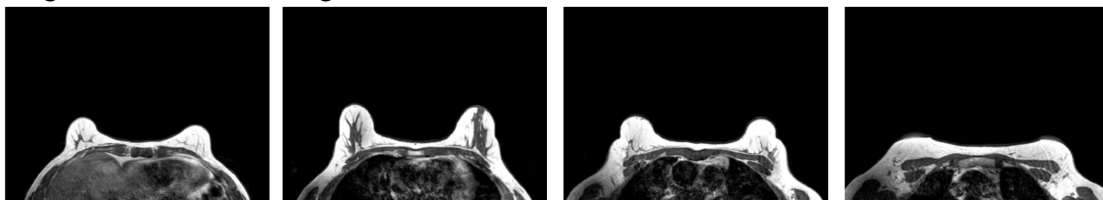
- Development of CVAE with more relevant features as conditions, such as breast volume, shape, or density. These features could be combined in one-hot vectors which would enable the controlled generation of shapes. To do so, a larger dataset than the one used in this dissertation would be necessary, along with the label features. While a larger dataset could be conceived from the VAE model developed and breast volume or shape can be determined by visual inspection, the breast density needs to be determined by specialists.

- Using the developed VAE models for generating semi-synthetic PCLs datasets that could be used to train more complex networks, such as the PointNet with the transform networks in the input. The exploration of the latent spaces from these models could also be explored as they may contain features that condition the decoded representations, similar to what happened in the developed models, allowing controlled generation.
- Exploration of other metrics for evaluating the generated MRI volumes. Visual Turing Tests made by radiologists could be valuable since they would offer the opinion of experts on the quality of the generated volumes. However, these tests are very time-consuming and hard to obtain.
- Creation of novel loss functions for evaluating the high-frequency content on the MRI volumes that could be used as loss functions for training DL models. This is especially important since the developed methods showed that even if the shape and structures can be captured by the model, the sudden changes in intensity and details of structures were not very well represented in the reconstructions. Using the Fourier Transform could give the models some insight on these components.
- In the case of patch-based methods, the algorithm for reconstruction of entire MRI volumes could be optimized so it would work faster and involve more predictions from the model (increasing overlap on the predictions and make a some average). Another possible method for improving the image quality of the reconstructed scans is to train a super-resolution network after with the original data and the reconstructed scans from the patch level method.
- Specifically for the 3D X-Net architecture, scenarios exploring reconstruction loss (with or without the use of adversarial learning) on the original resolution of the slices may help to capture some of the structures with more fidelity.
- Creation of technologies for manipulating input data, such as label maps in the case of MRI, to create scans according to specifications desired using generative models.
- Development of a methodology for generating corresponding MRI volumes from PCLs. While segmenting the MRI contour of the breast is a possible way to generate a breast PCLs, which was the method to develop the dataset used in this work, the contrary operation is much more intricate. The combination of the internal structures of the breasts (from annotations, for instance) with the corresponding PCL of the breast contour, could be used as input for generating a volumetric breast MRI with those specifications. A combination of a DL network that received both PCL (PointNet variation, for instance) and annotation volumes (Convolutional Network) could be used to generate the MRI volume. This would enable the creation of the final pipeline with models of both interior and exterior parts of the breasts, combining MRI and surface information.

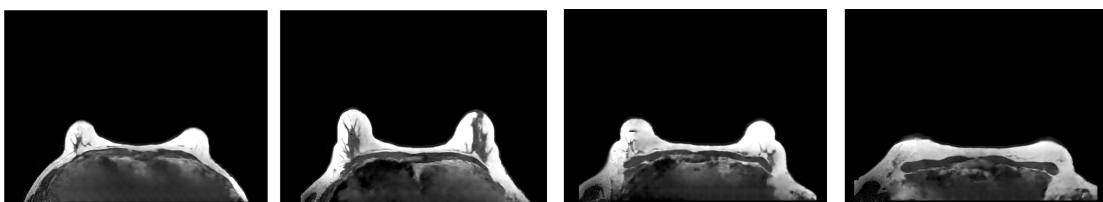
Appendix A

Visual Results

Original slices from training set scan



Reconstructed slices

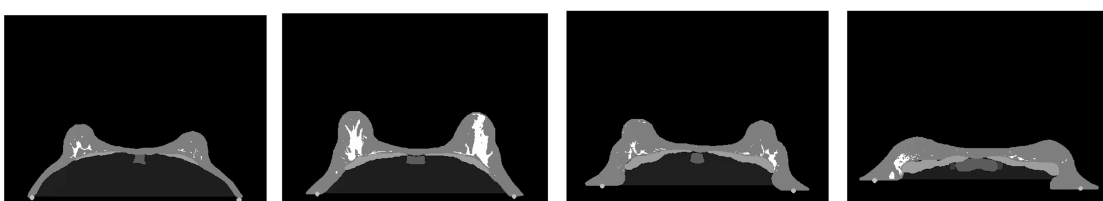


(a) MRI

Original slices from training set scan



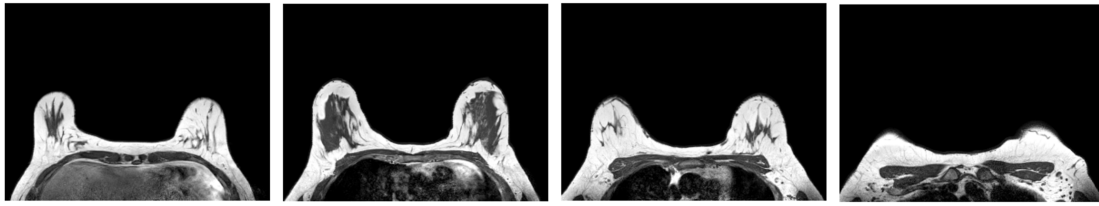
Reconstructed slices



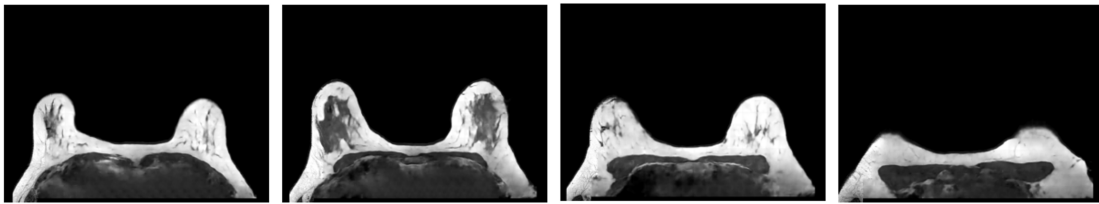
(b) Annotation

Figure A.1: Visual performance of 3D X-Net reconstructed slices with GT-M2 model and (64, 64, 4) patches - additional visual results on training scan.

Original slices from validation set scan

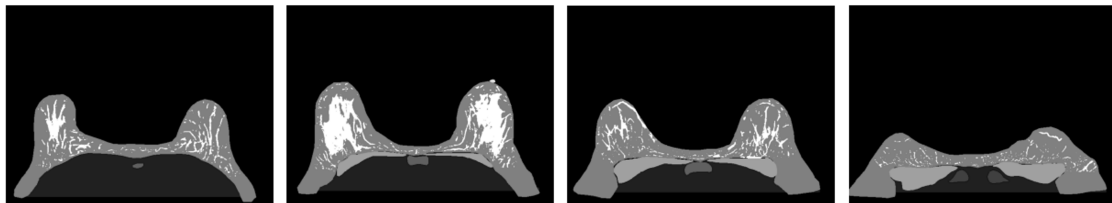


Reconstructed slices

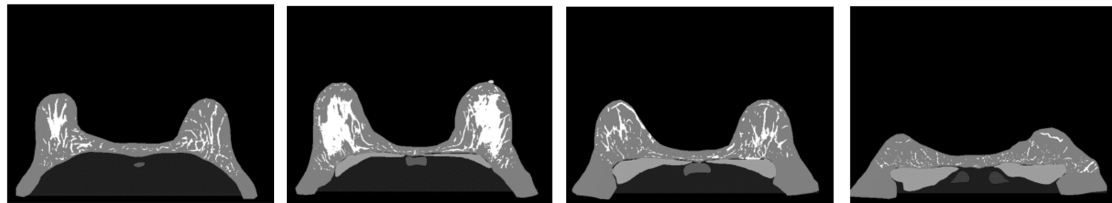


(a) MRI

Original slices from validation set scan

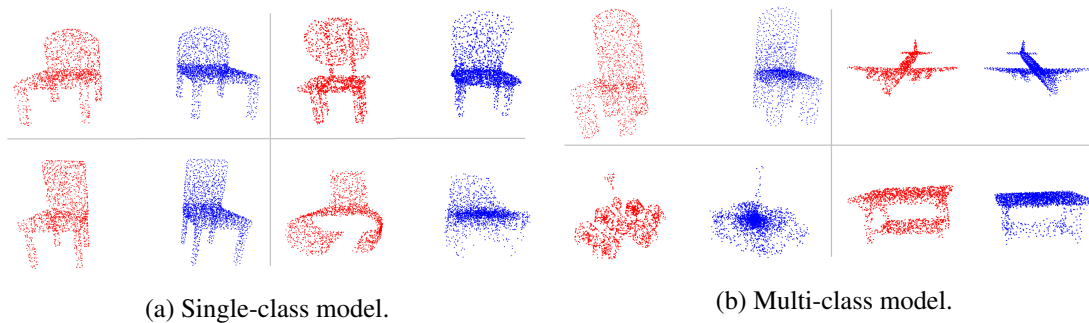


Reconstructed slices



(b) Annotation

Figure A.2: Visual performance of 3D X-Net reconstructed slices with GT-M2 model and (64, 64, 4) patches - additional visual results on validation scan.



(a) Single-class model.

(b) Multi-class model.

Figure A.3: Reconstruction of unseen shapes from ShapeNet dataset in the validation set. Ground truth objects are shown in red, reconstructions are shown in blue.

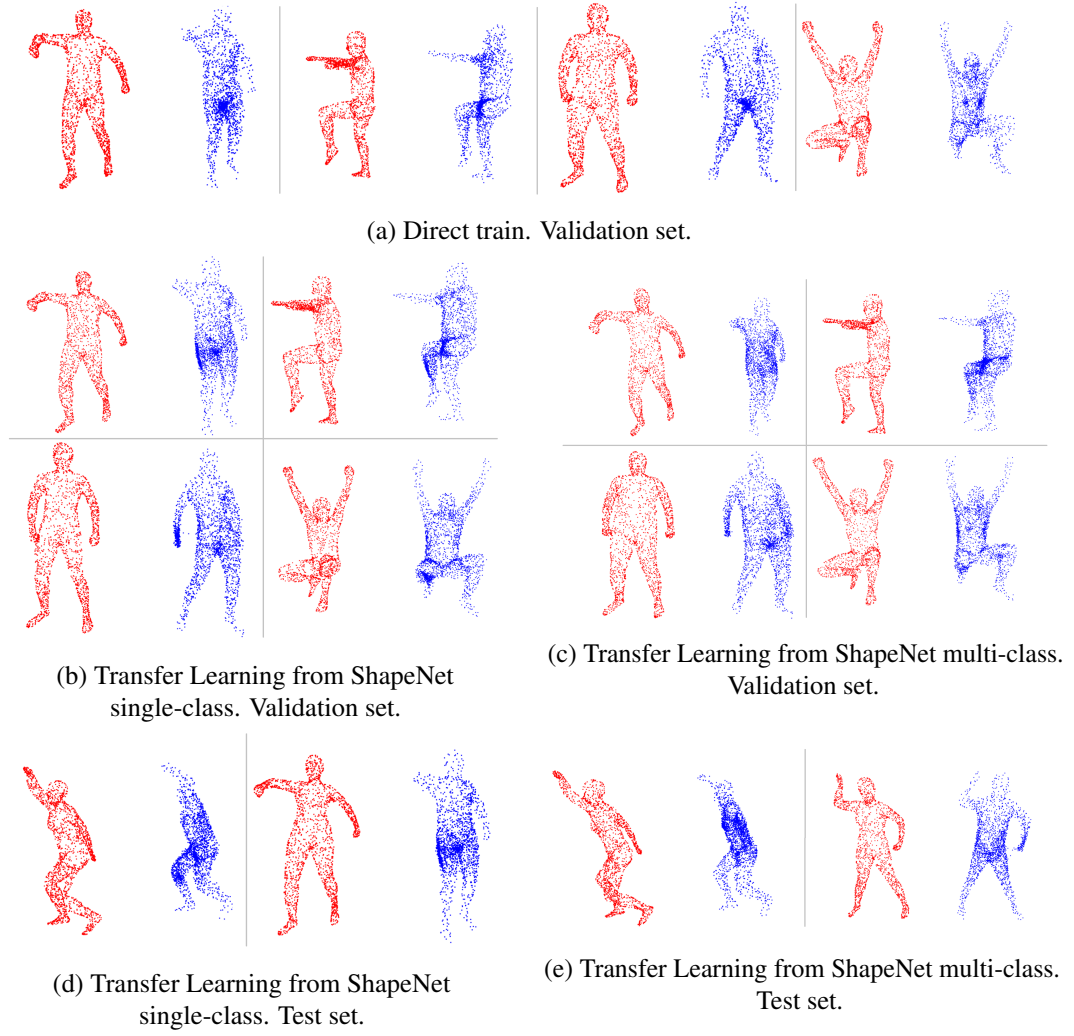


Figure A.4: Reconstruction of unseen shapes from SHREC14. Ground truth objects are shown in red, reconstructions are shown in blue.

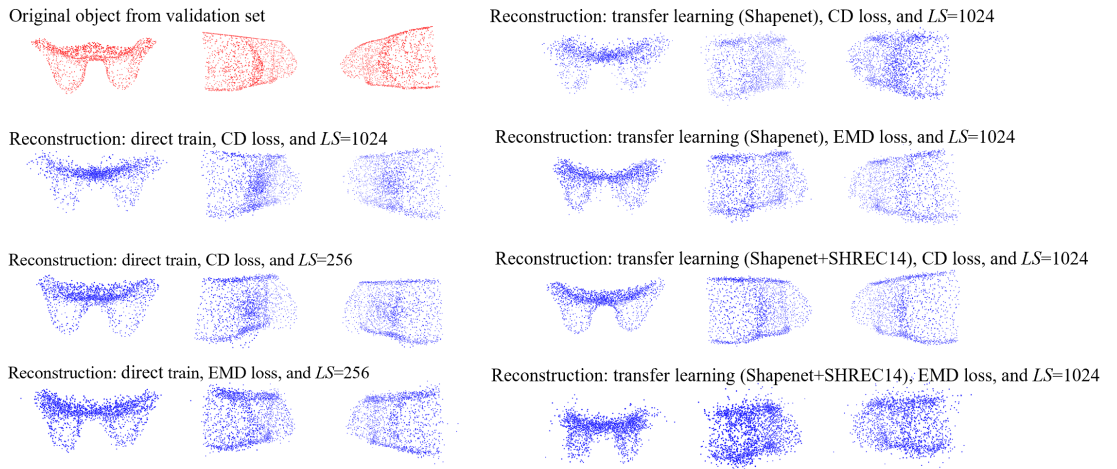


Figure A.5: Reconstruction of unseen shapes from 3D Breast dataset in the validation set. Ground truth objects are shown in red, reconstructions are shown in blue.

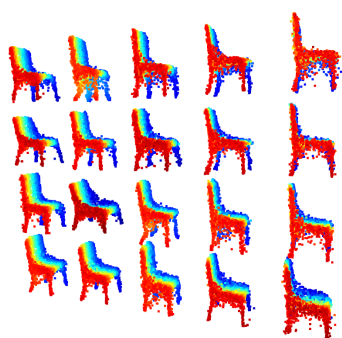


(a) Single-class model (chair - chair).

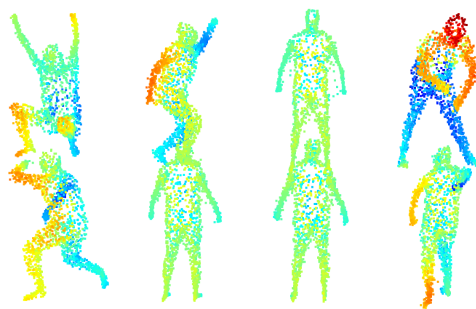


(b) Multi-class model (chair - table).

Figure A.6: Linear interpolation between objects from ShapeNet dataset.



(a) ShapeNet single-class.



(b) SHREC14 (male), transfer learning from ShapeNet single-class.

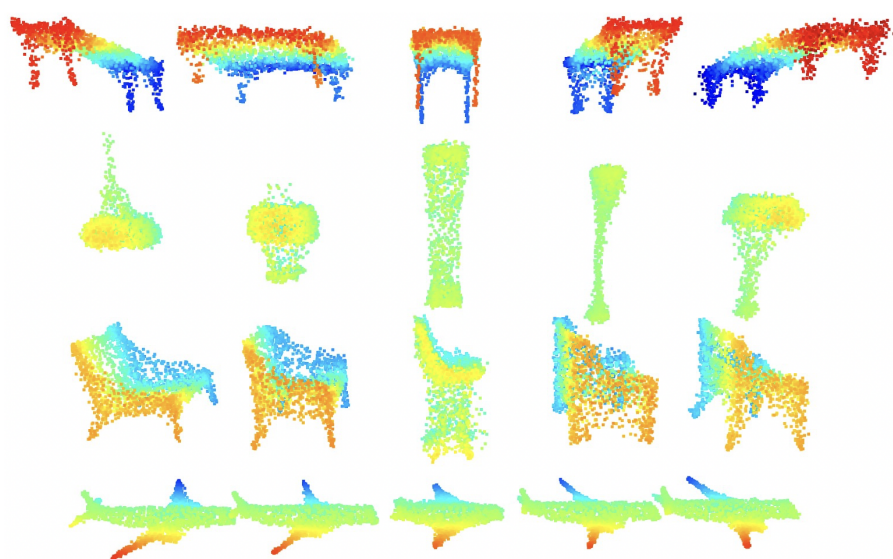
Figure A.7: Synthesis of shapes with β -VAE-EMD (LS=256).

Figure A.8: Conditional synthesis of objects in ShapeNet model.

Appendix B

Tables

Table B.1: Numerical performance of PointNet Autoencoder models: mean and standard deviation distances are shown. All values are multiplied by a factor of 10^{-3} . DT: Direct Train. TL: Transfer Learning.

(a) AE-CD (LS=1024): performance on the rigid ShapeNet the validation set. Distances in Unit Sphere normalization.

Train	Method Description	Euclidean↓		Hausdorff↓	
		$Rec \rightarrow Ori$	$Ori \rightarrow Rec$	$Rec \rightarrow Ori$	$Ori \rightarrow Rec$
DT	Single-Class	30.15 ± 9.08	28.27 ± 8.47	111.76 ± 50.69	94.21 ± 48.70
DT	Multi-Class	26.20 ± 9.98	24.87 ± 9.24	99.08 ± 54.84	82.95 ± 43.55

(b) AE-CD (LS=1024): performance on the non-rigid SHREC14 dataset. Validation set. Measures in meters.

Train	Method Description	Euclidean↓		Hausdorff↓	
		$Rec \rightarrow Ori$	$Ori \rightarrow Rec$	$Rec \rightarrow Ori$	$Ori \rightarrow Rec$
DT	-	19.83 ± 1.93	25.18 ± 2.56	79.18 ± 11.28	79.34 ± 9.15
TL	Single-class ShapeNet	19.63 ± 1.97	20.12 ± 1.81	67.06 ± 9.43	55.78 ± 6.02
TL	Multi-class ShapeNet	20.18 ± 2.01	20.42 ± 1.98	68.24 ± 8.80	58.94 ± 7.92

(c) AE-CD (LS=1024): performance on the non-rigid SHREC14 dataset. Test set. Distances in meters.

Train	Method Description	Euclidean		Hausdorff	
		$Rec \rightarrow Ori$	$Ori \rightarrow Rec$	$Rec \rightarrow Ori$	$Ori \rightarrow Rec$
DT	-	-	-	-	-
TL	ShapeNet single-class	20.84 ± 2.09	20.48 ± 2.02	66.88 ± 9.77	57.65 ± 10.02
TL	ShapeNet multi-class	21.01 ± 2.19	20.69 ± 2.10	65.24 ± 7.94	60.05 ± 8.57

Table B.2: Numerical performance of PointNet Autoencoder models with data augmentation: mean and standard deviation distances are shown. Distances in millimeters.

(a) Training Set (22 patients).

Train	Method		Loss	Euclidean↓		Hausdorff↓	
	LS	Loss		$Rec \rightarrow Ori$	$Ori \rightarrow Rec$	$Rec \rightarrow Ori$	$Ori \rightarrow Rec$
DT	256	CD	0.27	8.64 ± 5.50	6.96 ± 1.73	48.52 ± 58.48	21.37 ± 5.73
TL1	1024	CD	0.16	5.56 ± 0.90	5.18 ± 0.80	20.68 ± 4.94	17.32 ± 4.91

(b) Validation Set (4 patients).

Train	Method		Loss	Euclidean↓		Hausdorff↓	
	LS	Loss		$Rec \rightarrow Ori$	$Ori \rightarrow Rec$	$Rec \rightarrow Ori$	$Ori \rightarrow Rec$
DT	256	CD	9.20	11.23 ± 7.32	8.92 ± 2.66	57.92 ± 58.72	31.00 ± 12.46
TL1	1024	CD	0.42	7.32 ± 0.93	7.13 ± 0.53	23.51 ± 1.83	17.31 ± 4.91

TL1: transfer learning from model previously trained on single-class ShapeNet.

Table B.3: Numerical performance of PointNet Variational Autoencoder model β -VAE-EMD (LS=256): mean and standard deviation distances are shown. All values are multiplied by a factor of 10^{-3} .

(a) Performance on the rigid ShapeNet the validation set. Distances in Unit Sphere normalization.

Train	Method		β	N	Euclidean↓		Hausdorff↓	
	Description				$Rec \rightarrow Ori$	$Ori \rightarrow Rec$	$Ori \rightarrow Rec$	$Ori \rightarrow Rec$
DT	Single-Class	4	2048		57.24 ± 16.93	72.93 ± 28.44	232.49 ± 53.29	285.94 ± 80.07
DT	Single-Class	4	4096		47.45 ± 13.96	58.05 ± 20.53	217.58 ± 51.21	236.39 ± 79.37
DT	Single-Class	6	4096		49.76 ± 15.29	63.06 ± 23.72	224.40 ± 65.21	272.15 ± 70.61

(b) Performance on the non-rigid SHREC14 dataset. Validation set. Distances in meters.

Train	Method		N	β	Euclidean↓		Hausdorff↓	
	Description				$Rec \rightarrow Ori$	$Ori \rightarrow Rec$	$Rec \rightarrow Ori$	$Ori \rightarrow Rec$
TL	ShapeNet single-class		2048	4	23.40 ± 4.60	25.92 ± 7.20	86.30 ± 19.61	88.31 ± 19.61
TL	ShapeNet single-class		4096	6	19.99 ± 4.07	21.82 ± 5.92	100.20 ± 35.23	82.89 ± 28.63

(c) Performance on the non-rigid SHREC14 dataset. Test set. Measures in meters.

Train	Method		N	β	Euclidean↓		Hausdorff↓	
	Description				$Rec \rightarrow Ori$	$Ori \rightarrow Rec$	$Rec \rightarrow Ori$	$Ori \rightarrow Rec$
TL	ShapeNet single-class		2048	4	23.10 ± 4.02	25.37 ± 5.56	82.67 ± 16.17	81.33 ± 20.93
TL	ShapeNet single-class		4096	6	20.23 ± 3.83	21.65 ± 4.93	90.66 ± 27.64	78.80 ± 19.60

References

- [1] Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, and Leonidas Guibas. Learning representations and generative models for 3d point clouds. *35th International Conference on Machine Learning, ICML 2018*, 1:67–85, 2018. [arXiv:1707.02392](#).
- [2] L. Aerts, M. R. Christiaens, P. Enzlin, P. Neven, and F. Amant. Sexual functioning in women after mastectomy versus breast conserving therapy for early-stage breast cancer: A prospective controlled study. *Breast*, 23(5):629–636, 2014. URL: <http://dx.doi.org/10.1016/j.breast.2014.06.012>, doi:10.1016/j.breast.2014.06.012.
- [3] Eman Ahmed, Alexandre Saint, Abd El Rahman Shabayek, Kseniya Cherenkova, Rig Das, Gleb Gusev, Djamila Aouada, and Bjorn Ottersten. A survey on Deep Learning Advances on Different 3D Data Representations. 1(1):1–35, 2018. URL: <http://arxiv.org/abs/1808.01462>, [arXiv:1808.01462](#).
- [4] American Cancer Society. Breast Cancer | Breast Cancer Information; Overview, 2019. Accessed on October 19 2020. URL: <https://www.cancer.org/cancer/breast-cancer.html>.
- [5] Karim Armanious, Chenming Jiang, Marc Fischer, K Thomas, and Konstantin Nikolaou. MedGAN : Medical Image Translation using GANs. pages 1–16, 2018. [arXiv:arXiv:1806.06397v2](#).
- [6] Rowel Atienza. A Conditional Generative Adversarial Network for Rendering Point Clouds. *CVPR Workshops*, pages 10–17, 2019.
- [7] Isaac Bankman. *Handbook of medical image processing and analysis*. Elsevier, 2008.
- [8] Anna-Belle Beau, Per Kragh Andersen, Ilse Vejborg, and Elsebeth Lynge. Limitations in the effect of screening on breast cancer mortality. *Journal of Clinical Oncology*, 36(30):2988, 2018.
- [9] Brad Boehmke, Brandon Greenwell, Brad Boehmke, and Brandon Greenwell. Autoencoders. *Hands-On Machine Learning with R*, pages 377–395, 2020. [arXiv:2003.05991](#), doi:10.1201/9780367816377-19.
- [10] Ali Borji. Pros and cons of GAN evaluation measures. *Computer Vision and Image Understanding*, 179:41–65, 2019. [arXiv:1802.03446](#), doi:10.1016/j.cviu.2018.10.009.
- [11] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

- [12] Gene Mutation Carriers, Women With, and High Familial. Effectiveness of Breast Cancer Surveillance in BRCA1/2 Gene Mutation Carriers and Women With High. *American Society of Clinical Oncology*, 19(4):924–930, 2016.
- [13] Pedro Henrique Moreira Queirós Carvalho. Multimodal Breast Image Registration: Mapping MRI and Surface Data. Master’s thesis, Faculdade de Engenharia da Universidade do Porto, 2018. URL: <https://hdl.handle.net/10216/114249>.
- [14] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.
- [15] Agisilaos Chartsias, Thomas Joyce, Mario Valerio Giuffrida, and Sotirios A. Tsaftaris. Multimodal MR Synthesis via Modality-Invariant Latent Representation. *IEEE Transactions on Medical Imaging*, 37(3):803–814, 2018. doi:10.1109/TMI.2017.2764326.
- [16] Mark Chen, Alec Radford, Rewon Child, Jeffrey Wu, Heewoo Jun, David Luan, and Ilya Sutskever. Generative pretraining from pixels. In *International Conference on Machine Learning*, pages 1691–1703. PMLR, 2020.
- [17] Wanjing Chen, Xiaoi Lv, Xiaohong Xu, Xiufei Gao, and Bei Wang. Meta-analysis for psychological impact of breast reconstruction in patients with breast cancer. *Breast Cancer*, 25(4):464–469, 2018. doi:10.1007/s12282-018-0846-8.
- [18] Xi Chen, Yan Duan, Rein Houthoofd, John Schulman, Ilya Sutskever, and Pieter Abbeel. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. *Advances in Neural Information Processing Systems*, pages 2180–2188, 2016. arXiv:1606.03657.
- [19] Xi Chen, Yan Duan, Rein Houthoofd, John Schulman, Ilya Sutskever, and Pieter Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Advances in neural information processing systems*, pages 2172–2180, 2016.
- [20] D. R.H. Christie, M. Y. O’Brien, J. A. Christie, T. Kron, S. A. Ferguson, C. S. Hamilton, and J. W. Denham. A comparison of methods of cosmetic assessment in breast conservation treatment. *Breast*, 5(5):358–367, 1996. doi:10.1016/S0960-9776(96)90004-1.
- [21] DRH Christie, MY O’Brien, JA Christie, T Kron, SA Ferguson, CS Hamilton, and JW Denham. A comparison of methods of cosmetic assessment in breast conservation treatment. *The Breast*, 5(5):358–367, 1996.
- [22] Maria JM Chuquicusma, Sarfaraz Hussein, Jeremy Burt, and Ulas Bagci. How to fool radiologists with generative adversarial networks? a visual turing test for lung cancer diagnosis. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 240–244. IEEE, 2018.
- [23] Elodia B Cole, Zheng Zhang, Helga S Marques, R Edward Hendrick, Martin J Yaffe, and Etta D Pisano. Impact of computer-aided detection systems on radiologist accuracy with digital mammography. *American Journal of Roentgenology*, 203(4):909–916, 2014.

- [24] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A. Bharath. Generative Adversarial Networks: An Overview. *IEEE Signal Processing Magazine*, 35(1):53–65, 2018. [arXiv:1710.07035](https://arxiv.org/abs/1710.07035), [doi:10.1109/MSP.2017.2765202](https://doi.org/10.1109/MSP.2017.2765202).
- [25] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.
- [26] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [27] Emily L Denton, Soumith Chintala, Rob Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. In *Advances in neural information processing systems*, pages 1486–1494, 2015.
- [28] Mariana Ribeiro Dias. Master Thesis: Thoracic MRI emulation through texture synthesis and ground truth manipulation. 2019.
- [29] Won-Joon Do, Sunghun Seo, Yoseob Han, Jong Chul Ye, Seung Hong Choi, and Sung-Hong Park. Reconstruction of multicontrast mr images through deep learning. *Medical physics*, 47(3):983–997, 2020.
- [30] Fairview. Lumpectomy _ Saint Luke’s Health System. Accessed on October 19 2020. URL: <https://www.fairview.org/patient-education/83001>.
- [31] Yi Fang, Jin Xie, Guoxian Dai, Meng Wang, Fan Zhu, Tiantian Xu, and Edward Wong. 3D deep shape descriptor. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June-2015:2319–2328, 2015. [doi:10.1109/CVPR.2015.7298845](https://doi.org/10.1109/CVPR.2015.7298845).
- [32] Jesse R. Fann, Anne M. Thomas-Rich, Wayne J. Katon, Deborah Cowley, Mary Pepping, Bonnie A. McGregor, and Julie Gralow. Major depression after breast cancer: a review of epidemiology and treatment. *General Hospital Psychiatry*, 30(2):112–126, 2008. [doi:10.1016/j.genhosppsych.2007.10.008](https://doi.org/10.1016/j.genhosppsych.2007.10.008).
- [33] William Fedus, Mihaela Rosca, Balaji Lakshminarayanan, Andrew M Dai, Shakir Mohamed, and Ian Goodfellow. Many paths to equilibrium: Gans do not need to decrease a divergence at every step. *arXiv preprint arXiv:1710.08446*, 2017.
- [34] 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. *Neurocomputing*, 321:321–331, 2018. [arXiv:1803.01229](https://arxiv.org/abs/1803.01229), [doi:10.1016/j.neucom.2018.09.013](https://doi.org/10.1016/j.neucom.2018.09.013).
- [35] Maayan Frid-Adar, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Synthetic data augmentation using GAN for improved liver lesion classification. *Proceedings - International Symposium on Biomedical Imaging*, 2018-April:289–293, 2018. [arXiv:1801.02385](https://arxiv.org/abs/1801.02385), [doi:10.1109/ISBI.2018.8363576](https://doi.org/10.1109/ISBI.2018.8363576).
- [36] Jon Gauthier. Conditional generative adversarial nets for convolutional face generation. *Technical report*, 2014.

- [37] Donald Geman, Stuart Geman, Neil Hallonquist, and Laurent Younes. Visual turing test for computer vision systems. *Proceedings of the National Academy of Sciences*, 112(12):3618–3623, 2015.
- [38] Adam Golinski, Reza Pourreza, Yang Yang, Guillaume Sautiere, and Taco S. Cohen. Feedback recurrent autoencoder for video compression. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, November 2020.
- [39] Ian Goodfellow. NIPS 2016 Tutorial: Generative Adversarial Networks. 2016. URL: <http://arxiv.org/abs/1701.00160>, [arXiv:1701.00160](#).
- [40] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in Neural Information Processing Systems*, 3(January):2672–2680, 2014. [arXiv:arXiv:1406.2661v1](#).
- [41] Kan Guo, Dongqing Zou, and Xiaowu Chen. 3D mesh labeling via deep convolutional neural networks. *ACM Transactions on Graphics*, 35(1), 2015. [doi:10.1145/2835487](#).
- [42] Swaminathan Gurumurthy, Ravi Kiran Sarvadevabhatla, and R Venkatesh Babu. Deligan: Generative adversarial networks for diverse and limited data. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 166–174, 2017.
- [43] Azam Hamidinekoo, Erika Denton, Andrik Rampun, Kate Honnor, and Reyer Zwiggelaar. Deep learning in mammography and breast histology, an overview and future trends. *Medical image analysis*, 47:45–67, 2018.
- [44] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P. Recht, Daniel K. Sodickson, Thomas Pock, and Florian Knoll. Learning a Variational Network for Reconstruction of Accelerated MRI Data. *Physiology & behavior*, 176(3):139–148, 2019. [doi:10.1016/j.physbeh.2017.03.040](#).
- [45] Changhee Han, Hideaki Hayashi, Leonardo Rundo, Ryosuke Araki, Wataru Shimoda, Shinichi Muramatsu, Yujiro Furukawa, Giancarlo Mauri, and Hideki Nakayama. GAN-based synthetic brain MR image generation. *Proceedings - International Symposium on Biomedical Imaging*, 2018-April(Isbi):734–738, 2018. [doi:10.1109/ISBI.2018.8363678](#).
- [46] Karol Hausman, Yevgen Chebotar, Stefan Schaal, Gaurav Sukhatme, and Joseph J. Lim. Multi-modal imitation learning from unstructured demonstrations using generative adversarial nets. *Advances in Neural Information Processing Systems*, 2017-December:1236–1246, 2017. [arXiv:1705.10479](#).
- [47] M Heath, Kevin Bowyer, D Kopans, R Moore, and P Kegelmeyer. The digital database for screening mammography. *Proceedings of the Fourth International Workshop on Digital Mammography*, 01 2000. [doi:10.1007/978-94-011-5318-8_75](#).
- [48] Helga Henseler, Balvinder S. Khambay, Adrian Bowman, Joanna Smith, J. Paul Siebert, Susanne Oehler, Xiangyang Ju, Ashraf Ayoub, and Arup K. Ray. Investigation into accuracy and reproducibility of a 3D breast imaging system using multiple stereo cameras. *Journal of Plastic, Reconstructive and Aesthetic Surgery*, 64(5):577–582, 2011. URL: <http://dx.doi.org/10.1016/j.bjps.2010.08.044>, [doi:10.1016/j.bjps.2010.08.044](#).

- [49] 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 *Advances in neural information processing systems*, pages 6626–6637, 2017.
- [50] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. B-VAE: Learning basic visual concepts with a constrained variational framework. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, pages 1–22, 2017.
- [51] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics (ToG)*, 36(4):1–14, 2017.
- [52] Phillip Isola, Alexei A Efros, Berkeley Ai, and U C Berkeley. Image-to-Image Translation with Conditional Adversarial Networks. 2018. [arXiv:arXiv:1611.07004v3](https://arxiv.org/abs/1611.07004v3).
- [53] Mor Joseph-Rivlin, Alon Zvirin, and Ron Kimmel. Momenet: Flavor the moments in learning to classify shapes. *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019*, pages 4085–4094, 2019. [arXiv:1812.07431](https://arxiv.org/abs/1812.07431), [doi:10.1109/ICCVW.2019.00503](https://doi.org/10.1109/ICCVW.2019.00503).
- [54] Julien Despois. Autoencoders — Deep Learning bits #1 – Hacker Noon, 2017. Accessed on October 27 2020. URL: <https://hackernoon.com/autoencoders-deep-learning-bits-1-11731e200694>.
- [55] Bodo Kaiser and Shadi Albarqouni. MRI to CT Translation with GANs. 2019. URL: <http://arxiv.org/abs/1901.05259>, [arXiv:1901.05259](https://arxiv.org/abs/1901.05259).
- [56] Kevin M Kelly, Judy Dean, W Scott Comulada, and Sung-jae Lee. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *European Radiology*, pages 734–742, 2010. [doi:10.1007/s00330-009-1588-y](https://doi.org/10.1007/s00330-009-1588-y).
- [57] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, (ML):1–14, 2014. [arXiv:1312.6114](https://arxiv.org/abs/1312.6114).
- [58] Christiane K Kuhl, Simone Schradling, Claudia C Leutner, Nuschin Morakkabati-spitz, Eva Wardelmann, Rolf Fimmers, Walther Kuhn, and Hans H Schild. Mammography , Breast Ultrasound , and Magnetic Resonance Imaging for Surveillance of Women at High Familial Risk for Breast Cancer. *Journal of Clinical Oncology*, 23(33):8469–8476, 2014. [doi:10.1200/JCO.2004.00.4960](https://doi.org/10.1200/JCO.2004.00.4960).
- [59] Gihyun Kwon, Chihye Han, and Dae-shik Kim. Generation of 3d brain mri using auto-encoding generative adversarial networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 118–126. Springer, 2019.
- [60] Carson Lam, Caroline Yu, Laura Huang, and Daniel Rubin. Retinal lesion detection with deep learning using image patches. *Investigative ophthalmology & visual science*, 59(1):590–596, 2018.
- [61] Kusam Lata, Mayank Dave, and Nishanth K.N. Data Augmentation Using Generative Adversarial Network. *SSRN Electronic Journal*, pages 1–14, 2019. [arXiv:arXiv:1711.04340v3](https://arxiv.org/abs/1711.04340v3), [doi:10.2139/ssrn.3349576](https://doi.org/10.2139/ssrn.3349576).

- [62] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-January:105–114, 2017. [arXiv:arXiv:1609.04802v5](#), [doi:10.1109/CVPR.2017.19](#).
- [63] Carol H. Lee, D. David Dershaw, Daniel Kopans, Phil Evans, Barbara Monsees, Debra Monticciolo, R. James Brenner, Lawrence Bassett, Wendie Berg, Stephen Feig, Edward Hendrick, Ellen Mendelson, Carl D’Orsi, Edward Sickles, and Linda Warren Burhenne. Breast Cancer Screening With Imaging: Recommendations From the Society of Breast Imaging and the ACR on the Use of Mammography, Breast MRI, Breast Ultrasound, and Other Technologies for the Detection of Clinically Occult Breast Cancer. *Journal of the American College of Radiology*, 7(1):18–27, 2010. URL: <http://dx.doi.org/10.1016/j.jacr.2009.09.022>, [doi:10.1016/j.jacr.2009.09.022](#).
- [64] Constance D Lehman, Robert D Wellman, Diana SM Buist, Karla Kerlikowske, Anna NA Tosteson, Diana L Miglioretti, Breast Cancer Surveillance Consortium, et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA internal medicine*, 175(11):1828–1837, 2015.
- [65] Chuan Li and Michael Wand. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *European conference on computer vision*, pages 702–716. Springer, 2016.
- [66] Shujie Li, Yang Zhou, Haisheng Zhu, Wenjun Xie, Yang Zhao, and Xiaoping Liu. Bidirectional recurrent autoencoder for 3d skeleton motion data refinement. *Computers & Graphics*, 81:92–103, 2019. URL: <https://www.sciencedirect.com/science/article/pii/S0097849319300330>, [doi:https://doi.org/10.1016/j.cag.2019.03.010](#).
- [67] Xiaqing KL Li, Guangyan Zhang, and W Zheng. Deep learning and its parallelization. *Big Data*, 12, 2016.
- [68] Yangyan Li, Sören Pirk, Hao Su, Charles R. Qi, and Leonidas J. Guibas. FPNP: Field probing neural networks for 3D data. *Advances in Neural Information Processing Systems*, (Nips):307–315, 2016. [arXiv:1605.06240](#).
- [69] Z. Lian, A. Godil, B. Bustos, M. Daoudi, J. Hermans, S. Kawamura, Y. Kurita, G. Lavoué, H. V. Nguyen, R. Ohbuchi, Y. Ohkita, Y. Ohishi, F. Porikli, M. Reuter, I. Sipiran, D. Smeets, P. Suetens, H. Tabia, and D. Vandermeulen. SHREC’11 track: shape retrieval on non-rigid 3d watertight meshes. In *Proceedings of the 4th Eurographics conference on 3D Object Retrieval*, EG 3DOR’11, pages 79–88. Eurographics Association, 2011.
- [70] Ming Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. *Advances in Neural Information Processing Systems*, 2017-December(Nips):701–709, 2017. [arXiv:1703.00848](#).
- [71] Weiping Liu, Jia Sun, Wanyi Li, Ting Hu, and Peng Wang. Deep learning on point clouds and its application: A survey. *Sensors*, 19(19):4188, 2019.

- [72] MA Guevara Lopez, N Posada, Daniel C Moura, Raúl Ramos Pollán, José M Franco Valiente, César Suárez Ortega, M Solar, Guillermo Diaz-Herrero, IMAP Ramos, J Loureiro, et al. Bcdr: a breast cancer digital repository. In *15th International conference on experimental mechanics*, volume 1215, 2012.
- [73] Liga Portuguesa Contra o Cancro LPCC. Programa de Rastreio de Cancro da Mama - Rastreio e Diagnóstico Precoce : Liga Portuguesa Contra o Cancro, 2017. URL: <https://www.ligacontracancro.pt/servicos/detalhe/url/programa-de-rastreio-de-cancro-da-mama/>.
- [74] Andreas Maier, Christopher Syben, Tobias Lasser, and Christian Riess. A gentle introduction to deep learning in medical image processing. *Zeitschrift für Medizinische Physik*, 29(2):86–101, 2019.
- [75] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. Adversarial Autoencoders. 2015. URL: <http://arxiv.org/abs/1511.05644>, [arXiv:1511.05644](https://arxiv.org/abs/1511.05644).
- [76] Ioannis Maniadis, Vassilis Solachidis, Nicholas Vretos, and Petros Daras. Data Augmentation Using GANs for 3D Applications. pages 229–269, 2020. doi:10.4018/978-1-5225-5294-9.ch011.
- [77] Ritse M. Mann, Corinne Balleyguier, Pascal A. Baltzer, and ... Breast MRI: EUSOBI recommendations for women’s information, 2015. Accessed on October 14 2020. URL: <https://link.springer.com/article/10.1007/s00330-015-3807-z>.
- [78] Daniel Maturana and Sebastian Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 922–928. IEEE, 2015.
- [79] Luke Metz, Ben Poole, David Pfau, and Jascha Sohl-Dickstein. Unrolled generative adversarial networks, 2017. [arXiv:1611.02163](https://arxiv.org/abs/1611.02163).
- [80] Mehdi Mirza and Simon Osindero. Conditional Generative Adversarial Nets. pages 1–7, 2014. URL: <http://arxiv.org/abs/1411.1784>, [arXiv:1411.1784](https://arxiv.org/abs/1411.1784).
- [81] Inês C Moreira, Igor Amaral, Inês Domingues, António Cardoso, Maria Joao Cardoso, and Jaime S Cardoso. Inbreast: toward a full-field digital mammographic database. *Academic radiology*, 19(2):236–248, 2012.
- [82] Monica Morrow, Janet Waters, and Elizabeth Morris. MRI for breast cancer screening , diagnosis , and treatment. *The Lancet*, 378(9805):1804–1811, 2011. URL: [http://dx.doi.org/10.1016/S0140-6736\(11\)61350-0](http://dx.doi.org/10.1016/S0140-6736(11)61350-0), doi:10.1016/S0140-6736(11)61350-0.
- [83] Sachin Prasad N and Dana Houserikova. THE ROLE OF VARIOUS MODALITIES IN BREAST IMAGING. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub*, 151(2):209–218, 2007. doi:10.5507/bp.2007.036.
- [84] Zaid Nabulsi, Vineet Kosaraju, and Shuvam Chakraborty. MRNGAN: Reconstructing 3D MRI Scans Using A Recurrent Generative Model. URL: <http://vineetkosaraju.com/papers/mrngan.pdf>.

- [85] National Cancer Intelligence Network. The Second All Breast Cancer Report. *Group*, pages 1–47, 2011. URL: <http://www.google.co.uk/url?sa=t{&}rct=j{&}q={&}esrc=s{&}source=web{&}cd=1{&}ved=0CCMQFjAA{&}url=http://www.ncin.org.uk/view.aspx?rid=612{&}ei=sXd1UM0xPLSM0wWV5YGQDA{&}usg=AFQjCNFnTgxV6RfCyNAmP4hKZ2wpdX639A>.
- [86] Hong-Wei Ng, Viet Dung Nguyen, Vassilios Vonikakis, and Stefan Winkler. Deep learning for emotion recognition on small datasets using transfer learning. In *Proceedings of the 2015 ACM on international conference on multimodal interaction*, pages 443–449, 2015.
- [87] Trung Nguyen, Quang-Hieu Pham, Tam Le, Tung Pham, Nhat Ho, and Binh-Son Hua. Point-set distances for learning representations of 3d point clouds. *arXiv preprint arXiv:2102.04014*, 2021.
- [88] Dong Nie, Roger Trullo, Jun Lian, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with context-aware generative adversarial networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10435 LNCS:417–425, 2016. [arXiv:1612.05362](https://arxiv.org/abs/1612.05362), [doi:10.1007/978-3-319-66179-7_48](https://doi.org/10.1007/978-3-319-66179-7_48).
- [89] Dong Nie, Roger Trullo, Jun Lian, Li Wang, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical Image Synthesis with Deep Convolutional Adversarial Networks. *IEEE Trans Biomed Eng.*, 176(3):139–148, 2018. [doi:10.1109/TBME.2018.2814538](https://doi.org/10.1109/TBME.2018.2814538). [Medical](https://arxiv.org/abs/1804.04366).
- [90] Rachel L. O’Connell, Roger J.G. Stevens, Paul A. Harris, and Jennifer E. Rusby. Review of three-dimensional (3D) surface imaging for oncoplastic, reconstructive and aesthetic breast surgery. *Breast*, 24(4):331–342, 2015. URL: <http://dx.doi.org/10.1016/j.breast.2015.03.011>, [doi:10.1016/j.breast.2015.03.011](https://doi.org/10.1016/j.breast.2015.03.011).
- [91] Helder P. Oliveira, Jaime S. Cardoso, Andre Magalhaes, and Maria J. Cardoso. Methods for the Aesthetic Evaluation of Breast Cancer Conservation Treatment: A Technological Review. *Current Medical Imaging Reviews*, 9(1):32–46, 2013. [doi:10.2174/1573405611309010006](https://doi.org/10.2174/1573405611309010006).
- [92] Sara P. Oliveira, Pedro Morgado, Pedro F. Gouveia, João F. Teixeira, Sílvia Bessa, João P. Monteiro, Hooshiar Zolfagharnasab, Marta Reis, Nuno L. Silva, Diana Veiga, Maria J. Cardoso, Helder Oliveira, and Manuel João Ferreira. Three-dimensional planning tool for breast conserving surgery: a technological review. *Critical Reviews in Biomedical Engineering*, 46(6):523–580, 2018. [doi:10.1615/critrevbiomedeng.2018028476](https://doi.org/10.1615/critrevbiomedeng.2018028476).
- [93] Sahin Olut, Yusuf H. Sahin, Ugur Demir, and Gozde Unal. Generative Adversarial Training for MRA Image Synthesis Using Multi-contrast MRI. (Midl):147–154, 2018. [arXiv:1804.04366](https://arxiv.org/abs/1804.04366), [doi:10.1007/978-3-030-00320-3_18](https://doi.org/10.1007/978-3-030-00320-3_18).
- [94] Word Health Organization. WHO | Breat cancer, 2019. Accessed on June 1 2021. URL: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>.
- [95] Word Health Organization. WHO | Preventing cancer, 2019. Accessed on October 6 2020. URL: <http://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/>.

- [96] Olga Pawluczyk, Bindu J Augustine, Martin J Yaffe, Dan Rico, Jiwei Yang, Gordon E Mawdsley, and Norman F Boyd. A volumetric method for estimation of breast density on digitized screen-film mammograms. *Medical Physics*, pages 352–364, 2003. doi:[10.1118/1.1539038](https://doi.org/10.1118/1.1539038).
- [97] D. Pickup, X. Sun, P. L. Rosin, R. R. Martin, Z. Cheng, Z. Lian, M. Aono, A. Ben Hamza, A. Bronstein, M. Bronstein, S. Bu, U. Castellani, S. Cheng, V. Garro, A. Giachetti, A. Godil, J. Han, H. Johan, L. Lai, B. Li, C. Li, H. Li, R. Litman, X. Liu, Z. Liu, Y. Lu, A. Tatsuma, and J. Ye. SHREC’14 track: Shape retrieval of non-rigid 3d human models. In *Proceedings of the 7th Eurographics workshop on 3D Object Retrieval*, EG 3DOR’14. Eurographics Association, 2014.
- [98] Etta D Pisano, Constantine Gatsonis, D Ph, Edward Hendrick, D Ph, Martin Yaffe, D Ph, Janet K Baum, Suddhasatta Acharyya, D Ph, Emily F Conant, Laurie L Fajardo, Lawrence Bassett, Carl D Orsi, Roberta Jong, and Murray Rebner. Diagnostic Performance of Digital versus Film Mammography for Breast-Cancer Screening. *The New England Journal of Medicine*, pages 1773–1783, 2005.
- [99] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua:77–85, 2017. arXiv:[1612.00593](https://arxiv.org/abs/1612.00593), doi:[10.1109/CVPR.2017.16](https://doi.org/10.1109/CVPR.2017.16).
- [100] Charles R. Qi, Hao Su, Matthias Niebner, Angela Dai, Mengyuan Yan, and Leonidas J. Guibas. Volumetric and multi-view CNNs for object classification on 3D data. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem:5648–5656, 2016. arXiv:[1604.03265](https://arxiv.org/abs/1604.03265), doi:[10.1109/CVPR.2016.609](https://doi.org/10.1109/CVPR.2016.609).
- [101] Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems*, 2017-Decem:5100–5109, 2017. arXiv:[1706.02413](https://arxiv.org/abs/1706.02413).
- [102] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- [103] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, pages 1–16, 2016. arXiv:[1511.06434](https://arxiv.org/abs/1511.06434).
- [104] Fabio Redmondino and Sabry El-Hakim. IMAGE-BASED 3D MODELLING: A REVIEW. *The Photogrammetric Record*, 21(115):269–291, 2006. doi:[10.1111/j.1477-9730.2006.00383.x](https://doi.org/10.1111/j.1477-9730.2006.00383.x).
- [105] Joost M. Riphagen, Johan W. Van Neck, and Leon N.A. Van Adrichem. 3D surface imaging in medicine: A review of working principles and implications for imaging the unsedated child. *Journal of Craniofacial Surgery*, 19(2):517–524, 2008. doi:[10.1097/SCS.0b013e31811ec20a](https://doi.org/10.1097/SCS.0b013e31811ec20a).
- [106] Joseph Rocca. Understanding Variational Autoencoders (VAEs) | by Joseph Rocca | Towards Data Science, 2019. Accessed on

- October 27 2020. URL: <https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73>.
- [107] Cátia Rodrigues. Multimodal Image Registration : mapping MRI and Surface Data to Create 3D Models of the Breast. Master's thesis, Faculdade de Engenharia da Universidade do Porto, 2019.
- [108] Itamar Ronen and Dae Shik Kim. Principles of magnetic resonance imaging, 2008. Accessed on October 15 2020. URL: <https://www.uptodate.com/contents/principles-of-magnetic-resonance-imaging{#}H7>, doi: 10.1142/9789812814807_0005.
- [109] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [110] Mihaela Rosca, Balaji Lakshminarayanan, David Warde-Farley, and Shakir Mohamed. Variational approaches for auto-encoding generative adversarial networks. *arXiv preprint arXiv:1706.04987*, 2017.
- [111] D. Rotten and J. M. Levaillant. The value of ultrasonic examination to detect and diagnose breast carcinomas. analysis of the results obtained in 125 tumors using radiographic and ultrasound mammography. *Ultrasound in Obstetrics and Gynecology*, 2, 1992.
- [112] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, 2000.
- [113] Saint Luke's. Mastectomy. Accessed on October 19 2020. URL: <https://www.saintlukeskc.org/health-library/lumpectomy>.
- [114] Mehdi S.M. Sajjadi, Bernhard Scholkopf, and Michael Hirsch. EnhanceNet: Single Image Super-Resolution Through Automated Texture Synthesis. *Proceedings of the IEEE International Conference on Computer Vision*, 2017-October:4501–4510, 2017. [arXiv:1612.07919](https://arxiv.org/abs/1612.07919), doi:10.1109/ICCV.2017.481.
- [115] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. In *Advances in neural information processing systems*, pages 2234–2242, 2016.
- [116] Irina Sanchez and Veronica Vilaplana. Brain MRI super-resolution using 3D generative adversarial networks. (Midl):1–8, 2018. URL: <http://arxiv.org/abs/1812.11440>, [arXiv:1812.11440](https://arxiv.org/abs/1812.11440).
- [117] M. Savva, F. Yu, Hao Su, M. Aono, B. Chen, D. Cohen-Or, W. Deng, Hang Su, S. Bai, X. Bai, N. Fish, J. Han, E. Kalogerakis, E. G. Learned-Miller, Y. Li, M. Liao, S. Maji, A. Tatsuma, Y. Wang, N. Zhang, and Z. Zhou. SHREC'16 track large-scale 3D shape retrieval from ShapeNet Core55. *Eurographics Workshop on 3D Object Retrieval, EG 3DOR*, pages 89–98, 2016. doi:10.2312/3dor.20161092.
- [118] Irhum Shafkat. Intuitively Understanding Variational Autoencoders | by Irhum Shafkat | Towards Data Science. Accessed on October 27 2020. URL: <https://towardsdatascience.com/intuitively-understanding-variational-autoencoders-1bfe67eb5daf>.

- [119] Wenling Shang, Kihyuk Sohn, and Yuandong Tian. Channel-recurrent autoencoding for image modeling. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1195–1204. IEEE, 2018.
- [120] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *arXiv preprint arXiv:1506.04214*, 2015.
- [121] Hoo Chang Shin, Neil A. Tenenholtz, Jameson K. Rogers, Christopher G. Schwarz, Matthew L. Senjem, Jeffrey L. Gunter, Katherine P. Andriole, and Mark Michalski. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11037 LNCS:1–11, 2018. [arXiv:1807.10225](https://arxiv.org/abs/1807.10225), [doi:10.1007/978-3-030-00536-8_1](https://doi.org/10.1007/978-3-030-00536-8_1).
- [122] Mengying Shu. Deep learning for image classification on very small datasets using transfer learning. 2019.
- [123] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [124] Jake Snell, Karl Ridgeway, Renjie Liao, Brett D Roads, Michael C Mozer, and Richard S Zemel. Learning to generate images with perceptual similarity metrics. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 4277–4281. IEEE, 2017.
- [125] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In *Advances in neural information processing systems*, pages 3483–3491, 2015.
- [126] Subbhuraam Vinitha Sree, Eddie Yin-Kwee Ng, Acharya Rajendra U, and Oliver Faust. Breast imaging: A survey. *World Journal of Clinical Oncology*, 2(4):171–178, 2011. [doi:10.5306/wjco.v2.i4.171](https://doi.org/10.5306/wjco.v2.i4.171).
- [127] A Thomas Stavros. *Breast ultrasound*. Lippincott Williams & Wilkins, 2004.
- [128] Bernhard Stimpel, Christopher Syben, Tobias Würfl, Katrin Mentl, Arnd Dörfler, and Andreas Maier. MR to X-Ray Projection Image Synthesis. 2017. URL: <http://arxiv.org/abs/1710.07498>, [arXiv:1710.07498](https://arxiv.org/abs/1710.07498).
- [129] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953, 2015.
- [130] J Suckling, J Parker, D Dance, et al. The mias digital mammogram database. In *Proc. of the 2nd Int. Workshop on Digital Mammography, England*, 1994.
- [131] João F. Teixeira, Sílvia Bessa, Pedro F. Gouveia, and Hélder P. Oliveira. A Framework for Fusion of T1-Weighted and Dynamic MRI Sequences. In Aurélio Campilho, Fakhri Karray, and Zhou Wang, editors, *Image Analysis and Recognition*, pages 157–169, Cham, 2020. Springer International Publishing. [doi:10.1007/978-3-030-50516-5_14](https://doi.org/10.1007/978-3-030-50516-5_14).
- [132] João F. Teixeira, Mariana Dias, Eva Batista, Joana Costa, Luís F. Teixeira, and Hélder P. Oliveira. Adversarial data augmentation on breast mri segmentation. *Applied Sciences*,

- 11(10), 2021. URL: <https://www.mdpi.com/2076-3417/11/10/4554>, doi:10.3390/app11104554.
- [133] Lucas Theis, Aäron van den Oord, and Matthias Bethge. A note on the evaluation of generative models. *arXiv preprint arXiv:1511.01844*, 2015.
- [134] Tímea Tőkés, László Torgyík, Gyöngyvér Szentmártoni, Krisztián Somlai, Andrea Tóth, Janina Kulka, and Magdolna Dank. Primary systemic therapy for breast cancer: Does the patient’s involvement in decision-making create a new future? *Patient education and counseling*, 98(6):695–703, 2015.
- [135] Ilya O Tolstikhin, Sylvain Gelly, Olivier Bousquet, Carl-Johann Simon-Gabriel, and Bernhard Schölkopf. Adagan: Boosting generative models. In *Advances in Neural Information Processing Systems*, pages 5424–5433, 2017.
- [136] Ceren Guzel Turhan and Hasan Sakir Bilge. Recent Trends in Deep Generative Models: A Review. *UBMK 2018 - 3rd International Conference on Computer Science and Engineering*, (September 2018):574–579, 2018. doi:10.1109/UBMK.2018.8566353.
- [137] Salman Ul, Hassan Dar, Mahmut Yurt, Levent Karacan, Aykut Erdem, Erkut Erdem, and Tolga Çukur. Image Synthesis in Multi-Contrast MRI with Conditional Generative Adversarial Networks. *IEEE Transactions on Medical Imaging*, 90(312):1–1, 2018.
- [138] Dominic Zeng Wang and Ingmar Posner. Voting for voting in online point cloud object detection. *Robotics: Science and Systems*, 11, 2015. doi:10.15607/RSS.2015.XI.035.
- [139] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807, 2018.
- [140] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.
- [141] J. S. Wheat, S. Choppin, and A. Goyal. Development and assessment of a Microsoft Kinect based system for imaging the breast in three dimensions. *Medical Engineering and Physics*, 36(6):732–738, 2014. URL: <http://dx.doi.org/10.1016/j.medengphy.2013.12.018>, doi:10.1016/j.medengphy.2013.12.018.
- [142] Tom White. Sampling generative networks: Notes on a few effective techniques. *CoRR*, abs/1609.04468, 7, 2016.
- [143] Terese Winslow. Anatomy of the female breast., 2011. Accessed on October 13 2020. URL: <http://www.ncbi.nlm.nih.gov/pubmedhealth/PMHT0021901/?figure=1>.
- [144] Jelmer M. Wolterink, Anna M. Dinkla, Mark H.F. Savenije, Peter R. Seevinck, Cornelis A.T. van den Berg, and Ivana Išgum. Deep MR to CT synthesis using unpaired data. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10557 LNCS:14–23, 2017. arXiv:1708.01155, doi:10.1007/978-3-319-68127-6_2.

- [145] Jiajun Wu, Chengkai Zhang, Tianfan Xue, William T. Freeman, and Joshua B. Tenenbaum. Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. *Advances in Neural Information Processing Systems*, (Nips):82–90, 2016. [arXiv:1610.07584](#).
- [146] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3D ShapeNets: A deep representation for volumetric shapes. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June-2015:1912–1920, 2015. [arXiv:1406.5670](#), [doi:10.1109/CVPR.2015.7298801](#).
- [147] Jianwen Xie, Yifei Xu, Zilong Zheng, Song-Chun Zhu, and Ying Nian Wu. Generative PointNet: Energy-Based Learning on Unordered Point Sets for 3D Generation, Reconstruction and Classification. 2020. URL: <http://arxiv.org/abs/2004.01301>, [arXiv:2004.01301](#).
- [148] Qiangeng Xu, Zengchang Qin, and Tao Wan. Generative cooperative net for image generation and data augmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11471 LNAI(1):284–294, 2019. [arXiv:1705.02887](#), [doi:10.1007/978-3-030-14815-7_24](#).
- [149] Xiangyu Xu, Deqing Sun, Jinshan Pan, Yujin Zhang, Hanspeter Pfister, and Ming Hsuan Yang. Learning to Super-Resolve Blurry Face and Text Images. *Proceedings of the IEEE International Conference on Computer Vision*, 2017-October:251–260, 2017. [doi:10.1109/ICCV.2017.36](#).
- [150] Xiaomei Zhao, Yihong Wu, Guidong Song, Zhenye Li, Yazhuo Zhang, and Yong Fan. A deep learning model integrating fcnn and crfs for brain tumor segmentation. *Medical image analysis*, 43:98–111, 2018.
- [151] Zhiming Zhou, Han Cai, Shu Rong, Yuxuan Song, Kan Ren, Weinan Zhang, Yong Yu, and Jun Wang. Activation maximization generative adversarial nets. *arXiv preprint arXiv:1703.02000*, 2017.
- [152] 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*, pages 2223–2232, 2017.
- [153] Jun Yan Zhu, Richard Zhang, Deepak Pathak, Trevor Darrell, Alexei A. Efros, Oliver Wang, and Eli Shechtman. Toward multimodal image-to-image translation. *Advances in Neural Information Processing Systems*, 2017-December(1):466–477, 2017. [arXiv:1711.11586](#).
- [154] Hooshiar Zolfagharnasab, Sílvia Bessa, Sara P. Oliveira, Pedro Faria, João F. Teixeira, Jaime S. Cardoso, and Hélder P. Oliveira. A regression model for predicting shape deformation after breast conserving surgery. *Sensors (Switzerland)*, 18(1):1–29, 2018. [doi:10.3390/s18010167](#).