Longitudinal monitoring of metastatic breast cancer through PET image registration and segmentation based on trained and untrained networks

Constance Fourcade

To cite this version:

Constance Fourcade. Longitudinal monitoring of metastatic breast cancer through PET image registration and segmentation based on trained and untrained networks. Signal and Image processing. École centrale de Nantes, 2022. English. NNT : 2022ECDN0029 . tel-03895340

HAL Id: tel-03895340
https://theses.hal.science/tel-03895340v1

Submitted on 12 Dec 2022

THÈSE DE DOCTORAT DE

L'ÉCOLE CENTRALE DE NANTES

ÉCOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies
de l'Information et de la Communication
Spécialité : *Signal, Image, Vision*

Par
Constance FOURCADE

Suivi de l'évolution du cancer du sein métastasé via le recalage et la segmentation d'images TEP en utilisant des réseaux entraînés et non-entraînés

Thèse présentée et soutenue à l'Ecole Centrale de Nantes, le 27 juin 2022
Unité de recherche : UMR 6004, Laboratoire des Sciences du Numérique de Nantes (LS2N)

Rapporteurs avant soutenance :

Caroline PETITJEAN	Professeure des universités, Université de Rouen Normandie
Vincent NOBLET	Ingénieur de recherche HDR, Université de Strasbourg

Composition du Jury :

Présidente :	Caroline PETITJEAN	Professeure des universités, Université de Rouen Normandie
Examinateurs :	Irène BUVAT	Directrice de recherche CNRS, Institut Curie, Orsay
	Mathieu HATT	Directeur de recherche, Université Bretagne Occidentale
	Nicolas DUCHATEAU	Maître de conférences, Université Claude Bernard Lyon 1
Directrice de thèse :	Diana MATEUS LAMUS	Professeure des universités, Ecole Centrale de Nantes
Encadrants :	Mathieu RUBEAUX	Ingénieur de recherche, Keosys, Saint-Herblain
	Ludovic FERRER	Ingénieur de recherche, ICO, Saint-Herblain

Acknowledgement

I would first like to say thank you to Pr. Diana Mateus for directing my thesis. It was a real pleasure to work with her these past three years and a half. She provided invaluable support and encouragement, and constantly gave me new ideas and insight to help me move my research forward. Muchas gracias!

I am also grateful to Dr. Mathieu Rubeaux for supervising and guiding me during my thesis. He helped me obtain organized and sugnificant results. Without his perseverance, the amount of available data to process would have been much smaller. Merci beaucoup!

Sincere thanks to Dr. Ludovic Ferrer for generously sharing with me his medical expertise and knowledge of PET images and acquisition systems.

My thanks also go out to Drs Caroline Petitjean, Irène Buvat, Vincent Noblet, Nicolas Duchateau and Mathieu Hatt for accepting to be members of my jury and for giving me insightful advice and suggestions to help me build upon my findings.

My research would not have been possible without the assistance of Drs. Caroline Rousseau, Marie Lacombe and Vincent Fleury from the Institut de Cancérologie de l'Ouest of Nantes and Angers. The manual segmentation of images used during my thesis took them a considerable amount of time, and I am truly grateful for their help.

I would also like to thank everyone at Keosys for all the shared moments: during the lunch breaks, the Keorunnings and our afterworks!

I especially would like to thank and congratulate Noémie Moreau, who conducted her PhD thesis alongside me. Many thanks to Aislinn Brenan for reading my papers this manuscript and for correcting my English when needed. Keep running!

I am also thankful to everyone at the LS2N lab, especially the whole SIMS team. Special thanks to the people with whom I shared an office, Aurélie Moyon for the (almost) daily lunch runs, and Gwenaël Samin for the hilarious conversations. Special thanks also to Mickael Tardy, Ludivine Moran, Vanessa Gonzalez, Guillaume Pelluet and Mira Rizkallah for all the memorable talks and fruitful brainstorming sessions!

I am lucky to be surrounded by wonderful friends, and I cannot be thankful enough for the support and encouragement they gave me during my thesis, especially during the difficult times. My friends from Limoges, from Centrale, from Spain, from Nantes, from climbing, from paragliding,... I thank you all.

I would be remiss in not mentioning my flatmates: Pauline, Octave, Ludovic and Maxime. They motivated me to go running and cooked me meals, we laughed and played games,... In short, they were my amazing second family.

A special thank you to Thibaud, who was there every step of the way. Just thank you.

Finally, a heartfelt thank my family for always being by my side. Anatole for all the video chats, Tiphaine for the meaningful conversations in difficult times, and Thibault for his open-mindedness. And of course, my parents for being so supportive even with my ever-changing and not-always-regular carrrer plans.

To all these incredible people and to all the ones I forgot to mention, thank you.

Table of Contents

Acronyms

^{18}F-FDG	^{18}F-FluoroDeoxyGlucose
AI	Artificial Intelligence
AJCC	American Joint Committee on Cancer
ANSM	Agence Nationale de Sécurité du Médicament et des produits de santé
BCS	Breast-Conservative Surgery
BRCA	BReast CAncer
CAD	Computer-Aided Diagnosis
CC	Cross Correlation
CE	Contrast-Enhanced
CG	Conjugate Gradient
CNN	Convolutional Neural Network
CPP	Comité de Protection des Personnes
CR	Complete Response
CRt	Complete Response to therapy
CT	Computed Tomography
DIP	Deep Image Prior
DL	Deep Learning
DoF	Degrees of Freedom
EORTC	European Organization for Research and Treatment of Cancer
ER	Estrogen Receptor

FFD	Free-Form Deformation
FLAIR	Fluid Attenuated Inversion Recovery
fMRI	functional Magnetic Resonance Imaging
FOV	Field Of View
FP	False Positive
GAN	Generative Adversial Network
GD	Gradient Descent
HR	Hormone Receptor
ICO	Institut de Cancérologie de l'Ouest
INC	Institut National du Cancer
LBM	Lean Body Mass
LDDMM	Large Deformation Diffeomorphic Metric Mapping
LM	Levenberg-Marquardt
LOR	Line Of Response
MAE	Mean Absolute Error
MDCT	Multi-Detector Computed Tomophraphy
MI	Mutual Information
MICCAI	Medical Image Computing and Computer Assisted Intervention
MIRRBA	Medical Image Registration Regularized By Architecture
ML	Machine Learning
MRF	Markov Random Field
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
MTV	Metabolic Tumor Volume
NCC	Normalized Cross Correlation

NN	Neural Network
OAR	Organ At Risk
OSEM	Ordered Subset Expectation Maximization
PCA	Principal Component Analysis
PD	Progressive Disease
PERCIST	Position Emission tomography Response Criteria In Solid Tumors
PET	Positron Emission Tomography
PgR	Progesterone Receptor
PMB	Physics in Medicine and Biology
PPV	Positive Predictive Value
PR	Partial Response
PREDIST	PET Residual Disease in Solid Tumor
PVE	Partial Volume Effect
QN	Quasi-Newton
RBF	Radial Basis Function
RD	Residual Disease
RECIST	Response Evaluation Criteria In Solid Tumors
RNN	Recurrent Neural Network
SD	Stable Disease
SDM	Statistical Deformation Model
SGD	Stochastic Gradient Descent
SPECT	Single-Photon Emission Computed Tomography
SSD	Sum of Squared Distances
STN	Spatial Transformer Network
SUL	Standardized Uptake value Lean body mass
SUV	Standardized Uptake Value
SVF	Stationary Velocity Field

TLG	Total Lesion Glycosis
TMI	Transactions on Medical Imaging
TNM	Tumor Node Metastasis
TOF	Time Of Flight
TP	True Positive
TPS	Thin-Plate Splines
TRE	Target Registration Error
TRG	Tumor Regression Grade
TRUS	TransRectal UltraSound
TV	Total Variation
US	UltraSound
WBIR	Workshop on Biomedical Image Registration
WHO	World Health Organization

General introduction

Metastatic breast cancer is the most common cancer among women and presents a poor prognosis. It requires constant monitoring, usually performed with ^{18}F-FluoroDeoxy-Glucose PET/CT acquisitions. On these images, physicians interpret the tumor response to treatment following standardized guidelines, such as Response Evaluation Criteria In Solid Tumors (RECIST) or Position Emission tomography Response Criteria In Solid Tumors (PERCIST). However, these guidelines tend to focus only on a selection of lesions representing tumor burden or, in the case of PERCIST, on only one lesion (the one showing the highest uptake).

Assessing the total tumor burden would be challenging and time consuming. To help physicians monitor all lesions and evaluate tumor evolution more accurately, research has notably focused on image registration methods. Indeed, if PET images acquired at different time points are precisely registered, the assessment of all lesions becomes much easier.

This work was conducted in the context of the EPICURE$_{\text{seinmeta}}$[1] project, in collaboration with the Keosys company[2] and the "Institut de Cancérologie de l'Ouest" (ICO)[3]. The EPICURE$_{\text{seinmeta}}$ project aims to collect a wide range of data regarding the disease and environment of patients presenting metastatic breast cancer.

This thesis focuses on the monitoring of metastatic breast cancer using registration and deep learning-based (DL) approaches. The main objective is to assist physicians monitor metastatic breast cancer patients with longitudinal Positron Emission Tomography (PET) images, and improve tumor evaluation by providing them with tools to consider all regions showing a high uptake.

---

[1] https://projet-epicure.fr/

[2] https://www.keosys.com/

[3] https://www.institut-cancerologie-ouest.com/

Manuscript organization

We present the clinical context, and especially the metastatic breast cancer disease as well as the main monitoring steps, in **Chapter 1**.

In **Chapter 2**, we introduce conventional and DL-based methods for medical image registration. DL approaches perform deformable registration in almost real time thanks to the benefit of large databases and computational resources. However, the slower conventional registration methods still benefit from many years of development and perform better in most applications.

On our $\text{EPICURE}_{\text{seinmeta}}$ dataset, composed of a baseline (pre-treatement) and two follow-ups (early- or post-treatment) acquisitions for each patient, we aim to monitor each individual lesion, manually delineated by ICO experts on the baseline image. As described in **Chapter 3** and in orange in Figure 1, we start with a conventional registration method (SimpleElastix [112]) to perform this longitudinal registration task. Thanks to the deformation field obtained by the registration algorithm, we are able to propagate the baseline lesion segmentations to the follow-up images. We also perform this approach on a public inspiration – expiration lung CT dataset in the context of the Learn2Reg 2020 challenge[4], held in conjunction with the MICCAI 2020 conference, which led to a conference publication [57] and a journal paper under revision for Transactions on Medical Imaging (TMI) [88]. To the best of our knowledge, on the $\text{EPICURE}_{\text{seinmeta}}$ dataset, we are the first to perform multiple lesion tracking on full body PET images with a registration method. We implemented this conventional longitudinal registration method in the software of the Keosys company to ease the lesion segmentation task on new follow-up images.

Even though performing well, the registration method we used was not specifically designed to perform accurate longitudinal image registration and lesion propagation. To address the above issue, we propose in **Chapter 4** a new registration approach called *MIRRBA* for *Medical Image Registration Regularized By Architecture.* As illustrated in blue in Figure 1, MIRRBA is a hybrid approach aiming to bridge the performance gap between conventional and DL registration methods using a Deep Image Prior (DIP) setup. It parametrizes the deformation field between the images to register with a network inspired by state-of-the-art DL registration methods. The main difference is that this network is not optimized on a large database but solely on the pair of images to register, as

---

[4] https://learn2reg.grand-challenge.org/

done with conventional methods. This is similar to a DIP setup, except the network is trained to perform accurate registration until convergence. Thanks to this specific setup, we were able to isolate, understand and interpret the impact of different components of the network on registration, without being biased by a training database. MIRRBA is a patient-specific registration method. As presented in the paper submitted to the journal Physics in Medicine and Biology (PMB) and under review [54], our approach outperforms both conventional and DL registration methods regarding the accuracy of the lesion segmentations propagated through registration from one acquisition to the next. To evaluate the registration performance on bigger and more homogeneous objects than lesions, we also propagate high uptake organ segmentations, i.e. the brain and the bladder, with a good accuracy. As a complementary study, we register the longitudinal brain MRI dataset published in the context of the BraTSReg challenge [5] with promising results.

To further improve the propagation of segmentations, we expand MIRRBA introducing segmentation information within the pipeline. As shown in **Chapter 5**, we add this segmentation information at several levels: i) as a new segmentation term in the loss function, in the form of a Dice loss between segmentations, and ii) as extra input channels to the network, to directly integrate segmentation maps. Since traditional segmentation methods segment all lesions together, as a whole, individual lesion monitoring between acquisitions is not possible with these methods. Yet, we take advantage of the segmentation results of our published work [56, 150] to obtain segmentations of the lesions and high uptake organs on follow-up images. As illustrated in green in Figure 1, we use these follow-up segmentations to optimize our new MIRRBA method and to refine the propagated segmentations.

Finally, in **Chapter 6**, we extract several biomarkers from the propagated segmentations. These biomarkers, and especially the $\mathrm{SUL}_{\mathrm{peak}}$, show a high correlation coefficient with the biomarkers extracted from the manual segmentations. For each lesion, we compar the $\mathrm{SUL}_{\mathrm{peak}}$ values of the baseline and follow-up acquisitions. If the $\mathrm{SUL}_{\mathrm{peak}}$ is decreasing, the lesion is considered responsive and colored in green. On the contrary, if the $\mathrm{SUL}_{\mathrm{peak}}$ is increasing, the lesion is considered non-responsive (i.e. in progression) and colored in red. We finally ask physicians to use this visual tool to assess patient response with promising results.

---

[5] https://www.med.upenn.edu/cbica/brats-reg-challenge/

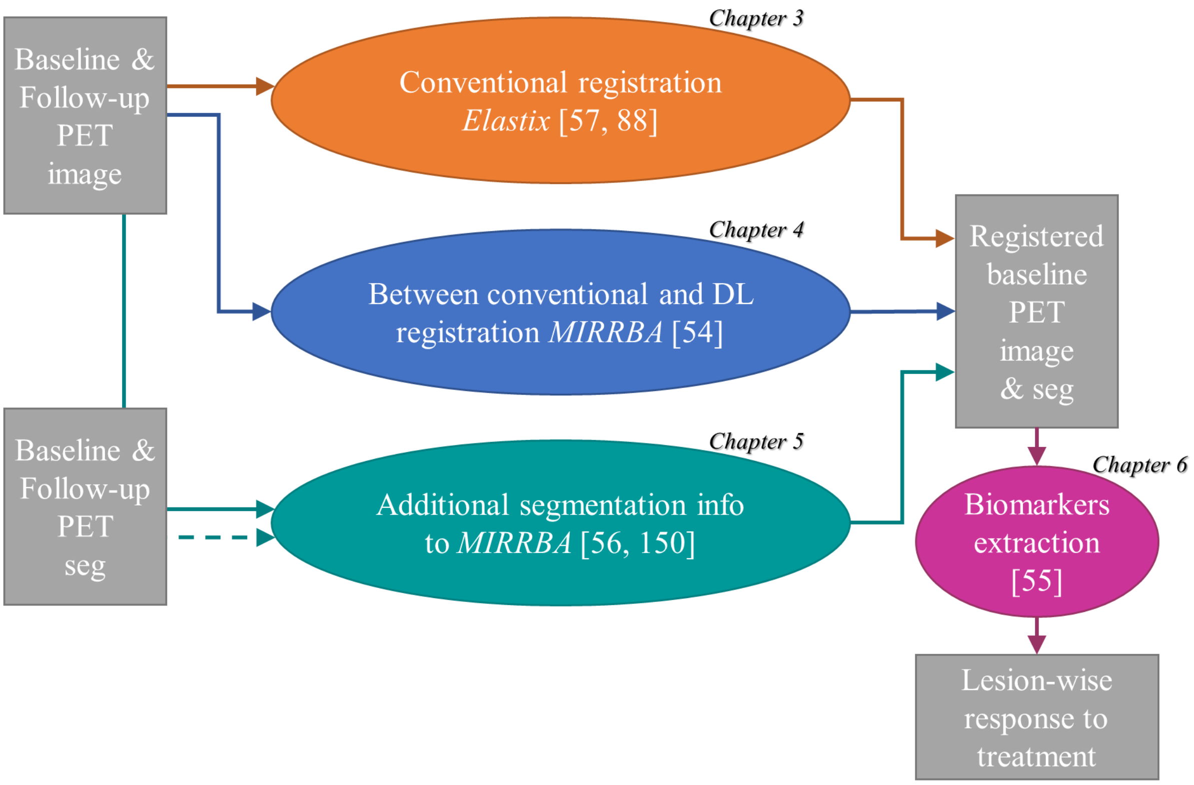


Figure 1 – Organization of the contributions. Chapters 3 (orange), 4 (blue), and 5 (green) are three registration methods. All take as input baseline and follow-up, i.e. moving and fixed, images. The method presented in Chapter 5 also takes the baseline and follow-up segmentations as input for its optimization. All three methods produce registered PET images and segmentations. From the registered segmentations, biomarkers are extracted in Chapter 6 (pink) to obtain a response to treatment for each lesion.

Summary of contributions

This thesis describes three contributions:

- In this thesis, we formulate the segmentation of lesions in follow-up images as an image registration problem. We first address longitudinal full-body PET image registration with conventional optimization-based methods [57, 88], and secondly with recent DL approaches. The first contribution of this thesis is the development of a novel method called MIRRBA (Medical Image Registration Regularized By Architecture), which combines the strengths of both conventional and DL-based approaches within a Deep Image Prior (DIP) setup. Optimizing a network (as DL-based method) on only the pair of images to register (as conventional methods), MIRRBA aims to bridge the gap between both approaches. Since the optimization of the network does not depend on a training database, we state that the architecture of the network acts as the only implicit regularizer of our registration method. We validate the three types of approaches (conventional, DL and MIRRBA) on a private longitudinal PET dataset obtained in the context of the EPICURE$_{\text{seinmeta}}$ project. Our proposed method performs better than all conventional and DL-based approaches. This contribution is currently under review for the journal PMB [54].

- The second contribution is a method for the automatic segmentation of active organs (brain, bladder, etc) based on a combination of superpixels and deep learning [56]. We show that combining superpixels and DL can improve specific medical tasks. Even if produced at the beginning of the thesis, this contribution is integrated in the overall work only at the end to improve the MIRRBA registration method.
- Finally, the third contribution is the evaluation of the biomarkers extracted from lesion segmentations obtained from the lesion registration step. We propose a protocol to evaluate tumor response in the case of a patient presenting multiple metastatic lesions. Our method provides a new tool for the monitoring of metastatic breast cancer and was accepted for an online publication at the conference ASCO 2022 [55].

Publications

The described contributions led to the following peer-reviewed publications.

Journal papers

C. Fourcade, L. Ferrer, N. Moreau, G. Santini, A. Brennan, C. Rousseau, M. Lacombe, V. Fleury, M. Colombié, P. Jézéquel, M. Campone, M. Rubeaux, and D. Mateus, « Deformable Image Registration with Deep Network Priors: a Study on Longitudinal PET Images », *in: arXiv preprint* (2021), pp. 1–25, DOI: 10.48550/ ARXIV.2111.11873. *Under major revision at Physics in Medicine and Biology (PMB)*

A. Hering, L. Hansen, T. C. W. Mok, A. C. S. Chung, H. Siebert, S. Häger, A. Lange, S. Kuckertz, S. Heldmann, W. Shao, S. Vesal, M. Rusu, G. Sonn, T. Estienne, M. Vakalopoulou, L. Han, Y. Huang, M. Brudfors, Y. Balbastre, S. Outard, M. Modat, G. Lifshitz, D. Raviv, J. Lv, Q. Li, V. Jaouen, D. Visvikis, **C. Fourcade**, M. Rubeaux, W. Pan, Z. Xu, B. Jian, F. De Benetti, M. Wodzinski, N., Gunnarsson, J. Sjölund, H. Qiu, Z. Li, C. Großbröhmer,

A. Hoopes, I. Reinertsen, Y. Xiao, B. Landman, Y. Huo, K. Murphy, N. Lessmann, B. van Ginneken, A. V. Dalca, and M. P. Heinrich, « Learn2Reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning », *in: arXiv preprint* (2021), pp. 1–15, DOI: 10.48550/ARXIV.2112.04489. *Under revision for the journal IEEE Transactions on Medical Imaging (TMI)*

N. Moreau, C. Rousseau, **C. Fourcade**, G. Santini, A. Brennan, L. Ferrer, M. Lacombe, C. Guillerminet, M. Colombié, P. Jézéquel, M. Campone, N. Normand, and M. Rubeaux, « Automatic segmentation of metastatic breast cancer lesions on ^{18}F-FDG PET/CT longitudinal acquisitions for treatment response assessment », *in: Cancers* 14.*1* (2022), DOI: 10.3390/CANCERS14010101.

Conference papers

C. Fourcade, L. Ferrer, G. Santini, N. Moreau, C. Rousseau, M. Lacombe, C. Guillerminet, M. Colombié, M. Campone, D. Mateus, and M. Rubeaux, « Combining Superpixels and Deep Learning Approaches to Segment Active Organs in Metastatic Breast Cancer PET Images », *in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* 2020-July (2020), pp. 1536–1539, DOI: 10.1109/EMBC44109.2020.9175683.

N. Moreau, C. Rousseau, **C. Fourcade**, G. Santini, L. Ferrer, M. Lacombe, C. Guillerminet, M. Campone, M. Colombié, M. Rubeaux, and N. Normand, « Deep learning approaches for bone and bone lesion segmentation on 18FDG PET/CT imaging in the context of metastatic breast cancer », *in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* 2020-July (2020), pp. 1532–1535, DOI: 10.1109/EMBC44109.2020.9175904.

C. Fourcade, M. Rubeaux, and D. Mateus, « Using Elastix to Register Inhale/ Exhale Intrasubject Thorax CT: A Unsupervised Baseline to the Task 2 of the *Learn2Reg Challenge* », *in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) for the Learn2Reg workshop* 12587 LNCS (2021), pp. 100–105, DOI: 10.1007/978-3-030-71827-5_13.

N. Moreau, C. Rousseau, **C. Fourcade**, G. Santini, L. Ferrer, M. Lacombe, C. Guillerminet, P. Jézéquel, M. Campone, N. Normand, and M. Rubeaux, « Comparison between threshold-based and deep learning-based bone segmentation on whole-body CT images », *in: Medical Imaging 2021: Computer-Aided Diagnosis* 11597 (2021), DOI: 10.1117/12.2580892.

N. Moreau, C. Rousseau, **C. Fourcade**, G. Santini, L. Ferrer, M. Lacombe, C. Guillerminet, P. Jézéquel, M. Campone, N. Normand, and M. Rubeaux, « Comparison between threshold-based and deep learning-based bone segmentation on whole-body CT images », *in: Medical Imaging 2021: Computer-Aided Diagnosis* 11597 (2021), DOI: 10.1117/12.2580892.

G. Santini, N. Moreau, **C. Fourcade**, C. Rousseau, L. Ferrer, M. Campone, M. Colombié, P. Jézéquel, and M. Rubeaux, « Quantification automatique de l'activité de fond pour le calcul du critère PERCIST », *in: Médecine Nucléaire* 45.4 (2021), pp. 213–214, DOI: 10.1016/J.MEDNUC.2021.06.080.

C. Fourcade, L. Ferrer, N. Moreau, G. Santini, A. Brennan, C. Rousseau, M. Lacombe, V. Fleury, M. Colombié, P. Jézéquel, M. Campone, M. Rubeaux, D. Mateus, B. Maucherat, and J. S. Frenel, « PERCIST-like automatic guided treatment response assessment on FDG PET images driven by full extension metastatic breast cancer », *in: American Society of Clinical Oncology (ASCO) conference* (2022)

Chapter 1

Metastatic breast cancer monitoring: clinical context

Contents

1.1 Breast cancer management

According to the World Health Organization (WHO) [227], breast cancer is the most frequent cancer among women, and was the cause of 15% of all female cancer deaths worldwide in 2018. According to the French “Institut National du Cancer” [96], if breast cancer is detected early enough, the five years survival rate is 99%. However, breast cancer discovered at a later stage, when already presenting metastases, has a poor prognosis with a survival rate of 25% (prognosis being the prospect of recovery from a disease [96]). Major breast cancer risk factors include sex (breast cancer is roughly 100 times more common in women than men), age (with increasing age, cells are more likely to mutate and cause cancer), and genetics (5% to 10% of breast cancers are associated to genes linked to higher risks).

1.1.1 Breast cancer types and sub-types

Harris et al. [77] describes multiple types of breast cancers. When planning breast cancer treatment, to increase the probability of success against the disease, it is important to first find out the type of cancer. The classification is done in two steps: i) assessment of the type of the tumor according to the area of the breast affected, and ii) determination of the cancer sub-type based on the mutation at the origin of the tumor. Different breast cancer types can be diagnosed:

- **In situ ductal carcinoma:** This cancer type reaches the breast milk ducts only, i.e. ducts conducting milk from mammary glands up to the nipple. It is characterized as an early breast cancer because, at the time of the diagnosis, it is limited to the breast.
- **Invasive ductal carcinoma:** As an situ ductal carcinoma, this cancer reaches the milk ducts but is more invasive. It will grow and spread in the breast tissues and can develop metastases. It is the most common type of breast cancer.
- **In situ lobular carcinoma:** With this type of cancer, abnormal cells will develop in mammary glands, i.e. glands producing breast milk. It is not a tumor, but this type of alteration increases the risks to develop a cancer afterwards.
- **Invasive lobular carcinoma:** This cancer type is less common. It starts in the mammary glands before extending to other breast tissues. Like other invasive cancers, it can develop metastases.

- **Inflammatory breast cancer:** It is a rare breast cancer type. Symptoms usually include a red and swollen breast. This cancer is rather aggressive, growing and spreading fast.

Once the breast cancer type has been defined, its sub-type must be determined. With laboratory analysis such as blood analysis, blood cell count or liver function test, it is possible to determine the mutation causing the development of the tumor:

- **Hormone Receptor positive (HR+):** Cell membranes have receptors enabling them to receive messages from hormones to adapt their function. In the breast, Estrogen Receptor (ER) and Progesterone Receptor (PgR) can be found. With specific analysis, it is possible to determine if the tumor is caused by an hyper-activation of the response of these receptors. These tumors are respectively qualified as ER-positive or PgR-positive. More generally, these cancers are called hormone-sensitive. Cancers with positive hormonal receptors are more frequent in menopaused women. This cancer sub-type represents 60% to 75% of all breast cancers.
- **HER2-positive (HER2+):** This sub-type represents 15% to 30% of all breast cancers. HER2 is a receptor located on the surface membrane that, when activated, transmits a growth signal to the cell. For this cancer sub-type, the HER2 receptor is overexpressed on the cell membrane and is generally over-activated. Growth signals transmitted to the cell are therefore multiplied excessively.
- **Triple negative:** When the analysis does not reveal any mutation on the above cited genes (ER, PgR or HER2), the tumor is “triple negative”. Triple negative breast cancers represent 15% of invasive breast cancers and are more frequent among young women [53].
- **BRCA (BReast CAncer):** Most breast cancers are not hereditary, except for the breast cancer sub-type BRCA. This tumor sub-type is less frequent as it represents only 5% to 10% of all cases. It is caused by a gene called BRCA (existing under the BRCA1 or BRCA2 form). People carrying this gene mutation present a predisposition to develop breast cancer.

1.1.2 Diagnosis of breast cancer

To deliver the best suited treatment to a patient, physicians need to accurately diagnose the gravity of the breast cancer. This is done according to two major criteria: the **grade** and the **stage**.

1.1.2.1 Grade

As defined in [92], **cancer grading** is a first indicator of the possible evolution of tumors. The higher it is, the faster the cancer is likely to grow and spread. It is used to help predict the outcome (prognosis) of the disease and assist in deciding on the best treatment to administer to the patient.

The grading of a cancer is done by looking at cells through a microscope. Indeed, cancerous cells progressively lose the features of normal breast cells until they end up in a state called undifferentiated. Hence, the grading system is based on *how much the cancer cells look like normal breast cells.*

In practice, a scale ranging from 1 to 3 is used to grade the cancer:

- **Grade 1 or well differentiated:** Slow-growing cells, looking like normal breast tissue.
- **Grade 2 or moderately differentiated:** Faster growing cells, which do not completely look like normal breast tissue.
- **Grade 3 or poorly differentiated:** Cancerous cells growing and spreading fast; they look very different from normal breast cells.

1.1.2.2 Stage

Once the grade is reported, physicians then define the **stage** of the disease. Staging defines *how much the cancer has spread within the body.* Looking at lesion size, as well as local and distant metastases, the staging helps determine how serious the cancer is and, consequently, what would be the best treatment. Staging is also used to estimate survival statistics [92].

Cancer stage is defined by a value ranging from 0 to IV. The lower the number, the less the cancer has spread. To help define it, the American Joint Committee on Cancer (AJCC) Tumor Node Metastasis (TNM) system is the most commonly used. It is based on the combination of seven elements:

- **The tumor size (T):** How extended is the cancer? Has it spread to nearby tissues?
- **The lymph nodes (N):** Has the cancer reached nearby lymph nodes? If so, how many?
- **The spread (or metastases) to distant sites:** Has the cancer reached distant parts of the body, such as the lungs, the liver or the bones?

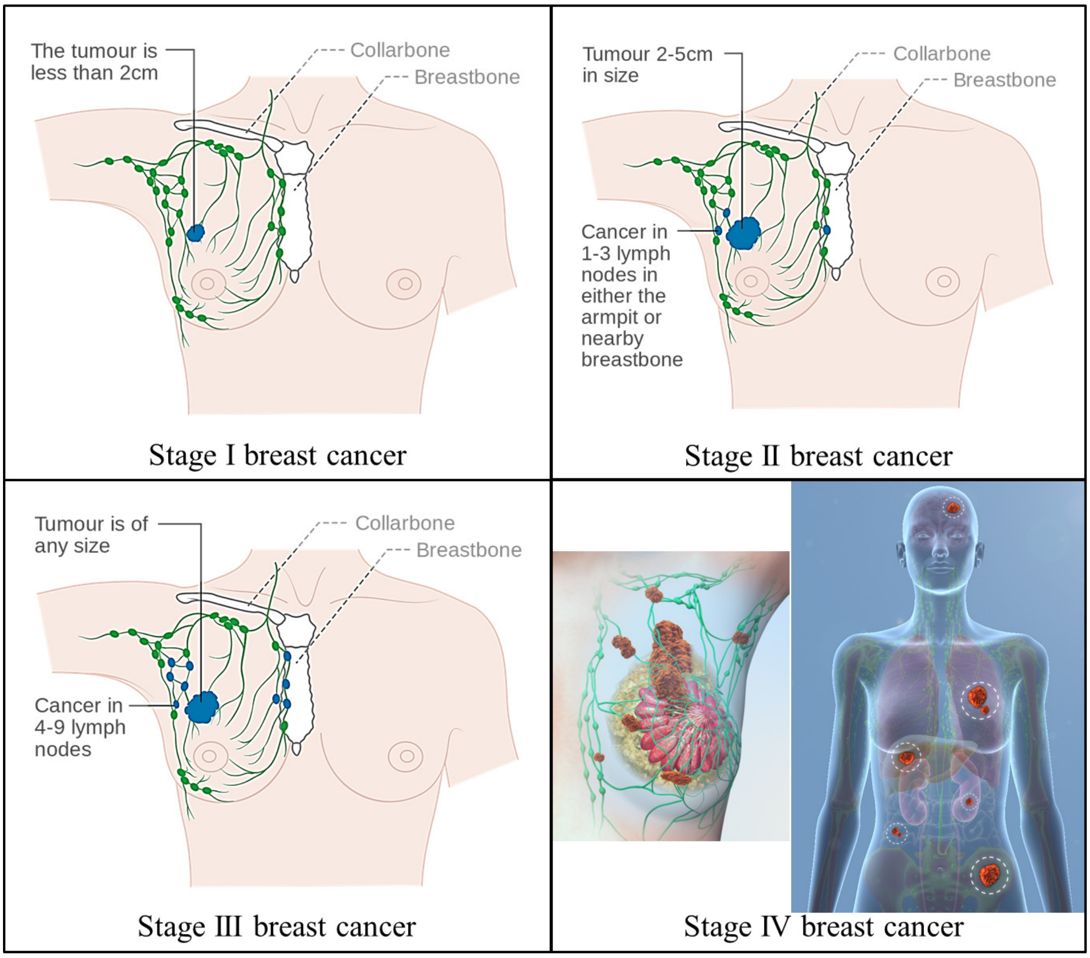


Figure 1.1 – Breast cancer stages, images adapted from an open knowledge project by Cancer Research UK.

- **The Estrogen Receptor status (ER):** Does the cancer present proteins called estrogen receptors?
- **The Progesterone Receptor status (PgR):** Does the cancer present proteins called progesterone receptors?
- **The HER2 status (HER2):** Does the cancer present an excessive production of proteins called HER2?
- **The grade of the cancer value (G):** How much do the breast cancer cells look like normal breast tissue cells?

As shown in Figure 1.1, a **stage 0** cancer will correspond to **pre-cancerous cells**, such as in-situ carcinomas. **Stages I to III** will describe **cancers spreading within**

the breast or to regional lymph nodes, while a **stage IV** cancer is a **metastatic** one. This means the cancer has spread beyond the breast and the surrounding lymph nodes to distant parts of the body. As already prsented in Section 1.1, stage IV breast cancers present worse prognosis than lower stage cancers.

1.1.3 Treatment of breast cancer

The choice of breast cancer treatment depends on the tumor extent (or stage), as well as the presence of hormone receptors (if the cancer is ER-positive or PgR-positive), the amount of HER2 expression (if the cancer is HER2-positive), the global health of the patient, the menopause status and the cancer growth rate (evaluated by the grade).

When diagnosed with a stage 0 cancer (non-invasive cancer limited to the inside of the milk duct), women can choose, in most cases, between a Breast-Conservative Surgery (BCS) or a mastectomy. They however would not need radiation nor drug therapies [158].

Treatments for stage I to III breast cancers usually include a mix of surgery and radiation therapy, associated with chemotherapy and other drug therapies [161].

Women with a metastatic (stage IV) breast cancer are mainly treated with systemic therapy [160], i.e. drugs that can reach and affect cells all over the body. This may include a combination of hormonotherapy, chemotherapy and targeted therapy. Systemic therapy is also often combined with local treatments, such as radiation or surgery. For such metastatic cancers with poor prognosis, it is important to monitor the cancer and assess its response to treatment. As presented hereafter, this can be done by acquiring at regular time points a combination of functional and anatomical images. The analysis of these images is the focus of this thesis. Indeed, this thesis is conducted in collaboration with the “Institut de Cancérologie de l’Ouest” (ICO) in the context of the EPICURE$_{\text{seinmeta}}$ project [32]. The EPICURE project focuses, among other, on stage IV metastatic breast cancer and on the evaluation of treatment response.

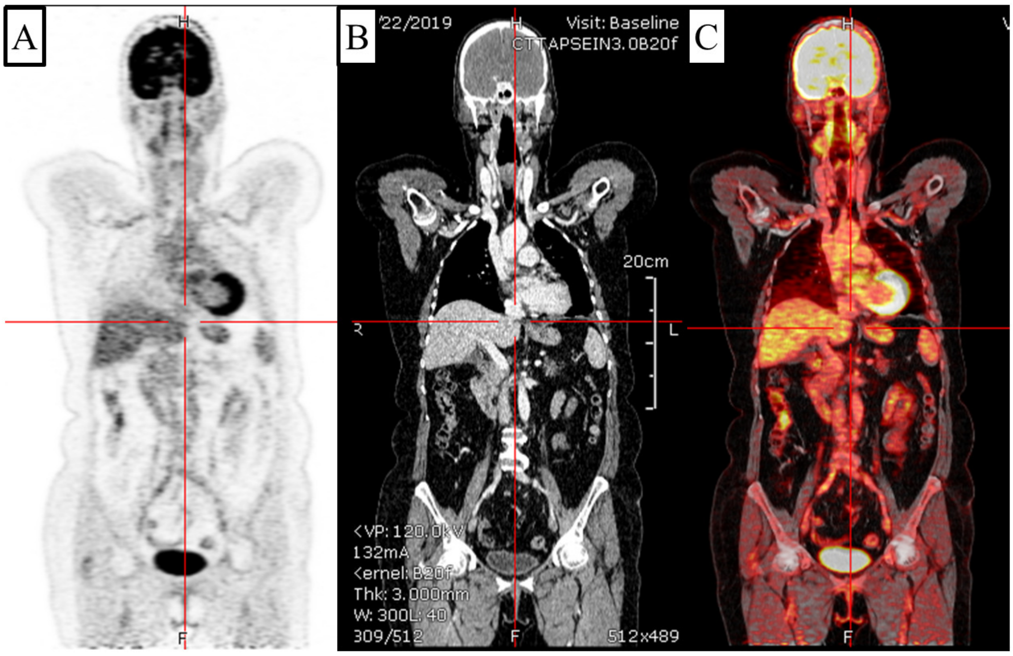
1.2 Imaging modalities for the monitoring of metastatic breast cancer

Different imaging techniques are used to diagnose breast cancer, assess the presence of metastases and evaluate response to treatment.

Since it is inexpensive, easy to set up and generates little amount of radiation, breast radiography (mammography) is usually the modality employed to screen and detect breast cancer at an early stage [59]. However, other imaging modalities such as Positron Emission Tomography (PET) are needed when physicians suspect that the cancer has spread outside the breast, as it is the study case in this thesis. Therefore, in the following sections, we describe the general principles of PET imaging (see Sections 1.2.1 and 1.3). In Section 1.4, we review current methods used to assess response to treatment with PET images.

1.2.1 PET imaging

PET is a functional medical imaging modality. It acquires data based on the body's metabolism, rather than on physical shapes and forms (see Figure 1.2.A). This imaging modality was originally used as a research tool for neurology and cardiology, but in recent years it has increasingly been integrated in clinical routine for the follow up of cancer patients. PET is currently employed to localize tumors, evaluate a cancer's stage, plan the treatment and determine its effects [19].



PET acquisition CT acquisition PET/CT acquisition

Figure 1.2 – CT (A), PET (B) and PET–CT (C) images from the EPICURE$_{\text{seinmeta}}$ study.

To acquire PET images, the patient is injected with a specific molecule, named tracer, which travels intravenously throughout the body. To be easily integrated into functional body processes, the tracer must be analogue to a molecule already present in the body. It should also present relevant physical and chemical properties, such as accumulating in tissues where the metabolic process of interest takes place, i.e. in cancerous lesions in our case [170]. To be able to quantify the local concentrations of this tracer *in vivo*, the tracer is combined with a beta+ emitting radioisotope. This combined component is named a *radiotracer*. The choice of radioisotope is important, since it should not disturb the tracer's chemical activity.

1.2.1.1 Photon emission

In the case of PET imaging, the tracer is labeled with a radioisotope emitter of positrons. This radioelement is characterized by an excess of protons in the nucleus that decays through disintegration to a stable state. The disintegration of the radioelement is accompanied by the emission of a positron and a neutrino.

After a disintegration in human tissues, an emitted positron only travels a short distance (few millimeters, depending on the radioisotope's energy spectrum and the density of the matter crossed by particles) before colliding with an electron of a nearby atom, i.e. its thermalization. This short distance is called the *positron range*. A positron being the anti-particle of an electron (same mass and opposite charge), the two particles annihilate each other to generate two 511-keV gamma-rays emitted at 180 degrees from each other (see Figure 1.3) [10]. These photons escape from the human body in opposite directions and can be recorded by external detectors (see Section 1.2.1.2).

The positron emitters used in nuclear medicine as radiotopes present short half-lives (from 1.26 min for the ^{82}Rb to 109.8 min for the ^{18}F), imposing a limited period of time between their production and use. However, these short half-lives also simplify radioprotection procedures, since the radioactivity emitted by the patient decays fast enough to pass the exposition threshold guidelines a few hours after the injection. Depending on the organs the physician wants to visualize, the type of radiotracer introduced will differ. ^{18}F-FluoroDeoxyGlucose (^{18}F-FDG) particularly suits the staging of breast cancer [64, 71].

1.2.1.2 Photon detection

PET detection consists in identifying the signal coming from two annihilation photons emitted at 180° with the same energy. These photons strike two opposite detectors, forming a Line Of Response (LOR). The location of the annihilation site is thus on this

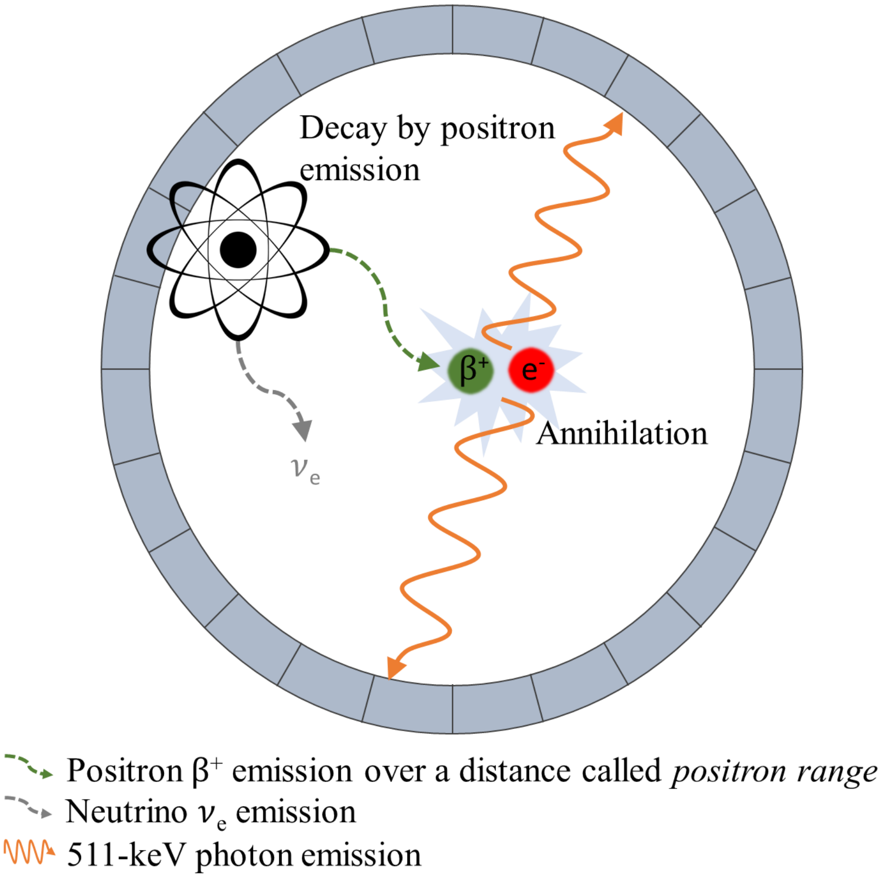


Figure 1.3 – Positron annihilation: the emitted positron collides with an electron in the patient's tissues after a distance called the positron range. Both anti-particles annihilate and generate the emission of two 511-keV photons in opposite directions (180°). A neutrino ν_e is also emitted. For readability, the figure is not on scale.

LOR. However, the distance from the positron emission site to the LOR depends on the positron range distance. The positron range is then one of the PET imaging limitations, since PET seeks to find the location of the emission of positrons from the collection of LORs, and the positron range adds an uncertainty to the LOR measurements.

The PET detection system is composed of detector blocks assembled as a ring. The complete PET system is made of several of such rings to cover a 15 - 20 cm field of view [205]. Detector blocks are composed of scintillation crystals associated with photomultiplier tubes. In scintillation crystals, photons are collected and converted into an electrical signal amplified by the photomultiplier tubes.

Specific PET systems, called Time Of Flight (TOF) PET, exploit the temporal information of two photons detection to constrain, during the reconstruction, the location of

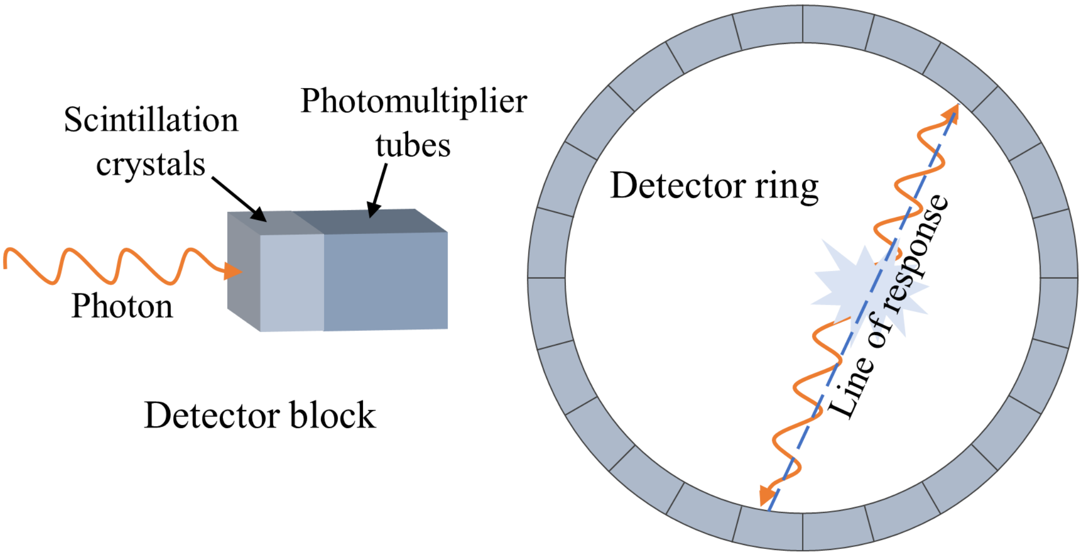


Figure 1.4 – In PET imaging systems. **Left:** Detector block composed of scintillation crystals receiving an emitted photon and photomultiplier tubes. **Right:** Detector ring composed of detector blocks. The line of response between two emitted photons is also represented.

the annihilation. Indeed, if the annihilation takes place at equal distance from the two detectors, both annihilation photons will be received simultaneously. However, if the annihilation is closer to one of the detectors, photons will be detected with a temporal gap, which brings additional information to the reconstruction and helps to better localize the emission site.

The overall quantitative information and radioactivity distribution represented by PET images are affected by several bias sources caused by the interaction between the photons and the surrounding environment, such as *attenuation* and *diffusion*. Some of these biases are addressed, in clinical practice, during the reconstruction process (see Section 1.2.1.3), or by correcting the data using information obtained with another imaging modality (see Section 1.2.2.2).

1.2.1.3 PET image reconstruction

PET reconstruction aims to generate, from all detected events on considered LORs, the tridimensional image of the radiotracer's distribution. Let f be the 3D discrete representation of the continuous object of interest. Most common PET reconstruction algorithms define R, a mathematical operator modeling the transformation between the object f and its projections p, called the system matrix. Knowing R and projections p measured in every direction, i.e. LOR, the log-likelihood $\mathcal{L}$ is maximized to compute f:

$$\begin{cases} \mathcal{L}(f) & = \sum_{i=1}^{I}(p_i \log(\overline{p}_i(f)) - \overline{p}_i(f)) \\ \overline{p}(f) & = Rf \end{cases} \tag{1.1}$$

The radiotracer's activity distribution showing the best correspondence with the measured projections is then identified, voxel by voxel. The inversion of the projection matrix R is not possible: the inverse problem is ill-posed, given the acquisition's limited angular sampling and the measurement's noise. Thus, there is no unique solution.

There is a vast literature on PET image reconstruction [172]. Algorithms used to achieve the reconstruction task can be divided into two main types: analytic and iterative. Most reconstruction methods used in clinical practice are based on iterative approaches [69]. To retrieve f, the distance between projections that would match the estimated activity distribution and the effectively acquired projections is defined with the log-likelihood defined in Eq. 1.1 and iteratively optimized. The system matrix R can model not only the geometry of detectors, but also physical phenomenons such as the positron range, the attenuation or the limited spatial resolution.

PET exams presented and used for analysis in the following parts of this manuscript are all reconstructed using the iterative Ordered Subset Expectation Maximization (OSEM) algorithm [237], taking in consideration TOF.

1.2.1.4 PET image limitations

Even though PET has become a standard imaging modality in oncology, it presents several limitations inherent to the acquisition process, which reduce the accuracy of the estimated tumoral activity. These limitations raise challenges that are addressed by the problematics of this thesis, as we are interested in delineating organs and lesions on PET images to help cancer monitoring.

Spatial resolution, sampling and partial volume effects PET images present poor spatial resolution (around 4 mm per voxel), while anatomic imaging modalities such as Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) can reach submillimetric spatial resolutions.

Indeed, PET image resolution depends on the spatial sampling of the acquired projections, themselves depending on the field of view's width and orientation, as well as on the location of each detector block. In addition, due to the positron's range (few millimeters), it is not possible to know the exact location of the positron emission, reducing PET image precision.

A sampling effect is added to the limited spatial resolution of the imaging system: PET image voxels used in clinical practice are about 4 mm cubes. Thus, the detected intensity information contained in a voxel can come from different tissues and tissue borders are poorly defined. This phenomenon is called the *tissue fraction effect*.

The combined effects of poor spatial resolution and rough image sampling are called *Partial Volume Effect (PVE)* [198]. PVE causes a mutual contamination of structures that have different activities. For instance, if high tracer uptake tissues are close to a structure of interest, a part of this high activity could artificially increase the structure's visible activity. The PVE impact on how a lesion is represented on an image depends on the object's size and compactness [198]. Small structures are more subject to PVE, especially those smaller than three times the spatial resolution. Various methods have been developed to correct the PVE [198, 201], applied either on raw data or on reconstructed images.

Noise PET images present an important noise level, making cancerous volumes difficult to precisely segment. The noise first comes from the random nature of radioactive decays and is related to the number of photons detected. It is usually modeled as a Poisson distribution. The chosen reconstruction algorithms also impact PET image noise as they modify its statistical distribution.

Patient motion In clinical practice, full body PET exams can last for about 20 min. Thus, the resulting PET image represents the average activity distribution acquired over several respiratory cycles. When a patient breathes, the motion of the chest, lungs and diaphragm results in additional blurring of the reconstructed images. Breathing motion is inherent to the modality acquisition process and makes the detection of lesions in the chest area challenging. Moreover, respiratory motion is different for every patient, and even for a given patient thoracic structures may deform non-homogeneously. Despite the fact that approaches based on the synchronization of gated PET acquisitions with the respiratory cycle have been developed [159, 219], PET–CT systems are currently used to help localize more precisely the lesions in blurred area, as shown in Section 1.2.2.2 [165].

1.2.2 PET–CT imaging

All PET images used for analysis were acquired with hybrid PET–CT systems. We present hereafter the working principle of CT imaging systems alone, before presenting hybrid systems.

1.2.2.1 CT imaging system

CT is the 3D generalization of X-ray and was the first imaging modality enabling the volumetric exploration of the body. Its first version was developed in 1972 with a brain scanner [61].

Through several technological and computational improvements over the years, the fourth CT scan generation was developed in 2000 (see in Figure 1.5). This Multi-Detector Computed Tomophraphy (MDCT) scan is integrated into clinical routine to visualize different body elements such as bones, organs or soft tissues.

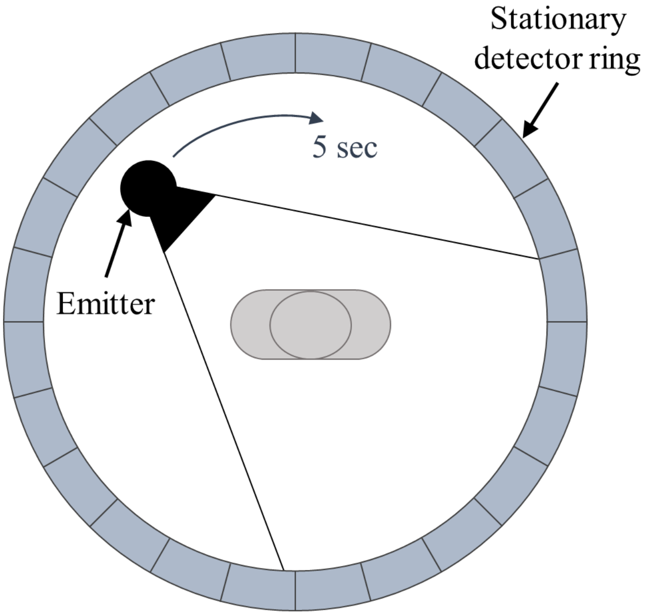


Figure 1.5 – 4th generation of CT systems, with a stationary detector ring and an emitter in continuous rotation.

To obtain a CT scan, the patient lies down on a table that moves inside the CT system. The main physical processes of CT are *attenuation*, *projection* and *image reconstruction*.

The *attenuation* is defined by the relation:

$$\log\left(\frac{I_0}{I}\right) = \nu x \tag{1.2}$$

with I_0 and I respectively the input (before crossing the body) and output (after crossing the body) beam intensities, ν the attenuation coefficient and x the object thickness. The input beam is sent by an X-ray tube. This emitter is a rotating anode with a thin source enabling the emission of a stable, continuous and homogeneous X-ray beam during the acquisition. The rotating MDCT system X-ray source spins in less than 500 ms around the patient, and makes slices thinner than 1 mm.

The *projection* is performed by the detector, which is made of a combination of scintillators, photo-diodes and photo-amplificators transforming the attenuated output X-ray beam into an amplified electrical signal proportional to the beam intensity.

Finally, the *image reconstruction* step solve an inverse problem to compute tomographic images from the amplified electrical signals. Each pixel of the tomographic image corresponds to an attenuation value and is measured in Hounsfield units (HU), which is related to the linear attenuation factor for a given X-ray tube energy spectrum. The CT scan being a transmission imaging modality, tissues are differentiated according to their attenuation properties. A windowing processing is applied to better visualize specific tissues, since $\nu_{water} = 0\text{HU}$, $\nu_{bone} = 1000\text{HU}$, and $\nu_{air} = -1000\text{HU}$ (see in Figure 1.6). Given their ability to differentiate tissues by their densities, CT scanners are categorized as **anatomical imaging systems** (see Figure 1.2.B).

Limited by mechanical constraints on spinning elements, the emitter rotation speed is currently 270 - 350 ms for most systems and the temporal resolution is 135 to 175 ms. CT temporal resolution depends on the displacement speeds of studied structures. Since the whole image acquisition process lasts around 15 min, the patient should remain as still as possible to obtain clear and informative images.

Even if CT is an irradiating imaging modality which often requires iodine injections to improve the contrast, the multiple evolutions from which it has benefited makes it a very efficient imaging system. CT is currently the most versatile diagnostic tool to visualize anatomical information. The advantages of CT are its high spatial resolution (less than 1 mm), almost isotropism, technical accessibility and moderate cost compared to alternative methods.

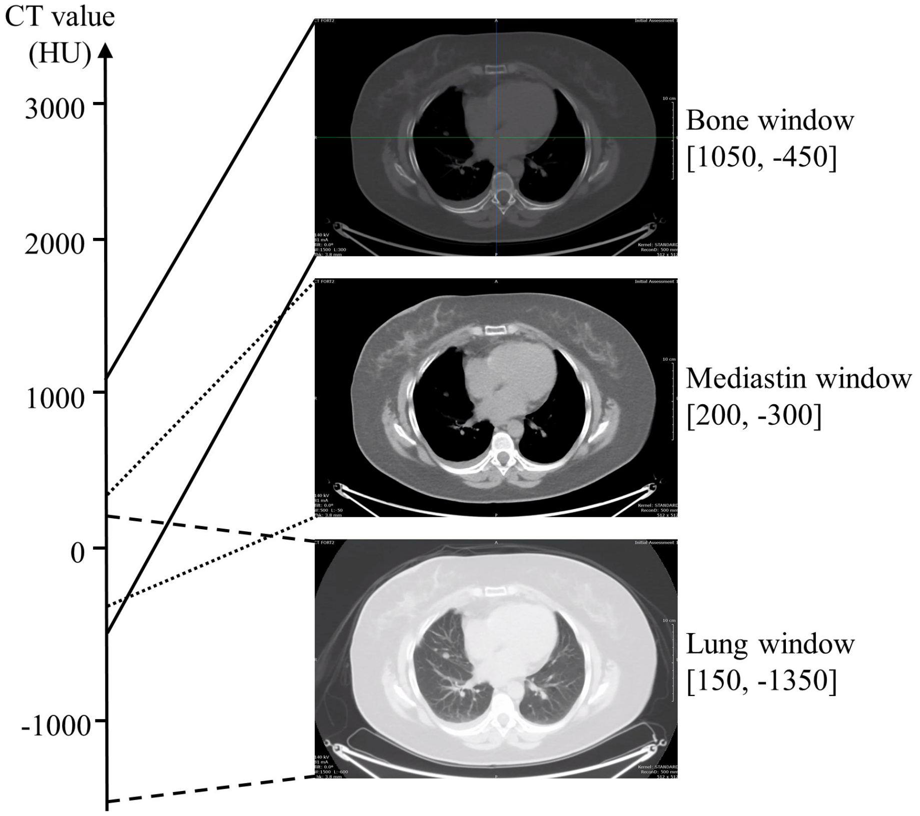


Figure 1.6 – Diagnostically relevant range of CT Hounsfield values depending on the organs to visualize. Adapted from Kalender [105] with EPICURE images.

1.2.2.2 Hybrid PET–CT imaging

PET–CT principle Since ^{18}F-FDG PET scans only allow to visualize areas of high FDG radiotracer uptake, it is considered as a **functional or metabolic imaging modality**, as visible in Figure 1.2.A. Because of their low spatial resolution, the precise location of objects of interest and the detection of very small tumors ($<$ 3 - 4 mm) is challenging (see Section 1.2.1.4). Therefore, combining PET and CT information [58] is a popular solution (see figure 1.2.C). Indeed, the combination provides the physician with information on areas of increased cell activity (with PET), as well as details on the structures found within these areas (with CT).

When anatomical and functional images are acquired on different systems, their combined use is only possible if they are precisely overlaid, i.e. registered, which presents a major challenge. In contrast, hybrid PET–CT systems acquire both exams sequentially, with the patient laying still on a single table (see Figure 1.7). Such hybrid systems provide a good correlation between the PET functional information and the CT anatomical information [4, 106]. The CT information acquired jointly with the PET helps correct some of PET image main drawbacks, such as image attenuation or the Partial Volume Effect (see Section 1.2.1.4).

PET–CT and cancer monitoring In the context of cancer evaluations, the areas explored by PET–CT hybrid system depends on the location of the primary tumor: for head and neck lesions the exam usually includes the views from the head to the abdomen, but for most other cancers the acquisition covers the area from the top of the head to the knees. For melanoma patients, the whole body is explored. After acquisition, images are automatically reconstructed and interpreted by a physician.

The fusion of both types of information allows to better distinguish a benign from a malign FDG fixation, as well as to localize and measure lesions without the need for an invasive intervention. Moreover, by comparing PET–CT scans acquired longitudinally, i.e. at different time points, physicians can follow the cancer evolution during the patient treatment. PET–CT scans can also help radiotherapists in positioning target lesions during a treatment plan, and are also useful for early detection of potential relapse.

Figure 1.7 – PET–CT sequential scanner. Image taken at the Nantes ICO.

1.3 Metabolic activity on PET images

1.3.1 ^{18}F-FDG radiotracer

In oncology, ^{18}F-FDG is a widely-used radiotracer to diagnose and visualize the extension of a cancer, including breast cancer [4, 64, 71].

As previously presented in Section 1.2.1, the PET image principle relies on the accumulation of a radiotracer in tissues where the metabolic process of interest takes place. In the case of cancer, the metabolic process of interest is the consumption of glucose. Indeed, body cells take in different amounts of sugar, depending on how fast they are dividing. Since cancerous cells present a high division activity, they consume an important amount of glucose compared to their healthy neighbors. When using a molecule analogue to glucose combined with a radiotope, tissues presenting an important metabolic activity show up on PET images as “hot spots” [60], enabling the analysis of the metabolic process in-vivo.

When injected into the patient’s body, the ^{18}F-FDG radiotracer is transported into cells. There, instead of being degraded by the glycolysis chemical reactions as the glucose molecule would normally undergo, the FDG tracer will remain trapped in the cells and accumulate. In order to let the FDG accumulate, a delay of an hour is respected between radiotracer injection and PET image acquisition.

The FDG tracer is consumed by all cells and accumulates within tissues presenting a high glucose uptake [243]: the amount of trapped ^{18}F-FDG is proportional to their uptake. Even though cancerous cells present a very disturbed metabolic activity with an inefficient glycolysis, this behavior is balanced by a high concentration of glucose receptors on their membrane [195]. For these reasons, cancerous but also brain cells present a high FDG uptake. The FDG accumulation within the patient’s heart depends on the insulin level and the fatty acid intake: after a fasting period, the heart degrades acids rather than glucose, thus presents a low FDG accumulation. In clinical practice, it is recommended to fast for 4 to 6 hours before a PET exam to limit artifacts caused by a strong heart FDG uptake. Since the radiotracer is eliminated through urine, the kidneys and the bladder may also present high fixation. The FDG uptake of other organs, such as the liver or the bowel, is limited but varies between patients. Cells presenting infections or inflammatory processes may also present a high radiotracer accumulation.

While the accumulation of FDG in cancerous cells is called *pathological hyperfixations*, other uptakes in the body are named *physiological hyperfixations*. Since physicians aim at identifying pathological hyperfixations on PET exams, the presence of physiological hyperfixations due to the non-specificity of the FDG tracer complicates the automatic completion of the task. We treat this problem in Chapter 5, where we propose an automatic method to segment these high uptake organs. Additionally, some cancer cells can sometimes not be identified by FDG uptake on PET exams because of their low glucose consumption. See Figure 1.8 for a visual representation of pathological and physiological hyperfixations on a PET image.

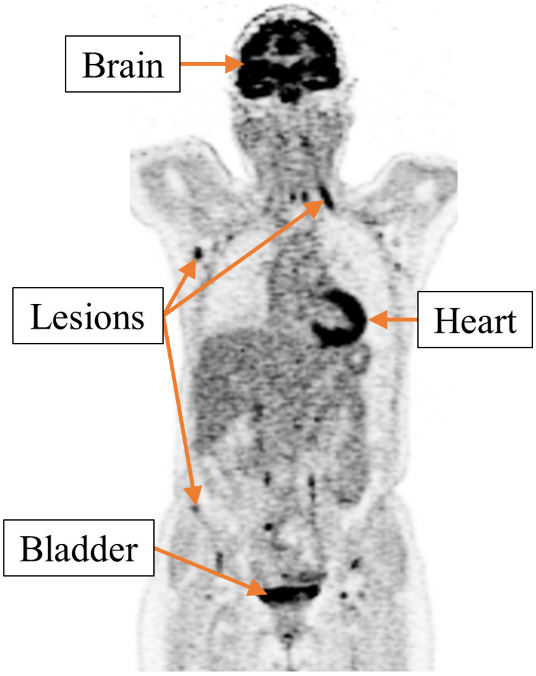


Figure 1.8 – PET high ^{18}F-FDG uptake sites: brain, heart, bladder and lesions. Image from the EPICURE$_{\text{seinmeta}}$ study.

1.3.2 Tumoral metabolic activity

To quantitatively characterize ^{18}F-FDG metabolic activity, several scores have been proposed and are used in clinical routine. The Standardized Uptake Value (SUV) was defined for every tracer and is the most used semi-quantitative index [109, 228]. Working with SUV values instead of image intensities allows to compare values across acquisitions. The SUV is defined as the ratio of the mean radiotracer activity in a defined area A_{local_mean} (kBq/mL), over the injected ^{18}F-FDG activity at acquisition time corrected by the radioactive decay A_{total} (kBq), itself normalized by the body volume V (mL) (see

Eq. 1.3). The SUV is unitless.

$$SUV = \frac{A_{local\_mean}}{\frac{A_{total}}{V}} \tag{1.3}$$

Since the SUV is easy to compute, it is available on all interpretation consoles used in clinical routine. However, SUV relies on several approximations.

SUV is computed from a single uptake measure, which depends on the post-injection duration. Yet, ^{18}F-FDG uptake only reaches a plateau two hours after the injection [75], while SUV is generally measured 1h after the injection, at a time of increasing tracer concentration in the body. Besides, SUV measures the ^{18}F-FDG metabolized by the tumor as well as the circulating one, which leads to an overestimation of the glucose metabolized by the lesions [62]. SUV also does not take into account the competition between glucose and ^{18}F-FDG. To reduce the influence of blood sugar on the lesion's tracer uptake [117], patients are asked to fast before PET acquisition, though glucose levels can remain high if patients suffer from severe diabetes.

The uniform tracer dilution hypothesis is also a coarse approximation. The tracer body distribution can noticeably vary depending on the patient's body composition, especially its adipose cell proportion. Since fat presents a very low metabolic activity, lesions of a patient presenting a high fat ratio will have a higher ^{18}F-FDG uptake than lesions from another patient. To take the patient's body composition in consideration, Morgan et al. [154] and Sugawara et al. [203] proposed to replace the volumetric normalization in Eq. 1.3 by a mass normalization. The Lean Body Mass (LBM) is used to perform this normalization (Standardized Uptake value Lean body mass (SUL)), as detailed in Eq. 1.4 and 1.5, with *weight* in kg and *size* in m.

$$\text{Woman:} \quad LBM = 1.07 \cdot weight - 148 \cdot \left(\frac{weight}{size}\right)^2 \tag{1.4}$$

$$\text{Man:} \quad LBM = 1.10 \cdot weight - 120 \cdot \left(\frac{weight}{size}\right)^2 \tag{1.5}$$

Indeed, considering the patient volumetric mass is identical to that of water (1 g/cm^3), the volume value in mL is the same as the mass in g. The body surface area (SUV$_{\text{bsa}}$) could also be used to normalize the activities ratio.

Despite all studies focusing on SUV, no consensus has emerged about its optimal computation area [132]. It can be computed on a fixed region around the lesion maximum

(SUV_{mean}) or in the maximum intensity voxel (SUV_{max}), and this choice has a major impact on the estimated value [20]. For instance, SUV_{mean} computed on manually defined fixed areas with a 50% or 75% threshold from the SUV_{max} can be respectively 35% or 15% lower than the SUV_{max} [114].

To address these disagreements, standardized guidelines have been proposed to measure the tumoral metabolic activity homogeneously from one center to another (see Section 1.4). These guidelines are widely used in clinical routine to monitor treatment response to cancer.

1.4 Response to treatment evaluation

1.4.1 Evaluation criteria principle

Response evaluation helps determine if a patient benefits from a treatment [23], allowing to regularly reevaluate prognosis during therapy. With imaging, cancer evolution at a lesion level can be observed even before a change in symptoms is reported. Evaluation criteria have therefore been established to measure treatment response from image data. These criteria should be quantitative, objective and reproducible. Criteria allow to categorize patients in three main categories:

- **Complete Response (CR)** or **Partial Response (PR)**: Patients presenting an objective response to treatment;
- **Progressive Disease (PD)**: Patients presenting a tumoral progression;
- **Stable Disease (SD)**: Patients presenting neither response nor progression.

CR, PR or SD patients benefit from the treatment: physicians consider that without it, tumoral progression would have been observed. PD patients do not benefit from the treatment. In these cases, the treatment is then suspended and adapted. During a clinical trial, evaluation criteria can also be used to characterize the efficiency of a treatment compared to another.

Since we focus in this work on metastatic breast cancer, we present hereafter the main therapeutic evaluation criteria used for the monitoring of solid tumors.

1.4.2 Morphological criteria

Morphological imaging evaluation criteria assess the change in lesion size over time to evaluate therapeutic response. These methods are easy to use and standardized.

While the objective treatment response is evaluated in comparison to the **baseline** acquisition (the one obtained just before the start of treatment), tumoral progression is evaluated against the **nadir** acquisition (see in Figure 1.9). This nadir corresponds to the acquisition for which the best treatment response was obtained. It may not be the last obtained exam.

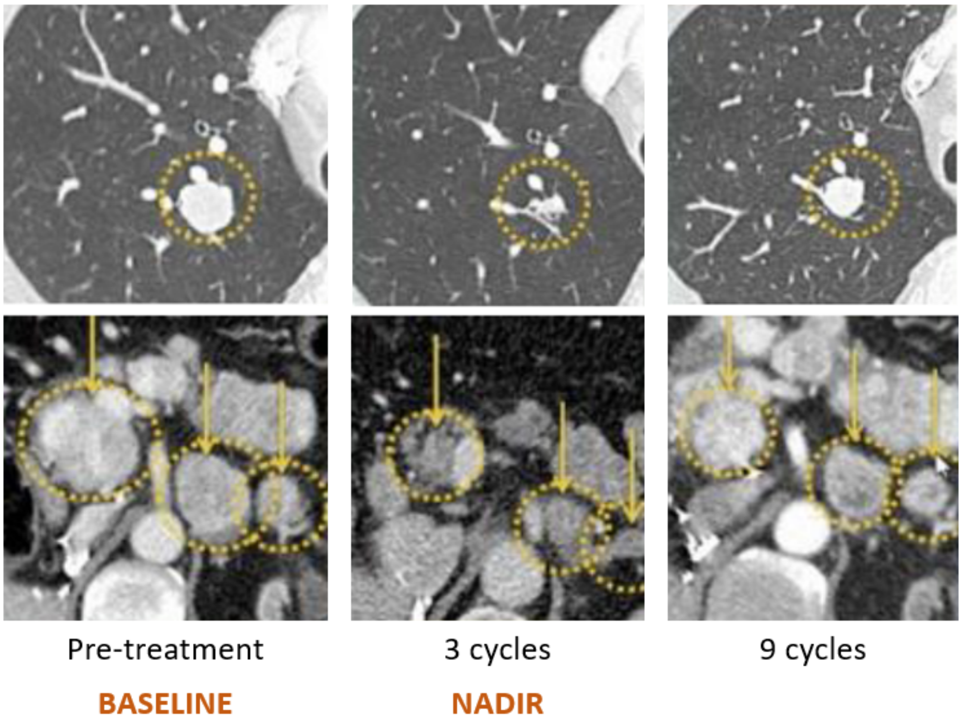


Figure 1.9 – Acquisitions obtained before treatment (left), after 3 cycles (middle) and after 9 cycles (right). Lesions are marked in yellow. The pre-treatment acquisition is the baseline, and here the acquisition after 3 cycles of treatment is the nadir. Images adapted from [140].

1.4.2.1 WHO criteria

The first imaging interpretation criteria were studied by the WHO [142], who proposed to measure lesions in 2D and to compute the sum of the product of the lesions' short and long axis. Several modifications have followed the WHO criteria, leading to the Response

Evaluation Criteria In Solid Tumors (RECIST) guidelines [206].

1.4.2.2 RECIST

RECIST guidelines are defined as a simple 2D evaluation method [206]. RECIST evaluate change in tumor diameters on CT exams acquired at different time points. A revised version, RECIST 1.1, was presented in Eisenhauer et al. [46]. RECIST 1.1 was then updated and clarified by Schwartz et al. [188]. The main goal of RECIST guidelines are to standardize and simplify the evaluation of tumor response to treatment. They are the most common morphological criteria used in clinical trials.

To evaluate treatment response, RECIST defines three categories of lesions: target, non-target and new lesions.

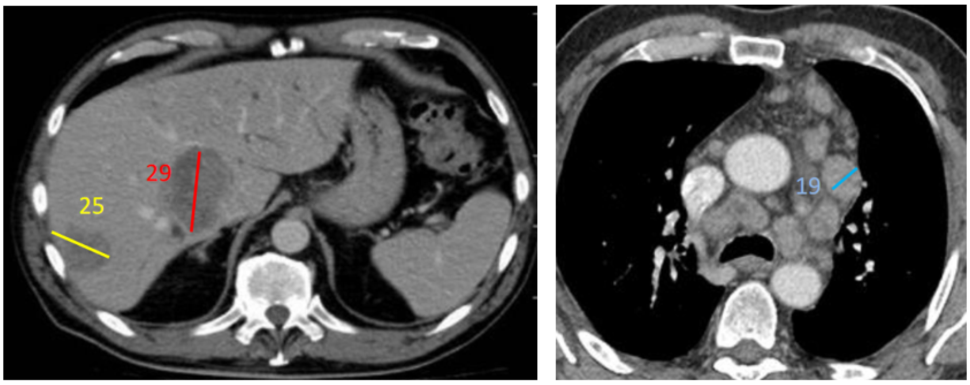
Target lesions Target lesions measure more than 1 cm diameter. A maximum of five target lesions can be selected per patient, with a maximum of two per organ. For example, if a patient only presents lesions in the liver, only two target lesions can be considered. If metastases can be found in the lymph nodes, liver and lungs, up to five target lesions representative of the patient's overall disease can be chosen. The 1 cm limit is not applied to lymph node lesions, which should measure at least 1.5 cm to be considered as target lesions [187]. Some target lesions are illustrated in Figure 1.10.

Target lesions are assessed quantitatively and therefore heavily contribute to tumoral evaluation. The sum of the small axis for nodal lesions and of the large axis for non-nodal lesions is computed at each visit. The result is used to determine *target lesion response* or *progression*:

- Lesion disappearance: CR;
- Diminution of the diameters sum of at least 30% compared to baseline: PR;
- Increase of the diameters sum of at least than 20% compared to nadir: PD;
- Neither responsive nor progressive: SD.

Only the evaluation of target lesions allows to draw the conclusion of partial response.

Non-target lesions Non-target lesions are all lesions which cannot be considered as target lesions: non-nodal lesions with a diameter smaller than 1 cm, nodal lesions with a small axis between 1 cm and 1.4 cm, as well as infiltrating or difficult to measure lesions. Lesions that fulfill the target lesion size requirements, but can't be selected as



Visceral lesions
Long axis measures

Lymph node lesion
Short axis measures

Figure 1.10 – Visceral (yellow and red, left) and lymph node (blue, right) diameters measurements on target lesion for RECIST. Images adapted from [140].

such because the maximum number of target lesions has already been reached should also be followed as non-target lesions. Primary digestive tumors are generally considered as non-target lesions as they are difficult to measure in a reproducible manner at different time points. Some non-target lesions are illustrated in Figure 1.11.

Non-target lesion are assessed qualitatively to determine the *non-target lesion response*:

- If all non-target lesions disappeared: CR;
- If lesions progressed unequivocally: PD;
- If there is no progression or disappearance, non-target lesions response is Non-CR/Non-PD.

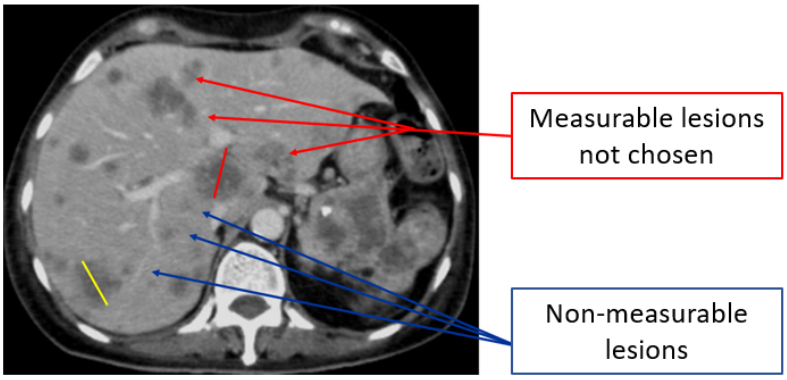


Figure 1.11 – Visceral non-target lesions for RECIST. Image adapted from [140].

New lesions New lesions indicate a progression: it is thus very important to be sure of their existence. New lesions are lesions that appeared unequivocally. The apparition of a pleural, pericardial or ascite effusion cannot be alone considered as a progression criteria. The apparition of condensing bone lesions should also be treated cautiously: they may indicate the healing process of lesions non detected on a first exam.

Overall response The combination of the responses of target and non-target lesions, as well as the presence or absence of new lesions, leads to the determination of an overall response, as defined in Table 1.1.

Table 1.1 – Overall RECIST response depending on the target and non-target lesion responses, as well as on the absence or presence of new lesions. *CR* = Complete Response, *PR* = Partial Response, *SD* = Stable Disease, *PD* = Progressive Disease. *Any* stands for any of the possible response of the considered lesions. Table from [46].

Target lesions	Non-target lesions	New lesions	Overall response
CR	CR	No	CR
CR	Non-CR/ Non-PR	No	PR
CR	Not evaluated	No	PR
PR	Non-PR or not all evaluated	No	SD
Not all evaluated	Non-PD or not all evaluated	No	Inevaluable
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

Progressive disease The categorization of a patient as PD is important since it may impact the patient's medical care [39]. It can be assessed:

- By the increase of the sum of the target lesion diameters of at least 20% (in that case, a 5 mm absolute increase is also necessary to avoid ambiguity on small lesions);
- By an unequivocal progression of non-target lesions;
- By the unequivocal appearance of new lesions.

Since the term "unequivocal" is subjective, progression is often considered uncertain when first observed. In those cases, physicians compare the imaging results with the rest of the exams and with all other available information. It is frequent for a patient in a good medical condition tolerating well the treatment to continue it, despite the presence of questionable lesions, until the next imaging evaluation. On the next time point, physicians will carefully look for confirmation of progression. This caution is needed as certain manifestations can be misinterpreted as tumoral progression, such as the modification of the liver morphology due to steatosis or a pulmonary infection.

RECIST limitations The criteria presented above are subject to inter- and intra-reader variability [13, 124, 204]. In addition, the measuring techniques and the size thresholds used to define response or progression are still debated.

With these criteria, measurements are done on a single slice in 2D, and thus only consider the axial dimension of the lesion(s). To improve the precision of response evaluation, Buerke et al. [22] and Heckel et al. [81] proposed to replace the 1D measurement by a 3D volumetric evaluation. However, with a 3D evaluation, part of the standardization is lost. These 3D measurements also slow down the image analysis process, which is problematic considering the increasing number of exams acquired in oncology. Software programs to help radiologists delineate volumes have been developed, but they do not yet provide fast, automatic and reliable segmentations. Indeed, while automatic segmentation methods provide good results for lung lesions, the delineation of the liver, pathological lymph nodes or infiltrating tumors remains challenging.

Tumor size measurements are an indirect reflection of their development. Yet, they often under-estimate the real response for two reasons. First, when treated, tumoral tissue can partially be replaced by necrotic or fibrous tissue, without causing any change in the global size of the lesion. Secondly, tumor size may remain stable even if its metabolism decreases or stops. Indeed, measurements taken on anatomical CT images are used for lesion monitoring because it is implied that change in lesion diameter is a marker for treatment response: as the treatment kills tumoral cells, cell retraction occurs. However this tumor shrinkage is often only observed at the end of the therapeutic action. Cell retraction (and subsequent death) are caused by functional changes disturbing metabolic cell functions such as glycolysis; anatomical changes are thus the consequence of metabolic changes in lesions. As described in Section 1.3, modifications in glucose metabolism are translated by less ^{18}F-FDG uptake in the cells: changes induced by therapy will thus be first visible on PET images, and then only after on CT images [213].

1.4.3 Metabolic criteria

PET imaging modality is now commonly used in oncology to monitor lesion evolution [43]. Since treatment response is visible on metabolic images before being visible on anatomic images, metabolic criteria to evaluate therapeutic response have been developed for ^{18}F-FDG PET images: the European Organization for Research and Treatment of Cancer (EORTC) [236] and, by analogy to RECIST, the Position Emission tomography Response Criteria In Solid Tumors (PERCIST) [223].

1.4.3.1 EORTC

EORTC guidelines are the first metabolic evaluation criteria developed for ^{18}F-FDG PET in the context of solid tumors [236]. They are based on the variation of the maximum $\mathrm{SUV_{bw}}$ (see Section 1.3.2). In practice, radiologists measure the maximum $\mathrm{SUV_{bw}}$ in 5 to 7 target lesions [189, 194], even if this number is not explicitly specified in the original EORTC paper [236].

The EORCT guidelines state that:

- Complete metabolic response (CR) is defined by the complete disappearance of FDG uptake in the tumoral volume, non visible within the surrounding background noise.
- Metabolic progression (PD) can be defined i) by an increase of at least 25% of the maximum $\mathrm{SUV_{bw}}$, or ii) by a visible increase of the size of the metabolic site (increase of the large axis by at least 20%), or iii) by the appearance of at least one new metabolic site corresponding to a new lesion.
- Partial metabolic response (PR) is defined by a decreasing maximum $\mathrm{SUV_{bw}}$ value of at least 25% (or at least 15% after a chemotherapy cycle).
- Metabolic stability (SD) corresponds to cases not defined above: between partial response and progression.

These recommendations have been formulated based on a small study number (10) and only on five cancer types: six of these studies were done on glioma and medulloblastoma (central nervous system lesions), two on breast cancer, one on head and neck cancer, and the last one on liver metastases from colorectal cancer. All these cancers have been treated with different chemotherapy strategies. Yet, EORTC guidelines are recommended for all types of cancer and chemotherapy, which is questionable [210]. In addition, EORTC guidelines suffer from inter- and intra-center variability because of their lack of standardization [93].

1.4.3.2 PERCIST

In analogy to the RECIST guidelines defined for anatomical images, PERCIST was developed and explained in [223] to evaluate the metabolic response of cancers to treatment with PET images. Contrary to the EORTC criteria, PERCIST use the SUL value (see Section 1.3.2). Since it is based on the lean body mass computation, it is not influenced by the patient's body weight. PERCIST guidelines are based on a meta-analysis of more than 3000 publications.

The metric used in PERCIST guideline is the $\mathrm{SUL}_{\mathrm{peak}}$. Considering a 1 cm^3 sphere in the lesion presenting the highest fixation, i.e. the **target lesion**, $\mathrm{SUL}_{\mathrm{mean}}$ is computed for every sphere position. $\mathrm{SUL}_{\mathrm{peak}}$ corresponds to the maximum value of these $\mathrm{SUL}_{\mathrm{mean}}$ values. The variation of $\mathrm{SUL}_{\mathrm{peak}}$ is measured with FDG PET acquisitions in 3D. In clinical practice, this measurement is now automatically computed: $\mathrm{SUL}_{\mathrm{peak}}$ presents a high inter-reader reproducibility. In addition, it does not suffer from the background noise influence, compared to $\mathrm{SUV}_{\mathrm{max}}$ or $\mathrm{SUL}_{\mathrm{max}}$ for which measurements are very sensible to noise, especially for small lesions.

To be considered as the target lesion for measurement, the selected lesion should have the highest FDG uptake and should have a $\mathrm{SUL}_{\mathrm{peak}}$ higher than 1.5 times the threshold value of the liver background noise added to two times its standard deviation. The threshold value of the liver background noise is defined by the $\mathrm{SUL}_{\mathrm{mean}}$ measured in a 3 cm diameter sphere on an homogeneous (without lesion) part of the liver (see in Figure 1.12). These minimum uptake thresholds limit interpretation errors and allows to better assess metabolic response. It is also recommended that the target lesion has a long axis of at least 1.5 cm to limit the partial volume effect (see Section 1.2.1.4). These thresholds were chosen empirically by authors [102].

To evaluate treatment response, the measurement of $\mathrm{SUL}_{\mathrm{peak}}$ is performed on the target lesion at each time point and it may not be the same lesion from one PET acquisition to the other:

- A complete metabolic response (CR) is defined by the visual disappearance of all lesions. This means that their metabolic activities are lower than the liver background noise threshold, and that they are non detectable from the surrounding background noise.
- A metabolic progression (PD) can be defined i) by an increase of the $\mathrm{SUL}_{\mathrm{peak}}$ of at least 30% and of 0.8 points in respect to the nadir exam, or ii) by a visible

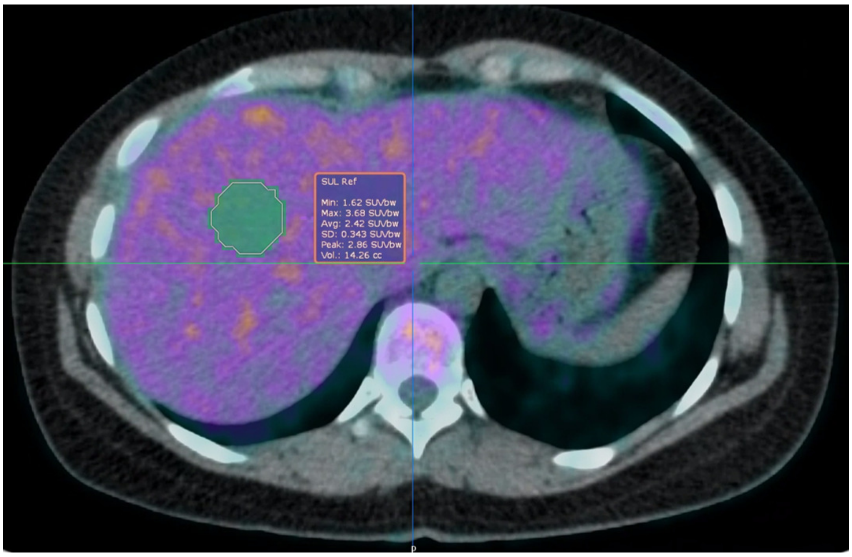


Figure 1.12 – Liver 3 cm diameter sphere placed in the liver as a reference for PERCIST background noise measure. Image from our EPICURE dataset.

increase of the metabolic site, or iii) by the apparition of at least one new lesion. Assessments are done on the target lesions. In case of a progression caused by the SUL_{peak} increase (without new lesion), the progression should be confirmed by a another acquisition once month later, except if a RECIST progression is confirmed.

- A partial response (PR) is defined by the decrease of at least 30% of the SUL_{peak} and 0.8 units.
- A stability (SD) is neither a partial response nor a progression.

In addition to these response evaluation criteria, criteria ensuring the comparability of acquisitions are also defined. The non-respect of these quality control criteria should be detailed in the clinical report.

- The difference between liver SUL_{mean} values in the liver should not exceed 20% and 0.3 units.
- The duration between the radiotracer injection and the image acquisition should not differ by more than 15 min between exams. Acquisition time should lie between 50 min and 70 min after the injection.
- The same imaging systems and acquisition protocols should be used for all PET image acquisition.
- Injected doses should not differ by more than 20%.

- Patients should be fasting for at least 4h and present a blood sugar lower than 2g/L.

Min et al. [144] showed that PERCIST and EORTC criteria present very similar performances, either regarding the patient's categorization [189, 241], or the overall survival [194]. Because of the quality control criteria, PERCIST guidelines are more difficult to integrate in clinical routine than EORTC's. Yet PERCIST standardization makes its results more reproducible: inter-reader variability is relatively low [51], below the simple visual analysis.

Comparing RECIST to PERCIST, PERCIST categorized more patients in CR or PD [144]. Hence, PERCIST is able to provide an earlier response to treatment. However, SUL measured according to the PERCIST criteria recommendations suffers from approximations (see Section 1.3.2). The relative variation of these indexes is subject to bias and precautions should be taken when comparing to fixed thresholds.

1.4.4 Criteria limitations

With most imaging criteria, the correlation between what is observed and what actually happens in tumors is not guarantied. For instance, if a patient is categorized CR, it does not mean that the tumor disappeared at the histologic level. In most cases, tumoral cells persist at a microscopical level and are at the origin of recurrence if the treatment is stopped [65]. This is especially important if a surgical treatment is considered, because non-removed lesions, even if mophologically disappeared, will cause local recurrence. This phenomenon is called "missing metastases". To avoid it, the surgical intervention should be planned while small isolated metastases are still visible.

These criteria should be used to monitor cancers and to evaluate a response to treatment. They are not meant to determine whether a patient has improved or not, as they are tumor-centered and not patient-centered. More information, such as global health, age, physical symptoms and biological markers should also be considered to provide the best treatment possible to the patient, and decide if there is a need for treatment adaptation.

1.5 Our clinical dataset: the EPICURE$_{\text{seinmeta}}$ data

This thesis is part of the ongoing prospective multicentric EPICURE$_{\text{seinmeta}}$ study [1] (NCT03958136) [32], in collaboration between the Keosys company [2] and the "Institut de Cancérologie de l'Ouest" (ICO) [3]. To better understand and describe breast cancers, the EPICURE project aims to collect diverse data about patients presenting breast metastases, such as clinical, biological, imaging and histological data, as well as personal and familial history and global environment. This prognostic study was approved by the French Agence Nationale de Sécurité du Médicament et des produits de santé (ANSM, #2018-A00959-46) and the Comité de Protection des Personnes (CPP) IDF I, Paris, France (#CPPIDF1-2018-ND40-cat.1), and a written informed consent was obtained from all patients.

In this large multidisciplinary project, we focused on the monitoring of the cancer's evolution, working on metastatic breast cancer images. Patients underwent between two to three PET–CT acquisitions (see in Figure 1.2). They correspond to pre-, early- (after about a month) and mid-treatment (after about three months) time points. Pre-treatment acquisitions correspond to the initial baseline images, while the early- and mid- treatment acquisitions to the follow-up ones.

Images were acquired in two different centers. In the Angers center, images were obtained using a Philips Vereos or a GE Discovery PET–CT imaging systems, while in the Nantes center images were acquired with two different dual-slice Siemens Biograph PET–CT scanners. Since this is a prospective study, the number of available images is still increasing. For this reason, we specify for each application the number of processed images, as well as their acquisition center.

As shown useful for metastatic breast cancer monitoring [8, 27], we worked only on PET images normalized by the SUV [109]. Ground truth segmentations of all lesions were manually delineated by expert physicians from ICO on all PET images.

1.6 Conclusion

Metastatic breast cancer is a disease with poor prognosis. To monitor its evolution and evaluate its response to treatment, PET–CT imaging is acquired at regular time points.

---

[1] https://projet-epicure.fr/

[2] https://www.keosys.com/

[3] https://www.institut-cancerologie-ouest.com/

While the CT part brings anatomical information, the PET part provides information on metabolic activity.

Using a glucose-based radiotracer, PET images help visualize high glucose uptake areas, i.e. cancerous lesions (pathological hyperfixations) and specific organs (physiological hyperfixations). These physiological hyperfixations can cause confusion, creating a challenge for disease monitoring.

For the monitoring of tumor response and progression, standardized evaluation criteria have been proposed. Yet, they only consider few lesions and can present reproducibility limitations. In this thesis, we address this issue by proposing an automatic tool which i) propagates lesion segmentations maps from baseline to follow-up images using registration (see Chapters 3, 4 and 5), and ii) extracts biomarkers to monitor each lesion individually (see Chapter 6). The methodological context of medical image registration is presented in Chapter 2.

Summary

✓ Despite several limitations, PET images are widely used for metastatic breast cancer monitoring.

✓ One of the limitations of PET imaging is that ^{18}F-FDG uptake is not specific to cancerous cells.

✓ Metastatic breast cancer evaluation and monitoring through morphological and metabolic criteria is time consuming and subject to inter- and intra-reader variability.

Chapter 2

Medical image registration: methodological context

Contents

2.1 Introduction

There is an extensive literature studying the image registration problem (see Figure 2.1), in particular its non-rigid form [91, 137, 139, 164, 171, 200, 214, 244]. Despite these efforts, deformable registration is still considered an unsolved problem [28].

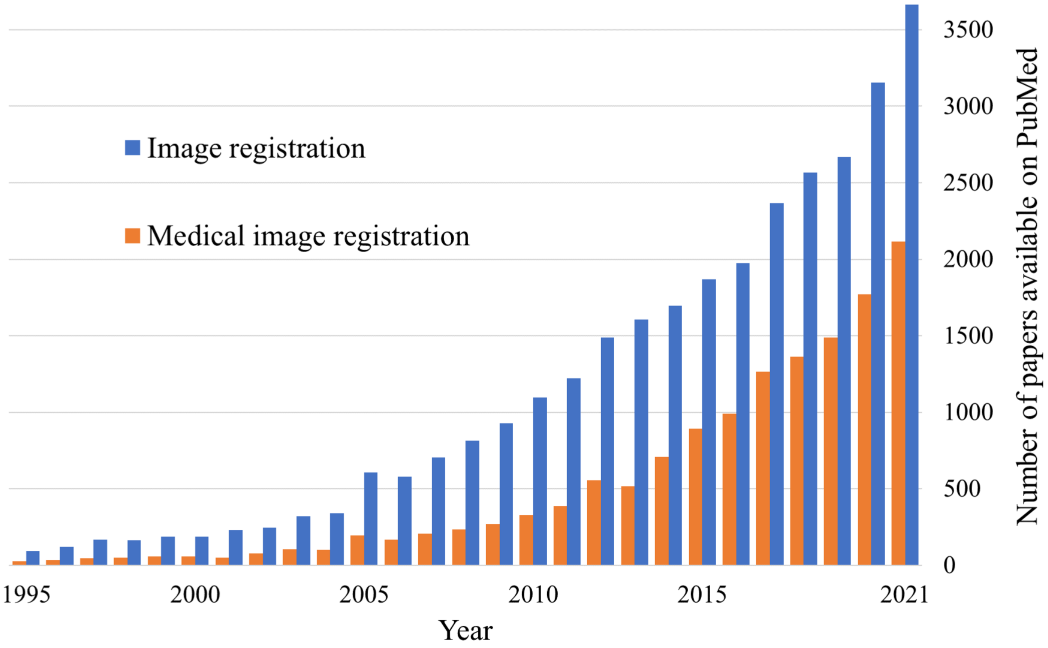
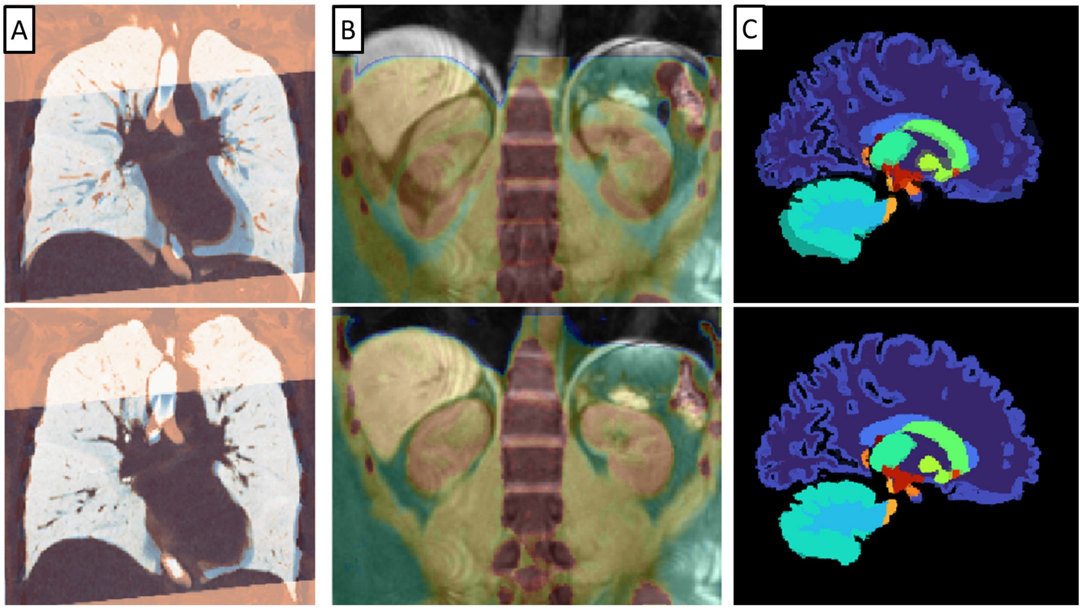


Figure 2.1 – Publication count in *PubMed* between 1992 and 2021 with the keywords “Image registration” and “Medical image registration”.

Registration plays an important role in medical image analysis. Its goal is to find corresponding functional or anatomical locations in two or more images. Depending on the application, registration can be *mono-modal* or *multi-modal*, as well as *intra-subject* or *inter-subject*:

- Mono-modal intra-patient registration (see Figure 2.2.A) may be used to spatially compensate for geometric (e.g. patient position) or intensity variations between different views. On longitudinal data taken at relatively long time intervals, registration has for instance been used in the oncological context to detect, characterize, quantify and monitor lesions or diseases (see Chapters 3 and 4).
- Multi-modal intra-patient registration (see Figure 2.2.B) allows the fusion of anatomical and/ or functional information obtained from different imaging modalities acquired from a single patient. These imaging modalities can be anatomic such



Mono-modal (CT) lung inspiration-expiration intra-patient registration

Multi-modal (CT/MRI) thorax-abdomen intra-patient registration

Mono-modality (MRI) whole brain inter-patient registration

Figure 2.2 – A. Mono-modal intra-patient image registration. B. Multi-modal intra-patient image registration. C. Mono-modal inter-patient image registration. First row shows the overlay of the fixed and moving images (columns A and B) or segmentations (column C), while the second row displays the overlaid fixed and warped data. Images from Siebert et al. [191].

as CT, MRI, angiography or UltraSound (US). They can also be functional such as PET, Single-Photon Emission Computed Tomography (SPECT) or functional Magnetic Resonance Imaging (fMRI). Multi-modal intra-patient registration helps to interpret images, as well as to plan (pre-operative) or perform (intra-operative) a surgery.

- Mono-modal inter-patient data registration (see Figure 2.2.C) is developed to find correspondences between real and generic patient models. It is used to automatically index anatomical structures and construct atlases.

In this thesis, we model patient response quantification in terms of a *mono-modal intra-patient registration problem for longitudinal whole-body PET images*. In this chapter, we introduce and formalize the general image registration problem (Section 2.2). Then we introduce conventional and learning-based approaches for registration (Section 2.3). We define here conventional registration methods as the non-learning-based ones, i.e. the ones not taking advantage of several images to perform registration. As we will see later, DL-based methods share some commonalities with conventional approaches. We will therefore mention how these commonalities have been transferred to DL-based methods while introducing conventional approaches. These will serve as a basis to our proposed method presented later in Chapter 4.

2.2 Conventional medical image registration

Medical image registration consists in finding a spatial transformation to align a *moving image* to another *fixed* or *target image*, as illustrated in Figure 2.3.

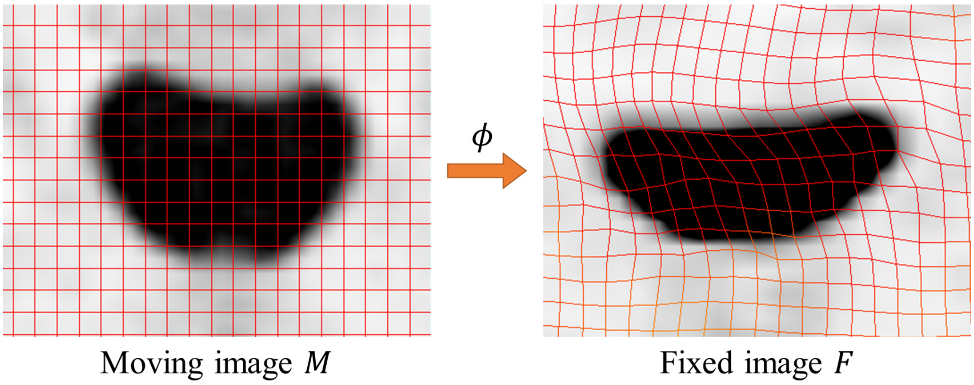


Figure 2.3 – Image registration principle: the moving image M is warped by the spatial transformation ϕ to match the fixed image F. The deformation is represented by an overlaid grid. PET images from our EPICURE$_{\text{seinmeta}}$ dataset [32].

Let M be the moving image and F the fixed one, both defined on the image domain Ω. Registering M on F consists in finding the spatial transformation $\phi : \Omega \to \Omega$ such that the warped image $W = M \circ \phi$ is similar to F according to a predefined criterion. ϕ is the spatial transformation, while $\circ$ describes the warping operation. In the more general case, the transformation ϕ can be defined using a displacement or deformation field $\mathbf{u}$ for every position $\mathbf{x} \in \Omega$: $\phi(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x})$. Thus, for every $\mathbf{x} \in \Omega$, we look for a field $\mathbf{u}$ such that $W(\mathbf{x}) = W(\mathbf{x} + \mathbf{u}(\mathbf{x}))$ is similar to $F(\mathbf{x})$.

Registration can be seen as an optimization problem aiming to find the optimal deformation field ϕ. Given $\mathcal{S}$, a dissimilarity measure between the moving image M and the fixed image F, registration algorithms aim at minimizing an objective function of the form:

$$\arg\min_{\phi} \mathcal{S}(F, M \circ \phi) \tag{2.1}$$

The problem stated Eq. 2.1 is ill-posed, i.e. it belongs to the non-linear transformation class [156]. To constrain the domain of admissible transformations and obtain plausible deformation fields, a regularizer $\mathcal{R}$ and its hyperparameter λ are added to Eq. 2.1 to enforce transformation properties, such as the desired level of transformation smoothness [200]:

$$\arg\min_{\phi} \mathcal{S}(F, M \circ \phi) + \lambda\mathcal{R}(\phi) \tag{2.2}$$

The regularization term can be interpreted as prior knowledge about the expected transformation, whereas the similarity measure, i.e. the opposite of the dissimilarity measure, can be viewed as a likelihood expressing the probability of a match between images [181].

Most image registration algorithms are composed of:

- A cost function, formed by chosen dissimilarity $\mathcal{S}$ and regularization $\mathcal{R}$ terms;
- A deformation/ transformation model, which determines the set of possible transformations $\mathcal{T}$;
- An optimization strategy suitable to minimize the cost function.

As presented in the following sections, similarity measures (Section 2.2.1), transformation models (Section 2.2.2), regularization terms (Section 2.2.3), optimization schemes (Section 2.2.4) and validation techniques (Section 2.2.5) are usually adapted to best address the medical registration problem at hand (mono- or multi-modal, intra- or inter-patient).

2.2.1 Similarity measures

2.2.1.1 Feature-based similarity measures

The first family of similarity measures are the feature-based similarities, which aim to match landmarks present in the images to register (see Figure 2.4). These landmarks correspond to salient structures in the images, such as points, contours or surfaces extracted either manually [155] or automatically [103]. The choice of landmark should be guided by properties, such as being easily and precisely detectable, being distributed on

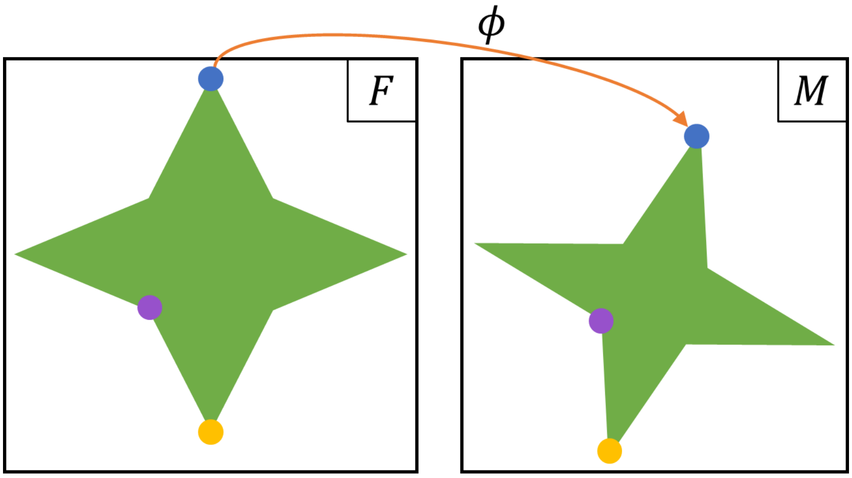


Figure 2.4 – Feature-based similarity measures compute the distance between matching landmarks placed on salient features (colored dots) of the two images to register F and M (green stars) providing a deformation field ϕ.

the whole image, as well as being robust to noise, artifacts and different changes inherent to the acquisition [244].

The main advantage of feature-based similarities is the manipulation of a compact image representation carrying high-level information based on anatomical considerations. As a consequence, the computation load is smaller than for the other family of similarity measures, e.g. intensity-based measures (see Section 2.2.1.2).

While there are numerous algorithms for landmark detection [33], the main drawback of feature-based similarities is the imprecision of the extracted landmarks. To address this issue, landmark extraction error can be included in the deformation model estimation [176]. Another limit regards the registration precision, which is guarantied only around landmarks. Since the obtained deformation field is sparse, i.e. computed only on landmark pixels, feature-based measures are usually employed to perform rigid transformation (see Section 2.2.2). In the case of non-rigid registration, the deformation field can also be extrapolated to the whole image, typically resulting in a loss in registration accuracy. These limitations explain the decreasing popularity of feature-based similarity measures for conventional registration methods [200]. Yet, DL-based registration methods developed lately have taken advantage of landmarks to improve the training of Neural Networks (NNs) (see Section 2.3.5).

2.2.1.2 Intensity-based similarity measures

The second similarity-measure family relies on image intensities. These similarities have become the reference to measure image alignment, largely due to their robustness and accuracy. Contrary to feature-based methods, intensity-based similarity measures compare the images intensity values either directly or after processing them (using for instance image differential statistics or a Fourier transform). The main advantage of intensity-based registration methods is therefore the use of the whole image information. Depending on the assumed relation between the intensities of the images to register, different measures can be used.

For mono-modal registration, assumptions on the relation between intensities and on the type of noise dictate the choice of the measure. For instance, the Sum of Squared Distances (SSD) measure assumes that both images (F, M) present the intensity and illumination consistency. If images are correctly aligned, the SSD between them should be zero, except for the noise. If this noise is Gaussian distributed, Viola et al. [218] demonstrated that SSD is the optimal similarity measure. With F and W respectively the fixed and warped images, and N the number of pixels or voxels:

$$SSD(F,W) = \frac{1}{N}\sum_{i=1}^{N}(F_i - W_i)^2 \tag{2.3}$$

When a linear relation between image intensities is assumed, the Cross Correlation (CC) or Normalized Cross Correlation (NCC) measures can be used to express the similarity between both images. Registration algorithms maximize correlation-based measures. With $\overline{x}$ and σ_x respectively the mean and standard deviation, NCC can be defined as:

$$NCC(F,W) = \frac{1}{N}\sum_{i=1}^{N}\frac{(F_i - \overline{F})(W_i - \overline{W})}{\sigma_F \sigma_W} \tag{2.4}$$

In multi-modal registration, similar structures are acquired with different imaging modalities. They may present very different intensity values, making mono-modal measures not suitable to perform accurate registration. The most popular measure in multi-modal registration is the Mutual Information (MI) [134, 218]. MI assumes a non-parametric statistical relationship between image intensities. Instead, MI is defined in terms of the images individual and joint entropies. Joint entropy being minimal when images are perfectly registered, a registration algorithm will seek to maximize the MI similarity:

$$MI(F, W) = \mathbf{H}(F) + \mathbf{H}(W) - \mathbf{H}(F, W) \tag{2.5}$$

with:

$$\begin{cases} \mathbf{H}(I) & = & -\sum p_i \log(p_i) \\ \mathbf{H}(I, J) & = & -\sum p_{i,j} \log(p_{i,j}) \end{cases} \tag{2.6}$$

and p_i and $p_{i,j}$ respectively the probability and co-occurence probability of i and j in term of gray values.

Despite their wide use, SSD, NCC or MI intensity-based measures suffer from drawbacks such as sensitivity to noise and image artifacts and may require pre-processing steps like image filtering [207]. They also present a high computational cost since they usually consider a large amount of image pixels or voxels, and the optimization problem is challenging to solve because often presenting several local minima. To decrease this cost and help solve the optimization problem, hierarchical approaches building image pyramids have been proposed (see Section 2.2.4). Moreover, some hybrid similarity measures were also proposed, using for instance feature-based measures to constrain intensity-based approaches [100, 199].

For our application, i.e. performing intra-patient PET image registration, we assume a linear relation between image intensities and use the intensity-based NCC similarity measure.

2.2.2 Transformation models

Once the similarity measure describing accurately the registration problem at hand is chosen, the transformation model should be defined. Spatial image transformations can model global or local transformations. A global transformation is uniformly applied to the whole image, while with a local one, image sub-sections present their own transformations. Most transformation models require a more or less complex interpolation step to effectively warp all image pixels into the fixed image grid.

The choice of the transformation model is very important as it determines the Degrees of Freedom (DoF) and complexity of the estimated deformation. This choice is guided by the application at hand and by prior information available on the nature of the deformation. Indeed, the transformation model can limit the set of admissible transformations $\mathcal{T}$ through different parametrizations. As shown hereafter, parametrizations of $\mathcal{T}$ can vary from having a small number to millions of parameters.

2.2.2.1 Rigid transformation models

Rigid transformation is the simplest global geometric model. It consists in the estimation of translations and rotations to reposition an object supposed rigid, while conserving distances, angles and parallelism (see Figure 2.5). The number of DoF is three in 2D (two translations and a rotation) and six in 3D (three translations and three rotations). It is commonly used in the context of medical image mono-modal intra-patient registration to correct position errors when no soft tissue deformation occurred [52], or as an initialization step for more complex transformations [173]. However, rigid transformation is generally not sufficient to perform accurate inter-subject registration [91].

2.2.2.2 Affine transformation models

An affine global transformation allows, in addition to rotations and translations, to correct the scaling and the shearing (see Figure 2.5). This transformation preserves parallelism. The number of DoF is six in 2D and twelve in 3D, representing translation, rotations, scaling and shearing. This model may be used for multi-modal inter-patient registration [99], even though its low number of DoF does not allow to precisely model anatomical variability.

2.2.2.3 Deformable transformation models

Since rigid and affine transformations only allow a small number of DoF, they are not suitable to represent complex deformations, e.g. in soft tissues such as the liver or breasts. Transformations using a higher number of DoF and which are able to locally modify the image are needed. These models, called non-rigid or deformable, can be used to perform inter-patient or atlas registration, as well as intra-subject registration if anatomical movement occurred between acquired images [52] (see Figure 2.5). Deformable models can be categorized as parametric or non-parametric.

Non-parametric deformable transformation models Non-parametric models are defined on the largest possible space of transformations. Their dense deformation fields are described by displacement fields, where an independent 2D or a 3D vector describes the motion of each pixel or voxel in the image. These displacement fields can be obtained

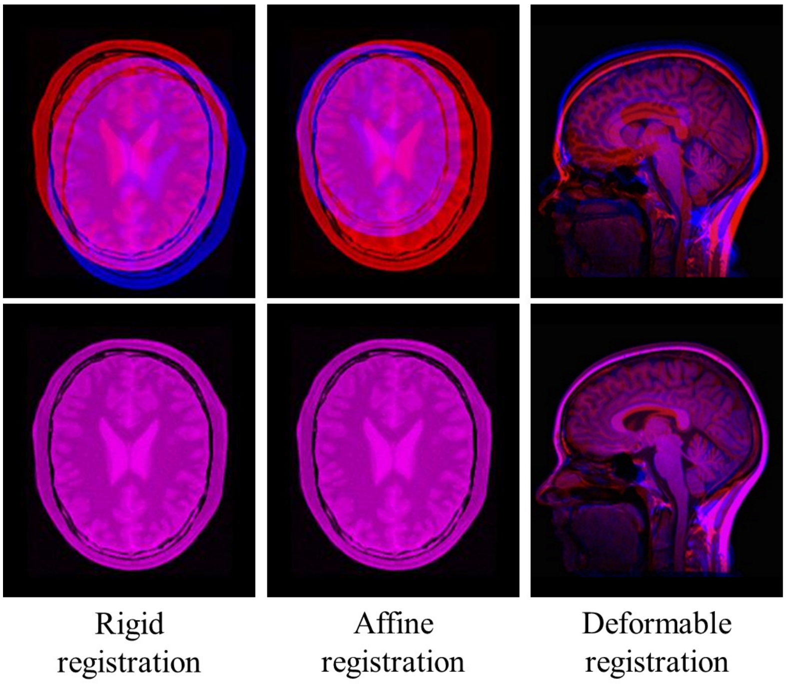


Figure 2.5 – The first row displays the fixed (red) and moving (blue) images, while the second row shows the fixed (red) and registered (blue), i.e. warped, images. The overlay of both images is visible in pink. From left to right, a rigid (translation and rotation), affine (translation, rotation, scaling and shearing), and non-rigid registration method was applied to the moving image so that the warped image matches the fixed one. Images from https://simpleelastix.readthedocs.io/.

either directly [91] or through the integration of a Stationary Velocity Field (SVF) overtime [5, 146]. Non-parametric deformable models include, for instance, the elastic model [40, 169], the fluid model [30] or the optical flow model [207]. Yet, they are complex and expensive to optimize.

Parametric deformable transformation models To parametrize the deformation field, the Free-Form Deformations (FFDs) model, for instance, is commonly used in medical image registration [175, 182, 185]. FFDs' main advantage lies in their aptitude to estimate very localized deformations.

From the subset of control points, the parametric transformation is interpolated to all the points of the image to obtain the global transformation. Several interpolation functions were developed for parametric deformable transformation models. Some apply to the whole image, such as the Radial Basis Functions (RBFs) [21], the Thin-Plate

Splines (TPS) function (a subset of RBFs), as well as polynomial [101], Gaussian [118] or trigonometric [6] functions. B-spline functions [182], on the contrary, do not apply to the whole image but only on an specific area of the image.

Diffeomorphic transformations, i.e. transformations reversible and which preserve the topology, can be parameterized by Large Deformation Diffeomorphic Metric Mappings (LDDMMs) [15, 26, 238], or Stationary Velocity Fields (SVFs) [5, 129, 215].

As detailed in Chapter 3, we perform deformable registration with a conventional method using a B-spline interpolation to address our intra-patient mono-modal registration problem.

2.2.3 Regularization

To cope with the ill-posed nature of the image registration problem and to reduce the range of admissible transformations, different explicit regularization terms $\mathcal{R}$ (see Eq. 2.2) have been proposed. In the medical image field, regularization helps produce plausible deformations, which means they help avoid problems such as non-smoothness, non-invertibility or folding. Moreover, it may be important for some applications to maintain tissue properties, e.g. bones rigidity, which can be included as prior knowledge.

To help produce a smooth deformation field, regularization terms based on the derivatives of the deformation field have been often proposed. Indeed, smooth deformations are related to small gradient values. Li et al. [123], for instance, used a Total Variation (TV) regularization on the deformation field aiming to reduce spatial incoherence, while Christodoulidis et al. [31] constrained the objective function using the L1-norm for more robustness to image noise. Also to enforce a smooth transformation and put more emphasis on avoiding large deformation values, Klein et al. [112] used the L2-norm on the gradients of the deformation field. This L2-norm regularization term, also called bending energy, is commonly employed in DL-based methods [11, 28] to enforce smoothness (see Section 2.3).

To obtain inverse consistent deformations, symmetric algorithms optimize the objective function in both directions: from the moving towards the fixed image and vice versa [7, 129, 216]. DL-based methods also seek reversible transformations by using similar objective functions, where fixed and moving images are both warped to match each other, and then deformed back to the original images [72, 108, 136].

Obtaining diffeomorphic transformations is interesting to avoid implausible foldings when performing medical image registration. Indeed, diffeomorphic transformations are reversible and preserve the topology. To avoid undesired foldings, the determinant of the Jacobian of the deformation field **J** [1] may be considered, since its negative values indicate folding areas. Thus, a positiveness constrain on **J** is a common strategy to limit the folding problem [63].

The determinant of the Jacobian of the deformation field can also be used to impose incompressibility constrains [175], for tissues such as the bones. Indeed, $\mathbf{J} > 1$ reflects a dilation and $0 < \mathbf{J} < 1$ a compression of the area.

In this thesis, we exploit the L2-norm on the gradients of the deformation to impose smoothness constrains, and add anti-fold regularization in our latter experiments. In addition to these explicit regularization terms and inspired by the implicit regularization imposed by parametric deformable transformation models (see Section 2.2.2), we state and develop in Chapter 4 that the architecture of DL network also acts as an implicit regularizer to parametrize the transformation.

2.2.4 Optimization

The image registration problem can be seen as an optimization problem aiming to minimize an energy function **E** defined as in Eq. 2.2:

$$\mathbf{E} = \mathcal{S}(F, M \circ \phi) + \lambda \mathcal{R}(\phi) \tag{2.7}$$

Since ϕ depends on the parameters θ of the transformation $\mathcal{T}$, optimizing **E** is equivalent to retrieving the optimal parameters θ^*. While rigid and affine models depend only on a small number of parameters (from three to twelve), this number rises to thousands for deformable models. For instance the number of parameters depends on the amount of chosen control points for parametric deformable models.

As presented hereafter, these optimal parameters θ^* are typically determined with continuous or discrete optimization schemes. Moreover, both optimization schemes can be performed in a mono- or multi-resolution setups. Mono-resolution setups are performed on full resolution images, which can be computationally expensive on large and/ or 3D data. Multi-resolution setups, on the contrary, optimize the transformation model on low resolution images before moving on full resolution ones. This is often referred as a pyramidal optimization strategy [244].

---

[1] $\mathbf{J} = \det(d\mathcal{T}/d\mathbf{x}) = \det(I + d\mathbf{u}/d\mathbf{x})$, with $\mathcal{T}$ the transformation, $\mathbf{x}$ spatial positions, I the identity matrix, and $\mathbf{u}$ a displacement field.

2.2.4.1 Continuous optimization

Continuous optimization approaches correspond to gradient-based methods. Even though the choice of an optimizer is not trivial [113], commonly used methods in continuous optimization have emerged over the years.

The simplest approach is the Gradient Descent (GD) method, which only requires the computation of the gradient of **E** to search for the steepest descent direction, and was used for instance by Johnson et al. [100] and Rueckert et al. [182]. The Conjugate Gradient (CG) method converges faster than GD and takes advantage of prior knowledge from gradient descent [208]. To speed up convergence and reach a higher convergence rate than CG, the Quasi-Newton (QN) method uses second-order information [128]. The Levenberg-Marquardt (LM) method combines the advantages of the GD and QN methods [104]. Finally, the widely used Stochastic Gradient Descent (SGD) method allows to reduce the burden of heavy calculation [202].

These methods reach a high accuracy. However, they require a differentiable function **E**, are difficult to control, and may converge to a local minima.

2.2.4.2 Discrete optimization

Discrete optimization is performed only on a discrete subset of the parameter space. It presents several advantages, such as the possibility to find a strong minima more often, a high computational efficiency, as well as a full control on the research space. However, the accuracy of the optimization is bounded by the discretization [67].

In medical image registration, discrete optimization methods commonly refer to the use of Markov Random Fields (MRFs) to conduct the optimization [68, 196]. The transformation parameters are represented by nodes in a fully connected graph where the edges model pairwise dependencies. Parameters are estimated through iterative discrete labeling and discrete optimization. Contrary to gradient-based methods, MRFs are robust to initialization [242].

In the remainder of this thesis, we use the SDG optimizer to solve our registration problem because of its interesting convergence and computational properties. As shown in Section 2.2.6, this optimizer is used for most conventional registration pipelines.

2.2.5 Validation

The validation of a registration algorithm is fundamental for clinical applications. While the comparison of computed segmentations with ground truth segmentations performed by experts makes the validation of medical image segmentation straightforward, registration methods are more challenging to validate.

Registration performed on simulated transformations can be validated directly by comparing the computed and the ground truth deformation fields. However, it is usually not possible to know the true motion of a patient/ organ, and whether such movements affect the clinical usefulness of the data [50]. Most registration methods are therefore validated indirectly, e.g. comparing the fixed and warped images using similarity measures such as SSD, NCC or MI. Yet, these measures evaluate the overall alignment of the images, not the local registration accuracy.

To evaluate the registration locally, landmarks placed manually or automatically can be used. However, they may be imprecise and are difficult to use in practice for whole body image registration.

Another indirect validation approach relies on segmentation masks: ground truth mask (on the fixed image) and warped mask (from the moving image) can be compared to evaluate the registration. Even though obtaining segmentation masks can be time consuming, this method allows to evaluate the registration precision on structures of interest, such as anatomical objects.

It is also interesting to evaluate the smoothness and diffeomorphism of the produced deformation fields, for instance through the analysis of their Jacobian determinants [28] (see Section 2.2.3). A high standard deviation of the Jacobian determinant translates into a non-smoothness of the deformation field. Moreover, negative values in the Jacobian determinant represent “folding” areas in the warped images, a phenomena that should be avoided in most medical image registration applications.

2.2.6 Conventional registration methods today

Through the years, several conventional methods have emerged and are commonly used as reference to validate new registration approaches. SimpleElastix[2] [112], for instance, is an intensity-based registration method using cubic B-spline optimized by the SGD algorithm in a multi-resolution optimization strategy. Using a similar optimization

---

[2] https://simpleelastix.readthedocs.io/

scheme, ANTS[3] [7] computes transformations from velocity fields. NiftyReg[4] [182] generates the deformation field with FFDs and optimizes its objective function with a CG algorithm. Among the four discussed here, Deeds[5] is the only using a MRF-based discrete optimization strategy [86].

Depending on the application at hand, each method presents its strengths and weaknesses. In this thesis, we aim to register 3D whole body longitudinal PET images, i.e. to perform intra-patient mono-modal registration. As previously presented, we use the NCC intensity-based similarity measure because we assume a linear relationship between the intensities of our images. Moreover, since we want to perform precise local registration, especially around the lesions, we compute deformable transformations, and choose the B-spline parametric method because of the large size of our images. To compute a smooth transformation, we add a bending energy term to the loss function, optimized using the SGD algorithm. These choices led us to use the SimpleElastix method, as detailed in Chapter 3.

Overall, conventional methods reach good registration performance. Yet, the choices of similarity measure, transformation function, regularization term or optimization scheme are not always straightforward. Moreover, they are often time-consuming and the choice of optimal parameters tends to be different for different pairs of images. To address some of these limitations and take advantage of large databases, learning-based methods are being developed, as presented in the remainder of this chapter.

2.3 Learning-based medical image registration

In the past seven years, learning-based and especially deep learning-based image registration methods have emerged to take advantage of large databases and faster computational resources (see Figure 2.6) [28, 63, 78, 209, 225]. DL registration methods have became an important research topic, with several workshops and challenges having taken place recently on this theme, such as the Workshop on Biomedical Image Registration (WBIR)[6] or the Learn2Reg[7] and BraTS-Reg[8] challenges.

---

[3] https://antspy.readthedocs.io/en/latest/

[4] https://sourceforge.net/p/niftyreg/git/ci/master/tree/

[5] https://github.com/mattiaspaul/deedsBCV

[6] https://www.wbir.info/

[7] https://learn2reg.grand-challenge.org/

[8] https://www.med.upenn.edu/cbica/brats-reg-challenge/

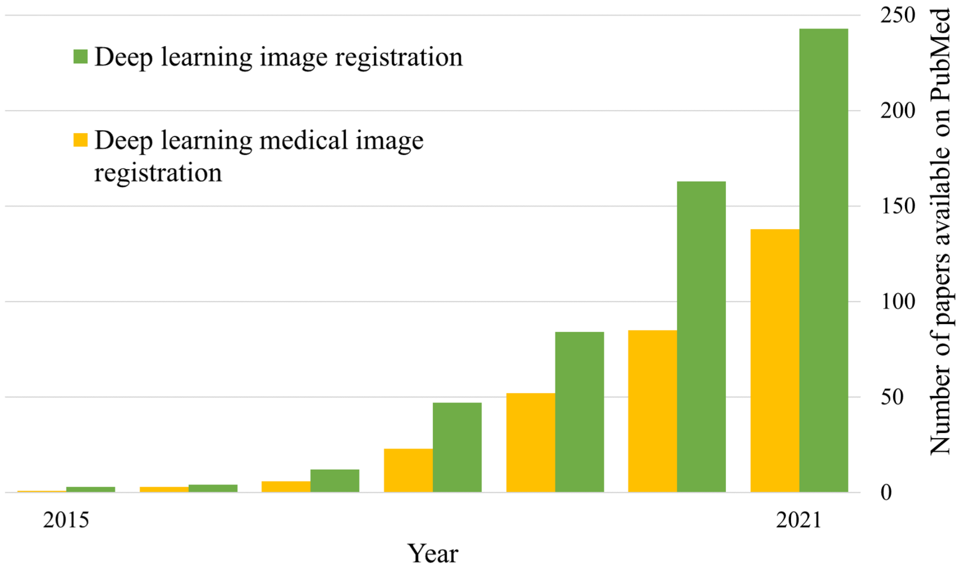


Figure 2.6 – Publication count in *PubMed* between 2015 and 2021 with the keywords "Deep learning image registration" and "Deep learning medical image registration".

In the following sections, we first present some Machine Learning (ML) registration methods (Section 2.3.1), before moving to DL-based approaches. We first explain the main characteristics of a Convolutional Neural Network (CNN) used to perform DL-based image registration (Section 2.3.2), before presenting different kinds of DL-based registration methods, such as supervised (Section 2.3.3), unsupervised (Section 2.3.4), as well as dual and weakly-supervised (Section 2.3.5) approaches.

2.3.1 Machine learning and image registration

Statistical Deformation Models (SDMs) can be considered as the first data-based unsupervised registration approaches. They reduce the deformation's dimensionality to be able to solve a simpler registration problem. Yet, SDMs need to be trained on a database, and are thus biased by the observations in this training set. Their goal is to find a compact representation of the plausible deformations by measuring them on a population database and identify the most representative directions of deformation, e.g. by using Principal Component Analysis (PCA) [180].

Later, learning-based approaches were developed to learn optimal similarity measures. Lee et al. [121], for instance, learned a similarity measure for multi-modal image registration using an algorithm derived from max-margin structured output learning. In Michel et al. [141], authors used boosted learning algorithms to embed data from different modalities into a common metric space. The metric of this space was then used to parametrize the similarity. Similarly, Simonovsky et al. [192] explored the use of supervised methods to learn similarity measures from a set of aligned examples. They presented a metric based on a CNN, which can be trained from scratch even from a few aligned image pairs.

Although the focus of learning-based approaches has been on approximating similarity measures, Gutiérrez-Becker et al. [74], proposed a regression method to learn optimization updates to perform multimodal registration.

Another approach consists in learning registration features, and inject them later into conventional registration algorithm to obtain deformation fields. Wu et al. [229], for instance, used a in-house conventional registration method with features learned using an auto-encoder instead of being manually extracted. Features obtained using a CNN were also used by Eppenhof et al. [48] to estimate the registration error.

These learning-based methods showed potential to outperform hand-crafted metrics. Yet, they learn only one element of the overall registration process, i.e. the similarity measure, the optimization step or the extraction of features, while recent computational developments allow the use of networks to produce deformation fields and perform fast and accurate registration. In the remainder of this chapter, we present such DL-based methods.

2.3.2 Convolutional neural network for image registration

While conventional registration methods rely on the pair of images to register, DL-based registration methods exploit larger amounts of data. These data are used to identify registration patterns to train a neural network. In general, a neural network can be considered as a function parameterized by a large number of parameters θ, i.e. the network weights. The straightforward DL-based approach to model the registration problem is to predict a deformation field ϕ given the image pairs (F, M) by training a function f_θ, such that:

$$\phi = f_\theta(F, M) \tag{2.8}$$

With this setup, the registration of a new image pair is possible in almost real-time, which is considered as one of the major benefits of DL-based registration. The selection of the optimal network parameters θ^* to learn the parameterized registration function f_θ is done optimizing a cost function, usually from a database $\{F_k, M_k \mapsto \phi_k\}_{k=1}^{K}$ with K

the number of image pairs in the database. As presented next, the cost function can be adapted to address the problem at hand [63].

The main type of neural networks used to perform image registration are CNNs. We next present these networks main characteristics. For an introduction about deep learning, we refer to the Annex A and to LeCun et al. [120]. The adverted reader may jump to Section 2.3.3, where we discuss specifically CNNs for registration.

2.3.2.1 Convolutional neural network layers

A CNN is a feed-forward network designed to process structured arrays of data. When fed as input to a CNN, images are represented in the form of matrices of pixel values. Images usually first go through a *convolutional layer*, which stacks convolution operations. A convolution is a simple mathematical operation generally used for image filtering (see Annex A.3). The values of the convolutional kernels are updated during the training phase of the network and the main function of a convolution layer is to extract feature from images.

Then, an *activation layer* applies a non-linear function to the feature maps extracted by the convolutional layer. The ReLU function is often chosen [235]. To better extract information and locate features, images are processed with multiple convolution filters in parallel and alternating convolution and ReLU layers (see an example in Figure 2.7).

Every feature map usually goes through a *pooling layer* after the activation layer. Pooling operations are used to down-sample feature maps. Reducing the dimensions of the feature map, and thus the number of network parameters, accelerates the computational time and improves the network generalization. The main pooling operations are *max-pooling* and *mean-pooling* [157]. Sliding a kernel (usually of size 2x2 or 2x2x2 with step 2) in the image, max-pooling takes the maximum value of each image tile, while mean-pooling returns a mean value for each image tile.

Finally, the output layer is designed to have the desired output dimension and range.

2.3.2.2 From CNN architectures to registration

Early CNNs architectures for classification include AlexNet [115], VGG [193], ResNet [79] or DenseNet [95]. In 2015, Ronneberger et al. [177] proposed the most widely used and studied architecture in medical image segmentation, the *U-Net*. This network is composed of an encoder and a decoder path (giving it its U shape), and skip-connections between the paths (see Figure 2.7). While the encoder reduces the input image dimensions with

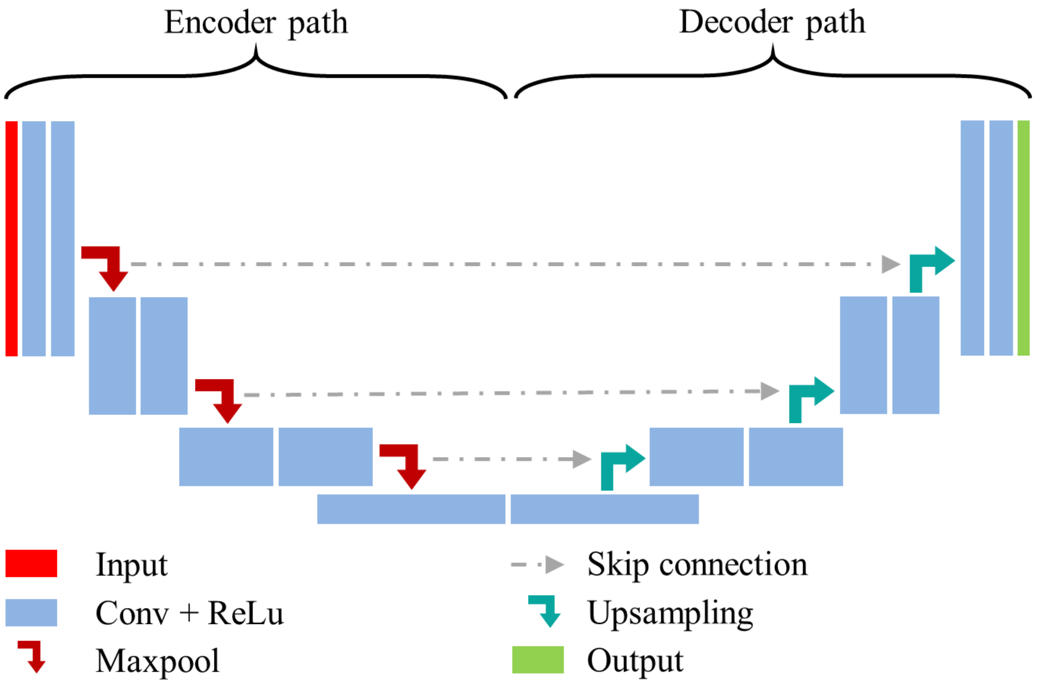


Figure 2.7 – U-Net architecture. The encoder path is on the left (blue blocks and red arrows) and the decoder path on the right (blue blocks and green arrows). Skip connections are visible in gray. Adapted from [177].

convolutional and pooling layers (*down-layers*), the same number of convolutional and upsampling layers (*up-layers*) compose the decoder, which reconstructs feature maps to original size of the input image. U-Net uses several down- and up-layers to learn features at different resolutions. Skip-connections transfer image information from the encoding to the decoding paths.

Today, U-Net is also the backbone for most state-of-the-art DL-based registration methods. However, other early attempts aimed to predict spatial geometric transformations. Spatial Transformer Networks (STNs), for instance, is a module designed to be incorporated into any existing network architecture to spatially manipulate data within the network. This module learns to spatially transform feature maps [98]. Even though it was not their initial purpose and because they are fully differentiable, STNs are now commonly used to warp the deformation field obtained in output of the CNN on the moving image in registration network.

Next, we discuss how the parameters of a registration network are learned in a supervised or unsupervised manner.

2.3.3 Supervised methods

In supervised registration, ground truth deformation fields ϕ_{GT} are available during training. They are either estimated using traditional registration methods or known from simulated transformations. This idea of working directly with ground truth dense deformation fields comes from optical flow, where large datasets with ground truth flow fields exist [42, 226]. In registration, the available ground truth deformation fields are used to guide the learning process. For all training image pairs (F, M) and associated ϕ_{GT}, the transformation $\phi = f_\theta(F, M)$ is predicted. These deformation fields can be represented as images with two or three channels (in 2D and 3D respectively). To compare the predicted and the ground truth deformation fields, a distance-based measure $\mathcal{S}_\phi$ such as SSD is used (see Figure 2.8). Thereby, the loss function to optimize takes the form:

$$\arg\min_\theta \mathcal{S}_\phi(\phi_{GT}, \phi) \tag{2.9}$$

Once the network is trained, supervised end-to-end approaches reach close to real-time performance. They can use ground truth deformation fields created either with random [48, 197] or statistically sampled [212] simulated transformations, or obtained with existing registration algorithms such as done by Cao et al. [25] with ANTs [7], or by Rohé et al. [174] with LCC log-Demons [129].

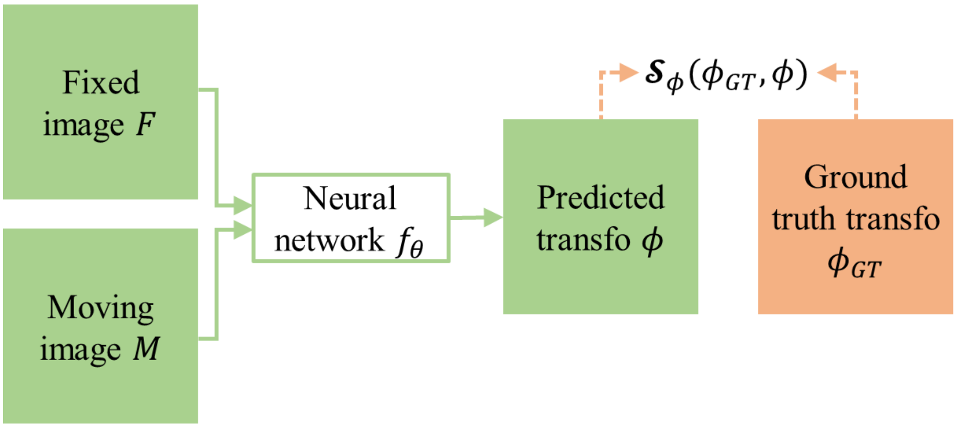


Figure 2.8 – Supervised registration. The predicted transformation ϕ, depending on θ, is directly compared to the ground truth ϕ_{GT}.

Supervised approaches do not need the definition of intensity similarity measures and may not require the definition of a regularization term. Their performance depends on the quality of the ground truth deformations. These transformations are difficult to obtain: while simulated deformation fields may not be realistic, creating them with existing registration algorithms is time consuming and depends on the performance of the baseline algorithm.

2.3.4 Unsupervised methods

Despite their relatively good performance, supervised methods suffer from the difficulty to obtain reliable ground truth transformations. Unsupervised approaches, on the contrary, do not depend on the need for deformation ground truth nor any expert labels.

Thanks to STNs [98], end-to-end DL-based methods can directly integrate transformation models in neural networks to perform an efficient and differentiable warping of the moving image (see Figure 2.9). These transformation models are based on B-splines or linear interpolation to produce dense deformation fields. The process being fully trainable, the registration objective function (see Eq. 2.2) optimizes the network parameters θ through the warped image $W = M \circ \phi$ with $\phi = f_\theta(F, M)$. Contrary to conventional registration methods, the registration parameters θ are not optimized over the image pair to register, but on all training image pairs (F, M), and any differentiable similarity measure $\mathcal{S}$ and regularization term $\mathcal{R}$:

$$\arg\min_{\theta} \mathcal{S}(F, W) + \lambda \mathcal{R}(\phi) \tag{2.10}$$

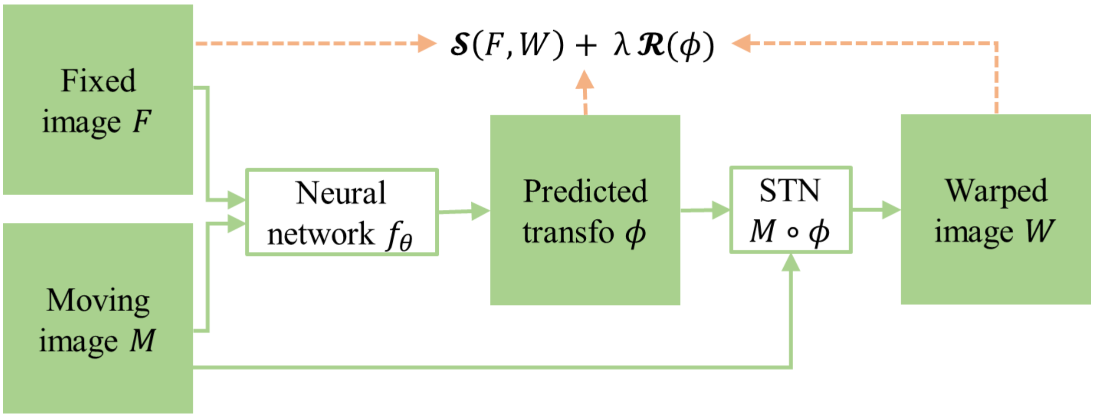


Figure 2.9 – Unsupervised registration. The fixed and warped images (F, W) are compared through $\mathcal{S}$. W is obtained with a Spatial Transformer Network (STN) and depends on θ through ϕ. A regularization term $\mathcal{R}$ and its hyperparameter λ are often added.

Obtaining an optimal deformation field using an unsupervised learning setup has generated significant interest from the research community recently. In this field, the VoxelMorph network [12] is now considered as a reference registration architecture. It is based on a U-Net shape network, which produces a deformation field enabling a spatial transformer to warp the moving image to the fixed one on brain MR volumes. VoxelMorph was further improved by Dalca et al. [37], who took advantage of SVFs to produce diffeomorphic deformation fields, as for instance the conventional method proposed by Ashburner [5]. Also in the unsupervised registration field, Vos et al. [221] performed coarse-to-fine registration staking affine and deformable image registration networks. Thereby, they performed global (i.e. affine) registration refined by local (i.e. deformable) transformations. On cardiac cine MRI and on chest CT and at the cost of an increased complexity, they reached performances similar to conventional registration methods, while being significantly faster. As with the pyramidal optimization scheme of conventional methods, multi-scale DL-based registration has been proposed by Mok et al. [149]. In this work, authors also took advantage of the SVF approach proposed by Dalca et al. [37] to produce diffeomorphic transformation.

All the above methods build on similarity measures and regularization terms taken from conventional registration methods (see Sections 2.2.1 and 2.2.3). The MSE and CC similarity measures were for instance used by Balakrishnan et al. [12], while Vos et al. [221] and Mok et al. [149] used the NCC. Regarding regularization terms, the bending energy, i.e. a smoothness constrain on the deformation field imposed by the gradients of the displacements, was used by all approaches.

2.3.5 Dual and weakly-supervised methods

Dual supervision refers to the use of both the transformation similarity $\mathcal{S}_\phi$ and the similarity measure $\mathcal{S}$ from unsupervised methods. While $\mathcal{S}_\phi$ is computed between the obtained ϕ and ground truth transformations ϕ_{GT}, $\mathcal{S}$ quantifies the fixed and warped (F, W) image similarity, with $W = M \circ \phi$ and $\phi = f_\theta(F, M)$ (see Figure 2.10). An additional regularization term $\mathcal{R}$ on ϕ is often added:

$$\arg\min_{\theta} \mathcal{S}(F, W) + \mathcal{S}_\phi(\phi_{GT}, \phi) + \lambda\mathcal{R}(\phi) \tag{2.11}$$

In the dual supervision category, Fan et al. [49] predicted the deformation field for 3D brain MR registration focusing first on learning from supervised deformation fields and later increasing the weight of the unsupervised objective $\mathcal{S}(F, W)$ for fine-tuning. Yan et al. [232] also used a dual approach with a Generative Adversial Network (GAN) to perform rigid registration between 3D MRI and TransRectal UltraSound (TRUS) volumes.

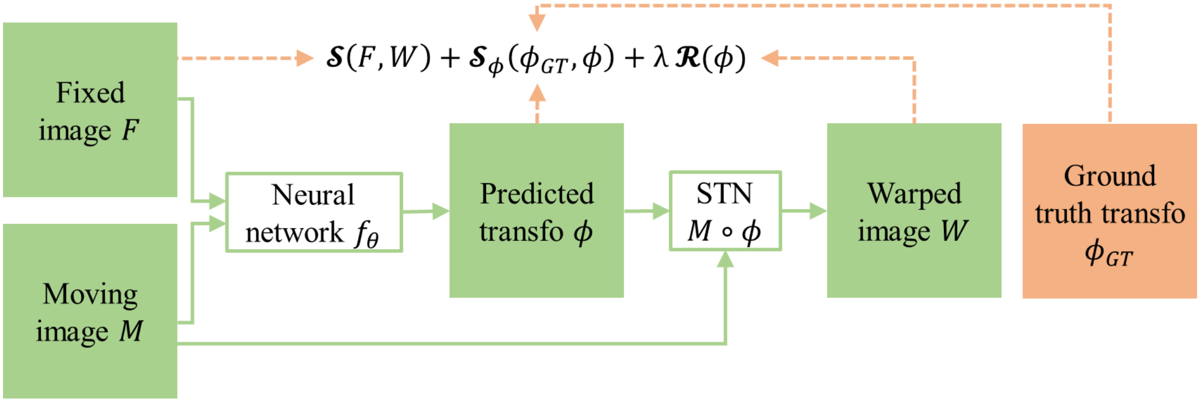


Figure 2.10 – Dual supervised registration. The predicted transformation ϕ, depending on θ, is compared to the ground truth ϕ_{GT}. The fixed and warped images (F, W) are also compared through $\mathcal{S}$. W is obtained with a Spatial Transformer Network (STN) block and also depends on θ through ϕ. A regularization term $\mathcal{R}$ on ϕ and its hyperparameter λ are often added.

Instead of ground truth transformations, weakly supervision combines the similarity measure between images $\mathcal{S}$, with a similarity measure $\mathcal{D}$ between segmentation masks or matching landmarks, here denoted (F_{labels}, M_{labels}) (see Figure 2.11). W_{labels} are warped labels obtained similarly to warped images: $W_{labels} = M_{labels} \circ \phi$:

$$\arg\min_\theta \mathcal{S}(F, W) + \mathcal{D}(F_{labels}, W_{labels}) + \lambda\mathcal{R}(\phi) \tag{2.12}$$

Hu et al. [94] combined the optimization of label matching based on a multi-scale Dice loss and a deformation regularization to perform MRI - TRUS registration. Building upon these methods, Hering et al. [89] introduced a similarity and segmentation loss respectively based on an edge-based normalized gradient fields distance measure and an SSD to register 2D cine–MRI images.

Although dual supervised approaches do not fully depend on ground truth deformation fields, they still require a certain amount of simulated transformations. For weakly supervised approaches, the anatomical guidance coming from segmentation masks or landmarks helps improve registration performance, especially in areas of interest. Unlike conventional methods using feature-based similarity measures, DL-based approaches require

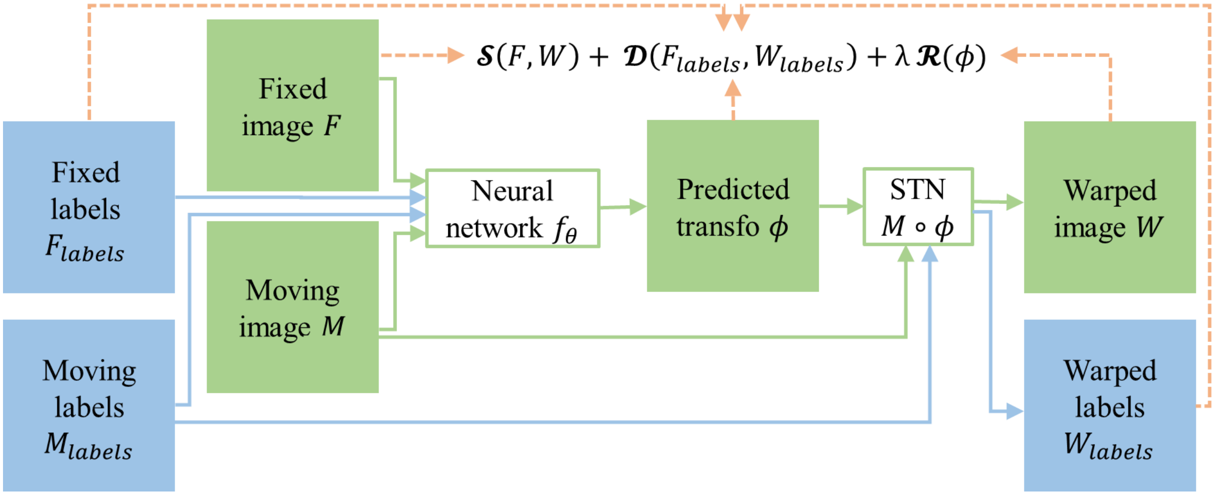


Figure 2.11 – Weakly supervised registration. The fixed and warped images (F, W) are compared through $\mathcal{S}$, while the fixed and warped segmentations (F_{labels}, W_{labels}) are compared through $\mathcal{D}$. W and W_{labels} are obtained with a Spatial Transformer Network (STN) and depend on θ through ϕ. A regularization term $\mathcal{R}$ and its hyperparameter λ are often added.

extra label information only at the training stage, and not for testing [28], and performing registration with a trained network is still done almost in real time.

Overall, DL-based registration methods are inspired by elements of conventional methods, such as the similarity measures, the regularization terms of the optimization schemes. Even if promising and reaching performances comparable to conventional methods in a shorter computational duration, DL-based methods present some limitations. Indeed, since deformations are learned by a network over a large database, it is important that this database represents the whole range of possible transformations. Moreover, it is easier to train a network on simple and homogeneous registration task than on complex and very variable deformations for which generalization may be a problem [67]. Last but not least, as for conventional methods, choices of the similarity measure, regularization term, optimization scheme or even on the network architecture need to be decided according to the considered registration problem.

2.4 Conclusion

Image registration is a fundamental problem in the field of medical image analysis. In oncology, it is used clinically to help physicians assess changes in lesions over time and to combine information captured by different imaging modalities. Image registration is complex and a major area of research where both conventional and DL-based registrations are today studied.

Yet, both approaches are challenged by the tuning of hyperparameters for the different applications. Moreover, conventional approaches require a consequent computational time, even though they are now commonly used in clinical practice. DL-based approaches, on the contrary, perform registration in almost real time but need a large database representative of the problem at hand. Moreover, image registration is one of the fields where DL-based methods do not largely surpass the performances of conventional methods.

In Chapter 4 of this thesis, we present a new registration method for lesion and organ tracking from full body PET images. This method, called Medical Image Registration Regularized By Architecture (MIRRBA), aims to bridge the gap between conventional and DL-based registration approaches.

Summary:

✓ Cancer evolution can be formulated as a longitudinal image registration problem.

✓ Image registration has been an active research field for over 30 years and DL-based registration for about 7 years.

✓ Conventional and DL-based methods reach comparable results, but DL-based methods are faster.

✓ Most state-of-the-art DL-based methods use features inspired by conventional method to a great extent.

✓ The performance of DL-based methods depends significantly on the size and representativeness of the training database.

Chapter 3

PET image registration and label propagation using conventional registration

Contents

3.1 Introduction

Metastatic breast cancer is usually monitored using the PERCIST guidelines presented in Section 1.4. Yet, these criteria suffer from some limitations, such as the need for manual lesion delineation. Moreover, PERCIST only considers the lesion presenting the highest uptake, which can be different from one acquisition to another.

To help better assess disease evolution at a global level, i.e. on the whole body and for each lesion, our goal is to follow the individual evolution of every lesion on all the acquisitions. However, performing lesion segmentations manually and finding correspondences between acquisitions is time consuming and unpractical. For this reason, we propose in this chapter a method to automatically obtain segmentations of cancerous lesions on a longitudinal dataset.

3.1.1 Related work: longitudinal data segmentation

"Longitudinal data" refers to data acquired on the same patient at different time points. They are, for instance, acquired in the context of disease monitoring or radiotherapy planning. While radiotherapy planning may involve the manual segmentation of Organs At Risk (OARs), disease monitoring may require the segmentation of cancerous lesions. Since manual delineations are time consuming and error-prone, automatic methods have been developed to specifically segment longitudinal data [190]. As presented in the following sections, these methods can be divided into two categories: segmentation and registration-based methods. Next, we review some recent DL contributions in both directions.

3.1.1.1 Segmentation-based methods

Segmentation-based methods seek to propagate a segmentation mask from one image to the next.

In the context of video processing, several recent approaches address the segmentation propagation problem in contiguous video frames [234]. In Oh et al. [163], for instance, a deep siamese encoder-decoder network was designed to take advantage of mask propagation and object detection at the same time. Mask propagation was also used in Xu et al. [230]. Both methods yielded good results, especially in the case of missing objects from one frame to the following. Lai et al. [116] propagated annotations, when given the first frame annotations. In practice, they exploited the spatio-temporal coherence of the annotations, and improved the model robustness using forward - backward consistency (training the network from t to $t+1$ and vice versa). This approach was later followed by some medical image registration methods to obtain consistent deformations (see Section 2.2.3).

However, while small changes occur between consecutive video frames, important modifications can be observed between images from a patient acquired at different time points. For this reason, video frame to frame segmentation methods are difficult to apply to disease monitoring and treatment planning. Moving to the medical image field, most pipelines comparing images perform registration as a preprocessing step. They often use conventional methods, such as ANTs [7] or Elastix [112] to this end.

Designing patient-specific networks, i.e. a network per set of images acquired on a single patient, Wang et al. [224] segmented lung tumors on longitudinal MRI sequences. Each network was trained on t_0, t_1, ..., $\mathrm{t}_{\mathrm{n-1}}$ images and corresponding segmentations ($n \sim 10$ in the paper), while the inference was ran on the t_n image to automatically obtain its segmentation. A similar approach was used in Elmahdy et al. [47], who pre-trained a CNN on a large CT prostate database. Then, the network was fine-tuned on images of the patient of interest only. The resultant patient-specific network was used to segment the new incoming image. Yet, with this method, there is no correspondence between the segmented regions, i.e. no per lesion follow-up monitoring.

Even though promising, these patient-specific networks require either at least 10 segmented acquisitions, or a large database to pre-train the network. During this thesis, we had access on average to two to three metastatic breast cancer acquisitions per patient, and to the best of our knowledge, no segmented public database was available for pre-training. Beside, these segmentation-based propagation methods are computed in two steps: a prior registration and a subsequent segmentation. This two-staged approach is error prone, since a non accurate registration will result in a poor segmentation.

3.1.1.2 Registration-based methods

In the radiotherapy planning field, Ghose et al. [66] reviewed methods to propagate prostate segmentations on longitudinal datasets. Several registration-based approaches have been used, such as Rueckert et al. [179] on MRIs with consecutive affine and deformable cubic B-spline registrations, as well as Acosta et al. [2] on CT images using affine and deformable parametric registrations. Later, Paganelli et al. [166] reviewed DL-based registration algorithms used to propagate the segmentations of OARs between acquisitions. Yet, authors insisted on the lack of accurate and efficient patient-specific validation of the registration accuracy.

Regarding disease monitoring, Zhuang et al. [240] developed a two-step conventional registration for the automatic determination of patients' cardiac morphology evolution. On MRIs, a first affine registration provides the correspondence of anatomical substructures, while the second free-form registration refines the local details using a constrained optimization scheme. Similarly, Calmon et al. [24] automatically measured changes in the volume of a structure of interest in successive 3D brain MR images using a two steps registration method. First a rigid registration was performed, followed by a deformable registration. Later, Heckemann et al. [83] also used nonrigid free-form deformations but to register healthy brain anatomical structures.

According to various reviews [28, 63, 78], few DL-based registration methods have been specifically developed and tested on longitudinal data. The recently proposed BraTS-Reg[1] challenge may reveal interesting approaches in this specific field of research at the publication of the results.

In Kemp et al. [107], the authors compared different methods (no registration, slice match and 3D rigid registration) to follow longitudinal changes in bone density and structure. The 3D registration approach best captured physiological longitudinal changes. Eijnatten et al. [45] performed registration between longitudinal CT images presenting bone metastasis from breast cancer. After removing the CT table and other extra-corporal components from the images, they used the registration network VoxelMorph [12]. To improve the network's results, the approach relies on an incremental learning strategy based on simulated deformations of consecutive CT images of the longitudinal dataset. The strategy consists in learning from a constantly arriving data-stream, which means that deformed CT images are sequentially presented to the network in chronological mini-batches per patient. Because of the small amount of data available, i.e. 5 to 10 acquisitions per patient, authors used simulated deformations, which may have limited the ability of registration algorithm to produce realistic transformations.

Longitudinal patient follow-up can be decomposed into a segmentation and registration component. Therefore, joint segmentation–registration approaches have been proposed. Li et al. [122] trained two networks, i.e. one for segmentation and one for registration, optimized jointly through a global loss. This loss presented segmentation and registration specific terms, as well as a regularization term. From 8045 longitudinal brain MRI of 3249

---

[1] https://www.med.upenn.edu/cbica/brats-reg-challenge/

individuals, the method aimed to segment white matter tracts, however the specificity of each patient was not considered or exploited.

Even though promising, DL-based approaches are not applicable to our metastatic breast cancer dataset and lesion monitoring goal. Indeed, the full body, multiple heterogeneous lesions and images acquired several months apart images of the $\text{EPICURE}_{\text{seinmeta}}$ dataset are challenging to register. Since we would like to avoid the limitations implied by synthetic transformations, our small number of available data (between two and three acquisitions per patient) is a challenge to accurately train a registration network that generalizes across patients. Yet, as shown by Paganelli et al. [166], most registration methods currently used to propagate segmentations on longitudinal data are based on conventional registration methods and are not specifically designed for this task.

3.1.2 Contributions

We formulate the longitudinal follow-up of metastatic breast cancer patients from whole body PET images as an image registration problem and propose a first solution based on conventional methods. For generality, we evaluate the performances on two datasets of 3D real data.

The proposed approach was first ran on a inhale–exhale lung CT public database. In this setup, the goal is to perform intra-patient registration from inhale (moving) to exhale (fixed) images, to propagate landmarks located in lung salient features. This work was presented in a conference paper [57] published in the context of the Learn2Reg 2020 challenge[2], held in conjunction with the MICCAI 2020 conference.

We then ran our longitudinal conventional registration pipeline on the private metastatic breast cancer dataset $\text{EPICURE}_{\text{seinmeta}}$. The proposed approach enables us to propagate segmentations previously obtained on the baseline acquisitions to the following ones, for each patient.

Our objective is to get a high performing conventional approach to compare later against DL-based methods (see Chapter 4). The main novelty of our method is the application domain, since conventional registration methods have not yet been used in $\text{EPICURE}_{\text{seinmeta}}$ clinical protocol, neither in research. We integrated the proposed conventional longitudinal image registration pipeline in the software of the Keosys company, to facilitate the segmentation of lesions on new follow-up data.

---

[2] https://learn2reg.grand-challenge.org/

3.2 Intra-patient longitudinal registration

The objective of intra-patient longitudinal registration is to find a warping function relying on an intensity-based similarity measure that aligns the fixed F and moving M images.

First global, i.e. rigid and affine, transformations are sough and optimized minimizing the cost function $\mathcal{L}_{non\_deformable}$ using an adaptive stochastic gradient descent optimizer [112]. Since both fixed and moving images (F, M) are from the same modality, $\mathcal{S}$ corresponds to the NCC measure:

$$\mathcal{L}_{non\_deformable}(\phi) = -\mathcal{S}(F, M \circ \phi) \tag{3.1}$$

To refine the global registration and obtain accurate local transformations, the deformable registration method minimized the loss function $\mathcal{L}_{deformable}$ also with the adaptive SGD optimizer. Moreover, to ensure smooth and realistic-looking deformable deformations, the similarity metric $\mathcal{S}$ is regularized by $\mathcal{R}_{smooth}$. It is a bending energy penalty term which corresponds to the second derivative of the deformation ϕ [182]:

$$\mathcal{L}_{deformable}(\phi) = -\mathcal{S}(F, M \circ \phi) + \lambda \mathcal{R}_{smooth}(\phi) \tag{3.2}$$

with:

$$\begin{split}\mathcal{R}_{smooth}(\phi) = \frac{1}{V} \int_0^X \int_0^Y \int_0^Z \Bigg[\left(\frac{\partial^2 \phi}{\partial x^2}\right)^2 + \left(\frac{\partial^2 \phi}{\partial y^2}\right)^2 + \left(\frac{\partial^2 \phi}{\partial z^2}\right)^2 + \\ 2\left(\frac{\partial^2 \phi}{\partial xy}\right)^2 + 2\left(\frac{\partial^2 \phi}{\partial xz}\right)^2 + 2\left(\frac{\partial^2 \phi}{\partial yz}\right)^2 \Bigg] dxdydz\end{split} \tag{3.3}$$

Our iterative registration method uses a B-spline transformation with several pyramidal resolutions. Using a pyramidal strategy refers to a coarse-to-fine optimization of the registration cost function, i.e. the optimization is performed first on low resolution images and then on higher resolution ones. Usually, the resolution of images is divided by two for each new pyramidal level.

Let's consider a strategy with three pyramidal levels, i.e. *low*, *mid* and *high*. The parameters of the registration function f are first optimized through the cost function on low resolution images (F_{low_res}, M_{low_res}) for a fixed number of iterations to obtain a low resolution deformation field ϕ_{low_res} (see Eq. 3.4). Then, the cost function is again optimized on higher resolution images $(F_{mid\_res}, W_{mid\_res})$, with $W_{mid\_res} = M_{mid\_res} \circ \phi_{low\_res\_up}$ and $\phi_{low\_res\_up}$ the deformation field previously obtained upsampled ϕ_{low_res} (see Eq. 3.5). This process is followed until original resolution images are registered (see Eq. 3.6).

$$\phi_{low\_res} = f(F_{low\_res}, M_{low\_res}) \tag{3.4}$$

$$\phi_{mid\_res} = f(F_{mid\_res}, W_{mid\_res}) \tag{3.5}$$

$$\phi_{high\_res} = f(F_{high\_res}, W_{high\_res}) \tag{3.6}$$

Once the optimal full resolution transformation ϕ^* is computed, it is applied to the moving image to get the warped one: $W = M \circ \phi^*$. Similarly, ϕ^* is applied to the moving segmentation M_{seg} to get the warped segmentation: $W_{seg} = M_{seg} \circ \phi^*$.

3.3 Experimental setup

3.3.1 Datasets

Lung CT Learn2Reg MICCAI 2020 satellite event is a registration challenge [38] consisting of four different tasks covering a wide range of medical image registration topics: multi-modality, noisy annotations, small datasets and large deformations. We focused on Task 2, which consists in registering the inhale (moving) to the exhale (fixed) image, thus to perform intra-subject inhale–exhale lung CT scans registration.

The dataset of the Task 2 of Learn2Reg challenge consists of 60 mono-centric thorax CT images from 30 subjects [90]. A pair of images (inhale and exhale scans) is available per subject (see Figure 3.1). A segmentation mask of the lungs is also available for each volume. Images size is 192 × 192 × 208 and spacing is 1.75 × 1.25 × 1.75 mm. Even if the dataset was split by challenge organizers into 17 training, 3 validation and 10 testing pairs, we treat every pair of images independently, since we do not use training images under the conventional approach.

All images provided by challenge organizers were already preprocessed to the same spatial dimension and voxel resolution, as well as affinely pre-registered.

As visible in Figure 3.1, the Field Of View (FOV) of exhale images is reduced compared to the one of inhale images. To reduce registration unrealistic deformations, we decided to align the FOV of both fixed and moving images. Thus, for each exhale image slice where the body of the subject in not visible, we set values of the corresponding inhale slice to 0.

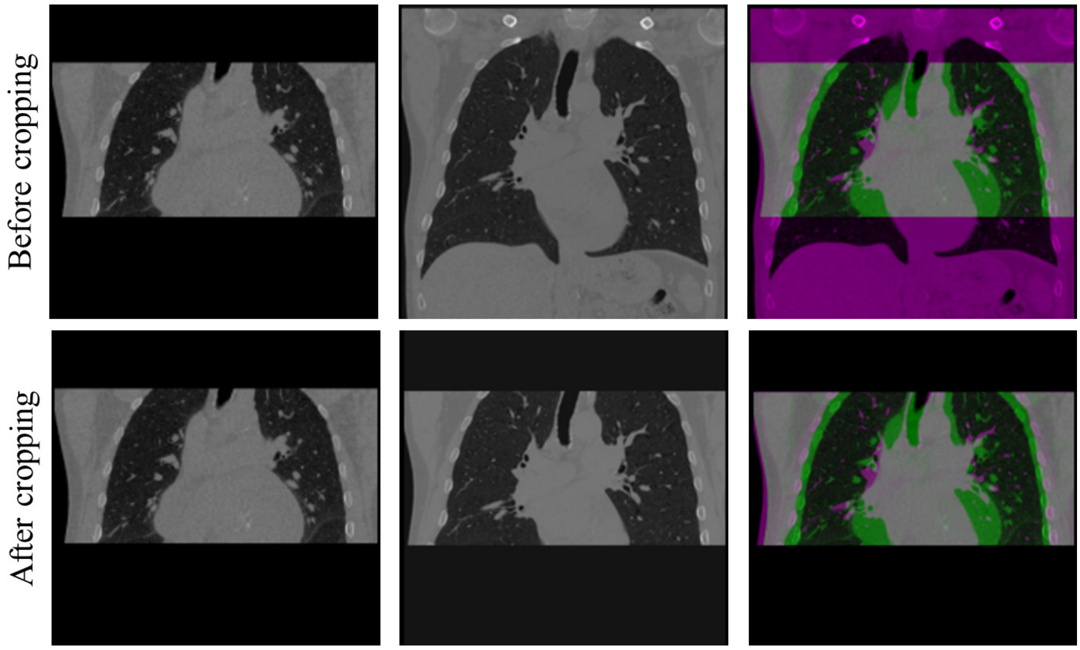
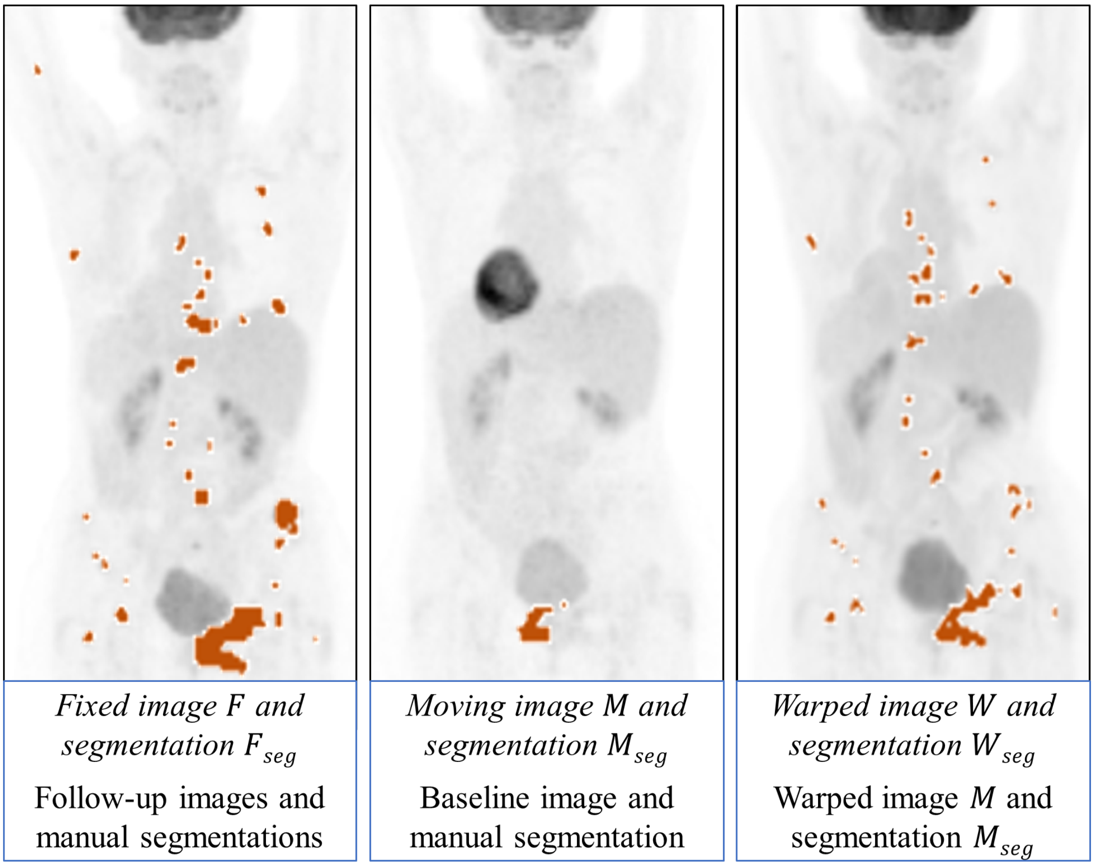


Figure 3.1 – Exhale, inhale and overlay of both images (1^{st}, 2^{nd} and 3^{rd} column respectively) before and after (1^{st} and 2^{nd} row) cropping the inhale image.

Since we modified voxel values only for null slices, some small FOV misalignment can persist in the image borders (see 3^{rd} column, Figure 3.4).

EPICURE$_{\text{seinmeta}}$ Following the previously presented work on inhale–exhale lung CT images, we performed deformable registration on our private EPICURE$_{\text{seinmeta}}$ dataset, already presented in Section 1.5, to propagate lesion segmentations from the baseline acquisitions to the follow-up ones. In this dataset, two to three PET images were acquired per patient and all lesions were delineated by experts on all images. To perform intra-patient registration, moving data corresponds to the baseline acquisitions and segmentations, while fixed data are the corresponding follow-up acquisitions and segmentations (see Figure 3.2).

Here, a total number of 115 image pairs were processed, with 61 image pairs corresponding to pre- and early-treatment acquisition times, and 54 image pairs corresponding to pre- and mid-treatment times. 58 and 57 pairs of images were respectively acquired in the Angers and Nantes centers.



Fixed image F and segmentation F_{seg}
Follow-up images and manual segmentations

Moving image M and segmentation M_{seg}
Baseline image and manual segmentation

Warped image W and segmentation W_{seg}
Warped image M and segmentation M_{seg}

Figure 3.2 – Fixed and moving PET images and segmentations (red) used to perform registration with the EPICURE$_{\text{seinmeta}}$ dataset.

3.3.2 Implementation details

For both datasets, the number of resolutions was chosen according to the initial image size and to ensure the main image features are visible even on the coarsest pyramidal level.

On the EPICURE$_{\text{seinmeta}}$ longitudinal dataset, we performed intra-patient registration in three steps. We first performed a rigid registration, followed by an affine and a deformable one. Both global registrations, i.e. rigid and affine, were performed with three pyramidal resolutions optimized for 500 iterations each. For the deformable registration, four pyramidal resolutions and 1000 iterations were used to better optimize the high number of parameters.

Since lung CT images are already pre-registered by the challenge organizers, we only performed the deformable registration step on this dataset, with similar setting as for the

EPICURE$_{\text{seinmeta}}$ dataset.

Hyperparameters such as the number of iterations or the bending energy parameter λ were determined empirically from the settings recommended in the state-of-the-art. The parameter files used to obtain these transformations are visible in Annex B. On the both datasets, we evaluated identity (i.e. no registration) and deformable registration.

3.3.3 Evaluation metrics

3.3.3.1 Lung CT

Two evaluation metrics were used within the scope of the challenge.

The Target Registration Error (TRE) evaluates registration precision. It is computed on 100 landmarks automatically selected on fixed images, while correspondences in moving images are manually annotated. The landmarks are extrapolated on lungs salient features, like vessels or nodules. For each landmark $k \in [1:K]$, the TRE measures the distance between the fixed $F_{landmarks}$ and warped $W_{landmarks}$ landmarks:

$$\text{TRE} = \sqrt{\sum_{i=1}^{K}(F_{landmarks}(k) - W_{landmarks}(k))^2} \tag{3.7}$$

The registration smoothness is quantified by the standard deviation of the logarithm of the Jacobian determinant of the deformation field (SDLogJ).

For both metrics, a lower value indicates a better registration.

3.3.3.2 EPICURE$_{\text{seinmeta}}$

The first criterion used to evaluate registration accuracy on the EPICURE$_{\text{seinmeta}}$ dataset is the detection rate. It is defined as the percentage of baseline and registered follow-up segmentation maps presenting an overlap equal or greater than 50%, as illustrated in Figure 3.3 [151].

To evaluate more precisely the registered segmentations, we also computed the Dice metric between the fixed F_{seg} and registered W_{seg} binary segmentation maps of the lesions:

$$Dice = \frac{2|F_{seg} \bigcap W_{seg}|}{|F_{seg}| + |W_{seg}|} \tag{3.8}$$

Both metric values lie between 0 and 1, the lower score being 0 and the best one 1.

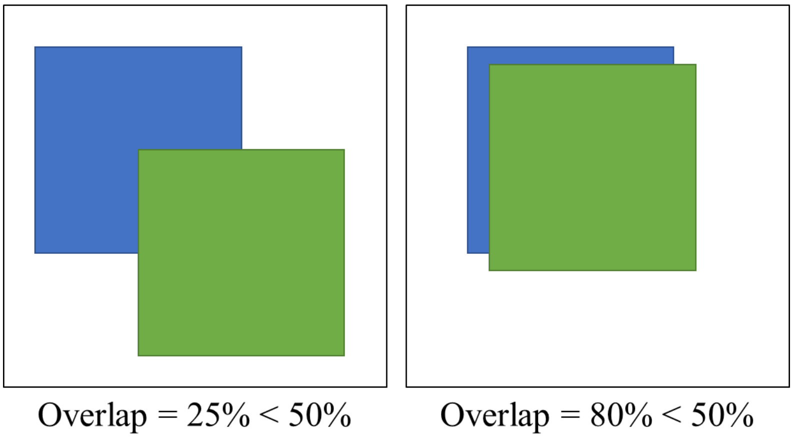


Figure 3.3 – Blue and green squares respectively represent two compared segmentation masks. Left: the overlay of both masks is lower than 50%, so the detection rate for this segmentation is 0. Right: the overlay of both masks is higher than 50%, so the detection rate for this segmentation is 1.

The average computational time to obtain these registrations on an Intel Core i7-8750H CPU is also given.

3.4 Results

3.4.1 Lung CT

With the proposed Elastix method, we obtained a TRE of 6.55 ± 2.69 mm and a SDLogJ of 0.07 ± 0.03 on all the different image pairs of the test set. Subject-wise results are visible Table 3.1. For reference, the initial error between fixed and moving images (after affine registration only) was 10.24 ± 2.72 mm. Table 3.2 shows these results, in comparison to DL-based methods participating to the challenge.

Table 3.1 – Results on the 10 subject from the test dataset.

Subject	1	2	3	4	5	6	7	8	9	10	Mean
TRE	7.78	6.84	7.42	5.19	9.57	11.70	5.19	1.35	5.61	4.82	6.55±2.69
SDLogJ	0.08	0.17	0.07	0.06	0.07	0.07	0.04	0.05	0.06	0.07	0.07±0.03

Our low value of SDlogJ confirms that the proposed method creates smooth deformations and realistic-looking images. However, the relatively higher TRE value reflects a lack of precise registration inside the lungs, as illustrated, in the 6th row of Figure 3.4.

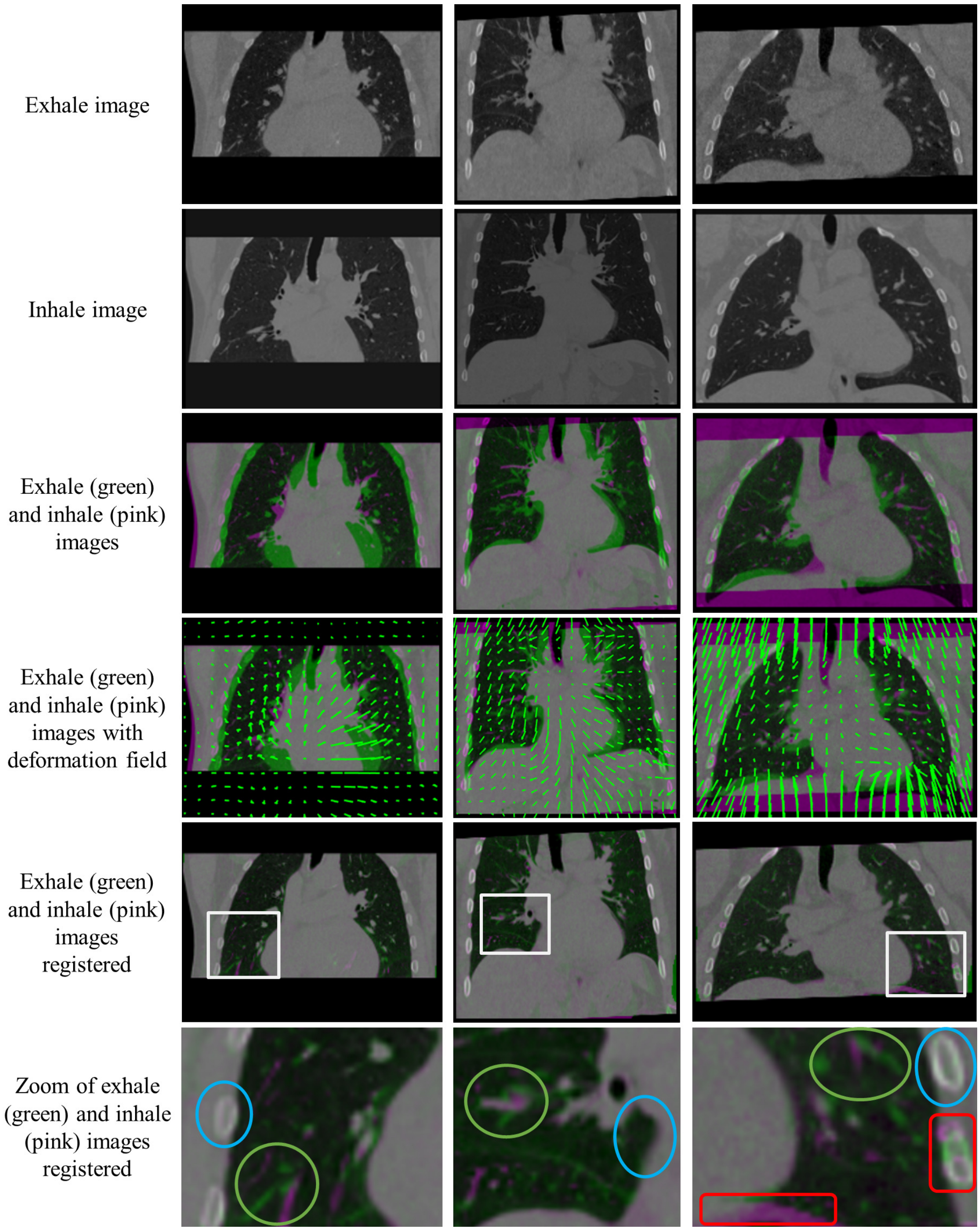


Figure 3.4 – For three validation subjects (resp. each column), exhale image, inhale image, overlay of exhale and inhale image, overlay of exhale and inhale images with the deformation field from inhale to inhale in green and overlay of exhale and registered inhale images (1st, 2nd, 3rd, 4th and 5th row resp.). The 6th row zooms inside white squares of the 5th row images. In this last row, body parts seem accurately registered (blue circles), but registration approximations are visible within the inner lung regions (green circles). Larger registration errors due to misaligned initial fields of view are highlighted in the bottom right image by red boxes.

Table 3.2 – Comparison of our results with the ones obtained using deep learning-based approaches for the Task 2 of the Learn2Reg challenge.

	TRE	SDLogJ
Identity	10.24	0.00
Tony (LapIRN)	3.24	0.06
Lasse (PDD-Net)	2.46	0.07
Niklas (Uppsala)	9.00	0.12
Elastix (ours)	**6.55**	**0.07**

3.4.2 EPICURE$_{\text{seinmeta}}$

A visual representation of the results is shown in Figure 3.5. While the identity registration did not perform well, the rigid and affine warped segmentations were not precise nor adapted to the lesions. The segmentation propagated with deformable registration, on the contrary, was close to the the manual annotations.

Detection rates and Dice scores values are presented in Table 3.3. For both metrics, the identity registration score is very low. Rigid and affine registrations present similar low values, while the deformable registration reached the best results, demonstrating its usefulness for the task.

As shown in Table 3.3, the computational time increased with the complexity of the method. Identity registration is simpler than a rigid transformation (9 parameters), which itself requires less parameters than an affine registration (12 parameters), while a deformable registration optimizes thousands of parameters. Deformable registration is thus the slowest method.

Table 3.3 – Detection rate, Dice score and computation times values obtained after performing identity, rigid, affine and deformable registration on 115 image pairs. Best detection rate and Dice score are marked in *italic*.

	Detection rate	Dice score	Computation time
Identity	0.046 ± 0.210	0.054 ± 0.111	5 sec
Rigid	0.376 ± 0.487	0.292 ± 0.202	27 sec
Affine	0.385 ± 0.489	0.303 ± 0.200	110 sec
Deformable	*0.835 ± 0.373*	*0.745 ± 0.272*	3000 sec

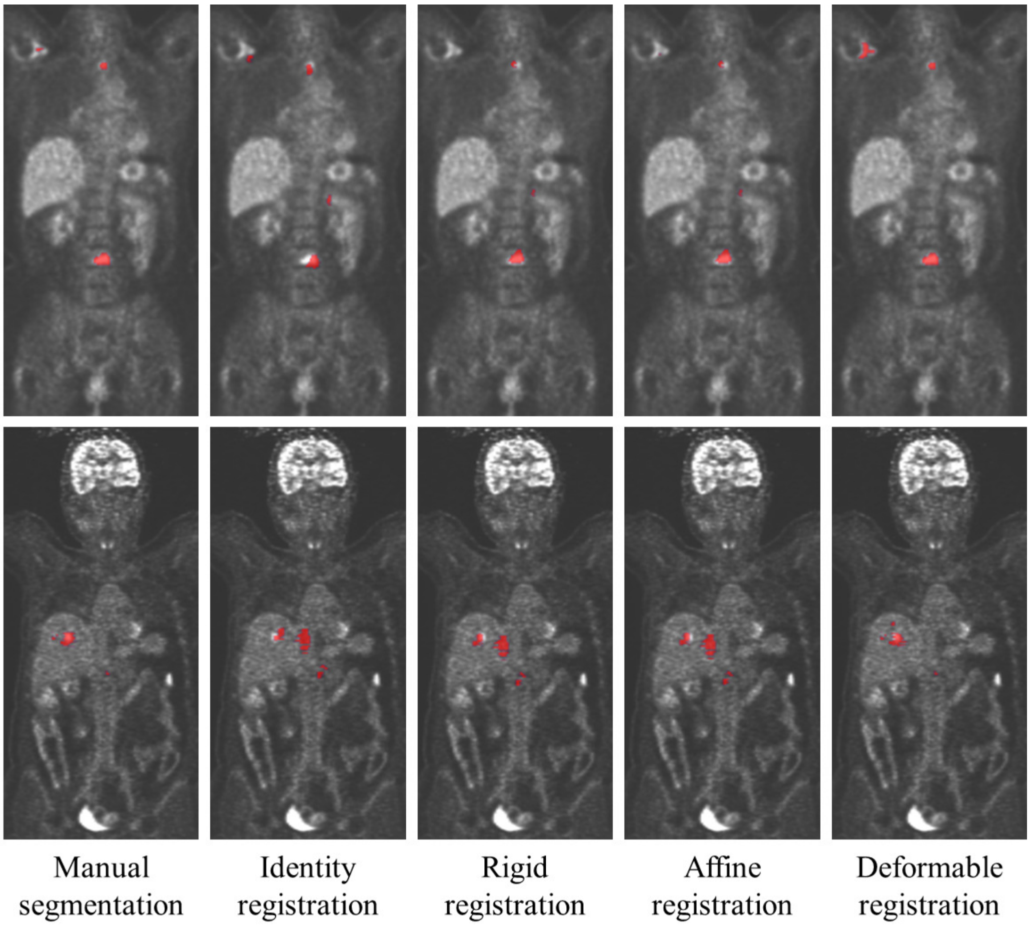


Figure 3.5 – Lesion segmentation (in red) on two EPICURE$_{\text{seinmeta}}$ follow-up acquisitions. The manually performed segmentation is visible in the 1st column. In the remaining columns, the manual segmentation of the initial image is propagated to the follow-up using identity (2nd column), rigid (3rd column), affine (4th column) and deformable (5th column) registration computed as described in Section 3.3.2.

3.5 Discussion

In this chapter, we propose a conventional registration method for the Task 2 of the Learn2Reg challenge and for our EPICURE$_{\text{seinmeta}}$ dataset, based on a NCC measure and a bending energy regularization term.

On the lung CT dataset, both fixed and moving images did not present the same FOV. Hence, we cropped images slices to obtain more realistic registration results. Without performing this prior FOV alignment, our method performs poorly, as illustrated on image borders by red boxes, last row of Figure 3.4. Hence, it seems important to align the FOV of fixed and moving images to reach better registration results. With the proposed loss, regularization and registration method, the TRE inside the lungs is high. Results could be improved masking the body to focus the registration only on the lungs in a second registration step.

With regards to the learning-based methods, our computation time is longer, yet we do not require time-expensive prior training. Since the available training dataset is relatively small, we show that a conventional method remains an option of interest. Also, our results reach similar smoothness, but slightly lower accuracy over the testing dataset. Regarding the use of hyperparameters, both conventional and learning-based methods require appropriate values for the number of resolutions/network depth, the number of iterations/number of epochs, the regularization weight, etc.

On the $\text{EPICURE}_{\text{seinmeta}}$ dataset, non-deformable registration methods do not accurately register metastatic breast cancer lesions in an intra-patient setup. Indeed, rigid registration only initiates global translations and rotations. Rigid global alignment may be useful to improve the overall overlapping of the bodies to register, but lesions do not globally translate or rotate within the body. Similarly, an affine transformation performs global translation, rotation, scaling and shearing. Yet, individual lesions located in different areas in a patient body may evolve differently, making a global registration method inadequate to follow their evolution.

On the other hand, deformable registration produces local deformations. This type of registration suits more the monitoring of individual lesions, since each one evolves differently. This is highlighted by the high deformable registration scores.

Overall, our approach provides reasonable registration results and our code is publicly available at: `https://github.com/fconstance/Learn2Reg_Task2_SimpleElastix`. We provide all the implementation details, and hyperparameters, such that the proposed method could serve as a non-learning based baseline for comparison.

3.6 Conclusion

Although not specifically designed to register longitudinal data, the proposed registration method to propagate landmark and segmentation obtained promising results.

Regarding the intra-patient lung registration, our conventional registration method reached similar smoothness but slightly lower accuracy than DL-based methods, such as Mok et al. [149] or Heinrich [84]. We also reached good accuracy on the $\text{EPICURE}_{\text{seinmeta}}$

data, successively performing conventional rigid, affine and deformable registrations.

One of the main advantages of using a conventional method to register data is the possibility to register image pairs presenting heterogeneous deformation patterns. On the contrary, trained DL-based registration methods are restricted by the deformation patterns learned during the training step. If the training dataset is not representative of the whole set of possible deformations, the registration of image pairs presenting unseen deformations will be challenging for the trained network.

Even if DL-based methods require time-expensive prior training, they are able to perform deformable registration in almost real-time, whereas our method requires about an hour of processing for a pair of full-body PET volumes.

In the following chapters, we propose a method aiming to connect conventional and DL-based methods by taking advantage of both approaches. Moreover, we aim to propose a registration method specifically developed for longitudinal data registration and segmentation propagation.

We also demonstrat in Chapter 6 how to use the proposed convolutional registration method to define a computer-aided solution to assist the evaluation of multi-lesion PERCIST criteria.

Summary:

✓ We demonstrate on a public and a private dataset how conventional registration methods can be used to propagate landmarks or segmentations between volumes.

✓ Quantitative results on the PET data show the need of modeling the problem with a deformable transformation.

✓ Comparable results between conventional and DL-based method for the challenge data show a trade-off between data and speed.

✓ We introduce direct and indirect criteria to evaluate the success of the propagation tasks.

Chapter 4

MIRRBA: a bridge between conventional and deep learning registration

Contents

4.1 Introduction

As presented in Chapter 3, image registration is a widely used approach to propagate segmentations between longitudinal data. Yet, only few methods have been specially developed to monitor patient-specific evolution (see Section 3.1.1). Among these, DL-based registration methods start to be competitive, but are still under development. In addition, despite trained networks being faster than conventional methods, the training step requires large databases which often restricts their use to a specific therapeutic area for which enough training data is available. Therefore, Computer-Aided Diagnosis (CAD) methods, if any, are based on conventional registration approaches in clinical practice [166].

As an alternative, our approach is to develop a registration method i) specifically designed for patient monitoring, and ii) that takes advantage of the strengths of both conventional and DL-based registration methods. In this chapter, we propose a new registration method formalized as a conventional registration approach, with a deformation field modeled by an untrained deep pyramidal network. We named our approach MIRRBA for *Medical Image Registration method Regularized By Architecture.*

We applied MIRRBA on our longitudinal PET images acquired for the evaluation of treatment response in patients with metastatic breast cancer and compared the results to registrations obtained with standard approaches. We also studied the impact of different architecture configurations (16 configurations) on the deformation field. This work was submitted and is currently under revision for publication in the journal Physics in Medicine and Biology (PMB) [54].

In addition to the private longitudinal PET dataset, we tested our method on a longitudinal brain MRI public dataset. Even though MIRRBA was designed to propagate metastatic breast cancer lesions, we did not make specific adaptations to our method for this new dataset, showing the generality of the proposed approach.

4.2 Related work

Longitudinal image registration can be a key step in lesion monitoring. As presented in Section 2.1, many conventional registration methods were proposed [111, 200]. Among the best performing ones, are the SimpleElastix method [112] based on cubic B-splines, and the Advanced Normalization Tools (ANTs) [7] which parametrizes the velocity field and enforces bi-directional diffeomorphisms. Both methods use a pyramidal coarse-to-fine optimization scheme, which has inspired recent works on DL-based registration, as presented hereafter. In this chapter, we use these two conventional registration methods as reference approaches.

Using DL, the monitoring of metastatic breast cancer can be achieved through unsupervised or weakly-supervised registration methods, avoiding the need for ground truth deformation fields. Indeed, the deformation fields between longitudinal acquisitions are not available. Among the unsupervised and weakly-supervised registration methods, various CNN architectures have been proposed, as presented in Section 2.3.4. Among them, LapIRN [149] is a state-of-the-art architecture which won the Learn2Reg 2021 MICCAI challenge[1]. LapIRN uses a pyramidal network trained in a coarse-to-fine manner to mimic best performing conventional approaches. Moreover, it regularizes the deformation field through its gradient to impose a smoothness constraint, and enforces diffeomorphic transformations using stationary vector fields under the Log-Euclidean framework, as in Dalca et al. [37].

Despite these recent efforts, no major gain in registration performance was reported when using DL-based registration techniques [85]. Larger databases may be needed to learn the network parameters and produce more accurate deformation fields. Moreover, we can argue that the generalization ability of the trained network is questionable when the deformation patterns do not repeat consistently across the dataset.

To alleviate the dependency to a dataset and the need for repeated deformation patterns, we adapted the Deep Image Prior (DIP) framework [211] to propose a learning-free method for deformable medical image registration. Contrary to standard DL-based approaches, DIP does not learn from a database but relies on a single image. It uses a deep architecture not to summarize the information across samples but as a prior. Effectively, the architecture plays the role of a parametric model in an optimization problem restricting the solution space. While DIP was initially designed for denoising and inpainting

---

1. https://learn2reg.grand-challenge.org/

tasks [211], we adapt it here to medical image registration.

Laves et al. [119] suggested image registration as a potential application of DIP, with preliminary results in the context of 2D brain MR images. DIP has also been used on medical data for the reconstruction of CT and PET images [9, 69]. To perform these reconstructions, an untrained deep network was used as a denoiser and iteratively optimized by a conventional algorithm. Apart from the use of an initial reconstruction and classical regularization, images from other modalities were also given as input to condition the network output.

With this work, we establish a link between conventional, deep learning and DIP-based approaches. In particular, we focus on the role of the untrained network in parameterizing the displacement field, showing that each architecture induces an implicit regularization when used within a conventional optimization scheme. Similar observations have been made in the context of DL for inverse problems [41, 131]. As a result, well-structured architectures (e.g. Mok et al. [149]) provide better over-parameterizations for the deformation fields, both in the trained and in the untrained cases. We also investigate the role of the input (random, moving or moving and fixed images) and potential interactions of the untrained network with conventional supervised approaches.

The contributions of this chapter are i) the proposition of a novel registration method regularized by architecture MIRRBA, ii) an extensive comparative study of the effects of different network components on the deformation field, and iii) a solution to register 3D whole body longitudinal PET images without the need for prior registration to facilitate the simultaneous monitoring of multiple lesions. We also briefly show that, without specific adaptation, our DIP-based registration method performes well on longitudinal brain MRI data.

4.3 Method

As in the previous chapter, we propose an approach for assisting the problem of multiple lesion segmentation propagation by modeling it as an image registration problem. Hence we seek the deformation field ϕ between the fixed F and moving M images, which for the main application of interest are PET volumes acquired several months apart. As presented hereafter, the main difference relies on the modelization of the deformation field, which makes a link between DL-based registration methods (see Section 4.3.1) and the DIP prior concept (see Section 4.3.2).

4.3.1 Background on DL-based image registration

As presented in Section 2.3, DL-based registration approaches model the deformation ϕ to be the output of a CNN trained on a dataset of image pairs (F, M) (see Eq. 4.1). The inputs of this network are the $(F,\ M)$ pairs, and the operations to get warped images W are performed using a spatial transformer layer [98], handling the transformation, the sampling and the interpolation. Regarding the regularization, DL-based registration methods rely on a term enforcing the smoothness of the transformation $\mathcal{R}_{smooth}$, as done in conventional approaches. Smoothness is often completed by $\mathcal{R}_{diffeo}$, which enforces the diffeomorphism of the transformation by penalizing the determinant of the Jacobian negative values. Regularization terms are weighted respectively by λ_{smooth} and λ_{diffeo} (see Section 2.2.3).

Moreover, we state that two additional regularizing priors are also implicitly added in the DL setup and influence the predicted deformation field ϕ: the first one $\mathcal{R}_{dataset}$ is induced by training on a domain-specific dataset, while the second one $\mathcal{R}_{archi}$ is entailed by the network architecture choice (see Figure 4.2.B). Due to the dataset dependency, one limitation of DL-based registration methods is the difficulty to generalize across organs or modalities.

DL image registration can thus be seen as an optimization problem such as:

$$\underset{\phi(\mathcal{R}_{dataset}, \mathcal{R}_{archi})}{\arg\min} \ \mathcal{S}(F, W) + \lambda_{smooth}\mathcal{R}_{smooth}(\phi) + \lambda_{diffeo}\mathcal{R}_{diffeo}(\phi) \tag{4.1}$$

Next, we recall the deep image prior concept, which removes the dataset dependency in Section 4.3.2, and present our method for registration based on an untrained network in Section 4.3.3.

4.3.2 Deep Image Prior

The DIP method proposed by Ulyanov et al. [211] uses a deep architecture to denoise images using a network without any prior learning. Supposing X_0 a distorted image and X the network output, the fitting process is characterized by Eq. 4.2, with $\mathcal{R}_{archi}$ the implicit prior captured by the network architecture and $\mathcal{S}_{DIP}$ a reconstruction function on a single image.

$$\underset{X(\mathcal{R}_{archi})}{\arg\min} \ \mathcal{S}_{DIP}(X_0, X) \tag{4.2}$$

DIP reconstructs a noisy image (e.g. with JPEG compression noise) from a white

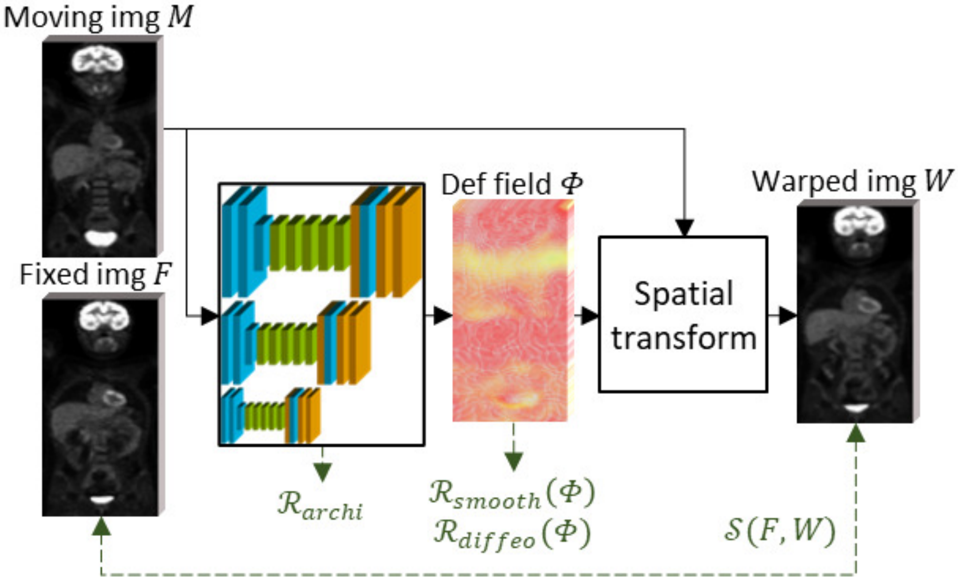


Figure 4.1 – The `MIRRBA` method, which is regularized implicitly by its architecture and directly by smoothing and diffeomorphic terms. The network architecture is visible with the encoder path (blue), the residual blocks (green) and the decoder path (orange with a blue layer from the encoder path). Input images are in 3D.

noise image by training a generator architecture to fit the noisy image. A denoised image is obtained by stopping the training before completely fitting the noise. We propose to adapt this idea to the registration of a pair of images, where we modify the moving image to match the fixed one. Contrary to the original DIP, we are interested in fitting all the way up to the finest deformations.

4.3.3 MIRRBA (Medical Image Registration method Regularized by Architecture)

We argue here that deep architectures, as parametric models with high capacity, are powerful representations for deformation fields. They can thus be exploited as implicit priors for image registration in iterative optimization schemes without a training stage [41, 82, 131]. Although DIP priors have been explored in the context of image reconstruction, there is no prior in-depth study of such architecture priors in the context of image registration.

Since there is no learning step, the training dataset no longer influences the transformation. Thus, we transform Eq. 4.1 into Eq. 4.3 to directly optimize a patient-specific

CNN for the pair of images of interest (F, M), as traditionally done with iterative optimization methods (see Figure 4.1):

$$\underset{\phi(\mathcal{R}_{archi})}{\arg\min}\, \mathcal{S}(F, W) + \lambda_{smooth}\mathcal{R}_{smooth}(\phi) + \lambda_{diffeo}\mathcal{R}_{diffeo}(\phi) \tag{4.3}$$

By optimizing Eq. 4.3, we find the best warped image allowed by the over-parametrization of the architecture. Carefully designing the architecture to be used, it is possible to integrate several of the tricks commonly used in conventional approaches [84]. We rely here on the LapIRN architecture [149], which incorporates filtering through convolutions, pyramidal coarse-to-fine refinement, and interpolation steps with down- and transpose convolutions. Eq. 4.3 is optimized with an SGD for an input pair of images. Our scheme applies to other architectures and optimization algorithms.

4.4 Experimental validation

4.4.1 Datasets description

EPICURE$_{\textbf{seinmeta}}$ dataset We ran our experiments on longitudinal PET images from the metastatic breast cancer dataset acquired in the context of the EPICURE$_{\text{seinmeta}}$ project and presented in Section 1.5. For this experiment, a total number of 110 pairs of images were obtained, a pair being composed of a pre-treatment and either an early- or a mid-treatment image (58 and 52 images respectively). 54 pairs of images were acquired in Angers and 56 pairs come from the Nantes center.

Since we are interested in lesion monitoring in the context of metastatic breast cancer, we worked only on full-body PET images, as shown useful in previous studies by Carlier et al. [27] and by Avril et al. [8]. Expert physicians manually segmented all lesions. 469 lesions were present on both baseline and follow-up pair of images and 155 lesions disappeared. Overall, a total of 624 lesions with a volume of 57.12 ± 189.65 cm^3 and 6 ± 4 lesions per patient were studied. The brain and the bladder were also delineated, since they can be useful to mask irrelevant regions for patient response assessment.

As preprocessing steps, all PET images were normalized by the SUV [109] and reshaped to size 200 × 200 × 200. The mean spacing is 3.66 × 3.66 × 4.96 mm. Besides, no prior registration of any kind was performed.

BraTSReg dataset We used the multi-institutional retrospective brain MRI data relieved in the context of the BraTSReg challenge[2]. The dataset is composed of pre-operative and follow-up image pairs acquired on 140 patients diagnosed and treated for brain glioma. For each time point, native (T1), Contrast-Enhanced T1-weighted (T1-CE), T2-weighted (T2) and T2 Fluid Attenuated Inversion Recovery (FLAIR) MR images are available. Here, we ran our method only on the FLAIR images. All images were de-identified and pre-processed to isotropic images of size $240 \times 240 \times 155$.

On this dataset, the goal is to perform intra-patient registration to establish correspondence between pre-operative and follow-up brain MRIs. Indeed, pre-operative brain tissues show heavy deformations because of the tumor pressure, which is relieved after its resection during the surgery. These heavy deformations are not limited to the tumor neighborhood but affect the whole brain. Moreover, treatment-induced changes, potential tumor recurrence, as well as modifications of the infiltrated tissue may also affect the brain tissue elasticity. Yet, the resected brain area is the main registration challenge, since it may cause missing correspondences between the pre- and post-operative scans.

To evaluate the registration accuracy, ground truth landmarks were placed on each pre-operative scan and corresponding follow-up images by expert clinical neuro-radiologists. These landmarks were placed on anatomical salient features, such as the anatomical shape of the cortex, blood vessel bifurcations or the midline of the brain. The total number of landmarks vary from case to case from 6 to 50 per scan.

4.4.2 Architectural implementation details

Our MIRRBA method relies on the LapIRN network architecture proposed in Mok et al. [149]. LapIRN is a pyramidal network with $N = 3$ depth levels, each level being composed of a feature encoder, a set of residual blocks and a feature decoder, as shown in Figure 4.2. Two convolutional layers with stride 1 and a convolutional layer with stride 2 compose the encoder path. For each level, they are $R = 5$ residual blocks, each containing two convolutional layers and a residual connection. The decoder path is composed of a transpose convolutional layer and two convolutional layers with stride 1, followed by a SoftSign activation. To prevent information loss, a skip connection from the encoder to the decoder path is added. Each convolution layer has a kernel size of 3, 4 filters, and is followed by a leaky rectified linear unit (LeakyReLU) activation [133] with a negative

---

[2] https://www.med.upenn.edu/cbica/brats-reg-challenge/

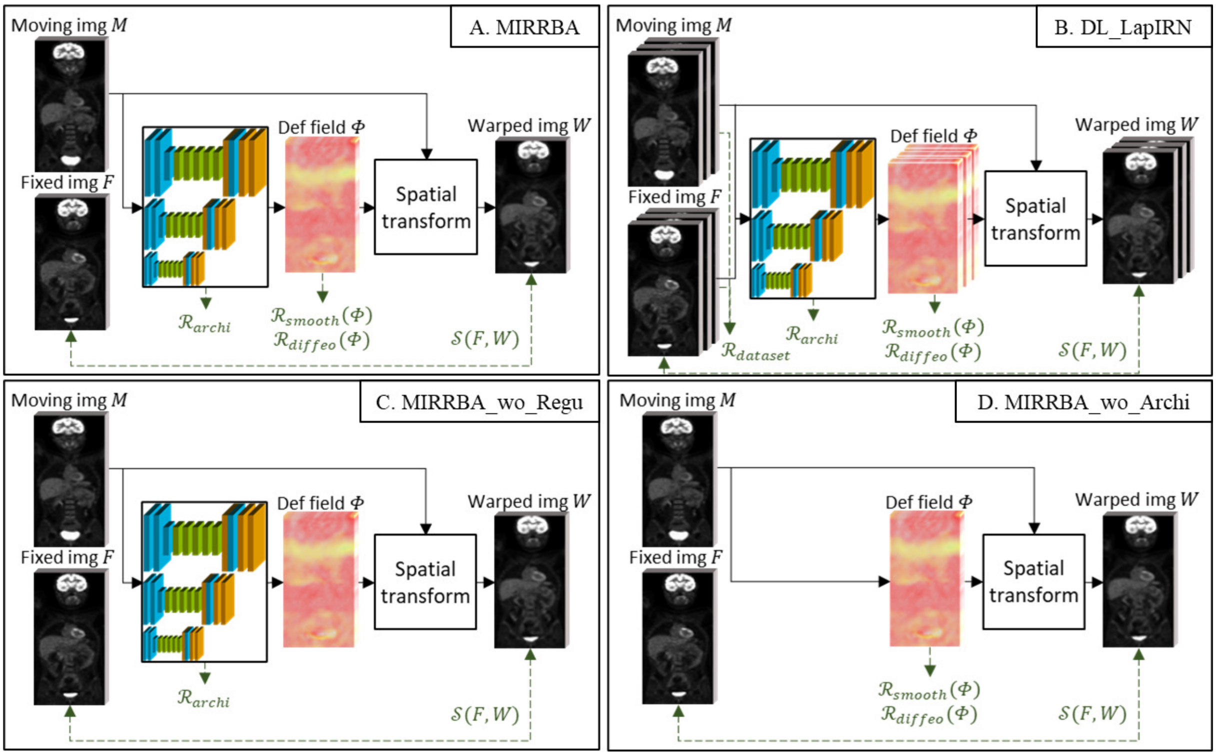


Figure 4.2 – Overview of A. `MIRRBA`, B. `DL_LapIRN`, C. `MIRRBA_wo_Regu` (`MIRRBA` without the registration-specific regularization terms) and D. `MIRRBA_wo_Archi` (`MIRRBA` without the network architecture and implicit regularization) methods. The network architecture is visible with the encoder path (blue), the residual blocks (green) and the decoder path (orange with a blue layer from the encoder path). Input images are in 3D.

slope of 0.2, except for the output convolution layers.

For each level $L_{i\in\{1,2,3\}}$, input images are downsampled by a factor $2^{(N-i)}$ using a trilinear interpolation. Hence, for the coarsest level L_1 the image resolution is divided by 4, while for the finest, L_3, it remains identical. Moreover, a scaling and squaring module enforces diffeomorphic deformations [37]. Network levels were trained in a coarse-to-fine manner, meaning the coarsest level L_1 is first trained alone, and then higher levels are progressively trained to refine the registration. To avoid unstable starts when training levels $L_{i>1}$, lower levels weights were frozen for a fixed number of epochs. Regarding the optimization process, the learning rate was set to 10^{-4}, and the Adam optimizer was used for 1000 iterations on the two lower levels and 2000 iterations on the finest. This number of iterations was chosen empirically as it allows the method to reach convergence.

We used the NCC similarity measure regularized by smooth and diffeomorphic terms. λ_{smooth} and λ_{diffeo} were set to 0.1 and 1.0 respectively for MIRRBA methods. We choose hyperparameters that minimized the cost function Eq. 4.3 i) over all image pairs for conventional and MIRRBA approaches, and ii) for each fold and across folds for DL-based methods. We followed the same hyperparameter set strategy over all compared methods (DL, conventional and MIRRBA) for fair comparison. Indeed, the loss function is unsupervised for all compared methods and expected to correlate with the performance measures.

Starting from the architecture in Figure 4.1, we performed an ablation study to measure the influence of each network component on the registration performance evaluated on the the EPICURE$_{\text{seinmeta}}$ dataset. As detailed in Table 4.1, we:

(i) Changed the depth of the network (network with 1, 2, 3 or 4 depth levels);

(ii) Computed the results after optimizing each level during the coarse-to-fine registration i.e. the coarsest, the intermediate and the finest;

(iii) Deleted the residual connections of residual blocks to transform them into simple convolutional blocks;

(iv) Replaced the down- and up-convolutions by respectively max-pooling and upsampling operations;

(v) Used deformable convolutions [36] in the finest level;

(vi) Used a Gaussian noise image $\mathcal{N}(0, 1)$ (as in Laves et al. [119]), and used both the fixed and moving images as input channels to the network;

(vii) Set both λ_{smooth} and λ_{diffeo} to 0 (individually and jointly) to remove the registration-specific regularization terms (see Figure 4.2.C).

We registered the images from the BraTSReg dataset with the best performing MIRRBA configuration in terms of Dice scores and detection rate, i.e. `MIRRBA_Best` as presented hereafter.

4.4.3 Reference methods implementation details

Regarding the methods used as reference, we used a grid search on all image pairs to find hyperparameters yielding best performance running the ANTs pipeline [7], i.e. with a three resolutions coarse-to-fine optimization, a gradient step of 0.2 and a symmetry transformation penalty. We used both MI and NCC as similarity measures, leading to `SyN` and `SyNCC` methods respectively. The method refining the results of an initial `SyNCC` registration with the best performing MIRRBA configuration is presented as `Combined`.

Table 4.1 – Loss terms and structural setups of the methods presented in the chapter. "Pyr. net." stands for "Pyramidal network", "Sym. diffeo. transfo." for "Symmetric diffeomorphic transformation", "Max." for "Maxpooling", "Up." for "Upsampling", and "Def. conv." for "Deformable convolutions". An absence of information in the *Other* column indicates the use of setups described in Sections 4.4.2 and 4.4.3.

	Loss terms				Structural choices				
	$\mathcal{R}_{dataset}$	$\mathcal{R}_{smooth}$	$\mathcal{R}_{diffeo}$	$\mathcal{R}_{archi}$	Model	Depth	Trained levels	Input images	Other
`MIRRBA`		x	x	x	Pyr. net.	3	3	Moving	-
`DL_LapIRN`	x	x	x	x	Pyr. net.	3	3	Fixed & moving	-
`DL_Voxelmorph`	x	x	x	x	U-shaped network	1	1	Fixed & moving	-
`Elastix`		x			Cubic B-splines	4	4	Moving	-
`SyN`		x			Sym. diffeo. transfo.	3	3	Moving	-
`SyNCC`		x			Sym. diffeo. transfo.	3	3	Moving	-
`Combined`		x	x	x	Sym. diffeo. transfo. & Pyr. net.	3 - 4	3 - 4	Fixed & moving	-
`MIRRBA_wo_Smooth`			x	x	Pyr. net.	3	3	Moving	-
`MIRRBA_wo_Diffeo`		x		x	Pyr. net.	3	3	Moving	-
`MIRRBA_wo_Regu`				x	Pyr. net.	3	3	Moving	-
`MIRRBA_wo_Archi`		x	x		Pyr. net.	3	3	Moving	-
`MIRRBA_Depth_1`		x	x	x	Pyr. net.	1	1	Moving	-
`MIRRBA_Depth_2`		x	x	x	Pyr. net.	2	2	Moving	-
`MIRRBA_Depth_4`		x	x	x	Pyr. net.	4	4	Moving	-
`MIRRBA_Level_1`		x	x	x	Pyr. net.	3	1	Moving	-
`MIRRBA_Level_2`		x	x	x	Pyr. net.	3	2	Moving	-
`MIRRBA_Max`		x	x	x	Pyr. net.	3	3	Moving	Max.
`MIRRBA_Up`		x	x	x	Pyr. net.	3	3	Moving	Up.
`MIRRBA_DefConv`		x	x	x	Pyr. net.	3	3	Moving	Def. conv.
`MIRRBA_NoiseImg`		x	x	x	Pyr. net.	3	3	White noise	-
`MIRRBA_FixImg`		x	x	x	Pyr. net.	3	3	Fixed & moving	-
`MIRRBA_Best`		x	x	x	Pyr. net.	4	4	Fixed & moving	Max. Up. Def. conv.

We also ran the Elastix pipeline [112] to perform successive rigid, affine and deformable image registration with four resolutions optimized with an adaptive SGD minimizing the NCC similarity term for 1000 iterations (`Elastix`). As for our method, we used a penalty on the bending energy as regularization to enforce smooth deformations (see Chapter 3 and Annex B for more implementation details).

To highlight the regularization power of the architecture in MIRRBA, we ran the registration pipeline without the network, optimizing directly the deformation field (initialized from a Gaussian noise $\mathcal{N}(0, 0.001)$) distribution with the Adam optimizer. The method is referred as to `MIRRBA_wo_Archi` (see Table 4.1 and Figure 4.2.D).

Regarding deep learning-based methods, `DL_LapIRN` [149] was ran with a similar regularization balance as MIRRBA, i.e. with $\lambda_{smooth} = 0.1$, while all other recommended settings were used. For `DL_Voxelmorph`, we used the latest diffeomorphic version to date [37], with the NCC loss and recommended settings. For methods relying on a training stage, we split our dataset into five folds, paying attention to balance data from different acquisition centers among folds. For each fold, we refer to $\mathcal{D}_{\text{train}}$ and $\mathcal{D}_{\text{test}}$ as the train and test dataset respectively. We trained the DL-based approaches on $\mathcal{D}_{\text{train}}$ before testing them on $\mathcal{D}_{\text{test}}$ for the five folds. We report the average performance scores over patients in each $\mathcal{D}_{\text{test}}$ and then across folds.

All architectures were implemented with PyTorch [168] and trained from scratch on a Nvidia V100 32GB SXM2 GPU.

4.4.4 Evaluation measures

On the $\text{EPICURE}_{\text{seinmeta}}$ dataset, the first criterion we used to evaluate the registration accuracy is the *detection rate*, defined as the percentage of detected lesions. A lesion is considered detected if the overlap between its warped and fixed segmentation maps is greater than 50% [151] (see Figure 3.3 in Section 3.3.3). A high detection rate is sought.

To evaluate the registration of the objects of interest, we computed *Dice scores* between fixed and warped i) brain and bladder and ii) lesions segmentation masks. A Dice score close to 1 suggests a precise local registration, while a Dice close to 0 is unsatisfactory. Since the detection rate represents the percentage of lesions correctly detected, it is positively correlated with the Dice score of the lesions. However, since some lesions (155) are cured over time and disappear on PET images, we removed them from the Dice score and detection rate computations to avoid erroneous null values. Instead, we evaluated the capacity of a method to effectively make lesions disappear by computing its *disappearing* *rate*, or percentage of volume reduction of a lesion induced by the registration, where a complete disappearance would mean a rate of 100%.

One of the interests of using the BraTSReg dataset is to have a direct registration performance measure (as opposed to indirect measure through registration of segmentation maps). Indeed, as advised by the BraTSReg challenge organizers, we quantitatively measured the performance of MIRRBA on the brain MRI dataset using the distance between the fixed $F_{landmark}$ and corresponding warped $W_{landmark}$ landmarks using the Mean Absolute Error (MAE) for each landmark $i \in [1:N]$:

$$\text{MAE}(F_{landmark}, W_{landmark}) = \frac{1}{N}\sum_{i=1}^{N} |F_{landmark\_i} - W_{landmark\_i}| \tag{4.4}$$

Moreover, we also evaluated the registration between the fixed and registered images themselves with the NCC.

Registration smoothness was evaluated on both datasets by measuring, for every deformation field, the standard deviation of its Jacobian determinant *SDJDet*. Null values indicate no transformation and high ones disorganized and incoherent displacements. Although an optimal value is difficult to define, we seek to obtain small positive values, known to characterize smooth deformations [149].

Finally, we evaluated the approximate running time of each approach, using a CPU for `SyN`, `SyNCC` and `Elastix`, and a GPU for all other methods.

4.4.5 Statistical analysis

To evaluate the statistical significance of our results, we first studied their distribution. According to the Shapiro-Wilk test (testing the null hypothesis that a sample comes from a normal distribution), we cannot reject the null hypothesis. Hence, we ran a paired sample t-test on our results and considered them statistically significant if $p < 0.05$.

4.5 Results and discussion

4.5.1 Regularization terms

In the first experiment, we looked at the influence of the different regularization terms from Eq. 4.1, i.e. $\mathcal{R}_{smooth}$, $\mathcal{R}_{diffeo}$, $\mathcal{R}_{dataset}$ and $\mathcal{R}_{archi}$. On top of smoothness $\mathcal{R}_{smooth}$ and diffeomorphisms $\mathcal{R}_{diffeo}$ explicit regularization terms, DL-based methods rely on a dataset

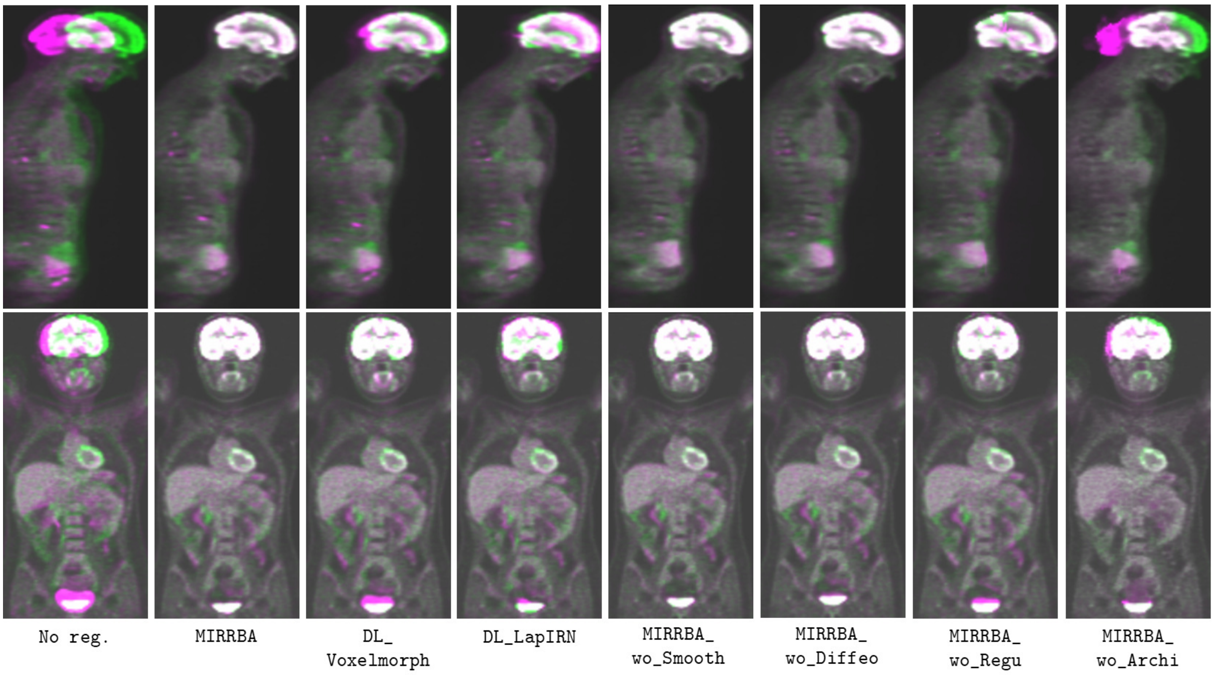


Figure 4.3 – Regularization terms – Overlay of the fixed (green) and warped (pink) images on two different patients after performing `No reg.`, `MIRRBA`, `DL_Voxelmorph`, `DL_LapIRN`, `MIRRBA_wo_Smooth`, `MIRRBA_wo_Diffeo`, `MIRRBA_wo_Regu`, and `MIRRBA_wo_Archi`. Grayscale color indicates good overlapping. It can be noticed that DL-based methods have difficulties to register the bladder because of its important deformation. Even without registration-specific regularization terms, `MIRRBA_wo_Diffeo`, `MIRRBA_wo_Regu` and `MIRRBA_wo_regu` warped images look realistic, unlike `MIRRBA_wo_archi`.

$\mathcal{R}_{dataset}$ and architecture-based $\mathcal{R}_{archi}$ implicit regularization. The MIRRBA method only uses the latter, as detailed in Table 4.1, Figure 4.2 and Eq. 4.3.

Dataset regularization We trained two state-of-the-art DL-based methods `DL_Voxelmorph` [37] and `DL_LapIRN` [149] with the four regularization terms from Eq. 4.1 (see Table 4.1 and Figure 4.2.B). Predictions were made on $\mathcal{D}_{\text{test}}$ images.

Even if both methods reached similar results quantitatively (see Table 4.2) and qualitatively (see Figure 4.3), `MIRRBA` performed better than both of them. Indeed, from Table 4.2, we note that running a patient-specific optimization, i.e. removing $\mathcal{R}_{dataset}$ and optimizing an untrained network `MIRRBA` compared to `DL_Voxelmorph` and `DL_LapIRN`, helps to improve the results, especially for the lesions. `MIRRBA` improved the organs' Dice score by 6% and 5% and the Dice of the lesions improved by 52% and 65%, w.r.t. `DL_Voxelmorph` and `DL_LapIRN`.

Table 4.2 – Comparison of `MIRRBA` to DL-based methods (Section 4.5.1) – Dice scores of the organs and of the lesions, detection and disappearing rates of the lesions, SDJDet and approximate computational time. Both training and inference computational times are indicated for DL-based methods. All pipelines were computed on GPU. Statistically significant improvement of our `MIRRBA` method over the others with $p < 0.05$ is indicated with *. Best results are marked in bold, except for SDJDet since no ideal value is defined.

	Dice organs ↑	Dice lesions ↑	Detec. rate (%) ↑	Disap. rate (%) ↑	SDJDet ↓	Time (min) ↓
`No reg.`	0.626 ± 0.138*	0.090 ± 0.115*	5.00	0.00*	0.000 ± 0.000	0
`MIRRBA`	**0.918 ± 0.126**	**0.425 ± 0.207**	**33.04**	9.36	0.124 ± 0.988	55
`DL_LapIRN`	0.878 ± 0.076*	0.258 ± 0.198*	11.76	**19.15**	0.464 ± 1.367	1450 – 3
`DL_Voxelmorph`	0.865 ± 0.077*	0.279 ± 0.192*	14.13	5.67	0.224 ± 0.186	1200 – 2

Moreover, `MIRRBA` presented lower SDJDet values, hence produced smoother deformations than training-based approaches.

These results show that not learning registration patterns from a dataset helps to obtain precise segmentations, especially at a lesion level. Indeed, while the size of organs and their locations are relatively consistent across patients, this is not the case with the lesions. Hence, fewer deformation patterns can be learned from a database. Since performing locally precise registration with a DL-based method is very challenging in this situation, our patient-specific method manages to adapt to each subject.

Registration-specific regularization To study the influence of the registration-specific regularization terms $\mathcal{R}_{smooth}$ and $\mathcal{R}_{diffeo}$, we removed them individually and jointly from Eq. 4.3 respectively in `MIRRBA_wo_Smooth`, `MIRRBA_wo_Diffeo` and `MIRRBA_wo_Regu` (see Table 4.1 and Figure 4.2.C).

As quantitatively shown in Table 4.3, `MIRRBA` presented Dice scores for the organs and lesions which are respectively 6% and 4% higher compared to `MIRRBA_wo_Regu`. Yet, the detection and disappearing rates were increased for `MIRRBA_wo_Regu` over `MIRRBA`. Moreover, the SDJDet of `MIRRBA_wo_Regu` was higher than the ones of all other configurations.

Regarding `MIRRBA_wo_Smooth` and `MIRRBA_wo_Diffeo`, both configurations presented similar results as `MIRRBA_wo_Regu` for the organs Dice score. For the measures of the lesions, the results of `MIRRBA_wo_Smooth` were slightly better than `MIRRBA_wo_Diffeo` and comparable to `MIRRBA`, except for the disappearing rate which was higher for the latter. Qualitative results are visible in Figure 4.3.

For similar Dice scores and detection rate, `MIRRBA_wo_Regu` has higher SDJDet than `MIRRBA`, indicating $\mathcal{R}_{smooth}$ and $\mathcal{R}_{diffeo}$ help smooth the deformation field, even though they have less impact than the regularization of the architecture. The SDJDet lower scores of `MIRRBA_wo_Smooth` and `MIRRBA_wo_Diffeo` confirm that even one of the regularization term independently improves the smoothness of the deformation. Moreover, it seems that removing the smooth regularization term helps with the registration around lesions.

As presented in Section 2.2, removing the registration-specific regularization terms is equivalent to solve an ill-posed problem in conventional registration methods. In our case, since we use a network to perform registration, $\mathcal{R}_{archi}$ is still present when running `MIRRBA_wo_Regu`, which may explain why the registration method converged to a plausible solution. Letting lesions disappear (which does not correspond to regular nor smooth deformations) and obtaining a smooth deformation field could be considered as antagonist goals. Because of the convergence of our method even without registration-specific regularization terms, further studies focusing on lesion disappearance may be pursue.

Architecture regularization Finally, to understand the impact of the regularization power of the network architecture on the registration, we looked at the results of `MIRRBA_wo_Archi`, which directly optimizes the deformation field without including any specific architecture, hence not including $\mathcal{R}_{archi}$ (see Table 4.1 and Figure 4.2.D).

Qualitative results visible in Figure 4.3 show that the lack of architecture negatively impacts the registration smoothness, as well as its detection rate and Dice scores, making questionable the convergence of the method without architecture.

According to Table 4.3, `MIRRBA_wo_Archi` was outperformed by `MIRRBA` for both organ and lesion segmentations by 22% and 78% respectively. Although `MIRRBA_wo_Archi` disappearing rate was higher than `MIRRBA`'s, the method without architectural regularization presented a low detection rate confirming the structural bias of convolutional generators.

Hence, we confirm that the architecture has a regularization effect on the registration, which helps to find an admissible solution.

4.5.2 Architectural choices

To study the regularization effect of a network architecture on registration, we compared the results of various architectural choices built from the pyramidal network presented in Mok et al. [149] (see Table 4.1). Quantitative results are presented in Table 4.3, while qualitative ones are visible in Figures 4.4 and 4.5.

Table 4.3 – Ablation study on the regularization terms, architectural choices and inputs (Sections 4.5.1 & 4.5.2) – Dice scores of the organs and of the lesions, detection and disappearing rates of the lesions, SDJDet and approximate computational time. All pipelines were computed on GPU. Statistically significant improvement of our `MIRRBA` method over the others with $p < 0.05$ is indicated with *. Best results are marked in bold, except for SDJDet since no ideal value is defined.

	Dice organs ↑	Dice lesions ↑	Detec. rate (%) ↑	Disap. rate (%) ↑	SDJDet ↓	Time (min) ↓
`No reg.`	0.626 ± 0.138*	0.090 ± 0.115*	5.00	0.00*	0.000 ± 0.000	0
`MIRRBA`	0.918 ± 0.126	0.425 ± 0.207	33.04	9.36	0.124 ± 0.988	55
`MIRRBA_wo_Smooth`	0.832 ± 0.168*	0.420 ± 0.217	33.71	18.61	0.177 ± 0.554	55
`MIRRBA_wo_Diffeo`	0.838 ± 0.112*	0.418 ± 0.203	33.68	15.11	0.051 ± 0.117	55
`MIRRBA_wo_Regu`	0.868 ± 0.199*	0.407 ± 0.219*	33.70	16.84	6.655 ± 30.709	55
`MIRRBA_wo_Archi`	0.753 ± 0.144*	0.239 ± 0.223*	20.54	**59.95**	1.247 ± 0.475	20
`MIRRBA_Depth_1`	0.748 ± 0.169*	0.239 ± 0.211*	11.74	4.82	0.016 ± 0.047	30
`MIRRBA_Depth_2`	0.873 ± 0.135*	0.364 ± 0.211*	23.26	7.56	0.038 ± 0.174	45
`MIRRBA_Depth_4`	0.945 ± 0.012	0.466 ± 0.199	40.22	15.12	0.057 ± 0.058	60
`MIRRBA_Level_1`	0.722 ± 0.091*	0.221 ± 0.157*	4.35	0.00*	0.013 ± 0.010	**2**
`MIRRBA_Level_2`	0.869 ± 0.114*	0.371 ± 0.198*	19.78	0.00*	0.035 ± 0.173	10
`MIRRBA_Max`	0.922 ± 0.098	0.426 ± 0.213	33.26	8.34	0.316 ± 2.934	55
`MIRRBA_Up`	0.922 ± 0.112	0.439 ± 0.204*	33.04	8.84	0.056 ± 0.309	55
`MIRRBA_DefConv`	0.487 ± 0.463*	0.257 ± 0.275*	35.22	12.23	0.027 ± 0.020	130
`MIRRBA_NoiseImg`	0.892 ± 0.081*	0.381 ± 0.194*	22.44	0.00*	0.007 ± 0.009	55
`MIRRBA_FixImg`	0.941 ± 0.017	0.451 ± 0.197	36.03	20.52	0.071 ± 0.095	60
`MIRRBA_Best`	**0.947 ± 0.010**	**0.467 ± 0.202**	**40.94**	19.48	0.080 ± 0.085	60

Depth of the pyramidal network Modifying the number of coarse resolution levels, i.e. the depth of the pyramidal architecture, we studied the impact of global structural choices to optimize the whole network. First, we modified the architecture to optimize a simple U-Net-shaped network (`MIRRBA_Depth_1`), and pyramidal ones with two (`MIRRBA_Depth_2`), three (`MIRRBA`), and four (`MIRRBA_Depth_4`) resolutions. `MIRRBA_Depth_1` was optimized on full resolution images, while all other architectures were trained using a coarse-to-fine strategy (see Section 4.4.2).

Results show that increasing the network depth improves the Dice results, as well as the detection and disappearing rates. Indeed, `MIRRBA_Depth_4` presented the second highest Dice scores for both the organs and the lesions among all MIRRBA setups. This is confirmed by the visual results from Figure 4.4, which show that a deeper pyramidal architecture reaches qualitative better results. Indeed, with four resolutions, the receptive field of the coarsest level of `MIRRBA_Depth_4` captures the whole image (i.e. $200 \times 200 \times$

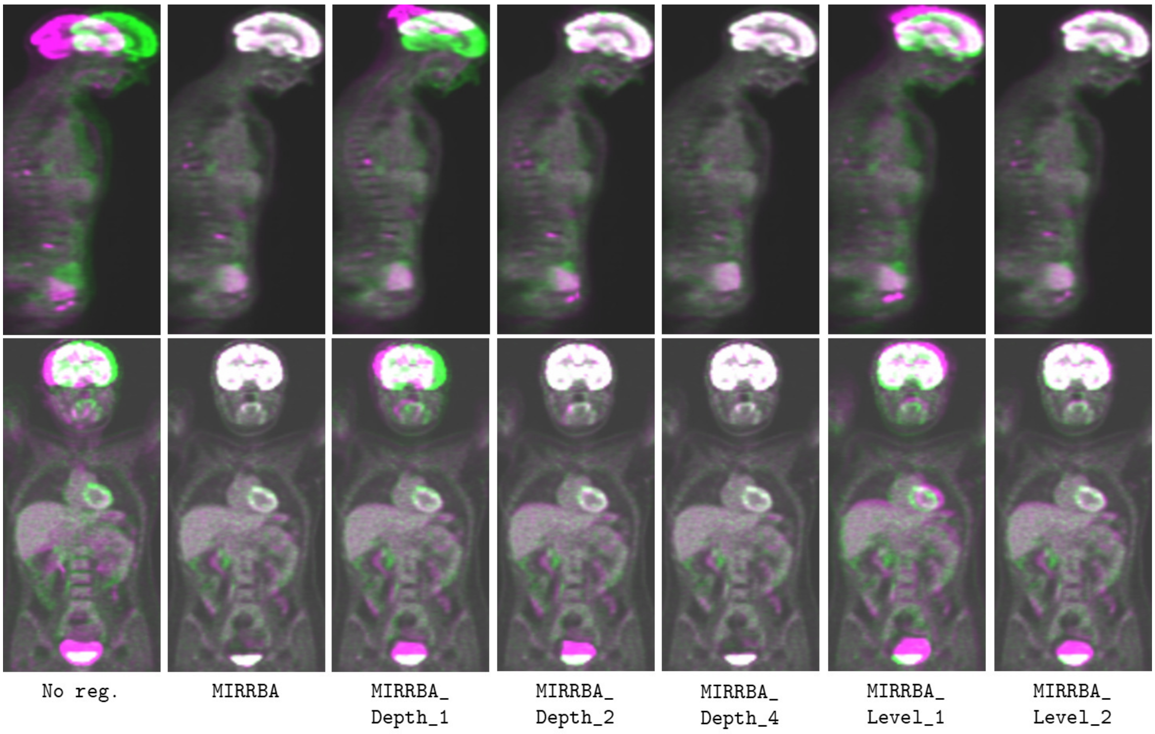


Figure 4.4 – Ablation study – Overlay of the fixed (green) and warped (pink) images on two different patients after performing `No reg.`, `MIRRBA` (corresponding to depth 3 and level 3), `MIRRBA_Depth_1`, `MIRRBA_Depth_2`, `MIRRBA_Depth_4`, `MIRRBA_Level_1` and `MIRRBA_Level_2`. Grayscale color indicates good overlapping. We can see that the higher the depth, the more precise the registration. A simple U-Net-shaped architecture as `MIRRBA_Depth_1` produces transformations of very low accuracy, while the more resolutions are used, the more precise the registration. We can also note that the coarsest level (`MIRRBA_Level_1`) performs global registration, while finest ones refine it (`MIRRBA_Level_2`) and achieve more precise registration (`MIRRBA`).

200), explaining the high Dice scores, as well as detection and disappearing rates. Indeed, successful conventional pipelines such as Elastix [112] or ANTs [7] also use this type of pyramidal strategy.

The higher SDJDet values obtained when more resolutions are used could be explained by the generation of more local transformations, hence a globally less regular deformation field, as illustrated in Figure 4.4 around the bladder.

Trained network level To understand the amount of information brought by each network level during the coarse-to-fine training strategy, we computed the registration after optimizing only the lower level (`MIRRBA_Level_1`) on coarse resolution images, both lower levels (`MIRRBA_Level_2`) on coarse and medium resolution images, and the whole network (`MIRRBA`) with the complete coarse-to-fine training strategy.

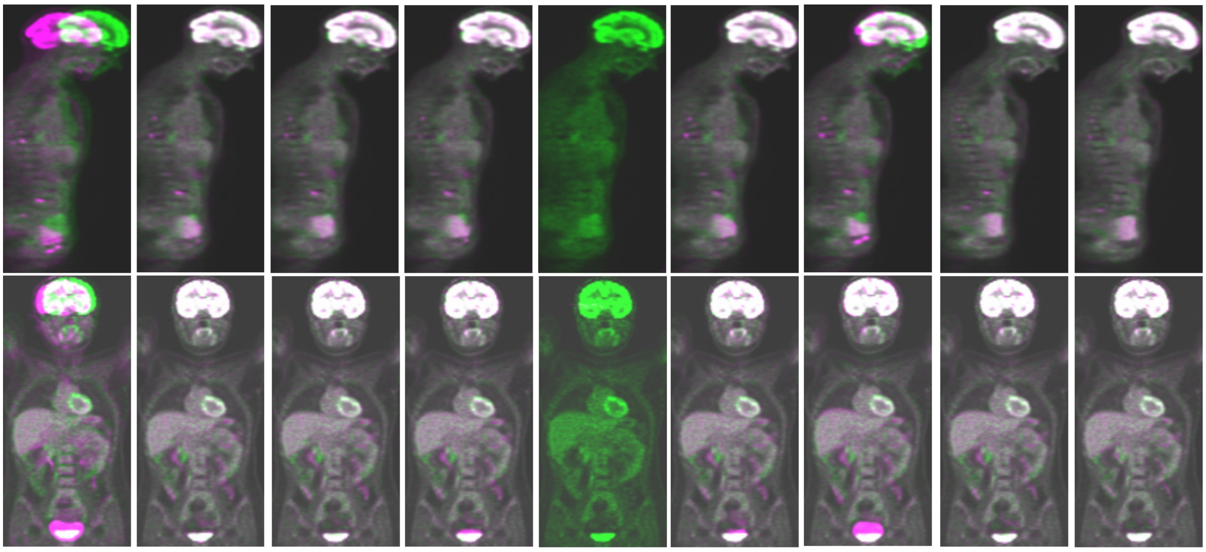
According to Table 4.3, training on all three levels of the network improved the registration accuracy over training only on low resolution images. Indeed, Dittmer et al. [41] affirms that an architecture ran in a DIP pipeline acts as a low-pass filter in the beginning of the optimization, allowing higher frequencies to pass only after lower ones. We observe similar results over the coarse-to-fine training strategy (see Figure 4.4), where low frequencies are registered first by the lower level, which performs a global registration. Higher frequencies are registered later by higher levels to refine the global registration with a local one. Moreover, the disappearing rate was very low when the optimization only occurred on low resolution images. In the same way, looking at the intermediate results during the 4-depth pyramidal optimization of `Elastix` (see in Figure 4.6), we observe a tendency to register global features before local ones.

Max-pooling and upsampling operations For each network level, we replaced the down-convolution (convolution with stride 2) by a max-pooling operation in `MIRRBA_Max`, and the transpose convolution by an upsampling in `MIRRBA_Up`. This reduced the amount of parameters to learn in the network.

As presented in Table 4.3 and in Figure 4.5, using these learning-free operations improved the registration accuracy both in terms of detection rate and Dice score. The disappearing rate was however slightly reduced. Indeed, even if the learnable down- and up-convolutions are now common to respectively increase or decrease image dimensions within a network, max-pooling and upsampling operations were originally used. As in conventional registration methods, these operations are not learnable. Hence, their results do not depend on optimization parameters, and they help to control the overfitting in a traditional trainable setup. Therefore, the good results obtained using max-pooling and upsampling operations in `MIRRBA_Max`, `MIRRBA_Up` and `MIRRBA_Depth_4_Max_Up` may be explained by the fewer number of parameters to fit (see Figure 4.6).

Residual blocks We removed the residual connections of the residual blocks (`MIRRBA_wo_RB`) to understand their influence in our architecture. Without them, we obtained null Dice scores and detection rate, as well as very high SDJDet value. Besides, warped images did not look realistic (see Figure 4.5).

Residual blocks can be related to diffeomorphic registration according to Rousseau



`No reg.` `MIRRBA` `MIRRBA_Max` `MIRRBA_Up` `MIRRBA_wo_RB` `MIRRBA_Def_Conv` `MIRRBA_NoiseImg` `MIRRBA_FixImg` `MIRRBA_Best`

Figure 4.5 – Ablation study – Overlay of the fixed (green) and warped (pink) images on two different patients after performing `No reg.`, `MIRRBA`, `MIRRBA_Max`, `MIRRBA_Up`, `MIRRBA_wo_RB`, `MIRRBA_Def_Conv`, `MIRRBA_NoiseImg`, `MIRRBA_FixImg` and `MIRRBA_Best`. Grayscale color indicates good overlapping. The all green `MIRRBA_wo_RB` image is due to the non-convergence of the registration algorithm and a warped image not registered to the fixed one. `MIRRBA_NoiseImg` does not reach local precise registration. All other approaches produce realistic looking and coherent transformations, even if `MIRRBA_DefConv` lacks a bit of precision around the bladder.

et al. [178]. Indeed, stacking residual blocks in ResNets [80] aims to incrementally map the embedding space to a new unknown space, each block being defined as $y = F(x) + x$, with x and y the respective input and output of the residual blocks, and F the residual mapping to be learned. Similarly, diffeomorphic registration models [15, 200] address the registration issue by piling up incremental diffeomorphic mappings. Making the link between ResNets and registration, the function F can be seen as a parametrization of an elementary deformation flow, and training a series of residual blocks as learning continuous and integral diffeomorphic operator. With our deep architecture, `MIRRBA_wo_RB` results indicate that the registration without residual blocks fails to converge. As explained above, residual blocks allow incremental diffeomorphic mappings, and removing them leads to gradient vanishing for our patient-specific method.

Deformable convolutions Regarding the convolutions, Heinrich [84] suggested deformable convolutions [36] to capture larger deformations. These convolutions add 2D or 3D offsets to the regular grid sampling of standard operations. If these offsets are set to zero, deformable convolutions become standard convolutions, otherwise they modify the receptive field. Since these offsets are learnable, deformable convolutions are trained to adapt their receptive fields, in order to focus on objects of interest in classification problems for instance.

Hence, to adapt the receptive field of the convolutions to the local scale of objects to be registered, we replaced those of the highest resolution level (i.e. level 3) by deformable convolutions (`MIRRBA_DefConv`).

According to the high detection rate compared to `MIRRBA`, `MIRRBA_DefConv` focused on main objects of interest. Yet, qualitative results in Figure 4.5 are not very precise, and the resulting low Dice scores show that `MIRRBA_DefConv` did not achieve precise local registration. This can be explained by the low SDJDet value, showing that only smooth and regular deformations occurred instead of locally irregular ones. Indeed, deformable convolutions might need a more complex integration in the architecture [127]. Moreover, the additional learnable parameters and adaptive receptive field of deformable convolutions are to some extent redundant with the pyramidal structure, and make it harder to train.

Input images Regarding the network inputs, instead of conditioning the network with the moving image, we fed it with a Gaussian noise, as in Laves et al. [119]. In Table 4.3, `MIRRBA_NoiseImg` shows worse results than `MIRRBA`, except for the SDJDet. This is confirmed by the visual result in Figure 4.5, where the bladder for instance is not correctly registered.

We also provided more information to the network by concatenating the fixed to the moving image in `MIRRBA_FixImg`, which significantly improved all `MIRRBA` results, especially the disappearing rate and reached visual accurate results (see Figure 4.5).

Contrary to Ulyanov et al. [211] and Laves et al. [119], Gong et al. [69] and Baguer et al. [9] respectively improved CT and PET DIP-based image reconstructions by providing acquisitions from other modalities to their networks. We made similar observations, as `MIRRBA_NoiseImg` reached less accurate results than when we conditioned the model with PET images, as in `MIRRBA`. Moreover, feeding the network with more patient information, i.e. using the fixed image as additional input to the moving image in `MIRRBA_FixImg`, improved the results by increasing the network conditioning on a single patient.

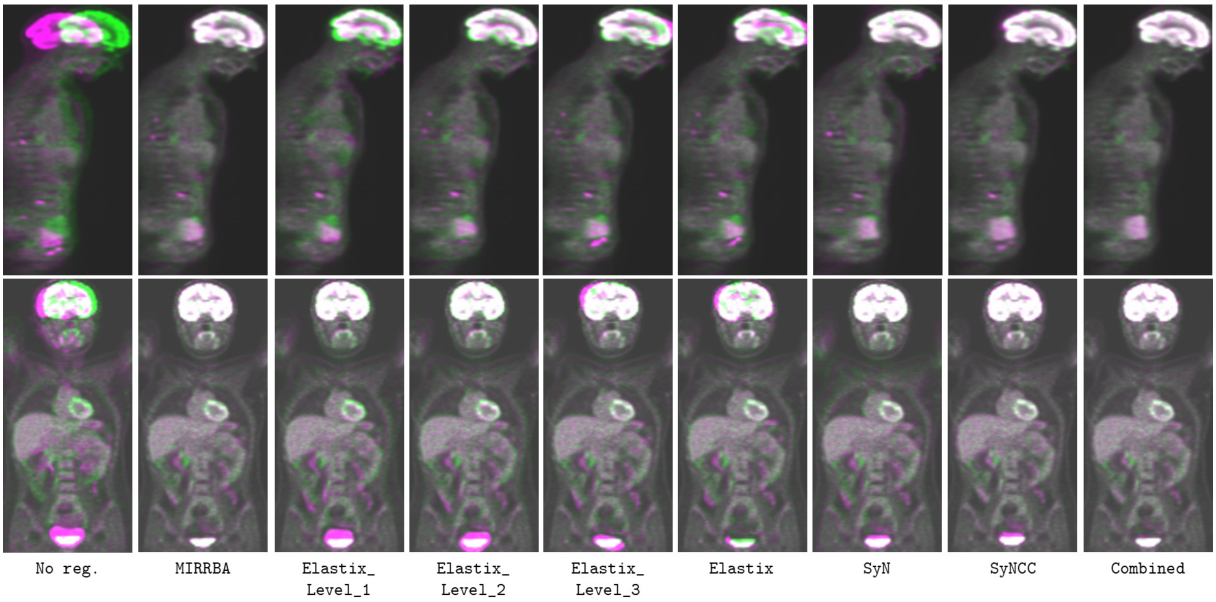


Figure 4.6 – Comparison to reference methods – Overlay of the fixed (green) and warped (pink) images on two different patients after performing `No reg.`, `MIRRBA`, `Elastix_Level_1`, `Elastix_Level_2`, `Elastix_Level_3`, `Elastix` (correponding to level 4), `SyN`, `SyNCC` and `Combined`. Grayscale color indicates good overlapping. The pyramidal optimization of `Elastix` acts as a progressive registration: global features are registered before more local and precise ones. `SyN` and `SyNCC` warped images look coherent, even if missing a bit of precision around the bladder, while `Combined` performs nice registration.

Combining best practices `MIRRBA` Dice scores were improved by the use of the fixed image as an additional input channel, as well as be max-pooling and upsampling operations. Moreover, adding a fourth depth to `MIRRBA` statistically improved its results. Hence, we combined the best architectural variations presented above to perform registration with four resolutions, both fixed and moving images as input channels, max-pooling and upsampling operations, as well as residual blocks, leading to the best performing MIRRBA-based method in terms of Dice scores. This configuration is called `MIRRBA_Best` hereafter. The SDJDet of `MIRRBA_Best` was smaller or similar to either one of the three other methods, while the disappearing rate lied between their values. Images of lesions registered with `MIRRBA_Best` are visible Figure 4.7, while whole-body images are visible in Figure 4.5.

Table 4.4 – Comparison of the reference `MIRRBA` configuration to conventional methods (Section 4.5.3) – Dice scores of the organs and of the lesions, detection and disappearing rates of the lesions, SDJDet and approximate computational time. While ANTs and Elastix pipeline were computed on CPU, `MIRRBA` ran on GPU. Statistically significant improvement of our `MIRRBA` method over the others with $p < 0.05$ is indicated with *. Best results are marked in bold, except for SDJDet since no ideal value is defined.

	Dice organs ↑	Dice lesions ↑	Detec. rate (%) ↑	Disap. rate (%) ↑	SDJDet ↓	Time (min) ↓
`No reg.`	0.626 ± 0.138*	0.090 ± 0.115*	5.00	0.00*	0.000 ± 0.000	0
`MIRRBA`	0.918 ± 0.126	0.425 ± 0.207	33.04	9.36	0.124 ± 0.988	55
`Elastix`	0.868 ± 0.124*	0.350 ± 0.191*	20.54	9.59	0.096 ± 0.044	25
`SyN`	0.936 ± 0.023	0.386 ± 0.210*	24.57	0.00	0.016 ± 0.018	**5**
`SyNNC`	0.944 ± 0.014	0.477 ± 0.211	39.57	4.26*	0.073 ± 0.066	60
`Combined`	**0.945 ± 0.012**	**0.481 ± 0.197**	**44.71**	**25.11**	0.077 ± 0.072	115

4.5.3 Comparison to conventional methods

Conventional registration According to Table 4.4, even after rigid and affine pre-registration, locally precise deformable registration is challenging on whole-body images for the conventional `Elastix` [112] pipeline (see Figure 4.6).

On the other hand, both `SyN` and `SyNCC` [7] statistically performed better than `MIRRBA` (our reference configuration, which used the NCC measure) on organ segmentation, while the NCC similarity measure allowed `SyNCC` to also reach a better accuracy on lesion segmentation. Regarding the disappearing rate, `MIRRBA` performed better than both SyN-based methods, whereas their SDJDet was lower than ours. Visually, `SyN` and `SyNCC` results are very similar (see Figure 4.6).

Combination of DIP and conventional registration Finally, we pushed the analysis by combining our best MIRRBA-based method with the best performing conventional `SyNCC` approach. To do so, we optimized `MIRRBA_Best`, using as input the deformation fields and already registered images obtained by `SyNCC`. Our assumption was that the conventional pre-registration would be improved by our method.

Results of this `Combined` pipeline are shown in Table 4.4 and in Figure 4.6. In terms of detection rate and Dice scores, the combined approach outperformed both `SyNCC` and `MIRRBA_Best`. In addition, the disappearing rate was significantly improved for both methods.

4.5.4 Lesion segmentation

Regarding lesion registration, Figure 4.7 shows different lesion evolution scenarios. We can see that the moving masks adapt to the fixed ones when the lesion does not disappear. Although the disappearing lesion from the first row was not deleted with `MIRRBA`, it was reduced by the registration algorithm and almost deleted with `MIRRBA_Best`. Indeed, lesion disappearance implies deformations which are not smooth nor diffeomorphic, hence the good disappearing rate of `MIRRBA_wo_Regu` and the probable need to adapt our registration strategy to the specific problem of lesion disappearance in future work.

Regarding other kind of lesion evolutions, i.e. diminution, stability or augmentation, it can be noticed that overall `MIRRBA` performs accurate registration and `MIRRBA_Best` seems slightly more precise.

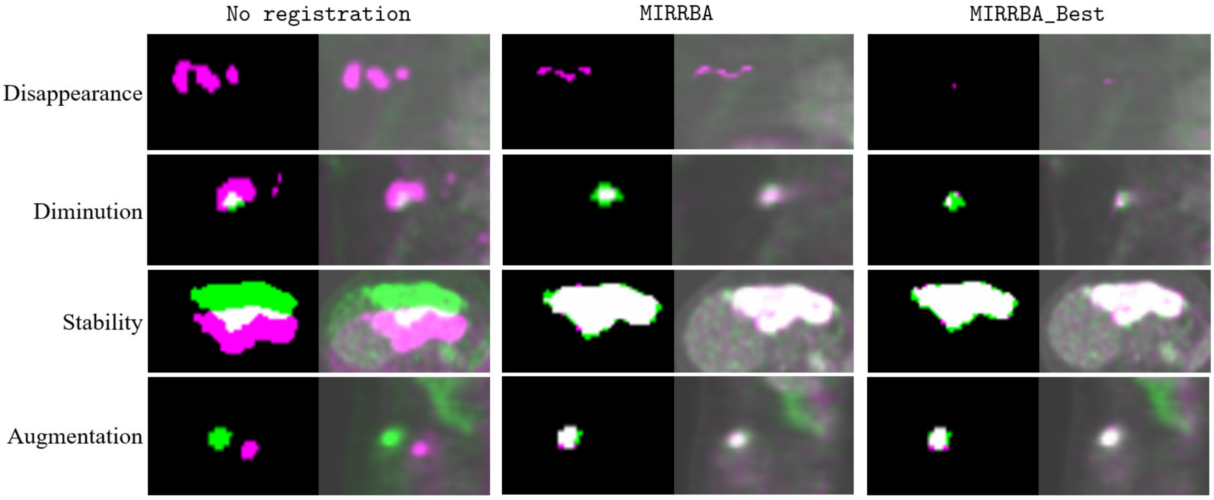


Figure 4.7 – Overlay of PET images and lesions (in columns) disappearing (1st row), reducing (2nd row), stable (3rd row) and growing (last row) without registration (`No registration`) and after performing `MIRRBA` and `MIRRBA_Best`. Fixed lesions can be viewed in green, while moving and warped ones in pink. White and gray-scale colors indicates overlapping.

4.5.5 BraTSReg dataset

To evaluate the generalization of our MIRRBA method, we tested it on the public BraTSReg dataset of pre- and post-operative brain MR images. Qualitative and quantitative results are respectively visible in Figure 4.8 and in Table 4.5. For state-of-the-art results, we will follow the results of the BraTSReg challenge in September 2022.

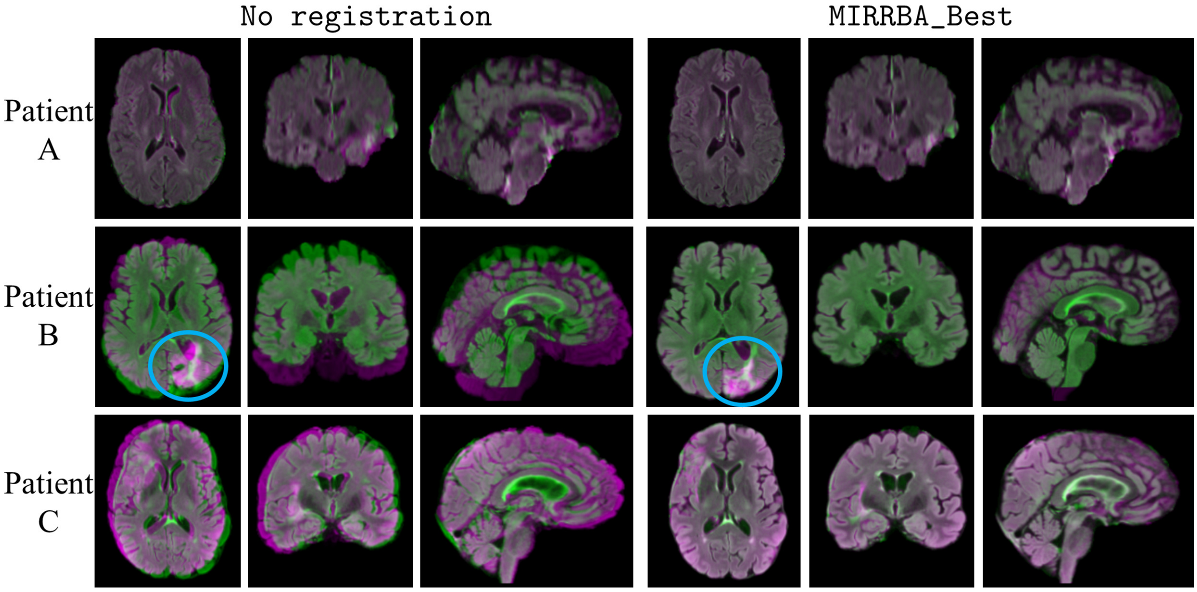


Figure 4.8 – No registration (left) and MIRRBA (right) registration results on three BraTSReg image pair (images A, B and C). The blue circle shows a resected area.

Table 4.5 – Mean Absolute Error (MAE) in mm, Normalized Cross Correlation (NCC) and standard deviation of the Jacobian determinant of the deformation field obtained after performing identity and MIRRBA registration on 140 flair MRI image pairs. Best MAE and NCC values are marked in *italic*.

Method	MAE (mm) ↓	NCC ↑	SDJDet
No registration	1.426 ± 0.962	0.058 ± 0.008	0.000 ± 0.000
MIRRBA	*1.347 ± 1.518*	*0.089 ± 0.011*	0.103 ± 0.070

Qualitative results show that `MIRRBA_Best` is able to correct the shift between images (see Figure 4.8). Yet, images do not seem perfectly registered in the resected area (blue circle in Figure 4.8). The MAE and NCC values reflect the visual analysis: `MIRBRA_Best` reduces the MAE and increases the NCC measure. The SDJDet value is small, meaning the produced deformation fields are relatively small.

Overall, MIRRBA improved the registration of MR FLAIR images. Yet, some areas with important modifications, such as the resected areas, are not very well registered. Since we did not make any adaptation to our MIRRBA method for this specific pre- and post-operative MR images intra-patient registration task, this is not surprising and could be improved in future work by learning the adapted hyperparameters for the task at hand, as in [148].

4.6 Conclusion

In this chapter, we propose an alternative method to perform image registration using a neural network without the typical learning stage on a database. We formalize the registration problem by following the conventional approaches relying on image-based similarities and regularization terms, but also explicitly consider the dataset and architecture bias. Indeed, our study is motivated by recent work on DIP, implying that neural networks create an inductive bias when learning from a database, but also create an intrinsic structural bias induced by the architecture [82]. These biases are shown sufficient to solve certain image processing tasks. Our proposition also resonates with Dittmer et al. [41], who suggested that deep convolutional neural networks process low-frequency information first, to later focus on the finer deformations, both desirable properties for a registration algorithm. We integrate the LapIRN network from Mok et al. [149], who developed an effective pyramidal architecture, tested in the DL-based registration set-up. Here, we further demonstrate that beyond any prior coming from the dataset learning step, the architecture design has an important effect on the registration results, acting as an implicit regularizer. Our study also shows the impact of some of the architecture components, particularly the residual blocks, and we justify this behavior by making a link with findings from Rousseau et al. [178]. Moreover, we find that for our application, a pyramidal architecture capturing the whole image with a limited amount of parameters to optimize, as conventional registration methods, provides precise registration results.

The architectural prior seems to be a better option than learning from data in cases where there are no positional consistency, which is the case with metastatic breast cancer lesions, which arbitrarily vary in position, size and number. Indeed, finding a set of network parameters allowing precise registration for a whole dataset is a challenging task. Our approach enables to correctly register active organs such as the brain and the bladder, which could be used to automatically propagate annotations masking regions irrelevant for patient response evaluation. Although the Dice scores are relatively low for the lesions, we obtain good detection values and improve the disappearance rate.

Our approach makes a step in bridging conventional and DL-based methods for image registration, and provides several suitable approaches for the challenging 3D full-body longitudinal registration problem. We demonstrated the possibility to perform both global and local registration on whole body medical images using a network but without suffering from dataset bias. In future work, we would study the feasibility of extracting registration-based feature from our method to monitor lesion evolution without depending on manually performed segmentations (see Chapter 6).

Moreover, it would be interesting to speed-up the overall process (currently lasting about an hour). Since the longer step is the optimization of the higher network level and, we want the registration to be especially precise around the lesions at this level, we could perform precise registration only on these very specific areas instead of performing deformable registration on the whole body.

Summary:

✓ MIRRBA is a registration method optimizing a CNN only on a pair of images in a Deep Image Prior setup.

✓ The choice of the network architecture impacts the registration results.

✓ To register heterogeneous objects, MIRRBA benefits from the flexibility of neural networks and from the absence of the training database bias.

✓ MIRRBA is generalizable, as shown by our preliminary results on pre- and post-operative brain MR images.

✓ MIRRBA links DL-based and conventional image registration methods, and can be used or combined with both.

Chapter 5

Improving PET image registration and label propagation using segmentations

---

Contents

5.1 Introduction

Our registration method presented in the previous chapter, MIRRBA, enables the registration of longitudinal data. Evaluated on both physiological high ^{18}F-FDG uptake organs (brain and bladder), as well as on metastatic breast cancer lesions, it reaches results comparable to conventional methods and outperforms DL-based ones.

To better monitor cancer, it is useful to accurately follow the individual evolution of each lesion. In addition, obtaining the individual segmentations of the physiological high ^{18}F-FDG uptake organs on longitudinal PET data enables i) to evaluate the registration accuracy more globally, and ii) to locate organs presenting similar uptake values as lesions. In this chapter, we focus on improving the registration accuracy around the cancerous lesions and high uptake organs, when performing longitudinal registration.

5.2 Related work

As introduced in Sections 2.3.5 and 3.1.1.2, it is possible to integrate segmentation information into a DL-based registration method to improve registration accuracy around segmentation masks. Several approaches integrate such segmentation masks into DL-based registration methods, either i) in the loss function alone, or ii) both in the architecture and in the loss.

In the loss function alone, weakly-supervised registration methods combine the classical registration similarity term with a label matching one (see Section 2.3.5), e.g. a Dice loss term [28]. Balakrishnan et al. [12] and Mok et al. [149], for instance, showed that adding segmentation information during the training step of their methods improved the overall registration performance of brain structures on MRI data.

Most joint segmentation–registration methods train two networks, respectively dedicated to the segmentation and registration tasks. Beljaards et al. [16], for instance, proposed "cross-stitch" units to linearly combine features maps from registration and segmentation branches using learnable parameters with good results. Reducing the need of a fully annotated dataset, Xu et al. [231] proposed to jointly perform weakly-supervised registration and semi-supervised segmentation. Unlabeled moving images were segmented by the segmentation network to get weak supervision for every training registration pair, while the registration similarity loss guided the segmentation learning on unlabeled data.

Note that none of the above methods [12, 16, 149, 231] fed segmentations directly into the network. Indeed, feeding the network with segmentations during training implies to also provide segmentations to the network during the inference. These segmentations can be obtained manually, but it is time consuming, subject to inter-operator variability and impractical in the particular case of multiple lesion tracking. One may also rely on automatic DL segmentation methods, for which there is an abundant literature in automatic medical image segmentation [145]. In the specific case of our EPICURE$_{\text{seinmeta}}$ dataset, we proposed in Moreau et al. [150] a method to automatically obtain segmentations of metastatic breast cancer lesions, feeding a U-Net-shaped network with the follow-up images, as well as with the image and lesion segmentations of the baseline acquisition. Even though the approach reached good precision, all lesions are segmented together, as a whole. The main drawback of such automatic segmentation method is its inadequacy to individually monitor lesions along time. Indeed, metastatic breast cancer may significantly evolve between acquisitions, making it challenging to match corresponding lesions for individual monitoring.

In the remainder of this chapter, we evaluate different approaches to improve longitudinal registration around objects of interest, i.e. metastatic lesions and active organs. We expand our MIRRBA method by integrating segmentation information both at the loss and architectural levels. While we use the conventional Dice score in the loss, we also directly feed the network architecture with segmentation maps as additional input channels.

To automatically obtain these delineations, lesions can be segmented by our approach presented in Moreau et al. [150] and published in the journal Cancers, while we propose an automatic segmentation method using superpixels [56] to delineate the high ^{18}F-FDG uptake organs (see Section 5.3). Then, we present the integration of the segmentations into the MIRRBA pipeline and the drawn conclusions respectively in Sections 5.4 and 5.5.

5.3 Automatic organ segmentation using superpixels

In clinical practice, metastatic breast cancer monitoring is performed using PERCIST guidelines through semi-automatic threshold-based measurements [70]. To avoid time-consuming interactions, automatic lesion segmentation algorithms on PET images have been developed [19, 87]. However, these approaches mostly focus on specific body parts, as the lungs or the brain, while metastatic breast cancer may spread over the entire body.

One of the challenges of automatic whole-body lesion segmentation on PET images is that areas of abnormal ^{18}F-FDG uptake can be obscured by active sites, presenting normal physiological uptake (brain, heart) and excretion (bladder) of the radioactive marker. Most methods proposing to localize and segment these organs use anatomical imaging techniques, such as MRI, CT [222, 130], or combine them with PET [18, 17].

However, methods based solely on PET images are less common: while PET imaging is recognized for metastatic activity characterization, it is considered lacking the spatial resolution required for precise organ segmentation. Hence, while segmenting directly on PET images is a challenge, it removes the dependence on error prone multi-modal PET–CT registration.

Unlike methods developed in Afshari et al. [3] and Vos et al. [222], we did not create bounding boxes around the target organs. Though more challenging, we seek a pixel-wise segmentation of the organ in order to preserve lesions close to the organs of interest for subsequent analysis. Moreover, our approach represents one of the first attempts to perform active organ segmentation using 3D CNNs.

One of the main limitation of developing DL-based methods to medical imaging problems is the lack of large datasets with expert annotations to train networks [126]. One alternative to by-pass this problem is to rely on larger but more heterogeneous datasets, collected e.g. from different sites and/or with different imaging systems [29, 76]. To cope with the resultant domain shift, and inspired by the work in Zhang et al. [239], we rely on a superpixel representation for segmenting active organs from whole-body PET volumes, in the context of metastatic breast cancer. Our hypothesis is that lesions often appear smaller than the normal active organs. Therefore, to segment active organs, it is enough to summarize the important anatomical information while blurring away other changes in image intensities coming from lesions or domain differences.

Our main contribution is the proposition of two approaches to segment active organs on 3D full-body PET images. Both approaches combine a classical deep segmentation network with a superpixel algorithm, but integrate it at different levels of the CNN: either as input to the network or within the loss function. To the best of our knowledge, this work is the first to perform automatic segmentation of high ^{18}F-FDG uptake on PET images and to take advantage of superpixel information to enforce networks to focus on the target objects.

5.3.1 Methods

Given a full body ^{18}F-FDG PET image, we aim to automatically determine which image voxels belong to the active organs (i.e. brain, heart and bladder). The selection of these sites is challenging due to potential surrounding tumors presenting a similar intensity, and to the variable tracer uptake amounts across the acquisitions and patients. In this section, we describe the baseline 3D U-Net method and propose two variants

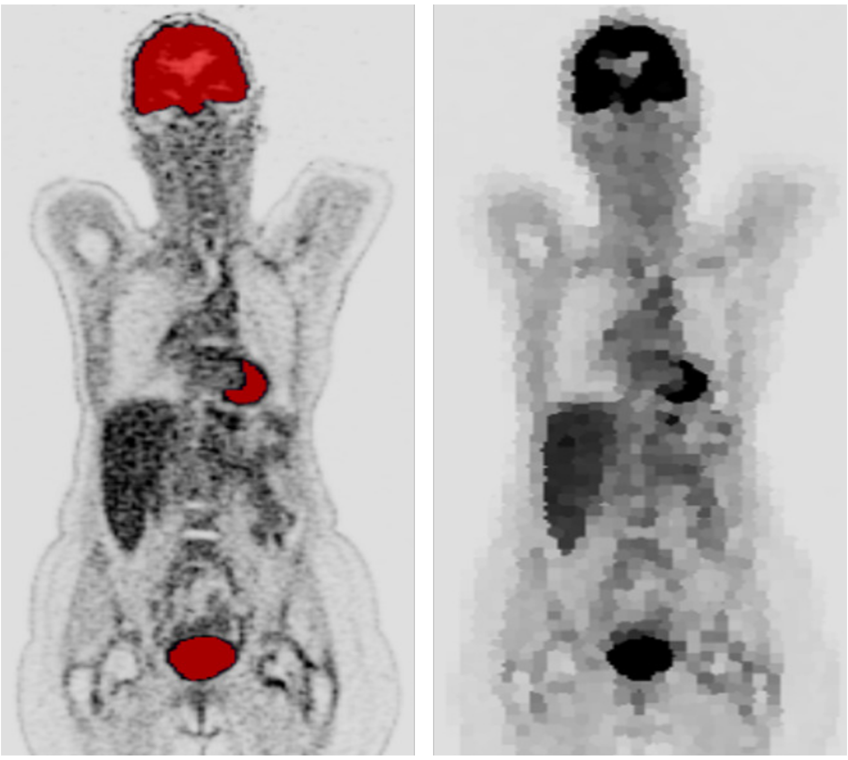


Figure 5.1 – Left: PET image and ground truth maps of the target organs (red). Right: Superpixelized PET image.

including superpixel information.

To segment the target organs, we rely on the 3D U-Net branch of the nnU-Net pipeline [97] as baseline. The training loss is a balanced combination of a cross-entropy $\mathcal{S}_{CE}$ and a Dice loss $\mathcal{S}_{Dice}$:

$$\mathcal{L}_{Total} = \mathcal{S}_{CE} + \mathcal{S}_{Dice}, \tag{5.1}$$

with $\mathcal{S}_{CE}$ as in Ronneberger et al. [177] and:

$$\mathcal{S}_{Dice} = -\frac{2}{|K|} \sum_{k \in K} \frac{\sum_{i \in I} u_i^k v_i^k}{\sum_{i \in I} u_i^k + \sum_{i \in I} v_i^k} \tag{5.2}$$

as in Isensee et al. [97], where u is the output map of the network, v the ground truth segmentation, $i \in I$ pixels in the training batch and $k \in K$ the classes. This loss reduces the influence of class imbalance of the labels [143].

To reduce the risk of confusing target active organs with other structures, especially tumors, we propose to incorporate superpixel information at different levels of the network: as input to the baseline model (`U-Net-SP-Input`) and in the loss function to train it

(`U-Net-SP-Loss`). An advantage of superpixel representations is they tend to smooth differences between images. Hence, using them in the training step of a CNN will increase the ability of the network to adapt from a dataset distribution to a similar yet different one (*domain adaptation*).

For the **first approach**, `U-Net-SP-Input`, we trained the reference network `U-Net` on superpixelized images (Figure 5.1, right). These superpixelized images were created using the reference Simple Linear Iterative Clustering (SLIC) algorithm [1] in 3D on the whole-body PET images. This method clusters voxels based on their gray level similarity and spatial proximity in the image, minimizing the following distance with a k-means algorithm:

$$D = d_{color} + \frac{m}{S} d_{image} \tag{5.3}$$

where D is the weighted sum of two pixel pairwise Euclidean distances, d_{color} for the color-space and d_{image} for the coordinate-space. S represents the approximate size of the superpixels, while m is introduced to control their compactness. Choosing the appropriate superpixel size and compactness parameters is important to emphasize the normal ^{18}F-FDG uptake on organs while blurring small lesions and objects.

In a **second approach**, `U-Net-SP-Loss`, we introduced superpixel information in the optimization of the reference network. The global loss is modified by balancing the cross-entropy loss with a superpixel-inspired regularization term:

$$\mathcal{L}_{Total_{SP}} = \mathcal{S}_{CE} + \mathcal{S}_{SP}, \tag{5.4}$$

where:

$$\mathcal{S}_{SP}(u, C) = \mathrm{MSE}(\arg\max(u), \Psi(\arg\max(u), C)), \tag{5.5}$$

with MSE the classical Mean Squared Error, u the softmax output of the network and C the correspondence map between pixels and superpixels. Ψ computes superpixel corrected segmentation maps, forcing all pixels within a superpixel to belong to the same class, as illustrated Figure 5.2.

The regularization term, inspired by Papadomanolaki et al. [167], enforces voxels belonging to the same superpixel to be associated with the same semantic object. In images, superpixels blur object smaller than their size, and highlight patterns of similar or bigger volumes. Since the organs to segment present homogeneous sizes in all images, we claim superpixels help networks to learn weights more generalizable across datasets.

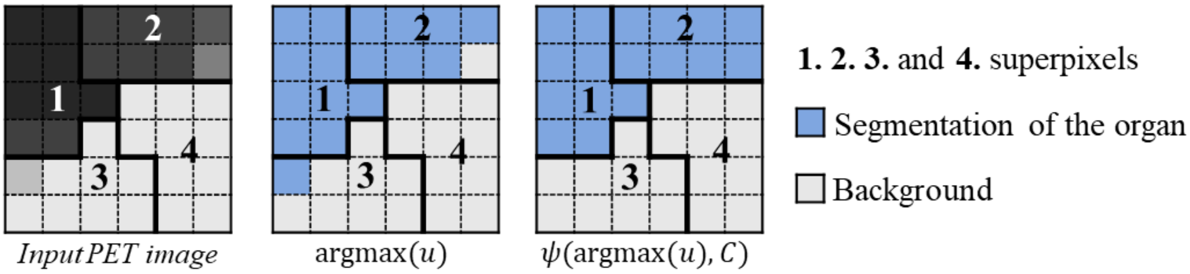


Figure 5.2 – From left to right: PET image; network softmax layer output map; and map corrected by the transformation Ψ. Superpixels **1.**, **2.**, **3.** and **4.** are also visible. On the transformed map, all pixels within a superpixel belong to the same class.

5.3.2 Experimental validation

5.3.2.1 Dataset description

The superpixel methods for active organ segmentation were developed at the beginning of this thesis and later connected to the MIRRBA method. In this context, we tested our methods on 60 images from the prospective EPICURE$_{\text{seinmeta}}$ study presented in Section 1.5, where 24 patients were recruited in the Angers center and 36 in the Nantes center.

All processed PET images where normalized by the SUV [109] (see Section 1.3.2).

5.3.2.2 Implementation details

The SLIC [1] algorithm was run in 3D, and the free parameters of the method, i.e. the superpixel size and the compactness, were set according to preliminary grid search results to $12 \times 12 \times 10$ mm and 5, respectively, to enhance target organs and to blur small lesions.

We followed the recommendations from Isensee et al. [97] regarding the preprocessing steps of the `U-Net` network, namely, cropping the input images to nonzero values, resampling them to the median voxel spacing of the dataset and applying a z-score normalization on each image. This helps to reduce the discrepancy between data domains. Considering the network itself, the architecture is similar to the benchmark U-Net [177] for 3D images, but uses leaky ReLU activation and instance normalization. All deep learning architectures were trained from scratch on a NVIDIA GeForce GTX 1080 Ti GPU. The input patch size (128×128×128) and the number of pooling operations (4 in each branch) were computed according to the preprocessed image size. The Adam optimizer was used for training. An epoch corresponded to an iteration over 250 training batches, each composed of two images. The learning rate was initialized at 3.10^{-4} and reduced by a factor 5 if the loss did not improve by at least 5.10^{-3} within the last 30 epochs. Once the learning rate reached 10^{-6}, the training stopped if the loss did not get better by at least 5.10^{-3} within the last 50 epochs. Random rotations, scaling, elastic deformations, gamma corrections and mirroring were performed on the fly for data augmentation purposes.

5.3.2.3 Evaluation metrics

The resulting segmentation maps of the three approaches were compared to ground truth masks containing the target organs (Figure 5.3, left), manually delineated with the Keosys Viewer[1] on PET images.

Since the outcome of our networks facilitates the posterior segmentation of lesions, it is essential that resulting segmentation maps do not contain any tumor located out of the target organs. Hence, the evaluation is two-fold.

Results were first assessed computing the Dice score between the segmented and reference masks of the normal uptaking sites, considering all three organs of interest together.

Moreover, the CNNs ability to segment target organs rather than other anatomical structures, i.e. to avoid false positive results, was assessed using the Positive Predictive Value (PPV). Indeed, this metric points out the ratio of True Positive (TP) pixels among all the positive predictions, characterized here by the sum of TP and False Positive (FP):

$$PPV = \frac{TP}{TP + FP} \tag{5.6}$$

5.3.2.4 Intra-domain experiment

To counteract the limited size of our dataset and maintain a reasonable computational time, all DL segmentation networks were initially validated on the images from the Nantes center using a 3-fold cross validation.

In that intra-domain experiment, the three CNNs present similarly high resulting Dice scores: 0.97 ± 0.01, 0.96 ± 0.01 and 0.96 ± 0.01 for the `U-Net`, `U-Net-SP-Input` and `U-Net-SP-Loss` approaches respectively. Hence, the three approaches accurately segment the target organs.

---

[1] https://www.keosys.com/

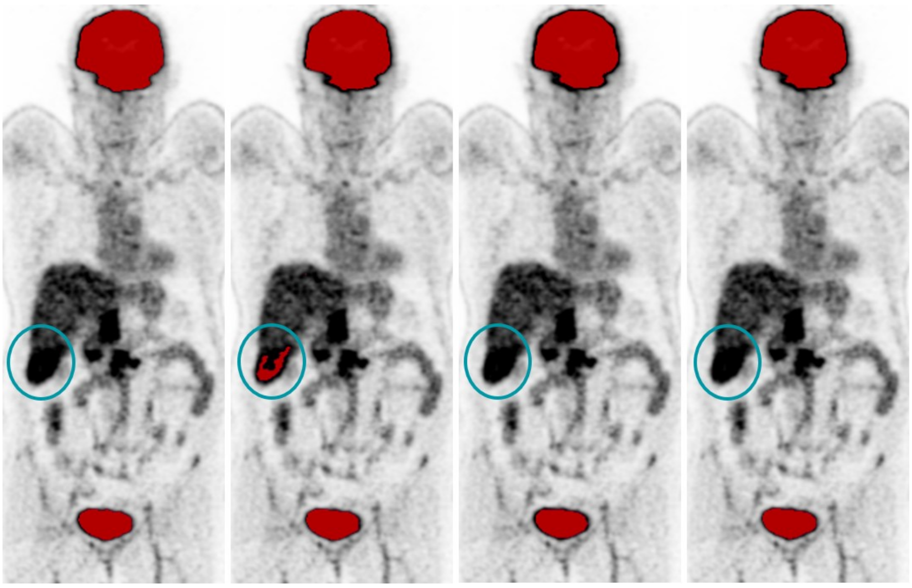


Figure 5.3 – PET image of a patient from the Angers center with overlaid masks corresponding to ground truth, `U-Net`, `U-Net-SP-Input` and `U-Net-SP-Loss` segmentations from left to right respectively. Blue circle marks a lesion erroneously segmented by the `U-Net` network.

5.3.2.5 Domain adaptation experiments

Over a second phase, the networks trained on the images from the Nantes center were tested on the images from Angers to evaluate their adaptability and generalization abilities. While slightly lower in the intra-domain case, the Dice scores obtained with the three approaches still remain high, with 0.92 ± 0.07, 0.92 ± 0.05 and 0.93 ± 0.05 for the `U-Net`, `U-Net-SP-Input` and `U-Net-SP-Loss` networks respectively.

Since we aim to remove the normal high ^{18}F-FDG uptake organs before segmenting lesions, we evaluated the methods ability to avoid the surrounding anatomical structures. Considering first a qualitative approach, the Figure 5.3 shows that networks using superpixel information are better suited to segment only the target organs. Quantitatively, we computed the PPV for the images from the Angers center. As illustrated in Figure 5.4, networks integrating superpixels information produce less FP than the reference `U-Net` network.

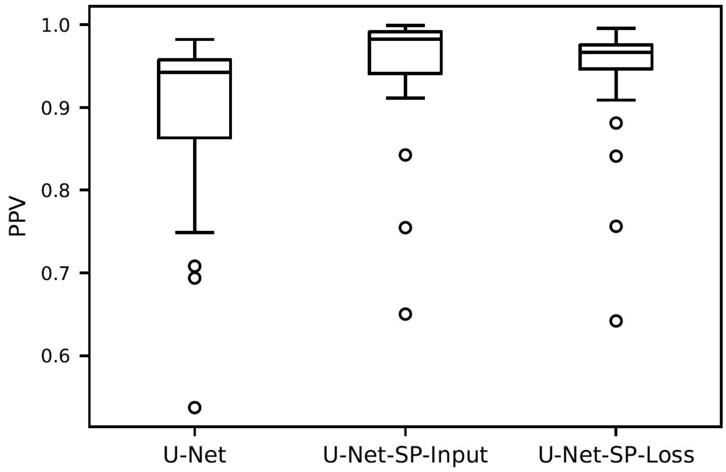


Figure 5.4 – Boxplot of the Positive Predictive Value (PPV) of the segmentations predicted by the `U-Net`, `U-Net-SP-Input` and `U-Net-SP-Loss` networks on the images from the Angers dataset.

5.3.3 Discussion

As shown by the obtained Dice scores, when the train and test datasets come from the same data distribution, the three approaches present similar results. In contrast, the superpixel regularization smooths variations between images acquired with different acquisition systems, improving segmentation performance on test data coming from a center unseen during training, showing the possibility of better generalization effect on unseen data. Moreover, according to the PPV measure, the integration of superpixel information in the training of a segmentation network reduces the amount of FP when using this network on unseen images.

From a computational point of view, the training time of the `U-Net-SP-Input` approach resulted to be faster ($\sim$ one day) than the one of `U-Net-SP-Loss` ($\sim$ two days). Indeed, superpixel computation is performed only once in the preprocessing step of the former network, while the latter requires superpixel computation for each epoch. However, the integration of the superpixel computation inside the training loop of `U-Net-SP-Loss` makes the process end-to-end, and potentially leaves room to integrate an automatic superpixel hyperparameters learning.

In the second part of this chapter, we integrate superpixel information at several levels of a deep learning approach to segment high ^{18}F-FDG uptake organs on PET images. The strength of this method is to enforce the network to focus on the organs of interest.

5.4 MIRRBA registration with segmentation masks

5.4.1 Segmentation integration

In the remainder of this chapter, we aim to improve the best performing MIRRBA configuration from Chapter 4. `MIRRBA_Best` corresponds to a 4-depth pyramidal network adapted from [149], optimized in a coarse-to-fine manner on a pair of fixed and moving images for a fixed number of iterations. Max-pooling and upsampling operations are used in this configuration. For more details, see Section 4.4.2.

To improve `MIRRBA_Best` registration performances around the objects of interest, we evaluate different integration strategies of the segmentations, as illustrated in Figure 5.5 and summarized in Table 5.1. These segmentations are obtained automatically, respectively with the method previously presented for the active organs [56], and with the method presented in [150] for the lesion segmentation.

For the integration of the segmentation maps in the loss function, we use an additional Dice score term $\mathcal{S}_{Dice}$ to evaluate the similarity between the fixed and warped segmentation maps (F_{seg}, W_{seg}):

$$\underset{\phi(\mathcal{R}_{archi})}{\arg\min}\, \mathcal{S}(F, W) + \mathcal{S}_{Dice}(F_{seg}, W_{seg}) + \lambda_{smooth}\mathcal{R}_{smooth}(\phi) + \lambda_{diffeo}\mathcal{R}_{diffeo}(\phi) \tag{5.7}$$

At the architecture level, we integrate segmentation maps as additional input channels to the network. These integration strategies can be used together of separately.

In the methods names, the integration of the Dice loss is specified with the suffix "`_Dice`", while the addition of input channels is noted with the suffixes "`_Fix`" and "`_Mov`", respectively for the fixed and moving segmentation maps.

Table 5.1 – Integration of the segmentations in the `MIRRBA` method to produce the `MIRRBA_Dice`, `MIRRBA_Dice_Fix`, `MIRRBA_Fix_Mov` and `MIRRBA_Dice_Fix_Mov` methods depending on the use of a Dice loss, as well as fixed and/ or moving segmentations as additional network channels.

	Dice loss	Fixed seg	Moving seg
`MIRRBA`			
`MIRRBA_Dice`	x		
`MIRRBA_Dice_Fix`	x	x	
`MIRRBA_Fix_Mov`		x	x
`MIRRBA_Dice_Fix_Mov`	x	x	x

As for MIRRBA, all architectures were implemented with PyTorch [168] and trained from scratch.

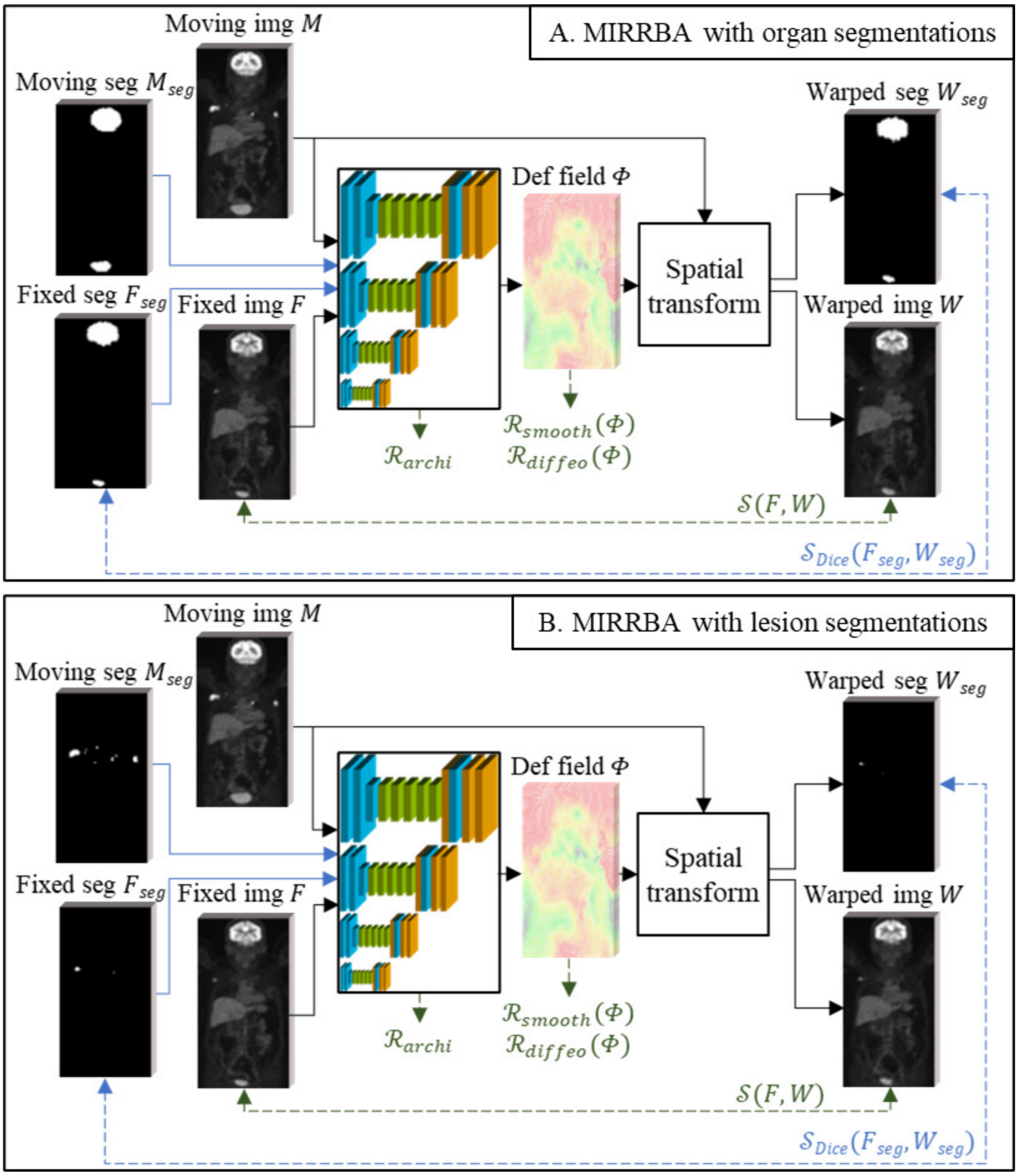


Figure 5.5 – Integration strategies of the organ (top image) and lesion (bottom image) segmentation maps in the MIRRBA method. The integration strategies are shown with the blue arrows. Segmentation maps can be integrated as additional input channels to the CNN. They can also be integrated through an additional loss function term in the form of a Dice score between the fixed and warped segmentations maps (F_{seg}, W_{seg}). These integration strategies can be used together or separately.

5.4.2 Experimental validation

5.4.2.1 Dataset description

To expand our MIRRBA method with segmentations, we used the $\text{EPICURE}_{\text{seinmeta}}$ dataset as in Chapter 4 and presented in details in Section 1.5. This dataset is composed of 110 pairs of PET images of patients presenting metastatic breast cancer. Brain, bladder and lesion segmentation maps were manually performed on all acquisitions.

All images were reshaped to size $200 \times 200 \times 200$ and their intensities were normalized by the SUV [109].

5.4.2.2 Evaluation metrics

We used the same evaluation metrics as the ones introduced in Section 4.4.4. While lesion registration is evaluated by the detection rate and the Dice score, the large deformations caused by their disappearance is quantified by the disappearing rate. Regarding the organ segmentations, they are also evaluated by the Dice score. The smoothness of the deformation field is measured by the standard deviation of its Jacobian determinant (SDJDet). Moreover, we measure the approximate running time of our method on a Nvidia V100 32GB SXM2 GPU.

5.4.3 Results and discussion

The quantitative results of the segmentation-based MIRRBA extension are presented in Table 5.2. Overall, Dice score values of the configurations including organ segmentations are very similar from one another. Indeed, the spatial shape and intensity distributions of the brain and the bladder make these organs easier to register than the more heterogeneous lesions.

The use of the Dice loss to integrate organ segmentation masks in the optimization process of the MIRRBA registration network does not improve the Dice scores nor the detection rate for neither the lesions nor the organs. Yet, the disappearing rate of the lesions increased, showing that this additional loss term may help register objects when large deformations occur.

Adding both the fixed and moving lesion segmentation maps as additional input channels to the network (`MIRRBA_Fix_Mov_Les` and `MIRRBA_Dice_Fix_Mov_Les`) leads to significantly better Dice scores for the lesions when compared to those obtained with

Table 5.2 – Dice score of the organs and of the lesions, detection and disappearing rates of the lesions and SDJDet. All pipelines were computed on GPU. Best results are marked in bold, except for SDJDet since no ideal value is defined.

	Dice organs ↑	Dice lesions ↑	Detec. rate (%) ↑	Disap. rate (%) ↑	SDJDet ↓
`No registration`	0.626 ± 0.138	0.090 ± 0.115	5.00	0.00	0.000 ± 0.000
`MIRRBA_Best`	0.947 ± 0.010	0.467 ± 0.202	40.94	19.48	0.080 ± 0.085
`MIRRBA_Dice_Les`	0.947 ± 0.010	0.466 ± 0.203	40.69	24.57	0.077 ± 0.069
`MIRRBA_Dice_Fix_Les`	0.947 ± 0.011	0.485 ± 0.204	44.35	24.26	0.081 ± 0.078
`MIRRBA_Dice_Fix_Mov_Les`	0.947 ± 0.010	0.489 ± 0.204	**47.12**	35.65	0.078 ± 0.070
`MIRRBA_Fix_Mov_Les`	**0.948 ± 0.010**	**0.490 ± 0.202**	46.37	**36.66**	0.090 ± 0.117
`MIRRBA_Dice_Org`	0.947 ± 0.010	0.466 ± 0.201	42.00	24.46	0.080 ± 0.070
`MIRRBA_Dice_Fix_Org`	**0.948 ± 0.010**	0.465 ± 0.207	40.94	25.03	0.084 ± 0.077
`MIRRBA_Dice_Fix_Mov_Org`	**0.948 ± 0.010**	0.468 ± 0.197	41.11	25.18	0.087 ± 0.084
`MIRRBA_Fix_Mov_Org`	**0.948 ± 0.010**	0.468 ± 0.203	42.95	26.06	0.096 ± 0.128

`MIRRBA_Best`. Indeed, providing the location of the lesions through their segmentations maps indicates the network the areas to mainly focus on and improves the registration. Moreover, both these methods improve the detection and the disappearing rate of the lesions compared to `MIRRBA_Best`, highlighting the help brought by segmentations as additional input channels.

Looking at the qualitative results of the segmentation-based extensions of the MIRRBA methods on whole body PET images, all images seem correctly registered, and no major differences can be noticed between the compared configurations (see Figure 5.6). For instance, all configurations manage to accurately register the deformed bladder. In conclusion, bringing any available segmentation information seems useful, regardless of the configuration (as additional input or in the Dice Loss).

Since we aim to accurately propagate lesion segmentations between acquisitions to help monitoring metastatic breast cancer, Figure 5.7 show some qualitative results of lesion registration using `MIRRBA_Best` and `MIRRBA_Fix_Mov_Les` (the best performing MIRRBA configuration including segmentation information, according to Dice scores and the disappearing rate Table 5.2). We can notice that the results of both MIRRBA-based methods are very similar: they make lesions disappear, diminish, translate or increase. However, `MIRRBA_Fix_Mov_Les` is slightly more accurate, especially to make lesion disappear (see 1st row in Figure 5.7).

Regarding computation times, all methods converged in about an hour. Hence, including the processing of segmentation information in MIRRBA does not significantly

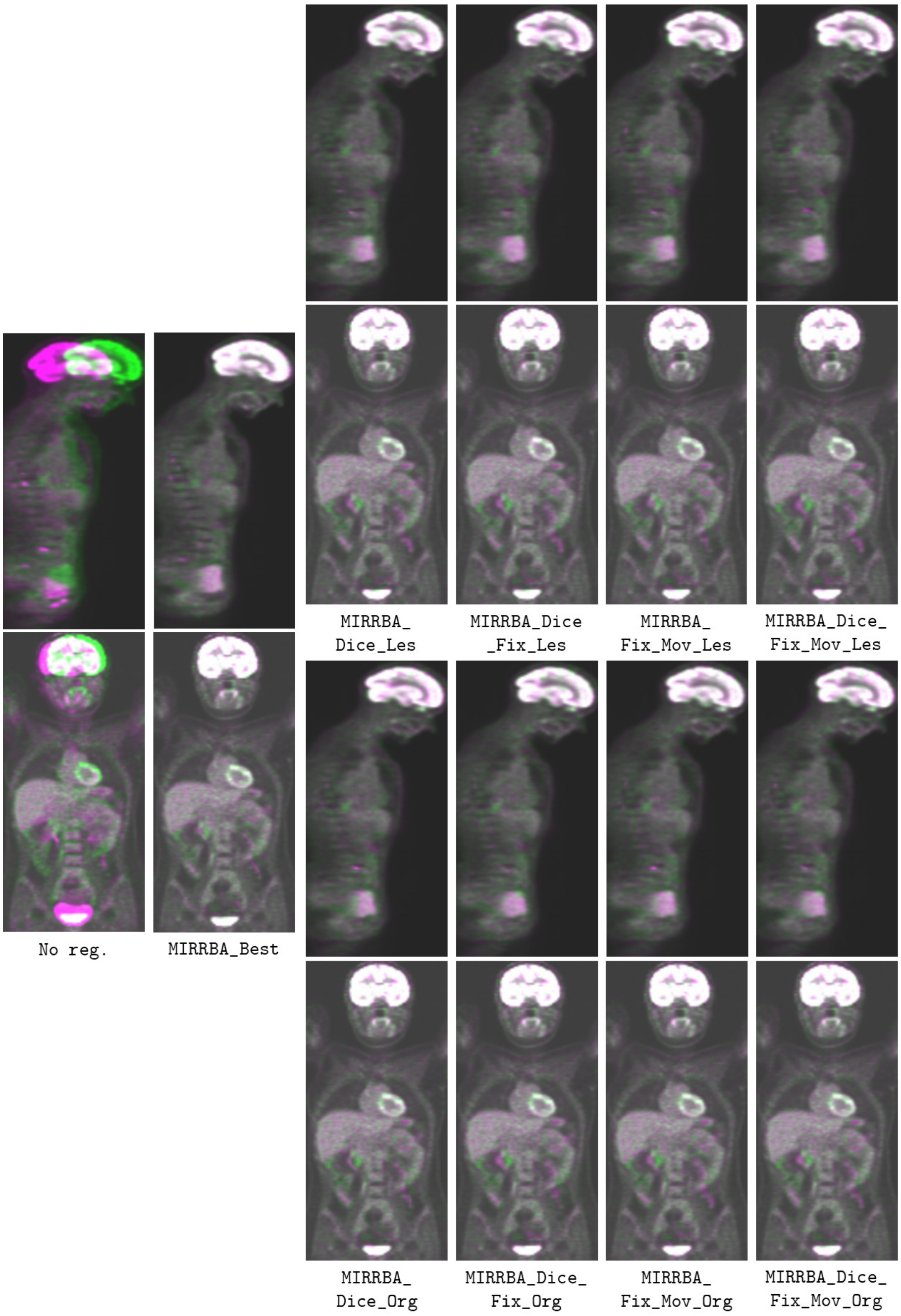


Figure 5.6 – Overlay of the fixed (green) and warped (pink) images on two different patients after performing `No reg.`, `MIRRBA`, `MIRRBA_Dice_Les`, `MIRRBA_Dice_Fix_Les`, `MIRRBA_Fix_Mov_Les`, `MIRRBA_Dice_Fix_Mov_Les`, `MIRRBA_Dice_Org`, `MIRRBA_Dice_Fix_Org`, `MIRRBA_Fix_Mov_Org` and `MIRRBA_Dice_Fix_Mov_Org`. Grayscale color indicates good overlapping. All the evaluated approaches including segmentation masks improve registration results. Qualitative differences are not perceptible to the human eye.

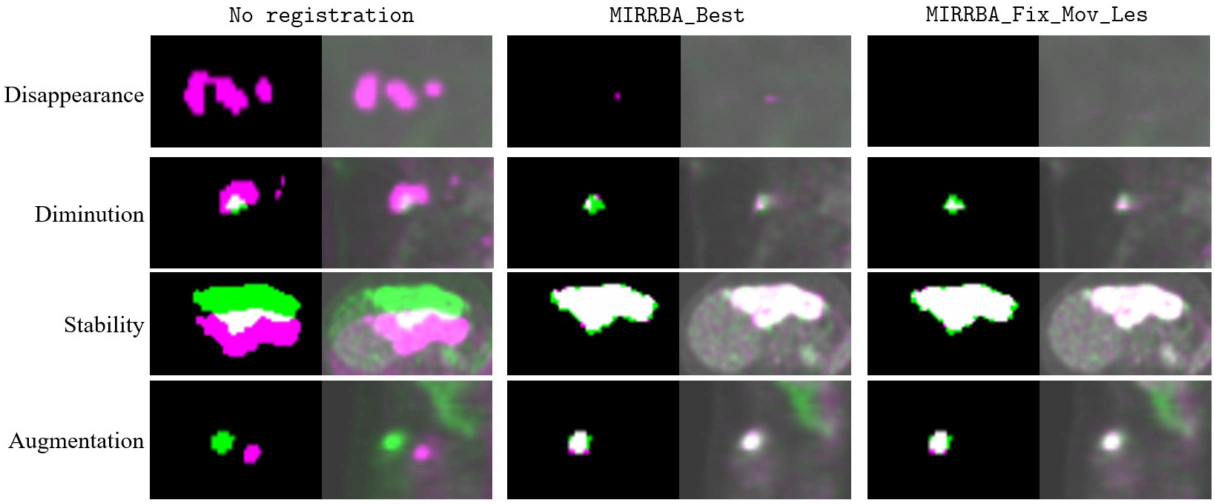


Figure 5.7 – Overlay of PET images and lesions (in columns) disappearing (1st row), reducing (2nd row), stable (3rd row) and growing (last row) without registration (`No registration`) and after performing `MIRRBA_Best` and `MIRRBA_Fix_Mov_Les`. Fixed lesions can be viewed in green, while moving and warped ones in pink. White and gray-scale colors indicates overlapping.

increases the overall computation time.

5.5 Conclusion

In this chapter, we first develop an automatic method to segment active organs (i.e. brain, heart and bladder) on whole-body PET images. Combining superpixels with a DL-based segmentation approach, both in the loss or as input of the network, we manage to accurately segment active organs on PET images, as well as to train networks generalizable across datasets. In a future work, it may be interesting to evaluate the performances of both superpixel-based approaches on a more significant dataset. Overall, our segmentation method is an important step towards bringing anatomical information and avoiding the confusion between high ^{18}F-FDG uptake organs and lesions, especially when they are contiguous.

In this thesis, we propose to take advantage of these automatically obtained organ segmentations as well as the lesion segmentations computed with our segmentation method [150] to improve the MIRRBA registration method. We found that including both the moving and the fixed segmentations to the registration pipeline effectively improves its performances, especially for the challenging task of the propagation of lesion segmentation. In future work, it would be interesting to develop a trainable end-to-end and simultaneous segmentation and registration approach to better take advantage of both methods.

In the last chapter of this thesis, we take advantage of the propagated segmentations masks to extract imaging biomarkers in order to help the monitoring of metastatic breast cancer in the clinical practice.

Summary:

✓ Combining superpixel and DL helps with the generalization ability of DL-based segmentation methods, in the context of TEP imaging.

✓ Prior information brought in the form binary segmentation masks improve the results of registration results. Such masks can be obtained with manual or automatic approaches.

Chapter 6

Metastatic breast cancer monitoring through registration

Contents

6.1 Introduction

In metastatic breast cancer, treatment response is often assessed using several information, such as markers from blood analysis, the patient general state, as well as the

PERCIST guidelines on ^{18}F-FDG PET images. These guidelines focus on appearing lesions and on changes in the $\mathrm{SUL}_{\mathrm{peak}}$ of the single hottest tumor identified on the baseline and follow-up images (which may be different from an acquisition to another). In this sense, PERCIST does not consider the response of the multiplicity of tumors that characterize metastatic breast cancer (see Section 1.4.3.2 for more details).

In this context, we question the fact that PERCIST represents the whole body response to treatment, and we propose a method to follow each lesion individually to monitor metastatic breast cancer locally. Moreover, ICO physicians hypothesized that acquisitions obtained early in the metastatic breast cancer course of treatment carry relevant information about patients' prognostics.

In this chapter, we automatically extract three biomarkers, i.e. the $\mathrm{SUL}_{\mathrm{peak}}$, the Metabolic Tumor Volume (MTV) and the Total Lesion Glycosis (TLG) from lesion segmentations obtained with registration, and evaluate their significance compared to biomarkers extracted from ground truth (manual) segmentations. From these biomarkers, we propose a protocol to evaluate tumor response on early- and mid-treatment acquisitions in the case where patients present multiple metastatic lesions, and compare it against the response determined by physicians in the clinical routine. Our method provides a new tool for the monitoring of metastatic breast cancer using only PET images, and was accepted for an on-line publication to the 2022 ASCO conference [55]. The remaining of this chapter is divided in two parts, one dedicated to the biomarker extraction (see Section 6.3.1), and the second to the evaluation response (see Section 6.3.2).

6.2 Related work

The response criteria obtained with PERCIST guidelines showed to be a good predictor for early therapeutic response to chemotherapy and prognosis of several types of cancer, such as esophageal cancer [233], non-small-cell lung cancer [189], metastatic prostate cancer [73], bone metastases [34].

Despite its good predictions of early treatment response, several authors have proposed some updates to the PERCIST system. As presented in Section 1.4.3.2, original PERCIST criteria classifies the response to a treatment within four categories: "complete response", "partial response", "stable disease" or "progression", according to the change of the $\mathrm{SUL}_{\mathrm{peak}}$ values between baseline and follow-up acquisitions. Maffione et al. [135] proposed to modify PERCIST guidelines into the PET Residual Disease in Solid Tumor

(PREDIST) criteria, for patients affected by rectal cancer. PREDIST considers two possible classes: the Complete Response to therapy (CRt) and the Residual Disease (RD), looking only at follow-up images. To this end, PREDIST compares PET FDG uptake on follow-up images to the SUV of the liver (i.e. to $1.5 \times \text{SUV}_{\text{mean\_liver}} + 2$ SD with $\text{SUV}_{\text{mean\_liver}}$ the mean SUV in the liver and SD its standard deviation in 3 cm of diameter spherical ROI). Maffione et al. showed that, contrary to PERCIST, PREDIST classification is statistically correlated to the Tumor Regression Grade (TRG) score, itself reflecting tumor response following a histopathological analysis [138].

On the other hand, Viglianti et al. [217] discussed the fact that brain and cardiac muscle store some of the injected FDG, which affects the distribution of the FDG marker in the organs and tissues of interest, i.e. tumors in our case. This effect is known and taken into consideration in PERCIST guidelines by measuring the uptake in a reference tissue (usually the liver or the aorta). If the difference between the reference uptakes of the studied acquisitions is not significant, PERCIST guidelines are considered valid, indicating the comparison between lesion uptakes is reliable. Yet, this condition is not always met. To diminish the FDG storage effect, and thus increase the amount of cases on which PERCIST guidelines are valid, Viglianti et al. [217] proposed to normalize the FDG uptake values of PET images by the uptake of the reference tissue, with promising results.

Focusing on metastatic breast cancer, Vogsen et al. [220] studied the feasibility and benefits of using the PERCIST guidelines on a retrospective study. In daily clinical practice, they found moderate agreement between PERCIST and visual assessment. Indeed, PERCIST is more sensitive than visual assessment to the detection of disease progression. This is important because an early identified progression may lead to the modification of a toxic and ineffective treatment. Yet, an early definition of a progressive state presents the risk of discarding a treatment too soon. Authors suggested that future studies should aim at comparing one-lesion against five-lesions PERCIST, to find a more optimal approach for the very widespread metastatic breast cancer disease.

Inspired by these updates, we propose here a computer aided automatic approach that considers for the first time all metastatic lesions, extracts biomarkers from them, and evaluates the response from ^{18}F-FDG PET acquisitions.

6.3 Method

On the EPICURE$_{\text{seinmeta}}$ dataset, baseline and follow-up PET images were acquired, and all baseline lesions were manually segmented by experts. In the approach presented in this chapter, we propagate these segmentation maps to the follow-up images, and extract biomarkers from each individual lesion segmentation to evaluate lesion-wise response to treatment. Finally we represent these responses through a visual tool. On overview of the proposed approach is visible Figure 6.1.

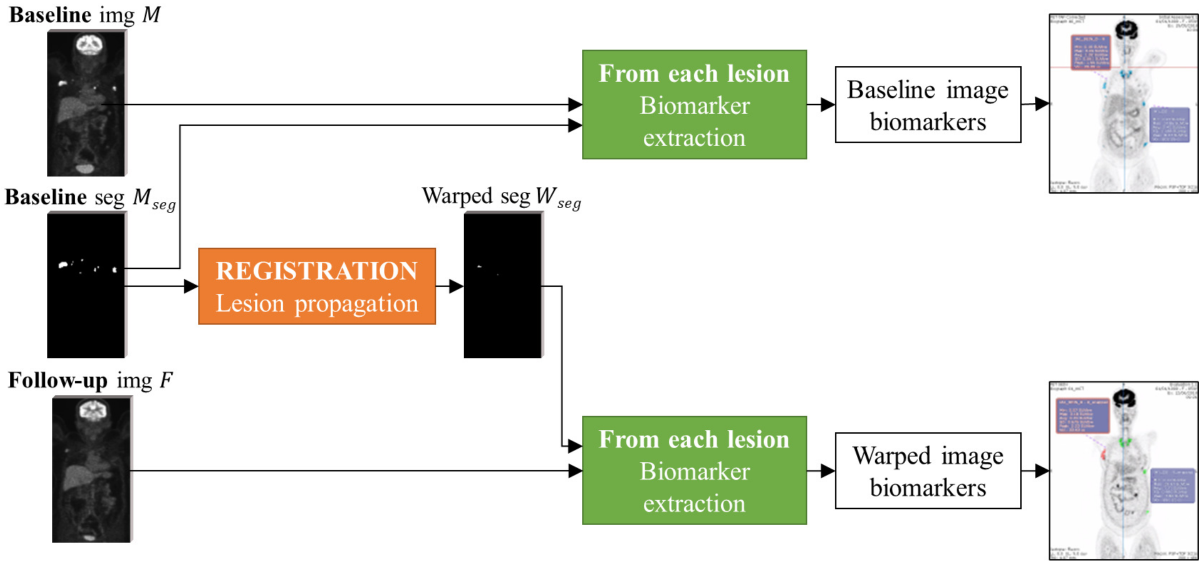


Figure 6.1 – Pipeline to obtain images representing the response to treatment of each lesion. Lesion segmentations from baseline images are propagated to follow-up images. Then biomarkers are extracted from each lesion on baseline and follow-up images. These biomarkers are finally compared to obtain the lesion-wise response to treatment and represent it visually.

6.3.1 Extraction of biomarkers

To individually monitor metastatic breast cancer lesions, we extracted three biomarkers from each lesion segmentation: the SUL$_{\text{peak}}$, the MTV and the TLG. We computed the SUL$_{\text{peak}}$ as in Section 1.4.3.2, i.e. seeking the 1 mm^3 sphere presenting the highest mean SUL, and extracting the maximum SUL value from this sphere. The MTV is evaluated as the number of voxels in a lesion mask multiplied by the voxel size. The TLG is a combination of the SUL$_{\text{peak}}$ and MTV biomarkers and represent both of them together:

$\text{TLG} = \text{SUL}_{\text{peak}} \times \text{MTV}$.

The biomarkers were extracted from lesions present in both baseline and warped segmentations maps (see Figure 6.1). These warped segmentations correspond to lesion segmentations propagated from baseline to follow-up images with four registration methods, i.e. `Affine`, `SyN`, `MIRRBA_Best` and `MIRRBA_Seg`:

- `Affine` is a global registration method performed in about a minute. We used a conventional registration implementation based on the SimpleElastix pipeline [110], as detailed in Chapter 3 with the registration parameters described in Annex B.2.
- `SyN` is a conventional registration method based on the ANTs pipeline using the MI similarity measure with a symmetry transformation penalty [7], as described in Section 4.4.3. In addition to good registration results, `SyN` performs deformable registration between two images in about 10 min, which is interesting for a potential integration of the tool in clinical routine.
- `MIRRBA_Best` refers here to the best configuration of our registration method, as described in our paper [54]. `MIRRBA_Best` performs registration overfitting deliberately a network on the pair of images to register. This network is pyramidal with four depths, considers fixed and moving image as inputs, and uses max-pooling and up-sampling operations. At the moment, the computation requires about an hour, which would need a pre-computation during clinical routine. Since PET is not a real time imaging system, this would not be a significant issue.
- `MIRRBA_Seg` expands `MIRRBA_Best` configuration integrating both fixed and moving segmentation maps automatically obtained as additional network channels, as described in Chapter 5 with the configuration `MIRRBA_Fix_Mov_Les`. The computation also requires about an hour.

As ground truth segmentation maps, we considered the lesion segmentations manually performed by expert physicians on follow-up images.

6.3.2 Response evaluation

In this section, we propose a new automatic tool to help physicians monitor metastatic breast cancer. From the SUL_{peak} values previously extracted, we classify each follow-up lesion as “Responsive” or “Non-responsive”, and colored them accordingly. Through this tool, we aim to study the clinical outcome information carried by early-treatment acquisitions.

Moreover, for each patient, we had access to the clinical outcome determined during the clinical routine by oncologists. This clinical outcome is defined according to i) tumoral markers visible through blood analysis, ii) the patient general state (eventual pain or masses during palpation), iii) RECIST guidelines, and iv) PERCIST guidelines applied by specialized physicians on the pre- and mid-treatment acquisitions. A unique feature of this study is the early-treatment images, which are usually not acquired nor used in clinical practice.

Once the SUL_{peak} is extracted from every baseline and follow-up after lesion segmentation, we used its values to evaluate the response to treatment of every individual lesion. Thus, we computed the differences between baseline and follow-up SUL_{peak} values. As presented in [55], our method also includes a visualization step, to facilitate the identification of responsive and progressive lesions. An example of the colored segmentation provided to physicians is visible Figure 6.2. On the baseline image, lesion segmentations are colored in blue. In follow-up images, lesions with a decreasing SUL_{peak} compared to the baseline are represented in green, while lesion with an increasing SUL_{peak} are visible in red. Since the representation is interactive, it is possible to move inside the patient to analyze different body parts.

6.4 Experimental validation

6.4.1 Dataset

We extracted biomarkers from the same 110 pairs of images on which we ran our MIRRBA registration method, as already introduced in Section 1.5 and presented in Chapter 4. These image pairs are composed of a baseline and a follow-up image, and were acquired in two different centers (Angers and Nantes). On these pairs of images, we registered the baseline image to the follow-up one.

Then, from the 110 pairs of images used to evaluate the pertinence of extracted biomarkers, we randomly choose 20 pairs of images, corresponding to 10 patients from the Nantes center. For each patient, two pairs of images are available: a pre- and early-treatment image pair, as well as a pre- and mid-treatment image pair.

6.4.2 Experiments

In this chapter, our first experiment consists in extracting biomarkers carrying information about the tumoral response to treatment from automatically propagated lesion segmentation maps.

Since there is no straightforward metric to evaluate the pertinence of the extracted biomarkers, we opted for a correlation measure with biomarkers extracted from manually performed segmentation maps, considered as a reference. First, to evaluate the normality of each biomarker distribution for all lesions, we used the statistical Shapiro-Wilk test [147]. Because distributions are not normally distributed, we measured the correlation between the ground truth and the registration-based biomarkers with the Lin correlation coefficient [125].

In our second experiment, we asked two expert physicians to blindly evaluate treatment response while seeing only these colored segmentations (and without any additional information about the patients). They evaluated the patient states as in “Responsive”, “Stable” or in “Progressive”.

Both expert physicians assessed the ^{18}F-FDG PET image, evaluating for each patient the baseline and one of the corresponding follow-ups, without knowing if they were working on the early- or on the mid-treatment follow-up. We then i) first compared early- to mid-treatment response, to study the information (similar or not) carried by both acquisitions; and ii) we then compared every follow-up response to the patient’s clinical outcome determined in clinical routine. Finally, we also compared both physicians answer to study their coherence.

6.5 Results and discussion

6.5.1 Biomarkers and registration method

Table 6.1 shows the Lin correlation coefficients for the studied methods (i.e. `Affine`, `SyN`, `MIRRBA_Best` and `MIRRBA_Seg`) and biomarkers (i.e. SUL_{peak}, MTV and TLG). Overall, biomarkers extracted from the propagated segmentations were similar to the ones extracted from the manual segmentations, with high Lin correlation coefficients.

Comparing biomarkers, the SUL_{peak} shows the highest correlation coefficient values and MTV the lowest correlation values among all registration methods. Since TLG is a combination of the SUL_{peak} and the MTV, its correlation coefficients are in between, as expected. Given $\mathrm{SUL}_{\mathrm{peak}}$'s higher correlation (see Table 6.1) and the familiarity of doctors with it, we retain this biomarker for the remainder of this chapter.

Table 6.1 – Lin correlation coefficients between the biomarkers obtained from the ground truth manual segmentations and the registration-based propagated segmentations for non appearing/disappearing lesions. Best result is highlighted in bold.

	`Affine`	`SyN`	`MIRRBA_Best`	`MIRRBA_Seg`
$\mathrm{SUL}_{\mathrm{peak}}$	0.890	0.892	0.979	**0.986**
MTV	0.770	0.787	0.864	0.867
TLG	0.824	0.869	0.921	0.922

Focusing now on the registration methods, all methods reached quite high results, even the global registration method `Affine`. Yet, both MIRRBA-based registration methods obtained higher Lin correlation coefficients than `Affine` and `SyN` for all biomarkers. Comparing MIRRBA-based methods, `MIRRBA_Seg` slightly outperforms `MIRRBA_Best`. This was expected, since the use of segmentations within `MIRRBA_Seg` registration pipeline helped the network to focus on areas of segmentation maps. Yet, the computational time of `SyN` is about 10 min, while `MIRRBA_Best` and `MIRRBA_Seg` require about an hour of computation. Having in mind this time and a potentially simpler integration within an industrial software, we worked with the `SyN` registration method in the remainder of this chapter.

We conclude that with segmentations derived automatically from registration algorithms, we can extract biomarkers well correlated to the ones obtained with manual segmentations, which suggests that both segmentations carry similar information. `SyN` fast registration and $\mathrm{SUL}_{\mathrm{peak}}$ computation can make it a useful tool in clinical routine, as shown hereafter.

6.5.2 Early- and mid-treatment acquisitions evaluation

For each patient, Table 6.2 shows each physician's evaluation for every follow-up image, as well as the clinical outcome. With the use of colored segmentations, early follow-up evaluations were predictive of mid-treatment response in 65% of the cases. This is especially the case if early follow-up were evaluated as responsive. In that case, the agreement between early- and mid-treatment evaluations is 100%. However, an early follow-up evaluated as "Stable" can either evolve to "Responsive" or "Progressive", as well as remain "Stable".

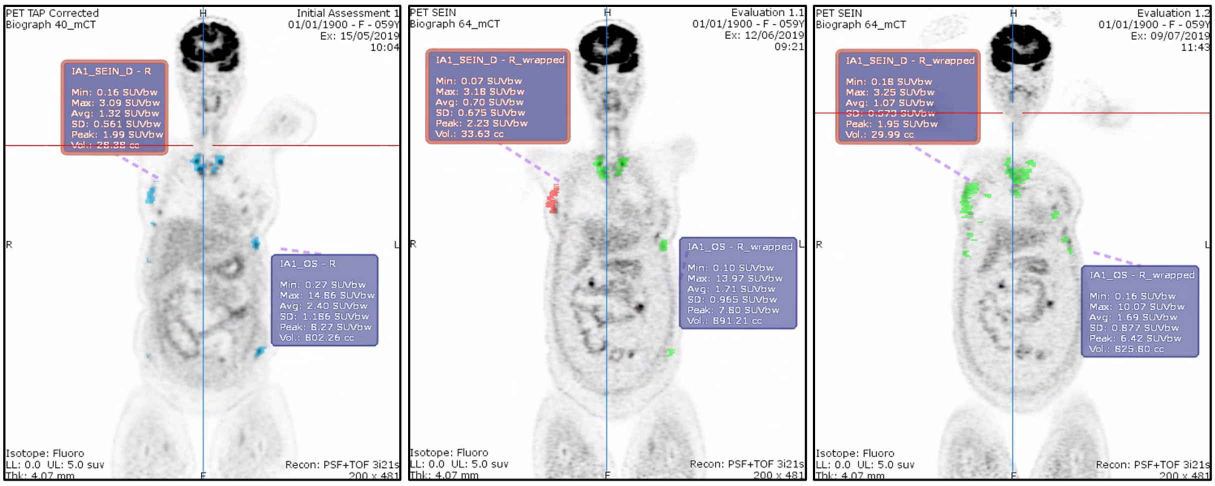


Figure 6.2 – Visual response evaluation. From left to right, pre-, early- and mid-treatment images. The pre-treatment image is the baseline and the early- and mid-treatment the follow-ups. On the baseline image, manually performed segmentations are visible in light blue. On the two follow-up images, propagated lesions are visible in green and red, respectively presenting a decreasing (responsive lesion) and an increasing (non-responsive lesion) SUL_{peak}.

6.5.3 Proposed visualization tool and clinical outcome

Physicians' follow-up readings are in accordance with the clinical outcome in 45% (physician 1) and 70% (physician 2) of the cases. The variable results are caused by the inter-operator variability. Indeed, as shown in Table 6.2, both physicians were given similar guidelines, but agreed only for 45% of the pairs of images. Physician 1 evaluated many images as “Stable” while the second one interpreted them as “Responsive”.

Despite the variability, the results of the tool are deemed promising, since it could complement other information usually available in clinical routine. Indeed, the routine clinical outcome is determined according to several information, such tumoral biomarkers, RECIST and PERCIST guidelines and the patient global health state. On the contrary, in our experiment, physicians evaluated the state of patient looking only at a 3D image with color-coded segmentations. Therefore, further evaluation considering the combination of both complementary types of information is an interesting idea for future work.

Table 6.2 – For each of the 20 studied pairs of images, evaluations performed by physicians 1 and 2 according to the colored segmentation visual representation, as well as the clinical outcomes from the clinical routine.

Patient	Physician 1	Physician 2	Clinical outcome
001_early	Responsive	Responsive	Responsive
001_pre	Responsive	Responsive	
002_early	Stable	Responsive	Responsive
002_pre	Responsive	Responsive	
003_early	Progressive	Stable	Stable
003_pre	Progressive	Progressive	
004_early	Stable	Responsive	Responsive
004_pre	Responsive	Responsive	
005_early	Stable	Responsive	Responsive
005_pre	Stable	Responsive	
006_early	Stable	Stable	Responsive
006_pre	Stable	Responsive	
007_early	Stable	Responsive	Responsive
007_pre	Responsive	Responsive	
008_early	Stable	Stable	Responsive
008_pre	Stable	Responsive	
009_early	Stable	Progressive	Stable
009_pre	Stable	Stable	
010_early	Stable	Responsive	Stable
010_pre	Stable	Responsive	

6.6 Conclusion

In this chapter, we propose a new automatic tool providing lesion-wise visual information about the tumor response to treatment.

From segmentations of the baseline acquisition propagated to follow-up images thanks to registration methods, we extract the SUL_{peak} biomarker. Then, for each lesion, the comparison of the biomarker values between baseline and follow-up segmentations allows the local evaluation of the response to treatment. This evaluation is represented through colored segmentation maps overlaid on the PET images. Finally, this colorful representation is shown to physicians to assist them during the monitoring of metastatic breast cancer.

The inter-expert variability of the response evaluations was high, while the agreement with the clinical outcome is relatively low. Such variability call for more precise definitions of the “Responsive”, “Stable” and “Progressive” response states. Moreover, PERCIST guidelines evaluate a patient as stable if he is not responsive nor in progression. Hence, it may be interesting to add a “Stable” state in our colored representation of the lesions.

Overall, lesion coloring helped evaluate treatment response, even though it should be noted that predicting the outcome from early- and mid-treatment images is a very difficult task [162]. In this sense, our method has a comparable performance to human experts, while being fully quantitative and reproducible. Moreover, early-treatment follow-up images were shown to be predictive of pre-treatment response, especially in the case of a responsive patient. Our study serves as a proof of concept, but similar experiments with more images and with more detailed qualitative feedback from the physicians are required in the future to develop a useful monitoring tool.

Summary:

✓ Biomarkers extracted from manually and automatically propagated segmentations are correlated, with the SUL_{peak} showing the highest correlation.

✓ Preliminary results show that automatic propagation applied to pre-treatment for predicting early-treatment response is a promising research direction.

✓ We propose a proof of concept visualization tool representing all individual lesion responses to assist the monitoring of metastatic breast cancer.

Conclusion and perspectives

In this thesis, we focused on improving the monitoring of metastatic breast cancer with longitudinal Positron Emission Tomography (PET) images. The main contributions of this thesis are summarized hereafter.

Our first and main contribution is a new registration method called Medical Image Registration Regularized By Architecture (MIRRBA), described in Chapter 4. It was designed to perform longitudinal registration on PET images. Based on the Deep Image Prior (DIP) scheme, our method optimizes the weights of a Convolutional Neural Network (CNN) (as in Deep Learning-based (DL) registration methods) but only for the pair of images to register (as in conventional registration methods). By design, our approach helps to bridge the gap between conventional and DL-based approaches. We provide an exhaustive experimental validation of our method on a private longitudinal metastatic breast cancer PET dataset, and to show its generality, on the public brain Magnetic Resonance Imaging (MRI) images.

MIRRBA reached results comparable to conventional registration methods, proving the efficiency of our approach. Regarding DL-based methods, we state that the training database and the neural network architecture introduce implicit regularization terms to the deformation model. MIRRBA is only implicitly regularized by the network architecture, since it is not optimized on a training database. The lack of data inductive biais may explain why MIRRBA reaches better results than DL-based methods on the private $\text{EPICURE}_{\text{seinmeta}}$ dataset. Indeed, metastatic breast cancer lesions are highly heterogeneous in shape, size and location across patients, making the learning of global registration patterns with full DL-based methods very challenging. Moreover, MIRRBA allows studying the parallel between components of conventional methods and architectural blocks, and we point at literature to support this parallel. Finally, we also show the possibility to combine MIRRBA with conventional or DL-based pre-computed deformation fields, to refine the registration.

Given its generality, MIRRBA could be studied on different registration problems, such as mono- or multi-modal, as well as intra- or inter-patient. The definition of an optimal neural network architecture as an over-parameterized representation of the deformation fields need further theoretical investigation.

Our second contribution is a method to automatically segment active organs on full body 3D PET images, bringing the concept of superpixels inside a DL framework either as input or in the loss function. We experimentally show that, in the context of PET imaging, superpixels provide an unsupervised way to focus the attention of a network towards objects of particular size. Superpixels also help homogenizing noise and differences across sites, improving generalizability.

The obtained segmentation can be used for different purposes, such as improving the automatic segmentation of cancerous regions in PET images by discarding high uptake organs, or enhancing the robustness of Position Emission tomography Response Criteria In Solid Tumors (PERCIST) guidelines.

In this thesis, we use the unsupervised superpixel-based segmentation approach to refine our in-house MIRRBA registration method, as described in Chapter 5. Indeed, cancerous regions, i.e. pathological ^{18}F-FluoroDeoxy-Glucose (^{18}F-FDG) uptake areas, present an uptake of similar intensity to the active organs, i.e. physiological ^{18}F-FDG uptake areas. If the locations of physiological ^{18}F-FDG uptake are known, we can ease the process of cancerous region detection by discarding the segmentations associated with active organs.

Going into another direction, Viglianti et al. [217] proposed to use the physiological ^{18}F-FDG uptake of the liver to normalize the Standardized Uptake Value (SUV) value and suggested updated PERCIST guidelines robust to uptake variations between acquisitions. Since our algorithm identifies active organs, it would be interesting to investigate the normalization of the SUV by ^{18}F-FDG uptake values w.r.t. other organs to see if it further improves the monitoring guidelines currently used.

Our third contribution is a new tool for the evaluation of biomarkers extracted from lesion segmentations. In Chapter 6, we propose an approach to extract biomarkers from the segmentations propagated between intra-patient acquisitions with a registration pipeline. We then propose a new visualization tool for the monitoring of metastatic breast cancer. This tool color-codes all metastatic breast cancer lesions according to their individual response to treatment. Even though tested on only 10 patients, this preliminary experiment showed promising results. It would be interesting to pursue a more thorough analysis to assess the tool's usefulness in clinical practice, and to further improve it by collaborating with physicians to include, for instance, stable lesions.

Moreover, for a patient presenting multiple lesions, it may be possible to develop a protocol to describe the global state of the patient from the individual responses of each lesion. Contrary to PERCIST which only consider the hottest lesion, this protocol could determine tumor response more accurately. This would be especially interesting when a patient presents heterogeneous metastatic lesions, i.e. lesions which do not belong to the same cancer sub-type and respond differently to treatments.

With these three contributions we achieved our goal of developing tools to improve the monitoring of metastatic breast cancer. We however consider the following limitations.

Overall, we mainly focused our registration methods on lesions which are always present (i.e. on lesions which neither appeared nor disappeared between two images). Indeed, from the registration point of view, lesion disappearance is not a smooth nor diffeomorphic problem since important modifications occur within the patient's body between two acquisitions. In the images, these modifications are translated as large deformations, which may correspond to folding areas. Hence, one of the trivial solutions would be to remove the smoothness and anti-folding regularization terms from our registration pipelines, at least locally around lesions. Full-removal is not an optimal solution, since a registration problem without regularization is ill-posed and the optimal solution is almost impossible to retrieve. Taking advantage of the known lesion locations thanks to segmentation methods, we could refine deformable registration very locally to improve the precision of our lesion propagation method. Another solution would be to develop a registration method allowing large displacement, such as Christensen et al. [30] who overcame the limitations of small deformation models by ensuring that transformations computed between images are diffeomorphic.

Regarding the appearance of new lesions, we found out that none of the registration pipelines could detect them. Addressing the challenge of registration and change detection, Dufresne et al. [44] proposed a *change loss* jointly optimized with a *registration loss* to monitor the evolution of multiple sclerosis lesions on brain MRI data. We adapted and implemented this loss in our MIRRBA pipeline. We obtained promising preliminary results so far, but a throughout analysis would be needed to evaluate the performances of this approach on the monitoring of metastatic breast cancer lesions.

As a conclusion, we made a first research step towards demonstrating the usefulness of early-acquired ^{18}F-FDG PET–CT images in the context of metastatic breast cancer. We hope this work will help physicians monitor patients presenting metastatic breast cancer.

Appendix A

Deep learning introduction

This annex aims at introducing DL. For more details and mathematical considerations, we refer to [120].

DL is a subset of Machine Learning (ML), which is a part of Artificial Intelligence (AI) (see Figure A.1). AI are programs which can learn, reason, act and adapt like humans. ML are algorithms with the ability to learn when exposed to more data over time, without much supervision. While ML uses simpler concepts, DL is an application of ML using algorithms in which Neural Networks (NNs) are trained on vast amounts of data to return accurate results. Until recently, NNs were limited in complexity by the computing power of available processors. Technological advancements have permitted larger and more sophisticated NNs, allowing computers to learn by example and improve on their own.

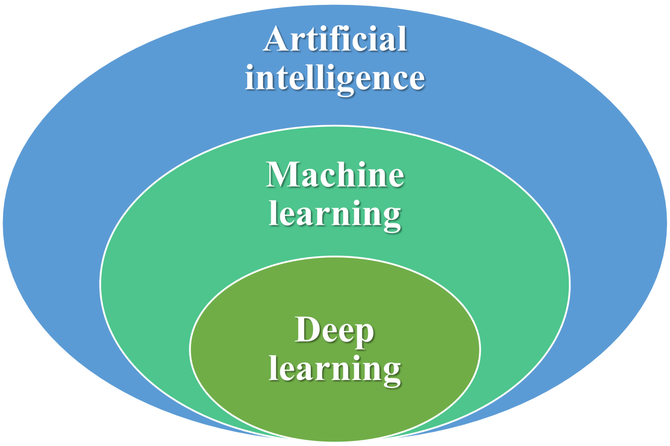


Figure A.1 – Deep Learning (DL) is a subset of Machine Learning (ML) which is also a subset of Artificial Intelligence (AI).

DL has a wide application range, from weather prediction to real-time bidding and targeted display advertising. DL has been developed to address problems such as image classification, language translation, speech recognition, pattern recognition,... with minimal human intervention.

A.1 Deep learning principle

To better grasp the DL principle, we consider hereafter a “pathological”/ “not-pathological” image classification problem.

“Pathological” and “non-pathological” images form together the *dataset*. This dataset is split into a *training dataset* and a *testing dataset*. For these images, the “pathological” and “non-pathological” classification is known: it is called the *ground truth*. This ground truth can be for instance manually determined. A large and diversified training dataset is recommended to train a more robust network.

Once the training dataset is created, images are converted into data and fed into the NN. These data move through the network, which delivers an output per image: “pathological” or “non-pathological”. This output is determined according to the network parameters, i.e. the weights and biases (see Section A.2 for more details).

NN output classifications are compared to the ground truth ones through the *cost function*. If they match, the output is confirmed. If not, the error is back-propagated through the network and parameters are adjusted to minimize the cost function, e.g. the accuracy in our example. Through multiple iterations and parameter updates, the NN tries to improve its classification performances, i.e. to optimize the cost function. Weights are adjusted until reaching a training limit such as a required performance criteria or a maximum number of iterations.

The next step is to test our trained NN. Without updating its parameters, the testing dataset images are fed to the trained network for a phase called the *inference*. Output classifications are again compared to the ground truth ones, but this time, testing images were not seen by the network during training. Hence, if the network performs well on this dataset, it proves its generalization performances and opens up the possibility to use it to classify as “pathological” or “not-pathological” images for which the ground truth is not known.

This type of training is called *supervised learning*, because ground truth classifications are provided to the network during training.

A.2 What are neural networks?

NN are made of layers. The first is the *input layer*, which receives the input data and transmits it to the next layer. Then, often more than one hidden layer is present. A network presenting more than two hidden layer is call a deep neural network. The *output layer* delivers the final result. This structure is illustrated in Figure A.2.

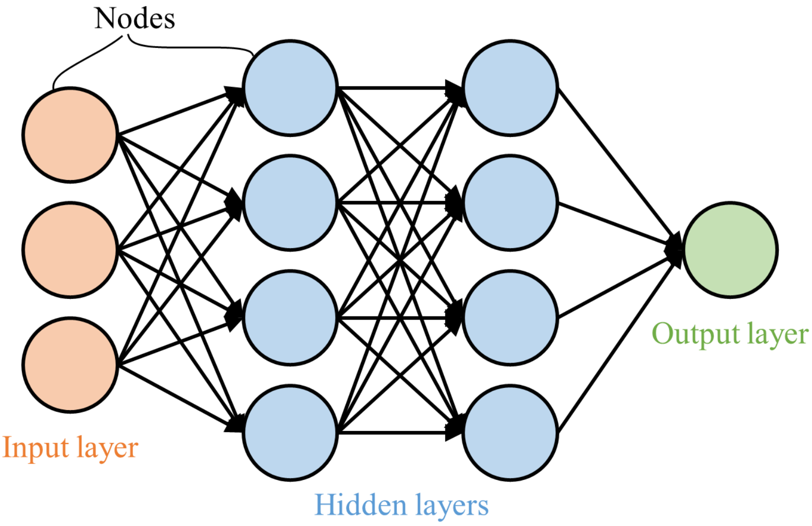


Figure A.2 – Networks are structured in layers: the input layer (orange), the hidden layers (blue) and the output layer (green). Layers are made of nodes (circles), transmitting information from one layer to the next one.

The layers of a NN are composed of neurons (or nodes). The term "neuron" was inspired by biological neurons, which receive information from thousands of other neurons through the *dendrites*, process this information in the *cell body*, and send the output signal to other neurons with the *axon*, to let the information flow. Just like the human brain made up of neurons, NNs are composed of nodes receiving, processing and transmitting information from one layer to another (see Figure A.2).

To process the information, each NN node performs the operations presented in Figure A.4. Let's consider a node j. First, each input value x_i for $i \in [1:n]$, e.g. each pixel value in image analysis, is multiplied by a weight w_{ij}. A heavier weighted node will have more effect on the next layer. All weighted products are then summed and a bias value b is added to all of them. An activation function f is finally applied to this biased weighted sum. The activation function is often non-linear to better represent complex data.

Frequently used non-linear activation functions are the sigmoid function (see Eq. A.1),

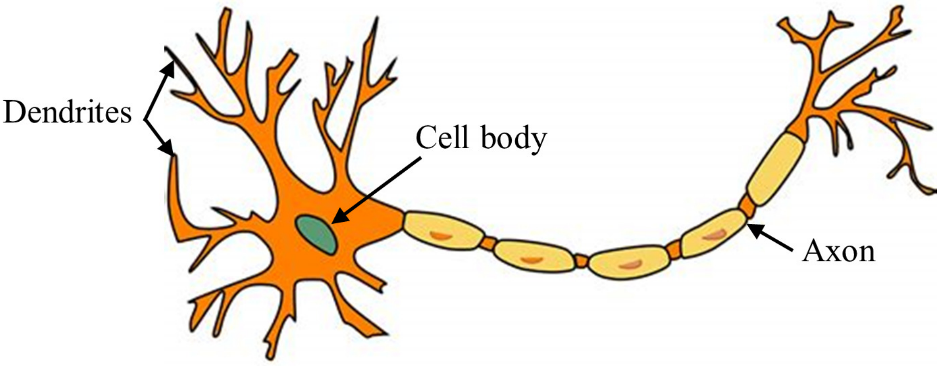


Figure A.3 – A biological neuron receives inputs from its dendrites, the information is processed in the cell body and output by the axon.

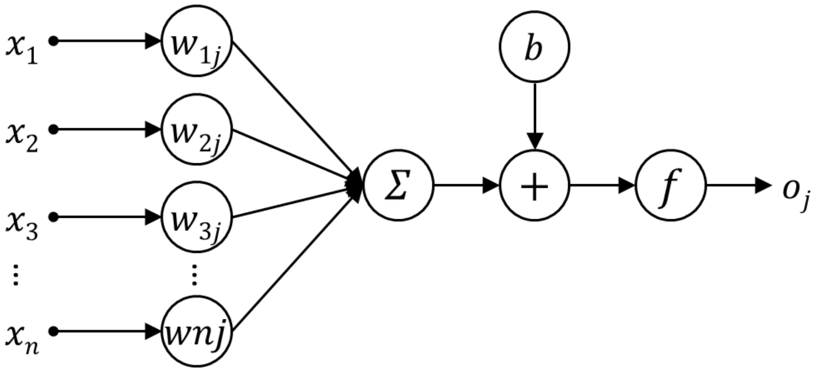


Figure A.4 – Operations in the node j: inputs x_i for $i \in [1 : n]$ are multiplied by a corresponding weight w_{ij}. These products are summed, and a bias b is added. This bias weighted sum is finally fed to an activation function f to obtain the node output o_j.

used for probability prediction, and the hyperbolic tangent function (see Eq. A.2), similar to the sigmoid function with a range of -1 to 1. The ReLU (rectified linear unit) function is the most commonly used these days, especially in Convolutional Neural Networks (CNNs) (see Section 2.3.2). ReLU performs an element-wise operation and sets all negative pixels to 0 (see Eq. A.3).

$$f(x) = \frac{1}{1+e^{-x}} \quad \text{(A.1)} \qquad f(x) = \frac{2}{1+e^{-2x}} - 1 \quad \text{(A.2)} \qquad f(x) = \begin{cases} 0 \text{ if } x < 0 \\ x \text{ if } x >= 0 \end{cases} \quad \text{(A.3)}$$

We have previously presented the generic elements of a network: its layers and nodes.

Yet, many combinations of these elements can be chosen depending on the application at hand.

Feed-forward networks [14], the simplest form of NN, are for instance used by almost all vision and speech recognition applications. In Recurrent Neural Networks (RNNs) [186], hidden layers save their outputs to be used for future prediction and become part of the new inputs. Their applications include text-to-speech conversion. Spatial Transformer Networks (STNs) learn to spatially transform feature maps [98], while Generative Adversial Networks (GANs) are the most widely used generative models for image synthesis and are used in the medical domain as tools for data augmentation [35]. As presented in the following section, CNNs are mainly used to process images.

A.3 Convolution operation

Convolution is the basic operation of any CNN. It is a simple mathematical operation generally used for image processing, acting as a filter. Convolution steps are detailed in Figure A.5 and hereafter:

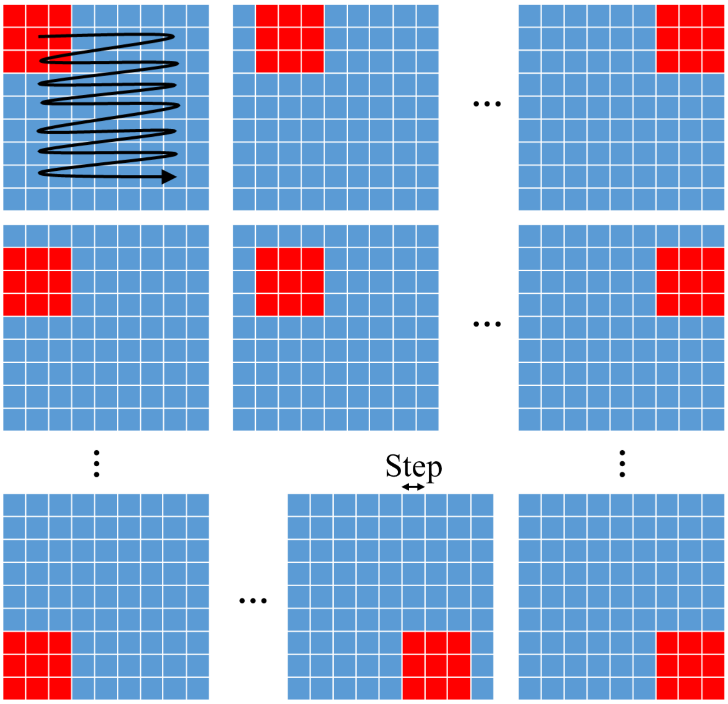


Figure A.5 – Convolution operation. The kernel (red) slides in an image (blue) by a step ($step = 1$ here) to obtain a feature map.

1. The *kernel* (top left red square Figure A.5) is first defined. Its size is generally 3x3 or 5x5 in 2D (and 3x3x3 or 5x5x5 in 3D). In figure A.5, the kernel size is 3x3.
2. The kernel progressively slides from left to right and from top to bottom with a pre-defined *step* until the whole image has been considered. Figure A.5, $step = 1$. The overlap of the considered image sub-parts, or tiles, improves the representation of the image.
3. A dot product is performed between the kernel and each image tile. As a result of all these operations, a *feature map* is obtained: it is an abstract representation of the image showing feature locations in images. Its values depend on the convolutional kernel parameters and on the pixel values of the input image. In the case of a CNN, the parameters of the kernel are learned during the training phase.

Appendix B

Elastix registration parameters

B.1 Rigid registration parameters

```
// ********** Image Types **********
(FixedInternalImagePixelType "float")
(FixedImageDimension 3)
(MovingInternalImagePixelType "float")
(MovingImageDimension 3)

// ********** Components **********
(Registration "MultiResolutionRegistration")
(Interpolator "BSplineInterpolator")
(ResampleInterpolator "FinalBSplineInterpolator")
(Resampler "DefaultResampler")

(FixedImagePyramid "FixedRecursiveImagePyramid")
(MovingImagePyramid "MovingRecursiveImagePyramid")

(Optimizer "AdaptiveStochasticGradientDescent")
(Transform "EulerTransform")
(AutomaticScalesEstimation "true")
(Metric "AdvancedNormalizedCorrelation")

// ********** Pyramid **********
```

```
// Total number of resolutions
(NumberOfResolutions 3)
(ImagePyramidSchedule 4 4 4 2 2 2 1 1 1)

// ********** Transform **********
(AutomaticTransformInitialization "true")
(AutomaticTransformInitializationMethod "CenterOfGravity")
(HowToCombineTransforms "Compose")

// ********** Optimizer **********
(MaximumNumberOfIterations 500)
(AutomaticParameterEstimation "true")
(UseAdaptiveStepSizes "true")

// ********** ImageSampler **********
//Number of spatial samples used to compute the mutual information in each
resolution level:
(ImageSampler "Random")
(NumberOfSpatialSamples 1500)
(NewSamplesEveryIteration "true")
(UseRandomSampleRegion "true")
(NumberOfHistogramBins 64)

// ********** Interpolator and Resampler **********
//Order of B-Spline interpolation used in each resolution level:
(BSplineInterpolationOrder 2)

//Order of B-Spline interpolation used for applying the final deformation:
(FinalBSplineInterpolationOrder 3)

//Default pixel value for pixels that come from outside the picture:
(DefaultPixelValue 0)

// ********** Other **********
```

```
(ErodeMask "false")
(UseDirectionCosines "true")
```

B.2 Affine registration parameters

```
// ********** Image Types **********
(FixedInternalImagePixelType "float")
(FixedImageDimension 3)
(MovingInternalImagePixelType "float")
(MovingImageDimension 3)

// ********** Components **********
(Registration "MultiResolutionRegistration")
(Interpolator "BSplineInterpolator")
(ResampleInterpolator "FinalBSplineInterpolator")
(Resampler "DefaultResampler")

(FixedImagePyramid "FixedRecursiveImagePyramid")
(MovingImagePyramid "MovingRecursiveImagePyramid")

(Optimizer "AdaptiveStochasticGradientDescent")
(Transform "AffineTransform")
(Metric "AdvancedNormalizedCorrelation")

// ********** Pyramid **********
// Total number of resolutions
(NumberOfResolutions 3)
(ImagePyramidSchedule 4 4 4 2 2 2 1 1 1)

// ********** Transform **********
(AutomaticTransformInitialization "false")
(FinalGridSpacingInPhysicalUnits 15.0 15.0 15.0)
(HowToCombineTransforms "Compose")
```

```
// ********** Optimizer **********
(MaximumNumberOfIterations 500)
(AutomaticParameterEstimation "true")
(UseAdaptiveStepSizes "true")

// ********** ImageSampler **********
//Number of spatial samples used to compute the mutual information in each
resolution level:
(ImageSampler "Random")
(NumberOfSpatialSamples 500)
(NewSamplesEveryIteration "true")
(UseRandomSampleRegion "true")
(NumberOfHistogramBins 64)

// ********** Interpolator and Resampler **********
//Order of B-Spline interpolation used in each resolution level:
(BSplineInterpolationOrder 2)

//Order of B-Spline interpolation used for applying the final deformation:
(FinalBSplineInterpolationOrder 3)

//Default pixel value for pixels that come from outside the picture:
(DefaultPixelValue 0)

// ********** Several **********
(ErodeMask "false")
(UseDirectionCosines "true")
```

B.3 Deformable registration parameters

```
// ********** Image Types **********
(FixedInternalImagePixelType "float")
```

```
(FixedImageDimension 3)
(MovingInternalImagePixelType "float")
(MovingImageDimension 3)

// ********** Components **********
(Registration "MultiMetricMultiResolutionRegistration")
(Interpolator "BSplineInterpolator")
(ResampleInterpolator "FinalBSplineInterpolator")
(Resampler "DefaultResampler")

(FixedImagePyramid "FixedRecursiveImagePyramid")
(MovingImagePyramid "MovingRecursiveImagePyramid")

(Optimizer "AdaptiveStochasticGradientDescent")
(Transform "BSplineTransform")
(Metric "AdvancedNormalizedCorrelation" "TransformBendingEnergyPenalty")
(Metric0Weight 1.0)
(Metric1Weight 1.0)

// ********** Pyramid **********
// Total number of resolutions
(NumberOfResolutions 4)
(ImagePyramidSchedule 8 8 8 4 4 4 2 2 2 1 1 1)

// ********** Transform **********
(AutomaticTransformInitialization "false")
(FinalGridSpacingInPhysicalUnits 15.0 15.0 15.0 15.0)
(HowToCombineTransforms "Compose")

// ********** Optimizer **********
(MaximumNumberOfIterations 1000)

// ********** ImageSampler **********
//Number of spatial samples used to compute the mutual information in each
```

```
resolution level:
(ImageSampler "Random")
(NumberOfSpatialSamples 2048)
(NewSamplesEveryIteration "false")
(NumberOfHistogramBins 32)

// ********** Interpolator and Resampler **********
//Order of B-Spline interpolation used in each resolution level:
(BSplineInterpolationOrder 1)

// Order of B-Spline interpolation used for applying the final deformation:
(FinalBSplineInterpolationOrder 3)

//Default pixel value for pixels that come from outside the picture:
(DefaultPixelValue 0)

// ********** Several **********
(ErodeMask "false")
(UseDirectionCosines "true")
```

Bibliography

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, « SLIC superpixels compared to state of the art superpixel methods », *in*: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.*11* (2012), pp. 2274–2281, DOI: `10.1109/TPAMI.2012.120`.

[2] O. Acosta, J. Dowling, G. Cazoulat, A. Simon, O. Salvado, de R. Crevoisier, and P. Haigron, « Atlas based segmentation and mapping of organs at risk from planning CT for the development of voxel-wise predictive models of toxicity in prostate radiotherapy », *in*: *International Workshop on Prostate Cancer Imaging* (2010), pp. 42–51, DOI: `https://doi.org/10.1007/978-3-642-15989-3_6`.

[3] S. Afshari, A. BenTaieb, and G. Hamarneh, « Automatic localization of normal active organs in 3D PET scans », *in*: *Computerized Medical Imaging and Graphics* 70 (2018), pp. 111–118, DOI: `10.1016/j.compmedimag.2018.09.008`.

[4] A. Almuhaideb, N. Papathanasiou, and J. Bomanji, « 18F-FDG PET/CT imaging in oncology », *in*: *Annals of Saudi Medicine* 31.*1* (2011), pp. 3–13, DOI: `10.4103/0256-4947.75771`.

[5] J. Ashburner, « A fast diffeomorphic image registration algorithm », *in*: *NeuroImage* 38.*1* (2007), pp. 95–113, DOI: `10.1016/j.neuroimage.2007.07.007`.

[6] J. Ashburner and K. J. Friston, « Nonlinear Spatial Normalization Using Basis Functions », *in*: *Human Brain Mapping* 266 (1999), pp. 254–266, DOI: `10.1002/(SICI)1097-0193(1999)7:4<254::AID-HBM4>3.0.CO;2-G`.

[7] B. B. Avants, N. J. Tustison, and H. Johnson, « Advanced Normalization Tools (ANTS) », *in*: *Insight Journal* 2.*365* (2009), pp. 1–35.

[8] S. Avril, R. F. Muzic, D. Plecha, B. J. Traughber, S. Vinayak, and N. Avril, « 18F-FDG PET/CT for Monitoring of Treatment Response in Breast Cancer », *in*: *Journal of Nuclear Medicine* 57 (2016), 34S–39S, DOI: `10.2967/jnumed.115.157875`.

[9] D. O. Baguer, J. Leuschner, and M. Schmidt, « Computed tomography reconstruction using deep image prior and learned reconstruction methods », *in*: *Inverse Problems* 36 (2020), DOI: `10.1088/1361-6420/aba415`.

[10] D. L. Bailey, D. W. Townsend, P. E. Valk, and M. N. Maisey, *Positron Emission Tomography*, 2005.

[11] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, « An unsupervised learning model for deformable medical image registration », *in*: *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), DOI: `10.1039/c5tc00349k`.

[12] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, « VoxelMorph: A Learning Framework for Deformable Medical Image Registration », *in*: *IEEE Transactions on Medical Imaging* (2018), pp. 1–13, DOI: `10.1109/TMI.2019.2897538`.

[13] M. Barkati, D. Simard, D. Taussky, and G. Delouya, « Magnetic resonance imaging for prostate bed radiotherapy planning: An inter- and intra-observer variability study », *in*: *Journal of Medical Imaging and Radiation Oncology* 60.*2* (2016), pp. 255–259, DOI: `10.1111/1754-9485.12416`.

[14] G. Bebis and M. Georgiopoulos, « Feed-forward neural networks », *in*: *Ieee Potentials* (1994), pp. 27–31.

[15] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes, « Computing large deformation metric mappings via geodesic flows of diffeomorphisms », *in*: *International Journal of Computer Vision* 61.*2* (2005), pp. 139–157, DOI: `10.1023/B:VISI.0000043755.93987.aa`.

[16] L. Beljaards, M. S. Elmahdy, Fons Verbeek, and M. Staring, « A Cross-Stitch Architecture for Joint Registration and Segmentation in Adaptive Radiotherapy », *in*: *arXiv preprint* (2020), pp. 1–12, DOI: `10.48550/arXiv.2004.08122`.

[17] L. Bi, J. Kim, D. Feng, and M. Fulham, « Multi-stage thresholded region classification for whole-body PET-CT lymphoma studies », *in*: *MICCAI* 8673 LNCS.*PART 1* (2014), pp. 569–576, DOI: `10.1007/978-3-319-10404-1_71`.

[18] L. Bi, J. Kim, A. Kumar, D. Feng, and M. Fulham, « Adaptive Supervoxel Patch-based Region Classification in Whole-Body PET-CT », *in*: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (2015).

[19] P. Blanc-Durand, A. Van Der Gucht, N. Schaefer, E. Itti, and J. O. Prior, « Automatic lesion detection and segmentation of 18F-FET PET in gliomas: A full 3D U-Net convolutional neural network study », *in*: *PLoS ONE* 13.*4* (2018), pp. 1–11, DOI: `10.1371/journal.pone.0195798`.

[20] R. Boellaard, N. C. Krak, O. S. Hoekstra, and A. A. Lammertsma, « Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values : a simulation study », *in*: *Journal of nuclear medicine* 45.*3* (2004), pp. 1519–1527.

[21] F. L. Bookstein, « Principal Warps: Thin-Plate Splines and the Decomposition of Deformations », *in*: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11.*6* (1989), pp. 567–585, DOI: `10.1109/34.24792`.

[22] B. Buerke, M. Puesken, S. Müter, M. Weckesser, J. Gerss, W. Heindel, and J. Wessling, « Measurement accuracy and reproducibility of semiautomated metric and volumetric lymph node analysis in MDCT », *in*: *American Journal of Roentgenology* 195.*4* (2010), pp. 979–985, DOI: `10.2214/AJR.10.4010`.

[23] M. Buyse, P. Thirion, R. W. Carlson, T. Burzykowski, G. Molenberghs, and P. Piedbois, « Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: A meta-analysis », *in*: *Lancet* 356.*9227* (2000), pp. 373–378, DOI: `10.1016/S0140-6736(00)02528-9`.

[24] G. Calmon and N. Roberts, « Automatic measurement of changes in brain volume on consecutive 3D MR images by segmentation propagation », *in*: *Magnetic Resonance Imaging* 18.*4* (2000), pp. 439–453, DOI: `10.1016/S0730-725X(99)00118-6`.

[25] X. Cao, J. Yang, J. Zhang, D. Nie, M. Kim, Q. Wang, and D. Shen, « Deformable image registration based on similarity-steered CNN regression », *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence*

and Lecture Notes in Bioinformatics) 10433 LNCS (2017), pp. 300–308, DOI: `10.1007/978-3-319-66182-7_35`.

[26] Y. Cao, M. I. Miller, R. L. Winslow, and L. Younes, « Large deformation diffeomorphic metric mapping of fiber orientations », *in*: *Proceedings of the IEEE International Conference on Computer Vision* II.*9* (2005), pp. 1379–1386, DOI: `10.1109/ICCV.2005.132`.

[27] T. Carlier and C. Bailly, « State-Of-The-Art and Recent Advances in Quantification for Therapeutic Follow-Up in Oncology Using PET », *in*: *Frontiers in Medicine* 2 (2015), DOI: `10.3389/fmed.2015.00018`.

[28] X. Chen, A. Diaz-pinto, N. Ravikumar, and A. F. Frangi, « Deep learning in medical image registration », *in*: *Progress in Biomedical Engineering* 3.*1* (2021), DOI: `10.1088/2516-1091/abd37c`.

[29] V. Cheplygina, M. de Bruijne, and J. P. W. Pluim, « Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis », *in*: *Medical Image Analysis* 54 (2019), pp. 280–296, DOI: `10.1016/j.media.2019.03.009`.

[30] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, « Deformable templates using large deformation kinematics », *in*: *IEEE Transactions on Image Processing* 5.*10* (1996), pp. 1435–1447, DOI: `10.1109/83.536892`.

[31] S. Christodoulidis, M. Sahasrabudhe, M. Vakalopoulou, G. Chassagnon, M.-P. Revel, S. Mougiakakou, and N. Paragios, « Linear and deformable image registration with 3D convolutional neural networks », *in*: *Image Analysis for Moving Organ, Breast, and Thoracic Images* (2018), pp. 13–22, DOI: `10.1007/978-3-030-00946-5_2`.

[32] M. Colombié, P. Jézéquel, M. Rubeaux, J. S. Frenel, F. Bigot, V. Seegers, and M. Campone, « The EPICURE study: a pilot prospective cohort study of heterogeneous and massive data integration in metastatic breast cancer patients », *in*: *BMC Cancer* 21.*1* (2021), pp. 1–9, DOI: `10.1186/s12885-021-08060-8`.

[33] T. F. Cootes and C. J. Taylor, « Anatomical statistical models and their role in feature extraction », *in*: *British Journal of Radiology* 77.*SPEC. ISS. 2* (2004), pp. 133–139, DOI: `10.1259/bjr/20343922`.

[34] C. M. Costelloe, H. H. Chuang, J. E. Madewell, and N. T. Ueno, « Cancer response criteria and bone metastases: RECIST 1.1, MDA and PERCIST », *in*: *Journal of Cancer* 1.*1* (2010), pp. 80–92, DOI: `10.7150/jca.1.80`.

[35] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, « Generative adversarial network: An overview », *in*: *IEEE Signal Processing Magazine* 35.*1* (2018), pp. 53–65, DOI: `10.19650/j.cnki.cjsi.J1804413`.

[36] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, « Deformable Convolutional Networks », *in*: *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 764–773, DOI: `10.1109/ICCV.2017.89`.

[37] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, « Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces », *in*: *Medical Image Analysis* 57 (2019), pp. 226–236, DOI: `10.1016/j.media.2019.07.006`.

[38] A. V. Dalca, A. Hering, L. Hansen, and M. Heinrich, *Learn2Reg Challenge*, 2020, URL: `https://learn2reg.grand-challenge.org/` (visited on 02/10/2020).

[39] J. E. Dancey, L. E. Dodd, R. Ford, R. Kaplan, M. Mooney, L. Rubinstein, L. H. Schwartz, L. Shankar, and P. Therasse, « Recommendations for the assessment of progression in randomised cancer treatment trials », *in*: *European Journal of Cancer* 45.*2* (2009), pp. 281–289, DOI: `10.1016/j.ejca.2008.10.042`.

[40] C. Davatzikos, « Spatial Transformation and Registration of Brain Images Using Elastically Deformable Models », *in*: *Computer Vision and Image Understanding* 66.*2* (1997), pp. 207–222, DOI: `10.1006/cviu.1997.0605`.

[41] S. Dittmer, T. Kluth, P. Maass, and D. O. Baguer, « Regularization by architecture : A deep prior approach for inverse problems », *in*: *Journal of Mathematical Imaging and Vision* 62.*3* (2020), pp. 456–470, DOI: `10.1007/s10851-019-00923-x`.

[42] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazırbas, and V. Golkov, « FlowNet: Learning Optical Flow with Convolutional Networks », *in*: *Proceedings of the IEEE international conference on computer vision* (2015), pp. 2758–2766.

[43] J. Dubreuil, F. Cachin, A. Berriolo-Ridinger, and A. Skanjeti, « Critères d'interprétation en imagerie cancérologique solide : RECIST, PERCIST… », *in*: *Medecine Nucleaire* 41.*3* (2017), pp. 241–248, DOI: `10.1016/j.mednuc.2017.02.010`.

[44] E. Dufresne, D. Fortun, B. Kumar, S. Kremer, and V. Noblet, « Joint Registration and Change Detection in Longitudinal Brain MRI », *in*: *Proceedings - International Symposium on Biomedical Imaging* 2020-April (2020), pp. 104–108, DOI: `10.1109/ISBI45749.2020.9098626`.

[45] M. van Eijnatten, L. Rundo, K. J. Batenburg, F. Lucka, E. Beddowes, C. Caldas, F. A. Gallagher, E. Sala, C.-B. Schönlieb, and R. Woitek, « 3D deformable registration of longitudinal abdominopelvic CT images using unsupervised deep learning », *in*: *Computer Methods and Programs in Biomedicine* (2020), DOI: `10.1016/j.cmpb.2021.106261`.

[46] E. A. Eisenhauer, P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe, and J. Verweij, « New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1) », *in*: *European Journal of Cancer* 45.*2* (2009), pp. 228–247, DOI: `10.1016/j.ejca.2008.10.026`.

[47] M. S. Elmahdy, T. Ahuja, U. A. Van Der Heide, and M. Staring, « Patient-Specific Finetuning of Deep Learning Models for Adaptive Radiotherapy in Prostate CT », *in*: *Proceedings - International Symposium on Biomedical Imaging* 2020-April (2020), pp. 577–580, DOI: `10.1109/ISBI45749.2020.9098702`.

[48] K. A. J. Eppenhof and J. P. W. Pluim, « Error estimation of deformable image registration of pulmonary CT scans using convolutional neural networks », *in*: *Journal of Medical Imaging* 5.*02* (2018), p. 1, DOI: `10.1117/1.jmi.5.2.024003`.

[49] J. Fan, X. Cao, P. T. Yap, and D. Shen, « BIRNet: Brain image registration using dual-supervised fully convolutional networks », *in*: *Medical Image Analysis* 54 (2019), pp. 193–206, DOI: `10.1016/j.media.2019.03.006`.

[50] J. M. Fitzpatrick, *Detecting failure, assessing success*, CRC Press Baton Rouge, 2001.

[51] J. Fledelius, A. Khalil, K. Hjorthaug, and J. Frøkiær, « Inter-observer agreement improves with PERCIST 1.0 as opposed to qualitative evaluation in non-small cell lung cancer patients evaluated with F-18-FDG PET/CT early in the course of chemo-radiotherapy », *in*: *EJNMMI Research* 6.*1* (2016), DOI: `10.1186/s13550-016-0223-6`.

[52] C. B. Fookes and M. Bennamoun, *Rigid and non-rigid image registration and its association with mutual information: A review*, 2002.

[53] W. D. Foulkes, I. E. Smith, and J. S. Reis-Filho, « Triple-Negative Breast Cancer », *in*: *New England journal of medicine* 363.*20* (2010).

[54] **C. Fourcade**, L. Ferrer, N. Moreau, G. Santini, A. Brennan, C. Rousseau, M. Lacombe, V. Fleury, M. Colombié, P. Jézéquel, M. Campone, M. Rubeaux, and D. Mateus, « Deformable Image Registration with Deep Network Priors: a Study on Longitudinal PET Images », *in*: *arXiv preprint* (2021), pp. 1–25, DOI: `10.48550/arXiv.2111.11873`.

[55] **C. Fourcade**, L. Ferrer, N. Moreau, G. Santini, A. Brennan, C. Rousseau, M. Lacombe, V. Fleury, M. Colombié, P. Jézéquel, M. Campone, M. Rubeaux, D. Mateus, B. Maucherat, and J. S. Frenel, « PERCIST-like automatic guided treatment response assessment on FDG PET images driven by full extension metastatic breast cancer », *in*: *American Society of Clinical Oncology (ASCO) conference* (2022).

[56] **C. Fourcade**, L. Ferrer, G. Santini, N. Moreau, C. Rousseau, M. Lacombe, C. Guillerminet, M. Colombié, M. Campone, D. Mateus, and M. Rubeaux, « Combining Superpixels and Deep Learning Approaches to Segment Active Organs in Metastatic Breast Cancer PET Images », *in*: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* 2020-July (2020), pp. 1536–1539, DOI: `10.1109/EMBC44109.2020.9175683`.

[57] **C. Fourcade**, M. Rubeaux, and D. Mateus, « Using Elastix to Register Inhale/Exhale Intrasubject Thorax CT: A Unsupervised Baseline to the Task 2 of the Learn2Reg Challenge », *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12587 LNCS (2021), pp. 100–105, DOI: `10.1007/978-3-030-71827-5_13`.

[58] J. V. Francés, « Fusión de imágenes médicas », *in*: *Universitat de Valencia - Escola tècnica superior enginyeria* 1.*Técnicas de Imágenes Biomédicas – MIB* (2017).

[59] J. V. Francés, « Imágenes por rayos X », *in*: *Universitat de Valencia - Escola tècnica superior enginyeria* 1.*Técnicas de Imágenes Biomédicas – MIB* (2017).

[60] J. V. Francés, « Medicina nuclear », *in*: *Universitat de Valencia - Escola tècnica superior enginyeria* 1.*Técnicas de Imágenes Biomédicas – MIB* (2017).

[61] J. V. Francés, « Tomografía Axial Computarizada », *in*: *Universitat de Valencia - Escola tècnica superior enginyeria* 1.*Técnicas de Imágenes Biomédicas – MIB* (2017), DOI: `10.3917/spi.031.0051`.

[62] N. M. T. Freedman, S. K. Sundaram, K. Kurdziel, J. A. Carrasquillo, M. Whatley, J. M. Carson, D. Sellers, S. K. Libutti, J. C. Yang, and S. L. Bacharach, « Comparison of SUV and Patlak slope for monitoring of cancer therapy using serial PET scans », *in*: *European Journal of Nuclear Medicine and Molecular Imaging* 30.*1* (2003), pp. 46–53, DOI: `10.1007/s00259-002-0981-4`.

[63] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, « Deep learning in medical image registration: a review », *in*: *Physics in Medicine and Biology* (2020), pp. 1–32, DOI: `10.1088/1361-6560/ab843e`.

[64] C. M. Gaeta, J. L. Vercher-Conejero, A. C. Sher, A. Kohan, C. Rubbert, and N. Avril, « Recurrent and metastatic breast cancer PET, PET/CT, PET/MRI: FDG and new biomarkers », *in*: *Quarterly Journal of Nuclear Medicine and Molecular Imaging* 57.*4* (2013), pp. 352–366.

[65] S. Gaujoux, D. Goéré, F. Dumont, A. Souadka, C. Dromain, M. Ducreux, and D. Elias, « Complete radiological response of colorectal liver metastases after chemotherapy: What can we expect? », *in*: *Digestive Surgery* 28.*2* (2011), pp. 114–120, DOI: `10.1159/000323820`.

[66] S. Ghose, A. Oliver, R. Martí, X. Lladó, J. C. Vilanova, J. Freixenet, J. Mitra, D. Sidibé, and F. Meriaudeau, « A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images », *in*: *Computer Methods and Programs in Biomedicine* 108.*1* (2012), pp. 262–287, DOI: `10.1016/j.cmpb.2012.04.006`.

[67] B. Glocker, A. Sotiras, N. Komodakis, and N. Paragios, « Deformable medical image registration: Setting the state of the art with discrete methods », *in*: *Annual Review of Biomedical Engineering* 13 (2011), pp. 219–244, DOI: `10.1146/annurev-bioeng-071910-124649`.

[68] B. Glocker, D. Zikic, N. Komodakis, N. Paragios, and N. Navab, « Linear image registration through MRF optimization », *in*: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (2009), pp. 422–425.

[69] K. Gong, J. Guan, K. Kim, X. Zhang, J. Yang, Y. Seo, G. El Fakhri, J. Qi, and Q. Li, « Iterative PET image reconstruction using convolutional neural network representation », *in*: *IEEE Transactions on Medical Imaging* 38.*3* (2019), pp. 675–685, DOI: `10.1109/TMI.2018.2869871`.

[70] D. Goulon, H. Necib, B. Henaff, C. Rousseau, T. Carlier, and F. Kraeber-Bodere, « Quantitative Evaluation of Therapeutic Response by FDG PET-CT in Metastatic Breast Cancer », *in*: *Frontiers in Medicine* 3 (2016), DOI: `10.3389/fmed.2016.00019`.

[71] D. Groheux, A. Cochet, O. Humbert, J.-L. Alberini, E. Hindie, and D. Mankoff, « 18F-FDG PET/CT for Staging and Restaging of Breast Cancer », English, *in*: *Journal of Nuclear Medicine* 57.*Supplement_1* (2016), 17S–26S, DOI: `10.2967/jnumed.115.157859`.

[72] D. Gu, X. Cao, S. Ma, L. Chen, G. Liu, D. Shen, and Z. Xue, « Pair-Wise and Group-Wise Deformation Consistency in Deep Registration Network », *in*: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (2020), pp. 171–180, DOI: `10.1007/978-3-030-59716-0_17`.

[73] M. Gupta, P. S. Choudhury, S. Rawal, H. C. Goel, and S. A. Rao, « Correction to: Evaluation of RECIST, PERCIST, EORTC, and MDA Criteria for Assessing Treatment Response with Ga68-PSMA PET-CT in Metastatic Prostate Cancer Patient with Biochemical Progression: a Comparative Study (Nuclear Medicine and Molecular Imaging, (2018), 52, 6, (420-429), 10.1007/s13139-018-0548-3) », *in*: *Nuclear Medicine and Molecular Imaging* 54.*5* (2020), p. 267, DOI: `10.1007/s13139-020-00662-7`.

[74] B. Gutiérrez-Becker, D. Mateus, L. Peter, and N. Navab, « Learning optimization updates for multimodal registration », *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9902 LNCS (2016), pp. 19–27, DOI: `10.1007/978-3-319-46726-9_3`.

[75] L. M. Hamberg, G. J. Hunter, N. M. Alpert, N. C. Choi, J. W. Babich, and A. J. Fischman, « The dose uptake ratio as an index of glucose metabolism: Useful parameter or oversimplification? », *in*: *Journal of Nuclear Medicine* 35.*8* (1994), pp. 1308–1312.

[76] Y. Han, J. Yoo, H. H. Kim, H. J. Shin, K. Sung, and J. C. Ye, « Deep learning with domain adaptation for accelerated projection-reconstruction MR », *in*: *Magnetic Resonance in Medicine* 80.*3* (2018), pp. 1189–1205, DOI: `10.1002/mrm.27106`.

[77] J. R. Harris, M. E. Lippman, C. K. Osborne, and M. Morrow, *Diseases of the Breast*, 2012.

[78] G. Haskins, U. Kruger, and P. Yan, « Deep Learning in Medical Image Registration: A Survey », *in*: *Machine Vision and Applications* 31.*1* (2019), p. 8.

[79] K. He, X. Zhang, S. Ren, and J. Sun, « Deep residual learning for image recognition », *in*: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* 2016-Decem (2016), pp. 770–778, DOI: `10.1109/CVPR.2016.90`.

[80] K. He, X. Zhang, S. Ren, and J. Sun, « Identity Mappings in Deep Residual Networks », *in*: *European conference on computer vision* (2016), pp. 630–645, DOI: `10.1007/978-3-319-46493-0`.

[81] F. Heckel, H. Meine, J. H. Moltz, J. M. Kuhnigk, J. T. Heverhagen, A. Kießling, B. Buerke, and H. K. Hahn, « Segmentation-based partial volume correction for volume estimation of solid lesions in CT », *in*: *IEEE Transactions on Medical Imaging* 33.*2* (2014), pp. 462–480, DOI: `10.1109/TMI.2013.2287374`.

[82] R. Heckel and M. Soltanolkotabi, « Compressive sensing with un-trained neural networks: Gradient descent finds the smoothest approximation », *in*: *International Conference on Machine Learning* (2020).

[83] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, « Automatic anatomical brain MRI segmentation combining label propagation and decision fusion », *in*: *NeuroImage* 33.*1* (2006), pp. 115–126, DOI: `10.1016/j.neuroimage.2006.05.061`.

[84] M. P. Heinrich, « Closing the Gap Between Deep and Conventional Image Registration Using Probabilistic Dense Displacement Networks », *in*: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* October (2019), pp. 50–58, DOI: `10.1007/978-3-030-32226-7_6`.

[85] M. P. Heinrich and L. Hansen, « Highly accurate and memory efficient unsupervised learning-based discrete CT registration using 2.5 D displacement search », *in*: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (2020), pp. 190–200, DOI: `10.1007/978-3-030-59716-0_19`.

[86] M. P. Heinrich, O. Maier, and H. Handels, « Multi-modal multi-atlas segmentation using discrete optimisation and self-similarities », *in*: *CEUR Workshop Proceedings* 1390.*January* (2015), pp. 27–30.

[87] T. Henry, C. Meyer, V. Chevance, V. Roblot, E. Blanchet, V. Arnould, G. Grimon, M. Chekroun, F. Parent, A. Seferian, R. Jovan, S. Bulifon, X. Jais, D. Montani, M. Humbert, P. Chaumet-Riffaud, V. Lebon, E. Durand, and F. Besson, « Automated PET segmentation for lung tumors: Can deep learning reach optimized expert-based performance? », *in*: *Journal of Nuclear Medicine* 59.*1* (2018), pp. 322–322.

[88] A. Hering, L. Hansen, T. C. W. Mok, A. C. S. Chung, H. Siebert, S. Häger, A. Lange, S. Kuckertz, S. Heldmann, W. Shao, S. Vesal, M. Rusu, G. Sonn, T. Estienne, M. Vakalopoulou, L. Han, Y. Huang, M. Brudfors, Y. Balbastre, S. Outard, M. Modat, G. Lifshitz, D. Raviv, J. Lv, Q. Li, V. Jaouen, D. Visvikis, **C. Fourcade**, M. Rubeaux, W. Pan, Z. Xu, B. Jian, F. De Benetti, M. Wodzinski, N. Gunnarsson, J. Sjölund, H. Qiu, Z. Li, C. Großbröhmer, A. Hoopes, I. Reinertsen, Y. Xiao, B. Landman, Y. Huo, K. Murphy, N. Lessmann, B. van Ginneken, A. V. Dalca, and M. P. Heinrich, « Learn2Reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning », *in*: *arXiv preprint* (2021), pp. 1–15, DOI: `10.48550/arXiv.2112.04489`.

[89] A. Hering, S. Kuckertz, S. Heldmann, and M. P. Heinrich, « Enhancing Label-Driven Deep Deformable Image Registration with Local Distance Metrics for State-of-the-Art Cardiac Motion Tracking », *in*: *Informatik aktuell* (2019), pp. 309–314, DOI: `10.1007/978-3-658-25326-4_69`.

[90] A. Hering, K. Murphy, and B. van Ginneken, *Lean2Reg Challenge: CT Lung Registration : CT Lung Registration - Training Data*, 2020, URL: `https://zenodo.org/record/3835682#.X2Nklos69hE` (visited on 09/19/2020).

[91] D. L. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, « Medical image registration », *in*: *Medical Image Registration* 46.*3* (2001), DOI: `10.1051/epn:2000401`.

[92] G. N. Hortobagyi, J. L. Connolly, C. J. D'Orsi, S. B. Edge, E. A. Mittendorf, H. S. Rugo, L. J. Solin, D. L. Weaver, D. J. Winchester, and A. Giuliano, *AJCC Cancer Staging Manual, Eighth Edition: Breast*, vol. 39, 2017, pp. 1355–1359, DOI: `10.1007/s00268-005-0585-9`.

[93] I. Hristova, R. Boellaard, W. Vogel, F. Mottaghy, S. Marreaud, S. Collette, P. Schöffski, R. Sanfilippo, R. Dewji, W. van der Graaf, and W. J. G. Oyen, « Retrospective quality control review of FDG scans in the imaging sub-study of PALETTE EORTC 62072/VEG110727: a randomized, double-blind, placebo-controlled phase III trial », *in*: *European Journal of Nuclear Medicine and Molecular Imaging* 42.*6* (2015), pp. 848–857, DOI: `10.1007/s00259-015-3002-0`.

[94] Y. Hu, M. Modat, E. Gibson, W. Li, N. Ghavami, E. Bonmati, G. Wang, S. Bandula, C. M Moore, M. Emberton, S. Ourselin, J. A. Noble, D. C. Barratt, and T. Vercauteren, « Weakly-supervised convolutional neural networks for multimodal image registration. », *in*: *Medical image analysis* 49 (2018), pp. 1–13, DOI: `10.1016/j.media.2018.07.002`.

[95] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, « Densely Connected Convolutional Networks », *in*: *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 4700–4708.

[96] INCa, *Quelques chiffres*, 2019, URL: `https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Quelques-chiffres` (visited on 07/30/2019).

[97] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, and K. H. Maier-Hein, « nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation », *in*: *Informatik aktuell* (2019), p. 22, DOI: `10.1007/978-3-658-25326-4_7`.

[98] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, « Spatial Transformer Networks », *in*: *Advances in neural information processing systems* 28 (2015), pp. 2017–2025, DOI: `10.5555/2969442.2969465`.

[99] M. Jenkinson and S. Smith, « A global optimisation method for robust affine registration of brain images », *in*: *Medical Image Analysis* 5.*2* (2001), pp. 143–156, DOI: `10.1016/S1361-8415(01)00036-6`.

[100] H. J. Johnson and G. E. Christensen, « Consistent landmark and intensity-based image registration », *in*: *IEEE Transactions on Medical Imaging* 21.*5* (2002), pp. 450–461, DOI: `10.1109/TMI.2002.1009381`.

[101] D. K. Jones, M. A. Horsfield, and A. Simmons, « Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging », *in*: *Magnetic Resonance in Medicine* 42.*3* (1999), pp. 515–525, DOI: `10.1002/(SICI)1522-2594(199909)42:3<515::AID-MRM14>3.0.CO;2-Q`.

[102] O. Joo Hyun, M. A. Lodge, and R. L. Wahl, « Practical percist: A simplified guide to PET response criteria in solid tumors 1.0 », *in*: *Radiology* 280.*2* (2016), pp. 576–584, DOI: `10.1148/radiol.2016142043`.

[103] L. Juan and O. Gwun, « A comparison of sift, pca-sift and surf », *in*: *International Journal of Image Processing (IJIP)* 3.*4* (2007), pp. 143–152, DOI: `10.1080/00420980020014884`.

[104] S. Kabus, T. Netsch, B. Fischer, and J. Modersitzki, « B-spline registration of 3D images with Levenberg-Marquardt optimization », *in*: *Medical Imaging 2004: Image Processing* 5370 (2004), p. 304, DOI: `10.1117/12.533976`.

[105] W. A. Kalender, *Computed tomography: fundamentals, system technology, image quality, applications*, John Wiley & Sons, 2011.

[106] V. Kapoor, B. M. McCook, and F. S. Torok, « An introduction to PET-CT imaging », *in*: *Radiographics* 24.*2* (2004), pp. 523–543, DOI: `10.1148/rg.242025724`.

[107] T. D. Kemp, C. M. J. de Bakker, L. Gabel, D. A. Hanley, E. O. Billington, L. A. Burt, and S. K. Boyd, « Longitudinal bone microarchitectural changes are best detected using image registration », *in*: *Osteoporosis International* 31.*10* (2020), pp. 1995–2005, DOI: `10.1007/s00198-020-05449-2`.

[108] B. Kim, J. Kim, June G. L., D. H. Kim, S. H. Park, and J. C. Ye, « Unsupervised Deformable Image Registration Using Cycle-Consistent CNN », *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11769 LNCS (2019), pp. 166–174, DOI: `10.1007/978-3-030-32226-7_19`.

[109] C. K. Kim, N. C. Gupta, B. Chandramouli, and A. Alavi, « Standardized uptake values of FDG: Body surface area correction is preferable to body weight correction », *in*: *Journal of Nuclear Medicine* 35.*1* (1994), pp. 164–167.

[110] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M. C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier, J. H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R. P. Woods, J. J. Mann, and R. V. Parsey, « Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration », *in*: *NeuroImage* 46.*3* (2009), pp. 786–802, DOI: `10.1016/j.neuroimage.2008.12.037`.

[111] S. Klein, J. P. W. Pluim, M. Staring, and M. A. Viergever, « Adaptive stochastic gradient descent optimisation for image registration », *in*: *International Journal of Computer Vision* 81.*3* (2009), pp. 227–239, DOI: `10.1007/s11263-008-0168-y`.

[112] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, « elastix: A Toolbox for Intensity-Based Medical Image Registration », *in*: *IEEE Transactions on Medical Imaging* 29.*1* (2010), pp. 196–205, DOI: `10.1109/TMI.2009.2035616`.

[113] S. Klein, M. Staring, and J. P. W. Pluim, « Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines », *in*: *IEEE Transactions on Image Processing* 16.*12* (2007), pp. 2879–2890, DOI: `10.1109/TIP.2007.909412`.

[114] N. C. Krak, R. Boellaard, Otto S. Hoekstra, J. W. R. Twisk, C. J. Hoekstra, and Adriaan A. Lammertsma, « Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial », *in*: *European Journal of Nuclear Medicine and Molecular Imaging* 32.*3* (2005), pp. 294–301, DOI: `10.1007/s00259-004-1566-1`.

[115] A. Krizhevsky, I. Sutskever, and G. E. Hinton, « ImageNet Classification with Deep Convolutional Neural Networks », *in*: *Advances in Neural Information Processing Systems 25 (NIPS 2012)* 25 (2012), pp. 45–1–45–16, DOI: `10.1201/9781420010749`.

[116] Z. Lai and W. Xie, « Self-supervised Learning for Video Correspondence Flow », *in*: *arXiv preprint* (2019), pp. 1–14, DOI: `10.48550/arXiv.1905.00875`.

[117] K. J. Langen, U. Braun, E. R. Kops, H. Herzog, T. Kuwert, B. Nebeling, and L. E. Feinendegen, « The influence of plasma glucose levels on fluorine-18-fluorodeoxyglucose uptake in bronchial carcinomas », *in*: *Journal of Nuclear Medicine* 34.*3* (1993), pp. 355–359.

[118] Yiu H. Lau, Michael Braun, and Brian F. Hutton, « Non-rigid image registration using a median-filtered coarse-to-fine displacement field and a symmetric correlation ratio », *in*: *Physics in Medicine and Biology* 46.*4* (2001), pp. 1297–1319, DOI: `10.1088/0031-9155/46/4/326`.

[119] M. H. Laves, S. Ihler, and T. Ortmaier, « Deformable medical image registration using a randomly-initialized CNN as regularization prior », *in*: *arXiv preprint 1* (2019), pp. 1–6, DOI: `10.48550/arXiv.1908.00788`.

[120] Y. LeCun, Y. Bengio, and G. E. Hinton, « Deep learning », *in*: *Nature* 521.*7553* (2015), pp. 436–444.

[121] D. Lee, M. Hofmann, F. Steinke, Y. Altun, N. D. Cahill, and B. Schölkopf, « Learning similarity measure for multi-modal 3D image registration », *in*: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (2009), pp. 186–193, DOI: `10.1109/CVPRW.2009.5206840`.

[122] B. Li, W. J. Niessen, S. Klein, M. de Groot, M. A. Ikram, M. W. Vernooij, and E. E. Bron, « A Hybrid Deep Learning Framework for Integrated Segmentation and Registration: Evaluation on Longitudinal White Matter Tract Changes », *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11766 LNCS (2019), pp. 645–653, DOI: `10.1007/978-3-030-32248-9_72`.

[123] H. Li and Y. Fan, « Non-rigid image registration using self-supervised fully convolutional networks without training data », *in*: *Proceedings - International Symposium on Biomedical Imaging* 2018-April (2018), pp. 1075–1078, DOI: `10.1109/ISBI.2018.8363757`.

[124] X. A. Li, A. Tai, D. W. Arthur, T. A. Buchholz, S. MacDonald, L. B. M., J. M. Moran, L. J. Pierce, R. Rabinovitch, A. Taghian, F. Vicini, W. Woodward, and J. R. White, « Variability of target and normal structure delineation for breast-cancer radiotherapy: a RTOG multi-institutional and multi- observer study », *in*: *International Journal of Radiation Oncology* 73.*3* (2009), pp. 944–951, DOI: `10.1038/jid.2010.15.Antagonistic`.

[125] L. I. Lin, « A Concordance Correlation Coefficient to Evaluate Reproducibility », *in*: *Biomatrics* 45.*1* (1989), pp. 255–268.

[126] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, « A survey on deep learning in medical image analysis. », *in*: *Medical Image Analysis* 42.*December 2017* (2017), pp. 60–88, DOI: `10.1016/j.media.2017.07.005`.

[127] F. Liu, D. Liu, J. Tian, X. Xie, X. Yang, and K. Wang, « Cascaded one-shot deformable convolutional neural networks: Developing a deep learning model for respiratory motion estimation in ultrasound sequences », *in*: *Medical Image Analysis* 65 (2020), DOI: `10.1016/j.media.2020.101793`.

[128] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens, « Nonrigid image registration using conditional mutual information », *in*: *IEEE transactions on medical imagingIEEE transactions on medical imaging* 29.*1* (2010), pp. 19–29.

[129] M. Lorenzi, N. Ayache, G. Frisoni, and X. Pennec, « LCC-Demons : a robust and accurate symmetric diffeomorphic registration algorithm To cite this version : HAL Id : hal-00819895 », *in*: (2014).

[130] X. Lu, D. Xu, and D. Liu, « Robust 3D Organ Localization with Dual Learning Architectures and Fusion Xiaoguang », *in*: *Deep Learning and Data Labeling for Medical Applications* (2016), pp. 12–20, DOI: `10.1007/978-3-319-46976-8`.

[131] A. Lucas, M. Iliadis, and R. Molina, « Using Deep Neural Networks for Inverse Problems in Imaging: Beyond Analytical Methods », *in*: *IEEE Signal Processing Magazine* 35.*1* (2018), pp. 20–36, DOI: `10.1109/msp.2017.2760358`.

[132] G. Lucignani, « SUV and segmentation: Pressing challenges in tumour assessment and treatment », *in*: *European Journal of Nuclear Medicine and Molecular Imaging* 36.*4* (2009), pp. 715–720, DOI: `10.1007/s00259-009-1085-1`.

[133] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng, « Rectifier nonlinearities improve neural network acoustic models », *in*: *ICML Workshop on Deep Learning for Audio, Speech and Language Processing* 28 (2013), DOI: `10.1.1.693.1422`.

[134] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, « Multimodality image registration by maximization of quantitative-qualitative measure of mutual information », *in*: *IEEE transactions on Medical Imaging* 16.*2* (1997), pp. 197–198, DOI: `10.1016/j.patcog.2007.04.002`.

[135] A. M. Maffione, A. Ferretti, S. Chondrogiannis, L. Rampin, M. C. Marzola, G. Grassetto, C. Capirci, P. M. Colletti, and D. Rubello, « Proposal of a new 18F-FDG PET/CT predictor of response in rectal cancer treated by neoadjuvant chemoradiation therapy and comparison with percist criteria », *in*: *Clinical Nuclear Medicine* 38.*10* (2013), pp. 795–797, DOI: `10.1097/RLU.0b013e3182a20153`.

[136] D. Mahapatra, Z. Ge, S. Sedai, and R. Chakravorty, « Joint Registration And Segmentation Of Xray Images Using Generative Adversarial Networks », *in*: *International Workshop on Machine Learning in Medical Imaging* 1 (2018), pp. 73–80.

[137] A. J. B. Maintz and M. A. Viergever, « A survey of medical image registration », *in*: *Medical Image Analysis* 2.*1* (1998), pp. 1–36, DOI: `10.1016/S1361-8415(01)80026-8`.

[138] A.-M. -M Mandard, F. Dalibard, J.-C. Mandard, J. Marnay, M. Henry-Amar, J.-F. Petiot, A. Roussel, J.-H. Jacob, P. Segol, G. Samama, J.-M. Ollivier, S. Bonvalot, and M. Gignoux, « Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations », *in*: *Cancer* 73.*11* (1994), pp. 2680–2686, DOI: `10.1002/1097-0142(19940601)73`.

[139] C. R. Maurer and J. M. Fitzpatrick, « A Review of Medical Image Registration », *in*: *Interactive Image-Guided Neurosurgery* (1993), pp. 17–44.

[140] Y. Menu, « Imagerie en cancérologie. Au-delà des critères RECIST », *in*: *Revue des Maladies Respiratoires Actualités* 5.*5* (2013), pp. 361–366.

[141] F. Michel, M. Bronstein, A. Bronstein, and N. Paragios, « Boosted metric learning for 3D multi-modal deformable registration », *in*: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro 1* (2011), pp. 1209–1214.

[142] A. B. Miller, B. Hoogstraten, M. Staquet, and A. Winkler, « Reporting results of cancer treatment », *in*: *Cancer* 47.*1* (1981), pp. 207–214, DOI: `10.1002/1097-0142(19810101)47:1<207::AID-CNCR2820470134>3.0.CO;2-6`.

[143] F. Milletari, N. Navab, and S. A. Ahmadi, « V-Net: Fully convolutional neural networks for volumetric medical image segmentation », *in*: *Int. Conf. on 3D Vision* (2016), pp. 565–571, DOI: `10.1109/3DV.2016.79`.

[144] S. J. Min, H. J. Jang, and J. H. Kim, « Comparison of the EORTC criteria and PERCIST in solid tumors: A pooled analysis and review », *in*: *Oncotarget* 7.*36* (2016), pp. 58105–58110, DOI: `10.18632/oncotarget.11171`.

[145] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, « Image Segmentation Using Deep Learning: A Survey », *in*: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8828.*c* (2021), pp. 1–20, DOI: `10.1109/TPAMI.2021.3059968`.

[146] M. Modat, P. Daga, M. J. Cardoso, S. Ourselin, G. R. Ridgway, and J. Ashburner, « Parametric non-rigid registration using a stationary velocity field », *in*: *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis* (2012), pp. 145–150, DOI: `10.1109/MMBIA.2012.6164745`.

[147] N. Mohd Razali and Y. Bee Wah, « Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests », *in*: *Journal of Statistical Modeling and Analytics* 2.*November* (2011), pp. 21–33.

[148] T. C. W. Mok and A. C. S. Chung, « Conditional Deformable Image Registration with Convolutional Neural Network », *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12904 LNCS (2021), pp. 35–45, DOI: `10.1007/978-3-030-87202-1_4`.

[149] T. C. W. Mok and A. C. S. Chung, « Large Deformation Diffeomorphic Image Registration with Laplacian Pyramid Networks », *in*: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (2020), pp. 211–221, DOI: `10.1007/978-3-030-59716-0_21`.

[150] N. Moreau, C. Rousseau, **C. Fourcade**, G. Santini, A. Brennan, L. Ferrer, M. Lacombe, C. Guillerminet, M. Colombié, P. Jézéquel, M. Campone, N. Normand, and M. Rubeaux, « Automatic segmentation of metastatic breast cancer lesions on ^{18}F-FDG PET/CT longitudinal acquisitions for treatment response assessment », *in*: *Cancers* 14.*1* (2022), DOI: `10.3390/cancers14010101`.

[151] N. Moreau, C. Rousseau, **C. Fourcade**, G. Santini, L. Ferrer, M. Lacombe, C. Guillerminet, M. Campone, M. Colombie, M. Rubeaux, and N. Normand, « Deep learning approaches for bone and bone lesion segmentation on 18FDG PET/CT imaging in the context of metastatic breast cancer », *in*: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology*

Society, EMBS 2020-July (2020), pp. 1532–1535, DOI: `10.1109/EMBC44109.2020.9175904`.

[152] N. Moreau, C. Rousseau, **C. Fourcade**, G. Santini, L. Ferrer, M. Lacombe, C. Guillerminet, M. Campone, M. Colombié, M. Rubeaux, and N. Normand, « Deep learning approaches for bone and bone lesion segmentation on 18FDG PET/CT imaging in the context of metastatic breast cancer », *in*: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* 2020-July (2020), pp. 1532–1535, DOI: `10.1109/EMBC44109.2020.9175904`.

[153] N. Moreau, C. Rousseau, **C. Fourcade**, G. Santini, L. Ferrer, M. Lacombe, C. Guillerminet, P. Jézéquel, M. Campone, N. Normand, and M. Rubeaux, « Comparison between threshold-based and deep learning-based bone segmentation on whole-body CT images », *in*: *Medical Imaging 2021: Computer-Aided Diagnosis* 11597 (2021), DOI: `10.1117/12.2580892`.

[154] D. J. Morgan and K. M. Bray, « Lean body mass as a predictor of drug dosage », *in*: *Clinical pharmacokinetics* 4.*26* (1994), pp. 292–307.

[155] K. Murphy, B. van Ginneken, S. Klein, M. Staring, B. J. de Hoop, M. A. Viergever, and J. P.W. Pluim, « Semi-automatic construction of reference standards for evaluation of image registration », *in*: *Medical Image Analysis* 15.*1* (2011), pp. 71–84, DOI: `10.1016/j.media.2010.07.005`.

[156] A. Myronenko and X. Song, « Adaptive Regularization of Ill-Posed Problems: Application to Non-rigid Image Registration », *in*: *arXiv preprint* (2009), pp. 1–10, DOI: `10.48550/arXiv.0906.3323`.

[157] J. Nagi, F. Ducatelle, G. A. Di Caro, D. Cireşan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, and L. M. Gambardella, « Max-pooling convolutional neural networks for vision-based hand gesture recognition », *in*: *2011 IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2011* (2011), pp. 342–347, DOI: `10.1109/ICSIPA.2011.6144164`.

[158] National Comprehensive Cancer Network, « Noninvasive Breast cancer », *in*: *Guidelines for patients* (2018), DOI: `10.1016/S0140-6736(94)90616-5`.

[159] S. A. Nehmeh, Y. E. Erdi, T. Pan, A. Pevsner, K. E. Rosenzweig, E. Yorke, G. S. Mageras, H. Schoder, P. Vernon, O. Squire, H. Mostafavi, S. M. Larson, and J. L. Humm, « Four-dimensional (4D) PET/CT imaging of the thorax », *in*: *Medical Physics* 31.*12* (2004), pp. 3179–3186, DOI: `10.1118/1.1809778`.

[160] National Comprehensive Cancer Network, « Brest Cancer Metastatic », *in*: *National Comprehensive Cancer Network* (2018).

[161] National Comprehensive Cancer Network, « Invasive breast cancer », *in*: *National Comprehensive Cancer Network* (2018).

[162] A. Nicolini, P. Ferrari, and M. J. Duffy, « Prognostic and predictive biomarkers in breast cancer: Past, present and future », *in*: *Seminars in Cancer Biology* 52.*September 2017* (2018), pp. 56–73, DOI: `10.1016/j.semcancer.2017.08.010`.

[163] S. W. Oh, J. Y. Lee, K. Sunkavalli, and S. J. Kim, « Fast Video Object Segmentation by Reference-Guided Mask Propagation », *in*: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (2018), pp. 7376–7385, DOI: `10.1109/CVPR.2018.00770`.

[164] F. P. M. Oliveira and J. M. R. S. Tavares, *Medical image registration: A review*, 2014, DOI: `10.1080/10255842.2012.670855`.

[165] M. M. Osman, C. Cohade, Y. Nakamoto, and R. L. Wahl, « Respiratory motion artifacts on PET emission images obtained using CT attenuation correction on PET-CT », *in*: *European Journal of Nuclear Medicine and Molecular Imaging* 30.*4* (2003), pp. 603–606, DOI: `10.1007/s00259-002-1024-x`.

[166] C. Paganelli, G. Meschini, S. Molinelli, M. Riboldi, and G. Baroni, « Patient-specific validation of deformable image registration in radiation therapy: Overview and caveats », *in*: *Medical Physics* 45.*10* (2018), e908–e922, DOI: `10.1002/mp.13162`.

[167] M. Papadomanolaki, M. Vakalopoulou, and K. Karantzalos, « A Novel Object-Based Deep Learning Framework for Semantic Segmentation of Very High-Resolution Remote Sensing Data: Comparison with Convolutional and Fully Convolutional Networks », *in*: *Remote Sensing* 11.*6* (2019), p. 684, DOI: `10.3390/rs11060684`.

[168] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, « Automatic differentiation in PyTorch », *in*: *31st Conference on Neural Information Processing System* (2017).

[169] X. Pennec, R. Stefanescu, V. Arsigny, P. Fillard, and N. Ayache, « Riemannian elasticity: A statistical regularization framework for non-linear registration », *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 3750 LNCS (2005), pp. 943–950, DOI: `10.1007/11566489_116`.

[170] R. D. Phair, « Development of kinetic models in the nonlinear world of molecular cell biology », *in*: *Metabolism: Clinical and Experimental* 46.*12* (1997), pp. 1489–1495, DOI: `10.1016/S0026-0495(97)90154-2`.

[171] J. P. W. Pluim, A. J. B. Maintz, and M. A. Viergever, « Mutual-information-based registration of medical images: A survey », *in*: *IEEE Transactions on Medical Imaging* 22.*8* (2003), pp. 986–1004, DOI: `10.1109/TMI.2003.815867`.

[172] A. J. Reader and H. Zaidi, « Advances in PET Image Reconstruction », *in*: *PET Clinics* 2.*2* (2007), pp. 173–190, DOI: `10.1016/j.cpet.2007.08.001`.

[173] D. Rey, G. Subsol, H. Delingette, and N. Ayache, « Automatic detection and segmentation of evolving processes in 3D medical images: Application to multiple sclerosis », *in*: *Medical Image Analysis* 6.*2* (2002), pp. 163–179, DOI: `10.1016/S1361-8415(02)00056-7`.

[174] M. M. Rohé, M. Datar, T. Heimann, M. Sermesant, and X. Pennec, « SVF-Net: learning deformable image registration using shape matching », *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10433 LNCS (2017), pp. 266–274, DOI: `10.1007/978-3-319-66182-7_31`.

[175] T. Rohlfing, C. R. Maurer, D. A. Bluemke, and M. A. Jacobs, « Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint », *in*: *IEEE Transactions on Medical Imaging* 22.*6* (2003), pp. 730–741, DOI: `10.1109/TMI.2003.814791`.

[176] K. Rohr, H. S. Stiehl, R. Sprengel, T. M. Buzug, J. Weese, and M. H. Kuhn, « Landmark-based elastic registration using approximating thin-plate splines », *in*: *IEEE Transactions on Medical Imaging* 20.*6* (2001), pp. 526–534, DOI: `10.1109/42.929618`.

[177] O. Ronneberger, P. Fischer, and T. Brox, « U-net: Convolutional networks for biomedical image segmentation », *in*: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* 9351 (2015), pp. 234–241, DOI: `10.1007/978-3-319-24574-4_28`.

[178] F. Rousseau, L. Drumetz, and R. Fablet, « Residual Networks as Flows of Diffeomorphisms », *in*: *Journal of Mathematical Imaging and Vision* 62.*3* (2020), pp. 365–375, DOI: `10.1007/s10851-019-00890-3`.

[179] D. Rueckert, P. Aljabar, R. A. Heckemann, J. V. Hajnal, and A. Hammers, « Diffeomorphic registration using B-splines », *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 4191 LNCS - II (2006), pp. 702–709, DOI: `10.1007/11866763_86`.

[180] D. Rueckert, A. F. Frangi, and J. A. Schnabel, « Automatic construction of 3D statistical deformation models using non-rigid registration », *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 2208.*8* (2003), pp. 77–84, DOI: `10.1007/3-540-45468-3_10`.

[181] D. Rueckert and J. A Schnabel, *Medical image registration*, Springer, 2010, pp. 131–154.

[182] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, « Nonrigid registration using free-form deformations: Application to breast mr images », *in*: *IEEE Transactions on Medical Imaging* 18.*8* (1999), pp. 712–721, DOI: `10.1109/42.796284`.

[183] G. Santini, **C. Fourcade**, N. Moreau, C. Rousseau, L. Ferrer, M. Lacombe, V. Fleury, M. Campone, P. Jézequel, and M. Rubeaux, « Unpaired PET/CT image synthesis of liver region using CycleGAN », *in*: *International Symposium on Medical Information Processing and Analysis* 11583 (2020), DOI: `10.1117/12.2576095`.

[184] G. Santini, N. Moreau, **C. Fourcade**, C. Rousseau, L. Ferrer, M. Campone, M. Colombié, P. Jézéquel, and M. Rubeaux, « Quantification automatique de l'activité de fond pour le calcul du critère PERCIST », *in*: *Médecine Nucléaire* 45.*4* (2021), pp. 213–214, DOI: `10.1016/j.mednuc.2021.06.080`.

[185] J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit, F. A. Gerritsen, D. L. G. Hill, and D. J. Hawkes, « A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations », *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 2208 (2001), pp. 573–581, DOI: `10.1007/3-540-45468-3_69`.

[186] M. Schuster and K. K. Paliwal, « Bidirectional recurrent neural networks », *in*: *IEEE Transactions on Signal Processing* 45.*11* (1997), pp. 2673–2681, DOI: `10.1109/78.650093`.

[187] L. H. Schwartz, J. Bogaerts, R. Ford, L. Shankar, P. Therasse, S. Gwyther, and E. A. Eisenhauer, « Evaluation of lymph nodes with RECIST 1.1 », *in*: *European Journal of Cancer* 45.*2* (2009), pp. 261–267, DOI: `10.1016/j.ejca.2008.10.028`.

[188] L. H. Schwartz, S. Litière, E. De Vries, R. Ford, S. Gwyther, S. Mandrekar, L. Shankar, J. Bogaerts, A. Chen, J. Dancey, W. Hayes, F. S. Hodi, O. S. Hoekstra, E. P. Huang, N. Lin, Y. Liu, P. Therasse, J. D. Wolchok, and L. Seymour, « RECIST 1.1 - Update and clarification: From the RECIST committee », *in*: *European Journal of Cancer* 62.*March* (2016), pp. 132–137, DOI: `10.1016/j.ejca.2016.03.081`.

[189] J. Shang, X. Ling, L. Zhang, Y. Tang, Z. Xiao, Y. Cheng, B. Guo, J. Gong, L. Huang, and H. Xu, « Comparison of RECIST, EORTC criteria and PERCIST for evaluation of early response to chemotherapy in patients with non-small-cell lung cancer », *in*: *European Journal of Nuclear Medicine and Molecular Imaging* 43.*11* (2016), pp. 1945–1953, DOI: `10.1007/s00259-016-3420-7`.

[190] F. Shi, Y. Fan, S. Tang, J. H. Gilmore, W. Lin, and D. Shen, « Neonatal brain image segmentation in longitudinal MRI studies », *in*: *Neuroimage* 49.*1* (2010), pp. 391–400, DOI: `10.1016/j.neuroimage.2009.07.066.Neonatal`.

[191] H. Siebert, L. Hansen, and M. P. Heinrich, « Fast 3D registration with accurate optimisation and little learning for Learn2Reg 2021 », *in*: *Biomedical Image Registration, Domain Generalisation and Out-of-Distribution Analysis* (2021), pp. 1–6, DOI: `10.1007/978-3-030-97281-3_25`.

[192] M. Simonovsky, B. Gutierrez-Becker, D. Mateus, N. Navab, and N. Komodakis, « A Deep Metric for Multimodal Registration », *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 1 (2016), p. 697, DOI: `10.1007/978-3-319-46726-9`.

[193] K. Simonyan and A. Zisserman, « Very deep convolutional networks for large-scale image recognition », *in*: *arXiv preprint* (2014), DOI: `10.2146/ajhp170251`.

[194] K. Skougaard, D. Nielsen, B. V. Jensen, and H. W. Hendel, « Comparison of EORTC criteria and PERCIST for PET/CT response evaluation of patients with metastatic colorectal cancer treated with irinotecan and cetuximab », *in*: *Journal of Nuclear Medicine* 54.*7* (2013), pp. 1026–1031, DOI: `10.2967/jnumed.112.111757`.

[195] T. A. D. Smith, « Facilitative glucose transporter expression in human cancer tissue.pdf », *in*: *British journal of biomedical science* 56.*4* (1999), p. 285.

[196] R. W. K. So, T. W. H. Tang, and A. C. S. Chung, « Non-rigid image registration of brain magnetic resonance images using graph-cuts », *in*: *Pattern Recognition* 44.*10-11* (2011), pp. 2450–2467, DOI: `10.1016/j.patcog.2011.04.008`.

[197] H. Sokooti, B. D. de Vos, F. Berendsen, B. P. F. Lelieveldt, I. Išgum, and M. Staring, « Nonrigid image registration using multi-scale 3D convolutional neural networks », *in*: *International conference on medical image computing and computer-assisted intervention (MICCAI)* 2 (2017), pp. 108–115, DOI: `10.1007/978-3-319-66182-7`.

[198] M. Soret, S. L. Bacharach, and I. Buvat, « Partial-volume effect in PET tumor imaging », *in*: *Journal of Nuclear Medicine* 48.*6* (2007), pp. 932–945, DOI: `10.2967/jnumed.106.035774`.

[199] C. Ó. S. Sorzano, P. Thévenaz, and M. Unser, « Elastic registration of biological images using vector-spline regularization », *in*: *IEEE Transactions on Biomedical Engineering* 52.*4* (2005), pp. 652–663, DOI: `10.1109/TBME.2005.844030`.

[200] A. Sotiras, C. Davatzikos, and N. Paragios, « Deformable medical image registration: A survey », *in*: *IEEE Transactions on Medical Imaging* 32.*7* (2013), pp. 1153–1190, DOI: `10.1109/TMI.2013.2265603`.

[201] S. M. Srinivas, T. Dhurairaj, S. Basu, G. Bural, S. Surti, and A. Alavi, « A recovery coefficient method for partial volume correction of PET images », *in*: *Annals of Nuclear Medicine* 23.*4* (2009), pp. 341–348, DOI: `10.1007/s12149-009-0241-9`.

[202] M. Staring, S. Klein, M. A. Viergever, J. P. W. Pluim, and U. A. van der Heide, « Registration of Cervical MRI Using Multifeature Mutual Information », *in*: *IEEE Transactions on Medical Imaging* 28.*9* (2009), pp. 1412–1421, DOI: `10.1109/TMI.2009.2016560`.

[203] Y. Sugawara, K. R. Zasadny, A. W. Neuhoff, and R. L. Wahl, « Reevaluation of the standardized uptake value for FDG: Variations with body weight and methods for correction », *in*: *Radiology* 213.*2* (1999), pp. 521–525, DOI: `10.1148/radiology.213.2.r99nv37521`.

[204] C. Suzuki, M. R. Torkzad, H. Jacobsson, G. Åström, A. Sundin, T. Hatschek, H. Fujii, and L. Blomqvist, « Interobserver and intraobserver variability in the response evaluation of cancer therapy according to RECIST and WHO-criteria », *in*: *Acta Oncologica* 49.*4* (2010), pp. 509–514, DOI: `10.3109/02841861003705794`.

[205] G. Tarantola, F. Zito, and P. Gerundini, « PET Instrumentation and Reconstruction Algorithms in Whole-Body Applications Continuing Education », *in*: *Journal of Nuclear Medicine* 44.*5* (2003), pp. 756–769.

[206] P. Therasse, J. Verweij, M. Van Glabbeke, A. T. Van Oosterom, S. G. Arbuck, R. S. Kaplan, L. Rubinstein, M. C. Christian, E. A. Eisenhauer, J. Wanders, and S. G. Gwyther, « New guidelines to evaluate the response to treatment in solid tumors », *in*: *Journal of the National Cancer Institute* 92.*3* (2000), pp. 205–216, DOI: `10.1093/jnci/92.3.205`.

[207] J. P. Thirion, « Image matching as a diffusion process: An analogy with Maxwell's demons », *in*: *Medical Image Analysis* 2.*3* (1998), pp. 243–260, DOI: `10.1016/S1361-8415(98)80022-4`.

[208] N. J. Tustison, B. B. Avants, and J. C. Gee, « Directly manipulated free-form deformation image registration », *in*: *IEEE Transactions on Image Processing* 18.*3* (2009), pp. 624–635, DOI: `10.1109/TIP.2008.2010072`.

[209] N. J. Tustison, B. B. Avants, and J. C. Gee, « Learning image-based spatial transformations via convolutional neural networks: A review », *in*: *Magnetic Resonance Imaging* 64.*May* (2019), pp. 142–153, DOI: `10.1016/j.mri.2019.05.037`.

[210] P. Tylski, « Caractérisation des tumeurs et de leur évolution en TEP / TDM à la 18F-FDG », PhD thesis, Université de Paris Sud XI, 2009.

[211] D. Ulyanov, A. Vedaldi, and V. Lempitsky, « Deep Image Prior », *in*: *International Journal of Computer Vision* 128.*7* (2020), pp. 1867–1888, DOI: `10.1007/s11263-020-01303-4`.

[212] H. Uzunova, M. Wilms, H. Handels, and J. Ehrhardt, « Training CNNs for image registration from few samples with model-based data augmentation », *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10433 LNCS (2017), pp. 223–231, DOI: `10.1007/978-3-319-66182-7_26`.

[213] P. E. Valk, D. Delbeke, D. L. Bailey, D. W. Townsend, and M. N. Maisey, « Positron emission tomography in clinical practice. », *in*: *The Journal of the Florida Medical Association* (2006), DOI: `10.1148/radiol.2442072524`.

[214] P. A. Van den Elsen, E. J. Pol, and M. A. Viergever, « Medical Image Matching—A Review with Classification », *in*: *IEEE Engineering in Medicine and Biology Magazine* 12.*1* (1993), pp. 26–39, DOI: `10.1109/51.195938`.

[215] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, « Diffeomorphic demons : Efficient non-parametric image registration », *in*: *NeuroImage* 45.*1* (2009).

[216] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, « Symmetric log-domain diffeomorphic registration: A demons-based approach », *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5241 LNCS.*PART 1* (2008), pp. 754–761, DOI: `10.1007/978-3-540-85988-8_90`.

[217] B. L. Viglianti, D. J. Wale, K. K. Wong, T. D. Johnson, C. Ky, K. A. Frey, and M. D. Gross, « Effects of Tumor Burden on Reference Tissue Standardized Uptake for PET Imaging: Modification of PERCIST Criteria », *in*: *Radiology* 287.*3* (2018), pp. 993–1002, DOI: `10.1148/radiol.2018171356`.

[218] P. Viola and W. M. Wells III, « Alignment by maximization of mutual information », *in*: *International journal of computer vision* 24.*2* (1997), pp. 137–154, DOI: `10.1096/fasebj.26.1_supplement.917.3`.

[219] P. Visvikis, C. Cheze Le Rest, and F. Lamare, « Le mouvement respiratoire en Imagerie Fonctionnelle du Cancer: une revue des effets et des méthodes de correction », *in*: *Traitement du Signal* 23 (2006), pp. 351–362.

[220] M. Vogsen, J. L. Bülow, L. Ljungstrøm, H. R. Oltmann, T. A. Alamdari, M. Naghavi-Behzad, P. E. Braad, O. Gerke, and M. G. Hildebrandt, « Fdg-pet/ct for response monitoring in metastatic breast cancer: The feasibility and benefits of applying percist », *in*: *Diagnostics* 11.*4* (2021), pp. 1–13, DOI: `10.3390/DIAGNOSTICS11040723`.

[221] B. D. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Išgum, « A deep learning framework for unsupervised affine and deformable image registration », *in*: *Medical Image Analysis* 52 (2019), pp. 128–143, DOI: `10.1016/j.media.2018.11.010`.

[222] B. D. de Vos, J. M. Wolterink, P. A. de Jong, M. A. Viergever, and I. Išgum, « 2D image classification for 3D anatomy localization: employing deep convolutional neural networks », *in*: *Medical Imaging 2016: Image Processing* 9784 (2016), 97841Y, DOI: `10.1117/12.2216971`.

[223] R. L. Wahl, H. Jacene, Y. Kasamon, and M. A. Lodge, « From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors », *in*: *Journal of Nuclear Medicine* 50.*Suppl 1* (2009), pp. 122–150, DOI: `10.2967/jnumed.108.057307.From`.

[224] C. Wang, N. Tyagi, A. Rimner, Y. C. Hu, H. Veeraraghavan, G. Li, M. Hunt, G. Mageras, and P. Zhang, « Segmenting lung tumors on longitudinal imaging studies via a patient-specific adaptive convolutional neural network », *in*: *Radiotherapy and Oncology* 131 (2019), pp. 101–107, DOI: `10.1016/j.radonc.2018.10.037`.

[225] M. Wang and P. Li, « A Review of Deformation Models in Medical Image Registration », *in*: *Journal of Medical and Biological Engineering* 39.*1* (2019), DOI: `10.1007/s40846-018-0390-1`.

[226] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, « DeepFlow: Large displacement optical flow with deep matching », *in*: *Proceedings of the IEEE International Conference on Computer Vision Section 2* (2013), pp. 1385–1392, DOI: `10.1109/ICCV.2013.175`.

[227] WHO, *Breast cancer*, 2019, URL: `https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/` (visited on 07/30/2019).

[228] H. Q. Woodard, R. E. Bigler, and B. Freed, « Expression of tissue isotope distribution. », *in*: *Journal of Nuclear Medicine* 16.*10* (1975), pp. 958–959.

[229] G. Wu, M. Kim, Q. Wang, B. C. Munsell, and D. Shen, « Scalable High Performance Image Registration Framework by Unsupervised Deep Feature Representations Learning », *in*: *IEEE Transactions on Biomedical Engineering* 63.*7* (2015), pp. 1505–1516, DOI: `10.1016/B978-0-12-810408-8.00015-8`.

[230] S. Xu, D. Liu, L. Bao, W. Liu, and P. Zhou, « MHP-VOS: Multiple hypotheses propagation for video object segmentation », *in*: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* 2019-June (2019), pp. 314–323, DOI: `10.1109/CVPR.2019.00040`.

[231] Z. Xu and M. Niethammer, « DeepAtlas: Joint Semi-supervised Learning of Image Registration and Segmentation », *in*: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) Dl* (2019), pp. 420–429, DOI: `10.1007/978-3-030-32245-8_47`.

[232] P. Yan, S. Xu, A. R. Rastinehad, and B. J. Wood, « Adversarial image registration with application for MR and TRUS image fusion », *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11046 LNCS (2018), pp. 197–204, DOI: `10.1007/978-3-030-00919-9_23`.

[233] M. Yanagawa, M. Tatsumi, H. Miyata, E. Morii, N. Tomiyama, T. Watabe, K. Isohashi, H. Kato, E. Shimosegawa, M. Yamasaki, M. Mori, Y. Doki, and J. Hatazawa, « Evaluation of response to neoadjuvant chemotherapy for esophageal cancer: PET response criteria in solid tumors versus response evaluation criteria in solid tumors », *in*: *Journal of Nuclear Medicine* 53.*6* (2012), pp. 872–880, DOI: `10.2967/jnumed.111.098699`.

[234] R. Yao, G. Lin, S. Xia, J. Zhao, and Y. Zhou, « Video Object Segmentation and Tracking: A Survey », *in*: *ACM Transactions on Intelligent Systems and Technology* 1.*1* (2019), DOI: `10.1145/3391743`.

[235] D. Yarotsky, « Error bounds for approximations with deep ReLU networks », *in*: *Neural Networks* 94 (2017), pp. 103–114, DOI: `10.1016/j.neunet.2017.07.002`.

[236] H. Young, R. Baum, U. Cremerius, K. Herholz, O. Hoekstra, A. A. Lammertsma, J. Pruim, and P. Price, « Measurement of clinical and subclinical tumour response using [18F]- fluorodeoxyglucose and positron emission tomography: Review and 1999 EORTC recommendations », *in*: *European Journal of Cancer* 35.*13* (1999), pp. 1773–1782, DOI: `10.1016/S0959-8049(99)00229-4`.

[237] J. Zeintl, A. H. Vija, A. Yahil, J. Hornegger, and T. Kuwert, « Quantitative accuracy of clinical 99mTc SPECT/CT using ordered-subset expectation maximization with 3-dimensional resolution recovery, attenuation, and scatter correction », *in*: *Journal of Nuclear Medicine* 51.*6* (2010), pp. 921–928, DOI: `10.2967/jnumed.109.071571`.

[238] M. Zhang and P. T. Fletcher, « Finite-dimensional lie algebras for fast diffeomorphic image registration », *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9123 (2015), pp. 249–260, DOI: `10.1007/978-3-319-19992-4_19`.

[239] Y. Zhang, L. Yang, H. Zheng, P. Liang, C. Mangold, R. G. Loreto, D. P. Hughes, and D. Z. Chen, « SPDA: Superpixel-based Data Augmentation for Biomedical Image Segmentation », *in*: *Medical Imaging with Deep Learning (MIDL)* (2019), pp. 1–15.

[240] X. Zhuang, K. S. Rhode, R. S. Razavi, D. J. Hawkes, and S. Ourselin, « A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI », *in*: *IEEE Transactions on Medical Imaging* 29.*9* (2010), pp. 1612–1625, DOI: `10.1109/TMI.2010.2047112`.

[241] D. Ziai, T. Wagner, A. El Badaoui, A. Hitzel, J. B. Woillard, B. Melloni, and J. Monteil, « Therapy response evaluation with FDG-ET/CT in small cell lung cancer: A prognostic and comparison study of the PERCIST and EORTC criteria », *in*: *Cancer Imaging* 13.*1* (2013), pp. 73–80, DOI: `10.1102/1470-7330.2013.0008`.

[242] D. Zikic, B. Glocker, O. Kutter, M. Groher, N. Komodakis, A. Kamen, N. Paragios, and N. Navab, « Linear intensity-based image registration by Markov random fields and discrete optimization », *in*: *Medical Image Analysis* 14.*4* (2010), pp. 550–562, DOI: `10.1016/j.media.2010.04.003`.

[243] S. Zincirkeser, E. Şahin, M. Halac, and S. Sager, « Standardized uptake values of normal organs on 18F-fluorodeoxyglucose positron emission tomography and computed tomography imaging », *in*: *Journal of International Medical Research* 35.*2* (2007), pp. 231–236, DOI: `10.1177/147323000703500207`.

[244] B. Zitová and J. Flusser, « Image registration methods: A survey », *in*: *Image and Vision Computing* 21.*11* (2003), pp. 977–1000, DOI: `10.1016/S0262-8856(03)00137-9`.

Titre : Suivi de l'évolution du cancer du sein métastasé via le recalage et la segmentation d'images TEP en utilisant des réseaux entraînés et non-entraînés

Mot clés : Cancer du sein métastasé, Recalage d'images, TEP, Apprentissage profond

Résumé : Le cancer du sein métastasé nécessite un suivi régulier. Au cours du traitement, des images de TEP- scan sont régulièrement acquises puis interprétées selon des recommandations telles que PERCIST pour décider d'un éventuel ajustement thérapeutique. Cependant, PERCIST se concentre seulement sur la lésion présentant l'activité tumorale la plus élevée. L'objectif de cette thèse est de développer des outils permettant de prendre en compte toutes les zones actives à l'aide du TEP-scan, afin de suivre au mieux l'évolution du cancer du sein. Notre première contribution est une méthode pour la segmentation automatique d'organes actifs (cerveau, vessie). Notre deuxième contribution formule la segmentation de lésions sur les images de suivi comme un problème de recalage d'images. Pour résoudre le recalage longitudinal d'images TEP corps entier, nous avons développé une nouvelle méthode nommée MIRRBA (Medical Image Registration Regularized By Architecture), qui combine les avantages des méthodes conventionnelles et de celles utilisant l'apprentissage profond. Nous avons validé trois approches (conventionnelle, apprentissage profond et MIRRBA) sur une base de données privées d'images TEP longitudinales obtenues dans le contexte de l'étude EPICURE. Finalement, notre troisième contribution est l'évaluation de biomarqueurs extraits des segmentations de lésions obtenues grâce au recalage. Nous proposons donc un nouvel outil automatisé pour améliorer suivi du cancer du sein métastasé.

Title: Longitudinal monitoring of metastatic breast cancer through PET image registration and segmentation based on trained and untrained networks

Keywords: Metastatic breast cancer, Image registration, PET, Deep learning

Abstract: Metastatic breast cancer requires constant monitoring. During follow-up care, PET images are regularly acquired and interpreted according to specific guidelines, such as PERCIST, to decide whether or not the treatment should be adapted. However, PERCIST focuses only on one lesion representing tumor burden. The objective of this PhD thesis is to assist physicians monitor metastatic breast cancer patients with longitudinal PET images and improve tumor evaluation by providing them tools to consider all regions showing a high uptake. Our first contribution is a method for the automatic segmentation of active organs (brain, bladder, etc). Our second contribution formulates the segmentation of lesions in the follow-up examination as an image registration problem. The longitudinal full-body PET image registration problem is addressed, in this thesis, with our novel method called MIRRBA (Medical Image Registration Regularized By Architecture), which combines the strengths of both conventional and DL-based approaches within a Deep Image Prior (DIP) setup. We validated the three types of approaches (conventional, DL and MIRRBA) on a private longitudinal PET dataset obtained in the context of the EPICURE project. Finally, the third contribution is the evaluation of the biomarkers extracted from lesion segmentations obtained from the lesion registration step. We propose a new tool for the monitoring of metastatic breast cancer.