

Titre: Synthèse d'images tomodensitométriques à partir d'IRM par des réseaux adverses génératifs pour le recalage 3D/2D de la colonne vertébrale
Title:

Auteur: Reda Oulbacha
Author:

Date: 2019

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Oulbacha, R. (2019). Synthèse d'images tomodensitométriques à partir d'IRM par des réseaux adverses génératifs pour le recalage 3D/2D de la colonne vertébrale [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie.
Citation: <https://publications.polymtl.ca/4162/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/4162/>
PolyPublie URL:

Directeurs de recherche: Samuel Kadoury
Advisors:

Programme: Génie biomédical
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Synthèse d'images tomodensitométriques à partir d'IRM par des réseaux
adverses génératifs pour le recalage 3D/2D de la colonne vertébrale**

REDA OULBACHA

Institut de génie biomédical

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

Génie biomédical

Décembre 2019

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Synthèse d'images tomodensitométriques à partir d'IRM par des réseaux
adverses génératifs pour le recalage 3D/2D de la colonne vertébrale**

présenté par **Reda OULBACHA**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Julien COHEN-ADAD, président

Samuel KADOURY, membre et directeur de recherche

Guillaume-Alexandre BILODEAU, membre

DÉDICACE

À ma famille

REMERCIEMENTS

Je voudrais avant tout remercier mon directeur de recherche, Pr. Samuel Kadoury, pour m'avoir accueilli au sein du laboratoire medICAL, pour m'avoir instruit en termes de démarches de recherche et de rigueur scientifique, et pour m'avoir guidé dans la bonne direction quand j'étais en difficulté.

Je tiens aussi à remercier tous les membres du laboratoire medICAL, tant pour les moments partagés que pour le partage de la connaissance et le travail collaboratif.

Merci Debi pour avoir été à mes côtés, soutenu et encouragé tout le long.

Mon séjour en maîtrise à Polytechnique n'aurait pas été le même sans tous les collègues et tous les amis que je me suis fait, ainsi que tous les moments que nous avons partagé, que ce soit entre les murs de cette école ou en dehors.

Je tiens à remercier l'ensemble de l'équipe des relations internationales du département de génie électrique à l'INSA Lyon pour leur suivi et leur soutien à distance. Je remercie aussi le cadre de Polytechnique pour m'avoir accueilli et offert l'opportunité d'effectuer cette maîtrise.

Enfin et surtout, je tiens à remercier tous les membres de ma famille qui m'ont beaucoup soutenu dans les moments difficiles, mais qui m'ont surtout inconditionnellement encouragé.

RÉSUMÉ

L'information structurelle tridimensionnelle apporte une aide précieuse aux procédures orthopédiques qui, le plus souvent, n'ont à portée de main que des modalités d'imagerie bidimensionnelles pour se guider. Non seulement cela aide-t-il à améliorer la précision des manoeuvres, mais cela permet aussi, dans plusieurs cas, de rendre les procédures moins invasives. La modalité d'imagerie bidimensionnelle sur laquelle se concentre ce mémoire est la fluoroscopie de bras en C. Celle-ci est très répandue dans les salles opératoires et permet une acquisition rapide et versatile pour guider les procédures orthopédiques. Cette modalité est le plus souvent fusionnée avec la tomodensitométrie par un recalage 2D/3D. Ces deux modalités reposent sur le principe d'absorption de rayons X, ce qui fait que l'on retrouve des similarités dans la géométrie et dans les intensités d'une modalité à l'autre. Cela simplifie le problème. Toutefois, la tomodensitométrie n'offre pas les meilleurs contrastes pour visualiser les organes, les nerfs et les tissus mous ; d'autant plus qu'elle cause une irradiation non négligeable au patient. En revanche, l'IRM se prête mieux à la visualisation des organes et des tissus mous. Elle a aussi l'avantage de ne pas irradier le patient. Le contraste de celle-ci est très différent de la tomodensitométrie, surtout au niveau de la colonne vertébrale ; ce qui rend le problème de recalage 2D/3D avec la fluoroscopie de bras en C plus difficile. Les dernières avancées en apprentissage profond montrent des résultats très prometteurs pour la tâche de translation d'image. Ceci est applicable à la génération de tomodensitométrie synthétique à partir d'IRM pour un recalage 2D/3D subséquent.

Nous proposons une méthode de recalage 2D/3D par Digitally Reconstructed Radiograph (DRR) entre l'imagerie par résonance magnétique et la fluoroscopie de bras en C, basée sur la synthèse de tomodensitométrie par des méthodes d'apprentissage profond. En premier lieu, nous explorons plusieurs architectures de réseaux adverses génératifs. Ces architectures-là ont montré d'impressionnants résultats pour la translation d'images non médicales. Nous expérimentons avec un ensemble de données public constitué de 18 volumes d'IRM et de CT. Nous constatons que l'architecture CycleGAN généralise mieux à des données non-observées que l'architecture cGAN, et que celle-ci tend à sur-apprendre. Nous introduisons aussi deux nouvelles composantes à l'architecture CycleGAN pour améliorer la résolution tridimensionnelle ainsi que la distribution des intensités pour les CT synthétiques. Enfin, nous effectuons un recalage 2D/3D par DRR en utilisant les tomodensitométries synthétiques, avec une erreur de recalage de $2.1 \pm 0.2\text{mm}$ pour valider notre méthode. La méthode proposée, de par ses composantes que par son application, présente un fort potentiel tant pour la synthèse d'images médicales que pour le recalage multimodal.

ABSTRACT

The structural information in three dimensions brings valuable insight and added precision to orthopedic interventions, which may otherwise only rely on bidimensional imaging modalities. Not only does 3D imaging help improve surgical accuracy, but it also helps reduce invasiveness. The bidimensional imaging modality on which we focus in this work is the C-Arm fluoroscopy. The latter is very common in operating theaters and allows for real-time versatile image acquisition to help guide interventions. That modality is often fused with CT scans to bring the added precision from the third dimension and lift the projective uncertainty on depth. CT and C-Arm fluoroscopy both rely on the physical principle of X-Ray absorption, which allows for their respective geometry and intensity distributions to be strongly correlated, and makes the registration problem relatively easier. However, computed tomography does not have the best contrast to visualize organs, nerves and soft tissue. It also involves a non-negligible radiation dose. On the other hand, MRI allows itself to a better visualization of those organs and of soft tissue. It also has the advantage of not exposing the patient to ionizing radiation. The MRI contrast being very different than that of the CT, especially for the spine, makes the 2D/3D registration problem much harder.

The latest advances in deep learning show promising results for the task of image translation, which is applicable to the generation of a synthetic CT from an MRI, for a subsequent 2D/3D registration to C-Arm fluoroscopy through Digitally Reconstructed Radiographs (DRR).

We propose such a method for 2D/3D registration between magnetic resonance imaging and C-Arm fluoroscopy, based on synthetic CT generation using deep learning methods. First, we explore numerous generative adversarial network architectures. Those architectures have shown impressive results for non-medical image translation. We experiment with a public dataset of 18 MRI and CT volumes. We notice that the CycleGAN architecture generalizes better to unseen data than the cGAN architecture does. The latter tends to overfit. We also introduce two new components to the CycleGAN architecture, which improves the tridimensional resolution as well as the voxel intensity distribution of the synthetic data. Finally, we perform DRR-based 2D/3D registration using the synthetic CT, and validate our method with a registration error of $2.1 \pm 0.2\text{mm}$. The proposed method, through its components and through its application, offers a strong potential for medical image synthesis as well as multimodal registration.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vi
TABLE DES MATIÈRES	vii
LISTE DES TABLEAUX	x
LISTE DES FIGURES	xi
LISTE DES SIGLES ET ABRÉVIATIONS	xv
LISTE DES ANNEXES	xvi
CHAPITRE 1 INTRODUCTION	1
1.1 Plan du mémoire	3
CHAPITRE 2 REVUE DE LITTÉRATURE	4
2.1 Structures anatomiques	4
2.1.1 Colonne vertébrale	4
2.1.2 Vertèbres	5
2.1.3 Sacrum et coccyx	5
2.2 Modalités d'imagerie	6
2.2.1 Imagerie fluoroscopique	6
2.2.2 Tomodensitométrie (CT Scan)	7
2.2.3 Imagerie par Résonance Magnétique (IRM)	9
2.3 Recalage 2D/3D, méthodes et modèles	11
2.3.1 Principes et types de recalage	11
2.3.2 Modèle géométrique de la caméra pour la fluoroscopie	11
2.3.3 Modèle physique pour la fluoroscopie et sa reconstruction digitale	15
2.3.4 Recalage 2D/3D manuel par marqueurs	18
2.3.5 Recalage 2D/3D par optimisation	20

2.3.6	Recalage de bout en bout par réseaux profonds	25
2.4	Méthodes et modèles de prédiction pour la synthèse IRM-CT	27
2.4.1	Vue globale sur l'état de l'art	27
2.4.2	Méthodes d'apprentissage machine	28
2.4.3	Méthodes d'apprentissage profond	30
2.5	Mot de synthèse	38
CHAPITRE 3	DÉMARCHE DE L'ENSEMBLE DU TRAVAIL DE RECHERCHE .	39
CHAPITRE 4	ARTICLE 1: MRI TO C-ARM SPINE REGISTRATION THROUGH PSEUDO3D CYCLE GANS WITH DIFFERENTIABLE HISTOGRAMS	41
4.1	Présentation de l'article	41
4.2	En-tête de l'article	41
4.3	Abstract	42
4.4	Introduction	43
4.5	Methods and materials	45
4.5.1	Pipeline architecture	45
4.5.2	Differentiable histogram loss	48
4.5.3	Training procedure	50
4.5.4	DRR Registration	51
4.6	Results	51
4.6.1	Experimental setup	51
4.6.2	Differentiable histograms	52
4.6.3	Image domain transfer	53
4.6.4	2D/3D Registration	56
4.7	Discussion	58
4.8	Conclusion	61
CHAPITRE 5	RÉSULTATS COMPLÉMENTAIRES	62
5.1	Histogrammes dérivables	62
5.1.1	Développements théoriques	62
5.1.2	Synthèse d'image	65
5.2	Recalage 2D/3D	66
CHAPITRE 6	DISCUSSION GÉNÉRALE	70
6.1	Synthèse IRM-CT	70
6.2	Recalage 2D/3D	72

6.3	Limitations	73
6.4	Perspectives futures	73
CHAPITRE 7 CONCLUSION		75
BIBLIOGRAPHIE		76
ANNEXES		86

LISTE DES TABLEAUX

Table 4.1	64x64 Fully convolutional discriminator architecture	46
Table 4.2	9-Block ResNet generator architecture	47
Table 4.3	Comparison of pseudo-3D models with and without histogram loss. Mean Absolute Error (MAE) in HU and Normalized Histogram Intersection (NHI) metrics between real to synthetic images (for CT and MRI), per patient.	55
Tableau 5.1	Valeurs des seuils $h(\alpha)$ pour différentes valeurs de alpha α . Les $h(\alpha)$ sont définis tel que $\sum_{i=0}^{n_{bins}-1} u_i(\tilde{X}_p) \in [h(\alpha), 1)$ pour $\tilde{X}_p \in [0, n_{bins} - 1]$	63
Tableau 5.2	Erreurs de recalage pour le fantôme de vertèbres lombaires .	68

LISTE DES FIGURES

Figure 2.1	Les différentes régions de la colonne vertébrale. Image tirée de OpenStax 2016	4
Figure 2.2	Schémas d'une vertèbre typique. Image tirée de Wikimedia Commons 2019	5
Figure 2.3	Bras en C biplan. Image tirée de Wikimedia Commons 2019 .	6
Figure 2.4	Schémas d'un système de radiographie typique. Inspiré de (Prince, 2006)	6
Figure 2.5	Deux exemples de distorsions radiale sur imagerie analogue. Image tirée de (Deserno, 2011).	7
Figure 2.6	Schéma de la tomodensitométrie. Wikimedia Commons 2019 .	8
Figure 2.7	Moments magnétiques nucléaires. À gauche, échelle microscopique, avant magnétisation externe. À droite, moment magnétique macroscopique après magnétisation externe.	9
Figure 2.8	Rotation du moment magnétique transversal après excitation par un champs polarisé à la fréquence de Larmor, pendant une durée TE correspondant à une rotation de 90°	9
Figure 2.9	Modèles de la relaxation des composantes longitudinale M_z (2.9a) et transversale M_\perp (2.9b) du moment magnétique M . Images tirées (Brown et al., 2014)	10
Figure 2.10	Structure typique d'un algorithme de recalage. Inspiré de (Johnson et al., 2015)	11
Figure 2.11	Modélisation du système de fluoroscopie en rayons X. Dans le modèle géométrique, la relation entre x et x' se traduit par $x' = \frac{f}{z}x$.	12
Figure 2.12	Différents repères pour modéliser la caméra projective. Tiré de (Sturm 2014)	13
Figure 2.13	Géométrie épipolaire d'une système de deux caméras avec des correspondances 2D/2D. Wikimedia Commons 2019	14
Figure 2.14	Schémas du problème de calcul du chemin radiologique exacte en 2D. Image tirée de (Jacobs, 1998).	17

Figure 2.15	Exemple de triangulation de 6 repères anatomiques d'une vertèbre sur deux vues différentes. Les repères sont les centres des plateaux inférieur et supérieur, ainsi que les extrémités des pédicules. Les croix rouges désignent les points marqués manuellement alors que les points verts désignent les points reconstruits en 3D puis rétroprojetés en 2D pour validation.	19
Figure 2.16	Récapitulatif des familles d'algorithmes de recalage 2D/3D, classés par objectif d'optimisation (intensité, géométrie ou gradients) et par type de transformation 2D/3D (projection, rétro-projection ou reconstruction). Image tirée de (Markelj et al. 2012)	20
Figure 2.17	Architecture de réseaux de neurones siamois pour le recalage d'images. Une couche de probabilités log-softmax est ajoutée au vecteur de corrélations. Image tirée de © IEEE 2019 (Beaupré et Bilodeau, 2019).	26
Figure 2.18	Exemple de machine à vecteur de support linéaire. Figure générée avec (https://github.com/ageron/handson-ml)	29
Figure 2.19	Exemple de jeu de données rendu linéairement séparable après l'application d'un noyau ϕ . Wikimedia Commons 2019	29
Figure 2.20	Modèle de régression par mélange de gaussiennes. Figure générée avec (https://github.com/AlexanderFabisch/gmr)	30
Figure 2.21	Schémas d'un perceptron multi-couche. Wikimedia Commons 2019	31
Figure 2.22	Illustration de la convolution 2D. Image tirée de (Dumoulin et Visin, 2016)	32
Figure 2.23	Architecture de transfert neuronal de style avec un ResNet qui apprend la procédure d'optimisation. Image tirée de (Johnson et al. 2016)	33
Figure 2.24	Exemple de photographies transformées dans des styles de tableaux. Images tirées de © IEEE 2015 (Gatys et al. 2015)	34
Figure 2.25	Architecture d'un réseau adverse génératif	35
Figure 2.26	Architecture du cGAN Pix2Pix. Image tirée de (Isola et al. 2017) © IEEE 2017	36
Figure 2.27	Architecture du CycleGAN (Zhu et al. 2017). La notation \hat{x}/\hat{y} désigne une reconstruction alors que le symbole \tilde{x}/\tilde{y} désigne une prédiction.	37

Figure 2.28	Schématisation chronologique, de gauche à droite, du processus d'apprentissage d'un réseau génératif adverse. La courbe noire désigne la distribution cible que l'on souhaite apprendre, la courbe verte représente la distribution apprise à l'itération courante alors que la courbe bleu désigne le pouvoir discriminatif du réseau. Progressivement, l'application de l'ensemble des \mathbf{z} vers l'ensemble des \mathbf{x} se rapproche de plus en plus de la distribution cible. Lorsque celles-ci se confondent, la courbe bleu illustre comment le discriminateur n'a plus aucun pouvoir discriminatif et qu'il ne peut plus faire la différence entre $G(\mathbf{z})$ et x . Image tirée de (Goodfellow et al., 2014)	38
Figure 4.1	Overview of the 3D/2D MRI to fluoroscopic image registration pipeline. The diagram shows the two-stage synthetic image generation procedure. First, a CycleGAN is trained to jointly learn the MRI to CT and CT to MRI mapping. Once the training is performed, synthetic CT volumes are generated for an unseen MRI. Finally, the synthetic CTs are registered to the C-Arm images through Digitally Reconstructed Radiographs of the synthetic CT.	45
Figure 4.2	For (a) and (b), $\alpha = 20$ was used. For (c), $\alpha = 6$ was used along with $n_{bins} = 9$. We see how the expression on (c) evaluates very closely to 1 on $[0, 8]$. If α was lower, the assumption would no longer be valid.	50
Figure 4.3	Experimental results for mutual information (MI) estimation with (a) additive white Gaussian noise channel, and (b) correlated bivariate Gaussians.	52
Figure 4.4	Absolute error maps of real and synthetic images (in HU). The first column shows images with the Pseudo-3D CycleGAN without the histogram loss, while the images in the second column were produced using the histogram loss. Images of two different patients are show in each row.	53
Figure 4.5	Sample voxel intensity histograms for two patient volumes, one for each row. Volumes are synthesized using the two pseudo-3D CycleGANs with and without the histogram loss term, one for each column.	54
Figure 4.6	Synthesis results from two different patients (each row), comparing results of different MRI to CT synthesis models to the real CT.	56
Figure 4.7	3D RMS registration errors on 6 vertebral anatomical landmarks of all vertebrae, for each patient.	57

Figure 4.8	3D RMS registration errors per vertebral level, for both pseudo-3D CycleGAN models with and without histogram loss.	58
Figure 4.9	(a)-(b) Two sample 3D/2D registration overlay results with Pseudo-3D CycleGAN, showing C-arm and DRR images. For each case, top row is without histogram loss, bottom row is with histogram loss. Left image is the lateral view, while the right image is at -45° . The corresponding MRI slices are depicted with the dashed green lines.	59
Figure 5.1	Schémas de la formulation alternative de f sans la restriction d'ensemble d'arrivée.	63
Figure 5.2	Comparaison de CT synthétiques sur le plan axial pour différentes épaisseurs de coupes sagitales, pour deux cas différents.	65
Figure 5.3	Erreurs absolues moyennes avec différentes épaisseurs de coupes sagitales pour l'ensemble du jeu de données	66
Figure 5.4	Résultats comparatifs pour différents modèles de synthèse IRM-CT pour des volumes de 4 patients différents, avec un volume par ligne. Les images, pour chaque patient, ont été réglées pour être visualisée sur la même plage de valeurs pour que les contrastes soient équitablement comparables.	67
Figure 5.6	Ensemble des 4 vues pour le recalage de la vertèbre L5 du fantôme. Le calque en rouge représente la radiographie synthétique de ce fantôme alors que le fond en nuances de gris représente la radiographie réelle.	67
Figure 5.5	Exemples d'histogrammes d'intensités pour 4 volumes de patients différents, avec un volume par colonne. La première ligne correspond aux images synthétisées par le modèle pseudo-3D sans perte d'histogrammes. La seconde ligne correspond aux images synthétisées par le modèle pseudo-3D avec perte d'histogrammes.	68
Figure 5.7	Exemples pour 3 cas différents. Pour chaque cas, les images de la premières lignes ont été générées avec des volumes issus du modèle avec perte d'histogrammes, alors que les images de la deuxième ligne ont été générées à partir des volumes issus du modèle pseudo-3D régulier. Les vues présentées sont la vue latérale et la vue à -45°	69
Figure A.1	Architecture du réseau de type FCN. Image tirée de (Drozdal et al., 2017)	86
Figure A.2	Architecture du réseau de type FC-ResNet. Image tirée de (Drozdal et al., 2017)	87

LISTE DES SIGLES ET ABRÉVIATIONS

IRM	Imagerie de résonance magnétique
CT	Computed Tomography
ICP	Iterative Closest Point
DRR	Digitally Reconstructed Radiograph
GAN	Generative Adversarial Network
CycleGAN	Cycle Generative Adversarial Network
cGAN	Conditional Generative Adversarial Network
CNN	Convolutional Neural Network
MLP	Multi-layer perceptron
DLT	Direct Linear Transform
SVD	Singular Value Decomposition
IG	Information de Gradients
CG	Corrélation de Gradients
GO	Gradient Orientation
NCC	Normalized Cross-Correlation
IM	Information Mutuelle
BFGS	Broyden–Fletcher–Goldfarb–Shanno
SVM	Support Vector Machine
VGG	Visual Geometry Group

LISTE DES ANNEXES

Annexe A	Architecture FCN et FC-ResNet	86
----------	---	----

CHAPITRE 1 INTRODUCTION

Dans le cadre de la chirurgie orthopédique, le recalage 2D/3D de modalités pré-opératoires et de modalités intra-opératoires est primordial pour assurer une bonne précision de la procédure et faire un suivi en temps réel des manoeuvres effectuées selon la planification prévue. Dans le cas contraire ; les modalités intra-opératoires, pour la plupart, permettent d’avoir uniquement l’information en deux dimensions. Malgré l’adoption de plusieurs angles de vue en simultané, lever l’ambiguïté projective dans la dimension de la profondeur reste difficile. La modalité la plus courante dans ce cas-ci est la fluoroscopie de bras en C. Celle-ci est même très répandue dans les salles opératoires. La problématique du recalage 2D/3D est d’autant plus critique pour la chirurgie orthopédique, que les marges d’erreurs sont de l’ordre de quelques millimètres et que les structures osseuses ciblées sont à proximité d’organes, de tissus et de nerfs critiques. Non seulement faut-il donc être précis par rapport aux structures osseuses, mais il faut l’être aussi par rapport aux autres structures avoisinantes qui sont à risque. Le problème de recalage comporte par ailleurs une complexité supplémentaire, associée au fait que la colonne vertébrale soit articulée, et qu’il faut donc effectuer un recalage différent pour chacune des vertèbres.

Il existe des méthodes relativement fiables pour le recalage 2D/3D de la tomodensitométrie et de la fluoroscopie de bras en C. Ce recalage est dit quasi-intramodal (Markelj et al., 2012). Cette quasi-similitude de modalité fait en sorte que le problème de recalage soit tractable avec des niveaux de précision cliniquement satisfaisants. Toutefois, quand il s’agit d’effectuer la fusion inter-modale de l’IRM et de la fluoroscopie de bras en C, il est bien plus difficile d’obtenir des précisions comparables au cas précédent. Ceci est dû aux différences intrinsèques aux deux modalités.

Pour l’IRM, Il existe des méthodes de recalage basées sur le marquage de repères anatomiques (Bijhold, 1993, Stokes et al., 1981), mais celles-ci sont très sensibles à la précision de l’opérateur, demandent une certaine expertise de ce dernier et sont contraignante en terme de temps et de fluidité. Hormis la méthode par marqueurs anatomiques, il existe des méthodes à marqueurs physiques attachés au patient (George et al., 2011). Celles-ci, malgré leur précision, ont le désavantage d’être très invasives. Par ailleurs, certaines méthodes, reposant sur des techniques d’optimisation, cherchent à faire correspondre des courbes et des surfaces d’une modalité à l’autre (Da Silva et al., 2017). Celles-ci utilisent une segmentation pré-opératoire de vertèbre, extraite de l’IRM. Même si elles prennent en compte l’aspect géométrique du problème, ces méthodes-là n’ont pas l’avantage d’utiliser des images aussi riches en relief et en texture, et aussi translucides qu’une radiographie. Lorsque cette richesse

d'information additionnelle est disponible pour le recalage, cela apporte autant plus d'éléments pour lesquels il faut satisfaire la contrainte de correspondance, et donc autant plus de précision atteignable. Pour palier à ce problème, une famille de méthodes de recalage effectue la génération de tomodensitométrie synthétique à partir d'IRM pour ensuite effectuer un recalage 2D/3D par intensités à l'aide de Radiographie Reconstituées Numériquement (DRR) (Van der Born et al. 2011). Ces méthodes utilisent toutefois des approches d'apprentissage machine classique, demandent des données paires difficiles à obtenir, et font uniquement de la régression par intensités sans incorporer d'information à plus haut niveau d'abstraction. Le travail de ce mémoire s'inscrit donc dans cette problématique de proposer une méthode de recalage 2D/3D entre l'imagerie de résonance magnétique et la fluoroscopie de bras en C, qui prenne en compte les avantages et les inconvénients des méthodes citées précédemment, et qui essaye le plus possible de les compléter.

Grâce aux dernières avancées en apprentissage profond, il est désormais possible d'utiliser des réseaux profonds pour la translation d'image. Cette tâche revient à transférer une image d'une modalité à une autre, tout en préservant le contenu qui leur est commun. Les techniques d'apprentissage machine offrent des outils qui peuvent partiellement résoudre ce problème. Toutefois, là où l'apprentissage profond excelle, c'est que les méthodes utilisées sont capables d'apprendre une certaine représentation des images à un certain niveau d'abstraction. Cette approche qu'on adopte avec l'apprentissage profond, de non pas directement modéliser les données numériques en tant que tel, mais plutôt de modéliser leur représentation, a la promesse d'être plus robuste aux légers changements de détails, de contraste, ou de décalage de covariable en général; tant que le contenu peut être représenté de la même manière. Le travail de (Zeiler et al., 2014) a démontré comment les réseaux convolutifs profonds peuvent apprendre différents niveaux de représentations des caractéristiques d'une image. Le travail de (Gatys et al., 2016) a montré comment cet apprentissage de représentation peut être utilisé pour mesurer des similarités de style et de contenu entre différentes images, et par la suite utiliser cette mesure pour effectuer la translation d'image par optimisation. Les méthodes de translation d'image les plus récentes de translation d'image utilisent des réseaux adverses génératifs (Goodfellow et al., 2014) qui ont montré d'impressionnants résultats pour cette tâche-là (Isola et al., 2017, Zhu et al., 2017).

L'objectif de ce mémoire est donc de développer une méthode automatique de recalage 2D/3D par radiographie numériquement reconstruite en utilisant des tomodensitométries synthétiques produites par des réseaux adverses génératifs.

1.1 Plan du mémoire

Le présent mémoire s'organise de la manière suivante. En premier lieux, nous effectuons une revue de littérature, au chapitre 2. Cette revue de littérature aborde les concepts importants à la compréhensions des différences et similitudes intrinsèques aux modalités de résonance magnétique, de tomodensitométrie, et de fluoroscopie de bras en C. Elle introduit ensuite les différents types de recalage 2D/3D, avant d'aller plus en détail dans les méthodes de recalage 2D/3D basées sur l'optimisation des intensités. Enfin, la revue de littérature conclu sur les différentes approches d'apprentissage machine et d'apprentissage profond pour générer des tomodensitométries synthétiques. Le méthodologie du travail subséquent est spécifiée au chapitre 3. Le chapitre 4 contient un article scientifique soumis au cours de ce projet de recherche. Le chapitre 5 discute globalement du projet de recherche alors que le chapitre 6 en fait une conclusion.

CHAPITRE 2 REVUE DE LITTÉRATURE

2.1 Structures anatomiques

2.1.1 Colonne vertébrale

La colonne vertébrale est constituée de 33 vertèbres, séparées pour la plupart par des disques intervertébraux en cartilage, et qui forment 4 régions : la région cervicale, la région thoracique, la région lombaire, le sacrum et le coccyx. La région cervicale est faite de 7 vertèbres, annotées de C1 à C7. La région thoracique est faite de 12 vertèbres annotées de T1 à T12. La dernière région, la région lombaire, comporte 5 vertèbres annotées de L1 à L5. Le sacrum et le coccyx sont en réalité le produit de la fusion de 5 et de 4 vertèbres respectivement (OpenStax, 2016).

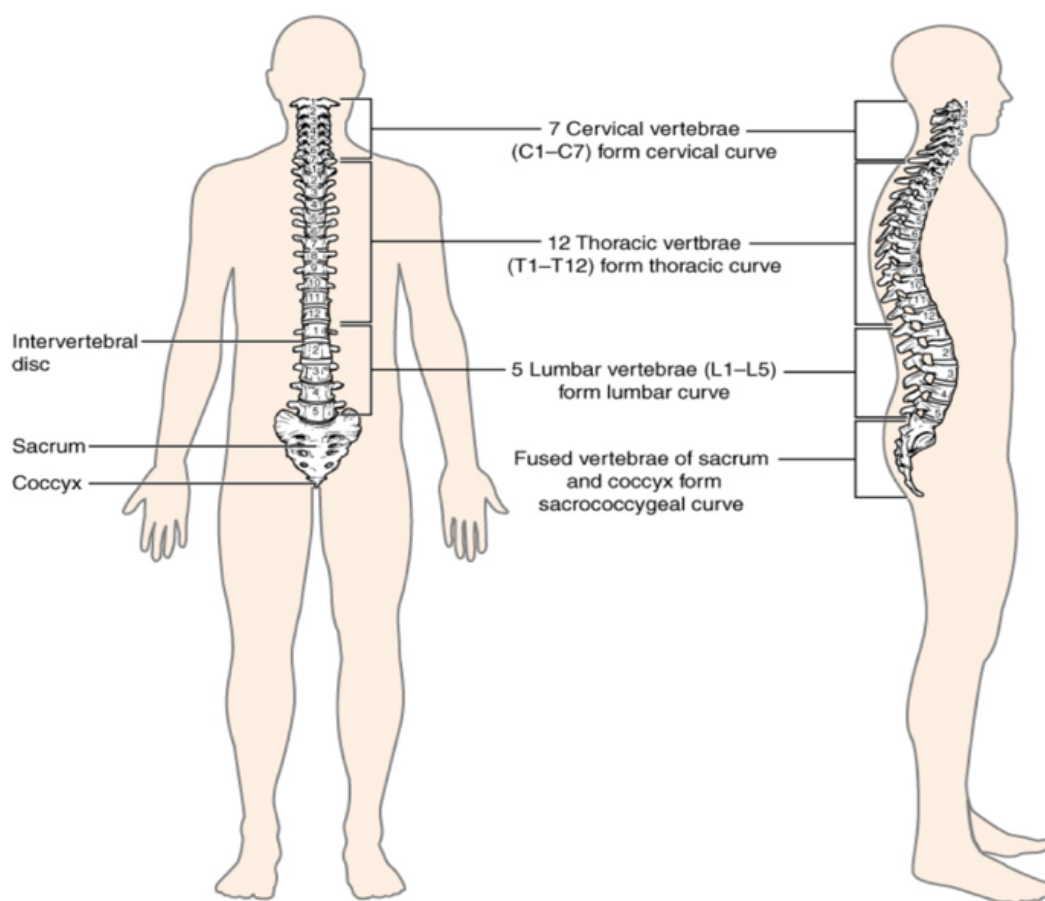


Figure 2.1 Les différentes régions de la colonne vertébrale. Image tirée de OpenStax 2016

2.1.2 Vertèbres

Dans cette partie, nous examinerons de plus près la structure anatomique de la vertèbre. La structure et les dimensions des vertèbres dépendent de la région dans laquelle elles se trouvent, mais elles ont toutes une structure globale commune. Une vertèbre est typiquement constituée d'un corps vertébral, de deux pédicules, deux processus articulaires supérieurs, deux processus articulaires inférieurs recouverts de cartilage, deux processus transverses et d'un processus épineux (O'Rahilly, Muller 1983).

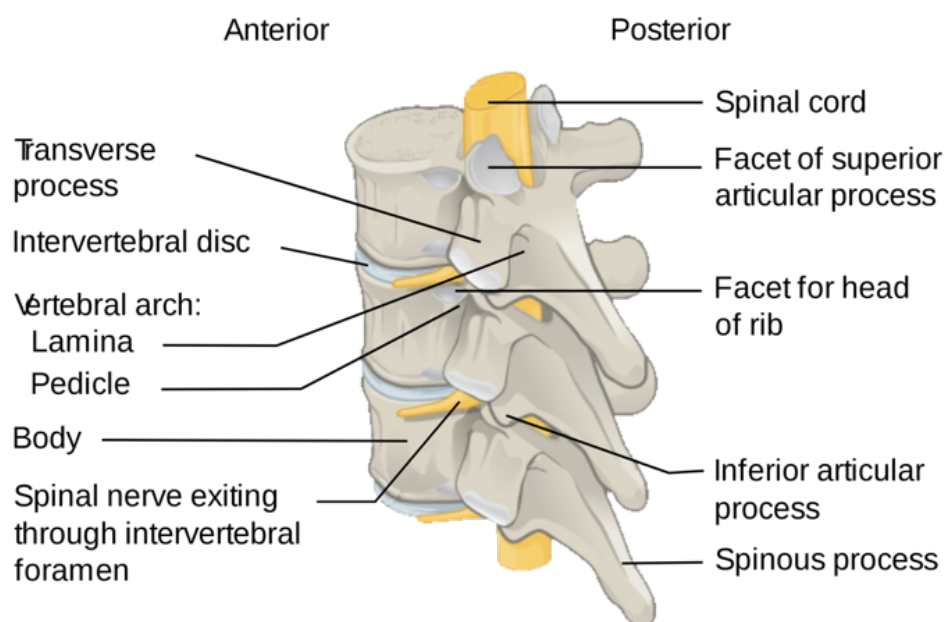


Figure 2.2 Schémas d'une vertèbre typique. Image tirée de Wikimedia Commons 2019

2.1.3 Sacrum et coccyx

Le sacrum est la structure osseuse qui se trouve à la base des vertèbres lombaires. Il est formé de la fusion des vertèbres S1-S5 et se situe en dessous de la vertèbre L5. Il se glisse dans le bassin osseux entre les deux os iliaques. Le coccyx est, quant à lui, le résultat de la fusion des 4 vertèbres coccygiennes. Il se situe en dessous du sacrum et a une forme élongée avec des discontinuités transversales qui marquent les bordures des vertèbres coccygiennes (O'Rahilly, Muller 1983).

2.2 Modalités d'imagerie

2.2.1 Imagerie fluoroscopique



Figure 2.3 Bras en C biplan. Image tirée de Wikimedia Commons 2019

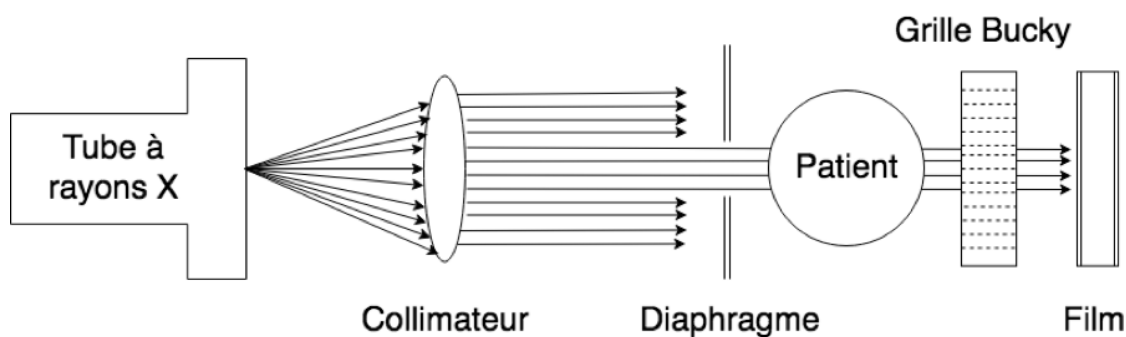


Figure 2.4 Schémas d'un système de radiographie typique. Inspiré de (Prince, 2006)

La fluoroscopie est la modalité d'imagerie la plus répandue en salle chirurgicale (De Silva et al., 2017). Son principal atout est l'acquisition simple et rapide. Elle permet aussi d'avoir de l'information sur la position des différents objets et des structures anatomiques les uns par rapport aux autres et ce, de manière quasi-instantanée (Patel et al. 2008).

L'appareil sur lequel on porte notre attention est le bras en C. Comme son nom l'indique, celui-ci est constitué d'un support et d'un bras mobile sur plusieurs degrés de liberté. Le support peut être fixé au sol, comme il peut être fixé au plafond. S'il ne se translate pas, c'est alors la table d'opération que l'on viendra bouger en translation le plus souvent.

Sur les deux extrémités opposées du bras en C se trouvent une source de rayons X et un récepteur, constitué d'un amplificateur de brillance. Des appareils plus récents disposent de capteurs digitaux plats.

La fluoroscopie repose sur le principe physique d'absorption qui suit la loi de Beer-Lambert,

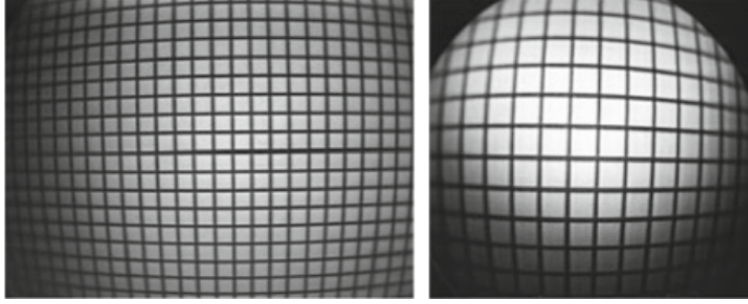


Figure 2.5 Deux exemples de distorsions radiale sur imagerie analogue. Image tirée de (Deserno, 2011).

de la forme :

$$I = I_0 e^{\int_0^{+\infty} \mu(u) du} \quad (2.1)$$

Où μ désigne le coefficient d'absorption linéique, I désigne l'intensité transmise, et I_0 désigne l'intensité incidente. L'opérateur a des degrés de liberté sur l'énergie d'alimentation de la source afin de contrôler la dose de radiation, ce qui a des effets sur la qualité de l'image mais aussi sur la radiation que reçoit le patient. Des filtres et des systèmes de collimation viennent améliorer le rapport signal sur bruit sans pour autant augmenter l'intensité émise (Prince, 2006). Cette correction se fait au cours du traitement d'images en sortie du système d'imagerie, à l'aide de modèles mathématiques et d'équations d'optiques qui considèrent le système de fluoroscopie comme une caméra projective (Zisserman et Hartley, 2003).

Dans le cas d'un système cathodique, l'image de sortie présente de la distorsion radiale qui peut être corrigée à l'aide d'un calibrage par fantôme de lignes parallèles (Deserno, 2011).

2.2.2 Tomodensitométrie (CT Scan)

La tomodensitométrie, ou CT-scan, est une modalité tridimensionnelle, basée elle aussi sur le principe d'absorption des rayons X. Celle-ci se fait par l'acquisition de tranches axiales successives, et l'empilement de ces tranches forme par la suite tout le volume en entier. Chaque tranche axiale est acquise par la rotation d'un émetteur et d'un récepteur autour de la coupe cible à imager. Ce phénomène est modélisé mathématiquement par une transformée de Radon de la carte des densités d'absorptions, appelée sinogramme (Prince, 2006).

La transformée de Radon est de la forme suivante, où δ désigne l'impulsion de Dirac :

$$g(l, \theta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \delta(x \cos(\theta) + y \sin(\theta) - l) dx dy \quad (2.2)$$

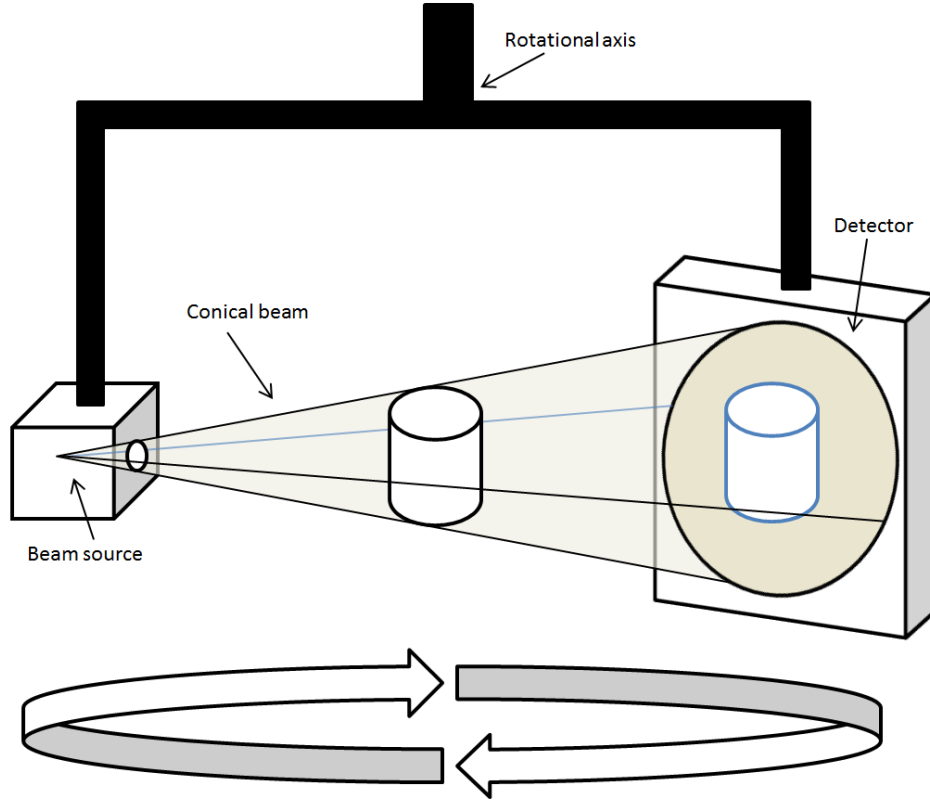


Figure 2.6 Schéma de la tomodynamométrie. Wikimedia Commons 2019

Les sinogrammes sont acquis avec un certain pas de rotation dénoté $\Delta\theta$, et ceux-ci sont filtrés, rétro-projetés, puis sommés pour obtenir la reconstruction finale. L'unité d'intensité conventionnellement utilisée est le Hounsfield (HU). Celle-ci est une fonction affine des coefficients d'absorption linéique μ , et a pour expression :

$$HU(\mu) = 1000 \frac{\mu - \mu_{\text{eau}}}{\mu_{\text{eau}} - \mu_{\text{air}}} \quad (2.3)$$

Les systèmes de tomodynamométrie sont classés en 7 générations (Prince, 2006). Les améliorations qui ont été graduellement apportées ont d'abord commencé par une meilleure collimation, tant au niveau de la source, qu'au niveau du détecteur. Tandis que les scanners de 1ère et 2ème génération effectuent des translations suivies par des rotations, les appareils de la 3ème génération sont les premiers à n'effectuer que des rotations, ce qui diminue le temps d'acquisition. Pour les générations suivantes, les capteurs sont répartis circulairement à l'intérieur du dispositif, et ce sont les émetteurs qui effectuent la rotation (Prince, 2006).

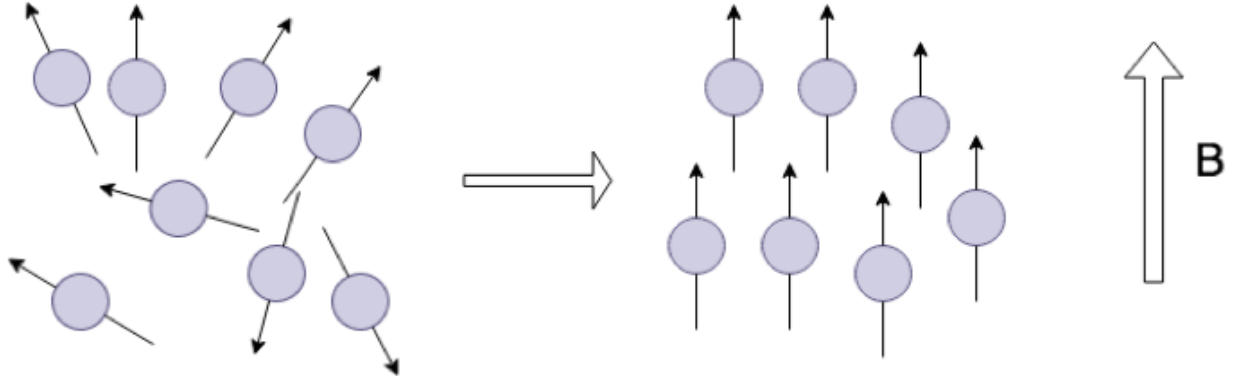


Figure 2.7 Moments magnétiques nucléaires. À gauche, échelle microscopique, avant magnétisation externe. À droite, moment magnétique macroscopique après magnétisation externe.

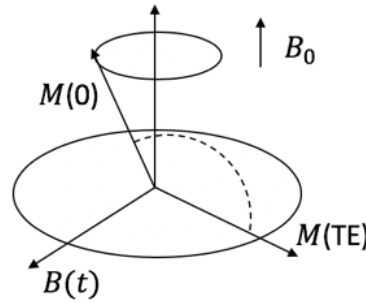


Figure 2.8 Rotation du moment magnétique transversal après excitation par un champs polarisé à la fréquence de Larmor, pendant une durée TE correspondant à une rotation de 90°

2.2.3 Imagerie par Résonance Magnétique (IRM)

Sur la figure 2.8, les spins magnétiques suivent un mouvement de précession autour de l'axe du champ magnétique externe, avec une composante longitudinale invariante dans le temps, mais une composante transversale rotative. Toutefois, la somme de ces moments magnétiques résulte à une composante macroscopique longitudinale invariante dans le temps (Brown et al., 2014). Contrairement à la tomodensitométrie, l'IRM ne produit pas de radiation dans le spectre des rayons X. Celle-ci repose sur le principe physique de la résonance magnétique. Cet avantage en fait l'une des modalités les plus courantes aujourd'hui (Da Silva et al., 2017). En IRM, on s'intéresse principalement au moment magnétique des noyaux d'hydrogène. Celui-ci existe à l'échelle microscopique, mais étant donné le caractère aléatoire de son orientation, il s'annule à l'échelle macroscopique. Cependant, sous l'influence d'un champ magnétique ex-

terne, le moment magnétique atomique a de plus en plus tendance à s'aligner avec le champ magnétique externe appliqué (Prince, 2006).

Lorsque le système est à l'équilibre, les moments magnétiques sont en précession autour de l'axe du champ magnétique longitudinal. La fréquence de rotation du moment magnétique transversal est caractéristique du spécimen, et elle porte le nom de la fréquence de Larmor (Weishaupt et al., 2008).

On peut exciter les spins magnétiques par un champs transversal rotatif à cette fréquence de Larmor. Ceci induira une rotation du moment magnétique et, selon la durée d'application de cette excitation, celle-ci aura faite tourner le moment magnétique d'un angle α .

À la suite de cette excitation, il va y avoir un retour à un état de stabilité, comme indiqué par M_0 sur la figure 2.8. Ce retour à l'équilibre suit un modèle exponentiel avec deux constantes de temps T_1 et T_2 pour les composantes longitudinales et transversales respectivement. Ces temps de relaxation sont caractéristiques de la matière à imager, et c'est ce qui permet de différencier les constituants d'un spécimen (Brown et al. 2014). Il est à noter que le champs magnétique B_0 n'est pas constant, mais suit un gradient d'intensité par palier G_z pour permettre un encodage spatial selon l'axe z (Brown et al. 2014). Les gradients G_x et G_y permettent à ce que la réponse à l'excitation magnétique, au niveau d'une coupe, soit modélisée par une transformée de Fourier spatiale en 2D. En imagerie par résonance magnétique, cet espace de Fourier 2D porte le nom de k-space (Prince, 2006). Par des séquences d'excitation et de mesure de signal bien appropriées, on peut lire des lignes dans ce k-space. L'opération finale pour retrouver le signal spatial est la transformée de Fourier inverse.

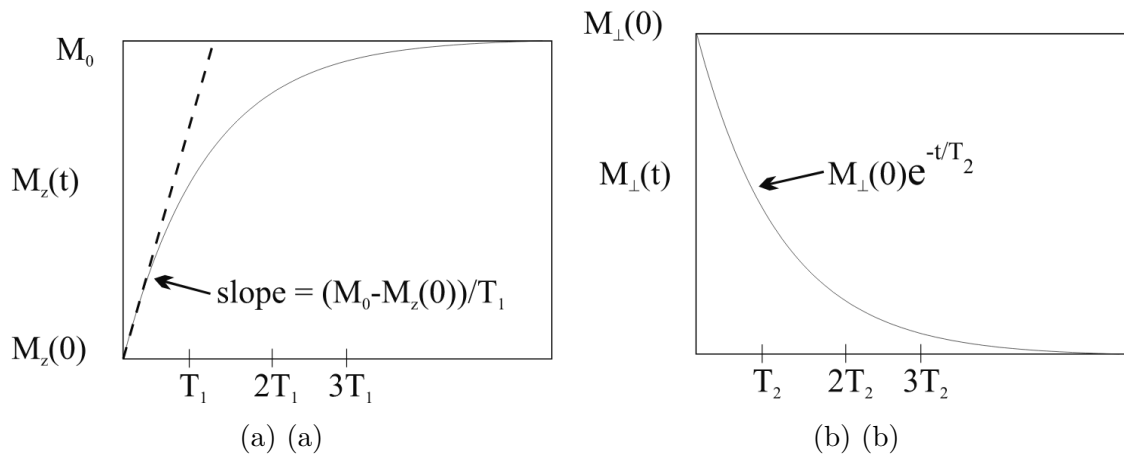


Figure 2.9 Modèles de la relaxation des composantes longitudinale M_z (2.9a) et transversale M_\perp (2.9b) du moment magnétique M . Images tirées (Brown et al., 2014)

2.3 Recalage 2D/3D, méthodes et modèles

2.3.1 Principes et types de recalage

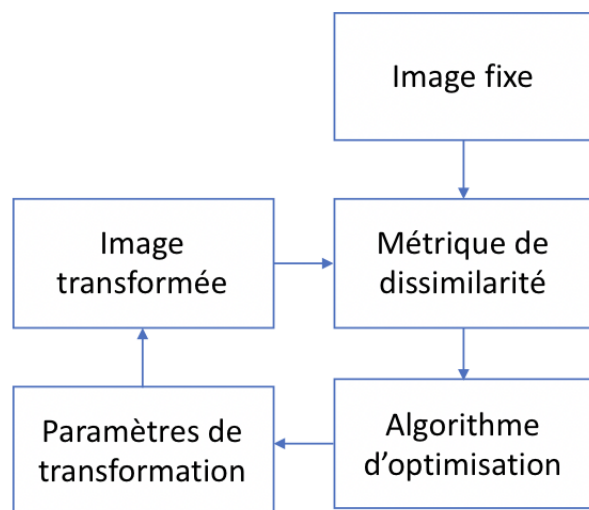


Figure 2.10 Structure typique d'un algorithme de recalage. Inspiré de (Johnson et al., 2015)

Lorsque l'on acquiert des images d'un même contenu, mais de deux manières différentes, la problématique qui resurgit le plus souvent est de trouver la correspondance entre ces deux images afin de pouvoir les fusionner. Ces images peuvent être de différentes modalités comme elles peuvent être de différentes dimensionnalités. Les transformations, elles aussi, peuvent être de différentes natures. Le plus souvent, les algorithmes de recalage suivent la même structure (Johnson et al., 2015). Ces algorithmes prennent deux images ainsi qu'une transformation paramétrée en entrée. On utilise une métrique de dissimilarité, en fonction de laquelle un algorithme d'optimisation ira chercher les paramètres de la transformation qui minimisent le plus cette métrique.

2.3.2 Modèle géométrique de la caméra pour la fluoroscopie

La formation d'une image fluoroscopique est modélisée à l'aide de la géométrie projective (Lilja, 2008). À l'opposé de la source de rayonnement se trouve un plan récepteur, comme indiqué sur la figure 2.11a. Sur ce plan récepteur sont projetés les rayons issus de la source. Ceux-ci sont atténués pendant leur passage au travers du spécimen à imager.

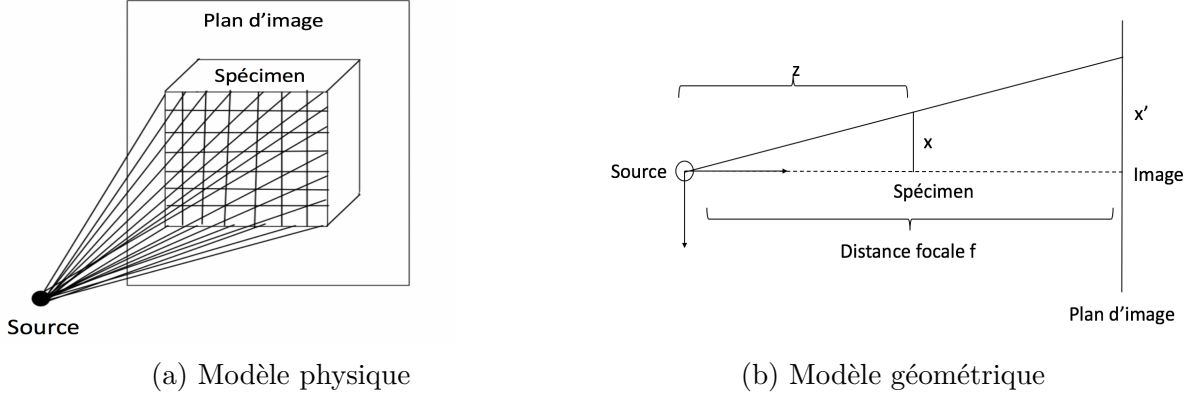


Figure 2.11 Modélisation du système de fluoroscopie en rayons X. Dans le modèle géométrique, la relation entre x et x' se traduit par $x' = \frac{f}{z}x$

Le modèle illustrée par le schéma de la figure 2.11b induit l'équation matricielle suivante :

$$\begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} fx \\ fy \\ z \end{pmatrix} = \begin{pmatrix} \frac{f}{z}x \\ \frac{f}{z}y \\ 1 \end{pmatrix} \quad (2.4)$$

Il est à noter que les coordonnées x, y et z de l'équation 2.4 sont dans le repère propre à la source, comme illustré figure 2.11b. Pour appliquer l'équation 2.4 à des points dans le référentiel fixe de base, il faut introduire une matrice de changement de base au préalable pour transformer tous les points dans le repère de la source. Par convention, on applique un changement d'origine dans le plan de projection 2D pour que celle-ci ne coïncide pas avec le point principal de l'image (Zisserman, Hartley 2003).

En utilisant la notation de la figure 2.12, les équations de projection se traduisent de la manière suivante, où R et T désignent respectivement la matrice de rotation et le vecteur de translation qui transforment un point du repère fixe, au repère de la caméra. La caméra a f pour distance focale et (c_x, c_y) pour point principal tel que $(u, v) = (x + c_x, y + c_y)$. Le point de coordonnées $(0, 0)$ est le plus souvent le coin haut gauche de l'image.

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} f & 0 & c_x & 0 \\ 0 & f & c_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_K \underbrace{\begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix}}_M \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = P \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (2.5)$$

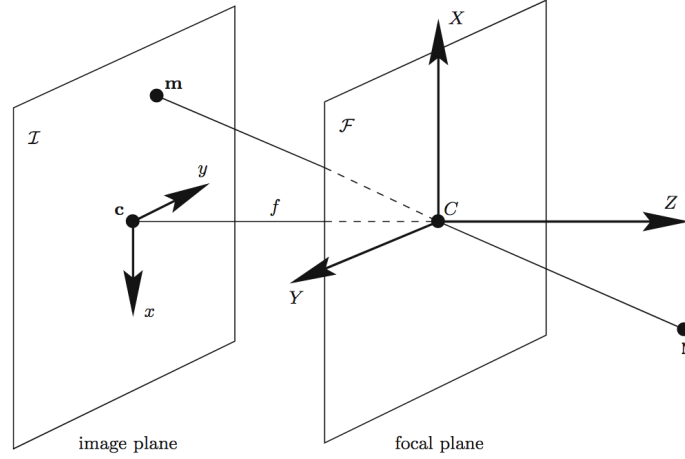


Figure 2.12 Différents repères pour modéliser la caméra projective. Tiré de (Sturm 2014)

Les paramètres de la caméra n'étant pas connus, il faut procéder à un calibrage pour retrouver les matrices de projection (Zisserman, Hartley, 2003). Dans ce projet, nous nous intéressons seulement au calibrage d'une caméra unique avec des correspondances de points 2D/3D ainsi qu'au calibrage d'une paire de caméras avec des correspondances 2D/2D.

Calibrage d'une paire de caméras

Dans la figure 2.13, il faut noter qu'un même point X donne lieu à deux images X_L et X_R . Ces points sont rétro-projetés aux rayons $[O_L, X_L)$ et $[O_R, X_R)$ respectivement. L'observation clé est que X, X_L, X_R, O_L et O_R sont tous coplanaires, et ce plan est dénommé π . Les lignes épipolaires forment l'intersection entre π et les plans images. Les points e_L et e_R sont les points d'intersection des lignes épipolaires avec le segment $[O_L, O_R]$, et ils portent le nom d'épipôles (Zisserman, Hartley, 2003). Étant donné que les points X_L et X_R sont les projections du même point X , il existe une homographie H_π tel que $X_R = H_\pi X_L$ (Zisserman, Hartley, 2003). La ligne épipolaire $l' = (e_R, X_R)$ est paramétrisée de la forme suivante :

$$l' = e_R \times X_R = \underbrace{([e_R]_\times H_\pi)}_F X_L = F X_L \quad (2.6)$$

La notation $[u]_\times$ désigne la matrice pour laquelle le produit matriciel $[u]_\times v = u \times v$ est de la forme $[u]_\times = \begin{pmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{pmatrix}$.

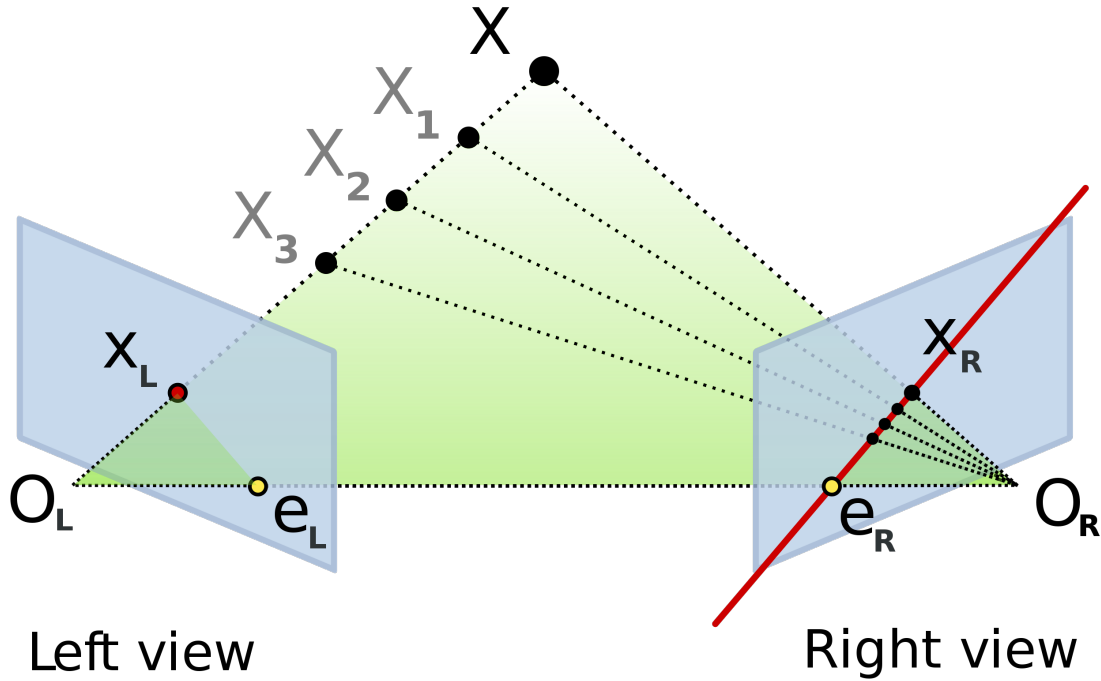


Figure 2.13 Géométrie épipolaire d'un système de deux caméras avec des correspondances 2D/2D. Wikimedia Commons 2019

Étant donné que x' appartient à l' , cela se traduit par l'équation :

$$X_R^T l' = X_R^T F X_L = 0 \quad (2.7)$$

La résolution de cette équation permet de retrouver la matrice F , dénommée matrice fondamentale. Étant donné que F est une matrice 3×3 , et que chaque correspondance donne lieu à une équation différente, et que ces équations sont égales à une constante multiplicative près, il faut au moins 8 équations pour retrouver F par résolution d'un système linéaire d'équations. Une fois F obtenue, les matrices de projection P et P' , sous forme canonique, sont obtenues tel que $P = \begin{pmatrix} I & | & 0 \end{pmatrix}$ et $P' = \begin{pmatrix} [e_R]_{\times} F & | & e_R \end{pmatrix}$ (Zisserman, Hartley 2003).

Calibrage d'une caméra unique

Dans le cas où l'on a besoin de calibrer une caméra indépendamment de toute autre caméra, il est alors nécessaire d'avoir des correspondances 2D/3D plutôt que des correspondances 2D/2D. Tout en gardant le modèle illustré figure 2.12, il nous est possible de résoudre l'équa-

tion $P \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} u/w \\ v/w \\ 1 \end{pmatrix}$. Pour cela, étant donné que chaque correspondance 2D/3D donne

lieu à deux équations, et que P est une matrice à 12 inconnues, il faut donc un minimum de 6 correspondances. En général, le système à résoudre est surdéterminé. Les équations peuvent être transformée à l'aide d'une Direct Linear Transform (DLT) (Sutherland, 1974), et ce, de la manière suivante :

$$\begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \iff \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{p_{11}X+p_{12}Y+p_{13}Z+p_{14}}{p_{31}X+p_{32}Y+p_{33}Z+p_{34}} \\ \frac{p_{21}X+p_{22}Y+p_{23}Z+p_{24}}{p_{31}X+p_{32}Y+p_{33}Z+p_{34}} \end{pmatrix} \quad (2.8)$$

$$\begin{pmatrix} X & Y & Z & 1 & 0 & 0 & 0 & 0 & -uX & -uY & -uZ & -u \\ 0 & 0 & 0 & 0 & X & Y & Z & 1 & -vX & -vY & -vZ & -v \end{pmatrix} \begin{pmatrix} p_{11} \\ \vdots \\ p_{34} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (2.9)$$

Equation 2.9 donne lieu à un système d'équations homogènes $AP = 0$, qui peut être résolu à l'aide d'une décomposition à valeurs singulières (SVD) (Zisserman, Hartley, 2003). La solution P^* qui est obtenue aura les propriétés $\|P^*\| = 1$ et $P^* = \operatorname{argmin}_P \|AP\|$.

Étant donné la décomposition en valeur singulière de A tel que $A = USV^T$, P^* est obtenue directement comme étant la dernière colonne de V , i.e le vecteur singulier de droit, ayant la valeur singulière correspondante la plus petite.

Au-delà des méthodes de calibrage explicite à résolution linéaire, il existe des méthodes de calibrage qui recourent à des techniques d'optimisation non-linéaire. On cherche à minimiser une fonction de coût. Celle-ci peut correspondre à l'erreur de projection 2D, à l'erreur de rétro-projection 3D, à l'erreur de projection 2D des rétro-projections 3D, ou à des combinaisons des erreurs précédentes. Ces techniques sont souvent désignées par 'bundle adjustment' (Zisserman, Hartley, 2003).

2.3.3 Modèle physique pour la fluoroscopie et sa reconstruction digitale

Rétroprojection des pixels 2D en rayons 3D

La figure 2.11a illustre comment la fluoroscopie est modélisée physiquement. Avec les équations 2.1 et 2.3, il est possible de simuler par le calcul le processus d'absorption des rayons X en projetant les rayons à travers la tomodesitométrie et en calculant l'intégrale-ligne discrète

pour l'équation de Beer-Lambert (Sherouse et al. 1990).

Avant cela, il faut d'abord retrouver les rayons. Connaissant, à partir de l'équation 2.5, la matrices de rotation R , la translation T , la distance focale f et le point principal (c_x, c_y) : un pixel de coordonnées (u, v) trace un rayon qui entre en intersection avec le plan focal, aux coordonnées (x, y, z) tel que :

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = R^{-1} \begin{pmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} fu \\ fv \\ f \end{pmatrix} - R^{-1}T \quad (2.10)$$

Le centre de la caméra a, quant la lui, des coordonnées (x_0, y_0, z_0) (Zisserman, Hartley, 2003) tel que :

$$\begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} = -R^{-1}T \quad (2.11)$$

Calcul de chemin radiologique exacte

Pour chaque rayon, il est possible de calculer le chemin radiologique exacte (Siddon, 1985) et (Jacobs, 1998).

Dans le cadre de ce mémoire, nous nous contenterons de présenter les algorithmes de Siddon (Siddon, 1985) et de Jacobs (1998), qui est une version améliorée du premier.

Supposons un volume V dans le repère (x, y, z) , muni d'une fonction de densité $\rho(i, j, k)$. Ce volume est traversé par un rayon r où $l(i, j, k)$ désigne la distance parcourue par ce rayon dans le parallélépipède délimité par les plans $z = k, z = k + 1, y = j, y = j + 1, x = i, x = i + 1$. Le chemin radiologique est donc défini par (Jacbos, 1998) :

$$d_{12} = \sum_{i,j,k} l(i, j, k) \rho(i, j, k) \quad (2.12)$$

Sans perte de généralité en 3D, le calcul du chemin radiologique 2D, comme illustré dans la figure 2.14, se fait entre les points p_1 et p_2 . Le carré est constitué d'une grille comportant un nombre N_x de plans perpendiculaires à x et un nombre N_y de plans perpendiculaires à y . Les espacement, en mm, de deux plans successifs selon x et y sont respectivement dénotés b_x et b_y . La première étape du calcul de ce chemin radiologique, que ce soit pour l'algorithme de Siddon ou de Jacobs, consiste à trouver les points d'entrée et de sortie du rayon, si celui-ci passe à travers le volume V . Le segment $[p_1, p_2]$ est paramétré par α , et $\alpha_{min}, \alpha_{max}$ correspondent aux paramètres d'entrée et de sortie du rayon, i.e $p_1 + \alpha_{min}(p_2 - p_1)$ et $p_1 + \alpha_{max}(p_2 - p_1)$

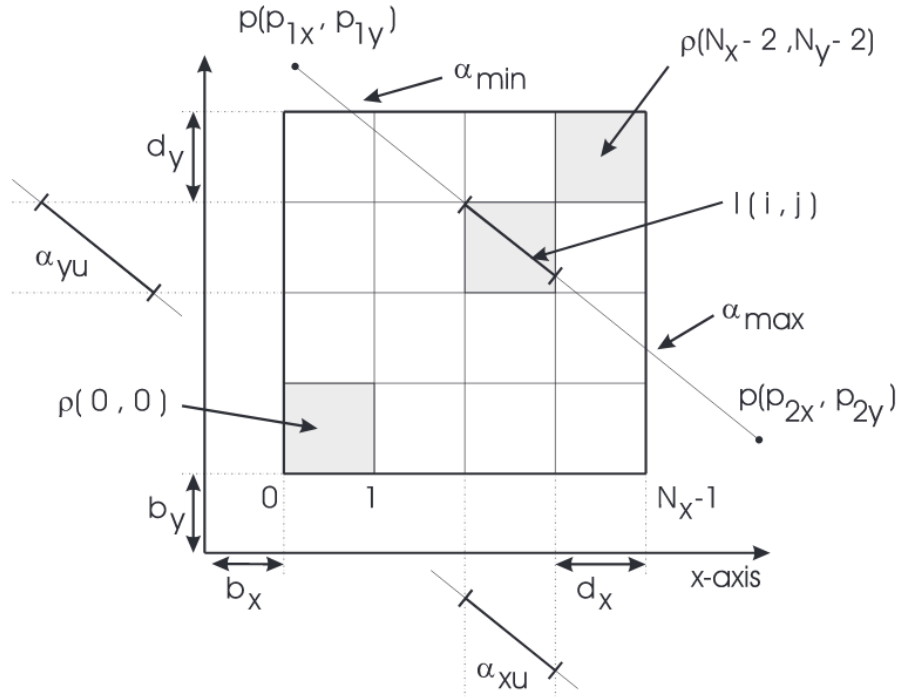


Figure 2.14 Schémas du problème de calcul du chemin radiologique exacte en 2D. Image tirée de (Jacobs, 1998).

sont les points d'entrée et de sortie du rayon dans le carré.

L'algorithme de Siddon, lui, consiste à trouver les valeurs des α pour lesquels le rayon croise tous les plans perpendiculaires à x et tous les plans perpendiculaires à y , entre le moment où le rayon pénètre le volume et le moment où il en sort. La différence entre deux α successifs donne directement la distance parcourue au sein d'un pixel. Enfin, il faut utiliser le paramètre α pour retrouver les coordonnées du pixel en question.

L'optimisation qu'apporte l'algorithme de Jacobs réside dans le fait qu'il n'ait pas besoin de calculer tous les α d'intersection pour ensuite retrouver les coordonnées des pixels. L'algorithme obtient les coordonnées du pixel traversé en même temps que l'on obtient les α associés. Ce calcul repose sur le fait que deux pixels traversés successivement sont forcément avoisinants, ce qui réduit le nombre d'opérations pour effectuer calcul des coordonnées d'un pixel avoisinant à l'autre à une simple incrémentation. Une fois le chemin radiologique calculé, il est possible de retrouver l'intensité reçue à l'aide des équations 2.3 et 2.1.

2.3.4 Recalage 2D/3D manuel par marqueurs

Une approche relativement simple pour effectuer un recalage 2D/3D est de procéder avec un marquage manuel (Bijhold, 1993). Il est nécessaire d'avoir calibré les caméras au préalable. La méthode de triangulation de ces points à partir de l'espace 2D vers l'espace 3D repose, comme pour le calibrage détaillé dans la section 2.3.2, sur la DLT. Cette fois-ci, nous devons résoudre pour trouver X plutôt que P . Étant donné des projections $x = \begin{pmatrix} u \\ v \end{pmatrix}$ et $x' = \begin{pmatrix} u' \\ v' \end{pmatrix}$ du point X sur les caméras de matrices P et P' , l'équation 2.8 peut être réarrangée de la manière suivante :

$$\underbrace{\begin{pmatrix} p_{11} - up_{31} & p_{12} - up_{32} & p_{13} - up_{33} \\ p_{21} - vp_{31} & p_{22} - vp_{32} & p_{23} - vp_{33} \\ p'_{11} - u'p'_{31} & p'_{12} - u'p'_{32} & p'_{13} - u'p'_{33} \\ p'_{21} - v'p'_{31} & p'_{22} - v'p'_{32} & p'_{23} - v'p'_{33} \end{pmatrix}}_A \underbrace{\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}}_X = \underbrace{\begin{pmatrix} up_{34} - p_{14} \\ vp_{34} - p_{24} \\ u'p'_{34} - p'_{14} \\ v'p'_{34} - p'_{24} \end{pmatrix}}_B \quad (2.13)$$

Deux correspondances donnent lieu à un système sur-contraint avec 4 équations et 3 inconnues, de la forme $AX = B$. Ce type d'équations admet une solution fermée, basé sur la pseudo-inverse de A , tel que $X = (A^T A)^{-1} A^T B$. Dans le cas où $A^T A$ est singulière, il est possible de faire une approximation de son inverse en éliminant les valeurs singulières qui sont beaucoup trop faibles. Son inverse est approximée de la manière suivante :

$$A^{-1} = (USV^T)^{-1} \approx VS_0^{-1}U^T \quad (2.14)$$

$$S_0 = \begin{cases} \frac{1}{s_i} & \text{si } s_i > t \\ 0 & \text{sinon} \end{cases}$$

La méthode décrite dans l'équation 2.14 porte le nom de SVD tronquée (Halko et al. 2009). La SVD tronquée n'est pas la seule manière de contourner le problème de mauvais conditionnement ou de singularité de la matrice $A^T A$. Il est aussi possible de procéder par régularisation de Tikhonov (Tikhonov et al. 1943). Dans ce cas-ci, la solution prend la forme suivante :

$$X = (A^T A + \lambda I)^{-1} A^T B \quad (2.15)$$

Le terme λ doit être choisi de manière à ce que $(A^T A + \lambda I)$ soit inversible et bien conditionnée. Le cas particulier de régularisation de Tikhonov illustré par l'équation 2.15 correspond à la régularisation $L2$ dans la littérature d'apprentissage machine.

À l'aide des méthodes décrites ci-dessus, la méthode de recalage par marqueurs consiste à

manuellement identifier les repères d'intérêt sur chacune des vues, et d'en faire une reconstruction par triangulation. Comme l'illustre la figure 2.15, il est possible de projeter à nouveau les points reconstruits afin de les comparer aux points marqués initialement, à des fins de validation.

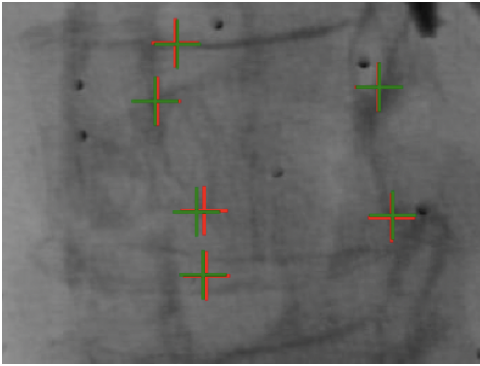
Une fois tous les repères reconstruits en 3D, il est nécessaire de trouver la transformation rigide optimale entre les repères de la modalité 3D et ceux qui ont été reconstruits à partir des images fluoroscopiques. Nous dénoterons par X et X' respectivement l'ensemble des points reconstruits en 3D à partir des images 2D et l'ensemble des points de la modalité 3D. Ces matrices sont de taille $3 \times n$ où n désigne le nombre de points à recalibrer. Notons qu'il faut au moins 3 points pour pouvoir effectuer un recalage rigide. Le problème de recalage se résume donc à minimiser l'équation suivante :

$$R^*, T^* = \operatorname{argmin}_{R, T} \|(RX + T) - X'\|_2 \quad (2.16)$$

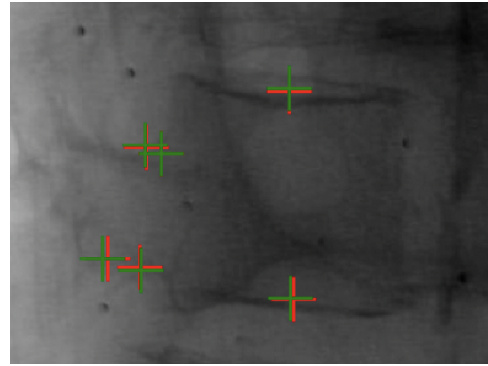
Dénotons par \bar{X} et \bar{X}' l'ensemble des points X et X' centrés sur leurs isobarycentres respectifs \bar{x} et \bar{x}' . Par SVD, les équations du recalage optimal sont les suivantes (Arun et al. 1987) :

$$R^* = VU^T \quad \text{tel que} \quad \bar{X}\bar{X}'^T = USV^T \quad (2.17)$$

$$T^* = -R^*\bar{x} + \bar{x}' \quad (2.18)$$



(a) Première vue



(b) Seconde vue

Figure 2.15 Exemple de triangulation de 6 repères anatomiques d'une vertèbre sur deux vues différentes. Les repères sont les centres des plateaux inférieur et supérieur, ainsi que les extrémités des pédicules. Les croix rouges désignent les points marqués manuellement alors que les points verts désignent les points reconstruits en 3D puis rétroprojetés en 2D pour validation.

2.3.5 Recalage 2D/3D par optimisation

Familles d'algorithmes de recalage 2D/3D

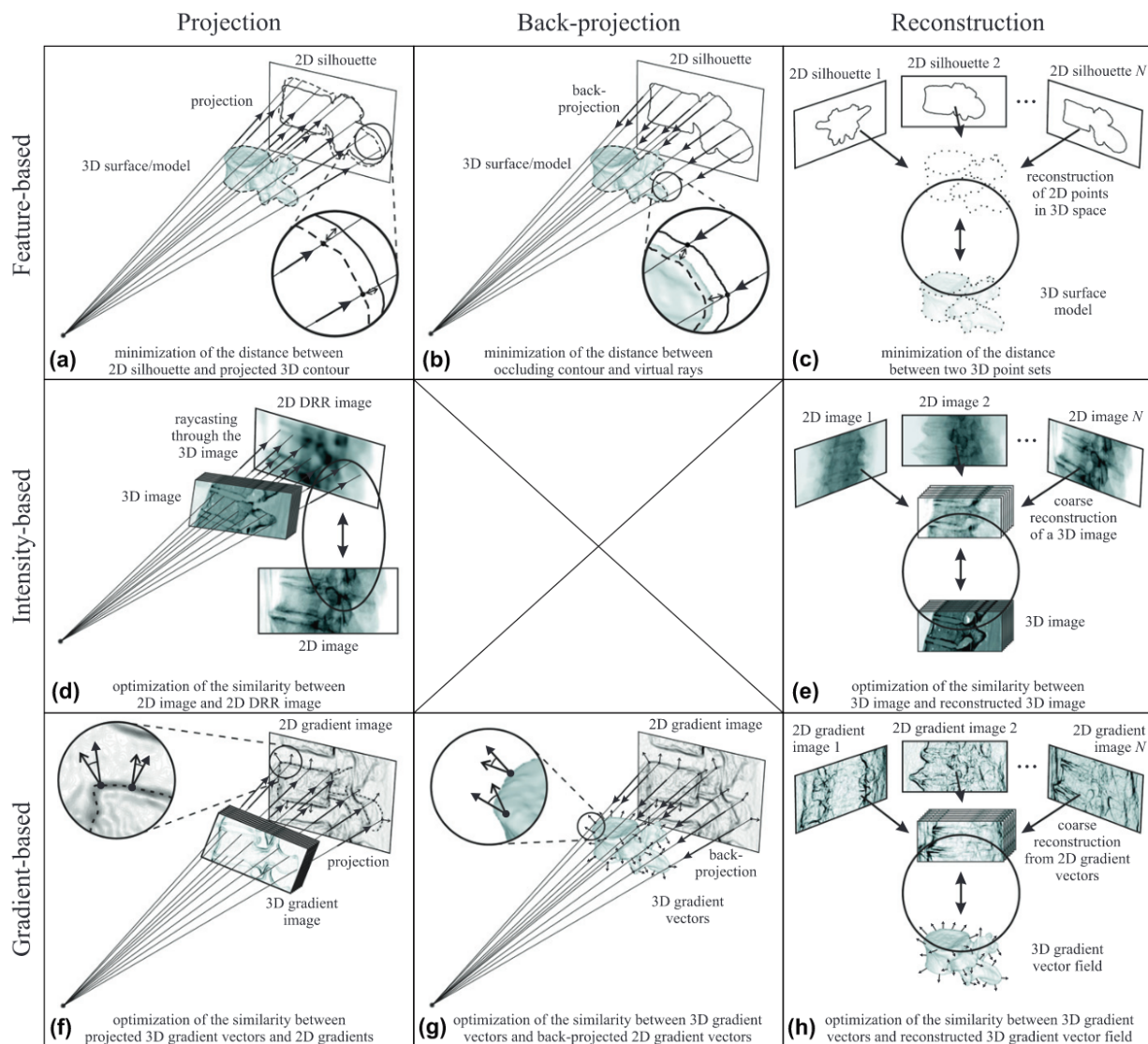


Figure 2.16 Récapitulatif des familles d'algorithmes de recalage 2D/3D, classés par objectif d'optimisation (intensité, géométrie ou gradients) et par type de transformation 2D/3D (projection, rétro-projection ou reconstruction). Image tirée de (Markelj et al. 2012)

Il existe plusieurs méthodes de recalage automatique et semi-automatique, mais elles reposent toutes sur une procédure d'optimisation pour l'un des trois éléments suivants : la géométrie, les intensités ou les gradients (Markelj, 2012).

Les méthodes de recalage par géométrie permettent d'optimiser soit des correspondances courbe-courbe (Bullit et al. 1999), ou des correspondances courbe-surface (Banks et al. 1996).

Le problème de correspondance point par point revient à la méthode de la section 2.3.4. Toutefois, lorsque ces correspondances ne sont pas connues au préalable, on a recours à des méthodes itératives d'optimisation, et notamment la méthode du point itératif le plus proche (ICP) (Besl et al. 1992). C'est une méthode qui, en deux étapes successives, estime les correspondances puis effectue le recalage et ce, jusqu'à la convergence des erreurs résiduelles. Le recalage par géométrie repose sur des modèles statistiques ou des segmentations de la structure d'intérêt (Fleute, 1999). La méthode par projection (Feldmar, 1997) consiste à projeter le modèle 3D sur l'image 2D. L'optimisation se fait alors par rapport aux contours de la silhouette projetée (Hermans, 2007). L'approche inverse revient à rétro-projeter la silhouette 2D en un faisceau de rayons 3D, puis de minimiser la distance orthogonale entre ces rayons et le modèle 3D (Lavalée, 1995).

Par ailleurs, le recalage peut se faire en optimisant des métriques associées aux intensités plutôt qu'à la géométrie. Une approche consiste à reconstruire le modèle d'intensités 3D à partir de plusieurs images 2D, puis de faire le recalage 3D en optimisant pour l'intensité (Zheng et al. 2006). L'approche contraire revient à effectuer le recalage en 2D plutôt qu'en 3D et repose sur la technique de la radiographie numériquement reconstruite (Sherouse et al. 1990), qui a été détaillée précédemment à la section 2.3.3.

Finalement, la troisième famille de méthodes de recalage 2D/3D est basée sur les gradients d'intensité (Markelj et al. 2012). Une première manière consiste à projeter les gradients d'intensité du volume 3D vers l'image 2D, afin d'optimiser une mesure de similarité entre ces gradients projetés et les gradients d'intensité de l'image 2D fixe (Livnyatan et al. 2002). Une seconde manière revient à rétro-projeter les gradients de l'image 2D en un faisceau de rayons, avec un rayon pour chacune des deux extrémités de chaque vecteur, et de minimiser une mesure de similarité entre les gradients rétro-projetés et les gradients du volume 3D (Tomazevic et al. 2003). La dernière méthode revient à reconstruire les gradients des images 2D dans l'espace 3D pour, là aussi, minimiser une mesure de similarité entre les deux champs vectoriels 3D (Markelj et al. 2008).

Fonctions d'objectif

Dans cette section, nous présenterons les métriques en tant que tel. Contrairement à ce qu'indique le schéma de la figure 2.10, ces métriques peuvent indiquer une similarité comme elles peuvent indiquer une dissimilarité. Il faut donc ajuster le signe de la métrique pour que celle-ci soit une métrique de dissimilarité.

La métrique d'Information des Gradients (Otake et al. 2013) est une métrique de similarité des gradients d'images 2D. Cette métrique est obtenue en sommant, pour chaque paire de vecteurs, les cosinus normalisés dans l'intervalle $[0, 1]$, avec une pondération par la norme du

vecteur de plus faible norme. On dénote par $\nabla_i X$ le gradient de X au pixel d'indice i .

$$\text{IG}(X, Y) = \frac{1}{N} \sum_{i=0}^{N-1} \min(|\nabla_i X|, |\nabla_i Y|) \frac{\cos(\theta_i) + 1}{2} \text{ avec } \theta_i = \frac{\nabla_i X \cdot \nabla_i Y}{\|\nabla_i X\| \|\nabla_i Y\|} \quad (2.19)$$

On peut remarquer que cette métrique ne pénalise pas les gradients alignés à l'opposé, à cause de la normalisation sur $[0, 1]$, mais qu'elle favorise surtout les gradients alignés dans le même sens. La pondération par la plus faible norme fait en sorte que cette métrique donne plus d'importance aux gradients qui ont la même orientation et qui sont tous deux de forte norme.

Une autre métrique, la Correlation des Gradients (Hipwell et al. 2003), revient à sommer la moyenne des corrélations croisées normalisées des gradients selon x et selon y .

$$\text{CG}(X, Y) = \frac{1}{2} \left(\frac{\text{Cov}(\nabla_x X, \nabla_x Y)}{\sqrt{\text{Var}(\nabla_x X) \text{Var}(\nabla_x Y)}} + \frac{\text{Cov}(\nabla_y X, \nabla_y Y)}{\sqrt{\text{Var}(\nabla_y X) \text{Var}(\nabla_y Y)}} \right) \quad (2.20)$$

Cette métrique, étant donné qu'elle ne prends pas directement en compte les différences de gradients en norme et en orientation, permet d'accommoder aux différences entre ces gradients, tant que ceux-ci restent linéairement corrélés (De Silva et al. 2016).

La métrique d'orientation des gradients (De Silva et al. 2016) se différencie des précédentes en ne prenant en compte que les orientations des gradients. Afin d'éliminer le bruit d'échantillonnage, la méthode ne considère que les gradients à norme au-delà des seuils t_X et t_Y . La constante de normalisation est, elle aussi, seuillée à une valeur minimale N_{LB}

$$\text{GO}(X, Y) = \frac{1}{\max(N, N_{LB})} \sum_{i \in V} 1 - \frac{\ln(1 + |\cos^{-1}(\frac{\nabla_i X \cdot \nabla_i Y}{\|\nabla_i X\| \|\nabla_i Y\|})|)}{2} \quad (2.21)$$

$$\text{avec } V = \{i \in \{1, \dots, N\} \mid |\nabla_i X| > t_X \cap |\nabla_i Y| > t_Y\}$$

D'autres métriques, reposant sur les intensités plutôt que leurs gradients, sont aussi utilisées dans la littérature. La corrélation croisée normalisée a une expression de la forme suivante (Penney et al. 1998) :

$$\text{NCC}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} \quad (2.22)$$

Cette métrique permet de quantifier la relation linéaire entre X et Y . L'Information Mutuelle, dans le cadre du recalage 2D/3D (Clippe et al. 2003), est une métrique issue de la théorie de l'information (Shannon 1948) et permet de quantifier la dépendance statistique entre X et Y . Les densités de probabilités dans l'équation 2.23 sont estimées par histogrammes ou par

noyaux (Parzen 1962).

$$\text{IM}(X, Y) = \int p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) dx dy \quad (2.23)$$

Algorithmes d'optimisation

Algorithme 1 : Nelder-Mead. Adapté de (Zisserman, 2011)

Résultat : Point de convergence des sommets du simplexe

Initialiser le simplexe avec $n + 1$ points dans le voisinage du point initial x_0 ;

tant que *Non convergence des sommets du simplexe* **faire**

Réordonner tous les points du simplexe par ordre décroissant tel que :

$f(x_{n+1}) > \dots > f(x_1)$;

Évaluer un point candidat x_r par *réflexion* tel que :

$x_r \leftarrow \bar{x} + \alpha(\bar{x} - x_{n+1})$ pour $\bar{x} = \frac{1}{n+1} \sum_{i=1}^{n+1} x_i$ et $\alpha > 0$;

si $f(x_1) < f(x_r) < f(x_n)$ **alors**

| $x_{n+1} \leftarrow x_r$;

sinon si $f(x_r) < f(x_1)$ **alors**

| Évaluer x_e par *expansion* tel que :

$x_e \leftarrow x_r + \beta(x_r - \bar{x})$ pour $\beta > 0$;

si $f(x_e) < f(x_r)$ **alors**

| $x_{n+1} \leftarrow x_e$

sinon

| $x_{n+1} \leftarrow x_r$

fin

sinon

| Évaluer x_c par *contraction* tel que :

$x_c \leftarrow \bar{x} + \gamma(x_{n+1} - \bar{x})$ pour $0 < \gamma < 1$;

tant que $f(x_c) \geq f(x_{n+1})$ **faire**

| $x_c \leftarrow \bar{x} + \gamma(x_c - \bar{x})$

fin

$x_{n+1} \leftarrow x_c$

fin

fin

Dans le cadre de cette section, nous nous limiterons aux algorithmes d'optimisation locale, c'est-à-dire au problème de la recherche d'optimum dans le voisinage d'une solution initiale. L'algorithme de Nelder-Mead (Nelder, Mead, 1965) est une méthode de minimisation sans gradient pour des fonctions scalaires à n variables. Cet algorithme est initialisé avec $n + 1$ points, formant une enveloppe convexe autour du point d'initialisation (Gao et al. 2012). Cette enveloppe convexe porte le nom de simplexe (Rudin, 1964). Trois opérations sont possibles : la réflexion, l'expansion et la contraction (Zisserman, 2011). Ces opérations se suc-

cèdent selon l'algorithme 1 jusqu'au point de convergence. Cet algorithme est récurrent dans la littérature du recalage 2D/3D (Fleute et al, 1999)(Fukoaka et al, 1999)(Markelj et al, 2012).

L'algorithme de Powell (Powell, 1964) est un autre algorithme d'optimisation sans gradient. Celui-ci va chercher à partiellement minimiser la fonction d'intérêt par rapport n directions différentes, avec n étant le nombre de variables qu'admet la fonction d'intérêt (Mathews 1992). Cet algorithme ajuste non seulement ses estimations des minimums partiels, mais il ajuste aussi ses estimations des directions de recherche locale. Ces direction sont initialisées comme étant les n directions orthogonales de la base standard. À chaque itération de l'algorithme, on remplace l'une des directions de recherche par la direction du déplacement de la précédente solution à la solution courante. Ainsi, à chaque nouvelle estimation de solution, on incorpore la direction du dernier déplacement parmi les directions de recherche de minimum partiel. L'intuition derrière cette approche est que, étant donné que la solution d'une itération est toujours meilleure que celle de la précédente, la direction du déplacement est supposée privilégiée pour trouver de meilleurs minimums (Mathews, 1992). Finalement, nous

Algorithme 2 : Powell. Adapté de (Mathews, 1992)

Résultat : Point de convergence des minimums successifs

```

 $x \leftarrow x_0;$ 
 $u_1 \dots u_n \leftarrow$  base standard de  $\mathbb{R}^n$ ;
tant que Non convergence de  $x$  et de  $f(x)$  faire
     $p \leftarrow x;$ 
    pour  $i \leftarrow 1$  à  $n$  faire
         $p \leftarrow p + \lambda^* u_i$  tel que  $\lambda^* = \operatorname{argmax}_{\lambda} (f(p + \lambda u_i));$ 
    fin
     $u_1, \dots, u_{n-1}, u_n \leftarrow u_2, \dots, u_n, (p - x);$ 
     $x \leftarrow x + \lambda^* u_n$  tel que  $\lambda^* = \operatorname{argmax}_{\lambda} (f(x + \lambda u_n))$ 
fin

```

présenterons la méthode de descente du gradient qui, elle aussi, est évoquée dans la littérature comme algorithme d'optimisation pour le recalage 2D/3D (Fu et al. 2008). La méthode repose sur l'estimation du gradient en évaluant la fonction f à plusieurs points et en utilisant les développements limités pour estimer son gradient. On peut judicieusement évaluer f à des points où on ira annuler les termes d'ordre plus élevés dans le développement limité pour

Algorithme 3 : Broyden–Fletcher–Goldfarb–Shanno (BFGS). Adapté de (Gramfort, 2014)

Résultat : Point de convergence

$x_0 \in \mathbb{R}^n$, H_0 définie positive;

pour $i \leftarrow 0$ **à** $n_{iterations} - 1$ **faire**

si $\|\nabla_{x_i} f\| < \epsilon$ **alors**

retourner x_i

$d_i \leftarrow -H_i^{-1} \nabla_{x_i} f$;

$\rho_i \leftarrow \operatorname{argmin}_{\rho} f(x_i + \rho d_i)$;

$x_{i+1} \leftarrow x_i + \rho_i d_i$;

$s_i \leftarrow \rho_i d_i$;

$y_i \leftarrow \nabla_{x_{i+1}} f - \nabla_{x_i} f$;

$H_{i+1} = H_i + \frac{y_i y_i^T}{y_i^T s_i} + \frac{H_i s_i s_i^T H_i}{s_i^T H_i s_i}$;

fin

retourner $x_{n_{iterations}}$;

avoir de meilleurs approximations du gradient (Mansfield, 2009), comme par exemple :

$$\begin{aligned}
 f(x+h) &\approx f(x) + \nabla_x f^T h + \frac{1}{2} h^T H_x h + o(\|h\|^2) \\
 f(x-h) &\approx f(x) - \nabla_x f^T h + \frac{1}{2} h^T H_x h + o(\|h\|^2) \\
 f(x+h) - f(x-h) &\approx 2 \nabla_x f^T h + o(\|h\|^2)
 \end{aligned} \tag{2.24}$$

La règle de descente du gradient prend donc la forme de $x_{k+1} = x_k - \alpha \nabla_{x_k} f$ où α désigne le taux d'apprentissage (Gramfort, 2014). Il existe aussi des méthode qui utilisent des calculs de la matrice Hessienne H_x , et la règle de descente du gradient prend alors la forme de $x_{k+1} = x_k - H_x^{-1} \nabla_x f$ (Zisserman, 2011). Le calcul de la Hessienne est toutefois très coûteux, il existe donc des méthodes pour approximer cette Hessienne, dont la méthode Broyden–Fletcher–Goldfarb–Shanno (BFGS) (Fletcher, 2013). La méthode est présentée dans l'algorithme 3.

2.3.6 Recalage de bout en bout par réseaux profonds

En apprentissage supervisé, les réseaux de neurones profonds peuvent être abstraits et vus comme étant des fonctions paramétriques qui interpolent la relation entre des entrées et des sorties de dimensions élevées. Le problème de recalage peut donc être interprété en tant que tel, si l'on dispose d'un jeu de données où le recalage optimal est connu. Un exemple de méthode basée sur des réseaux siamois a été proposée dans la littérature (Beaupré et Bilodeau, 2019). Celle-ci repose sur l'encodage de parcelles d'images multimodales par ces

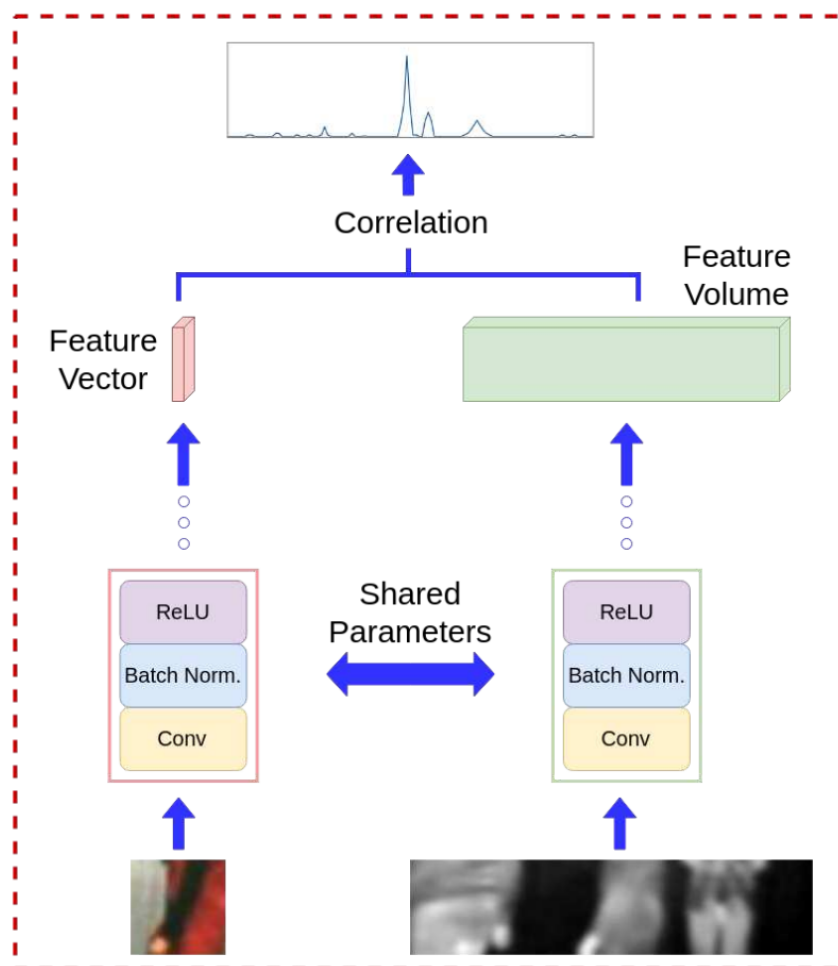


Figure 2.17 Architecture de réseaux de neurones siamois pour le recalage d'images. Une couche de probabilités log-softmax est ajoutée au vecteur de corrélations. Image tirée de © IEEE 2019 (Beaupré et Bilodeau, 2019).

mêmes réseaux profonds. Des vecteurs de probabilités d'emplacement d'une petite parcelle au sein d'une autre parcelle de plus grande taille sont produits, à partir de ces encodages. Cela revient à un problème de classification, où les différentes classes correspondent aux différents emplacements possibles de la petite parcelle au sein de la plus grande. Cette méthode a été appliquée au problème de recalage d'images en couleur avec des images infrarouges. Le recalage est fait dans les deux sens pour assurer une consistance entre la transformée directe et son inverse.

D'autres méthodes sont basées sur de la régression par réseaux convolutifs profonds (Li et al. 2017). Ces méthodes-là sont inspirées des méthodes de recalage par optimisation et par estimation discrète du gradient. La différence-clé réside dans le fait que la rétro-propagation

est utilisée comme algorithme d’optimisation. Les calculs effectués sont sauvegardés dans des graphes, alors que des réseaux profonds sont utilisés pour mémoriser la procédure d’optimisation durant l’entraînement.

2.4 Méthodes et modèles de prédiction pour la synthèse IRM-CT

2.4.1 Vue globale sur l’état de l’art

L’imagerie par résonance magnétique offre de meilleurs contrastes que la tomodensitométrie, ainsi qu’une meilleure visibilité des tissus et des organes à risque (Slobodan, 2012). Toutefois, pour une bonne visibilité des structures osseuses et pour un bon recalage 2D/3D, le CT reste indispensable (Patel et al. 2008). De nombreux travaux ont été effectués durant les dernières années pour la synthèse de CT synthétiques à partir d’IRM (Price et al. 2017, Edmund et al. 2017, Johnstone et al. 2018), à des fins de recalage 2D/3D.

Une première catégorie de méthodes repose sur le recalage déformable entre un atlas de tomodensitométrie et une IRM (Uh et al. 2014), alors que d’autres méthodes utilisent des séquences d’IRM à très court temps d’écho comme substitut au CT (Edmund et al. 2017).

En ce qui concerne les méthodes basées sur l’apprentissage machine, des méthodes basées sur des modèles d’apprentissage par instance, tel que l’algorithme des N voisins les plus proches (Van der Bom et al. 2011), ont été proposées. D’autres modèles basés sur de la régression par arbres aléatoires (Huynh et al. 2015) et par mélanges de gaussiennes (Johansson et al. 2018) ont aussi été proposés.

Enfin, ce sont les dernières avancées en terme d’apprentissage profond qui ont le plus marqué les dernières tendances pour ce type d’applications en particulier (Lei et al. 2019). En effet, comme nous le verrons dans la section 2.4.3, l’avantage de l’apprentissage profond est qu’il met à disposition des modèles capables d’apprendre des représentations perceptuelles à plus haut niveau d’abstraction, contrairement aux modèles classiques de statistiques ou d’apprentissage machine (Johnson et al. 2016), qui restent limités à de l’estimation point par point. Ceci a l’avantage de produire des images plus vraisemblables, et de réduire les artefacts d’interpolation que l’on retrouve souvent avec des modèles plus simples. Des approches d’apprentissage profond classique pour la génération de CT synthétiques ont été proposées (Li et al. 2014), tandis que d’autres reposent sur des fonctions de pertes adverses (Wolterink et al. 2017, Nie et al. 2017).

Les modèles d’apprentissage profond classique effectuent de l’apprentissage supervisé, ce qui nécessite donc des images d’IRM et de CT qui soient alignées. De plus, malgré l’apprentissage de représentation au sein des filtres des réseaux convolutifs, l’optimisation se fait tout de même point par point, et le bruit statistique associé à cette correspondance point par

point influe sur l'apprentissage. Les modèles adverses génératifs ont l'avantage d'imposer des contraintes perceptuelles à plus haut niveau sur l'apparence globale des images synthétisées, ce qui tend à corriger les problèmes d'artefacts au sein des images générées.

Dans la suite de ce mémoire, nous présenterons une brève revue des concepts de base apprentissage machine pour le problème de régression et de translation d'image, avant d'exposer les méthodes d'apprentissage profond d'intérêt pour ce travail.

2.4.2 Méthodes d'apprentissage machine

Régression linéaire

Le modèle de régression le plus simple est la régression linéaire. Celle-ci s'exprime de la manière suivante :

$$\hat{y} = w^T x + b \quad (2.25)$$

Régression par machine à vecteur de support

La machine à vecteur de support (SVM) linéaire (Vapnik, 1963), celle-ci va chercher à trouver un hyperplan qui soit le plus proche possible de tous les points du jeu de données d'entraînement. La distance entre cet hyperplan et ces points ne doit pas dépasser un ϵ que l'on appelle marge (Géron, 2017).

Pour mieux comprendre les machines à vecteur de support et la technique des noyaux, il faut utiliser la formulation duale (Zisserman, 2015) :

$$\begin{aligned} \text{Formulation primale : } \hat{y} &= w^T x + b \\ \text{Formulation duale : } \hat{y} &= \sum_{i=1}^N \alpha_i y_i (x_i^T x) + b \end{aligned} \quad (2.26)$$

Comme illustré figure 2.18, l'objectif d'optimisation de la machine à vecteur de support est de compter le plus d'exemples d'entraînement possible à l'intérieur de la marge ϵ . Ceci se traduit par l'objectif d'optimisation suivant (Sayad 2019) :

$$\begin{aligned} w^* &= \operatorname{argmin}_w \|w\|^2 \\ \text{Tel que : } \forall i \in \{0, \dots, n\}, \quad |y_i - w^T x + b| &< \epsilon \end{aligned} \quad (2.27)$$

Un avantage de la méthode de régression par machine à vecteur de support est qu'elle permet de faire de la régression sur des jeux de données qui ne suivent pas nécessairement un modèle

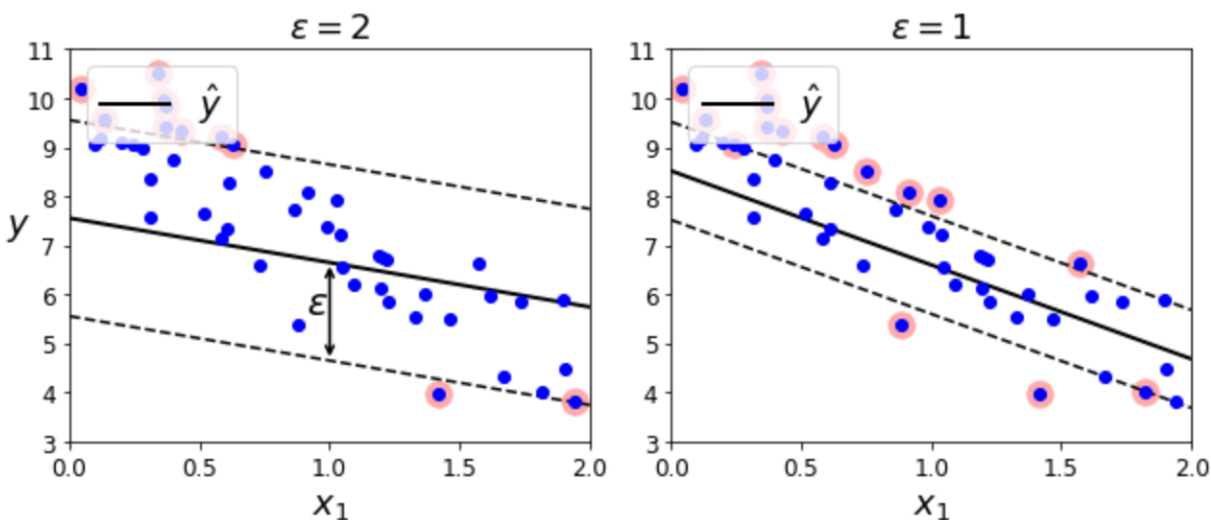


Figure 2.18 Exemple de machine à vecteur de support linéaire. Figure générée avec (<https://github.com/ageron/handson-ml>)

linéaire, et ce, à l'aide de la méthode des noyaux (Vapnik, 1963). Celle-ci revient à faire une application du jeu de données vers un espace vectoriel à dimensions plus élevées afin de pouvoir rendre ce jeu de données-là linéaire dans cet espace vectoriel. Un noyau est une application $k : x_i, x_j \mapsto \phi(x_i)^T \phi(x_j)$. La régression par machine à vecteur de support prend donc la forme :

$$\hat{y} = \sum_{i=1}^N \alpha_i y_i (\phi(x_i)^T \phi(x)) + b \quad (2.28)$$

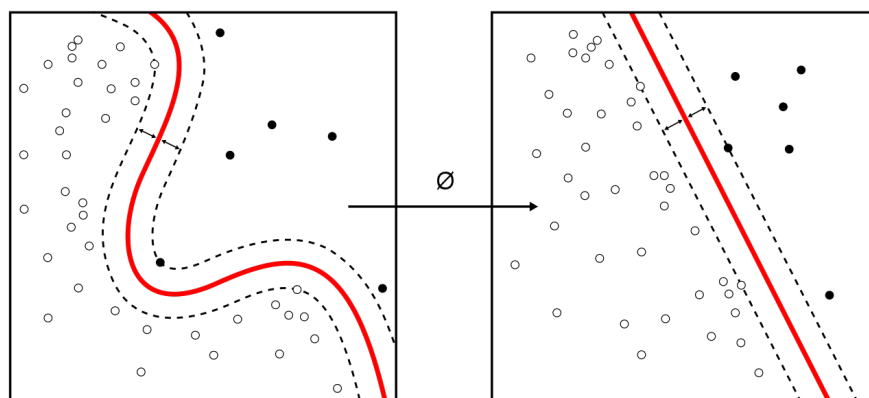


Figure 2.19 Exemple de jeu de données rendu linéairement séparable après l'application d'un noyau ϕ . Wikimedia Commons 2019

Régression par mélange de gaussiennes

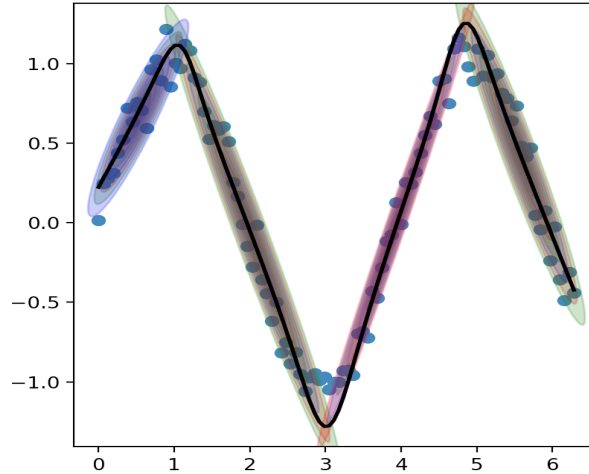


Figure 2.20 Modèle de régression par mélange de gaussiennes. Figure générée avec (<https://github.com/AlexanderFabisch/gmr>)

Le modèle de mélange de gaussiennes (Murphy 2012) est un modèle qui cherche à modéliser un jeu de données comme ayant été conditionnellement généré par plusieurs loi normales possibles.

Mathématiquement, le modèle du mélange de gaussiennes s'exprime de la manière suivante (Murphy 2012) :

$$p(y_i|\mathbf{x}_i, \boldsymbol{\theta}) = \sum_k p(z_i|\mathbf{x}_i, \boldsymbol{\theta})p(y_i|z_i, \mathbf{x}_i, \boldsymbol{\theta}) \quad (2.29)$$

À l'aide de la formulation de $p(y_i|\mathbf{x}_i, \boldsymbol{\theta})$, il est possible d'apprendre les paramètres $\boldsymbol{\theta}$ par estimation de maximum de vraisemblance, par estimation avec à priori Bayésien, ou par descente du gradient (Koller 2009).

2.4.3 Méthodes d'apprentissage profond

Perceptron multi-couches

Le perceptron multi-couche (Rosenblatt 1961) est un modèle de réseau de neurones artificiels. Chaque couche comporte une opération matricielle affine suivie de l'application d'une fonction f non-linéaire. Une couche effectue donc une opération de la forme $f(W^T x + b)$. Le perceptron multi-couche se résume en la composition fonctionnelle de plusieurs de ces couches-là. Le choix de la taille des matrices W détermine la taille du vecteur en sortie de chaque couche. La taille de la sortie de chaque couche doit être compatible avec la taille à

l'entrée de celle qui la succède. Pour les couches qui ne sont ni en entrée ni en sortie, on parle de couches cachées (Géron 2017).

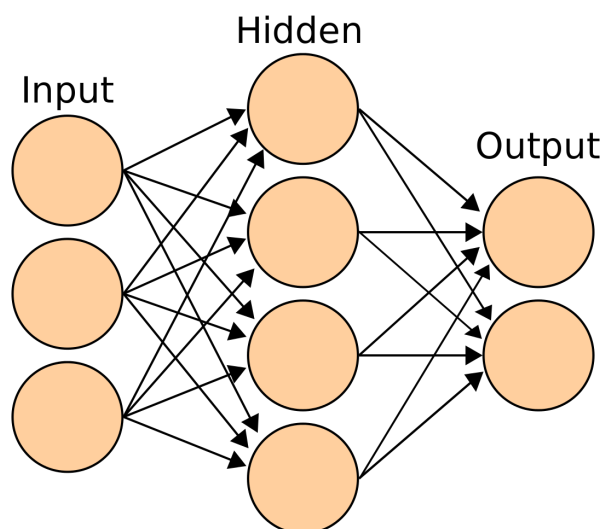


Figure 2.21 Schémas d'un perceptron multi-couche. Wikimedia Commons 2019

Réseaux de neurones convolutifs

Les réseaux de neurones convolutifs (LeCun et al. 1998) suivent aussi le paradigme de couches qui se succèdent. Toutefois, contrairement au perceptron multi-couches, un réseau convolutif incorpore une opération de convolution plutôt qu'une multiplication matricielle. Une couche de réseau convolutif effectue donc une opération de la forme $y = f(W * x + b)$ où $*$ dénote l'opération de convolution. La puissance des réseaux convolutifs est non seulement due au fait qu'ils se prêtent bien à l'analyse d'images par nature (Witten et al. 2016), mais c'est aussi dû au fait qu'une convolution nécessite bien moins de paramètres qu'une multiplication matricielle complète (Witten et al. 2016). En effet, la convolution est elle-même souvent implémentée sous forme de matrices de Toeplitz T à paramètres partagés (Witten et al. 2016), tel que $f(x) = Tx + b$. Par exemple, une matrice de Toeplitz pour une convolution de

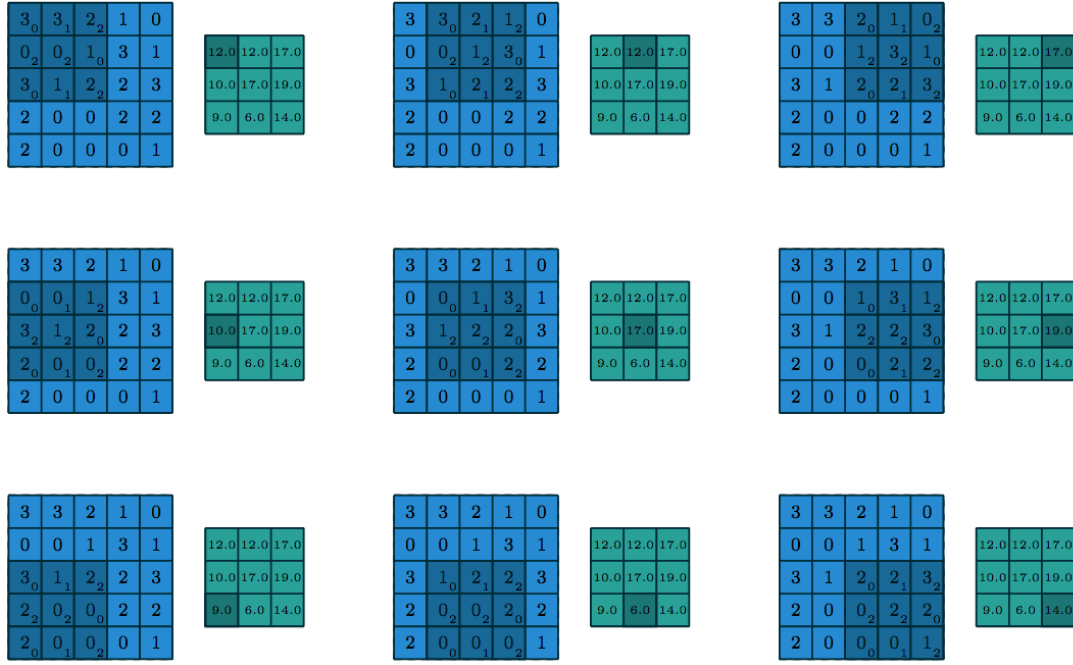


Figure 2.22 Illustration de la convolution 2D. Image tirée de (Dumoulin et Visin, 2016)

taille 3x3 a 6 paramètres avec une image de taille 6x6 s'écrit de la forme :

$$\begin{pmatrix}
 a & b & c & 0 & 0 & 0 & d & e & f & 0 & 0 & 0 & g & h & i & 0 & \dots & \dots & \dots & \dots & 0 \\
 0 & a & b & c & 0 & 0 & 0 & d & e & f & 0 & 0 & 0 & g & h & i & 0 & \dots & \dots & \dots & 0 \\
 \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
 \vdots & \dots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
 \vdots & \dots & \dots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
 \vdots & \dots & \dots & \dots & 0 & a & b & c & 0 & 0 & 0 & d & e & f & 0 & 0 & 0 & g & h & i & 0 \\
 0 & \dots & \dots & \dots & \dots & 0 & a & b & c & 0 & 0 & 0 & d & e & f & 0 & 0 & 0 & g & h & i
 \end{pmatrix} \quad (2.30)$$

Cette convolution a 9 paramètres, contrairement à la multiplication matricielle qui en a 384. Il faut noter que les termes de chaque ligne de T ne sont pas toujours translatés à droite d'une colonne. En effet, chaque fois que le filtre arrive à l'extrémité droite de l'image, il faut s'assurer de revenir au début de la ligne suivante de cette image. Cela revient à sauter un nombre de colonnes égal à la différence entre la taille du filtre et la taille de l'image.

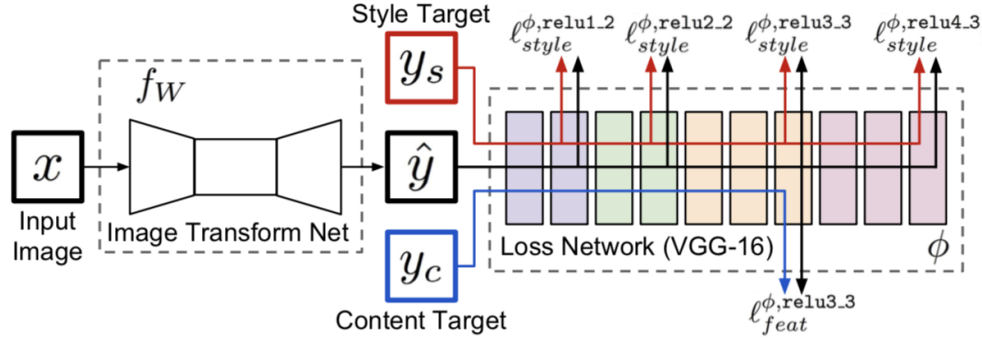


Figure 2.23 Architecture de transfert neuronal de style avec un ResNet qui apprend la procédure d'optimisation. Image tirée de (Johnson et al. 2016)

Transfert de domaine d'image

Le problème de de translation d'image consiste à transformer une image d'une modalité à une autre, tout en préservant le contenu et l'information qui est commune à ces deux modalités. Une première approche, appliquée aux arts visuels, a été introduite par (Gatys et al. 2015) : c'est le transfert neuronal de style.

Cette approche repose sur l'idée-clef qu'un réseau de neurones, pour pouvoir faire de la classification, a du apprendre une certaine représentation de ce qu'il observe, et l'encoder dans les filtres de ses couches cachées (Gatys et al. 2015). Une méthode pour comprendre la représentation qu'apprennent les réseaux de neurones de classification a été proposée par (Zeiler et al. 2013). Cette méthode consiste à inspecter les activations des couches intermédiaires et de voir quelles parties de l'image d'entrée sont les plus responsables pour ces fortes activations. Pour le transfert de style par neurones (Gatys et al. 2015), l'objectif d'optimisation est découpé en deux composantes : le contenu et le style. Le réseau profond utilisé pour encoder le style est de type VGG (Simonyan, Zisserman, 2014). Le réseau VGG pré-entraîné constitue une fonction de perte en soi (Gatys et al. 2015), puisque celui-ci aura appris les fonctions perceptuelles de base pour pouvoir de la classification. Le travail de (Gatys et al. 2015) rapporte que la profondeur corrèle avec des attributs stylistiques à plus haut niveau d'abstraction, contrairement aux première couches où beaucoup de détail et de texture et de géométrie sont encore conservés.

Cette méthode donne d'impressionnants résultats, mais elle est reste basée sur l'optimisation plutôt que sur l'apprentissage. Par ailleurs, avec la forme sous laquelle la méthode a été originalement proposée, le transfert de style n'utilise qu'une seule image de style pour effectuer la procédure d'optimisation. Cela ne garantit donc pas d'apprendre des éléments de style qui ne sont pas présents dans cette image-là, mais qui auraient très bien pu l'être.



Figure 2.24 Exemple de photographies transformées dans des styles de tableaux. Images tirées de © IEEE 2015 (Gatys et al. 2015)

Une approche complémentaire à celle précédemment discutée vient palier au problème de non-apprentissage, en introduisant un réseau de neurones qui va capturer et apprendre la procédure d'optimisation de transfert neuronal de style (Johnson et al. 2016).

Cette méthode utilise le même paradigme de décomposition de la fonction de perte en une fonction de style et une fonction de contenu. La différence entre cette méthode et la précédente est due à l'interposition d'un ResNet (He et al. 2016) entre l'image d'entrée et le réseau VGG. Ce ResNet génère une prédiction \hat{y} , et c'est sur celle-ci que s'effectuera l'optimisation. L'effet que cela peut avoir est que le ResNet apprenne et mémorise cette procédure d'optimisation (Johnson et al. 2016). Après l'entraînement, on utilise le ResNet en isolation pour effectuer le transfert de domaine d'image.

Réseaux adverses génératifs

Les réseaux génératifs adverses (Goodfellow et al. 2015) est un type d'architecture qui couple deux types de réseaux : un réseau discriminatif et un réseau génératif. Les deux sont entraî-

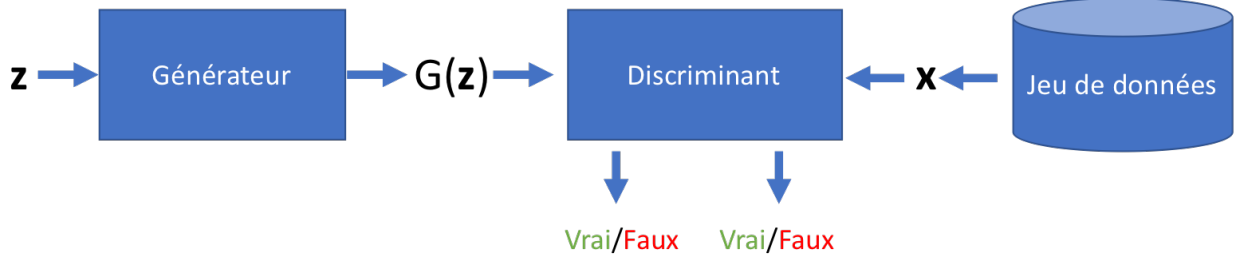


Figure 2.25 Architecture d'un réseau adverse génératif

nés conjointement de manière adverse, avec pour objectif que le réseau discriminatif puisse distinguer les vraies données des données générées par le générateur. Ce dernier est entraîné pour tromper le discriminateur. Il s'agit d'un jeu adverse de type minimax (Goodfellow et al. 2015) où le premier joueur, le réseau discriminatif, cherche à maximiser la précision de ses prédictions alors que le second joueur, le générateur, cherche à la lui minimiser. Le générateur utilise un bruit \mathbf{z} pour générer les données synthétiques $G(\mathbf{z})$. La formulation mathématique est la suivante :

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})} [\log(D(\mathbf{x}))] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))] \quad (2.31)$$

Le but de la procédure d'entraînement de ce type d'architecture est de pouvoir atteindre le point où le discriminateur est bien capable de classifier les vraies données comme étant vraies, alors que le générateur arrive tout de même à la tromper (Goodfellow et al. 2015). On peut alors isoler le générateur et l'utiliser en tant que tel pour générer des données synthétiques par échantillonnage de \mathbf{z} . L'algorithme d'entraînement d'un réseau adverse est le suivant : La

Algorithme 4 : Réseau adverse génératif. Adapté de (Goodfellow et al., 2014)

pour $i \leftarrow 1$ à $n_{iterations}$ **faire**

Échantillonner m vecteurs de bruit aléatoire $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m\}$;

Échantillonner m exemples $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$;

Mettre à jour les paramètres du réseau discriminatif en maximisant la fonction suivante :

$$\mathbb{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})} [\log(D(\mathbf{x}))] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))];$$

Échantillonner m vecteurs de bruit aléatoire $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m\}$;

Mettre à jour les paramètres du réseau discriminatif en minimisant la fonction suivante :

$$\mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

fin

littérature (Goodfellow et al. 2014) rapporte que le terme $\log(1 - D(G(\mathbf{z})))$ soit difficile à

minimiser pour des raisons de faible gradient. L'approche proposée par (Goodfellow et al., 2014), et qui est très adoptée dans la littérature, est de plutôt maximiser $\log(D(G(\mathbf{z})))$.

La version du GAN que nous présentons utilise une fonction de perte d'entropie croisée. D'autres variantes utilisent des fonctions de perte, telle qu'une perte par moindres carrés (Mao et al., 2017) ou une perte de Wasserstein (Arjovsky et al., 2017). La perte de Wasserstein est rapportée avoir des gradients plus lisses que la plupart des autres pertes (Gulrajani et al., 2017) et donc plus facile d'optimisation. Celle-ci consiste à optimiser la distance suivante :

$$W(\mathbb{P}_x, \mathbb{P}_y) = \inf_{\gamma \in \Pi(\mathbb{P}_x, \mathbb{P}_y)} \mathbb{E}_{(x,y) \sim \gamma(x,y)} [\|x - y\|] \quad (2.32)$$

La notation $\Pi(\mathbb{P}_x, \mathbb{P}_y)$ désigne l'ensemble de densités de probabilité conjointes qui se marginalisent aux densités \mathbb{P}_x et \mathbb{P}_y . Intuitivement, l'équation traduit une espérance de la distance $\|x - y\|$ pondérée par la fonction $\gamma(x, y)$. Cette fonction traduit une densité de masse. L'équation 2.32 exprime donc une intégrale des distances de transport, pondérées par les densités de masse transportées. La distance de Wasserstein est donc la distance la distance de transport pondérée la plus faible.

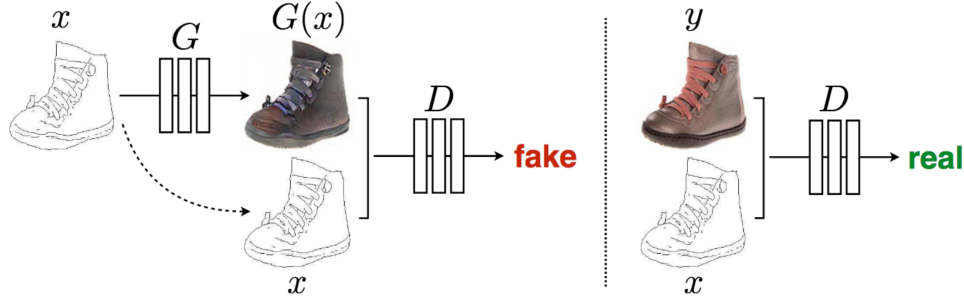


Figure 2.26 Architecture du cGAN Pix2Pix. Image tirée de (Isola et al. 2017) © IEEE 2017

Dans la section précédente, nous avons présenté des modèles pour la translation d'image. Une d'architectures pour le transfert de domaine d'image que nous explorerons dans cette partie est basée sur les réseaux adverses génératifs. Le concept est le même que pour les réseaux adverses génératifs classiques, à la différence près que nous avons une modalité X plutôt qu'un bruit Z en entrée du générateur, et que l'on optimise pour avoir une modalité Y à sa sortie.

L'architecture Pix2Pix (ou cGAN) proposée par (Isola et al. 2017) s'inscrit dans ce paradigme. Le générateur adopte une architecture en UNet (Ronneberg et al. 2015). Cette

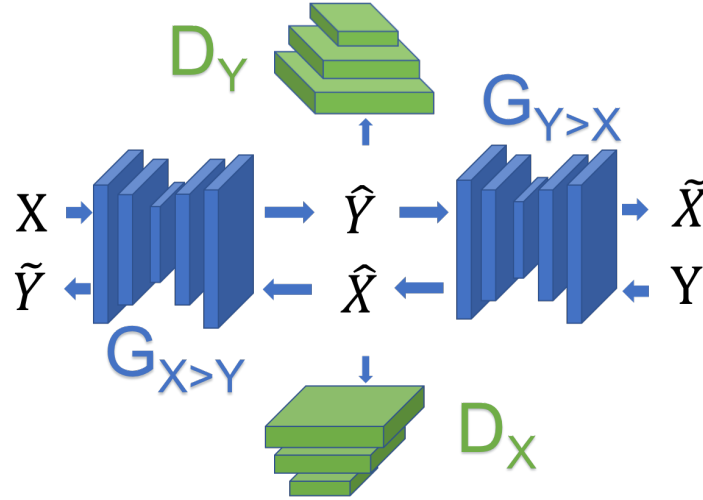


Figure 2.27 Architecture du CycleGAN (Zhu et al. 2017). La notation \hat{x}/\hat{y} désigne une reconstruction alors que le symbole \tilde{x}/\tilde{y} désigne une prédiction.

architecture-là a l'avantage de combiner les tenseurs en sortie des quelques premières couches avec les tenseurs en entrée des quelques dernières couches. Ceci permet de propager l'information d'amont en aval, en évitant que celle-ci ne soit perdue en chemin (Ronnenberg et al. 2015). Le discriminateur, quant à lui, est un réseau convolutif qui effectue des prédictions sur des fenêtres de taille 70x70 (Isola et al. 2017). La fonction de perte adverse est combinée avec une fonction de perte L1. Les deux sont complémentaires étant donné que la perte L1 permet l'apprentissage global de la synthèse, tandis que la perte adverse raffine et fait ressortir les détails. Cela revient à dire qu'appliquer uniquement la perte L1 donne un modèle qui produit des images floues (Isola et al. 2017).

L'architecture Pix2Pix nécessite un jeu de données paillé à l'entraînement. À l'inverse, le CycleGAN (Zhu et al. 2017) n'impose pas cette contrainte. Celui-ci a l'objectif d'apprentissage suivant : synthétiser des images dans la modalité Y à partir de leur équivalent dans la modalité X , tout en étant capable de reconstruire l'image originale à partir de sa prédiction. Cela est valable aussi dans l'autre sens, c'est à dire que le CycleGAN apprend la synthèse $X - Y$ et la synthèse inverse $Y - X$ conjointement. Le CycleGAN est donc implémenté comme deux réseaux adverses génératifs qui apprennent les deux synthèses adverses en même temps, en plus d'une contrainte de reconstruction qui les lie entre eux.

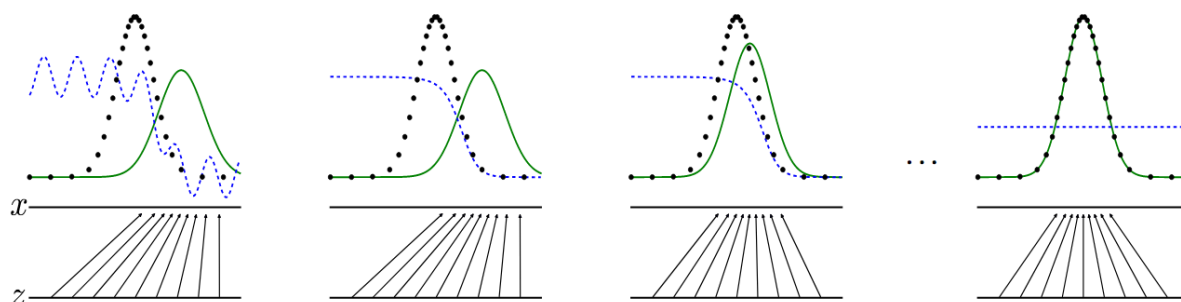


Figure 2.28 Schématisation chronologique, de gauche à droite, du processus d'apprentissage d'un réseau génératif adverse. La courbe noire désigne la distribution cible que l'on souhaite apprendre, la courbe verte représente la distribution apprise à l'itération courante alors que la courbe bleu désigne le pouvoir discriminatif du réseau. Progressivement, l'application de l'ensemble des \mathbf{z} vers l'ensemble des \mathbf{x} se rapproche de plus en plus de la distribution cible. Lorsque celles-ci se confondent, la courbe bleu illustre comment le discriminateur n'a plus aucun pouvoir discriminatif et qu'il ne peut plus faire la différence entre $G(\mathbf{z})$ et x . Image tirée de (Goodfellow et al., 2014)

2.5 Mot de synthèse

La revue de littérature nous a permis d'exposer les différentes modalités d'imagerie d'intérêt pour ce projet de recherche, ainsi que les différents types de recalage et les techniques utilisées. L'élément le plus important reste toutefois l'utilisation de l'apprentissage profond pour la translation d'image. Nous avons été capables de présenter une multitude de modèles et de techniques génératives. Ces techniques ont montré, dans la littérature, des résultats très prometteurs pour la translation d'image.

Armés des éléments vus, nous présenterons au chapitre suivant la méthodologie adoptée pour ce projet de recherche.

CHAPITRE 3 DÉMARCHE DE L'ENSEMBLE DU TRAVAIL DE RECHERCHE

Le cheminement du projet est le suivant. Ce projet débute avec un ensemble de données public de 18 volumes d'IRM et CT de patients différents.

En premier lieu, nous aurons effectué des expériences comparatives de synthèse d'image pour les architectures cGAN et CycleGAN. Étant donné que ceux-ci reposent sur différents paradigmes de translation d'images appairées et d'images non-appairées, il a été intéressant de comparer les deux. Nous aurons constaté par la suite que c'est le CycleGAN qui généralise le mieux à l'ensemble de validation, et nous l'aurons choisi comme notre modèle de référence pour la suite de ce projet.

Le CycleGAN ne produisant lui-même toujours pas d'images suffisamment satisfaisantes à l'oeil, il nous aura fallu expérimenter et être créatifs. Nous avons d'abord exploré des possibilités d'augmentation de données par synthèse d'images non-appairées. Nous avons collecté différents jeux de données de différentes sources. Cette approche-là n'a pas été fructueuse, et nous pensons que cela a été dû au fait que les données collectées soient beaucoup trop hétérogènes.

Après plusieurs expérimentations, nous avons trouvé qu'une approche pseudo-3D améliore la qualité des images produites suffisamment pour pouvoir ensuite poursuivre avec le recalage 2D/3D par DRR. Nous avons par ailleurs proposé une approche de calcul des histogrammes d'intensité qui puisse s'intégrer avec l'algorithme de rétro-propagation.

En ce qui concerne le protocole de validation de la synthèse d'image, celle-ci s'est faite en recalant les différents niveaux vertébraux par marquage anatomique. Ce recalage est validé avec le coefficient DICE des vertèbres post-recalage. Le recalage s'est fait manuellement en utilisant 6 marqueurs anatomiques suivants : centres des plateaux inférieur et supérieur, et extrémités inférieure et supérieure des pédicules droit et gauche.

Nous comparons par ailleurs les images synthétiques avec les vraies images en calculant les erreurs moyennes absolues sur les voxels ainsi que les coefficients normalisés d'intersections d'histogrammes. Nous effectuons un recalage 2D/3D à 4 vues où des DRR de vrai CT sont substitués aux fluoroscopies de bras en C. Les 4 vues sont les suivantes : antéropostérieur, latérale, $+45^\circ$ et -45° . Enfin, nous intégrons la composante de synthèse d'image avec la composante de recalage 2D/3D pour des modèles pseudo-3D avec et sans le terme de perte d'histogrammes afin de valider toute la méthode de recalage au complet.

L'erreur moyenne de recalage obtenue est de $2.1 \pm 0.2\text{mm}$, alors que les erreurs que l'on peut retrouver dans l'état de l'art pour le recalage d'IRM et de fluoroscopie sont de 4.8 mm

(Harmouche et al., 2012) et de 4.3 ± 2.6 mm (Da Silva et al., 2017).

La méthodologie proposée a mené à la soumission d’un article de journal que l’on peut retrouver au chapitre 4, d’un article de conférence ainsi qu’un brevet. Les en-tête sont les suivants :

—**R.Oulbacha**, S.Kadoury, “MRI to C-arm Spine Registration through Pseudo3D Cycle GANs with Differentiable Histograms”, *IEEE Transaction on Medical Imaging*, soumis, Novembre 2019

—**R.Oulbacha**, S.Kadoury, “MRI TO CT SYNTHESIS OF THE LUMBAR SPINE FROM A PSEUDO-3D CYCLE GANS”, *IEEE International Symposium on Biomedical Imaging*, soumis, Octobre 2019

—R.Oulbacha, S.Kadoury “INTRA-OPERATIVE REGISTRATION OF DIAGNOSTIC SPINE MRI WITH C-ARM IMAGES USING IMAGE SYNTHESIS WITH CYCLE GENERATIVE ADVERSARIAL NETWORKS (GAN)”, 05015490-144USPR, soumis, Mars 2019

CHAPITRE 4 ARTICLE 1: MRI TO C-ARM SPINE REGISTRATION THROUGH PSEUDO3D CYCLE GANS WITH DIFFERENTIABLE HISTOGRAMS

4.1 Présentation de l'article

Cet article présente la méthode de recalage 2D/3D entre l'IRM et la fluoroscopie de bras en C au complet. L'article inclus les expériences pour les différents modèles avec et sans les composantes pseudo-3D. L'article présente aussi une comparaison avec et sans la fonction de perte d'histogrammes. Les résultats de recalage sont présentés pour ces modèles-là.

4.2 En-tête de l'article

Titre de l'article

"MRI TO C-ARM SPINE REGISTRATION THROUGH PSEUDO3D CYCLE GANS WITH DIFFERENTIABLE HISTOGRAMS", IEEE Transactions on Medical Imaging, soumis le 14 Novembre 2019

Auteurs

Oulbacha Reda^a, Samuel Kadoury^{a,b}

Affiliations

^amedICAL Laboratory, Polytechnique Montréal, Montréal, QC, Canada

^bCHU Sainte-Justine Research Center, Montréal, QC, Canada

4.3 Abstract

Image-guided spine surgery increasingly relies on diagnostic MRI for tool navigation, as it provides a visualization of the nerves and soft tissue during screw insertion. However registration to C-arm images remains difficult due to differences in visible tissue. In this paper, we introduce a fully unsupervised 3D/2D registration approach of the MR lumbar spine using CT images synthesized from T2-weighted images acquired for diagnostic purposes. Our approach incorporates a trainable pre-processing pipeline using FC-ResNets combined with a low-capacity fully convolutional network to normalize the input MRI data and segment the vertebral bodies and pedicles. A pseudo-3D cycle GAN architecture is proposed to include neighboring slices in the synthesis process, introducing a new cyclic loss function ensuring consistency between MRI and CT synthesis by differential histograms to match multimodal distributions. As a final step, a multi-planar digitally reconstructed radiograph (DRR) registration approach aligns the 3D and 2D modalities. Clinical experiments were performed on a public dataset, which includes 18 patients with both MRI and CT. Quantitative comparison to expert CT segmentations yields an average Dice score of 84 ± 1.5 on synthetic CTs, while the DRR registration experiments yielded a target localization error of $2.1 \pm 0.2\text{mm}$. Intensity distributions and mean absolute errors in Hounsfield units also show promising results, illustrating the strong potential and versatility of the pipeline by achieving clinically viable CT scans which can be used for surgical guidance.

Keywords: CT-MRI synthesis, Cycle GAN, Differentiable histograms, 3D/2D registration, Image-guided spine surgery

4.4 Introduction

For image-guided spine interventions, registration of diagnostic MRI with intra-operative imaging (C-arm images) is a crucial step to visualize the third dimension of the anatomy and provide an axial view of the vertebra, which is important for pedicle screw insertion and to avoid any critical structures and soft tissues (nerves, veins) during instrumentation insertion. While registration of pre-operative CT to intra-operative C-arm is more common as both modalities are based on Roentgen's radiographic principles (bone attenuation) and often available in commercial systems, fusion of MRI to intra-operative imaging remains an open problem. This is mainly due to the difference in poses of the spine between the modalities (prone vs supine) and because there is no direct physical relationship between bone attenuation and the water excitation due to the magnetic field in the MRI. Indeed, an MRI is typically acquired in a supine position for diagnostic purposes, while intra-operative cone beam CT (CBCT) and C-arm images are taken in prone for surgical purposes with the back exposed.

Currently, T2-weighted MRI is the most commonly acquired imaging sequence for preoperative diagnosis (with good image quality of the spinal cord and CSF), while during surgery, multi-view fluoroscopic images are most frequently acquired. Accurate registration of the MRI with fluoroscopic images on a multi-level basis would allow to improve the surgical guidance, spatial localization of the anatomy and enable an adoption to a larger patient population undergoing spine surgery. With regards to multi-modal spine registration, most methods in the literature focus on CT to intra-operative X-ray registration [1]. Registration methods often depends on a pre-segmentation of the CT to perform a registration with the 2D radiographs [2]. In previous studies, a basic gradient-based 3D-2D registration framework was proposed with a parallel implementation for digitally reconstructed radiograph (DRR) generation and a similarity metric computation [3].

An MRI to X-ray registration approach based on inter-vertebral deformable models was presented for the scoliotic spine [4]. However the technique was based on the assumption landmarks were precisely detected on the vertebra, in order to match them with the pre-operative model. Several assumptions were made with regards to the curvature of the spine which may have limitation when applying the technique to a larger population of spines. Other registration techniques were proposed by corresponding projected 3D surface-normals to 2D gradients, and optimized with non-linear optimizers [5]. On the other hand, the accuracy of these techniques was limited when in presence of important differences in the image content between the MRI and X-ray images. This increased the number of views required for registration. However, the performance of these methods was compromised

when strong content mismatch is observed between MR and radiograph images, and multiple radiographic views (as many as nine) per registration were required for robust registration. Recently, an approach referred to as MR-LevelCheck was presented to perform 3D– 2D registration based on a segmented preoperative MRI [6], however it was based on contour alignment, which lacks robustness and relies heavily on segmentation.

Alternatively, simulation-based pseudo-CT data [7, 8] from multi-parametric MRI was proposed and then used in a straightforward 3D– 2D registration framework. The major disadvantage of this approach is the need to obtain specialized MRI pulse sequences that were not within the standard diagnostic imaging protocols for spine, in order to robustly synthesize a simulated CT image.

In the past few years, image synthesis has regained interest with the emergence of deep learning, with methods training image-to-image translation CNNs with unpaired natural images, for example with DualGAN [9] and CycleGAN [10] from an MRI. The goal of these CNNs is to map information from one image domain to another within latent space. Examples include the CT synthesis in the brain [11] and pelvic area [12]. These CNNs were trained by minimizing voxel-wise differences with respect to reference CT volumes that are rigidly aligned with the input MR images. However, slight voxel-wise misalignment of MR and CT images may lead to synthesis of blurred images. To address this, Nie et al. [12] proposed to combine the voxel-wise loss with an image-wise adversarial loss in a generative adversarial network (GAN) [13], with [14] proposing an unpaired method for MRI to CT synthesis of brain images. However image synthesis methods were primarily developed for the generation of medical images with normal anatomy [15], and provided single sliced prediction without incorporating context from surrounding slices of structures or attempt to generate similar intensity distributions. Few applications were dedicated for the spine anatomy, which exposes several pathologies and disk degeneration hindering the labelling process.

In this work, we propose a 3D/2D registration approach based on automated MRI to CT synthesis of the lumbar spine from a new Cycle GAN architecture. The model learns the appearance of vertebral bodies and pedicles from a training set of MR images and corresponding CT images, which incorporates contextual information with a pseudo-3D prediction from neighboring slices. We propose a novel differentiable histogram loss function enforcing the cycle consistency between both modalities during adversarial training. This ensures similar distributions of intensities between MRI and synthetic CT, which matches the physical properties of CT.

4.5 Methods and materials

4.5.1 Pipeline architecture

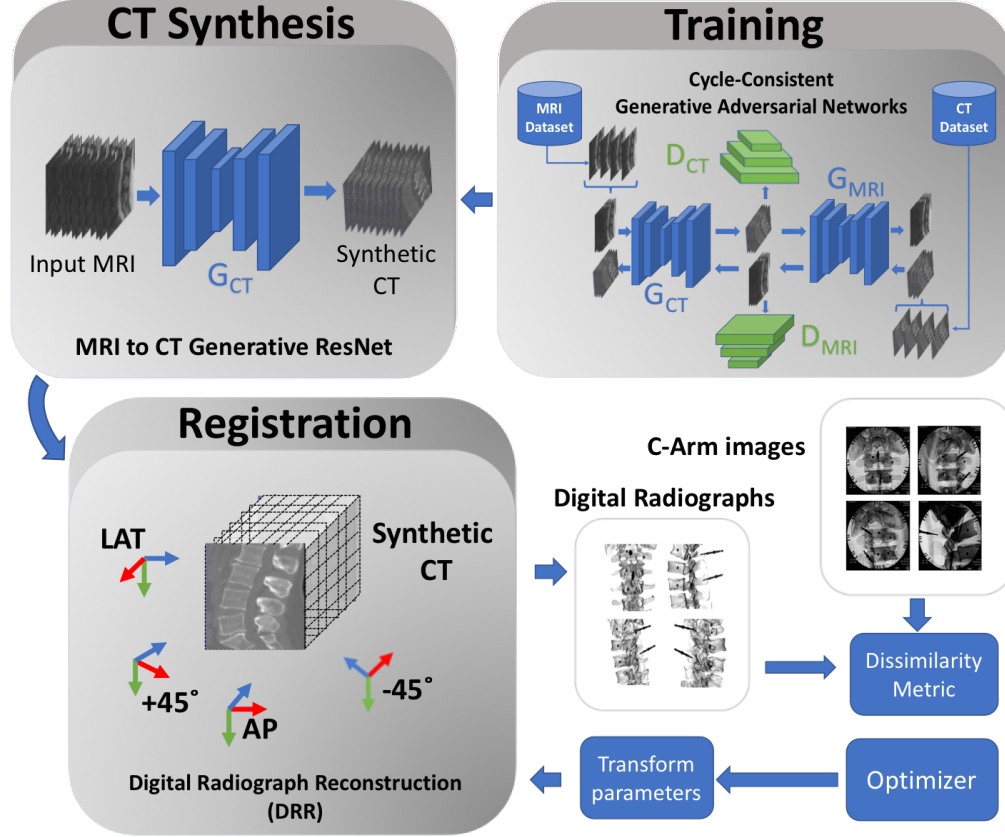


Figure 4.1 Overview of the 3D/2D MRI to fluoroscopic image registration pipeline. The diagram shows the two-stage synthetic image generation procedure. First, a CycleGAN is trained to jointly learn the MRI to CT and CT to MRI mapping. Once the training is performed, synthetic CT volumes are generated for an unseen MRI. Finally, the synthetic CTs are registered to the C-Arm images through Digitally Reconstructed Radiographs of the synthetic CT.

Overview

The 3D/2D registration pipeline integrates three main components, including a two-stage synthetic image generation procedure (Fig. 4.1). The first stage consists of a new cycle-consistent Generative Adversarial Network (CycleGAN) [10] for synthetic CT generation. In our case, the proposed network learns not only to generate a CT from an MRI, but it also learns to generate an MRI from a CT as a by-product. In the second stage, we use

the synthetic CT for Digital Radiograph Reconstruction (DRR) [16] using 4 different angles: anteroposterior, lateral, 45° and -45° . The pose of all 4 cameras is adjusted to determine the best match between the synthetic radiographs and the fluoroscopic images. The final camera transform parameters yield the MRI to C-Arm rigid registration transform.

Data pre-processing

As a first step to training the synthesis network, the vertebral bodies and pedicles were automatically segmented from the MRI volumes, with a trainable pre-processing pipeline using FC-ResNets [17]. That pipeline showcases a low-capacity fully convolutional network (FCN) [18] as a pre-processor to normalize the MRI data. The pipeline forms an iterative process where a normalized image is subsequently refined by the FC-ResNet to generate a segmentation prediction of the vertebral bodies and pedicles [19]. The volumes were then cropped around the lumbar spine region, rigidly registered using normalized mutual information and resampled in the same physical space with a sagittal slice thickness of 1.5mm. Finally, MRIs were standardized to a zero mean and unit variance. The CTs and MRIs were downsampled to voxel values in the open range $(-1, 1)$ using the arbitrary bounds $[-10, 10]$ and $[-1000, 4000]$ for the MRIs and CTs respectively, as in [14].

Image domain transfer

The first stage of the synthetic image generation follows the architecture of a CycleGAN, consisting of 4 fully convolutional neural networks (CNN) [20]: two generators and two discriminators. In this paper, we denote the MRI to CT generator by G_{CT} , and the CT to MRI generator by G_{MRI} . D_{MRI} and D_{CT} are the MRI and CT discriminators respectively. The discriminators follow a fully convolutional PatchGAN [21] architecture, with a patch size of 64x64 and a stride of 16. It is a succession of 4x4 filters with a stride of 2 and a padding of 1, followed by a final 4x4 kernel with zero padding and unit stride which gives those networks such a property (Table 4.1). The generators follow an architecture similar to

Table 4.1 64x64 Fully convolutional discriminator architecture

Block	Description
Conv - Dropout - LeakyReLU	64 4x4 kernels, stride 2, pad 1
Conv - InstanceNorm - Dropout - LeakyReLU	128 4x4 kernels, stride 2, pad 1
Conv - InstanceNorm - Dropout - LeakyReLU	256 4x4 kernels, stride 2, pad 1
Conv - InstanceNorm - Dropout - LeakyReLU	512 4x4 kernels, stride 2, pad 1
Conv - Sigmoid	1 4x4 kernel, stride 1, pad 0

that of a ResNet [22], which was used previously in [23] and [10]. It is composed of a series of downsampling convolutional layers, followed by 9 residual blocks and a series of upsampling transpose convolutional layers [24] to recover the original dimensions of the input (Table 4.2). A residual block's forward path consists of the following: unit reflection pad, 3x3 convolution with a zero pad and a unit stride, InstanceNorm [25], ReLU, Dropout [26], unit stride zero pad 3x3 convolution, and a unit reflection pad. The forward path's output is summed with its input.

Table 4.2 9-Block ResNet generator architecture

Block	Description
2D Reflection Padding	Padding size 3
Conv - InstanceNorm - ReLU	64 7x7 kernels, stride 1, pad 0
Conv - InstanceNorm - ReLU	128 3x3 kernels, stride 2, pad 1
Conv - InstanceNorm - ReLU	256 3x3 kernels, stride 2, pad 1
9 x Residual block	-
ConvTranspose - InstanceNorm - ReLU	128 3x3 kernels, stride 2, pad 1
ConvTranspose - InstanceNorm - ReLU	64 3x3 kernels, stride 2, pad 1
2D Reflection Padding	Padding size 3
Conv - Tanh	7x7 kernels, stride 1, pad 0

We separate the loss for the generators and discriminators as:

$$\begin{aligned}
\mathcal{L}(G_{CT}, G_{MRI}) = & \|D_{CT}(G_{CT}(MRI)) - 1\|_2^2 + \\
& \|D_{MRI}(G_{MRI}(CT)) - 1\|_2^2 + \\
& \lambda_{cycle} \left[|G_{CT}(G_{MRI}(CT)) - CT|_1 + \right. \\
& \left. |G_{MRI}(G_{CT}(MRI)) - MRI|_1 \right]
\end{aligned} \tag{4.1}$$

$$\begin{aligned}
\mathcal{L}(D_{CT}, D_{MRI}) = & \|D_{CT}(CT) - 1\|_2^2 + \\
& \|D_{MRI}(MRI) - 1\|_2^2 + \\
& \|D_{CT}(G_{CT}(MRI))\|_2^2 + \\
& \|D_{MRI}(G_{MRI}(CT))\|_2^2
\end{aligned} \tag{4.2}$$

At each iteration, Eq.(1) is minimized with respect to the generators, and then Eq.(2) is

minimized with respect to the discriminators. Rather than performing an image domain transfer on single-slice 2D images, we adopt a pseudo-3D strategy which is a computationally effective alternative to 3D convolutions. We group neighbouring sagittal slices in quadruplets and stack them along the channel’s dimension, thus forming a thick sagittal slice [27]. Similar performance was obtained with thicknesses of 3 or 4 slices. Larger thickness was found to degrade the model’s performance, thus we chose to use quadruplets.

The GAN [13] is trained to perform image domain transfer on the thickness ened slices rather than 2D images. Such an approach has the advantage of factoring in neighbouring 3-dimensional information, while still maintaining a reasonable memory footprint, contrary to 3D convolutions [27]. Unlike the original network which was designed for individual 2D images, we propose to train on a 3D volume consisting of quadruplets of neighbouring 2D slices stacked together along the channel dimensions, forming a thickened slice. A volume of N singleton slices equivalates to $N - 3$ thickened slices. To recombine the thickened slices into a volume of N singleton slices, we take the first three slices of the thickened slice. For each of following thickened slices, we chose to keep only the 3rd slice. Finally, the last thickened slice is created from remaining slices to complete a total of N slices. This recombination strategy ensures that all the slices that are not in the extremities come from thick slices with 3 overlapping slices, allowing a continuity in the shared 3D information.

4.5.2 Differentiable histogram loss

Matching histograms of intensity distribution between multimodal images during the synthesis process can help adversarial learning towards a realistic translation model. However, computing histograms for continuous variables involves sequential operations which are not differentiable: quantization and binning. This prohibits the use of backpropagation. A histogram conveys important information on intensity distributions, and being able to use it in the framework of backpropagation has a strong potential in a wide range of applications. In our approach, we perform quantization and binning using a smooth set of functions, thus allowing us to backpropagate through the computation graph. The mathematical formalism of non-differentiable histogram computation is as follows. Let X and Y be two real-valued vectors of size n , where all values are bounded in the (m, M) interval. Let:

$$\tilde{X} = \lfloor \frac{X - m}{M - m} n_{bins} \rfloor \quad \tilde{Y} = \lfloor \frac{Y - m}{M - m} n_{bins} \rfloor \quad (4.3)$$

Here, \tilde{X} and \tilde{Y} are vectors with integer values between 0 and $n_{bins} - 1$. Constructing the frequency count histograms of \tilde{X} , \tilde{Y} and the joint (\tilde{X}, \tilde{Y}) , we assume a Dirichlet prior on the joint histograms where each (\tilde{X}, \tilde{Y}) has k imaginary observations. For $i, j \in \{0, \dots, n_{bins} - 1\}$,

and without loss of generality for $H^{\tilde{Y}}$, we have the following frequency counts for histogram vector $H^{\tilde{X}}$ and joint histogram matrix $H^{\tilde{X}\tilde{Y}}$:

$$H_i^{\tilde{X}} = \frac{\sum_{p=0}^{n-1} \Delta_i(\tilde{X}_p) + kn_{bins}}{n + kn_{bins}^2} \quad (4.4)$$

$$H_{i,j}^{\tilde{X}\tilde{Y}} = \frac{\sum_{p=0}^{n-1} \Delta_{i,j}(\tilde{X}_p, \tilde{Y}_p) + k}{n + kn_{bins}^2} \quad (4.5)$$

where Δ_i is the discrete Dirac Delta shifted to the integer i . One can verify the Dirichlet prior by marginalizing the joint histograms matrix, by either \tilde{X} or \tilde{Y} , and verifying for equality with the corresponding marginal histograms vector. Looking closely at this formalism, two functions are problematic for differentiability and backpropagation: the floor function $\lfloor \cdot \rfloor$ and the Dirac delta Δ_i . As substitutes for the latter two, we propose two other functions that are continuous and differentiable, and that allows the network to backpropagate through any scalar-valued function which uses histograms. For some $\alpha > 0$, let g and f be:

$$g : x \mapsto \frac{1}{2} (1 + \tanh(\alpha x)) \quad f : x \mapsto \sum_{i=1}^{n_{bins}-1} g(x - i) \quad (4.6)$$

Figure 4.2 shows how f can be a natural substitute to the floor function. For the Dirac delta, we should recall that, when composed with the floor function, it serves as an indicator function for values in the $(i, i + 1)$ interval. Since f maps any $(i, i + 1)$ interval to the $(i - \frac{1}{2}, i + \frac{1}{2})$ interval, our substitute for the Dirac delta will be a standalone smooth indicator for values in that interval. That way, the composition of both of the smooth substitute functions performs as a soft count of values in the $(i, i + 1)$ interval.

$$u_i : x \mapsto g\left(\left(x + \frac{1}{2}\right) - i\right) - g\left(\left(x - \frac{1}{2}\right) - i\right) \quad (4.7)$$

$$u_{ij} : x, y \mapsto u_i(x)u_j(y) \quad (4.8)$$

From Eqs. (4.7) and (4.8), u_i is the univariate indicator function, while u_{ij} represents the bivariate indicator function. The latter is used to compute the joint histogram X, Y . Not only do the u_i 's serve as indicator functions, but they also produce valid histograms that naturally integrate to 1. The proof for that relies on the fact that $\sum_{i=0}^{n_{bins}-1} u_i(\tilde{X}_p)$ is a telescopic sum that evaluates very closely to one for $\tilde{X}_p \in [0, n_{bins} - 1]$ if α is not too small. The proposed histogram loss term integrated in the adversarial training is as follows:

$$\begin{aligned} \mathcal{L}_{hist}(G_{CT}, G_{MRI}) = \lambda_{hist} & \left[|H^{G_{MRI}(G_{CT}(MRI))} - H^{MRI}|_1 \right. \\ & \left. + |H^{G_{CT}(G_{MRI}(CT))} - H^{CT}|_1 \right]. \end{aligned} \quad (4.9)$$

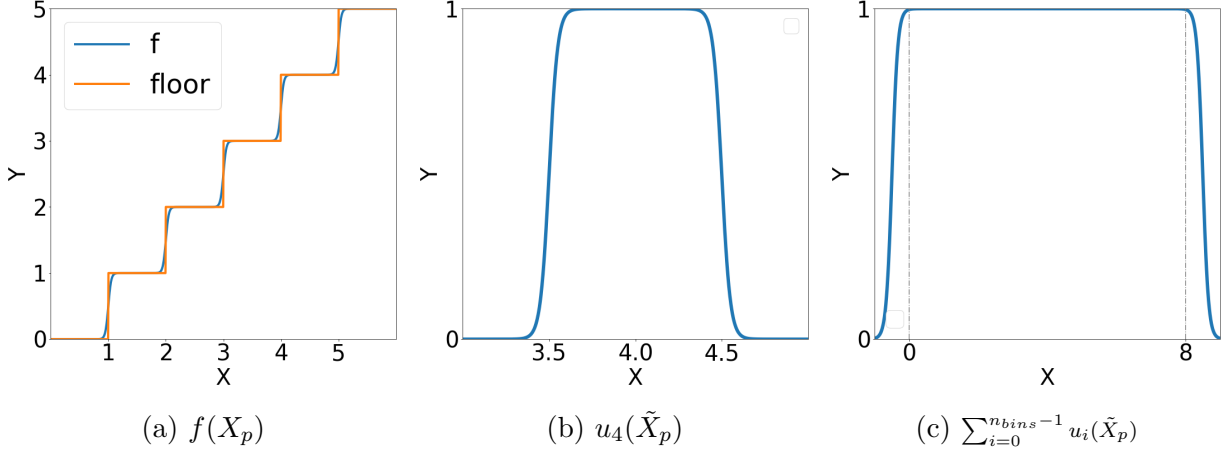


Figure 4.2 For (a) and (b), $\alpha = 20$ was used. For (c), $\alpha = 6$ was used along with $n_{bins} = 9$. We see how the expression on (c) evaluates very closely to 1 on $[0, 8]$. If α was lower, the assumption would no longer be valid.

Using this approach, we are able to use these histogram vectors and matrices to optimize functions such as mutual information, KL-divergence or any other metric that requires marginal or joint histograms.

4.5.3 Training procedure

During the training process, all volumes are reoriented along the same anatomical axes and reformatted in the sagittal plane, with the patient’s back on the right side of the image and the spine in the upright position. Data augmentation was performed with random crops of sizes ranging from 25% to 100% of the size of the original image, at a probability of 0.5. Random flipping was done in the dimensions of the width, height and depth. The final output of the data pipeline was resized to a width and height of $[324 \times 324]$ before being fed to the networks. We used the Adam optimizer for our experiments [28], with a β_1 of 0.5. The initial learning rate for all optimizers was 0.0002. Training was done with 600 epochs, with a linear learning rate decaying to zero, starting at epoch 200. The discriminator’s objective functions are scaled by a factor of 0.5 to reduce the learning rate of the discriminator relative to the generator [21]. The overall loss function \mathcal{L} with hyperparameters $\lambda_{cycle} = 10$, $\lambda_{hist} = 0.1$, $k = 1$, $n_{bins} = 32$ and $\alpha = 20$ is the following:

$$\mathcal{L} = \mathcal{L}(G_{CT}, G_{MRI}) + \mathcal{L}(D_{CT}, D_{MRI}) + \mathcal{L}_{hist}(G_{CT}, G_{MRI}) \quad (4.10)$$

At training time, the discriminators are shown random images drawn from a history of the last 50 images produced by the generators rather than the latest ones, for regularization purposes [29].

4.5.4 DRR Registration

For the generation of Digitally Reconstructed Radiographs (DRR), we use the Siddon-Jacobs [30] algorithm to compute the exact radiological path through the synthesized CT volume, which was implemented on GPU using CUDA. We use the linear absorption coefficients for an equivalent energy of 30 keV. The C-Arm is modeled as a pinhole camera, where each view has an intrinsic and an extrinsic camera matrix [31]. We use a setup of 4 views, where the relative camera poses remain fixed, and a single rigid transformation moves the entire space for registration. After selecting the vertebral level, a ROI is defined around the vertebral body and the pedicles to initiate the optimization process, which minimizes an objective function summing the negative Normalized Cross Correlation (NCC) over all 4 views.

4.6 Results

4.6.1 Experimental setup

A dataset of MRI and CT volumes on the lumbar spine from 18 different patients was used for the experiments. The dataset was extracted from the public SpineWeb library (<http://spineweb.digitalimaginggroup.ca>). This dataset includes T2-w MRI and CT scans with different dimensions, voxel spacing, and intensity ranges. For the image domain transfer task, we performed a cross-validation with 6 different splits, where each model only sees 15 out of the 18 volumes, and test that model on the 3 remaining volumes. Aggregating the test set synthetic image predictions of each of those models, a synthetic CT is obtained for 18 unseen test cases.

All vertebrae visible on the real and synthetic CTs were annotated by an expert with the following 6 anatomical landmarks: top and bottom endplate center, left and right pedicle extremities. These markers were used to compute residual landmark localization errors on each vertebra. Quality of the image synthesis was also assessed with the DICE score computed from segmentations. The purpose of these experiments was to validate how much structure is preserved when generating synthetic CTs. For the 3D/2D registration procedure, we use DRRs of the ground-truth CT as a substitute for the C-Arm fluoroscopic images. For each vertebra in each one of the real CT volumes, we took images from 4 different angles: posterior, lateral, $+45^\circ$ and -45° . These images were use as target X-rays for 3D/2D registration.

Finally, the post-registration error on the anatomical landmarks cited above was computed for each vertebra.

4.6.2 Differentiable histograms

In this section, we experiment the behavior of the differentiable histogram loss. To do so, we estimate the mutual information for two simulated cases where the true theoretical value of that metric is known. The parameters for the differentiable histogram estimator are $\alpha = 20$, $k = 0.1$, $n_{bins} = 64$. The sample size for the estimators is $n = 100000$. We set $m = \min(\inf(X), \inf(Y))$ and $M = \max(\sup(X), \sup(Y))$. For the first case, we consider an input signal $X \sim \mathcal{N}(0, \sigma_x^2)$, a noise $N \sim \mathcal{N}(0, \sigma_n^2)$ and an output $Y = X + N$. Setting $s = \frac{\sigma_x^2}{\sigma_n^2}$ (SNR), it can be shown that $MI(X, Y) = \frac{1}{2} \log_2(1 + s)$. For the second simulated case, we consider a bivariate Gaussian distribution with zero mean, unit variance and a correlation coefficient $\rho > 0$. It can be shown that $MI(X, Y) = -\frac{1}{2} \log_2(1 - \rho^2)$.

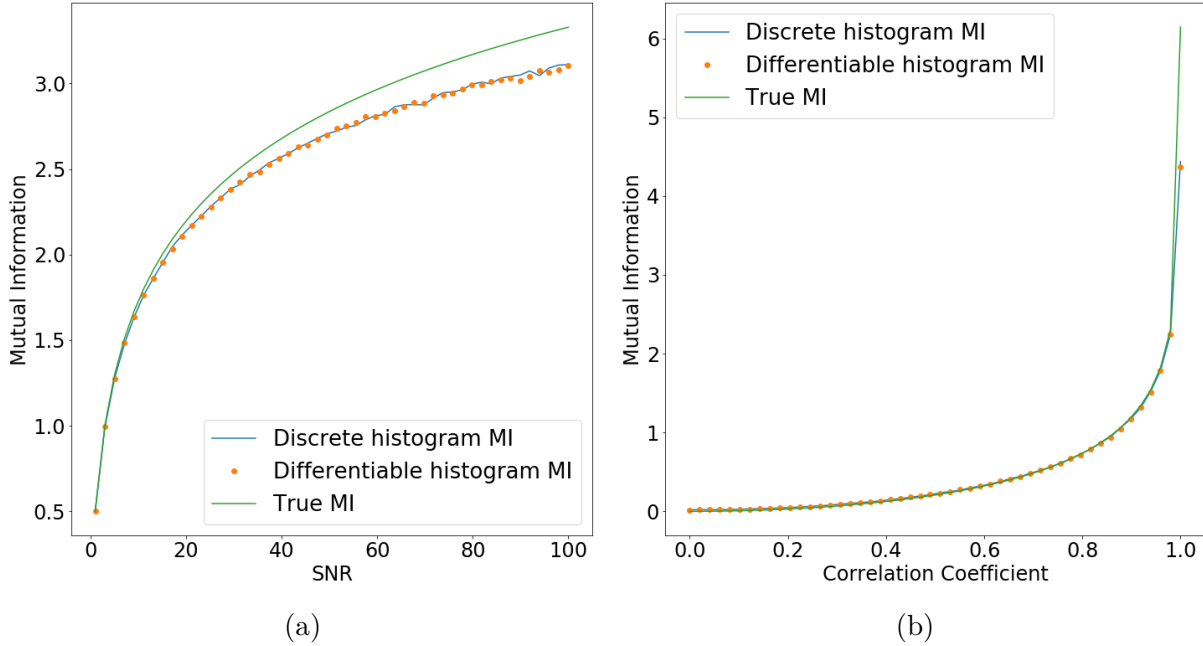


Figure 4.3 Experimental results for mutual information (MI) estimation with (a) additive white Gaussian noise channel, and (b) correlated bivariate Gaussians.

Although Fig. 4.3a shows the discrete non-differentiable histogram density estimator doesn't converge to the exact theoretical value of the mutual information, there is a clear correlation between the two. Most importantly, we see how the differentiable estimator approximates the non-differentiable one. Fig. 4.3b shows a closer approximation of the true underlying

mutual information than Fig. 4.3a, yet the differentiable estimator still approximates the discrete one very closely.

4.6.3 Image domain transfer

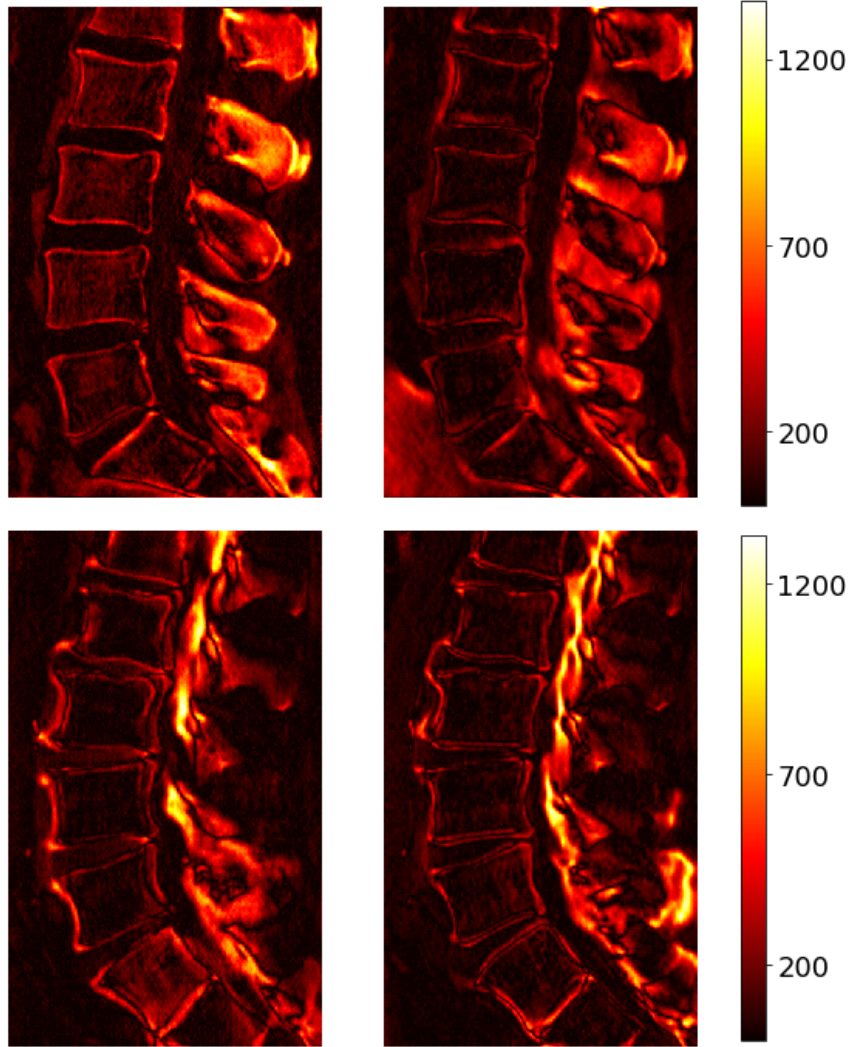


Figure 4.4 Absolute error maps of real and synthetic images (in HU). The first column shows images with the Pseudo-3D CycleGAN without the histogram loss, while the images in the second column were produced using the histogram loss. Images of two different patients are shown in each row.

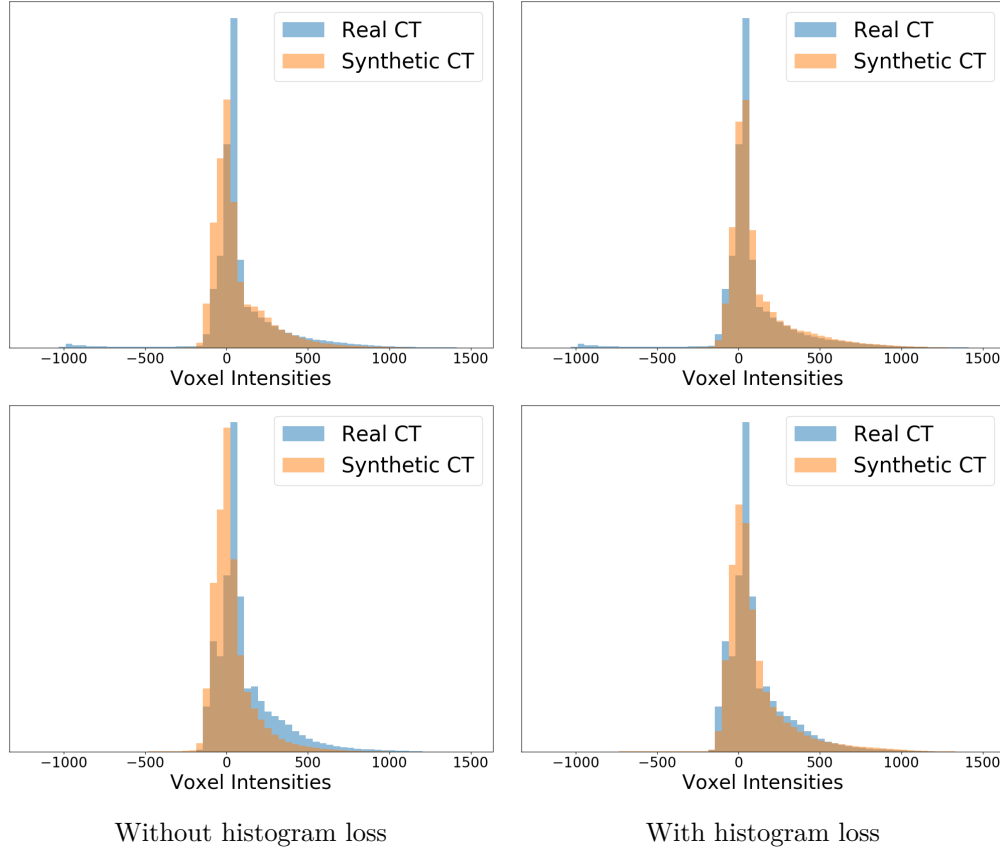


Figure 4.5 Sample voxel intensity histograms for two patient volumes, one for each row. Volumes are synthesized using the two pseudo-3D CycleGANs with and without the histogram loss term, one for each column.

In this section, we compare the performances of four different synthesis models: a conditional GAN (cGAN) [21], a CycleGAN, and a pseudo-3D CycleGAN, without and with the proposed histogram loss term. All models were trained with the same hyper-parameters and use the same backbone neural networks, except for the conditional GAN which adopts a UNet [32] architecture for its generators and a slightly different objective function, as in [21].

Fig. 4.4 compares HU error maps from synthetic CTs of two different patients, where we can observe how the model with the histogram loss improves on some of the voxel intensity errors in the region of the vertebrae. The observations regarding the DRR threshold value along with those made on Fig. 4.4 correlates with Fig. 4.5, where we can see how the histograms are translated towards the higher voxel intensities, which correspond to the bone structures.

Table 4.3 shows the Normalized Histogram Intersection (NHI) and Mean Absolute Error (MAE) for all patients. While the MAE metric shows no significant difference between both

Metric	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Average
NHI (synCT) w/o histogram loss	0.917	0.946	0.908	0.814	0.762	0.893	0.808	0.772	0.778	0.847	0.783	0.768	0.688	0.738	0.856	0.878	0.867	0.761	0.821 ± 0.061
NHI (syn. CT) w histogram loss	0.922	0.955	0.751	0.949	0.914	0.841	0.908	0.882	0.845	0.936	0.843	0.903	0.670	0.834	0.832	0.885	0.862	0.733	0.859 ± 0.067
MAE (syn. CT) w/o histogram loss	116.45	140.06	124.7	126.34	113.64	99.7	177.86	150.46	128.38	148.66	144.49	134.93	147.21	155.00	114.36	122.64	92.73	149.72	132.63 ± 20.06
MAE (syn. CT) w histogram loss	127.37	146.86	153.76	121.08	108.33	122.46	164.62	141.00	120.31	123.32	133.08	96.88	162.67	116.51	129.3	132.93	108.27	160.41	131.62 ± 15.97
NHI (syn. MRI) w/o histogram loss	0.880	0.867	0.866	0.740	0.751	0.638	0.788	0.876	0.760	0.631	0.543	0.476	0.849	0.860	0.727	0.847	0.718	0.874	0.761 ± 0.095
NHI (syn. MRI) w histogram loss	0.863	0.846	0.833	0.901	0.890	0.712	0.530	0.558	0.841	0.903	0.882	0.684	0.760	0.677	0.850	0.842	0.697	0.835	0.783 ± 0.088
MAE (syn. MRI) w/o histogram loss	0.464	0.658	0.525	0.744	0.748	0.748	0.905	0.914	0.916	0.970	0.803	0.866	0.854	0.889	0.766	0.505	0.688	0.751	0.769 ± 0.119
MAE (syn. MRI) w histogram loss	0.637	0.729	0.705	0.583	0.605	0.652	0.754	0.723	0.648	0.706	0.561	0.646	0.708	0.656	0.662	0.699	0.708	0.637	0.668 ± 0.045

Table 4.3 Comparison of pseudo-3D models with and without histogram loss. Mean Absolute Error (MAE) in HU and Normalized Histogram Intersection (NHI) metrics between real to synthetic images (for CT and MRI), per patient.

pseudo-3D models, a noticeable difference is seen with regards to NHI. A one-tailed and paired t-test yields a p-value of 0.032. This confirms the observations in Fig. 4.5 with the overlap of histograms with the ground-truth CTs. When generating DRRs from the CTs that were produced with the model integrating the proposed histogram loss, no adjustment in terms of contrast levels was needed to take into account the line integral computation, as opposed to the baseline model.

Fig. ?? shows improved DICE scores for the histogram loss model, with an average score of 84 ± 1.5 , compared 82 ± 1.8 . The latter is coherent with the previous observations, as the automatic segmentation platform we used was trained on real data, and heavily relies on voxel intensities. Fig. 4.6 shows how the pseudo-3D improves on the synthetic images, and that is mostly visible when looking at axial slices. We notice how the cGAN fails to produce visually coherent images, given the small size of the training dataset.

Looking at the images produced by the proposed model with the histogram loss, we notice a slightly higher contrast when analyzing the contours of the vertebra in the axial views, as opposed to the regular pseudo-3D CycleGAN. The overall contrast level also seems to be preserved. The visual improvements that the histogram loss brings remain subtle when only looking at the 3D volumes qualitatively, though as shown in the following section, improved distribution of voxel intensities can lead to improvements when performing DRR registration.

Not only are the contours of the vertebra smoothed out, but the intensities within the ver-

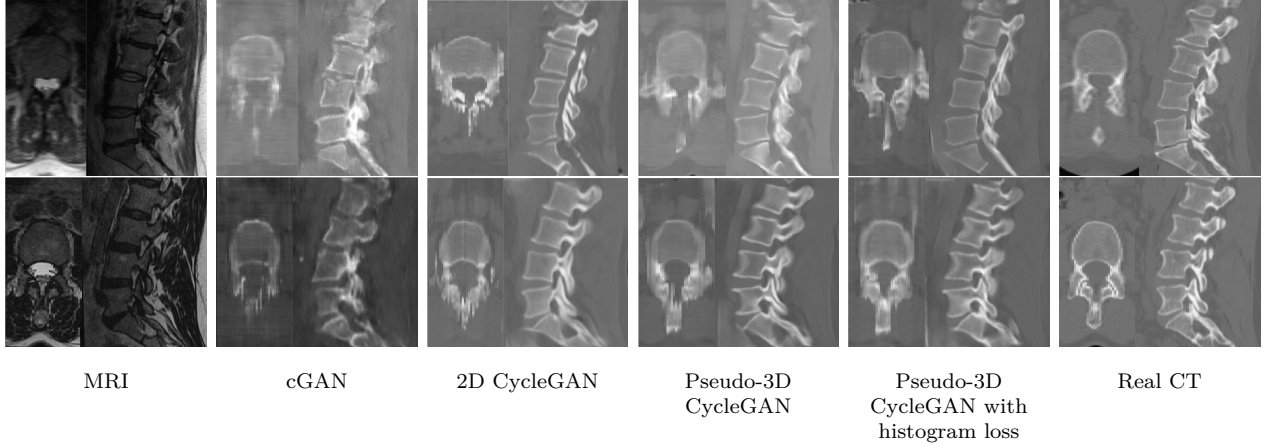


Figure 4.6 Synthesis results from two different patients (each row), comparing results of different MRI to CT synthesis models to the real CT.

tebral body also seem more homogeneous, contrary to images produced by the 2D model, for which the sagittal misalignment in space and contrast is visible in the axial plane. The pseudo-3D CycleGANs perform better than the normal CycleGAN, even when only considering the sagittal slices. There are cases where the 2D CycleGAN was not able to produce a volume with visually coherent sagittal slices. This may be caused by the added pseudo-3D component which acts as a form of data augmentation, as the model sees the same slice in different representations. Every single slice can take 4 different positions in 4 different quadruplets. The latter does not apply to extremities of a volume however.

4.6.4 2D/3D Registration

Finally, we present registration results using two versions of the pseudo-3D CycleGAN. In each experiment, an approximate focus of the cameras around the vertebra of interest is done, by extracting the physical coordinate of the center of that vertebra from the CT. This is followed by a translation of the 4 cameras views such that the center of the vertebrae coincide with the intersection point of the optical axes of all cameras. This enables the vertebra to be at the center in all views. The intersection point of all camera optical axes is obtained through Direct Linear Transform (DLT) [33] triangulation. We define the three rotation axis to consist of the optical axis of the camera capturing the anteroposterior view, the optical axis of the camera capturing the lateral view, and the axis that is perpendicular to both. Our translation axis remain X , Y and Z . Such parameterization allows us to adjust the anteroposterior and lateral views in rotation independently, without changing

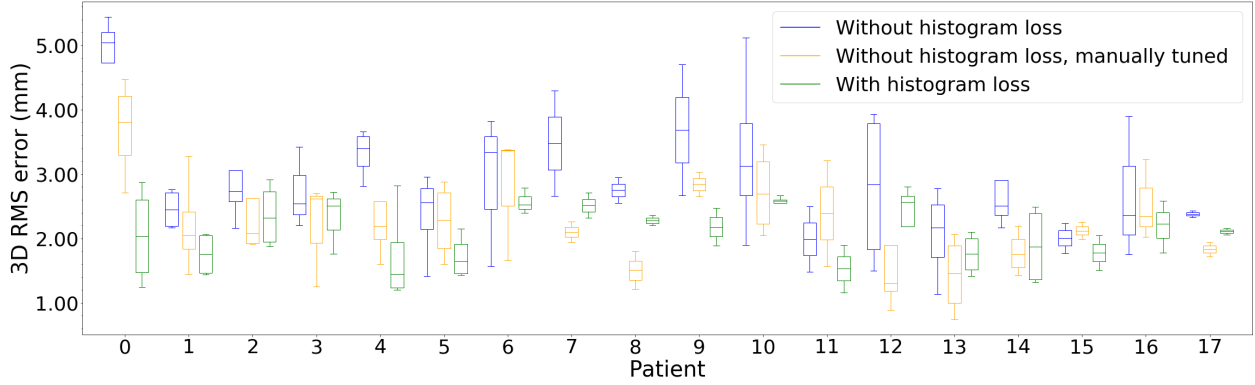


Figure 4.7 3D RMS registration errors on 6 vertebral anatomical landmarks of all vertebrae, for each patient.

the point on which all cameras are focused. Once the rough positioning is performed, we define a rectangular region of interest (ROI) mask around the vertebra. The dissimilarity metric, the negative Normalized Cross Correlation (NCC), is thus computed only within the ROI. We use SciPy’s [34] Powell optimizer to locally minimize the dissimilarity metric with our parameters being the 6 rigid degrees of freedom. Three registrations were performed in total, using synthetic CTs produced from the two pseudo-3D CycleGAN with and without the extra histogram loss term. Manually tuned registration involves manual adjustment of voxel thresholding along with careful alignment of the texture, visible contour and anatomical structure in all four views. An ROI masks the vertebra that is targeted for registration.

Figures 4.7 and 4.8 show median errors in landmark localization within the 2-3mm range, where the registrations performed using synthetic CTs produced by the CycleGAN with no histogram loss have a relatively higher error if no manual fine tuning is performed. Average 3D RMS errors are $2.89 \pm 0.57\text{mm}$ for the model with no histogram loss, $2.30 \pm 0.49\text{mm}$ for the model with no histogram loss and with manual tuning, and $2.08 \pm 0.23\text{mm}$ for the model with the proposed histogram loss. The manual tuning is done by adjusting the DRR generation threshold along with fine tuning of the transform parameters while aiming for superposition of the vertebra’s contours, geometry and visible anatomical structure. For the other experiments, we only perform the initial positioning, and then perform a final visual assessment after the automatic registration is performed, with no adjustment. As it was shown in Fig. 4.5, the lower voxel intensities if no histogram loss is used. This tends to produce synthetic CTs for which intensities in the bone structures fall below the DRR generation threshold value, thus needing manual adjustment to improve the registration.

Fig. 4.9 shows sample registration results. The overlay in orange corresponds to the DRR

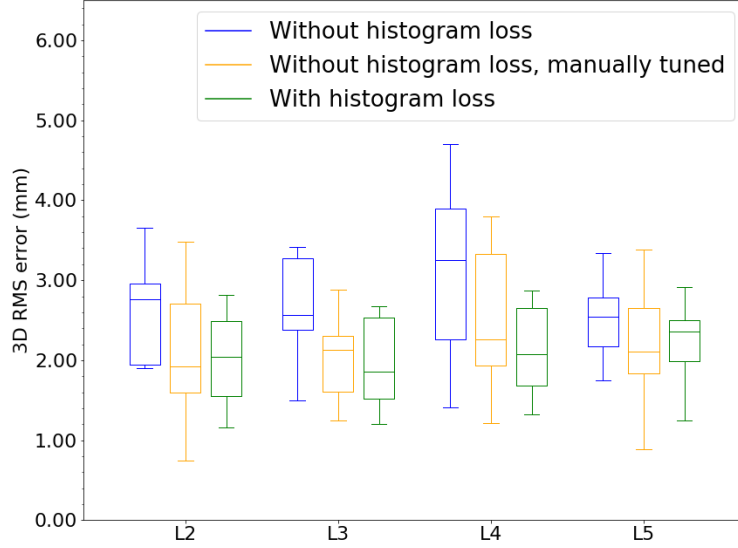


Figure 4.8 3D RMS registration errors per vertebral level, for both pseudo-3D CycleGAN models with and without histogram loss.

generated from the synthetic CT. Without histogram loss, the subsequent DRRs are sparse and may lack important contour and geometry information. We can observe the vertebra's contours are blurred compared with histogram loss, where the contours appear more clearly. Indeed, the voxel intensities in the synthetic CT fall below the threshold value, thus producing a sparser DRR where much of the information on the contours and the geometry lacks. In bottom rows of Fig. 4.9(a)-(b), the DRRs are much more dense and we can see more of the vertebra's contour and structure, improving registration. In Fig. 4.6, synthetic CTs produced by the pseudo-3D model with no histogram loss can show lower contrast between bone and surrounding tissue. This results in a blurry DRR, which may come from the contribution of the surrounding tissue voxels to the line integral computation. The comparison of top and bottom rows in Fig. 4.9(a) and 4.9(b) illustrates this phenomena.

4.7 Discussion

Based on the experimental results presented in the previous section, we can observe that synthetic images generated from the differentiable histogram loss yields improved segmentation and registration accuracy compared to synthetic CTs where manual adjustments of the projected images are needed. However, when performing manual adjustments of contrast levels, noticeable differences in registration errors are found. Our hypothesis for such differences

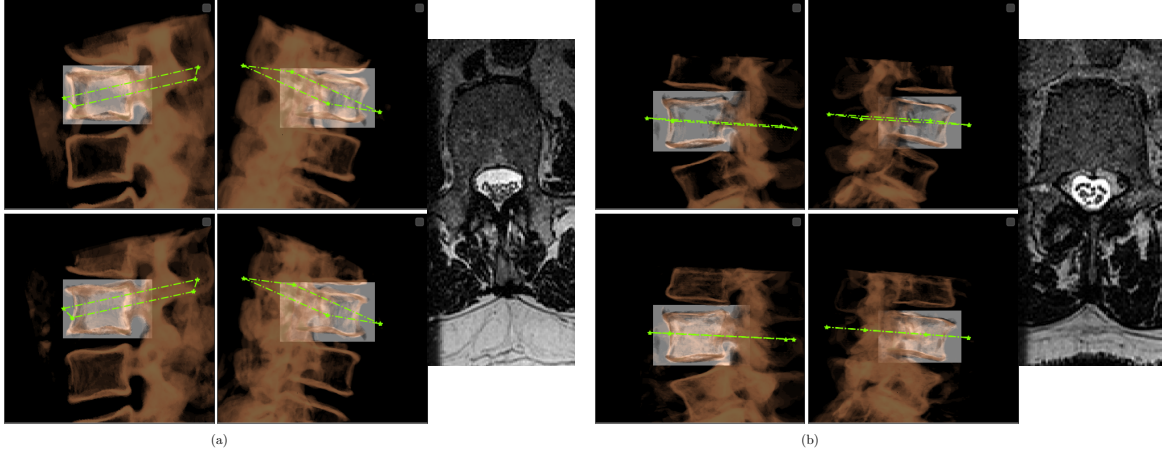


Figure 4.9 (a)-(b) Two sample 3D/2D registration overlay results with Pseudo-3D CycleGAN, showing C-arm and DRR images. For each case, top row is without histogram loss, bottom row is with histogram loss. Left image is the lateral view, while the right image is at -45° . The corresponding MRI slices are depicted with the dashed green lines.

lies from the fact the histogram loss yields CTs with voxel intensity distributions similar to that of real data, and thus allows for DRRs of improved quality that are easier to register. However, given the relatively small size of training data, this observation would need to be confirmed with more extensive datasets.

The histogram loss contributes in recreating synthetic volumes for which the intensity distribution mimics that of the real data. For our synthetic CT, it helps produce a DRR that is more similar to what would have been generated from a real CT, and that is our hypothesis for the improvement to the automatic registration accuracy. Fig. 4.8 shows a registration error per vertebra that is relatively higher for the experiment with no histogram loss and no manual tuning. Fig. 4.7 shows a relatively high variance for that experiment compared to the other two, which translates into a lower accuracy. We suspect the lower quality of the generated DRRs to be the cause of the added variance. Indeed, a blurry DRR will generally lack the distinguishing geometry, features and contours which is what allows the registration to be done accurately and precisely, whether it be manually or automatically. Not having that visibility adds randomness and uncertainty to how the dissimilarity metric is optimized. Manual fine tuning can allow for that error to be minimized. However, such tuning is still prone to error. It sometimes requires the operator to be able to distinguish the vertebra's features from within a blurry image, and often extrapolate the location of those features with their prior knowledge of the vertebra's anatomy.

When comparing different CycleGANs, the improvement when adding a pseudo-3D compo-

ment is mostly visible on the axial view, where noise between neighbouring sagittal slices is smoothed. A subtle improvement in the colorization and some of the details can be seen too when adding the extra histogram loss term. Overall, we can see that the pseudo-3D approach clearly shows improvement over the 2D approach when visually inspecting the images that were synthesized. Furthermore, the extra histogram loss term allows for a voxel intensity distribution that resembles more that of our real data. That extra loss term has repercussions on how the CycleGAN learns to capture the dynamic range of intensities from our CT data. The synthetic data is thus also easier to segment automatically using models that were trained on real data.

From Fig. 4.6, we can observe how the cGAN produces poor synthetic images. We suspect this is also caused by the limited size of the dataset, where the total number of different slices in the training set is less than 800. This itself is a limitation regardless of which model we chose. When considering the cGAN model, we see it imposes a very strong optimization objective for inputs with their counterpart ground truth images compared to its adversarial objective [21], thus assuming point-to-point correspondences. Many image domain adaptation tasks do not include a well-defined expected output [10], which adds to the challenge of validating such models. Our data, although having paired MRI and CT images of the same patients, inherently contains imperfect point-to-point correspondences due to the patient’s movement from one modality to another. Rigid registration minimizes that phenomena, though it is still present and very noticeable in some cases. One example is the CT of patient containing clamps and screws, whereas the MRI doesn’t, then the model still tries to optimize for a synthetic CT with clamps and screws, and that is undesired behavior.

The CycleGAN, however, does not require voxel-to-voxel or even image-to-image correspondences. Its optimization objective is to attempt to reproduce plausible synthetic images from which the original images can be reconstructed, using the same cyclical neural networks. As previously mentioned, in select cases where the ground truth CT has surgical clamps and screws, the synthetic CT that is produced from MRI excludes these tools. We also find that the CycleGAN performs better to unseen examples with the smaller dataset than the cGAN does. However, when inspecting synthetic samples produced from the training set, we notice that the cGAN easily captures its complexity and overfits rather quickly. This observation leads us to chose the CycleGAN over the cGAN in our 3D/2D registration pipeline even if our data is paired. Although the CycleGAN produces synthetic sagittal slices that are visually coherent, grouping these slices in a 3D volume that preserves spatial coherence is a challenge. Our initial approach for image domain transfer performs learning on 2D sagittal slices. The 3D volumes that are produced by this approach are noisy in the axial and coronal planes due to the abrupt transition from one sagittal slice to another, without continuity in the 3D

structure. The latter is most noticeable especially in images with high contrast such as CT scans of the spine, where the gradient of the intensities can be very strong. The intuition behind the pseudo-3D approach is that, to preserve 3D structure and continuity, a model only needs to look at a few of the neighbouring slices [27] rather than several, and optimize for piece-wise continuity. The entire volume will be smoothed out in direction perpendicular to the sagittal plane by transitivity.

4.8 Conclusion

We presented a fully unsupervised registration approach integrating an MRI-based synthesis of CT images of the lumbar spine from T2-weighted MR images used for image-guided spine surgery. We showed the significant differences in image quality between a classical CycleGAN approach using a differentiable histogram loss, and a pseudo-3D model, which takes into account each slice’s neighbouring spatial information. The proposed model yields improved target registration errors for landmarks, demonstrating a good potential for surgical guidance workflows. Future work will include training and testing on larger datasets, using unpaired CT and MRI image sets, as well as conduct a prospective trial to evaluate the registration accuracy in a surgical setting.

CHAPITRE 5 RÉSULTATS COMPLÉMENTAIRES

Étant donné que le corps de ce mémoire est constitué d'un seul article de journal scientifique, nous avons du faire preuve d'un esprit de synthèse pour rédiger ce dernier. Toutefois, ce mémoire nous donnant plus de liberté à l'exhaustivité, nous allons y présenter quelques résultats complémentaires.

Nous commencerons par quelques développements théoriques associés à la méthode des histogrammes dérivables. Ensuite, nous présenterons davantage de résultats qualitatifs et quantitatifs de synthèse IRM-CT pour davantage de cas. Enfin, nous présenterons des résultats complémentaire de recalage 2D/3D.

5.1 Histogrammes dérivables

5.1.1 Développements théoriques

Commençons par démontrer que les histogrammes dérivables donnent bien des histogrammes qui définissent naturellement une densité de probabilité dont l'intégrale s'évalue à 1. Faisant la démonstration pour $H^{\tilde{X}}$, sans perte de généralité pour $H^{\tilde{Y}}$:

$$\sum_{i=0}^{n_{bins}-1} H_i^{\tilde{X}} = \sum_{i=0}^{n_{bins}-1} \frac{\sum_{p=0}^{n-1} u_i(\tilde{X}_p) + kn_{bins}}{n + kn_{bins}^2} = 1 \iff \sum_{i=0}^{n_{bins}-1} \sum_{p=0}^{n-1} u_i(\tilde{X}_p) = n$$

Inversant les indices de la somme, nous obtenons :

$$\sum_{i=0}^{n_{bins}-1} \sum_{p=0}^{n-1} u_i(\tilde{X}_p) = \sum_{p=0}^{n-1} \sum_{i=0}^{n_{bins}-1} u_i(\tilde{X}_p)$$

Nous allons nous intéresser à la somme intérieure et, par changement d'indice, allons annuler les termes intermédiaires de la somme télescopique.

$$\begin{aligned} \sum_{i=0}^{n_{bins}-1} u_i(\tilde{X}_p) &= \sum_{i=0}^{n_{bins}-1} g\left(\tilde{X}_p + \frac{1}{2} - i\right) \\ &\quad - \sum_{i=0}^{n_{bins}-1} g\left(\tilde{X}_p - \frac{1}{2} - i\right) \\ &= g\left(\tilde{X}_p + \frac{1}{2}\right) - g\left(\tilde{X}_p - \frac{1}{2} - (n_{bins} - 1)\right) \\ &= \frac{1}{2} \left[\tanh\left(\alpha\left(\tilde{X}_p + \frac{1}{2}\right)\right) - \tanh\left(\alpha\left(\tilde{X}_p + \frac{1}{2} - n_{bins}\right)\right) \right] \end{aligned}$$

Il faut se référer à la figure 4.2c pour voir que l'expression $\sum_{i=0}^{n_{bins}-1} u_i(\tilde{X}_p)$ décrit une fonction qui s'évalue très près de 1 sur l'intervalle $[0, n_{bins} - 1]$. En réalité, il suffit de l'évaluer à 0 pour voir si l'hypothèse $\sum_{i=0}^{n_{bins}-1} u_i(\tilde{X}_p) = 1$ est justifiée ou non. La figure 4.2c illustre cette idée que $x \in [0, n_{bins}-1] \implies \sum_{i=0}^{n_{bins}-1} u_i(x) \in \left[\sum_{i=0}^{n_{bins}-1} u_i(0), 1 \right)$.

Définissons h tel que :

$$h(\alpha) = \sum_{i=0}^{n_{bins}-1} u_i(0) = \frac{1}{2} \left[\tanh\left(\frac{\alpha}{2}\right) - \tanh\left(\alpha\left(\frac{1}{2} - n_{bins}\right)\right) \right] \quad (5.1)$$

$$h(\alpha) \underset{n_{bins} \gg 0}{\sim} \frac{1}{2} \left[\tanh\left(\frac{\alpha}{2}\right) + 1 \right]$$

α	1	2	3	4	5	6	7	8	9
$h(\alpha)$	0.7311	0.8808	0.9526	0.9820	0.9933	0.9975	0.9991	0.9997	0.9999

Tableau 5.1 Valeurs des seuils $h(\alpha)$ pour différentes valeurs de alpha α . Les $h(\alpha)$ sont définis tel que $\sum_{i=0}^{n_{bins}-1} u_i(\tilde{X}_p) \in [h(\alpha), 1)$ pour $\tilde{X}_p \in [0, n_{bins} - 1]$

Ainsi, il est possible d'utiliser l'expression $h(\alpha)$ pour quantifier à quel point est-ce que la méthode produit toujours des histogrammes constituant une densité, en fonction de l'hyperparamètre α .

Dans un second temps, nous allons présenter un cas de figure simple pour illustrer l'effet

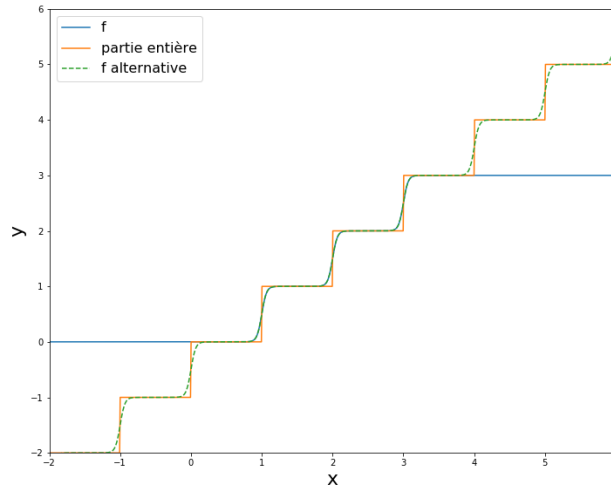


Figure 5.1 Schémas de la formulation alternative de f sans la restriction d'ensemble d'arrivée.

que produit une optimisation directe par différence d'histogrammes et le discuter.

Supposons un vecteur X de taille n dont toutes les entrées ont été échantillonnées selon $\mathcal{N}(0, 1)$ et où n est suffisamment grand pour que les histogrammes produits soient symétriques et qu'ils définissent une courbe gaussienne. Soit $Y = -X$. Cherchons à optimiser les histogrammes avec le gradient par rapport à Y . Nous avons toutefois $|H^Y - H^X| = 0$ et le gradient est nul puisqu'on est déjà à un minimum.

Nous présentons ce cas de figure limite pour illustrer le fait suivant : la métrique d'histogrammes se contente d'optimiser les histogrammes en tant que tel, mais l'information de correspondance entre les histogrammes de X et les histogrammes de Y est perdue.

Une approche que nous avons esquissé, pour ce type de problème en particulier, repose sur l'utilisation de la matrice d'histogrammes conjoints $H^{X,Y}$. Cette approche se résume à minimiser, par rapport à Y , tous les termes de $H^{X,Y}$ qui ne sont pas sur la diagonale. Une approche équivalente revient à maximiser les termes sur la diagonale.

Pour illustrer l'intuition derrière cette approche, revenons au cas précédent avec $Y = -X$, contre un autre cas où $Y = X$. Pour le dernier cas, $H^{X,Y}$ est une matrice diagonale. Pour le premier cas, $H^{X,Y}$ est la même matrice que le premier cas, avec les mêmes termes sur chaque ligne, mais ces termes-là sont décalés à d'autres colonnes. On peut voir pourquoi minimiser les termes qui ne sont pas sur la diagonale aboutit à une correspondance d'histogrammes, puisque pour chaque ligne, la somme des termes reste constante : les histogrammes de X sont constants. Parmi tous les indices p de X et Y ayant des intensités contribuant au terme H_i^X , l'approche va sélectionner les indices des valeurs qui contribuent au terme H_i^X sans contribuer au terme H_i^Y . L'objectif d'optimisation est donc de faire en sorte que les valeurs à ces indices-là contribuent à H_i^Y jusqu'à ce qu'on ait $H_i^Y = H_i^X$. C'est cette sélection d'indices qui permet de retrouver la correspondance entre les histogrammes de X et de Y . Autrement dit, de retrouver comment est répartie, sur Y , la densité de masse provenant de X . Il est important de noter que ceci reste une esquisse d'approche, et que cette dernière est à explorer et à expérimenter.

Enfin, nous présenterons un dernier résultat pour la méthode des histogrammes. Ce résultat constitue une formulation alternative de notre fonction f qui se substitue à la fonction partie entière. Cette formulation a l'avantage de ne pas utiliser de somme et offre donc une meilleure complexité algorithmique. Cette formulation est la suivante :

$$f : x \mapsto \min \left(\max \left(0, \lfloor x \rfloor - 1 + g(x - \lfloor x \rfloor) + g((x - 1) - \lfloor x - 1 \rfloor - 1) \right), n_{bins} - 1 \right) \quad (5.2)$$

Il s'agit d'une composition de fonctions choisies avec attention pour reproduire le résultat de la formulation originale. Les opérateurs min et max servent à restreindre l'ensemble d'arri-

vée de l'expression qui est y enchâssée à l'intervalle $[0, n_{bins}-1]$. Cette expression-là définit la fonction périodique illustrée à la figure 5.1

5.1.2 Synthèse d'image

Par souci d'exhaustivité, nous présenterons dans cette partie davantage de résultats qualitatifs comparant les images synthétisées par les différents modèles.

Tout d'abord, la figure 5.2 montre l'effet de créneau produit par l'utilisation de coupes sa-

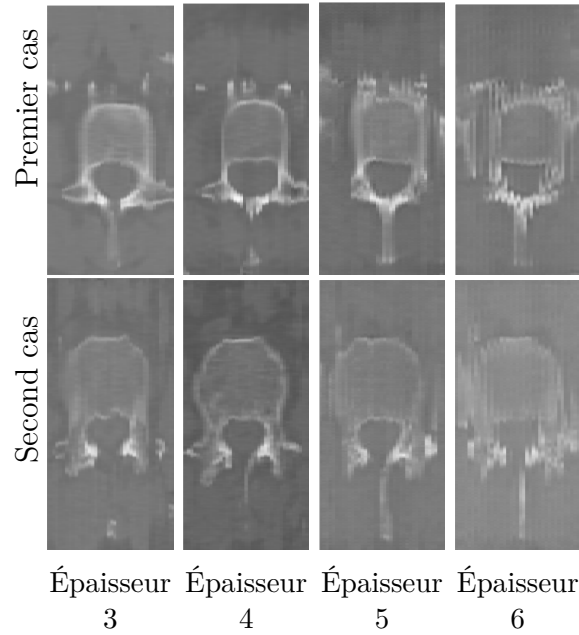


Figure 5.2 Comparaison de CT synthétiques sur le plan axial pour différentes épaisseurs de coupes sagittales, pour deux cas différents.

gittales bien trop épaisses. La figure 5.3 appuie cette observation de manière quantitative. Avec ces résultats-là, nous pouvons affirmer que l'utilisation de coupes d'épaisseurs 3 ou d'épaisseur 4 donne des résultats comparables, d'où le choix arbitraire d'utiliser des coupes d'épaisseur 4.

La figure 5.4 montre une différence entre les modèles 2D et les modèles Pseudo-3D. Cette différence-là est visible tant sur les coupes axiales que les coupes sagittales. Les images synthétisées par les CycleGAN pseudo-3D, que ce soit avec ou sans la fonction de perte par histogrammes, sont comparables à quelques subtiles nuances près. Il nous est possible d'observer un épaississement des contours des vertèbres pour l'approche à perte d'histogrammes sur les quelques images que nous nous présentons. Il est important de noter que, malgré les

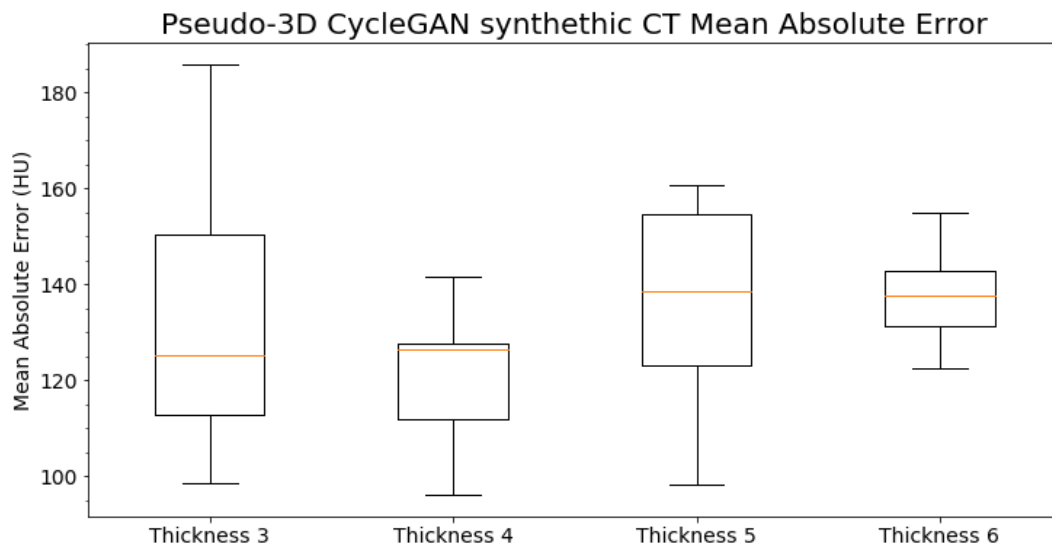


Figure 5.3 Erreurs absolues moyennes avec différentes épaisseurs de coupes sagitales pour l'ensemble du jeu de données

observations que nous présentons, il faudrait un ensemble de données bien plus important pour pouvoir en tirer des conclusions.

Nous inspectons de plus près davantage d'histogrammes d'images synthétisées par les modèles pseudo-3D avec et sans la perte par histogrammes. Nous remarquons, lorsque le modèle utilise la fonction de perte à histogrammes, que les modes de distribution d'intensité ont une légère tendance à être mieux alignés avec la référence du vrai CT.

5.2 Recalage 2D/3D

Nous avons des radiographies et une tomodensitométrie d'un fantôme à disposition. Ce dernier a été utilisé au cours du développement, du test et de la validation de l'outil de recalage, en isolation du reste de la méthode. Les images utilisées sont d'une taille de 768x768 pixels. Nous avons utilisé le même protocole de marquage des 6 repères anatomique que celui du chapitre 4, pour mesurer les erreurs de recalage.

Nous présentons ci-dessous des résultats de test sur ce fantôme.

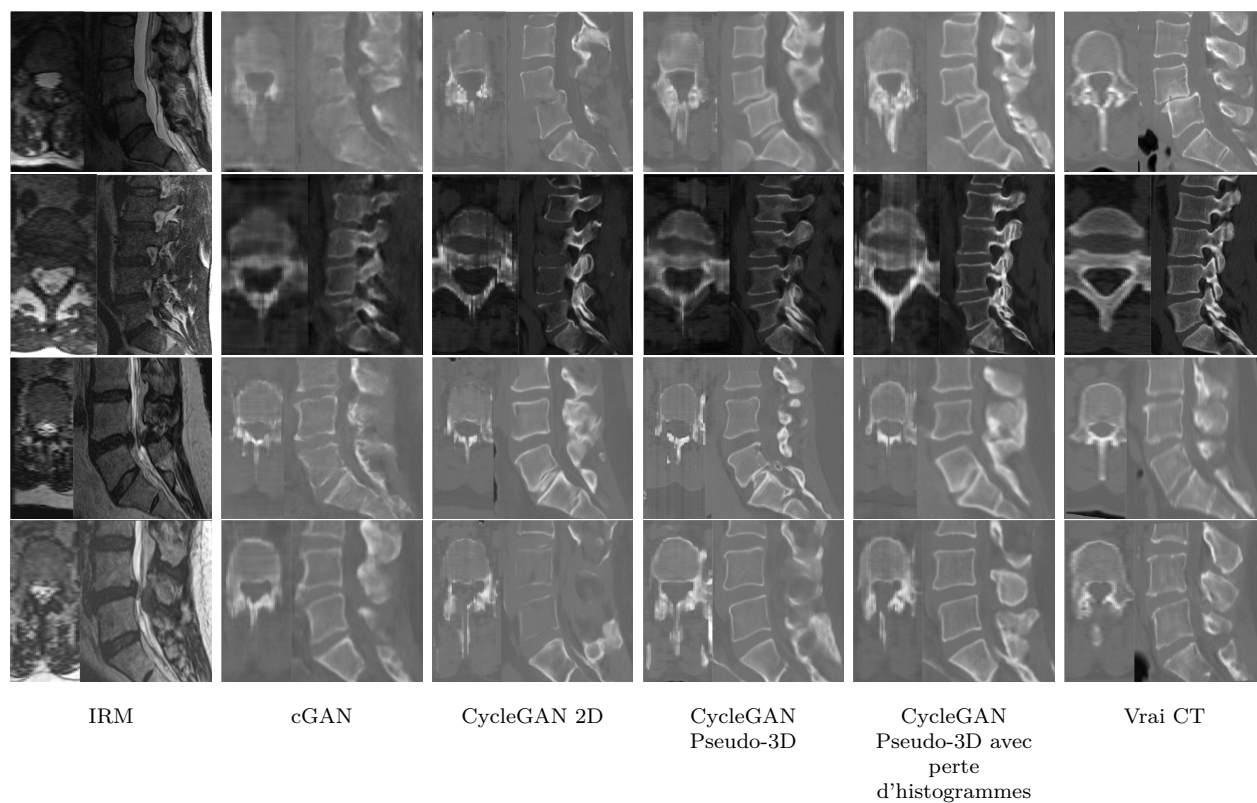


Figure 5.4 Résultats comparatifs pour différents modèles de synthèse IRM-CT pour des volumes de 4 patients différents, avec un volume par ligne. Les images, pour chaque patient, ont été réglées pour être visualisée sur la même plage de valeurs pour que les contrastes soient équitablement comparables.

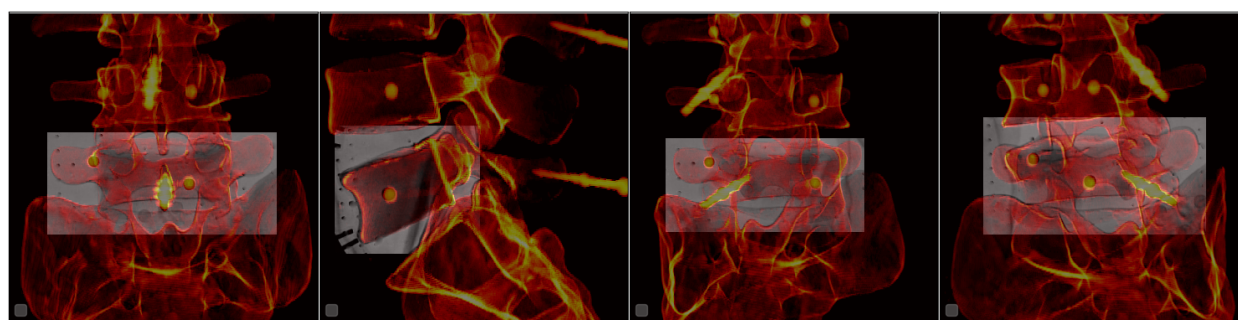


Figure 5.6 Ensemble des 4 vues pour le recalage de la vertèbre L5 du fantôme. Le calque en rouge représente la radiographie synthétique de ce fantôme alors que le fond en nuances de gris représente la radiographie réelle.

Le CT du fantôme étant d'une résolution bien meilleure que les CT du jeu de données, on

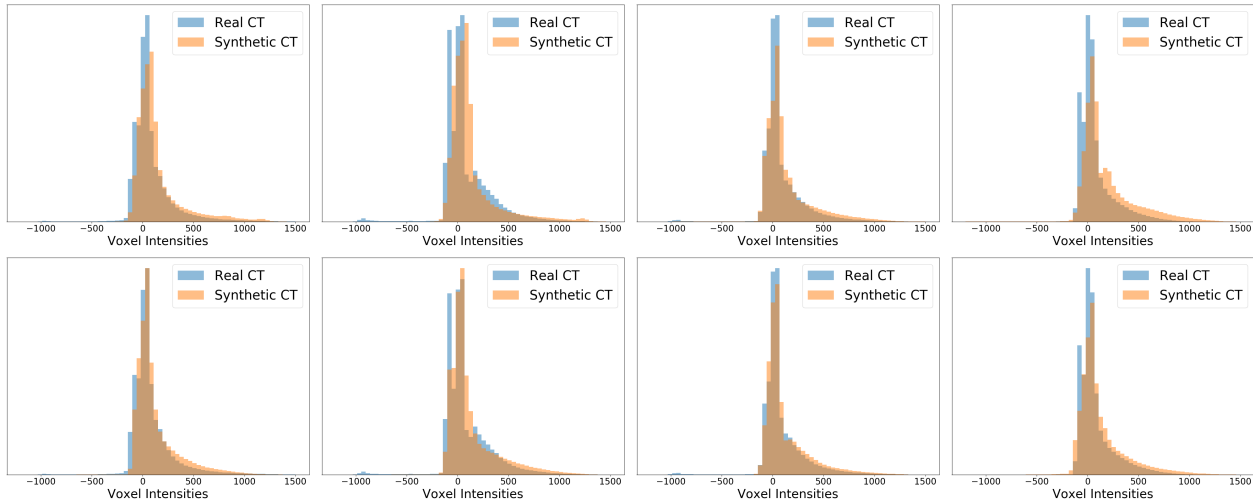


Figure 5.5 Exemples d’histogrammes d’intensités pour 4 volumes de patients différents, avec un volume par colonne. La première ligne correspond aux images synthétisées par le modèle pseudo-3D sans perte d’histogrammes. La seconde ligne correspond aux images synthétisées par le modèle pseudo-3D avec perte d’histogrammes.

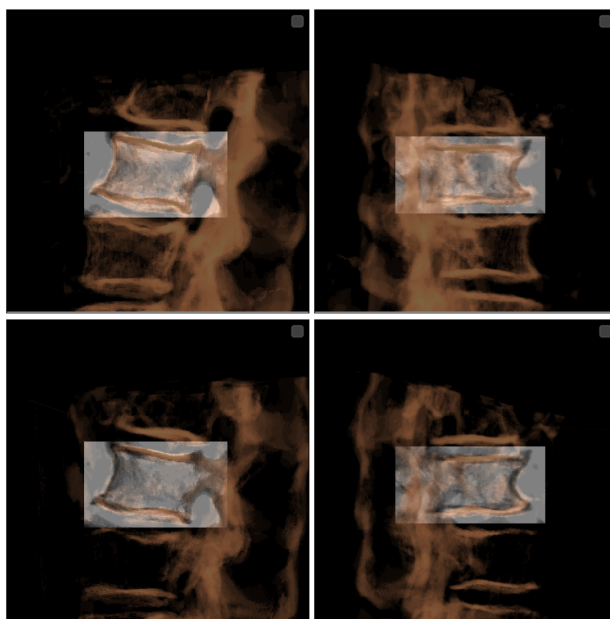
Erreur	L3	L4	L5
Erreur RMS 3D (mm)	2.14	2.39	2.75
Erreur RMS 2D (pixels)	10.92	12.11	14.00

Tableau 5.2 Erreurs de recalage pour le fantôme de vertèbres lombaires

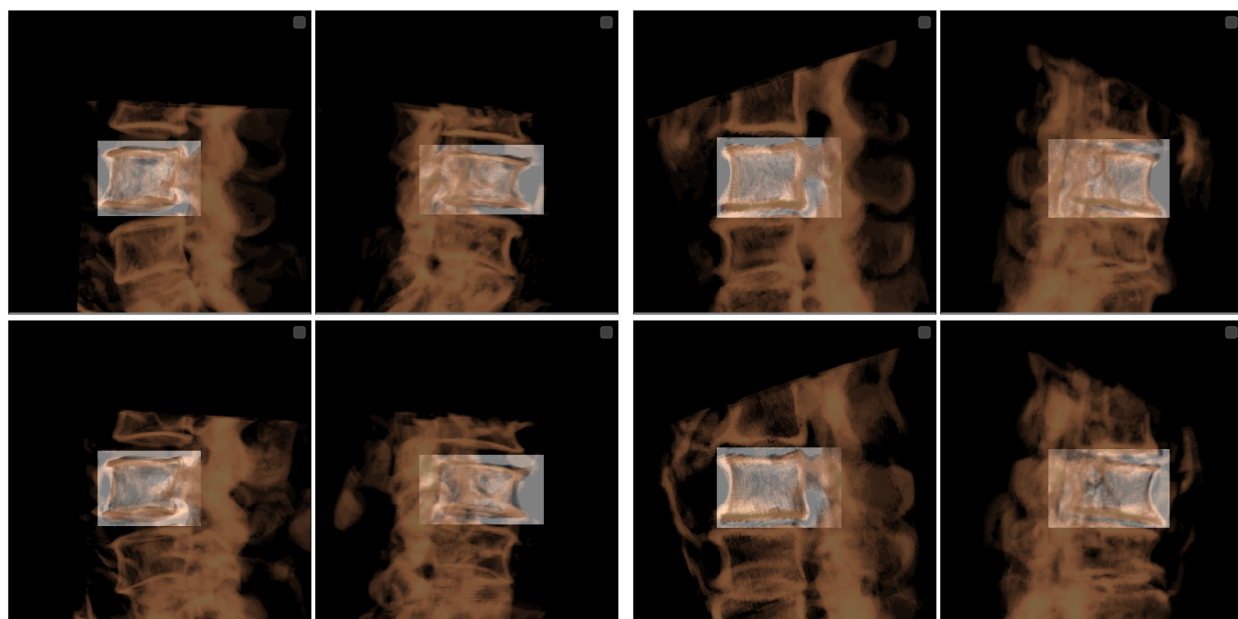
remarque un meilleur rendu au niveau des DRR.

Enfin, et toujours dans le même souci d’exhaustivité, nous présentons davantage de cas de recalage 2D/3D pour la méthode au complet. Ces résultats sont obtenus en utilisant les tomodensitométries synthétiques.

La figure 5.7 présente les mêmes éléments de différence de DRR énoncés précédemment au chapitre 4. On observe que le modèle pseudo-3D régulier tend à produire des volumes pour lesquels les intensités au niveau des structures osseuses tombe en dessous du seuil des 300 HU que l’on utilise pour générer les DRR. Cela est surtout visible au niveau de la figure 5.7a. Dans ces cas-là, le recalage a surtout été possible grâce à une bonne initialisation manuelle ainsi qu’à l’alignement des 4 vues. Le fait d’avoir ces 4 vues en simultané permet d’alléger effets de manque de contours. Ces effets ne sont tout de même pas désirables et produisent une configuration qui n’est pas optimale. Les images issues des deux modèles présentent quelques artefacts qui peuvent parfois rendre le recalage difficile. Ces artefacts-là sont observables et se manifestent sous la forme de tâches isolées. Un dernier point que l’on observe est le manque de netteté au niveau des apophyses.



(a) Premier cas



(b) Second cas

(c) Troisième cas

Figure 5.7 Exemples pour 3 cas différents. Pour chaque cas, les images de la première ligne ont été générées avec des volumes issus du modèle avec perte d'histogrammes, alors que les images de la deuxième ligne ont été générées à partir des volumes issus du modèle pseudo-3D régulier. Les vues présentées sont la vue latérale et la vue à -45°

CHAPITRE 6 DISCUSSION GÉNÉRALE

Le travail autour duquel tourne la rédaction de ce mémoire porte sur le sujet bien défini et bien délimité du développement d’une méthode de recalage 2D/3D d’IRM et de fluoroscopie de bras en C, en utilisant la synthèse de tomodensitométrie par apprentissage profond. Ce sujet de recherche, malgré le cadre structuré dans lequel il se définit, pose un défi en terme d’implémentation et d’expérimentation pour atteindre les objectifs souhaités. Ceci est principalement dû à la complexité intrinsèque aux méthodes de réseaux génératifs adverses, ainsi qu’à l’intégration de plusieurs composantes pour former la méthode au complet : jeu de données, procédure d’entraînement, procédure de validation, synthèse d’image, recalage automatique, validation du recalage par l’opérateur. Cette nature composite de la méthode introduit plusieurs sources d’erreur possibles.

Pour minimiser toutes ces sources d’erreurs, il a longuement fallu expérimenter avec différentes configurations pour différents paramètres, et voire même essayer d’intégrer de nouvelles composantes. D’autres approches ont été essayées, mais nous n’avons toutefois retenu que celles qui sont présentées dans ce mémoire. Ceci dit, nous avons été capables de présenter une méthode au complet, la tester sur notre ensemble de données, et la valider de manière quantitative et qualitative. Le travail présenté incorpore aussi deux nouvelles composantes qui poussent cette méthode un peu plus loin. Nous allons, dans ce chapitre, discuter plus en détail du travail de recherche effectué et des résultats obtenus pour les différentes composantes de la méthode, avant de conclure ce mémoire au chapitre suivant.

6.1 Synthèse IRM-CT

Pour la synthèse de tomodensitométrie, nous nous sommes tout d’abord dirigés vers le modèle cGAN, étant donné que nous disposions de données appairées et que ce modèle-là est reconnu comme étant supérieur aux autres modèles tel que le CycleGAN. Ceci dit, contrairement à d’autres tâches telles que la classification, la validation quantitative, durant l’entraînement, d’images produites par transfert de domaine est un problème ouvert. Le moyen le plus fiable reste l’inspection visuelle des données synthétisées, mais celui-ci est coûteux en terme de temps. Cela est d’autant plus vrai qu’un modèle prend typiquement quelques jours pour compléter son entraînement. Les premières expériences ont d’abord montré une capacité du cGAN à capturer la complexité de l’ensemble d’entraînement, constitué d’environ 800 coupes sagitales. Nous nous sommes rendus compte par la suite que ce modèle-là ne généralise pas bien aux données non observées. C’est là que nous nous sommes tournés vers le CycleGAN. Ce

dernier démontrait une meilleure capacité de généralisation pour un ensemble de données de la taille du nôtre. L’hypothèse que l’on émet pour expliquer cette observation est la suivante. Étant donné que le cGAN suppose une correspondance pixel à pixel, la procédure d’optimisation va ajuster une partie non-négligeable du bruit statistique, associé à l’imperfection de cette correspondance ainsi qu’à la faible quantité de données. Le CycleGAN n’a toutefois pas besoin de correspondance pixel à pixel, ce qui élimine déjà cette source d’erreur-là. Par ailleurs, sa procédure d’optimisation se contente de générer des images vraisemblables qui puissent être utilisées pour reconstruire les images en entrée. Cette contrainte est plus faible que celle du cGAN, mais elle reste suffisante pour obtenir le résultat escompté.

L’approche classique pour appliquer l’apprentissage profond à un volume 3D est d’utiliser les coupes 2D de ce dernier. C’est cette approche que nous avons d’abord adopté. Dans notre cas, étant donné que le CycleGAN apprend une translation d’image bidirectionnelle, nous pouvions aussi inspecter les IRM synthétisées à partir des CT. Ces IRM synthétiques ne comportaient pas, ou peu de discontinuités au niveau des coupes axiales et coronales, contrairement aux CT synthétiques. Cette incohérence a soulevé des questionnements. Nous avons par la suite remarqué que ces discontinuités étaient tout de même présentes dans les IRM synthétiques, mais qu’elles étaient subtiles. C’est en réalité le fait que le CT aie un contraste bien plus élevé au niveau des structures osseuses qui fait ressortir ces discontinuités-là. Les gradients d’intensité peuvent être très forts au sein d’un CT de la colonne vertébrale, contrairement à une IRM.

Ces observations nous ont poussé à essayer l’approche pseudo-3D pour le CycleGAN, et cette approche-là a prouvé une amélioration de la qualité des images synthétiques tant pour les coupes sagitales que pour les coupes axiales et coronales. Nous avons aussi proposé une stratégie de combinaison de quadruplets de coupes pour former un volume, inspirée du concept d’interpolation par morceaux sur des intervalles qui se chevauchent. L’idée est de pouvoir adapter un modèle par morceaux à une séquence de données, et que ces morceaux-là se chevauchent. Quant à l’amélioration de la synthèse pour les coupes sagitales, nous nous sommes d’abord demandé quelle pouvait bien en être la raison. L’hypothèse que l’on émet pour expliquer ceci est que la pseudo-3D est aussi une forme d’augmentation de données. En effet, la même coupe peut se trouver à différentes positions dans différents quadruplets. Ainsi, le même modèle voit la même coupe, mais de plusieurs manières différentes.

Enfin, nous nous sommes inspirés des méthodes de recalage multimodal par information mutuelle pour travailler sur une fonction de perte d’information mutuelle optimisable par rétropropagation. Cette fonction de perte nécessite une estimation des densités de probabilités par histogrammes, mais cette opération-là n’est pas dérivable. C’est là que nous avons développé notre méthode d’histogrammes dérivables. Les premières expériences avec cette

métrique d'information mutuelle, avec notre modèle et avec notre ensemble de données n'ont pas montré de bons résultats. Une réflexion un peu plus poussée nous a mené à penser la chose suivante. Optimiser l'information mutuelle entre X et Y revient à optimiser le fait que Y se rapproche le plus possible d'une fonction déterministe de X tel que $Y = f(X)$. Dans le cas du recalage multimodal, cela fonctionne bien. Dans le nôtre, la fonction f fait parti du problème plutôt que de la solution, f est en réalité le générateur de notre réseau adverse. Nous avons toutefois gardé la composante d'histogrammes dérivables. Nous l'avons utilisée pour optimiser la distribution des intensités d'images synthétiques afin qu'elle soit similaire à celle des données réelles. Les résultats ont montré une légère amélioration des distributions d'intensités.

6.2 Recalage 2D/3D

La partie subséquente de la méthode reposant sur un recalage 2D/3D en utilisant les tomodensitométries synthétiques, nous avons initialement constaté une différence de contraste au niveau des DRR générées à partir des CT synthétiques. Il faut noter que la distribution des intensités est importante pour la génération de ces DRR, et qu'elle influe beaucoup sur leur apparence et sur la visibilité des détails qui y sont. Ces détails et reliefs sont très importants et déterminent la qualité de recalage obtenu, tant pour la procédure d'optimisation que pour la validation visuelle finale. Pour plusieurs cas de recalage, sans avoir utilisé la fonction de perte par histogrammes, nous nous sommes rendu compte d'une différence d'échelle au niveau des intensités. Les effets observés ont été tel que certains contours de vertèbres manquaient de netteté, alors que d'autres parties n'étaient juste pas très visibles.

Ces effets-là étaient moins marqués lorsque nous avons expérimenté avec le modèle utilisant la fonction de perte par histogrammes. Lorsqu'on génère une radiographie synthétique, on fixe un seuil d'intensité en dessous duquel on ne prend pas un voxel en compte pour le calcul de l'intégrale-ligne. Nous n'avons pas eu besoin d'ajuster ce seuil pour le modèle à histogrammes, contrairement aux expériences faites avec le modèle régulier. Il a fallu donc plus d'ajustements manuels pour avoir des erreurs acceptables en utilisant ce dernier.

Cela dit, ces observations, toutefois prometteuses, restent qualitatives, subjectives et empiriques, d'autant plus que l'ensemble de données utilisé n'est pas assez exhaustif pour que l'on puisse affirmer une amélioration significative apportée par la fonction de perte à histogrammes.

6.3 Limitations

La première limitation du travail effectué durant ce projet de recherche reste la non-exhaustivité du jeu de données. Les effets de cette limitation sont doubles. Tout d’abord, étant donné que les performances des modèles d’apprentissage profond dépendent directement des données utilisées, nous ne pouvons jamais réellement établir un résultat et l’extrapoler ailleurs. Il faut donc être conscient que les résultats présentés ont été établis pour le jeu de données utilisé. D’autre part, les expériences de test et de validation dépendent de ces données-là. Même en supposant que les modèles soient entraînés avec des données suffisamment larges et diversifiées, la capacité de généralisation de ces modèles à des données qui n’ont jamais été vues auparavant reste à prouver de manière plus exhaustive. Dans notre cas, étant donné que nous avons utilisé des radiographies synthétiques comme substituts de fluoroscopies de bras en C, une limitation de la méthode proposée apparaît dans le fait qu’elle n’a pas été testée sur des radiographies réelles.

En ce qui concerne la méthode en elle-même, nous avons noté que le temps d’entraînement (4 jours) est problématique. Ceci rend les différentes itérations très difficiles et très coûteuses en terme de temps. Cela est d’autant plus vrai que le domaine dans lequel s’inscrit ce travail est fortement compétitif et connaît un rythme effréné.

Par ailleurs, comme l’ont montré la figure 5.7 ainsi que la figure 5.4 au chapitre précédent, les apophyses peuvent ne pas être fidèlement synthétisés. Cela vient s’ajouter à la présence de d’artefacts aux bords de quelques images.

Le facteur de complexité de cette méthode est aussi une limitation en soi. Ce travail s’inscrit dans une vision et une ambition d’application clinique avec un impact sur la fluidité des procédure. L’étape de recalage automatique reste non triviale, mais nous espérons que les futures et potentielles améliorations pour la synthèse d’image viendront palier à cela.

6.4 Perspectives futures

Durant les développements de ce travail et durant les différentes expérimentations qui ont été faites, nous avons abordé la possibilité d’utilisation de données non appairées. Nous avons collecté des jeux de données publiques et effectué quelques expériences. Le jeu de données obtenu, étant constitué plusieurs jeux de données de sources différentes, comportait une hétérogénéité intrinsèque. Les expériences n’ont pas été fructueuses. Nous espérons toutefois pouvoir revisiter la méthode de nouveau avec des données non appairées, mais cette fois-ci avec plus de contrôle sur ces données-là. Cela serait une éventuelle solution au problème de non exhaustivité des données, puisque les données appairées sont généralement plus difficiles

à obtenir. Par ailleurs, des esquisses d'architectures ont été faites, reposant sur la combinaisons de données appairées avec des données non appairées. Celles-ci seraient intéressantes à revisiter elles aussi. Finalement, des tests et des évaluations plus exhaustives, dans un contexte clinique, donneraient à la méthode présentée, un réel aboutissement.

CHAPITRE 7 CONCLUSION

Ce mémoire a commencé par introduire le contexte dans lequel se place la problématique du projet de recherche à son centre. Nous avons présenté la plus-value recherchée ainsi que les différents enjeux, tant sur le plan technique que sur le plan clinique.

La difficulté du recalage 2D/3D d'imagerie par résonance magnétique et de fluoroscopie de bras en C a longtemps été problématique.

Ce sont les récentes avancées et applications en terme d'apprentissage profond qui ont donné naissance à l'idée de ce projet de recherche. Premièrement, nous avons exposé la problématique clinique, les objectifs et l'impact souhaité. Par la suite, nous avons présenté le contexte technologique et scientifique pour l'approche de cette même problématique.

Durant la réalisation de ce projet, nous avons exploré différents modèles de type adverses génératifs pour la synthèse de tomodensitométrie à partir d'IRM. Pour notre ensemble de données, nous avons expérimenté avec des modèles existants. Nous y avons par la suite ajouté deux composantes : une composante pseudo-3D et une composante de perte d'histogrammes. Nous avons été capables de montrer des différences entre les différents modèles à différentes composantes. Notre méthode atteint des coefficients DICE de 84 ± 1.5 entre les segmentations d'images synthétiques et les segmentations d'images réelles.

Une fois capables de générer des tomodensitométries synthétiques, nous avons effectué nos expériences de recalage 2D/3D. Nous avons atteint une erreur moyenne de 2.1 ± 0.2 mm. Là aussi, nous avons été capables de montrer comment les différences constatées pour la synthèse d'image transparaissent au moment du recalage.

La méthode présentée comporte toutefois plusieurs limitations ; la non-exhaustivité des données étant la plus constatable. Des améliorations sont aussi encore possibles pour la synthèse d'image et la robustesse du recalage. De plus, étant donné la difficulté d'entraînement des réseaux adverses génératifs, et du temps que demandent les différentes itérations, il est encore possible de perfectionner les modèles de synthèse d'image.

Enfin, nous souhaitons conclure sur l'espoir que les avancées à venir en apprentissage profond viennent davantage compléter la présente méthode dans ses perspectives futures d'aboutissement clinique et d'innovation scientifique.

BIBLIOGRAPHIE

- [1] P. Markelj, D. Tomaževič, B. Likar et F. Pernuš, “A review of 3d/2d registration methods for image-guided interventions,” *Medical image analysis*, vol. 16, n°. 3, p. 642–661, 2012.
- [2] A. Hamadeh, S. Lavalée et P. Cinquin, “Automated 3-dimensional computed tomographic and fluoroscopic image registration,” *Computer Aided Surgery : Official Journal of the International Society for Computer Aided Surgery (ISCAS)*, vol. 3, n°. 1, p. 11–19, 1998.
- [3] Y. Otake, A. S. Wang, A. Uneri, G. Kleinszig, S. Vogt, N. Aygun, L. L. Sheng-fu, J.-P. Wolinsky, Z. L. Gokaslan et J. H. Siewerdsen, “3d–2d registration in mobile radiographs : algorithm development and preliminary clinical evaluation,” *Physics in Medicine & Biology*, vol. 60, n°. 5, p. 2075, 2015.
- [4] R. Harmouche, F. Cheriet, H. Labelle et J. Dansereau, “3d registration of mr and x-ray spine images using an articulated model,” *Computerized Medical Imaging and Graphics*, vol. 36, n°. 5, p. 410–418, 2012.
- [5] D. Tomazevic, F. Pernus *et al.*, “Robust gradient-based 3-d/2-d registration of ct and mr to x-ray images,” *IEEE Transactions on Medical Imaging*, vol. 27, n°. 12, p. 1704–1714, 2008.
- [6] T. De Silva, A. Uneri, M. Ketcha, S. Reaungamornrat, J. Goerres, M. Jacobson, S. Vogt, G. Kleinszig, A. J. Khanna, J. Wolinsky *et al.*, “Registration of mri to intraoperative radiographs for target localization in spinal interventions,” *Physics in Medicine & Biology*, vol. 62, n°. 2, p. 684, 2017.
- [7] M. Van der Bom, J. Pluim, M. Gounis, E. van de Kraats, S. Sprinkhuizen, J. Timmer, R. Homan et L. Bartels, “Registration of 2d x-ray images to 3d mri by generating pseudo-ct data,” *Physics in Medicine & Biology*, vol. 56, n°. 4, p. 1031, 2011.
- [8] A. Torrado-Carvajal, J. L. Herraiz, E. Alcain, A. S. Montemayor, L. Garcia-Cañamaque, J. A. Hernandez-Tamames, Y. Rozenholc et N. Malpica, “Fast patch-based pseudo-ct synthesis from t1-weighted mr images for pet/mr attenuation correction in brain studies,” *Journal of Nuclear Medicine*, vol. 57, n°. 1, p. 136–143, 2016.
- [9] Z. Yi, H. Zhang, P. Tan et M. Gong, “Dualgan : Unsupervised dual learning for image-to-image translation,” dans *Proceedings of the IEEE international conference on computer vision*, 2017, p. 2849–2857.

- [10] J.-Y. Zhu, T. Park, P. Isola et A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” dans *Proceedings of the IEEE international conference on computer vision*, 2017, p. 2223–2232.
- [11] X. Han, “Mr-based synthetic ct generation using a deep convolutional neural network method,” *Medical physics*, vol. 44, n^o. 4, p. 1408–1419, 2017.
- [12] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang et D. Shen, “Medical image synthesis with context-aware generative adversarial networks,” dans *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, p. 417–425.
- [13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville et Y. Bengio, “Generative adversarial nets,” dans *Advances in neural information processing systems*, 2014, p. 2672–2680.
- [14] J. M. Wolterink, A. M. Dinkla, M. H. Savenije, P. R. Seevinck, C. A. van den Berg et I. Išgum, “Deep mr to ct synthesis using unpaired data,” dans *International Workshop on Simulation and Synthesis in Medical Imaging*. Springer, 2017, p. 14–23.
- [15] A. Frangi, S. Tsafaris et J. Prince, “Simulation and synthesis in medical imaging,” *IEEE Trans. on Med. Imag.*, vol. 37, n^o. 3, p. 673, 2018.
- [16] G. W. Sherouse, K. Novins et E. L. Chaney, “Computation of digitally reconstructed radiographs for use in radiotherapy treatment design,” *International Journal of Radiation Oncology* Biology* Physics*, vol. 18, n^o. 3, p. 651–658, 1990.
- [17] M. Drozdal, E. Vorontsov, G. Chartrand, S. Kadoury et C. Pal, “The importance of skip connections in biomedical image segmentation,” dans *Deep Learning and Data Labeling for Medical Applications*. Springer, 2016, p. 179–187.
- [18] J. Long, E. Shelhamer et T. Darrell, “Fully convolutional networks for semantic segmentation,” dans *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, p. 3431–3440.
- [19] M. Shakeri, I. Nahle, E. Finley et S. Kadoury, “Inter-vertebral disk modelling from pairs of segmented vertebral models using trainable pre-processing networks,” dans *Proc. IEEE ISBI*. IEEE, 2018, p. 1122–1125.
- [20] A. Krizhevsky, I. Sutskever et G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” dans *Advances in neural information processing systems*, 2012, p. 1097–1105.
- [21] P. Isola, J.-Y. Zhu, T. Zhou et A. A. Efros, “Image-to-image translation with conditional adversarial networks,” dans *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, p. 1125–1134.

- [22] K. He, X. Zhang, S. Ren et J. Sun, “Deep residual learning for image recognition,” dans *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, p. 770–778.
- [23] J. Johnson, A. Alahi et L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” dans *European conference on computer vision*. Springer, 2016, p. 694–711.
- [24] V. Dumoulin et F. Visin, “A guide to convolution arithmetic for deep learning,” *arXiv preprint arXiv :1603.07285*, 2016.
- [25] D. Ulyanov, A. Vedaldi et V. Lempitsky, “Instance normalization : The missing ingredient for fast stylization,” *arXiv :1607.08022*, 2016.
- [26] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever et R. Salakhutdinov, “Dropout : a simple way to prevent neural networks from overfitting,” *The journal of machine learning research*, vol. 15, n°. 1, p. 1929–1958, 2014.
- [27] Q. Yu, Y. Xia, L. Xie, E. K. Fishman et A. L. Yuille, “Thickened 2d networks for 3d medical image segmentation,” *arXiv preprint arXiv :1904.01150*, 2019.
- [28] D. P. Kingma et J. Ba, “Adam : A method for stochastic optimization,” 2014.
- [29] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang et R. Webb, “Learning from simulated and unsupervised images through adversarial training,” dans *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, p. 2107–2116.
- [30] F. Jacobs, E. Sundermann, B. De Sutter, M. Christiaens et I. Lemahieu, “A fast algorithm to calculate the exact radiological path through a pixel or voxel space,” *Journal of computing and information technology*, vol. 6, n°. 1, p. 89–94, 1998.
- [31] R. Hartley et A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [32] O. Ronneberger, P. Fischer et T. Brox, “U-net : Convolutional networks for biomedical image segmentation,” dans *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, p. 234–241.
- [33] I. E. Sutherland, “Three-dimensional data input by tablet,” *Proceedings of the IEEE*, vol. 62, n°. 4, p. 453–461, 1974.
- [34] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.*, “Scipy 1.0—fundamental algorithms for scientific computing in python,” *arXiv preprint arXiv :1907.10121*, 2019.
- [35] J. L. Prince et J. M. Links, *Medical imaging signals and systems*. Pearson Prentice Hall Upper Saddle River, 2006.

- [36] D. Weishaupt, V. D. Köchli et B. Marincek, *How does MRI work ? : an introduction to the physics and function of magnetic resonance imaging*. Springer Science & Business Media, 2008.
- [37] A. A. Patel, P. G. Whang et A. R. Vaccaro, “Overview of computer-assisted image-guided surgery of the spine,” dans *Seminars in Spine Surgery*, vol. 20, n°. 3. Elsevier, 2008, p. 186–194.
- [38] R. W. Brown, Y.-C. N. Cheng, E. M. Haacke, M. R. Thompson et R. Venkatesan, *Magnetic resonance imaging : physical principles and sequence design*. John Wiley & Sons, 2014.
- [39] G. Bradski, “The OpenCV Library,” *Dr. Dobb’s Journal of Software Tools*, 2000.
- [40] N. Halko, P.-G. Martinsson et J. A. Tropp, “Finding structure with randomness : Stochastic algorithms for constructing approximate matrix decompositions,” 2009.
- [41] A. N. Tikhonov, “On the stability of inverse problems,” dans *Dokl. Akad. Nauk SSSR*, vol. 39, 1943, p. 195–198.
- [42] K. S. Arun, T. S. Huang et S. D. Blostein, “Least-squares fitting of two 3-d point sets,” *IEEE Transactions on pattern analysis and machine intelligence*, n°. 5, p. 698–700, 1987.
- [43] R. L. Siddon, “Fast calculation of the exact radiological path for a three-dimensional ct array,” *Medical physics*, vol. 12, n°. 2, p. 252–255, 1985.
- [44] J. Bijhold, “Three-dimensional verification of patient placement during radiotherapy using portal images,” *Medical physics*, vol. 20, n°. 2, p. 347–356, 1993.
- [45] E. Bullitt, A. Liu, S. R. Aylward, C. Coffey, J. Stone, S. K. Mukherji, K. E. Muller et S. M. Pizer, “of 3d cerebral vessels with 2d digital angiograms : Clinical evaluation,” *Academic Radiology*, vol. 6, n°. 9, p. 539–546, 1999.
- [46] D. Zikic, M. Groher, A. Khamene et N. Navab, “Deformable registration of 3d vessel structures to a single projection image,” dans *Medical Imaging 2008 : Image Processing*, vol. 6914. International Society for Optics and Photonics, 2008, p. 691412.
- [47] S. A. Banks et W. A. Hodge, “Accurate measurement of three-dimensional knee replacement kinematics using single-plane fluoroscopy,” *IEEE Transactions on Biomedical Engineering*, vol. 43, n°. 6, p. 638–649, 1996.
- [48] P. J. Besl et N. D. McKay, “Method for registration of 3-d shapes,” dans *Sensor fusion IV : control paradigms and data structures*, vol. 1611. International Society for Optics and Photonics, 1992, p. 586–606.

- [49] J. Feldmar, N. Ayache et F. Betting, “3d–2d projective registration of free-form curves and surfaces,” *Computer vision and image understanding*, vol. 65, n°. 3, p. 403–424, 1997.
- [50] M. Fleute et S. Lavallée, “Nonrigid 3-d/2-d registration of images using statistical models,” dans *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 1999, p. 138–147.
- [51] J. Hermans, P. Claes, J. Bellemans, D. Vandermeulen et P. Suetens, “Robust initialization for 2d/3d registration of knee implant models to single-plane fluoroscopy,” dans *Medical Imaging 2007 : Image Processing*, vol. 6512. International Society for Optics and Photonics, 2007, p. 651208.
- [52] S. Lavallée et R. Szeliski, “Recovering the position and orientation of free-form objects from image contours using 3d distance maps,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 17, n°. 4, p. 378–390, 1995.
- [53] G. Zheng, M. Á. Ballester, M. Styner et L.-P. Nolte, “Reconstruction of patient-specific 3d bone surface from 2d calibrated fluoroscopic images and point distribution model,” dans *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2006, p. 25–32.
- [54] H. Livyatan, Z. Yaniv et L. Joskowicz, “Robust automatic c-arm calibration for fluoroscopy-based navigation : a practical approach,” dans *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2002, p. 60–68.
- [55] D. Tomazevic, B. Likar, T. Slivnik et F. Pernus, “3-d/2-d registration of ct and mr to x-ray images,” *IEEE Transactions on Medical Imaging*, vol. 22, n°. 11, p. 1407–1416, 2003.
- [56] P. Markelj, D. Tomaževič, F. Pernuš et B. Likar, “Robust 3-d/2-d registration of ct and mr to x-ray images based on gradient reconstruction,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 3, n°. 6, p. 477–483, 2008.
- [57] Y. Otake, A. S. Wang, J. W. Stayman, A. Uneri, G. Kleinszig, S. Vogt, A. J. Khanna, Z. L. Gokaslan et J. H. Siewerdsen, “Robust 3d–2d image registration : application to spine interventions and vertebral labeling in the presence of anatomical deformation,” *Physics in Medicine & Biology*, vol. 58, n°. 23, p. 8535, 2013.
- [58] J. H. Hipwell, G. P. Penney, R. A. McLaughlin, K. Rhode, P. Summers, T. C. Cox, J. V. Byrne, J. A. Noble et D. J. Hawkes, “Intensity-based 2-d-3-d registration of cerebral angiograms,” *IEEE transactions on medical imaging*, vol. 22, n°. 11, p. 1417–1426, 2003.

- [59] T. De Silva, A. Uneri, M. Ketcha, S. Reaungamornrat, G. Kleinszig, S. Vogt, N. Aygun, S. Lo, J. Wolinsky et J. Siewerdsen, “3d–2d image registration for target localization in spine surgery : investigation of similarity metrics providing robustness to content mismatch,” *Physics in Medicine & Biology*, vol. 61, n°. 8, p. 3009, 2016.
- [60] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. Hill *et al.*, “A comparison of similarity measures for use in 2-d-3-d medical image registration,” *IEEE transactions on medical imaging*, vol. 17, n°. 4, p. 586–595, 1998.
- [61] S. Clippe, D. Sarrut, C. Malet, S. Miguet, C. Ginestet et C. Carrie, “Patient setup error measurement using 3d intensity-based image registration techniques,” *International Journal of Radiation Oncology* Biology* Physics*, vol. 56, n°. 1, p. 259–265, 2003.
- [62] C. E. Shannon, “A mathematical theory of communication,” *Bell system technical journal*, vol. 27, n°. 3, p. 379–423, 1948.
- [63] E. Parzen, “On estimation of a probability density function and mode,” *The annals of mathematical statistics*, vol. 33, n°. 3, p. 1065–1076, 1962.
- [64] J. A. Nelder et R. Mead, “A simplex method for function minimization,” *The computer journal*, vol. 7, n°. 4, p. 308–313, 1965.
- [65] F. Gao et L. Han, “Implementing the nelder-mead simplex algorithm with adaptive parameters,” *Computational Optimization and Applications*, vol. 51, n°. 1, p. 259–277, 2012.
- [66] W. Rudin *et al.*, *Principles of mathematical analysis*. McGraw-hill New York, 1964, vol. 3.
- [67] OpenStax, *Anatomy Physiology*. Houston, TX : OpenStax CNS, 2016. [En ligne]. Disponible : <http://cnx.org/contents/14fb4ad7-39a1-4eee-ab6e-3ef2482e3e22>
- [68] W. Commons. (2019) Bras en c biplanair. [En ligne]. Disponible : <https://upload.wikimedia.org/wikipedia/commons/d/d8/Biplanar.jpg>
- [69] ——. (2019) Vertèbre typique. [En ligne]. Disponible : https://upload.wikimedia.org/wikipedia/commons/thumb/e/ed/Vertebra_Posterolateral-en.svg/480px-Vertebra_Posterolateral-en.svg.png
- [70] ——. (2019) Géométrie épipolaire. [En ligne]. Disponible : https://upload.wikimedia.org/wikipedia/commons/thumb/1/14/Epipolar_geometry.svg/440px-Epipolar_geometry.svg.png
- [71] ——. (2019) Transformation par noyau. [En ligne]. Disponible : https://upload.wikimedia.org/wikipedia/commons/thumb/f/fe/Kernel_Machine.svg/2880px-Kernel_Machine.svg.png

- [72] ——. (2019) Perceptron multi-couche. [En ligne]. Disponible : https://upload.wikimedia.org/wikipedia/commons/thumb/e/e4/Artificial_neural_network.svg/350px-Artificial_neural_network.svg.png
- [73] ——. (2019) Schémas d'acquisition d'une tomodensitométrie. [En ligne]. Disponible : https://upload.wikimedia.org/wikipedia/commons/thumb/b/bb/Cone_Beam_CT_principle.png/440px-Cone_Beam_CT_principle.png
- [74] R. O'Rahilly et F. Müller, *Basic human anatomy : a regional study of human structure*. WB Saunders Company, 1983.
- [75] R. D. A. W. V. A. Mitchell, *Gray's Anatomy for Students*. Elsevier, 2019.
- [76] A. Zisserman, "Lecture notes in optimisation," 2011. [En ligne]. Disponible : <http://www.robots.ox.ac.uk/~az/lectures/opt/2011/lect2.pdf>
- [77] —, "Lecture notes in machine learning," 2015. [En ligne]. Disponible : <http://www.robots.ox.ac.uk/~az/lectures/ml/lect3.pdf>
- [78] T. M. Deserno, "Biomedical image processing," 2011.
- [79] S. Sayad, "Lecture notes in data science," 2019.
- [80] Y. Fukuoka, A. Hoshino et A. Ishida, "A simple radiographic measurement method for polyethylene wear in total knee arthroplasty," *IEEE transactions on rehabilitation engineering*, vol. 7, n^o. 2, p. 228–233, 1999.
- [81] M. J. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *The computer journal*, vol. 7, n^o. 2, p. 155–162, 1964.
- [82] J. H. Mathews, *Numerical methods for mathematics, science and engineering*. Prentice-Hall, 1992.
- [83] D. Fu et G. Kuduvali, "A fast, accurate, and automatic 2d–3d image registration for image-guided cranial radiosurgery," *Medical physics*, vol. 35, n^o. 5, p. 2180–2194, 2008.
- [84] E. L. Mansfield et G. R. W. Quispel, "On the construction of discrete gradients," 2009.
- [85] A. Gramfort, "Quasi-newton optimization," 2014.
- [86] R. Fletcher, *Practical methods of optimization*. John Wiley & Sons, 2013. [En ligne]. Disponible : http://www.lix.polytechnique.fr/bigdata/mathbigdata/wp-content/uploads/2014/09/notes_quasi_newton.pdf
- [87] L. A. Gatys, A. S. Ecker et M. Bethge, "Image style transfer using convolutional neural networks," dans *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, p. 2414–2423.

- [88] K. Simonyan et A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv :1409.1556*, 2014.
- [89] M. D. Zeiler et R. Fergus, “Visualizing and understanding convolutional networks,” dans *European conference on computer vision*. Springer, 2014, p. 818–833.
- [90] N. Jones, “Computer science : The learning machines,” *Nature News*, vol. 505, n°. 7482, p. 146, 2014.
- [91] S. Devic, “Mri simulation for radiotherapy treatment planning,” *Medical physics*, vol. 39, n°. 11, p. 6701–6711, 2012.
- [92] R. G. Price, J. P. Kim, W. Zheng, I. J. Chetty et C. Glide-Hurst, “Image guided radiation therapy using synthetic computed tomography images in brain cancer,” *International Journal of Radiation Oncology* Biology* Physics*, vol. 95, n°. 4, p. 1281–1289, 2016.
- [93] J. M. Edmund et T. Nyholm, “A review of substitute ct generation for mri-only radiation therapy,” *Radiation Oncology*, vol. 12, n°. 1, p. 28, 2017.
- [94] E. Johnstone, J. J. Wyatt, A. M. Henry, S. C. Short, D. Sebag-Montefiore, L. Murray, C. G. Kelly, H. M. McCallum et R. Speight, “Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging–only radiation therapy,” *International Journal of Radiation Oncology* Biology* Physics*, vol. 100, n°. 1, p. 199–217, 2018.
- [95] J. Uh, T. E. Merchant, Y. Li, X. Li et C. Hua, “Mri-based treatment planning with pseudo ct generated through atlas registration,” *Medical physics*, vol. 41, n°. 5, p. 051711, 2014.
- [96] A. Johansson, M. Karlsson et T. Nyholm, “Ct substitute derived from mri sequences with ultrashort echo time,” *Medical physics*, vol. 38, n°. 5, p. 2708–2714, 2011.
- [97] T. Huynh, Y. Gao, J. Kang, L. Wang, P. Zhang, J. Lian et D. Shen, “Estimating ct image from mri data using structured random forest and auto-context model,” *IEEE transactions on medical imaging*, vol. 35, n°. 1, p. 174–183, 2015.
- [98] Y. Lei, J. Harms, T. Wang, Y. Liu, H.-K. Shu, A. B. Jani, W. J. Curran, H. Mao, T. Liu et X. Yang, “Mri-only based synthetic ct generation using dense cycle consistent generative adversarial networks,” *Medical physics*, 2019.
- [99] R. Li, W. Zhang, H.-I. Suk, L. Wang, J. Li, D. Shen et S. Ji, “Deep learning based imaging data completion for improved brain disease diagnosis,” dans *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2014, p. 305–312.

- [100] A. Géron, *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems.* " O'Reilly Media, Inc.", 2017.
- [101] V. Vapnik, "Pattern recognition using generalized portrait method," *Automation and remote control*, vol. 24, p. 774–780, 1963.
- [102] K. P. Murphy, *Machine learning : a probabilistic perspective.* MIT press, 2012.
- [103] W. L. Buntine, "Operations for learning with graphical models," *Journal of artificial intelligence research*, vol. 2, p. 159–225, 1994.
- [104] H. G. Sung, "Gaussian mixture regression and classification," Thèse de doctorat, Rice University, 2004.
- [105] D. Koller et N. Friedman, *Probabilistic graphical models : principles and techniques.* MIT press, 2009.
- [106] F. Rosenblatt, "Principles of neurodynamics. perceptrons and the theory of brain mechanisms," Cornell Aeronautical Lab Inc Buffalo NY, Rapport technique, 1961.
- [107] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner *et al.*, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, n°. 11, p. 2278–2324, 1998.
- [108] I. H. Witten, E. Frank, M. A. Hall et C. J. Pal, *Data Mining : Practical machine learning tools and techniques.* Morgan Kaufmann, 2016.
- [109] A. Vedaldi et K. Lenc, "Matconvnet : Convolutional neural networks for matlab," dans *Proceedings of the 23rd ACM international conference on Multimedia.* ACM, 2015, p. 689–692.
- [110] P. Sturm, *Pinhole Camera Model.* Boston, MA : Springer US, 2014, p. 610–613. [En ligne]. Disponible : https://doi.org/10.1007/978-0-387-31439-6_472
- [111] H. J. Johnson, M. M. McCormick et L. Ibanez.
- [112] M. Lilja, "Fast and accurate voxel projection technique in free-form cone-beam geometry with application to algebraic reconstruction," dans *2008 First International Symposium on Applied Sciences on Biomedical and Communication Technologies.* IEEE, 2008, p. 1–5.
- [113] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang et S. Paul Smolley, "Least squares generative adversarial networks," dans *Proceedings of the IEEE International Conference on Computer Vision*, 2017, p. 2794–2802.
- [114] M. Arjovsky, S. Chintala et L. Bottou, "Wasserstein gan," *arXiv preprint arXiv :1701.07875*, 2017.

- [115] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin et A. C. Courville, “Improved training of wasserstein gans,” dans *Advances in neural information processing systems*, 2017, p. 5767–5777.
- [116] D.-A. Beaupre et G.-A. Bilodeau, “Siamese cnns for rgb-lwir disparity estimation,” dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, p. 0–0.
- [117] H. Li et Y. Fan, “Non-rigid image registration using fully convolutional networks with deep self-supervision,” *arXiv preprint arXiv :1709.00799*, 2017.
- [118] M. Drozdal *et al.*, “Learning normalized inputs for iterative estimation in medical image segmentation,” *Medical image analysis*, vol. 44, p. 1–13, 2018.
- [119] R. Hecht-Nielsen, “Theory of the backpropagation neural network,” dans *Neural networks for perception*. Elsevier, 1992, p. 65–93.
- [120] R. Korez *et al.*, “Model-based segmentation of vertebral bodies from mr images with 3d cnns,” dans *Proc. Int’l Conf. on MICCAI*. Springer, 2016, p. 433–441.
- [121] P. Richards et J. George, “Diagnostic ct radiation and cancer induction,” *Skeletal radiology*, vol. 39, n^o. 5, p. 421–424, 2010.
- [122] D. Tomazevic, B. Likar et F. Pernus, “3-d/2-d registration by integrating 2-d information in 3-d,” *IEEE transactions on medical imaging*, vol. 25, n^o. 1, p. 17–27, 2005.

ANNEXE A ARCHITECTURE FCN ET FC-RESNET

Dans cette partie, nous présentons l'architecture du FCN et FC-ResNet tirés de (Drozdal et al., 2017). Celle-ci repose sur l'utilisation de deux réseaux profonds en cascade. Le FCN, en premier, sert à normaliser les images en entrée. Le FC-ResNet, quant à lui, effectue la segmentation avec une fonction de perte DICE de la forme :

$$\mathcal{L}_{DICE}(X, Y) = -\frac{\sum_i X_i Y_i + 1}{\sum_i X_i + \sum_i Y_i + 1} \quad (\text{A.1})$$

Layer name	Block type	Output resolution	Output width	Repetition number
Input	–	512×512	1	–
Down 1	conv 3×3	512×512	16	2
Pooling 1	maxpooling	256×256	16	1
Down 2	conv 3×3	256×256	32	2
Pooling 2	maxpooling	128×128	32	1
Down 3	conv 3×3	128×128	64	2
Pooling 3	maxpooling	64×64	64	1
Down 4	conv 3×3	64×64	128	2
Pooling 4	maxpooling	32×32	128	1
Across	conv 3×3	32×32	256	2
Up 1	upsampling	64×64	256	1
Merge 1	concatenate	64×64	384	1
Up 2	conv 2×2	64×64	128	1
Up 3	conv 3×3	64×64	128	2
Up 4	upsampling	128×128	128	1
Merge 2	concatenate	128×128	192	1
Up 5	conv 2×2	128×128	64	1
Up 6	conv 3×3	128×128	64	2
Up 7	upsampling	256×256	64	1
Merge 3	concatenate	256×256	96	1
Up 8	conv 2×2	256×256	32	1
Up 9	conv 3×3	256×256	32	2
Up 10	upsampling	512×512	32	1
Merge 4	concatenate	512×512	48	1
Up 11	conv 2×2	512×512	16	1
Up 12	conv 3×3	512×512	16	2
Output	conv 3×3	512×512	1	1

Figure A.1 Architecture du réseau de type FCN. Image tirée de (Drozdal et al., 2017)

