

Méthode automatique pour la localisation et l'identification de
structures osseuses dans les radiographies EOS

par

Roseline OLORY AGOMMA

THÈSE PRÉSENTÉE À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
COMME EXIGENCE PARTIELLE À L'OBTENTION
DU DOCTORAT EN GÉNIE
Ph.D.

MONTRÉAL, LE 15 NOVEMBRE 2021

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC



Roseline Olory Agomma, 2021



Cette licence Creative Commons signifie qu'il est permis de diffuser, d'imprimer ou de sauvegarder sur un autre support une partie ou la totalité de cette oeuvre à condition de mentionner l'auteur, que ces utilisations soient faites à des fins non commerciales et que le contenu de l'oeuvre n'ait pas été modifié.

PRÉSENTATION DU JURY

CETTE THÈSE A ÉTÉ ÉVALUÉE

PAR UN JURY COMPOSÉ DE:

M. Carlos Vazquez, directeur de thèse
Département de génie logiciel et des TI à l'École de technologie supérieure

M. Jacques de Guise, co-directeur
Département de génie des systèmes à l'École de technologie supérieure

M. Matthiew Toews, co-directeur
Département de génie des systèmes à l'École de technologie supérieure

M. Ismail Ben Ayed, président du jury
Département de génie des systèmes à l'École de technologie supérieure

M. Marco Pedersoli, membre du jury
Département de génie des systèmes à l'École de technologie supérieure

M. Benjamin De Leener, examinateur externe indépendant
Département de génie informatique et génie logiciel à Polytechnique Montréal

ELLE A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 08 NOVEMBRE 2021

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Je tiens à remercier mon directeur de recherche Carlos Vázquez pour son encadrement, sa disponibilité, ses conseils et son soutien indéfectible. Merci de m'avoir épaulée, de m'encourager et d'avoir cru en moi. Je remercie mon co-directeur Jacques de Guise de m'avoir donné l'opportunité de faire partie du laboratoire de recherche en imagerie et orthopédie (LIO). Merci pour tes conseils et suggestions. Je remercie également mon co-directeur Matthew Toews pour ses recommandations et moments d'échanges. Grand merci à vous trois. Vous êtes des professeurs extrêmement passionnés dans vos champs d'expertise respectifs, compréhensifs et soucieux du bien-être de vos étudiants. Collaborer avec vous a été très enrichissant.

Je souhaite également remercier Thierry Cresson pour sa supervision continue tout au long de ma thèse, ses suggestions pertinentes et pour ces magnifiques moments d'échanges fructueux.

J'aimerais remercier les professeurs Ismail Ben Ayed, Marco Pedersoli et Benjamin De Leener d'avoir accepté de faire partie de mon jury de thèse.

Je remercie mes collègues du LIO, particulièrement Ted et Amiel, avec qui j'ai passé de bons moments de partage, de connaissance et d'amusement. Merci à Christine, Lulu, Magali et Francis pour leurs soutiens. Je remercie également mes amis Maxime, Atma, Agathe, Mathieu, Yvan, Philip et Junior pour leur soutien indéniable. Merci à Cathérine pour les bonnes bouffes camerounaises et les périodes d'amusement.

J'aimerais exprimer toute ma gratitude à la compagnie EOS Imaging, à MEDTEQ, à MITACS et au CRSNG RDC, pour avoir financé ce projet de recherche. Merci également à la Fondation Arbour de m'avoir comptée parmi ses boursiers.

Je remercie ma mentore Diane de Champlain, pour sa sollicitude, ses conseils avisés, son désir ardent de m'aider à atteindre mon plein potentiel. À chacune de nos rencontres, je me sens invincible, plus confiante et prête à relever de nouveaux défis.

Merci à mon frère, Carel Olory et à Charbel Boco, pour leurs précieux conseils et soutiens tout au long de ce périple que représente le doctorat. Dans les moments d'intempéries, vous avez su me motiver, me pousser à donner le meilleur de moi-même et à toujours garder la tête hors de l'eau lorsque je risquais de me noyer.

J'aimerais exprimer toute ma gratitude à l'endroit de mon père, Théophile Olory Agomma, et de ma mère, Cécile Ahounou, pour m'avoir soutenue et encouragée durant cette thèse. Vous êtes mes modèles. Ce que je suis actuellement, je vous le dois.

Finalement, je remercie également mon frère, Suzon Olory, et mes sœurs, Eunice Olory et Cilia Olory, pour leur soutien indéfectible et leur amour sans limites. Merci d'avoir été de tout cœur avec moi tout au long de cette thèse.

Je tiens également à remercier tous ceux, qui de près ou de loin, m'ont soutenue durant cette thèse de doctorat.

Méthode automatique pour la localisation et l'identification de structures osseuses dans les radiographies EOS

Roseline OLORY AGOMMA

RÉSUMÉ

Le calcul de paramètres cliniques et la reconstruction tridimensionnelle de structures osseuses à partir de radiographies biplans en pratique clinique sont des tâches essentielles dans le domaine de l'imagerie médicale, particulièrement en orthopédie, pour établir le diagnostic et planifier le geste médical. Les processus de mesures cliniques et de reconstruction 3D sont basés sur la tâche d'extraction d'informations (par exemple : identifier sur des radiographies de face et de profil le centre de la tête fémorale, segmenter le tibia). Cette tâche d'extraction, principalement réalisée manuellement par des professionnels de la santé, est répétitive, chronophage et peut conduire à des résultats avec une grande variabilité. De ce fait, elle pourrait entraîner des erreurs lors du processus de reconstruction 3D ou entraîner un mauvais calcul des paramètres cliniques (par exemple : l'angle cervico-diaphysaire, l'angle fémoro-tibial), importants lors de la prise de décision avant, pendant et après la chirurgie ; ce qui pourrait compromettre la santé du patient.

Dans les stéréoradiographies EOS (images issues du système de radiologie EOS à faible dose d'irradiation) présentant de fortes superpositions osseuses, différents points de vue (images de corps entier, du membre inférieur ou images avec structures partielles) et différentes orientations (0° , 20° , 45° , 70° ou 90°), cette tâche d'extraction est d'autant plus ardue. Cela explique la non-existence d'une méthode automatique d'extraction d'informations des membres inférieurs couvrant cette diversité d'images EOS. La non-résolubilité de cette problématique explique qu'actuellement en pratique clinique, la phase d'initialisation des processus de mesures cliniques et de reconstruction tridimensionnelle de structures osseuses pour les membres inférieurs, phase consistant à identifier des primitives d'intérêts (exemple : sphère, points) sur chaque structure osseuse soit encore réalisée manuellement.

L'objectif principal de ce travail est de pallier cette problématique en automatisant la tâche d'extraction d'informations liées aux membres inférieurs sur des images EOS présentant différents points de vue et orientations en vue d'initialiser automatiquement les processus de mesures cliniques et de reconstruction 3D pour les radiographies EOS.

Nous avons développé une méthode entièrement automatique de segmentation de structures osseuses et de localisation de points d'intérêt des membres inférieurs pour les radiographies EOS. L'approche de segmentation proposée est basée sur une approche d'augmentation de données permettant la génération d'images contenant à la fois des structures partielles et complètes et sur une nouvelle architecture neuronale appelée RobustNet. L'approche de localisation des points d'intérêt est principalement basée sur un réseau siamois qui prend en entrée des paires d'images EOS et prédit en sortie des points 3D.

Les méthodes de segmentation de structures osseuses et de localisation de points d'intérêt des membres inférieurs proposées ont été respectivement testées sur 70 et 30 images EOS présentant

différents points de vue (images frontales et sagittales du membre inférieur et de corps entier) et orientations du patient (0° , 90° et 45°). Les résultats obtenus sont prometteurs pour pouvoir être intégrés dans la plateforme logicielle d'analyse d'images EOS.

Ce travail permettra d'extraire automatiquement des structures osseuses, des régions ou points d'intérêt des membres inférieurs dans des banques de données d'images EOS. Il permettra d'initialiser automatiquement les processus de mesures cliniques et de reconstruction tridimensionnelle des os des membres inférieurs et permettra une diminution du temps de traitement effectué sur les radiographies EOS.

Mots-clés: méthode automatique, segmentation d'os, identification de points, points stéréo-correspondants, réseaux de neurones, reconstruction 3D, calcul de paramètres cliniques

Automatic method for localization and identification of bone structures in EOS radiographs

Roseline OLORY AGOMMA

ABSTRACT

The calculation of clinical parameters and the three-dimensional reconstruction of the bone structures from biplane radiographs in clinical practice are essential tasks in the field of medical imaging, particularly in orthopedics, to establish the diagnosis and plan the medical procedure. The processes of clinical measurements and 3D reconstruction are based on the task of extracting information (for example : identifying on frontal and lateral radiographs the center of the femoral head, segmenting the tibia). This extraction task, mainly performed manually by healthcare professionals, is repetitive, time consuming and can lead to results with greater variability. As a result, it could lead to 3D reconstruction error and miscalculation of clinical parameters (for example : neck-shaft angle, femoro-tibial angle), which are important when making a decision before, during and after surgery ; which could compromise the patient's health.

In EOS stereo-radiographs (images from the low-dose EOS radiology system) presenting strong bone superpositions, different fields of view (full-body images, lower limb images or images with partial structures) and orientations (0° , 20° , 45° , 70° or 90°), this extraction task is more difficult. This explains the non-existence of an automatic method of extracting lower limb information covering this diversity of EOS images. The non-resolvability of this problem explains why currently in clinical practice, the initialization step of the processes of clinical measurement and three-dimensional reconstruction of the lower-limb bone structures step consisting in identifying the primitives of interest (for example : sphere, points) on each bone structure is still performed manually.

The main objective of this work is to overcome this problem by automating the task of extracting information related to the lower limbs on EOS images presenting different points of view and orientations in order to automatically initiate the processes of clinical measurements and 3D reconstruction for EOS radiographs.

We have developed a fully automatic method of segmenting bone structures and identifying landmarks of the lower limbs for EOS radiographs. The proposed segmentation approach is based on a data augmentation approach, allowing the generation of images containing both partial and complete structures and on a new neural architecture called RobustNet. The landmark identification approach is mainly based on a Siamese network which takes EOS image pairs as input and predicted three-dimensional points as output.

The proposed methods for bone structures segmentation and lower limb landmark identification were respectively tested on 70 and 30 EOS images presenting different points of view (frontal and sagittal images of the lower limb and of the whole body) and the orientations of the patient (0° , 90° and 45°). The results obtained are promising to be integrated into the EOS image analysis software platform for future evaluation.

This work will make it possible to automatically extract bone structures, regions or points of interest of the lower limbs in databases of EOS images. It will make it possible to automatically initiate the process of clinical measurements and three-dimensional reconstruction of the lower limb bone structures and will allow a reduction in the processing time performed on EOS radiographs.

Keywords: automatic method, bone segmentation, point identification, stereo-corresponding points, neural network, 3D reconstruction, calculation of clinical parameters

TABLE DES MATIÈRES

	Page
INTRODUCTION	1
0.1 Structure du document	4
CHAPITRE 1 NOTIONS FONDAMENTALES ET CONTEXTE CLINIQUE	5
1.1 Références anatomiques	5
1.2 Anatomie du membre inférieur	6
1.2.1 Anatomie du fémur	7
1.2.2 Anatomie du tibia	9
1.3 Pathologies liées aux membres inférieurs	10
1.4 Système EOS	12
1.5 Processus disponibles dans le logiciel commercial SterEOS	14
1.6 Conclusion	16
CHAPITRE 2 REVUE DE LITTÉRATURE	21
2.1 Méthodes conventionnelles d'extraction d'informations	21
2.1.1 Méthodes conventionnelles de segmentation d'objets	21
2.1.1.1 Méthodes classiques	21
2.1.1.2 Modèles déformables	22
2.1.1.3 Méthodes par atlas	23
2.1.1.4 Méthode de coupe de graphe	24
2.1.2 Méthodes conventionnelles d'extraction de caractéristiques	24
2.1.2.1 DéTECTEUR de Harris	25
2.1.2.2 L'algorithme SIFT	25
2.1.2.3 L'algorithme ASIFT	26
2.1.2.4 La transformée de Hough	26
2.1.3 Conclusion	27
2.2 Méthodes d'apprentissage machine pour l'extraction d'informations	27
2.2.1 Modèles de forme statistiques	28
2.2.2 Machines à vecteurs de support	28
2.2.3 Forêts aléatoires	29
2.2.4 Réseaux de neurones	30
2.2.4.1 Réseaux de neurones convolutionnels	31
2.2.4.2 Segmentation par réseau de neurones	39
2.2.4.3 Extraction de caractéristiques à partir de réseaux neuronaux	45
2.3 Extraction d'informations dans les images médicales	50
2.3.1 Segmentation de structures dans les images médicales	50
2.3.2 Identification de points d'intérêt dans les images médicales	54
2.4 Méthodes d'augmentation de données	60
2.5 Synthèse	63

CHAPITRE 3 OBJECTIFS ET CHOIX MÉTHODOLOGIQUES	69
3.1 Contexte et objectifs de recherche	69
3.1.1 Objectifs	70
3.1.1.1 Objectifs spécifiques	70
3.2 Choix méthodologiques	71
CHAPITRE 4 APPROCHE MÉTHODOLOGIQUE GLOBALE	73
4.1 Ensemble de données et définition de la vérité terrain	74
4.1.1 Ensemble de données d'images	74
4.1.1.1 Ensemble de données pour la méthode d'identification et de segmentation des os des membres inférieurs	74
4.1.1.2 Ensemble de données pour la méthode d'identification de points d'intérêt	75
4.1.2 Définition de la vérité terrain	77
4.1.2.1 Définition de la vérité terrain de la méthode d'identification et de segmentation des os du MI	77
4.1.2.2 Définition de la vérité terrain de la méthode d'identification de points d'intérêt	78
4.2 Méthode automatique pour l'identification et la segmentation de structures osseuses des membres inférieurs	79
4.2.1 Détails méthodologiques	81
4.2.1.1 Phase 1 : Méthode d'augmentation de données	82
4.2.1.2 Phase 2 : Architecture neuronale proposée : RobustNet	84
4.2.2 Détails d'implémentation	90
4.3 Méthode automatique pour la localisation et l'identification de points d'intérêt des membres inférieurs	90
4.3.1 Détails méthodologiques	91
4.3.1.1 Phase 1 : Extraction de boîtes englobantes	91
4.3.1.2 Phase 2 : Entraînement du réseau siamois	94
4.3.2 Détail d'implémentation	95
4.4 Méthode de validation	95
4.4.1 Métrique de validation de la méthode d'identification et de segmentation des os du MI	96
4.4.2 Métrique de validation de la méthode de localisation et d'identification des points d'intérêt	97
4.5 Résumé	97
CHAPITRE 5 RÉSULTATS	99
5.1 Résultats de la méthode d'identification et de segmentation de structures osseuses des membres inférieurs	99
5.1.1 Résultats RobustNet	99
5.1.2 Influence du choix de techniques d'augmentation des données sur la prédiction	100
5.1.3 RobustNet versus SegNet	106

5.2	Résultats de la méthode de localisation et l'identification des points d'intérêt des membres inférieurs	107
5.2.1	Boîtes englobantes et masques comme données d'entrée	109
5.2.2	Boîtes englobantes comme données d'entrée	111
CHAPITRE 6 DISCUSSION ET CONTRIBUTIONS DE LA THÈSE		117
6.1	Discussion	117
6.1.1	Discussion sur la méthode d'identification et de segmentation de structures osseuses	117
6.1.2	Discussion sur la méthode de localisation et d'identification de points d'intérêt	120
6.2	Contributions de la thèse	123
CONCLUSION ET RECOMMANDATIONS		125
7.1	Conclusion	125
7.2	Limitations et recommandations	126
7.3	Publications scientifiques	128
7.3.1	Revue scientifique et communications internationales	129
7.3.2	Congrès et communications locaux	129
7.3.3	Prix et distinction	131
ANNEXE I	IDENTIFICATION DE POINTS D'INTÉRÊT DANS LES IMAGES EOS VIA L'ALGORITHME ASIFT ET LA TRANSFORMÉE DE HOUGH	133
BIBLIOGRAPHIE		137

LISTE DES TABLEAUX

	Page	
Tableau 2.1	Synthèse de quelques approches neuronales de segmentation	47
Tableau 2.2	Synthèse de performance de différentes méthodes de segmentation dans les images médicales	55
Tableau 2.3	Synthèse de performance de différentes méthodes d'identification de points d'intérêt dans les images médicales (distance euclidienne utilisée comme métrique)	60
Tableau 4.1	Couleur et pourcentage moyen de pixels associés à chaque classe	77
Tableau 4.2	Liste des points d'intérêt et leurs numéros associés	80
Tableau 4.3	Champ réceptif de chaque couche de SegNet et de RobustNet dans la partie encodeur.	89
Tableau 5.1	Résultats par classe pour la totalité de notre ensemble de données de test	100
Tableau 5.2	Résultats par classe pour les images frontales de notre ensemble de données	100
Tableau 5.3	Résultats par classe pour les images latérales de notre ensemble de données	100
Tableau 5.4	Comparaison des résultats de segmentation selon la méthode d'augmentation de données utilisée (%).....	103
Tableau 5.5	RobustNet vs SegNet pour l'ensemble de données de test (%).....	106
Tableau 5.6	Erreurs (mm) par point pour le fémur (cas où boîtes englobantes et masques considérés comme entrée)	109
Tableau 5.7	Erreurs (mm) par point pour le tibia (cas où boîtes englobantes et masques considérés comme entrée)	111
Tableau 5.8	Erreurs (mm) par point pour le fémur (cas où uniquement boîtes englobantes considérées comme entrée)	115
Tableau 5.9	Erreurs (mm) par point pour le tibia (cas où uniquement boîtes englobantes considérées comme entrée)	115

LISTE DES FIGURES

	Page	
Figure 0.1	Radiographies EOS en orientation $0^\circ/90^\circ$ et $45^\circ/45^\circ$	3
Figure 1.1	Repères anatomiques Tirée de Chaibi (2010).....	6
Figure 1.2	Squelette du membre inférieur Tirée de Soins-infirmiers (2021).....	7
Figure 1.3	Anatomie du fémur	8
Figure 1.4	Anatomie du tibia	9
Figure 1.5	Arthrose de la hanche.....	10
Figure 1.6	Arthrose du genou	11
Figure 1.7	Genus varum, normal et valgum Tirée de Hossenbaccus (2021)	12
Figure 1.8	Processus d'acquisition de radiographies EOS Tirée de Illés & Somoskeöy (2012).....	13
Figure 1.9	Exemples d'images radiographiques EOS	14
Figure 1.10	Initialisation manuelle des processus de mesures cliniques et de reconstruction 3D du MI Tirée de SterEOS	15
Figure 1.11	Estimation des paramètres cliniques des MI Tirée de SterEOS.....	16
Figure 1.12	Déformation du modèle générique surfacique Tirée de SterEOS.....	17
Figure 1.13	Reconstruction 3D personnalisée des os des MI Tirée de SterEOS	18
Figure 2.1	Architecture d'un réseau de neurones Tirée de Truong (2016)	30
Figure 2.2	Répartition de papiers publiés de 2012 à 2017 Tirée de Litjens, Kooi, Bejnordi, Setio, Ciompi, Ghafoorian, Van Der Laak, Van Ginneken & Sánchez (2017)	32
Figure 2.3	Architecture d'un CNN	32
Figure 2.4	Couche de convolution Tirée de Cohen-Hadria & Peeters (2016)	34
Figure 2.5	Opération de convolution sur une image de profondeur 3	36
Figure 2.6	Opération de convolution dilatée avec un noyau de taille 3×3	36

Figure 2.7	Exemples d'opérations de pooling	37
Figure 2.8	Positionnement général du pooling dans les CNN	38
Figure 2.9	Opération de dropout	38
Figure 2.10	Architecture DeepMask Tirée de Pinheiro, Collobert & Dollár (2015)	40
Figure 2.11	Architecture Mask R-CNN Tirée de He, Gkioxari, Dollár & Girshick (2017).....	40
Figure 2.12	Architecture RetinaMask Tirée de He <i>et al.</i> (2017)	41
Figure 2.13	Architecture YOLACT Tirée de He <i>et al.</i> (2017).....	41
Figure 2.14	Architecture FCN Tirée de Long, Shelhamer & Darrell (2015)	43
Figure 2.15	Architecture DCAN Tirée de Chen, Qi, Yu, Dou, Qin & Heng (2017)	43
Figure 2.16	Architecture U-Net Tirée de Ronneberger, Fischer & Brox (2015).....	45
Figure 2.17	Architecture Deep ResNet Tirée de Zhang, Liu & Wang (2018)	46
Figure 2.18	Architecture SegNet Tirée de Badrinarayanan, Kendall & Cipolla (2017).....	47
Figure 2.19	Réseau siamois	49
Figure 4.1	Processus d'entraînement de la méthode proposée.....	74
Figure 4.2	Images présentant différents points de vue et orientations	76
Figure 4.3	Radiographie EOS, résultat de projection des modèles 3D du MI et masques de vérité terrain associés	78
Figure 4.4	Résultat de projection des points 3D du MI sur de images EOS Tirée de SterEOS	79
Figure 4.5	Position des points d'intérêt du fémur. Reproduite et adaptée avec l'autorisation de LIO (2017)	80
Figure 4.6	Position des points d'intérêt du tibia. Reproduite et adaptée avec l'autorisation de LIO (2017)	81
Figure 4.7	Processssus d'entraînement de la méthode d'identification et de segmentation des os des MI	82

Figure 4.8	Exemples d'erreurs produites par SegNet (Image obtenue à partir de Olory Agomma, Vázquez, Cresson & de Guise (2019))	85
Figure 4.9	Architecture RobustNet	86
Figure 4.10	Visualisation du champ réceptif dans le bloc contextuel ($RF(L2 L0)$ signifie champs réceptif de la couche $L2$ par rapport à la couche $L0$)	88
Figure 4.11	Processus d'entraînement de la méthode de localisation et d'identification des 13 points 3D du fémur	92
Figure 5.1	Résultats qualitatifs de radiographies entières	101
Figure 5.2	Résultats qualitatifs d'images avec structures partielles	102
Figure 5.3	Images partielles contenant un petit morceau d'os	103
Figure 5.4	Cas où la segmentation comporte des erreurs importantes	104
Figure 5.5	Prédiction de RobustNet utilisant ADLG, ADL et ADG	105
Figure 5.6	Courbes de convergence de RobustNet et de SegNet	106
Figure 5.7	Comparaison qualitative entre RobustNet et SegNet	107
Figure 5.8	Quelques limitations de SegNet corrigées par RobustNet	108
Figure 5.9	Identification de points d'intérêt stéréo-correspondants du fémur dans les radiographies EOS	110
Figure 5.10	Identification de points d'intérêt stéréo-correspondants du tibia dans les radiographies EOS	112
Figure 5.11	Images présentant de fortes erreurs de localisation de points d'intérêt stéréo-correspondants du fémur dans les radiographies EOS	113
Figure 5.12	Images présentant de fortes erreurs de localisation de points d'intérêt stéréo-correspondants du tibia dans les radiographies EOS	114
Figure 6.1	Images de genoux sain et arthrosique	122

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

2D	Bidimensionnel
3D	Tridimensionnel
ADG	Augmentation de Données avec transformations Globales
ADL	Augmentation de Données avec transformations Locales
ADLG	Augmentation de Données avec transformations Locales et Globales
AP	Antéro-Postérieure
ASIFT	Affine Scale-Invariant Feature Transform
SAM	Statistical appearance models (Modèle statistique d'apparence)
CCAM	Criss-cross Attention Module
CNN	Convolutional Neural Network (Réseau de neurones convolutionnel)
COCO	Common Objects in Context
CRF	Conditional Random Fields
CT-Scan	Computerized Tomography Scan
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DRR	Digitally Reconstructed Radiographs
ÉTS	École de Technologie Supérieure
FCN	Fully Convolutional Network (Réseau entièrement convolutionnel)
HOG	Histogram of Oriented Gradients
IRM	Imagerie par Résonance Magnétique
LAT	Latéral
LIO	Laboratoire de recherche en Imagerie et Orthopédie
mAP	Mean Average Precision
MICCAI	Medical Image Computing and Computer-Assisted Intervention

MI	Membre Inférieur
MLP	MultiLayer Perceptron (Perceptron multicouches)
PA	Postéro-Antérieure
PACS	Picture Archiving and Communication System (système d'archivage et de transmission d'images)
PASCAL VOC	PASCAL Visual Object Classes
RANSAC	RANdom SAmple Consensus
ReLU	Rectified Linear Unit
ROI	Region of Interest (Région d'intérêt)
R-CNN	Region-based Convolutional Neural Networks
RF	Random Forest (Forêt aléatoire)
RMSD	Root-Mean-Squared Distance
SCP	Stereo-Corresponding Points
SIFT	Scale-Invariant Feature Transform
SSC	Sparse Shape Composition (compositions de formes clairsemées)
SS	Scatter Search (recherche par dispersion)
SAM	Statistical Appearance Model (Modèle statistique d'apparence)
SSM	Statistical Shape Model (Modèle statistique de formes)
SPIE	Society of Photo-Optical Instrumentation Engineers
SURF	Speeded Up Robust Features
SVM	Support Vector Machine (Machine à vecteurs de support)
TMI	IEEE Transactions on Medical Imaging
YOLO	You Only Look Once

INTRODUCTION

L'imagerie médicale est à ce jour un outil diagnostique incontournable grâce à son potentiel à mettre en lumière l'état des structures internes du corps humain. Grâce aux progrès rapides observés dans le domaine du traitement d'images, il est désormais possible d'assister le chirurgien orthopédiste avant, durant et après ses interventions chirurgicales, grâce à l'accès aux mesures de paramètres cliniques (exemples : longueur de l'axe du col fémoral, angle cervico-diaphysaire, etc.). Ces paramètres cliniques peuvent être mesurés à partir d'informations identifiées dans l'image 2D (Kwolek, Brychey, Kwolek & Marczyński, 2019) ou par le biais de reconstructions tridimensionnelles de structures osseuses (Chaibi, Cresson, Aubert, Hausselle, Neyret, Hauger, De Guise & Skalli, 2012).

Dans le cadre de ce projet de recherche, nous travaillons en partenariat avec la compagnie EOS Imaging qui a développé en collaboration avec notre groupe de recherche, le laboratoire de recherche en imagerie et orthopédie (LIO) et le laboratoire de biomécanique (LBM) de l'École nationale supérieure d'arts et métiers de Paris, le système d'imagerie EOS qui permet de faire des acquisitions simultanées à faible dose de radiation de radiographies biplans EOS calibrées.

En pratique clinique, le logiciel commercial de la compagnie EOS (SterEOS) permet entre autres de calculer des paramètres cliniques du patient, et de reconstruire en 3D des structures osseuses des membres inférieurs grâce à l'information extraite des radiographies EOS.

Au niveau du membre inférieur (MI), les radiographies EOS sont habituellement prises sur deux plans perpendiculaires face/profil en orientation 0°/90° (Figure 0.1(a)) et le patient est placé en position debout avec les pieds joints ou décalés. Dans certaines situations, il est aussi possible d'effectuer des acquisitions obliques 45°/45° (patient tourné de 45° par rapport à la position standard face/profil (Figure 0.1(b))) pour éviter la superposition des structures osseuses (superposition entre les deux fémurs et les deux tibias) observée en vue de profil (en orientation 90°) ; ce qui permet une meilleure visibilité des structures osseuses. De plus, pour l'étude d'une

région spécifique comme le genou (Sarazin, 2018) ou la hanche (Ouertani, 2015), la région à analyser peut différer et on peut ainsi retrouver des images contenant des structures partielles (Figure 0.1(c)). Cela explique l'existence d'images présentant différentes configurations relatives au point de vue (comme par exemple : images de corps entier, images des membres inférieurs, images contenant des structures partielles) et à l'orientation (comme par exemple : images en orientation $0^\circ/90^\circ$ et $45^\circ/45^\circ$).

Mettre en place une approche de segmentation de structures osseuses et d'identification de points d'intérêts capable de couvrir cette variété d'images EOS n'est pas une tâche triviale. Il n'existe d'ailleurs pas à ce jour d'approche automatique d'extraction de structures osseuses, de localisation et d'identification de points d'intérêts sur une telle variété d'images radiographiques. La plupart des approches d'extraction d'informations existantes effectuées sur les radiographies couvrent une seule ou parfois deux orientations d'images (exemple : image antéro-postérieure) (Kwolek *et al.*, 2019 ; Tsai, 2021) ou sont spécifiques à une région d'intérêt comme le fémur (Chen, Ee, Leow & Howe, 2005), le genou (Kordon, Fischer, Privalov, Swartman, Schnetzke, Franke, Lasowski, Maier & Kunze, 2019), le fémur proximal (Guillen, Cerquin, Obando & Castro-Gutierrez, 2018 ; Lindner, Thiagarajah, Wilkinson, Wallis, Cootes, arcOGEN Consortium *et al.*, 2013) et le bassin (Chen & Zheng, 2014 ; Ding, Leow & Howe, 2007), de sorte qu'un algorithme donné est applicable spécifiquement à une structure donnée ou à une orientation donnée d'image. Pour ce qui est de l'extraction d'informations en vue d'initialiser les processus de mesures cliniques et de reconstruction 3D, la méthode d'extraction d'informations implantée dans SterEOS et en lien avec le MI est basée sur une approche manuelle. Cela rend ces différents processus du MI chronophage (10 min en moyenne pour la reconstruction des deux membres inférieurs) et utilisateur dépendant (Chaibi, 2010).

Plusieurs facteurs rendent difficile l'extraction d'informations dans les radiographies, notamment : le bruit (Stolojescu-Crișan & Holban, 2013), le changement de contraste (Stolojescu-

CriŞan & Holban, 2013), le recouvrement des structures osseuses par des tissus mous, la superposition de structures osseuses (Chaibi *et al.*, 2012) et la présence de structures similaires ou adjacentes (Ouertani, Vazquez, Cresson & de Guise, 2015). Dans les radiographies EOS, ces tâches d'extraction sont complexes à mettre en œuvre, dans le cas d'une approche automatique pour le membre inférieur, en raison des facteurs précédemment cités et de l'absence de connaissance par rapport à l'orientation du patient.

Le but de ce travail de thèse est de développer une méthode capable d'identifier et de segmenter des structures osseuses des MI et de localiser et d'identifier des points d'intérêts des MI sur les différentes configurations d'images EOS en vue d'initialiser automatiquement les différents processus EOS dédiés au MI et présents dans SterEOS.

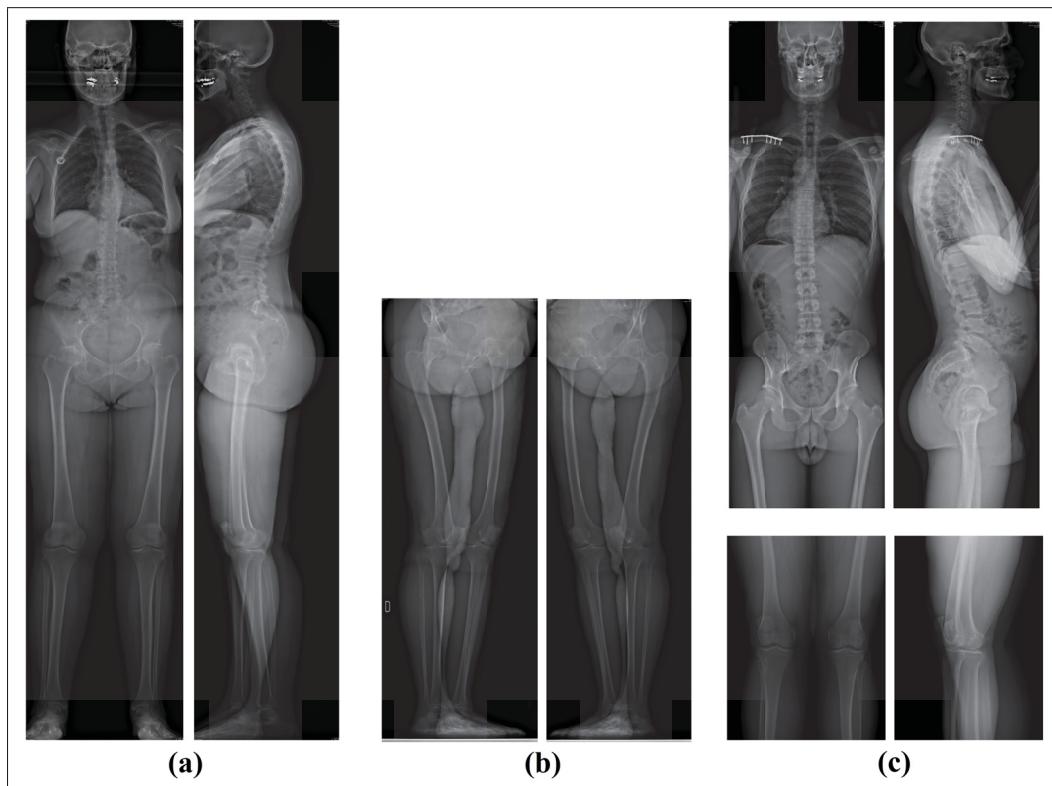


Figure 0.1 Radiographies EOS en orientation $0^\circ/90^\circ$ et $45^\circ/45^\circ$. (a) Images de corps entier en orientation $0^\circ/90^\circ$. (b) Images du MI en orientation $45^\circ/45^\circ$. (c) Images avec structures partielles en orientation $0^\circ/90^\circ$

0.1 Structure du document

Ce travail de thèse est organisé en six chapitres :

- **Chapitre 1 : Notions fondamentales.** Ce chapitre présente les différentes notions et termes techniques utilisés dans le document.
- **Chapitre 2 : Revue de littérature.** Ce chapitre présente une revue de littérature approfondie sur les méthodes de segmentation d'objets et d'identification de points d'intérêt dans les images avec emphase sur les méthodes d'apprentissage automatique. Cela nous permettra de ressortir les limitations des méthodes existantes afin de mieux justifier l'utilité de ce projet de recherche.
- **Chapitre 3 : Objectifs et choix méthodologiques.** Ce chapitre présente les objectifs et choix méthodologiques de cette thèse.
- **Chapitre 4 : Approche méthodologique globale.** Ce chapitre présentera l'approche proposée pour la segmentation des os et l'identification des points d'intérêt.
- **Chapitre 5 : Résultats.** Ce chapitre présentera les différents résultats obtenus.
- **Chapitre 6 : Discussion et contributions de la thèse.** Dans ce chapitre, une discussion sur les différentes méthodes proposées sera effectuée suivie d'une présentation des différentes contributions de cette thèse.
- Pour finir, une **conclusion** mettant en exergue l'originalité, les impacts et retombées de ce travail sera par la suite présentée. Nous finirons par adresser quelques **recommandations** et directions futures.

CHAPITRE 1

NOTIONS FONDAMENTALES ET CONTEXTE CLINIQUE

Dans ce chapitre, quelques notions importantes ainsi que le contexte clinique lié au membre inférieur seront présentés pour une bonne compréhension des travaux réalisés dans le cadre de cette thèse. Nous présenterons dans un premier temps les références anatomiques utilisées comme repère dans l'étude des structures anatomiques. Par la suite, nous présenterons l'anatomie du membre inférieur, suivie d'une présentation de différentes pathologies liées aux membres inférieurs. Nous terminerons ce chapitre par une présentation du système EOS et de la méthode d'initialisation manuelle de marqueurs spécifiques utiles pour les applications présentes dans le logiciel commercial (SterEOS) de EOS Imaging en lien avec l'extraction de mesures 2D/3D et la reconstruction 3D.

1.1 Références anatomiques

Les références anatomiques sont constituées d'axes, de plans et de directions (Figure 1.1). Nous pouvons distinguer trois axes : l'axe vertical Y , qui est perpendiculaire au sol et dirigé vers le haut ; l'axe transverse Z , qui est horizontal et dirigé vers la droite du sens du regard ; et l'axe sagittal X , qui est normal aux axes vertical et transverse et dirigé dans le sens du regard (Chaibi, 2010). L'axe vertical et l'axe transverse (\vec{Y}, \vec{Z}) forment le plan frontal ; l'axe sagittal et l'axe transverse (\vec{X}, \vec{Z}) forment le plan transversal ; et l'axe sagittal et l'axe vertical (\vec{X}, \vec{Y}) forment le plan sagittal. Les directions servent à définir l'emplacement d'une structure anatomique. Antérieur et postérieur signifient respectivement vers l'avant et vers l'arrière dans le sens du regard. Un élément est dit interne ou médial s'il est situé près du plan sagittal et externe ou latéral dans le cas inverse (Südhoff, 2007). Le terme proximal, pour une structure donnée, fait référence à la partie située en amont du tronc de la structure tandis que le terme distal désigne la partie éloignée du tronc (Südhoff, 2007).

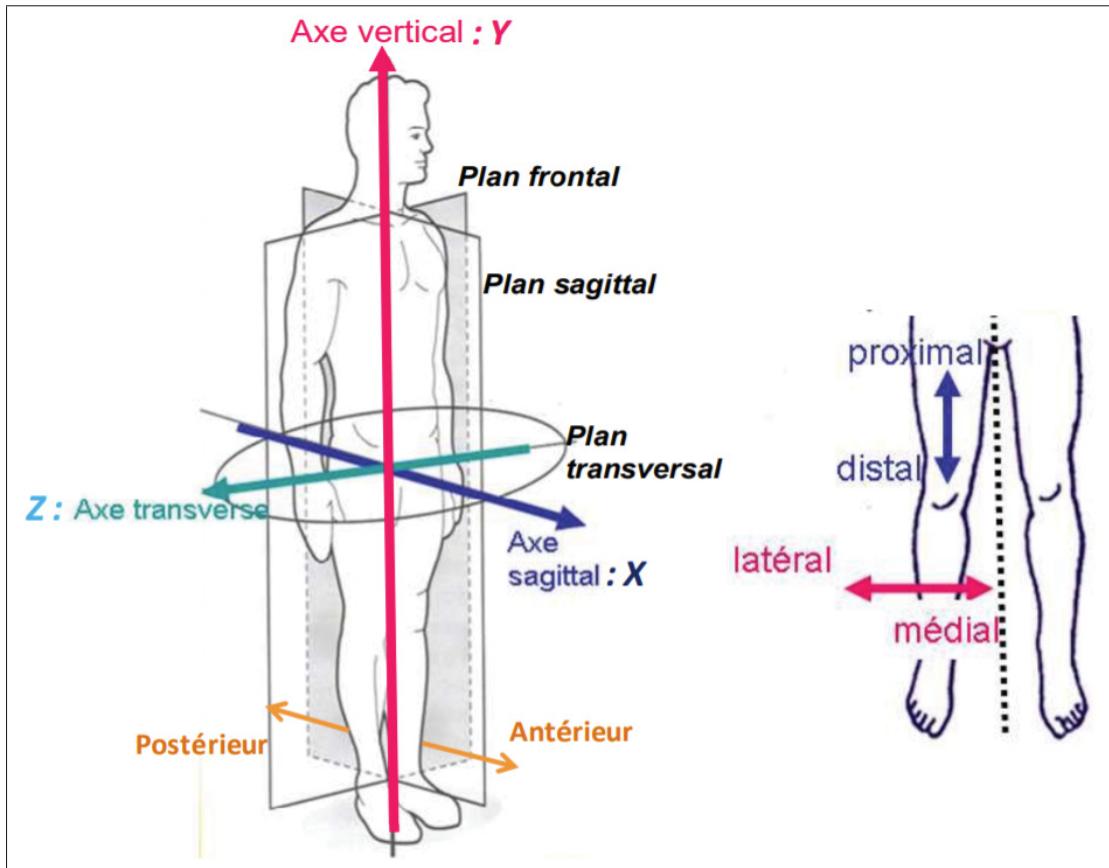


Figure 1.1 Repères anatomiques

Tirée de Chaibi (2010)

1.2 Anatomie du membre inférieur

Le membre inférieur (Figure 1.2) a pour principal rôle de soutenir le poids du corps chez l'humain et de lui permettre de se déplacer. Anatomiquement, le membre inférieur peut être subdivisé en six régions, à savoir : la ceinture pelvienne qui rattache le fémur à la hanche, la cuisse qui renferme le fémur, le genou qui rattache le fémur et le tibia via la rotule, la jambe qui renferme le tibia et le péroné, la cheville et la région du pied. Dans ce travail, nous focaliserons notre étude sur le fémur et le tibia.

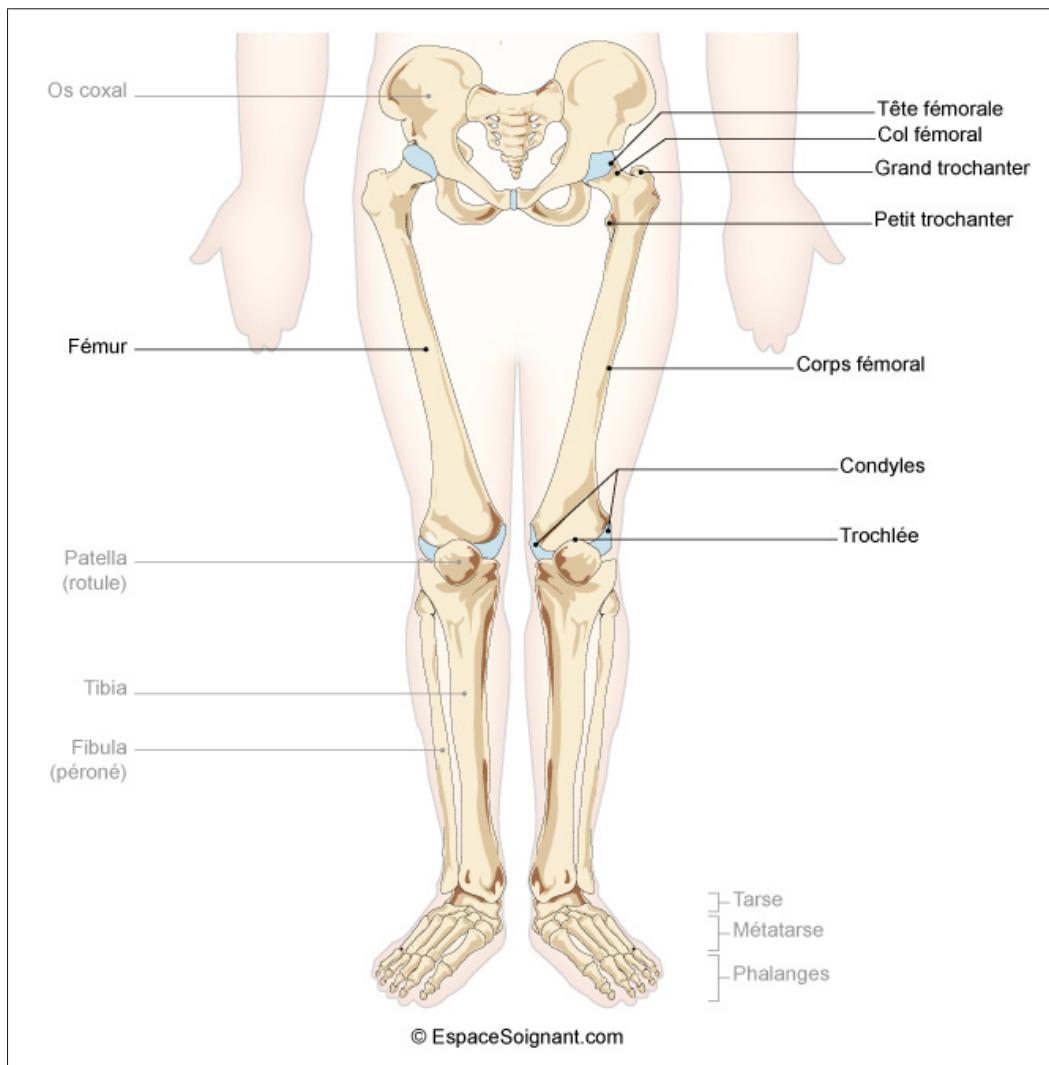


Figure 1.2 Squelette du membre inférieur
Tirée de Soins-infirmiers (2021)

1.2.1 Anatomie du fémur

Le fémur (Figure 1.3), os situé au niveau de la cuisse, est l'os le plus grand, le plus gros et le plus lourd du corps humain (Josse, 1832). Il peut être subdivisé en trois parties à savoir : l'épiphyse proximale, la diaphyse et l'épiphyse distale. L'épiphyse proximale comprend quatre parties à savoir : la tête fémorale, le col du fémur, le grand trochanter et le petit trochanter. La tête fémorale a une forme presque sphérique et occupe en proportion 2/3 d'une sphère d'environ 25 mm de rayon (Akkoul Berkache, 2013). De forme cylindrique, le col du fémur sépare la tête fémorale

de la diaphyse d'un angle d'environ 130° (Vasquez Barahona, 2004). Le grand trochanter et le petit trochanter sont des tubérosités volumineuses situées dans la partie proximale du fémur. La diaphyse est la partie la plus longue du fémur qui relie l'épiphyse proximale à l'épiphyse distale. L'épiphyse distale est constituée de condyle latéral et de condyle médial séparés par la fosse inter-condylienne. La droite liant le centre de la tête fémorale et le centre de la trochlée est l'axe mécanique fémoral tandis que la droite liant le centre de la trochlée au point situé sur le tiers de la longueur de la diaphyse représente l'axe anatomique du fémur (Chaibi, 2010).

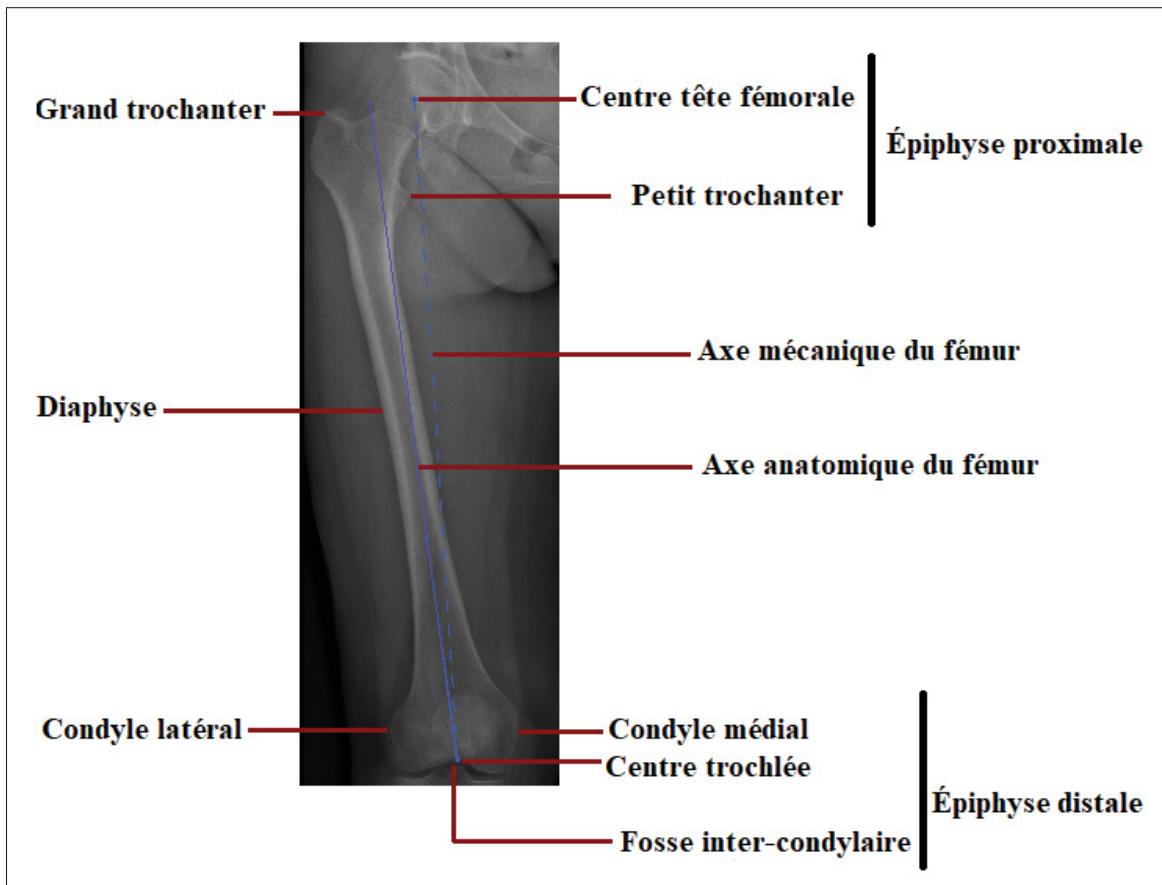


Figure 1.3 Anatomie du fémur

1.2.2 Anatomie du tibia

Le tibia (Figure 1.4), os le plus grand de la jambe, est constitué d'une épiphysé proximale, d'une diaphysé et d'une épiphysé distale. L'épiphysé proximale du tibia est constituée d'un condyle latéral et d'un condyle médial séparés par l'éminence inter-condylaire (Tortora & Derrickson, 2018). La diaphysé de forme cylindrique relie l'épiphysé proximale à l'épiphysé distale. L'épiphysé distale, quant à elle, est constituée du pilon tibial et de la malléole médiale qui est située à 90° de l'épiphysé distale. La droite liant le centre des épines tibiales au centre du pilon tibial représente l'axe mécanique ou anatomique du tibia (Chaibi, 2010).

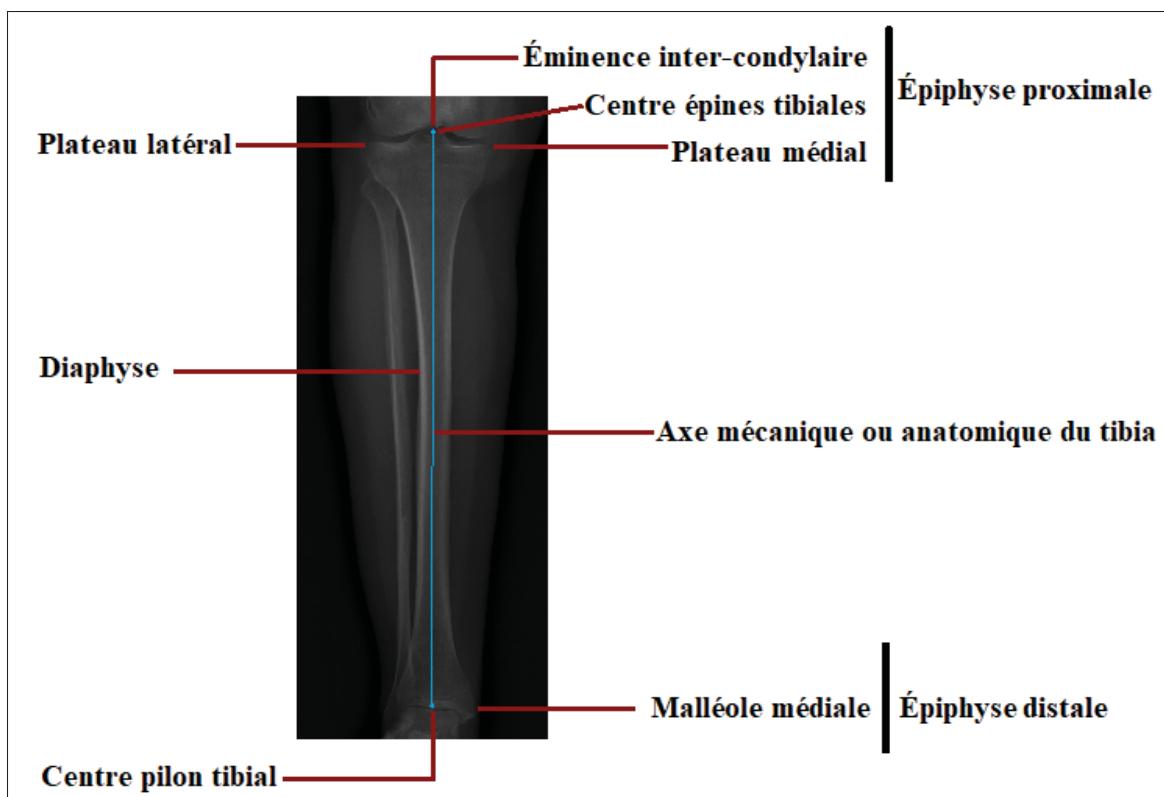


Figure 1.4 Anatomie du tibia

1.3 Pathologies liées aux membres inférieurs

Il existe plusieurs pathologies reliées aux membres inférieurs. Une des plus courantes est l'arthrose qui est une maladie qui touche environ 13.6% de la population canadienne âgée de 20 ans et plus (ASPC, 2020) et qui se caractérise par l'usure du cartilage présent au niveau d'une articulation. Cette maladie survient lorsque le corps n'arrive pas à réparer les tissus articulaires endommagés par l'obésité, des charges articulaires anormales et des blessures articulaires (ASPC, 2020). L'arthrose de la hanche (Figure 1.5) encore appelée coxarthrose se caractérise par une disparition progressive du cartilage existant entre le fémur et l'os du coxal (os du bassin) tandis que l'arthrose du genou (Figure 1.6) encore appelée gonarthrose se caractérise par l'usure du cartilage existant entre le fémur et le tibia.

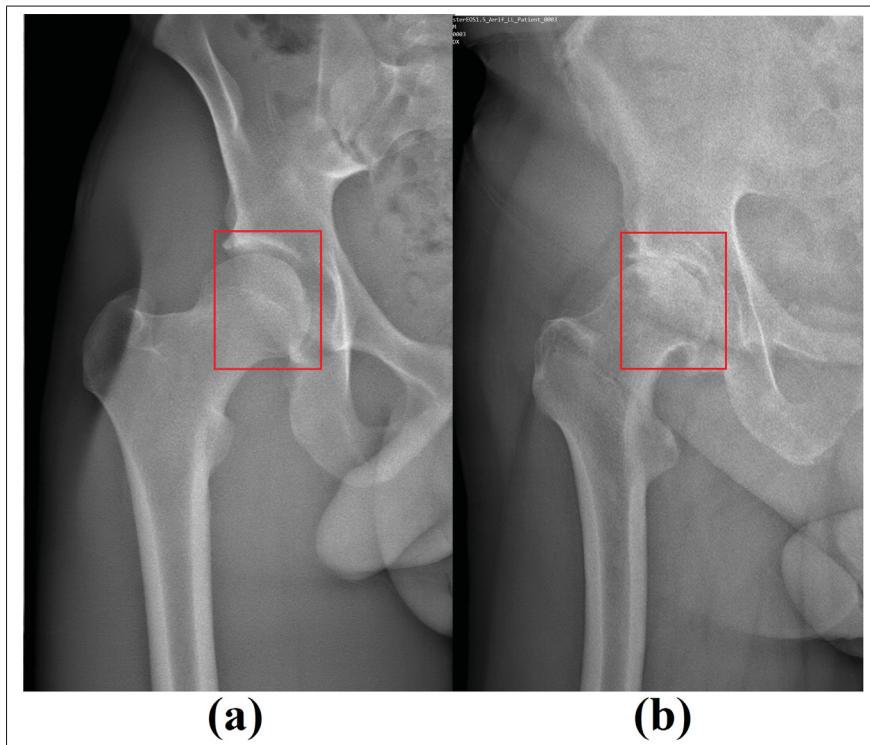


Figure 1.5 Arthrose de la hanche. (a) Hanche saine. (b) Hanche arthrosique

L'arthrose occasionne des douleurs durant la marche en raison de la réduction du cartilage causant le frottement entre les os adjacents. Deux autres pathologies des membres inférieurs considérées comme une arthrose débutante ou modérée du genou sont le genu valgum et le genu varum. Elles se caractérisent par une déviation entre les axes anatomiques du fémur et du tibia avec le fémur ou le tibia dévié vers l'intérieur ou vers l'extérieur (Figure 1.7).

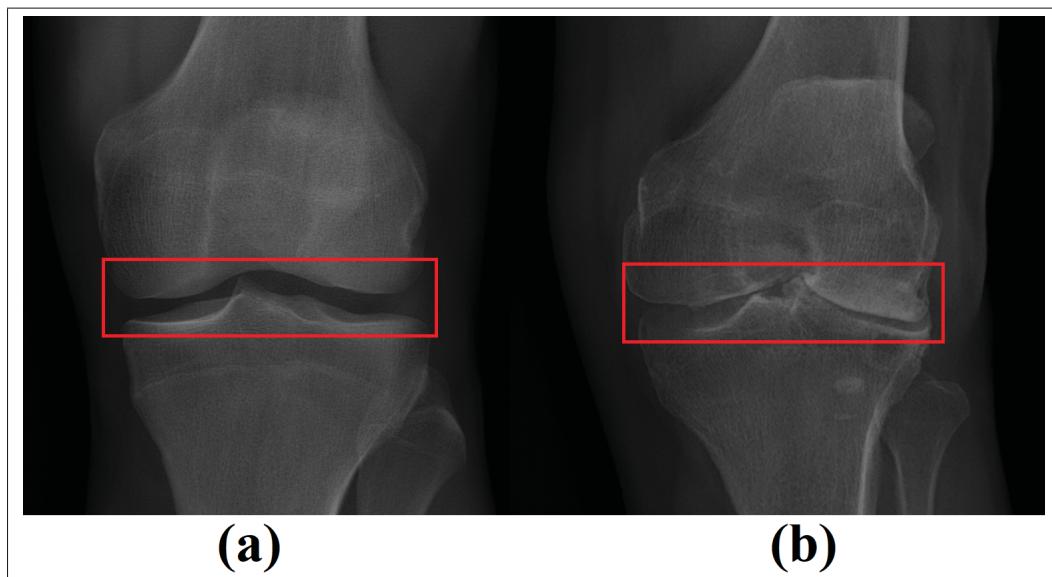


Figure 1.6 Arthrose du genou. (a) Genou sain. (b) Genou arthrosique

D'autres pathologies liées au membre inférieur comme l'arthrite inflammatoire qui se caractérise par une inflammation du cartilage au niveau des articulations, les fractures osseuses ou l'ostéopose qui se caractérise par une fragilisation des os, affectent négativement les structures liées aux membres inférieurs, occasionnant une détérioration de la mobilité.

L'orthopédiste a besoin d'avoir recours à un bilan radiographique pour confirmer son diagnostic, quantifier l'ampleur de la pathologie ou pour déterminer les indications d'ostéotomies (chirurgie réparatrice réalisée sur l'os) et des corrections à réaliser (Langlais & Thomazeau, 1999). Dans ces différents cas, les mesures cliniques ou la reconstruction 3D de l'os pathologique pourraient être requises ; ce qui nécessiterait de segmenter l'os pathologique ou de localiser des points d'intérêt liés à cet os.

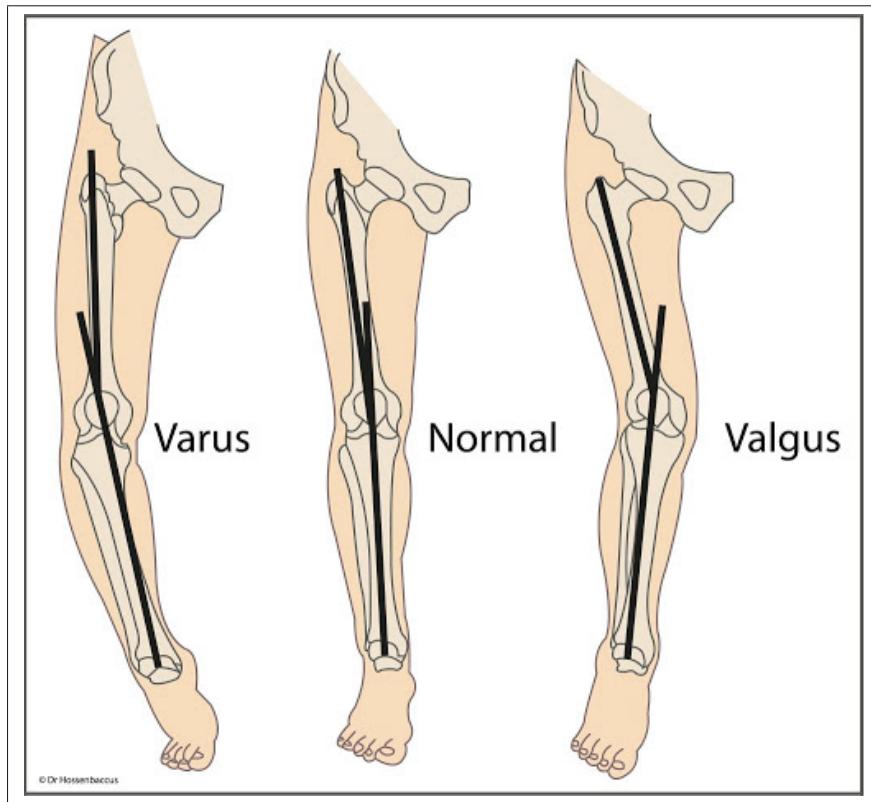


Figure 1.7 Genus varum, normal et valgum
Tirée de Hossenbaccus (2021)

1.4 Système EOS

Le système radiographique EOS (Figure 1.8) est un appareil de radiologie numérique très peu irradiant permettant l'acquisition simultanée de la tête aux pieds, de deux images orthogonales (face et profil) calibrées. Sa dose d'irradiation est 8 à 10 fois inférieure à celle de la radiographie conventionnelle et 800 à 1000 fois inférieure à celle du CT-Scan (Dubousset, Charpak, Dorion, Skalli, Lavaste, Deguise, Kalifa & Ferey, 2005). L'imageur EOS comporte deux sources de rayons X qui balaiennent linéairement de haut vers le bas le patient en position fonctionnelle debout (position en charge). Avoir le patient debout, en position en charge durant l'acquisition, représente un grand avantage en orthopédie pour l'étude des troubles de torsion (déformation de l'os) dans les structures osseuses, des troubles de la rotation au niveau des articulations ou de pathologies comme l'arthrose (Chaibi, 2010). Le protocole standard d'acquisition pour

le membre inférieur (MI) consiste à acquérir simultanément deux images face et profil (en orientation 0° et 90° respectivement) avec le patient ayant les pieds joints ou les pieds décalés. Cela occasionne une superposition de structures osseuses dans les images de profil en orientation 90° (voir Figure 1.9 (b), (f), (g)). Pour éviter cette superposition, les radiologistes de certains hôpitaux produisent des images à 45° en positionnant le patient en position $45^\circ/45^\circ$ dans la cabine EOS. Ainsi, les structures osseuses acquises sont bien séparées (voir Figure 1.9 (c), (d)) ; ce qui facilite la segmentation des os et l'identification de points d'intérêt du MI. Pour l'étude d'une région spécifique comme le fémur proximal ou le genou, des images contenant des structures partielles sont produites. Il existe donc une diversité d'images EOS (Figure 1.9) à savoir : les images des membres inférieurs, les images contenant des structures partielles ou les images du corps entier avec le patient (ayant les pieds joints ou les pieds décalés) en orientation ($0^\circ/90^\circ$, $45^\circ/45^\circ$).



Figure 1.8 Processus d'acquisition de radiographies EOS
Tirée de Illés & Somoskeöy (2012)

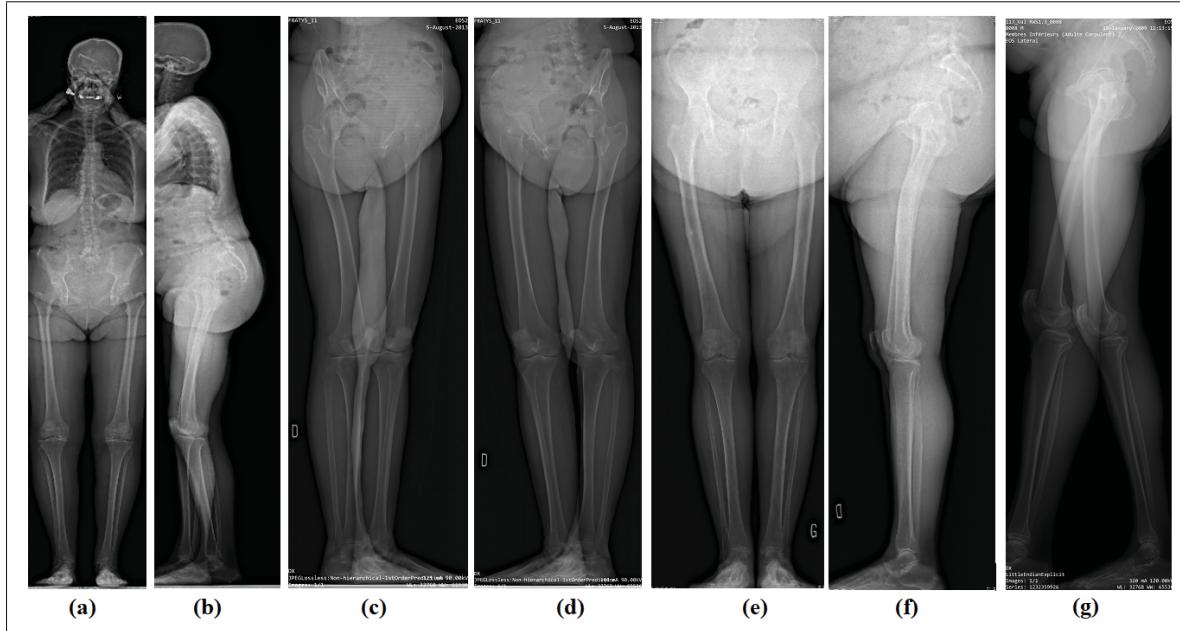


Figure 1.9 Exemples d’images radiographiques EOS. (a) Radiographie EOS corps entier en orientation 0° , (b) Radiographie EOS corps entier avec pieds légèrement décalés en orientation 90° , (c) Radiographie EOS des MI en orientation 45° , (d) Radiographie EOS des MI en orientation 45° , (e) Radiographie EOS des MI en orientation 0° , (f) Radiographie EOS des MI avec pieds joints en orientation 90° , (g) Radiographie EOS des MI avec pieds très décalés en orientation 90°

1.5 Processus disponibles dans le logiciel commercial SterEOS

La compagnie EOS Imaging produit une plateforme logicielle (SterEOS) constituée de processus permettant l’analyse d’images, la prise de mesures cliniques et la reconstruction 3D personnalisée des os du MI à partir de deux images biplans EOS. Ces différents processus nécessitent une phase d’initialisation où un opérateur est amené à numériser manuellement sur les structures osseuses (fémur et tibia) de chaque paire d’images EOS, des points, des contours osseux ou autres primitives (Figure 1.10). À partir des primitives positionnées par l’opérateur, des mesures cliniques peuvent être estimées par le processus de mesures cliniques. Des analyses 3D peuvent également être faites à partir de ces primitives, sans nécessiter une reconstruction 3D comme c’est le cas pour la pangonométrie du MI. Dans le cadre du processus de reconstruction 3D, les différentes primitives numérisées permettent de déformer des modèles génériques surfaciques

des structures osseuses des MI (Figure 1.13) pour avoir des modèles 3D finaux des os des MI (Figure 1.13) (Chaibi, 2010). Des mesures cliniques peuvent également être estimées à partir de cette reconstruction 3D.

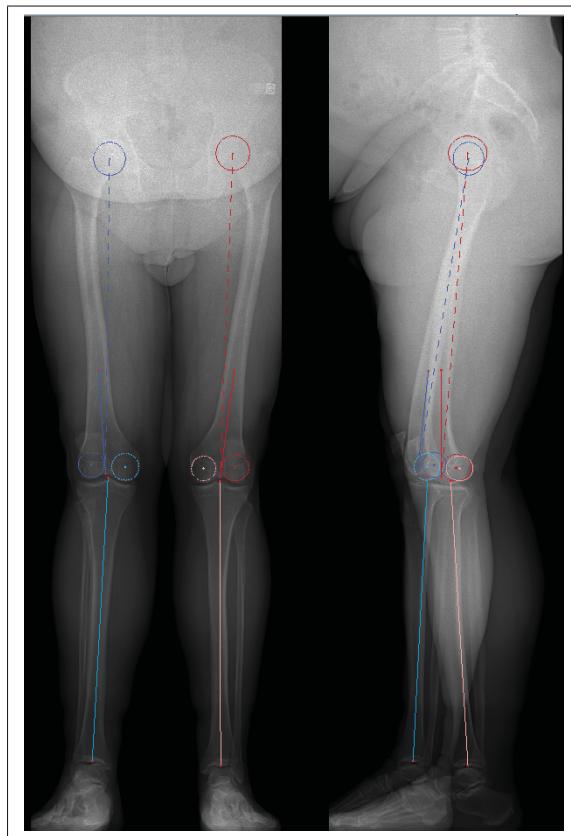


Figure 1.10 Initialisation manuelle des processus de mesures cliniques et de reconstruction 3D du MI
Tirée de SterEOS

Les processus disponibles dans SterEOS pour le MI sont fondés sur l'obligation de numériser manuellement des informations dans les images EOS ; ce qui est chronophage. La compagnie EOS Imaging souhaite rendre ces processus moins chronophages en proposant un approche permettant d'initialiser automatiquement tous ces processus ; ce qui reviendrait à segmenter les os et à identifier les points d'intérêts des MI dans chaque paire d'images EOS de façon automatique. L'opérateur n'aura donc plus que des ajustements à effectuer si nécessaire une fois qu'une première initialisation est proposée.

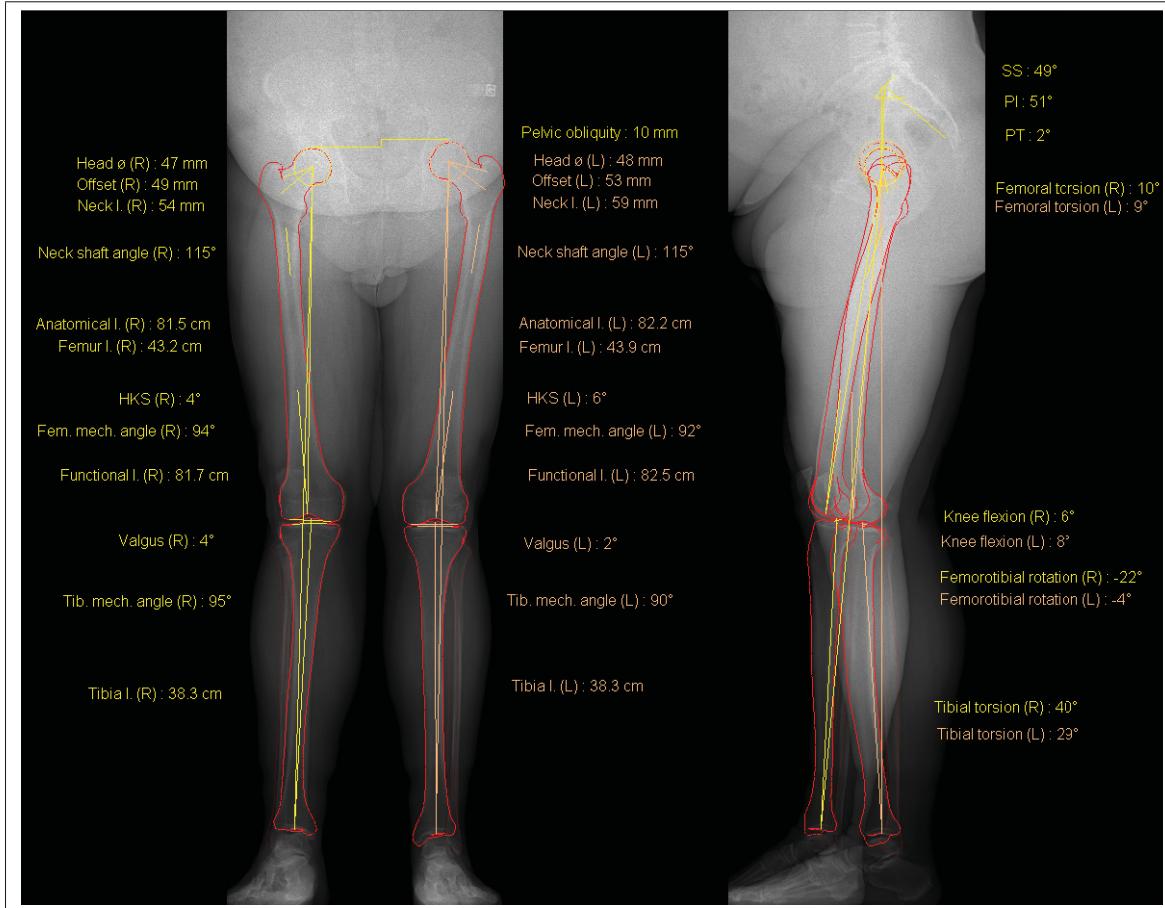


Figure 1.11 Estimation des paramètres cliniques des MI
Tirée de SterEOS

1.6 Conclusion

Les tâches de segmentation des os, de localisation et d'identification de points d'intérêt dans les images EOS sont essentielles en orthopédie pour les processus de mesures cliniques ou de reconstruction 3D des os en vue de planifier le geste médical. L'automatisation de ces tâches est importante pour rendre moins chronophages les processus disponibles actuellement dans la plateforme SterEOS.

En pratique clinique, ces tâches, réalisées sur les images EOS sont principalement basées sur la perception visuelle de l'expert et sur son habileté à identifier manuellement ces informations dans l'image. Cela constitue un potentiel risque d'erreur de diagnostic à cause de la capacité visuelle

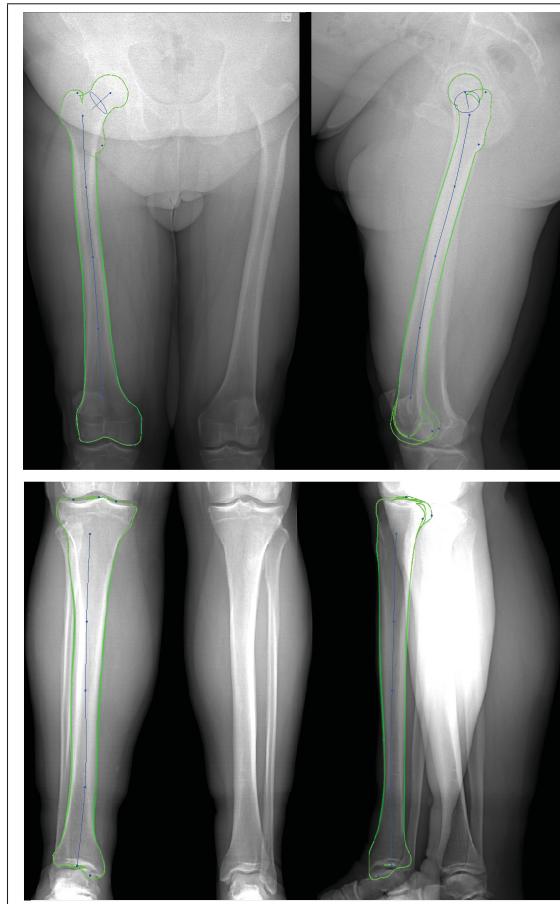


Figure 1.12 Déformation du modèle générique surfacique
Tirée de SterEOS

limitée de l'être humain (Ketenci, 2013), de la variabilité de performance dans l'exécution des tâches manuelles d'un individu à un autre (Chaibi, 2010) et du fort bruit présent dans les radiographies (Chaibi, 2010). D'autres facteurs comme le flux grandissant de données pour un nombre restreint de radiologues et la capacité limitée de l'humain à traiter l'information, occasionnent chez le radiologue la fatigue visuelle, augmentent les délais de traitement et rendent son travail moins efficace. Ces facteurs font accroître les délais d'attente (Gargouri, 2009), de même qu'une baisse de la qualité du diagnostic. Une étude réalisée par Pinto & Brunese (2010) a montré que 41% à 80% de fractures sont mal diagnostiquées en service d'urgence et que les blessures orthopédiques constituent à elles seules 75% des diagnostics manqués. Une

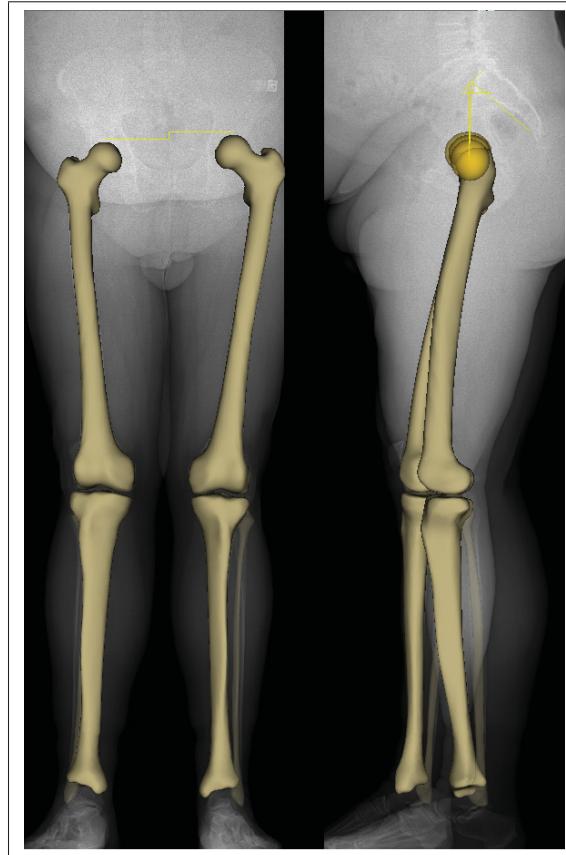


Figure 1.13 Reconstruction 3D personnalisée des os des MI
Tirée de SterEOS

solution potentielle pour aider le radiologue dans ses tâches quotidiennes serait d'entièrement automatiser ces tâches de localisation et d'identification manuelles.

Le système de radiographie EOS permet de faire des acquisitions du MI avec des points de vue différents (images de corps entier, image du membre inférieur, images avec structures partielles) et des orientations différentes du patient dans la cabine ($0^\circ/90^\circ$, $45^\circ/45^\circ$ ou $20^\circ/70^\circ$ par exemple). Ce système est présent sur plusieurs sites hospitaliers et les acquisitions EOS peuvent varier d'un site à l'autre. Proposer une approche capable de fonctionner sur des images EOS variables en fonction de la pose, de l'orientation, du point de vue ou selon la pathologie serait très important.

La mise en place d'un processus entièrement automatisé de segmentation de structures osseuses et d'identification de points d'intérêt du MI, applicable sur la variété d'images EOS n'est pas triviale, car on peut avoir des images EOS très différentes en termes de points de vue et de l'orientation du patient (Figure 1.9). Par exemple, les images en orientation 90° pieds joints (Figure 1.9(f)) présentent plus de superposition de structures osseuses comparées aux images pieds décalés (Figure 1.9(g)) en orientation 90°. De plus, la mise en place d'un processus automatique requiert l'accès à un large ensemble de données représentatives surtout si l'on vise à utiliser des méthodes d'apprentissage automatique. L'utilisation d'un petit ensemble de données dans ce cas pourrait donc être problématique, car cet ensemble de données pourrait ne pas être assez représentatif, et donc, pourrait limiter la performance décisionnelle de l'approche développée. En imagerie médicale, il est souvent difficile d'avoir accès à un large ensemble de données. Particulièrement, dans cette thèse, nous avons accès à un ensemble limité d'images EOS.

Les algorithmes d'extraction d'informations (segmentation de structures et identifications de points) existants (Chen & Zheng, 2014 ; Guillen *et al.*, 2018 ; Kordon *et al.*, 2019 ; Kwolek *et al.*, 2019 ; Ouertani, 2015 ; Sarazin, 2018 ; Tsai, 2021) traitent spécifiquement une région donnée du MI (exemple fémur) ou une orientation donnée d'images (image en orientation 0°). Jusqu'à présent, le problème de segmentation automatique de structures osseuses, de localisation et d'identification de points d'intérêt des membres inférieurs sur la diversité d'images EOS n'est pas résolu.

Un des besoins de la compagnie EOS Imaging est de rendre moins chronophages leurs applications dédiées à l'extraction de mesures cliniques et de reconstructions 3D pour le MI qui nécessitent jusqu'à présent l'intervention d'un opérateur pour leur initialisation.

Le but de ce travail est de développer une approche automatique capable de segmenter des structures osseuses des MI et de localiser et d'identifier des points d'intérêts des MI dans des images EOS présentant différents points de vue et orientations afin d'initialiser automatiquement

les processus dédiés à l'extraction de mesures cliniques et à la reconstructions 3D implantés dans SterEOS.

Une revue de littérature sera présentée au prochain chapitre sur différentes méthodes utilisées dans la littérature pour segmenter des structures et identifier des points d'intérêt dans les images. Étant donné que dans ce travail, nous avons accès à un ensemble limité d'images EOS, nous aborderons également les méthodes d'augmentation de données utilisées dans la littérature pour résoudre le problème relatif au manque de données dans les méthodes d'apprentissage automatique.

CHAPITRE 2

REVUE DE LITTÉRATURE

Dans ce chapitre, une revue sur les méthodes de segmentation de structures osseuses et d'identification de points d'intérêt existantes dans la littérature sera effectuée. Nous commencerons notre revue par présenter quelques approches conventionnelles d'extraction d'informations (segmentation d'objets et identification de points) dans l'image. Ensuite, nous présenterons les méthodes d'extraction basées sur les méthodes d'apprentissage machine. Une revue sera par la suite faite spécifiquement sur différents travaux abordant la tâche de segmentation de structures et d'identification de points d'intérêts dans les images médicales. Pour finir, nous aborderons différentes méthodes d'augmentation de données utilisées dans la littérature.

2.1 Méthodes conventionnelles d'extraction d'informations

Dans cette section, nous présentons les principales méthodes conventionnelles de segmentation d'objets et d'identification de points d'intérêt dans différents types d'images (exemples : images optiques, images satellitaires).

2.1.1 Méthodes conventionnelles de segmentation d'objets

2.1.1.1 Méthodes classiques

L'une des approches précurseures pour faire de la segmentation dans la littérature est la méthode de segmentation par seuillage. Cette méthode consiste à segmenter une image en régions grâce à une ou deux valeurs de seuil d'intensité fixées (Stolosescu-Crișan & Holban, 2013). Cette méthode souffre d'une sensibilité au bruit dans l'image puisqu'elle ne tient pas compte de l'information spatiale présente dans l'image, mais reste uniquement focalisée sur les valeurs d'intensité des pixels (Stolosescu-Crișan & Holban, 2013). De plus, les seuils sont souvent choisis manuellement de façon empirique ou en se basant sur l'histogramme de l'image (Lebore, 2007). Pour éviter l'ajustement manuel du seuil, certaines méthodes comme k-means (Chen,

Luo & Parker, 1998) sont utilisées pour regrouper les pixels d'une image en des groupes selon la similarité de leurs intensités. Cette méthode reste cependant sensible au bruit et à la présence de structures d'apparence similaire (Chartrand, 2017). Une autre approche classique de segmentation est la méthode de croissance de région (Stolosescu-CriŞan & Holban, 2013). À partir d'un pixel de départ choisi, cette méthode fait croître une région en y intégrant des pixels voisins sur la base d'un critère de similarité (même valeur de pixels ou valeurs de pixels situées dans une plage fixée). Les principaux problèmes de cette méthode résident dans le choix du pixel de départ et du critère de similarité (Martin, 2008) qui ont un impact direct sur la qualité de la segmentation (un mauvais choix entraîne une mauvaise segmentation). Cette approche fonctionne bien pour la segmentation de structures homogènes, mais reste inadaptée à la segmentation de structures présentant de fortes inhomogénéités (Martin, 2008). D'autres méthodes telles Canny, Sobel, Prewitt (Stolosescu-CriŞan & Holban, 2013), permettent de retrouver des frontières entre différentes régions dans les images. Ces méthodes, principalement basées sur la valeur du gradient des intensités de pixels dans l'image, sont également sensibles au bruit dans l'image et ont tendance à trouver des frontières qui n'ont aucun rapport avec les frontières réelles (Stolosescu-CriŞan & Holban, 2013). L'utilisation de ces différentes méthodes classiques reste donc très limitée sur les radiographies EOS en raison du niveau élevé de bruit et de changement d'intensité qu'elles présentent.

2.1.1.2 Modèles déformables

Les modèles déformables (exemples : contour actif (Kass, Witkin & Terzopoulos, 1988), level set (Osher & Sethian, 1988)) sont des techniques de segmentation, très utilisées en traitement d'images pour retrouver des formes dans les images. Ils sont un mixage de trois disciplines mathématiques, à savoir : la géométrie qui sert à représenter les formes d'objets, la physique pour imposer des contraintes sur la variation des formes dans le temps et dans l'espace et la théorie de l'approximation qui fournit les fondements formels des mécanismes d'ajustement des modèles aux données mesurées (McInerney & Terzopoulos, 1996). Les modèles déformables ont une phase d'initialisation consistant à sélectionner dans l'image des pixels se situant proches de

l'objet recherché. La région définie par ces pixels (durant la phase d'initialisation) croît et s'arrête grâce à une contrainte de variation de forme définie. Une application des modèles déformables peut être vue dans le travail de Amrapur et al. (2019) qui segmentent des visages humains à l'aide de la méthode des contours actifs. Les modèles déformables présentent l'avantage de converger rapidement, mais souffrent par contre d'un problème d'initialisation qui doit généralement être faite proche de l'objet recherché (Medjeded & Mahmoudi, 2009 ; Tsechpenakis, 2011).

2.1.1.3 Méthodes par atlas

En imagerie, un atlas est une image étiquetée conçue à partir d'un ou de plusieurs images (Martin, 2008). La méthode basée sur l'atlas est utilisée pour les tâches de segmentation d'objets (Sharma, Aggarwal, Roy, Gupta, Vashist & Sidhu, 2018). La méthodologie suivie consiste à trouver une transformation mettant en correspondance l'image à segmenter avec l'atlas (Martin, 2008). La précision de la segmentation dépend donc du fait que la mise en correspondance soit bien réalisée. Cette méthode présente l'avantage d'éviter les faux positifs (objet segmenté à tort) ou les faux négatifs (objet non segmenté à tort) (Martin, 2008). De plus, le recalage d'une structure présente dans l'atlas est facilité par la position des autres structures contenues dans l'atlas. Par contre, la méthode par atlas n'est pas souvent très précise, car une régularisation de forme n'est pas faite (Martin, 2008). De plus, elle est sensible au bruit. En effet, l'utilisation d'un atlas non représentatif des images de test ou mal étiqueté pourrait générer des erreurs lors de la phase de mise en correspondance et donc dans le résultat final de segmentation (Aljabar, Heckemann, Hammers, Hajnal & Rueckert, 2009). Dans le cas de l'approche multi-atlas, l'effet de cette erreur peut être réduit lorsqu'on fusionne les estimations de segmentation de chacun des atlas constituant le groupe d'atlas (Aljabar *et al.*, 2009). Cependant, le coût de calcul de la segmentation augmente linéairement avec la taille de la base de données de l'atlas (Aljabar *et al.*, 2009).

2.1.1.4 Méthode de coupe de graphe

La méthode de coupe de graphe (graph cut) est une approche d'optimisation globale utilisée pour segmenter itérativement des images en attribuant une étiquette à chaque pixel de l'image (Ayed, Chen, Punithakumar, Ross & Li, 2010 ; Feng, Wang & Liu, 2021). L'approche considère l'image comme un graphe où les nœuds correspondent aux pixels de l'image, et les arêtes aux connexions pondérées entre des paires de pixels adjacents. Les poids attribués aux arêtes reflètent la similitude existant entre les pixels (Feng *et al.*, 2021). Deux sommets additionnels désignant le premier plan (objet à segmenter) et l'arrière-plan sont ajoutés au graphe. L'étiquetage de l'image est alors réalisé en trouvant la coupe minimale (partition avec poids minimal) qui sépare les pixels de premier plan des pixels d'arrière-plan (Preim & Botha, 2014). Des exemples d'application de la méthode de coupe de graphe peuvent être vus dans les travaux de Feng *et al.* (2021) et de He, Peng & Chen (2019) qui segmentent des objets dans des images de scènes naturelles. Le principal problème avec cette méthode, c'est son manque de précision à segmenter les contours de l'objet dans l'image (Candemir & Akgul, 2011) et sa grande consommation en mémoire (Lermé, Malgouyres, Létocart & Rocchisani, 2011).

2.1.2 Méthodes conventionnelles d'extraction de caractéristiques

Les méthodes d'extraction de caractéristiques sont très utilisées pour réaliser les tâches d'identification de points d'intérêt dans l'image et de mise en correspondance d'images (Karsh, Das, Swetha, Medhi, Laskar, Arya & Agarwal, 2016 ; Yu & Morel, 2011). Différentes méthodes (exemples : SIFT (Lowe, 2004), la transformée de Hough (Hough, 1962)) permettent de retrouver des caractéristiques dans l'image. Les détecteurs de points caractéristiques (exemples : détecteur de Harris (Harris & Stephens, 1988), ASIFT (Yu & Morel, 2011)) permettent de mettre en évidence des zones d'intérêt (zones renfermant de l'information pertinente) caractérisant l'image (Lowe, 2004). La plupart du temps, des descripteurs associés à chaque point ou zone d'intérêt sont dans un premier temps générés, ensuite une mise en correspondance des descripteurs est faite à l'aide d'une mesure de similarité (Yu & Morel, 2011).

2.1.2.1 DéTECTEUR DE HARRIS

Le détecteur de Harris (Harris & Stephens, 1988), développé par Harris et Stephen est l'un des détecteurs de points d'intérêts qui a révolutionné le traitement d'images. Ce sont des extracteurs de coins stables (points d'intérêts où les bords présentent un changement brusque de direction) qui présentent une invariance aux changements affines d'intensité et à la rotation. Akbar, Maulana & Yuniarti (2018) ont utilisé une technique de mise en correspondance d'images basée sur les détecteurs de Harris pour détecter l'emplacement des plaques d'immatriculation de véhicules indonésiens dans l'image. Les détecteurs de Harris ne sont pas invariants à l'échelle, ce qui reste problématique dans notre cas, car dans les radiographies en général, les structures anatomiques d'un individu à l'autre sont de tailles différentes et situées à des distances variables l'une de l'autre ; ce qui représente une variation d'échelle. L'algorithme SIFT vient combler cette lacune de Harris.

2.1.2.2 L'algorithme SIFT

SIFT (scale-invariant feature transform) (Lowe, 2004) est un algorithme qui, à partir d'une image, extrait des points clés (caractérisés par 128 descripteurs) invariants à l'échelle, la rotation, l'orientation et au point de vue. Il comble par ailleurs la lacune d'invariance à l'échelle du détecteur de Harris et est robuste dans le processus de recherche d'objets ou de similarités dans des images prises sous différents points de vue. Cependant, SIFT ne couvre pas tous les paramètres de transformation affine, ce qui fait qu'il n'arrive pas toujours à détecter certains points correspondants entre deux images et réalise parfois de fausses mises en correspondance de points sur différentes images (Yu & Morel, 2011). Il est donc souvent associé à d'autres algorithmes tels que la transformée de Hough (Hough, 1962), RANSAC (Fischler & Bolles, 1981), afin de supprimer ces fausses mises en correspondance. Comme exemple d'application de l'algorithme SIFT, Lenc & Král (2015) ont proposé un système de reconnaissance faciale basé sur l'algorithme SIFT. Cet algorithme a un faible coût de calcul, utilise moins de puissance de traitement, est invariant au flou (Morel & Yu, 2009) et utilise moins de mémoire de stockage comparativement aux réseaux de neurones. Il n'est par contre pas robuste aux transformations

non linéaires et est partiellement invariant aux changements d'intensité et de point de vue (Morel & Yu, 2009).

2.1.2.3 L'algorithme ASIFT

Développé par Yu & Morel (2011), ASIFT (Affine-SIFT) est une variante de l'algorithme SIFT. Cette méthode est plus robuste que SIFT dans le processus de mise en correspondance de points caractéristiques sur diverses images puisqu'il couvre entièrement tous les six paramètres (rotation, translation (composantes x et y), mise en échelle (composantes x et y), cisaillement) de la transformation affine contrairement à SIFT qui n'en couvre que quatre (Morel & Yu, 2009). La robustesse de ASIFT par rapport à SIFT a été démontrée par Karsh *et al.* (2016) dans leur travail de détection de falsification dans les images. Ils ont obtenu une différence drastique entre les points clés détectés par SIFT et ASIFT prouvant ainsi la supériorité de ASIFT sur SIFT quand il s'agit de transformations affines. L'algorithme ASIFT est entièrement invariant aux transformations affines et au flou (Morel & Yu, 2009). Cet algorithme a été étudié sur les images EOS durant notre phase exploratoire et a donné des erreurs d'identifications assez conséquentes (voir ANNEXE I). Cela pourrait être dû à sa sensibilité aux valeurs d'intensité dans l'image.

2.1.2.4 La transformée de Hough

Introduit en 1959 par Paul Hough, la transformée de Hough est une technique permettant de retrouver des formes géométriques dans l'image. Initialement, elle avait été conçue pour retrouver des lignes dans l'image. Des formes géométriques plus complexes (comme de cercles, des ellipses ...) peuvent également être identifiées en utilisant la transformée généralisée de Hough, introduite par Richard Duda et Peter Hart en 1972. Le principe qui sous-tend la transformée de Hough est que chaque image peut être représentée dans l'espace paramétrique. Les informations présentes dans l'image peuvent donc être représentées sous une forme paramétrique. Les applications de la transformée de Hough dans la littérature sont variées. Elle est utilisée pour la détection de linéaments dans des images de scènes naturelles (Zhao, Han, Zhang, Xu & Cheng, 2021) ou des images satellitaires (Poncelet & Cornet, 2010), l'extraction du chemin de navigation pour les

robots cueilleurs de concombres de serre (Chen, Qiang, Wu, Xu & Wang, 2021), etc. Bien que la transformée de Hough permette de retrouver des primitives dans l'image, elle présente certaines limitations. En effet, elle détecte un taux élevé de faux positifs (Mammeri, Boukerche & Lu, 2014) et a un temps de calcul élevé (Chiu, Liaw & Lin, 2010).

2.1.3 Conclusion

Nous avons abordé les principales méthodes conventionnelles de segmentation d'objets et d'identification de points d'intérêt dans les images. Ces méthodes présentent plusieurs limitations. En effet, elles sont sensibles au bruit et sont principalement basées sur les valeurs d'intensité des pixels dans l'image. L'utilisation de ces méthodes sur les radiographies EOS reste très limitée en raison du niveau élevé de bruit et de changement d'intensité présent dans les images EOS. D'un autre côté, nous avons les approches d'apprentissage machine qui sont robustes au bruit et qui tiennent compte de l'information spatiale dans l'image pour effectuer une tâche. La section qui suit sera consacrée à la présentation des méthodes d'apprentissage machine.

2.2 Méthodes d'apprentissage machine pour l'extraction d'informations

Les méthodes d'apprentissage machine (exemple : les réseaux de neurones profonds, les forêts aléatoires), sont très utilisées en imagerie grâce à leur précision de prédiction et leur capacité à apprendre des représentations abstraites. Elles sont utilisées pour les tâches de segmentation d'objets (Chen *et al.*, 2017; Sirinukunwattana, Pluim, Chen, Qi, Heng, Guo, Wang, Matuszewski, Bruni, Sanchez *et al.*, 2017) et de localisation et d'identification de points d'intérêt (Ding, Kang, Yuan, Shan & Cai, 2021) dans les images. Elles peuvent être utilisées en apprentissage supervisé (ou la classe des données d'apprentissage est connue et préalablement étiquetée par un expert) comme en apprentissage non supervisé (ou la classe des données d'apprentissage n'est pas connue) (Bonacorso, 2017). Dans cette section, nous présenterons quelques méthodes d'apprentissage machine utilisées pour les tâches de segmentation d'objets et d'identification de points d'intérêt dans les images.

2.2.1 Modèles de forme statistiques

Les modèles statistiques sont très utilisés pour réaliser la tâche de segmentation (Abhyankar & Schuckers, 2006 ; Mirikharaji & Hamarneh, 2018) et d'identification de points d'intérêt (Iqtait, Mohamad & Mamat, 2018 ; Zhuo & Gou, 2020) grâce à leur habileté à incorporer des connaissances de formes spécifiques à la structure à segmenter (Heimann & Meinzer, 2009). Ils sont utilisés pour caractériser la variation de forme d'un objet présent dans des images présentant des caractéristiques semblables. Diverses méthodes statistiques comme les modèles de formes actives (Active Shape Model) (Cootes, Taylor, Cooper & Graham, 1995), les modèles de contraintes locales (Constrained local Model) (Cristinacce & Cootes, 2008), les compositions de formes clairsemées (Sparse shape composition) (Zhang, Zhan, Dewan, Huang, Metaxas & Zhou, 2012) existent dans la littérature. L'approche adoptée en général consiste à apprendre les diverses contraintes ou variations de chaque région cible des images d'apprentissage pour ensuite faire une généralisation. Par exemple, Yahiaoui (2017) et Butakoff & Frangi (2010) ont, dans un premier temps, fait une initialisation, qui a par la suite servi de repère pour le modèle de forme permettant de pouvoir segmenter la forme d'intérêt dans l'image. Le point négatif avec les modèles de forme statistiques, c'est qu'ils ne sont pas capables d'estimer des variations de formes autres que celles présentes dans leur base d'apprentissage (Piazzese, Carminati, Pepi & Caiani, 2017). Avec un petit ensemble d'apprentissage, les modèles de forme statistiques ne pourront pas estimer toutes les variations de formes possibles (Piazzese *et al.*, 2017).

2.2.2 Machines à vecteurs de support

Les machines à vecteurs de support (SVM), introduites par Vapnik (2013) sont une méthode d'apprentissage machine supervisée destinée à résoudre des problèmes de discrimination, de classification et de régression. Elles sont une généralisation des classificateurs linéaires et sont très adaptées pour les données de grandes dimensions. Une des notions les plus importantes avec les SVM est la notion de marge maximale. La marge est la distance entre la frontière de séparation et les échantillons les plus proches. Dans les SVM, la frontière de séparation est choisie comme celle qui maximise la marge. Le problème à résoudre dans cet algorithme est donc de trouver la

frontière séparatrice optimale à partir d'un ensemble d'apprentissage. Pour traiter les cas où les données ne sont pas linéairement séparables, les SVM transforment l'espace de représentation des données d'entrées en espace de plus grande dimension dans lequel il est probable qu'il existe une séparation linéaire (Vapnik, 2013). Les SVM ont une bonne capacité à détecter les vrais négatifs et les faux positifs, ce qui explique leur très grande utilité en classification. Niu & Suen (2012) ont remplacé la dernière couche (couche de classification) de leur réseau de neurones convolutionnel (CNN) par un SVM, ce qui leur a permis d'avoir un fort taux de classification (99.81%) d'écritures manuscrites de l'ensemble de données MNIST (LeCun, Cortes & Burges, 1994). Le seul bémol est le fait que cette méthode hybride consomme plus de temps et d'espace mémoire qu'un simple CNN ; ce qui semble raisonnable du fait de l'introduction d'un classifieur SVM plus sophistiqué en remplacement de la couche de sortie simple du CNN. Les SVM sont efficaces dans les espaces à grandes dimensions et dans le cas où le nombre de dimensions est supérieur au nombre d'échantillons (Alpaydin, 2020). Un de leurs points faibles est que si le nombre de points d'intérêts est supérieur au nombre d'échantillons, la méthode peut donner de mauvaise performance (Alpaydin, 2020). De plus, ils ne fournissent pas directement d'estimation de probabilité.

2.2.3 Forêts aléatoires

Une forêt aléatoire, comme son nom l'indique, est constituée de plusieurs arbres décisionnels avec comme plus, la notion d'aléatoire (Breiman, 2001). Un arbre décisionnel se construit itérativement en découplant la population en deux sous-ensembles différents. Dans une forêt aléatoire (Random forest (RF)), le résultat final est l'agrégation des résultats issus de chacun des arbres le constituant. Dans le cas d'une prédiction, cela revient à considérer la moyenne des résultats tandis que dans le cas d'une classification cela revient à retenir la majorité de votes. Une application de forêts aléatoires peut être vue dans le travail de Bélouard, Schmuck, Nagelisen & Guyon (2018) qui identifient des défoliations forestières sur des images satellites. Les forêts aléatoires sont computationnellement efficaces grâce à leur propriété de calcul conditionnelle (Ioannou, Robertson, Zikic, Kortschieder, Shotton, Brown & Criminisi, 2016).

Aussi, les arbres (constituant la forêt) sont construits de façon indépendante. Ils ne sont par contre pas précis comparativement aux CNN (Ioannou *et al.*, 2016).

2.2.4 Réseaux de neurones

Les réseaux de neurones, inspirés du fonctionnement des neurones biologiques, sont une méthode d'apprentissage machine. Leur principal but est d'apprendre des informations abstraites à partir d'exemples afin de résoudre des problèmes complexes. Ils ont une architecture commune (Figure 2.1), constituée d'une couche d'entrée, d'une couche de sortie et d'une ou plusieurs couches cachées constituées de neurones (Parizeau, 2004).

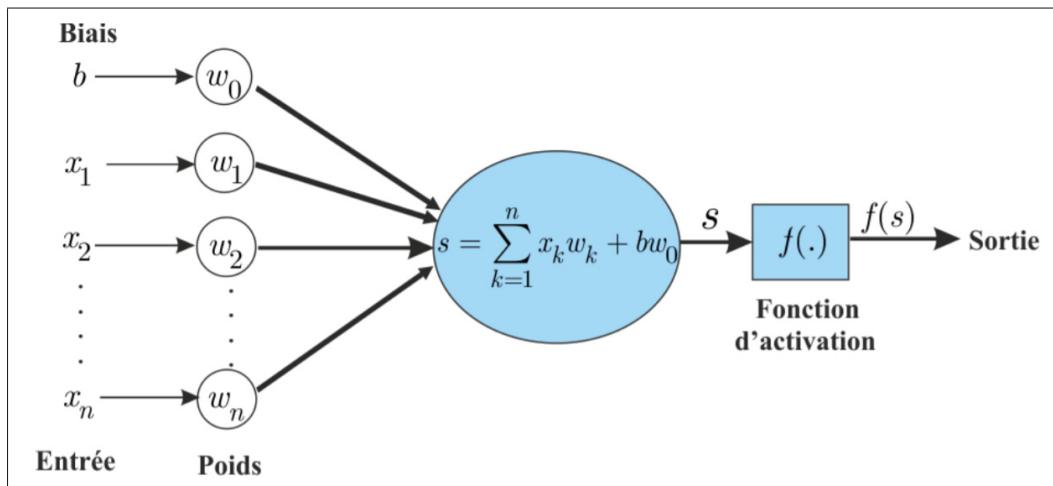


Figure 2.1 Architecture d'un réseau de neurones
Tirée de Truong (2016)

Il existe différents types de réseaux de neurones. Nous pouvons entre autres citer (Alpaydin, 2020 ; Parizeau, 2004) : les perceptrons multicouches, les réseaux de neurones convolutionnels. Chaque neurone d'un réseau de neurones est muni d'un poids synaptique lui permettant de stocker l'information apprise lors de l'apprentissage. Il comprend aussi une fonction d'activation pouvant entre autres être la fonction ReLU ($f(x) = \max(0, x)$), la fonction sigmoïde ($f(x) = 1/(1 + \exp(-x))$), la fonction tangente hyperbolique ($f(x) = \tanh(x)$). Les réseaux de neurones sont très utilisés pour les tâches de classification (Molchanov, Yang, Gupta, Kim, Tyree & Kautz,

2016), d'identification (Polat & Yıldırım, 2008), de segmentation (Badrinarayanan *et al.*, 2017) et de détection d'objets (Cai, Fan, Feris & Vasconcelos, 2016) dans l'image. Ils offrent une représentation compacte, permettent de décomposer des données en niveaux d'abstraction, permettent l'apprentissage de fonctions compliquées pouvant représenter des abstractions de haut niveau, sont robustes au bruit (Portaz, 2018) et ont la capacité d'analyser des relations spatiales et topologiques (Virole, 2001). Ils sont par contre lents durant la phase d'entraînement (Lemley, Bazrafkan & Corcoran, 2017), demandent une grande puissance de calcul (Courbariaux, 2015), de mémoire de stockage pour les poids et ont besoin d'être entraînés sur une grande quantité de données représentatives (Delassus, 2018).

2.2.4.1 Réseaux de neurones convolutionnels

Les réseaux de neurones convolutionnels (CNN), très utilisés ces dernières années (Figure 2.2), ont montré de très bons résultats en imagerie pour des tâches de localisation, d'identification et de segmentation (Litjens *et al.*, 2017). Les applications des CNN sont nombreuses comme l'identification d'objets dans les images (Krizhevsky, Sutskever & Hinton, 2012), la segmentation d'objets dans les images (Yuan, 2018) ou la mise en correspondance d'images (Balntas, Johns, Tang & Mikolajczyk, 2016). Un exemple d'architecture CNN classique est donné à la Figure 2.3.

Les CNN (Figure 2.3) sont des réseaux de neurones multicouches constitués d'une couche d'entrée (une image dans notre cas), d'une partie permettant l'extraction de caractéristiques (encadré en rouge dans Figure 2.3) et d'une partie (avec la couche de sortie à la fin) permettant de faire de la prédiction (encadré en bleu dans Figure 2.3). Le début du réseau (la partie encadrée en rouge dans Figure 2.3), c'est-à-dire la partie permettant d'extraire des caractéristiques dans l'image, est constitué d'une succession de couches de convolution parfois intercalées par des couches de pooling. La fin du réseau (la partie encadrée en bleu dans Figure 2.3) est constituée de couches entièrement connectées permettant de faire de la prédiction (exemple : classification). Le résultat (image) en sortie d'une couche de convolution est appelé carte de caractéristiques. La carte de caractéristiques est obtenue en convoluant l'image par un noyau (encore appelé filtre) qui est représenté par une matrice de dimension inférieure à celle de l'image d'entrée

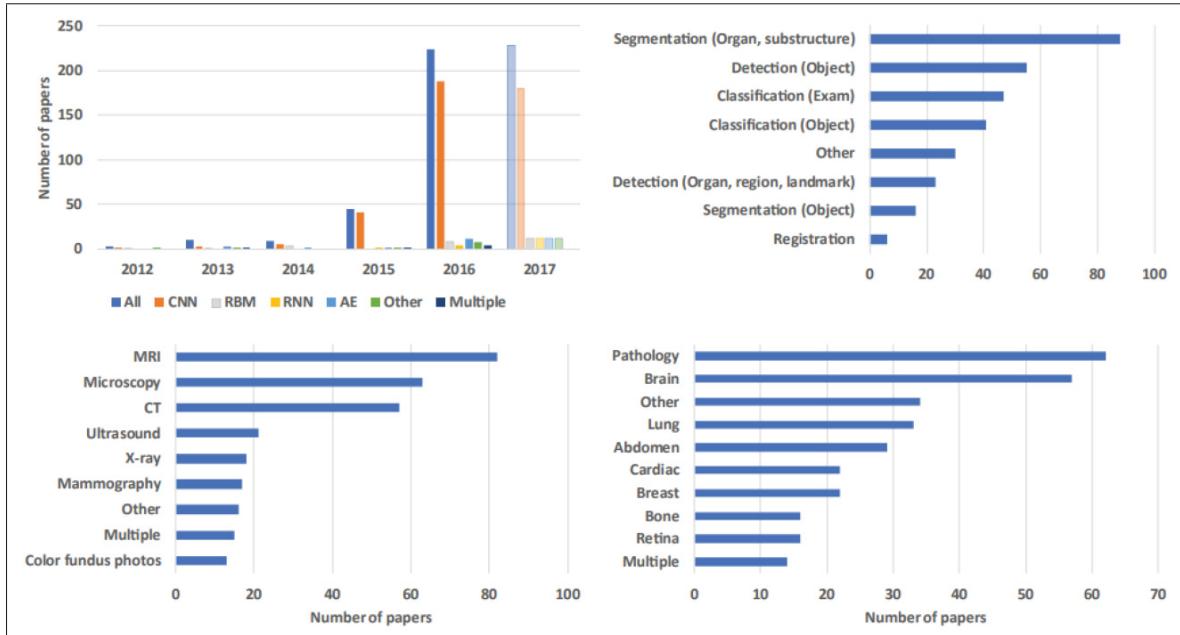


Figure 2.2 Répartition de papiers publiés de 2012 à 2017
Tirée de Litjens *et al.* (2017)

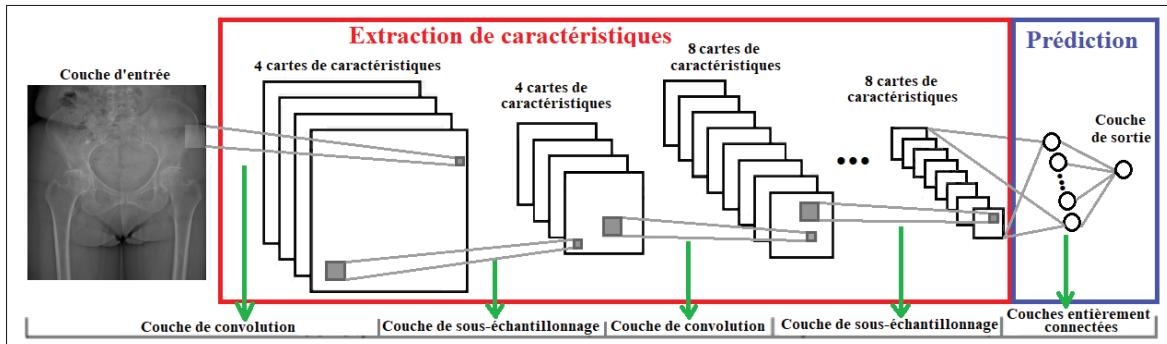


Figure 2.3 Architecture d'un CNN

et dont les éléments sont appelés poids (Delalleau, 2008). Ces poids sont déterminés durant l’entraînement du réseau par rétropropagation du gradient (Delalleau, 2008). Dans les CNN, les poids sont partagés de sorte que le réseau effectue des opérations de convolution sur les images (Litjens *et al.*, 2017). Le modèle n'a donc pas besoin d'apprendre des détecteurs séparés pour une même structure se trouvant à différentes positions dans l'image (Litjens *et al.*, 2017). Cela réduit considérablement la quantité de paramètres appris par le réseau. La convolution dilatée,

une convolution qui insère des trous dans le filtre de convolution, est parfois intégrée dans les CNN dans le but d'augmenter leur habileté à prendre en compte le contexte global présent dans l'image. Des couches de "dropout" (opération permettant de réduire le surapprentissage) sont parfois intégrées dans les CNN afin d'éviter du surapprentissage du réseau (Portaz, 2018).

Les CNN sont adaptés au traitement de données en forme de tableaux multidimensionnels et donnent de grandes précisions dans l'apprentissage de représentation (Ioannou *et al.*, 2016). Ils utilisent peu de phases de prétraitement, rendant le réseau responsable de faire évoluer tout seul ses propres filtres (connexions locales et partage de poids) (Buyssens & Elmoataz, 2016). En plus de l'information liée aux intensités des pixels dans l'image, les CNN exploitent également l'information de contexte présente dans l'image pour prendre des décisions (Cheng, Zhang, Yang, Liu & Zhang, 2017).

2.2.4.1.1 L'information contextuelle dans les réseaux convolutionnels

Les CNN sont très utilisés pour les tâches de prédiction, mais ne tirent pas toujours pleinement profit de l'information de contexte (encore appelée information contextuelle) présente dans l'image (Cheng *et al.*, 2017). L'information contextuelle peut être améliorée dans le réseau neuronal en fusionnant (exemple d'opération de fusion : sommation élément par élément) des cartes de caractéristiques à plusieurs niveaux comme dans (Chen *et al.*, 2017; Kayalibay, Jensen & van der Smagt, 2017; Ronneberger *et al.*, 2015) pour avoir des cartes de probabilité précises (Chen *et al.*, 2017). Chen *et al.* (2017), par exemple, combinent les cartes de caractéristiques intermédiaires de leur réseau avec leurs cartes de caractéristiques de sortie. Ronneberger *et al.* (2015) concatènent progressivement les cartes de caractéristiques de la partie encodeur de leur réseau à leur correspondant dans la partie décodeur. Zhao, Shi, Qi, Wang & Jia (2017); Zhao, Qi, Shen, Shi & Jia (2018a) considèrent différents redimensionnements de l'image comme entrées de leur réseau et combinent les cartes de caractéristiques en sortie.

Une façon d'augmenter l'information contextuelle dans le réseau neuronal est d'augmenter le champ réceptif. Soit C_p une carte de caractéristiques issue d'une couche profonde, le champ

réceptif d'un élément (ou pixel) p de C_p est l'ensemble des éléments de l'image d'entrée (située dans la couche de base la moins profonde) qui l'influence. Un réseau avec un grand champ réceptif intègre une information contextuelle plus large lui permettant de capturer l'information contextuelle globale, ce qui est important pour augmenter la précision de prédiction d'objets relativement grands (Wang, Chen, Yuan, Liu, Huang, Hou & Cottrell, 2018a). La taille du champ réceptif d'un élément peut être augmentée en empilant de nombreuses couches pour rendre le réseau plus profond, en utilisant des opérateurs de sous-échantillonnage comme le pooling ou en utilisant des convolutions dilatées (Luo, Li, Urtasun & Zemel, 2016).

2.2.4.1.2 La couche de convolution

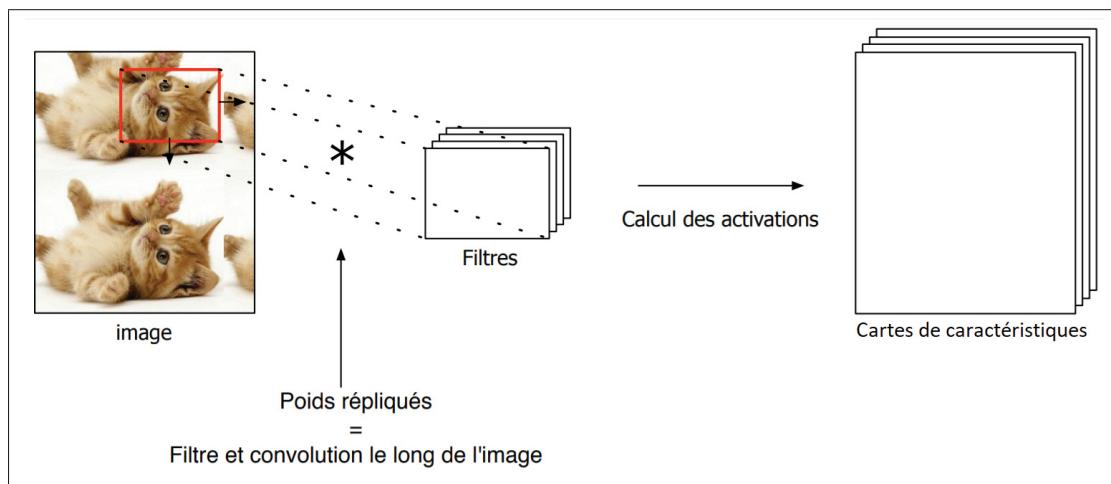


Figure 2.4 Couche de convolution
Tirée de Cohen-Hadria & Peeters (2016)

La couche de convolution (Figure 2.4) permet d'extraire les caractéristiques présentes dans une image. Soient $I : \mathbb{Z}^2 \rightarrow \mathbb{R}^3$ une fonction discrète, $k : J^2 \subset \mathbb{Z}^2 \rightarrow \mathbb{R}^3$ un filtre discret (encore appelé noyau) et $p, q, n \in \mathbb{Z}^2$. La convolution discrète est définie comme $(I * k)(n) = \sum_{n=p+q} (I(p)k(q))$ où $*$ est l'opérateur de convolution. L'opération de convolution est réalisée en faisant glisser un filtre sur une image d'entrée (Figure 2.4) avec un pas (stride) défini et en réalisant le produit de convolution de ce filtre avec l'image sous-jacente. La profondeur (nombre de canaux) du filtre est la même que celle de l'image d'entrée. La matrice de sortie de la couche

de convolution est de taille inférieure à la matrice d'entrée étant donné que le filtre ne peut être appliqué aux coins de la matrice d'entrée (l'opération de convolution n'est réalisée que si chaque valeur du filtre est appliquée à une valeur de la matrice d'entrée). Cela crée une perte d'informations (sur les bords) pouvant être contrée par la méthode de rembourrage (padding) qui consiste généralement à ajouter la valeur 0 sur les bords de l'image d'entrée. La Figure 2.5 présente le résultat en sortie d'une couche de convolution avec un 0-padding de 1 et un biais de 1 sur une image couleur RVB (rouge, vert, bleu) de profondeur 3 (on a donc 3 canaux : le premier canal pour R, le deuxième canal pour V et le troisième canal pour B). Le résultat à la sortie de l'opération de convolution est ensuite passé à une fonction d'activation f pour former la carte de caractéristiques de sortie. Dans le neurone de convolution, le niveau d'activation des neurones est calculé comme suit (Bouvierie, 2006) :

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \quad (2.1)$$

avec M_j une sélection de cartes de caractéristiques d'entrée, k le noyau de convolution et b le biais.

2.2.4.1.3 La convolution dilatée

La convolution dilatée est une convolution avec un noyau modifié par un facteur de dilatation. Elle augmente le champ réceptif dans le réseau (en insérant des « trous » dans les noyaux de convolutions) tout en préservant l'information contextuelle globale dans l'image et sans réduire la dimensionnalité (Mittal, Hooda & Sofat, 2018 ; Wang *et al.*, 2018a ; Yu & Koltun, 2016). Soit $I : \mathbb{Z}^2 \rightarrow \mathbb{R}^3$ une fonction discrète et soit $k : J^2 \subset \mathbb{Z}^2 \rightarrow \mathbb{R}^3$ un filtre discret. Soit $p, q, n \in \mathbb{Z}^2$ et soit $*_d$ une convolution dilatée d'un facteur d , la convolution dilatée discrète est définie comme $(I *_d k)(n) = \sum_{n=p+dq} (I(p)k(q))$. Quand le facteur de dilatation est 1, la convolution dilatée est équivalente à la convolution classique. Plus le facteur de dilatation est élevé, plus l'information globale est préservée et moins l'information locale est préservée (Wang *et al.*, 2018a). La Figure 2.6 présente des convolutions dilatées de facteurs différents avec des filtres de taille 3×3 . Les pixels en rouge (croix rouges) dans l'image représentent les éléments du noyau de convolution

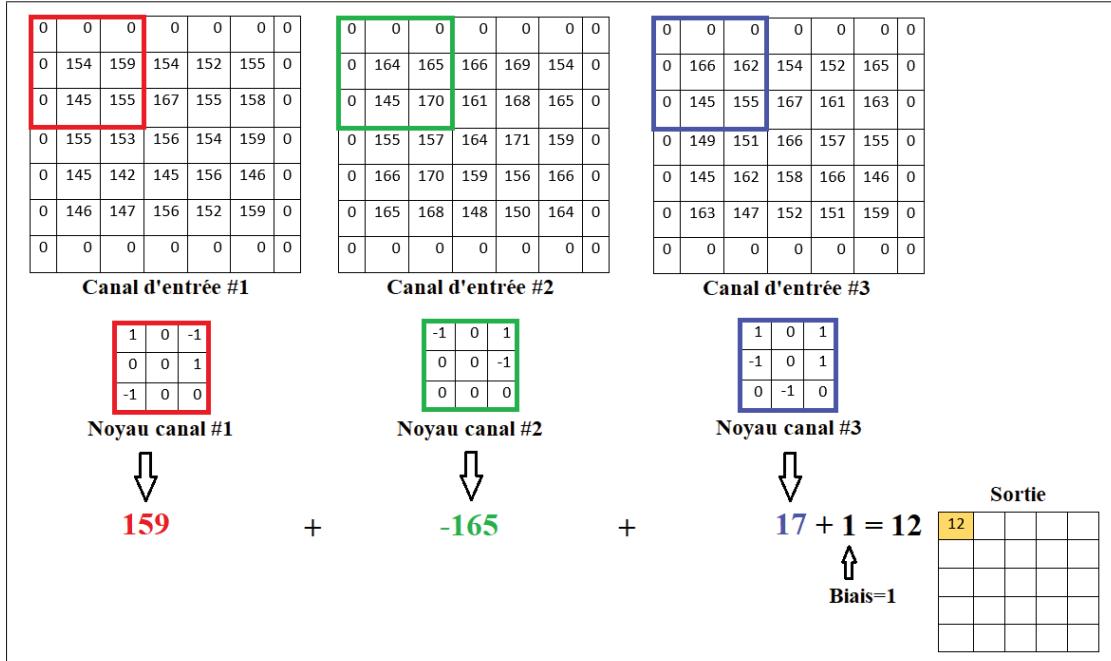


Figure 2.5 Opération de convolution sur une image de profondeur 3

qui participent à l'opération de convolution. La zone en jaune représente la taille du champ réceptif des pixels des cartes de caractéristiques en sortie des couches C_i avec $i \in \{0, 1, 2\}$.

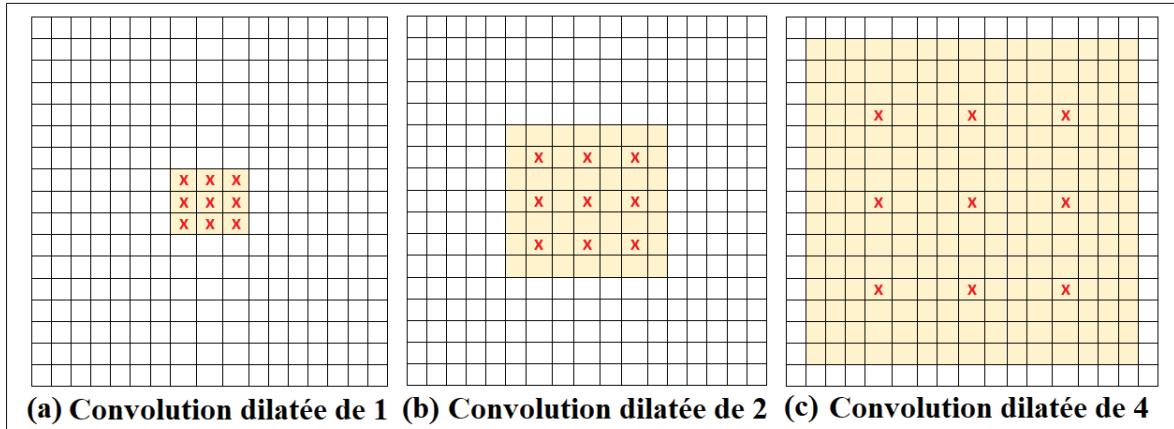


Figure 2.6 Opération de convolution dilatée avec un noyau de taille 3×3 . (a) C_1 obtenue en convoluant C_0 par une convolution dilatée de facteur 1. (b) C_2 obtenue en convoluant C_1 par une convolution dilatée de facteur 2. (c) C_3 obtenue en convoluant C_2 par une convolution dilatée de facteur 4

2.2.4.1.4 Le pooling

Le pooling est une méthode de sous-échantillonnage permettant de compresser l'information. Largement utilisé dans les réseaux convolutionnels, le pooling permet une invariance (de l'image en entrée) à la rotation et à la translation (Buyssens & Elmoataz, 2016). La couche de pooling réduit la dimensionnalité (Haque & Neubert, 2020 ; Saeedan, Weber, Goesele & Roth, 2018), le coût de calcul (Haque & Neubert, 2020 ; Saeedan *et al.*, 2018) et évite le surapprentissage surtout quand on apprend de grands ensembles de données contenant plusieurs objets (Haque & Neubert, 2020).

Deux opérations de pooling souvent utilisées sont :

- le pooling moyen (average Pooling) (Figure 2.7(a)) qui calcule la valeur moyenne pour chaque signal d'entrée ;
- le maximum Pooling or Max pooling (Figure 2.7(b)) qui calcule la valeur maximale pour chaque signal d'entrée.

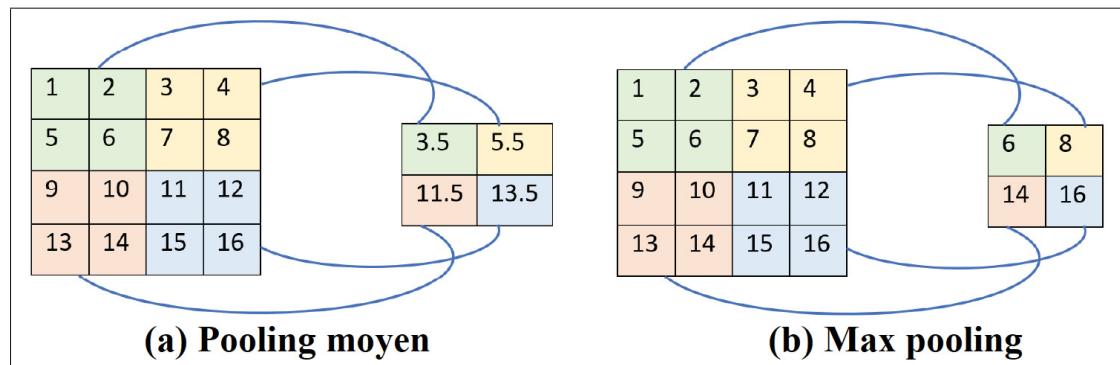


Figure 2.7 Exemples d'opérations de pooling. (a) Pooling moyen avec un filtre 2×2 et un pas de 2 (b) Max pooling avec un filtre 2×2 et un pas de 2

Le max pooling préserve les détails présents dans l'image contrairement au pooling moyen qui en perd (Saeedan *et al.*, 2018). Une section du réseau VGG (un réseau classique très utilisé en traitement d'images) (Simonyan & Zisserman, 2014) est présenté à la Figure 2.8 afin de montrer le positionnement général des couches de pooling dans les CNN.

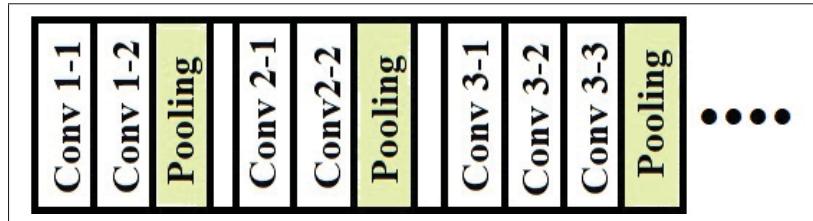


Figure 2.8 Positionnement général du pooling dans les CNN

2.2.4.1.5 Le Dropout

Le dropout est une forme de régularisation intégrée dans les réseaux de neurones dans le but de le rendre plus robuste et d'éviter le surapprentissage (Portaz, 2018). L'utilisation du dropout dans le réseau neuronal permet de désactiver aléatoirement certains neurones du réseau avec une probabilité fixe définie (0.5 par exemple). Cela force donc le réseau à s'adapter au manque d'informations et à être robuste. Une illustration de l'opération de dropout peut être vue à la Figure 2.9. La Figure 2.9(a) illustre un réseau entièrement connecté standard tandis que la Figure 2.9(b) illustre un réseau entièrement connecté après application du dropout avec sept neurones désactivés.

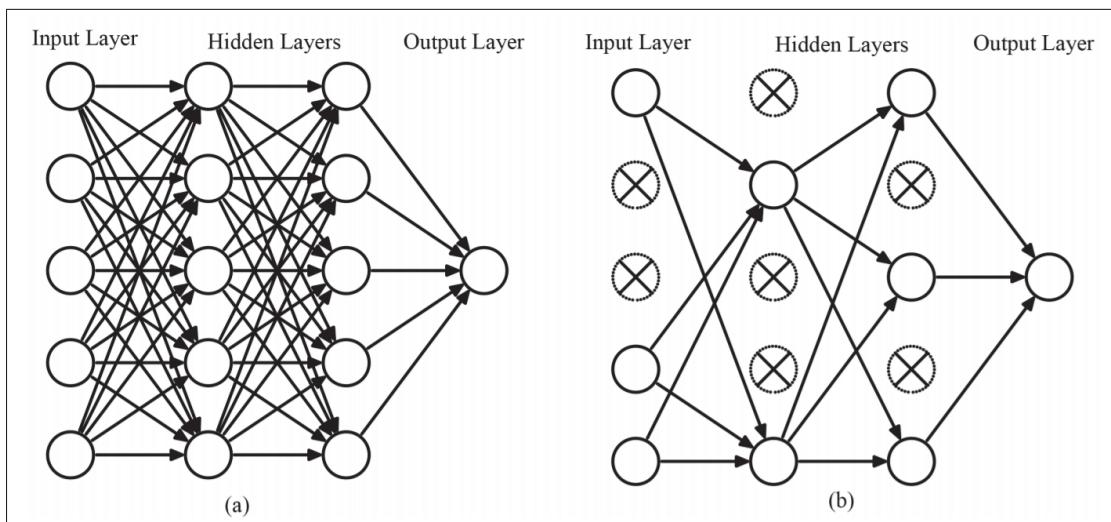


Figure 2.9 Opération de dropout. (a) Réseau de neurones sans dropout (b) Réseau de neurones après application du dropout
Tirée de Wang *et al.* (2020)

2.2.4.2 Segmentation par réseau de neurones

Dans l'état de l'art actuel, les méthodes offrant de meilleurs résultats pour les tâches de segmentation d'objets dans les images sont celles basées sur les réseaux de neurones (Chen *et al.*, 2017; Sirinukunwattana *et al.*, 2017). Nous pouvons distinguer deux types de segmentations : la segmentation d'instances et la segmentation sémantique. Les méthodes de segmentation d'instances reposent sur les méthodes de détection d'objets qui leur permettent d'obtenir les instances des classes présentes dans l'image (Ruiz-Santaquiteria, Bueno, Deniz, Vallez & Cristobal, 2020). Ils sont une combinaison des méthodes de détection d'objets et de segmentation sémantique (Ruiz-Santaquiteria *et al.*, 2020). La segmentation d'instances segmente individuellement chaque objet d'une image, donc attribue une étiquette unique à chacun des objets présents dans l'image, que les objets soient du même type ou non, tandis que la segmentation sémantique classifie chaque pixel de l'image et ne fait pas de différence entre deux instances d'un même objet (Ruiz-Santaquiteria *et al.*, 2020).

2.2.4.2.1 Segmentation d'instances

Différentes méthodes permettent de faire de la segmentation d'instances. DeepMask (Figure 2.10) (Pinheiro *et al.*, 2015) est un réseau neuronal constitué d'un module d'extraction de caractéristiques basé sur VGG-A (Simonyan & Zisserman, 2014) et d'un deuxième module neuronal à deux branches lui permettant de prédire des régions segmentées et leurs classes (Girshick, 2015). Ils obtiennent un rappel (recall) moyen de 18.3% sur l'ensemble de données MS COCO (Lin, Maire, Belongie, Hays, Perona, Ramanan, Dollár & Zitnick, 2014). Le petit bémol ici est que le réseau génère le masque d'un seul objet présent dans l'image même lorsqu'il y en a plusieurs dans l'image.

Mask R-CNN (Figure 2.11) (He *et al.*, 2017) est une méthode de segmentation d'instances qui effectue la segmentation sémantique d'un objet contenu dans une boîte englobante. Il est composé d'un bloc basé sur Fast R-CNN (Girshick, 2015) lui permettant de détecter des objets dans l'image via des boîtes englobantes de même que leur classe et d'un module de FCN (réseaux

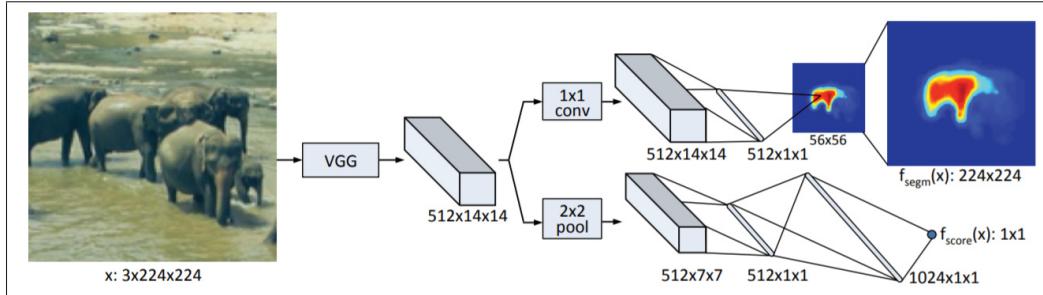


Figure 2.10 Architecture DeepMask
Tirée de Pinheiro *et al.* (2015)

entièlement convolutionnel) (Long *et al.*, 2015) lui permettant de segmenter sémantiquement chaque objet précédemment détecté. Ils obtiennent une précision moyenne (mAP : mean Average Precision) de segmentation de 37.1% sur la base de données COCO (Lin *et al.*, 2014).

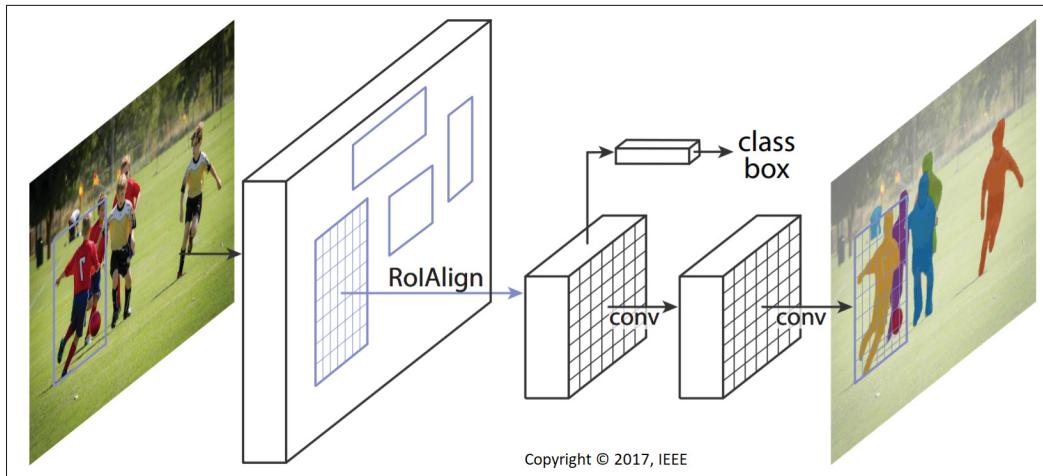


Figure 2.11 Architecture Mask R-CNN
Tirée de He *et al.* (2017)

RetinaMask (Fu, Shvets & Berg, 2019) est basé sur le réseau RetinaNet dans lequel un module de prédiction de masques d'instance a été intégré. Fu *et al.* (2019) obtiennent une précision de détection (mAP) de 36.4% sur l'ensemble de données COCO.

YOLOACT (Bolya, Zhou, Xiao & Lee, 2019) (Figure 2.13) est constitué de deux branches. La première utilise un FCN pour produire des masques tandis que la deuxième branche produit un

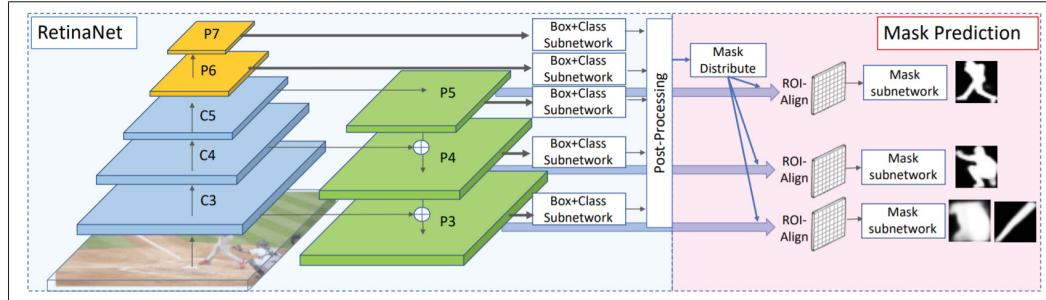


Figure 2.12 Architecture RetinaMask
Tirée de He *et al.* (2017)

vecteur de coefficients de masque d’instance. Ces deux branches sont par la suite combinées linéairement pour produire des masques d’instance. Bolya *et al.* (2019) obtiennent une précision de détection (mAP) de 31.2% sur l’ensemble de données COCO. YOLACT est très rapide (33.5 fps) mais moins précis que Mask-RCNN (mAP de 31.2% vs 35.7%) sur l’ensemble de données MS COCO (Lin *et al.*, 2014).

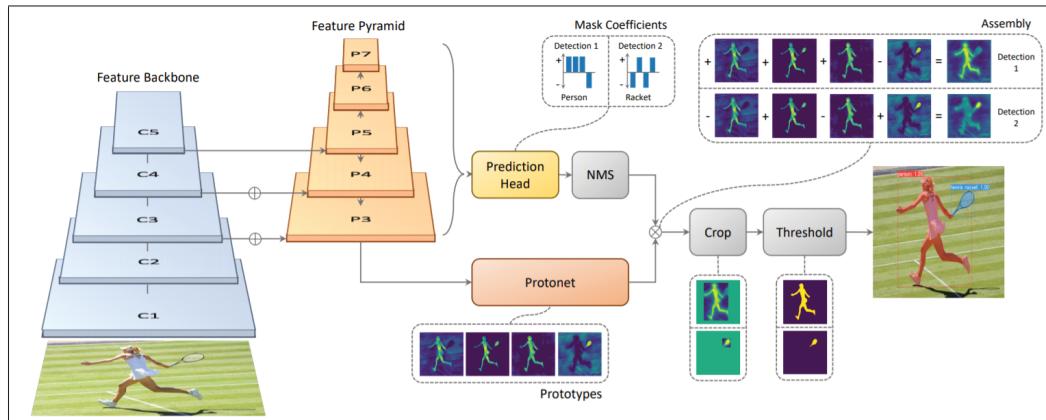


Figure 2.13 Architecture YOLACT
Tirée de He *et al.* (2017)

Le principal problème avec les méthodes de segmentation d’instances est que leur performance dépend de la performance de la méthode de détection d’objets utilisé. En effet, vu qu’elles se basent sur les méthodes de détection d’objets pour trouver les instances dans les images, seules les instances détectées sont segmentées. Ces approches ne sont pas toujours précises et les

boîtes de délimitation prédites n’englobent pas toujours les structures d’intérêt ou englobent seulement une partie de la structure d’intérêt (Fu *et al.*, 2019 ; Pinheiro *et al.*, 2015) ; ce qui n’est pas forcément intéressant dans notre cas, étant donné qu’on aimerait extraire l’ensemble des structures osseuses dans l’image.

2.2.4.2.2 Segmentation sémantique

L’introduction des réseaux entièrement convolutionnels (FCN) par (Long *et al.*, 2015) a permis une avancée notable en imagerie pour les tâches de segmentation sémantique. Long *et al.* (2015) ont adopté une architecture entièrement connectée afin de résoudre le problème de segmentation. L’architecture FCN (Figure 2.14) prend en entrée une image de n’importe quelle taille et produit en sortie la segmentation (image de même taille que l’entrée où chaque pixel est classifié). Elle est constituée de couches de sous-échantillonnage (conv+pool) pour extraire les informations représentant une abstraction de haut niveau et de couches de suréchantillonnage (conv+deconv) pour produire des masques de scores (pour compenser la perte engendrée par la couche de pooling). Dans le but de travailler avec des images de différentes tailles, ils adoptent un réseau n’ayant pas de couches entièrement connectées (fully connected layers). Ils remplacent donc les couches entièrement connectées par des couches de convolutions de noyau 1x1. Sur la base de données PASCAL VOC 2011, Long *et al.* (2015) obtiennent en utilisant comme métrique l’intersection des pixels sur l’union (mean intersection over union (IoU)), un IoU moyen de 62.7%.

Ce travail a inspiré différents autres travaux comme celui de Chen *et al.* (2017), qui ont proposé l’architecture DCAN (Figure 2.15), basée sur les FCN, pour segmenter des structures histologiques telles que des glandes et des noyaux cancéreux dans des images histologiques. À l’architecture FCN, Chen *et al.* (2017) ont ajouté la notion d’extraction de caractéristiques multi-échelles dans le but d’avoir une segmentation plus précise. DCAN est une architecture multitâches qui prédit à la fois les masques des structures histologiques et leurs contours. Elle a donc deux sorties : une pour la prédiction des masques et le second pour la prédiction des contours. Ces deux prédictions sont fusionnées par sommation élément pas élément pour obtenir

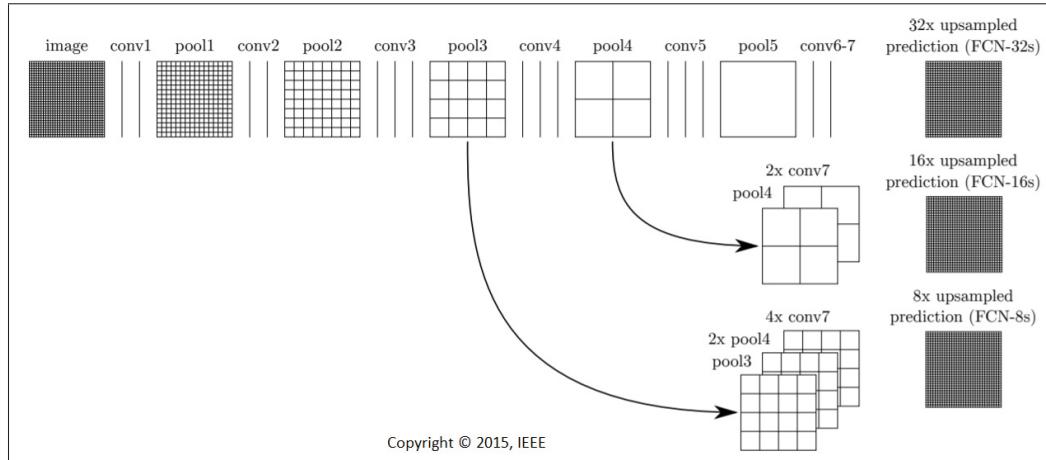


Figure 2.14 Architecture FCN
Tirée de Long *et al.* (2015)

la segmentation finale. Cette architecture leur a permis d'obtenir une segmentation plus précise que celle des autres participants de la compétition de segmentation d'objets histologiques de MICCAI 2015 (Chen *et al.*, 2017) ; ce qui leur a fait obtenir la première place. Ils obtiennent sur les ensembles de données histologiques (glandes et noyaux) de MICCAI 2015 un Dice de 86.8% pour les glandes et de 81.2% pour les noyaux.

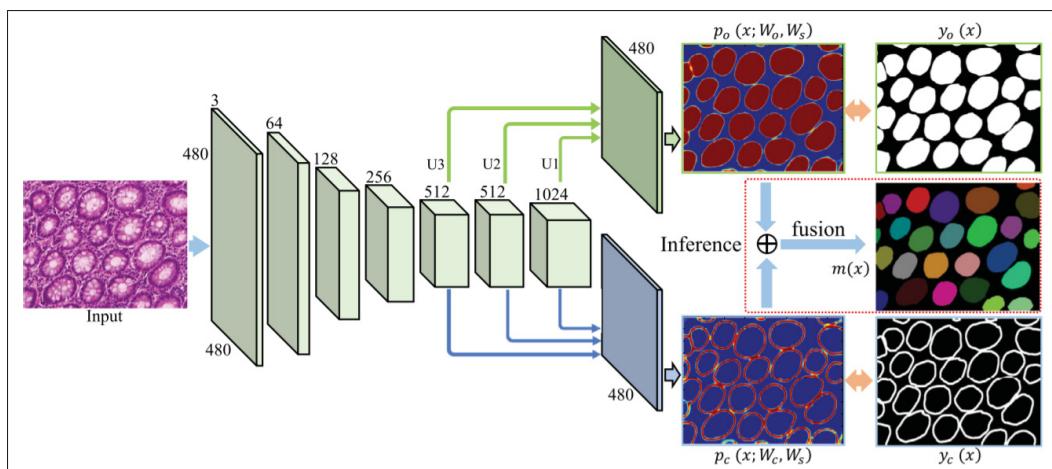


Figure 2.15 Architecture DCAN
Tirée de Chen *et al.* (2017)

Yu & Koltun (2016) ont quant à eux proposé un module convolutionnel constitué uniquement de couches de convolutions et de convolutions dilatées. Ce module a été placé à la suite d'un réseau neuronal basée sur le réseau VGG-16 (Simonyan & Zisserman, 2014). Cela leur a permis d'obtenir sur la base de données de PASCAL VOC 2012 un IoU moyen de 67.6% comparé à un IoU moyen de 62.2% pour l'architecture FCN.

Une autre architecture inspirée des FCN est l'architecture encodeur-décodeur U-Net (Figure 2.16) de Ronneberger *et al.* (2015). Largement utilisée, cette architecture en forme de U est composée d'une partie encodeur (première moitié du U sur la Figure 2.16) constituée d'opérations de convolution et de sous-échantillonnage (pooling) et d'une partie décodeur (deuxième moitié du U sur la Figure 2.16) constituée d'opérations de suréchantillonnage (déconvolution) et de convolution. Elle ne contient pas de couches entièrement connectées permettant la possibilité d'entraîner le réseau avec des images de différentes tailles. Un des points forts de cette architecture est que les cartes de caractéristiques issues de la partie encodeur sont directement insérées dans les cartes de caractéristiques de la partie décodeur (Ronneberger *et al.*, 2015). Ronneberger *et al.* (2015) obtiennent sur les données PhC-U373 et DIC-HeLa de l'ensemble de données cellulaire ISBI 2015 des IoU de 92.02% et 77.56% respectivement.

Zhang *et al.* (2018) intègrent dans l'architecture U-Net des blocs résiduels pour faciliter l'entraînement du réseau (Figure 2.17). Ils obtiennent sur des données publiques de routes du Massachusetts une précision de segmentation de 91.87%.

Badrinarayanan *et al.* (2017) intègrent dans leur réseau encodeur-décodeur SegNet (Figure 2.18), la notion d'indice de pooling rendant ainsi leur réseau beaucoup plus rapide à entraîner que U-Net avec des résultats similaires à ceux de U-Net (Alam, Wang, Guangpei, Yunrong & Chen, 2021 ; Hussain, Naqvi, Loh & Lee, 2021 ; Oliveira & dos Santos, 2018 ; Şahin, Alpaslan & Hanbay, 2021). En effet, au lieu de transmettre l'intégralité de la carte des caractéristiques de la partie encodeur dans la partie décodeur comme dans U-Net (Ronneberger *et al.*, 2015), Badrinarayanan *et al.* (2017) transmettent plutôt des indices de pooling de sorte que leur réseau (SegNet) utilise moins de mémoire. Ils obtiennent sur l'ensemble de données CamVid un IoU de 60.10%.

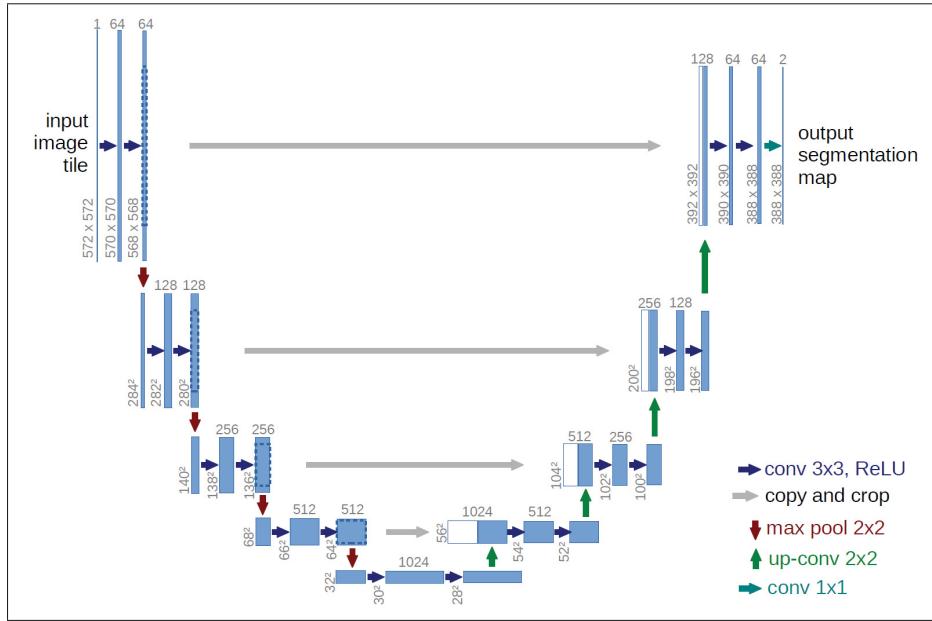


Figure 2.16 Architecture U-Net
Tirée de Ronneberger *et al.* (2015)

Une synthèse des approches présentées se retrouve dans le Tableau 2.1. Les approches de segmentation d’instances sont en premier lieu présentées suivie des approches de segmentation sémantique. Des deux approches, nous pouvons remarquer que les approches de segmentation sémantiques, particulièrement celles basées sur les réseaux entièrement convolutionnels (FCN, DCAN, U-Net, SegNet, ResUnet) sont celles donnant les meilleurs résultats. L’exploitation d’une approche de segmentation sémantique basée sur un réseau entièrement convolutionnel pourrait aider à répondre à notre problématique de segmentation des os dans les radiographies EOS.

2.2.4.3 Extraction de caractéristiques à partir de réseaux neuronaux

Contrairement aux méthodes d’extraction de caractéristiques décrites à la section 2.1.2 qui requièrent d’extraire dans un premier temps des points caractéristiques dans l’image avant d’effectuer une tâche donnée, les réseaux neuronaux transforment directement l’image en caractéristiques à mesure qu’on évolue dans le réseau, et cela, pendant qu’il effectue une tâche.

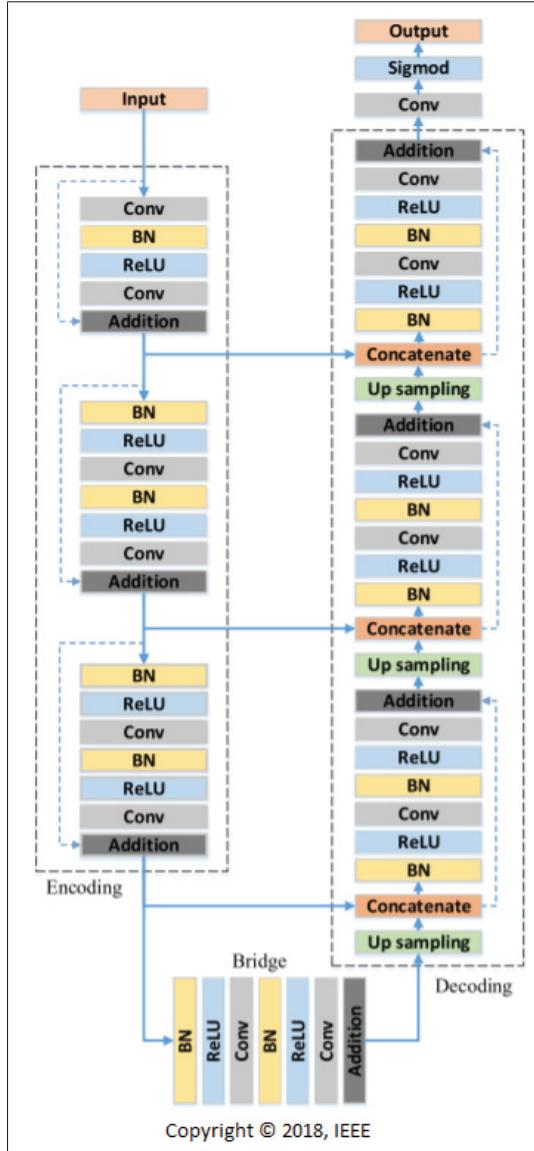


Figure 2.17 Architecture Deep ResNet
Tirée de Zhang *et al.* (2018)

Les réseaux de neurones donnent d'ailleurs les meilleurs résultats dans la littérature pour les tâches d'extraction de caractéristiques et de recherche de similarités dans les images. Différents travaux (Balntas *et al.*, 2016 ; Dosovitskiy, Fischer, Springenberg, Riedmiller & Brox, 2016 ; Razavian, Azizpour, Sullivan & Carlsson, 2014 ; Rocco, Arandjelović & Sivic, 2019) ont montré que les descripteurs issus des architectures neuronales offraient de meilleurs résultats que ceux issus des méthodes classiques comme SIFT, SURF, HOG (Histogram of oriented gradients), etc.

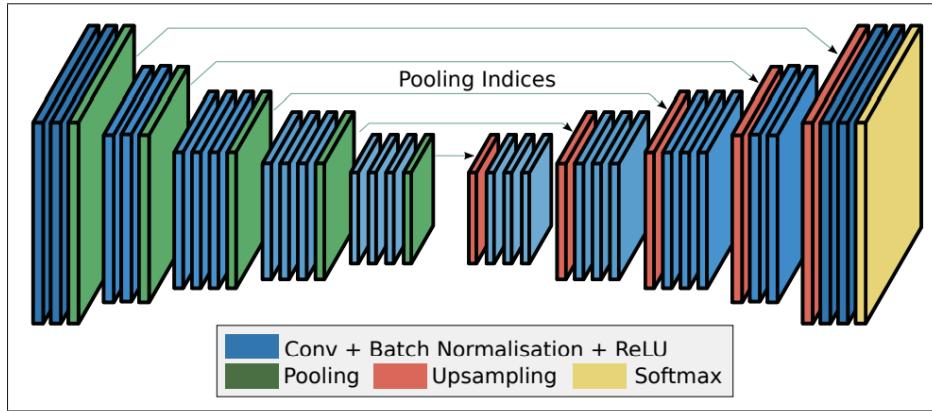


Figure 2.18 Architecture SegNet
Tirée de Badrinarayanan *et al.* (2017)

Tableau 2.1 Synthèse de quelques approches neuronales de segmentation

Méthode	Approche	Métrique	Ensemble de données	Performance
(Pinheiro <i>et al.</i> , 2015)	DeepMask	Rappel moyen	COCO	18.3%
(He <i>et al.</i> , 2017)	Mask R-CNN	mAP	COCO	37.1%
(Fu <i>et al.</i> , 2019)	RetinaMask	mAP	COCO	36.4%
(Bolya <i>et al.</i> , 2019)	YOLACT	mAP	COCO	31.2%
(Long <i>et al.</i> , 2015)	FCN	IoU	PASCAL VOC 2011 2012	62.7% 62.2%
(Chen <i>et al.</i> , 2017)	DCAN	Dice	2015 MICCAI Segmentation Challenge	86.8% & 81.2%
(Ronneberger <i>et al.</i> , 2015)	U-Net	IoU	PhC-U373 & DIC-HeLa	70.0% 68.4%
(Badrinarayanan <i>et al.</i> , 2017)	SegNet	IoU	CamVid	60.10%
(Zhang <i>et al.</i> , 2018)	ResUnet	précision	Données de route	91.87%

Yi, Trulls, Lepetit & Fua (2016), via leur architecture neuronale LIFT (Learned Invariant Feature Transform), extraient des points caractéristiques (avec leurs orientations et descripteurs) qui performent mieux que celles des méthodes SIFT, SURF, KAZE, ORB,... pour la tâche de mise en correspondance d'images. Balntas *et al.* (2016) ont montré que l'utilisation de descripteurs de même dimensionnalité que les descripteurs SIFT, extraits à partir de leur réseau de neurones convolutionnel PN-Net, pouvait réduire les erreurs de mise en correspondance d'images de 26% (SIFT) à 7% avec un temps d'extraction qui est 40 fois plus rapide que SIFT.

D'autres approches neuronales comme openPose (Cao, Hidalgo, Simon, Wei & Sheikh, 2021) ou openFace (Baltrušaitis, Zadeh, Lim & Morency, 2018) permettent de retrouver des points caractéristiques dans l'image. OpenPose est un réseau de neurones convolutionnel qui prédit à la fois la position des différentes parties du corps ainsi que l'association existant entre elles. Le problème avec OpenPose est que le niveau de précision des estimations de points d'intérêt est limité en raison de ses sorties à faible résolution ; ce qui le rend moins adapté aux applications exigeantes en précision comme en médical (Groos, Ramampiaro & Ihlen, 2021). OpenFace est un outil qui permet de détecter des points d'intérêt du visage via l'utilisation d'un réseau de neurones et d'un algorithme de détection de points d'intérêt du visage. OpenFace est basé sur la méthode CE-CLM (Convolutional Experts Constrained Local Model) de Zadeh, Lim, Baltrušaitis & Morency (2017) qui est composé d'un réseau convolutionnel qui prédit l'emplacement des points d'intérêt et d'un modèle de distribution de points qui capture les variations de formes des points d'intérêt. Ils optimisent la méthode CE-CLM de sorte à être plus rapide avec une réduction minimale de la précision.

Les méthodes extrayant des caractéristiques et faisant la tâche de recherche de similarité entre les images à partir de réseaux neuronaux utilisent le plus souvent un type de réseau de neurones appelé réseaux siamois (Balntas *et al.*, 2016; Rocco *et al.*, 2019).

Réseaux de neurones siamois

Les réseaux de neurones siamois (Figure 2.19) sont un type de réseaux utilisé pour extraire des caractéristiques et pour mesurer la similarité entre deux entrées indépendantes X_1 et X_2 partageant une certaine relation de similarité. Ils sont composés de sous-réseaux identiques appelés réseaux jumeaux partageant les mêmes paramètres et poids (Dey, Dutta, Toledo, Ghosh, Lladós & Pal, 2017). Les réseaux jumeaux produisent en sortie des représentations vectorielles V_1 et V_2 (Figure 2.19) qui sont comparées à l'aide d'une métrique de similarité pouvant être par exemple le cosinus de l'angle entre deux vecteurs (Bromley, Guyon, LeCun, Säckinger & Shah, 1994) ou bien la distance euclidienne (Dey *et al.*, 2017). Différents travaux (Balntas *et al.*, 2016; Han, Leung, Jia, Sukthankar & Berg, 2015; Rocco *et al.*, 2019; Simo-Serra, Trulls, Ferraz,

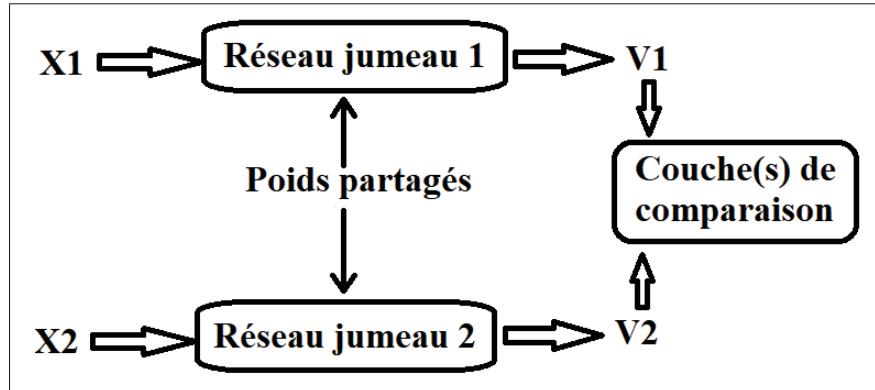


Figure 2.19 Réseau siamois

Kokkinos, Fua & Moreno-Noguer, 2015 ; Yi *et al.*, 2016 ; Zagoruyko & Komodakis, 2015) se sont intéressés aux processus d'appariement d'images en développant une architecture siamoise neuronale (prenant des paires d'images). Han *et al.* (2015) ont développé l'architecture MatchNet, un réseau siamois à deux branches qui dans un premier temps extrait des caractéristiques présentes dans chaque paire d'imagettes pour par la suite calculer la similarité entre ces caractéristiques grâce à un réseau entièrement connecté. Yi *et al.* (2016) ont développé la méthode LIFT (Learned Invariant Feature Transform) basée sur un réseau siamois à quatre branches qui dans un premier temps extrait des points caractéristiques puis utilise la distance euclidienne comme mesure de similarité. Rocco *et al.* (2019) quant à eux adoptent une approche neuronale pour estimer la transformation géométrique existante entre deux paires d'images. Leur réseau neuronal est constitué d'un réseau siamois permettant de mettre en correspondance les paires d'images, suivi d'un réseau de neurones convolutionnel qui prédit les paramètres de la transformation géométrique existant entre les paires d'images. L'avantage d'utiliser des réseaux siamois réside dans leur capacité à apprendre la similarité entre les images (Dey *et al.*, 2017).

2.3 Extraction d'informations dans les images médicales

2.3.1 Segmentation de structures dans les images médicales

Différents travaux dans la littérature se sont intéressés à l'automatisation de la tâche de segmentation de structures dans les images médicales en utilisant entre autres les modèles de formes (modèles déformables, les atlas, les modèles statistiques), les forêts aléatoires et les réseaux de neurones.

Parmi les travaux basés sur les modèles de forme, nous avons Linguraru, Sandberg, Li, Shah & Summers (2010) qui ont développé une méthode basée sur les atlas pour segmenter le foie et la rate sur des images CT-scan. L'approche adoptée consiste à construire dans un premier temps, un atlas du foie et de la rate à partir d'images CT. Une mise en correspondance entre l'atlas construit et l'image à segmenter est ensuite réalisée de sorte à produire une segmentation grossière qui est par la suite raffinée par un contour actif. Ils obtiennent un Dice moyen de segmentation de $95.7 \pm 1\%$. Ding *et al.* (2007), pour leur part, ont utilisé une approche basée sur un atlas pour segmenter automatiquement le fémur proximal dans des radiographies antéro-postérieures. Dans un premier temps, un atlas contenant le pelvis et le fémur proximal est construit à partir de 20 radiographies d'entraînement. L'atlas construit est ensuite aligné par recalage sur les bords correspondants dans l'image de test. La méthode "Level set" avec contrainte de forme est appliquée pour déformer l'atlas aligné dans le but d'extraire les contours des fémurs. Ils obtiennent un RMSD moyen (Root-mean-squared distance) de 3.02 ± 0.90 pixels. Dans le but de segmenter des images médicales telles que des images histologiques de microscopie, des images CT-scan et IRM, Mesejo, Valsecchi, Marrakchi-Kacem, Cagnoni & Damas (2015) ont développé une approche basée sur la méthode "Level Set". Cette approche combine l'information basée sur les bords et régions et la connaissance de la forme préalable de l'objet cible obtenue par l'intermédiaire d'un atlas pour guider l'évolution du «Level Set». Premièrement, un apprentissage du "level set" est fait au moyen d'un algorithme génétique, et ensuite la segmentation proprement dite est faite, avec l'algorithme métahéuristique, recherche par dispersion (Scatter Search). Ils obtiennent un Dice moyen de $87.75 \pm 5.4\%$. Ce travail est intéressant du fait que leur méthode

permet d'effectuer une segmentation robuste sur différentes modalités d'images médicales. Elle est par contre coûteuse computationnellement. Un autre travail basé sur les modèles de forme est celui de Guillen *et al.* (2018) qui utilisent des modèles statistiques de forme (SSM) et d'apparence (SAM) pour segmenter le fémur proximal dans des radiographies pelviennes antéro-postérieures avec un Dice moyen de 80%. Chen *et al.* (2005), quant à eux, utilisent un modèle 2D de contour du fémur à 5 sommets pour segmenter le fémur sur de radiographies de la hanche. Ils détectent dans un premier temps des diaphyses et têtes fémorales candidates dans l'image, de même qu'un point caractéristique situé sur le grand trochanter. Le modèle 2D de contour du fémur est ensuite aligné sur l'image radiographique en fonction des caractéristiques détectées. L'algorithme des contours actifs avec contrainte de forme est par la suite appliqué pour identifier le contour du fémur. Ils obtiennent un taux de succès de 81.4% sur 172 images de test. Le fémur est considéré comme extrait avec succès si l'erreur maximale du point le plus proche au contour est inférieure à 8 pixels. Cette méthode est intéressante, mais ne fonctionne pas sur les fémurs présentant une forme un peu différente de la normale. Cette limitation est probablement due à l'utilisation du contour actif qui est sensible au bruit et qui a tendance à être attiré par des contours différents des contours recherchés (Chen *et al.*, 2005). Une autre explication est que le modèle 2D de contour du fémur utilisé n'est pas représentatif (ne représente pas les fémurs ayant une forme différente) ou que le nombre de sommets (5 sommets) utilisé pour caractériser le modèle 2D de contour du fémur est trop petit.

Certains travaux ont combiné les modèles de formes avec les forêts aléatoires pour effectuer la tâche de segmentation d'images. C'est le cas de Chen & Zheng (2014) qui ont combiné les compositions de formes clairsemées (SSC : "Sparse Shape Composition") et les forêts aléatoires pour segmenter le fémur proximal gauche et le bassin gauche dans les radiographies antéro-postérieures du bassin. La position approximative des points d'intérêt caractérisant chacune des structures d'intérêt est dans un premier temps régressée à travers une forêt aléatoire puis utilisée dans une composition de forme clairsemée pour obtenir une forme régularisée. Ils ont obtenu un taux de succès de 96.7% et considèrent la segmentation comme réussie lorsque la distance moyenne des points à la courbe est inférieure à 4 mm. Lindner *et al.* (2013), quant à eux,

ont utilisé les forêts aléatoires et un modèle de forme statistique pour segmenter automatiquement le fémur proximal sur les radiographies pelviennes antéro-postérieures. Leur approche comprend une recherche globale où une forêt aléatoire prédit le centre d'une boîte candidate contenant le fémur proximal et une recherche locale où une autre forêt aléatoire et un modèle de forme statistique prédisent un ensemble de points d'intérêt autour du fémur permettant ainsi de le segmenter. Ils obtiennent une erreur moyenne point à la courbe (Point-to-Curve error) inférieure à 0.9mm dans 99% des cas.

Les méthodes basées sur des modèles de formes (Chen & Zheng, 2014; Ding *et al.*, 2007; Guillen *et al.*, 2018; Linguraru *et al.*, 2010; Mesejo *et al.*, 2015) telles que les atlas, les modèles statistiques par exemple, sont robustes mais complexes à mettre en place dans le cas d'images présentant différents points de vue et orientations comme c'est le cas avec les images EOS. En effet, il faudrait mettre en place un atlas ou un modèle statistique selon la région que l'on souhaite segmenter. Dans le cas où on serait intéressé par le membre inférieur au complet ou la hanche ou le genou par exemple, cela impliquerait de mettre en place un modèle de formes spécifique pour le membre inférieur complet ou la hanche ou le genou. Leur utilisation n'est donc pas optimale dans notre cas, étant donné qu'on aimerait mettre en place une unique méthode pouvant fonctionner sur différentes régions.

D'un autre côté, nous avons les méthodes basées sur les réseaux de neurones qui segmentent différentes régions simultanément dans l'image avec de bons résultats. Parmi les approches de segmentation basées sur les réseaux de neurones, nous pouvons distinguer les méthodes de segmentation d'instances et les méthodes de segmentation sémantique. Parmi les travaux faisant de la segmentation d'instances nous avons Thang, Dung, Duc, Nguyen, Nguyen, Anh, Cuong, Hanh, Phu, Phuong *et al.* (2021) qui ont proposé une méthode utilisant le réseau Mask R-CNN pour segmenter des régions anormales et classifier le niveau d'anormalité de la région segmentée sur des images radiographiques de mammographie. Leur méthode comprend trois étapes à savoir : le prétraitement des données (retirer l'étiquette de l'image, séparer l'arrière-plan de la mammographie, augmenter le contraste et réduire le bruit), l'extraction d'imagettes et l'entraînement du réseau Mask R-CNN. Ils obtiennent une précision de détection

(mAP) de 81.09% et une précision de classification de 58.44%. Wessel, Heinrich, von Berg, Franz & Saalbach (2019), quant à eux, utilisent le réseau Mask R-CNN pour segmenter les côtes dans des radiographies thoraciques antéro-postérieures avec un coefficient de Dice de $73.3 \pm 20.9\%$. Ceron, Chang, Ochoa-Ruiz & Ali (2021), pour leur part, introduisent des modules d'attention CCAM (Criss-cross Attention Module) (Huang, Wang, Wei, Huang, Shi, Liu & Huang, 2020) dans l'architecture YOLACT leur permettant de faire de la segmentation temps réel des instruments médicaux dans des images issues d'instruments laparoscopiques. Ils obtiennent un coefficient de Dice de 31.3% sur l'ensemble de données ROBUST-MIS. Cette méthode n'est pas précise, mais permet de faire de la segmentation en temps réel (49 images par seconde).

Les méthodes de segmentation d'instances sont souvent utilisées pour des applications temps réels grâce à leur rapidité de calcul (Ruiz-Santaquiteria *et al.*, 2020). Elles sont par contre pas très précises (Ruiz-Santaquiteria *et al.*, 2020).

D'un autre côté, nous avons les méthodes de segmentation sémantiques basées sur les réseaux entièrement convolutionnels (une architecture très utilisée en segmentation sémantique) qui donnent des résultats très intéressants. Un réseau entièrement convolutionnel a été utilisé par Dolz, Desrosiers & Ayed (2018) pour segmenter des structures sous-corticales du cerveau dans des images IRM du cerveau avec un Dice moyen de 83.75% et de 89% respectivement sur les ensembles de données publiques d'images IRM du cerveau ABIDE et ISBR. Oliveira & dos Santos (2018) ont utilisé les architectures SegNet et UNet pour segmenter le poumon, le coeur et les clavicules dans des radiographies thoraciques postéro-antérieures. Ils obtiennent un indice de Dice moyen de $95.08 \pm 40.33\%$ et de $94.9 \pm 61.67\%$ en utilisant SegNet et UNet respectivement. Dans le but d'évaluer le rétrécissement de l'espace articulaire (dans le genou) permettant de quantifier le degré de l'arthrose du genou, Swiecicki, Said, O'Donnell, Buda, Li, Jiranek & Mazurowski (2020) ont utilisé l'architecture neuronale U-net pour segmenter le genou (fémur distal et tibia proximal) avec un Dice moyen de 96.59% sur des radiographies postéro-antérieures du genou. Kwolek *et al.* (2019) dans le but de mesurer l'alignement des membres inférieurs et l'orientation des articulations, segmentent les fémurs et les tibias dans des

radiographies frontales des membres inférieurs. La segmentation des os est réalisée à partir de l'architecture U-Net. Ils obtiennent un Dice moyen de 96.8%. Les fémurs et les tibias ne sont par contre pas différenciés, c'est-à-dire une même classe est associée aux fémurs et aux tibias. Une modification de l'architecture U-net a été utilisée par Lianghui, Gang, Yang & Bin (2019) pour segmenter les fémurs proximaux gauche et droit dans les radiographies frontales de hanche. La modification apportée à U-Net a été d'enlever les couches de pooling et d'y intégrer des couches de batch normalisation. Ils obtiennent un Dice de 96.60%. Par contre, les fémurs segmentés ne sont pas différenciés (les fémurs gauche et droit ont la même classe). Shen, Xu, Zhang, Sun, Ma, Ma, Zhou, Guo & Wang (2020), quant à eux, ont développé un réseau neuronal basé sur U-Net dans lequel des convolutions dilatées ont été intégrées, pour segmenter les os fémoraux et tibiaux dans les images radiographiques en vue frontale. Ils ont obtenu un indice de Dice moyen de 97.26%. Les deux fémurs et les deux tibias ne sont par contre pas dissociés.

Une synthèse des méthodes présentées se retrouve dans le Tableau 2.2. Les approches basées sur des méthodes conventionnelles sont en premier lieu présentées suivies des approches basées sur des méthodes d'apprentissage machine. Nous pouvons remarquer dans le Tableau 2.2 que les approches donnant les meilleurs résultats sont celles basées sur les méthodes d'apprentissage machine en particulier celles basées sur les réseaux entièrement convolutionnels encodeur-décodeurs en forme de U tels que U-Net ou SegNet. L'exploitation d'une méthode encodeur-décodeur en forme de U pourrait aider à répondre à notre problématique de segmentation de structures osseuses.

2.3.2 Identification de points d'intérêt dans les images médicales

Différentes approches se sont intéressées à l'automatisation de la tâche d'identification de points d'intérêt dans les images médicales en se basant sur les approches conventionnelles d'identification de caractéristiques telles que l'algorithme SIFT et la transformée de HOUGH. Nous avons nous-même exploré cette piste pour détecter automatiquement le centre de la tête fémorale par recalage avec l'algorithme ASIFT et la transformée de Hough dans les radiographies frontales EOS. La stratégie adoptée consiste à trouver dans un premier temps

Tableau 2.2 Synthèse de performance de différentes méthodes de segmentation dans les images médicales

Travaux	Méthode	Type d'image	Structure identifiée	Métrique	Résultat
(Linguraru <i>et al.</i> , 2010)	Atlas + contour actif	CT-scan	Foie & rate	Dice	$95.7 \pm 1\%$
(Ding <i>et al.</i> , 2007)	Atlas + "level set"	Radio	Fémur proximal	RMSD	3.02 ± 0.90 pixels
(Mesejo <i>et al.</i> , 2015)	"level set + SS "	Images médicales	Differentes structures	Dice	$87.75 \pm 5.4\%$
(Guillen <i>et al.</i> , 2018)	SSM + SAM	Radio	Fémur proximal	Dice	80%
(Chen <i>et al.</i> , 2005)	modèle 2D+contour actif	Radio	Fémur proximal	Taux de succès	81.4%
(Chen & Zheng, 2014)	SSC + forêt aléatoire	Radio	Fémur proximal & Bassin	Taux de succès	96.7%
(Lindner <i>et al.</i> , 2013)	Forêt aléatoires + SSM	Radio	Fémur proximal	Point à la courbe	0.9mm
(Thang <i>et al.</i> , 2021)	Mask R-CNN	radiographie	Seins	mAP	81.09%
(Wessel <i>et al.</i> , 2019)	Mask R-CNN	radiographie	Côtes	Dice	$73.3 \pm 20.9\%$
(Ceron <i>et al.</i> , 2021)	YOLACT	laparoscopie	Instruments médicaux	Dice	31.3%
(Dolz <i>et al.</i> , 2018)	FCN	IRM	Cerveau	Dice	83.75% & 89%
(Oliveira & dos Santos, 2018)	SegNet UNet	Radio	Poumon,coeur,clavicule	Dice	95.08% 94.9%
(Swiecicki <i>et al.</i> , 2020)	U-Net	Radio	Genou	Dice	96.59%
(Kwolek <i>et al.</i> , 2019)	U-Net	Radio	Femur et Tibia	Dice	96.2%
(Lianghui <i>et al.</i> , 2019)	U-Net	Radio	Fémur proximal	Dice	96.6%
(Shen <i>et al.</i> , 2020)	U-Net	Radio	Fémur & Tibia	Dice	97.26%

l'image d'entraînement ayant la plus forte similarité avec l'image de test en se basant sur les descripteurs ASIFT. La transformée de Hough est par la suite utilisée pour retrouver la position du centre de la tête fémorale sur l'image de test (voir ANNEXE I pour plus de détail). L'erreur moyenne d'identification obtenue est assez grande 94.86 ± 169.74 mm avec un taux de succès de 62%. Ce fort taux d'erreur est probablement dû à la sensibilité de l'algorithme ASIFT et de la transformée de HOUGH (des approches conventionnelles en général) aux valeurs d'intensité de pixels et au bruit dans l'image.

Pour surmonter ces limitations, certaines approches combinent des approches conventionnelles d'identification de caractéristiques avec des approches d'apprentissage machine. C'est le cas de Wang, Li, Li, Zhang & Zou (2018b) qui ont développé une approche combinant les méthodes

SIFT et CART. Ils extraient dans un premier temps des vecteurs caractéristiques SIFT sur des imagettes générées aléatoirement sur une grille multi-résolution. L'arbre décisionnel CART est ensuite utilisé pour apprendre le lien existant entre ces vecteurs caractéristiques et les vecteurs de déplacement quittant les centres des imagettes vers la position des points d'intérêt céphalométriques. Ils obtiennent une erreur moyenne de 1.69 ± 1.43 mm sur l'ensemble de données de radiographies céphalométriques latérales du Challenge ISBI 2015 (Wang, Huang, Lee, Li, Chang et al., 2016). Guo, Mu, Xue, Li, Chen, Yan, Xu & Wang (2021), pour leur part, ont proposé une méthode basée sur des descripteurs SIFT et un perceptron multicouches de régression à trois couches cachées pour identifier quatre points d'intérêt du pied sur des radiographies latérales du calcanéum avec une erreur de rayon moyen (Mean Radius Error) de 1.85 ± 1.61 mm. Afin d'identifier le centre des vertèbres L1 à L5 dans des images IRM, Suzani, Rasoulian, Seitel, Fels, Rohling & Abolmaesumi (2015) utilisent un perceptron multicouches à trois couches cachées. Un seuillage d'Otsu est ensuite effectué pour raffiner les positions des centres vertébraux prédits par leur réseau. Ils obtiennent une erreur moyenne de 3.0 ± 2.4 mm. Dans (Olory Agomma, Vázquez, Cresson & De Guise, 2018), nous avons développé une méthode automatique capable d'identifier la position de treize points d'intérêt sur les radiographies frontales EOS. L'approche suivie consiste à identifier dans un premier temps des points caractéristiques générés sur la base des différences des niveaux d'intensité (DNI) dans l'image (Irrera, 2015), puis à apprendre, par l'intermédiaire d'un réseau de neurones aléatoire ou d'une forêt aléatoire, le déplacement de chaque point caractéristique vers la position de treize points d'intérêt préalablement identifiés par un expert. Pour chaque emplacement possible de points d'intérêt, il y a donc autant de prédictions qu'il y a de points caractéristiques. Les différentes prédictions sont par la suite combinées grâce à l'algorithme DBSCAN (Ester, Kriegel, Sander & Xu, 1996) pour permettre de prédire la position des points d'intérêt. Cette méthode est robuste, mais n'est cependant pas très précise (erreur d'identification de 29 ± 18 mm) et a été uniquement appliquée aux radiographies EOS frontales.

Les approches combinant les méthodes conventionnelles et d'apprentissage machine utilisent généralement les méthodes conventionnelles pour extraire des caractéristiques et les méthodes

d'apprentissage machine pour apprendre le lien existant entre les caractéristiques extraites et l'emplacement des points d'intérêt (Guo *et al.*, 2021 ; Wang *et al.*, 2018b). Différents travaux ont montré la supériorité de certaines méthodes d'apprentissage machine (exemple : CNN) par rapport à certaines méthodes conventionnelles d'extraction de caractéristiques (exemple : SIFT) (Balntas *et al.*, 2016 ; Rocco *et al.*, 2019).

Cela a poussé certains chercheurs à développer des approches basées uniquement sur les méthodes d'apprentissage machine pour la prédiction de l'emplacement des points d'intérêt dans les images médicales. C'est le cas de Criminisi, Robertson, Pauly, Glocker, Konukoglu, Shotton, Mateus, Möller, Nekolla & Navab (2013) qui ont utilisé les forêts de régression pour prédire simultanément l'emplacement d'organes d'intérêts (exemples : cœur, estomac) sur des images 3D tomographiques et de résonance magnétique. Pour ce faire, ils ont modélisé les positions des organes dans les scans par des boîtes englobantes 3D, de sorte que chaque boîte englobante contienne un organe d'intérêt, pour ensuite estimer simultanément le centre de ces boîtes par régression. Cette méthode n'est par contre pas très précise ($13.5 \pm 13mm$ pour les images CT). Aubert, Vazquez, Cresson, Parent & de Guise (2019) ont, quant à eux, développé une méthode automatique de reconstruction 3D du rachis dans laquelle un modèle statistique du rachis est recalé sur des paires de radiographies EOS du rachis, à l'aide d'un CNN. Cela leur a permis d'obtenir une erreur moyenne (3D) de 1.9 ± 1.33 mm pour l'identification des centres des corps vertébraux, des plateaux et des pédicules dans les radiographies EOS du rachis. Ce travail donne des résultats intéressants, mais requiert par contre une connaissance a priori (modèle statistique 3D) de la structure concernée. Dans le cas où l'on disposeraient de différentes structures prises sous différentes orientations, cela reviendrait à mettre en place un modèle statistique par structure (exemples : fémur, tibia) ; ce qui n'est pas optimal. Dans le but d'évaluer l'angle de Cobb de la scoliose sur des radiographies antéro-postérieures, Yi, Wu, Huang, Qu & Metaxas (2020), ont développé une méthode d'identification de points d'intérêt des vertèbres (17 vertèbres de la colonne thoracique et lombaire). Leur approche consiste à identifier dans un premier temps le centre des vertèbres via un réseau de neurones puis, sur la base de celui-ci, à régresser les quatre points d'intérêt situés aux coins de la vertèbre. Ils obtiennent une erreur de détection

moyenne de 50.11 (unité pas indiquée). Leur réseau n'arrive pas à identifier les vertèbres ayant une morphologie différente de celle des autres vertèbres dans l'image. Payer, Štern, Bischof & Urschler (2016) ont étudié la capacité des CNN à localiser plusieurs points d'intérêt en régressant des cartes thermiques dans des données d'images 2D et 3D de la main. Ils ont proposé pour cela une nouvelle architecture (SpatialConfiguration-Net) qui combine l'apparence locale avec des configurations de repères spatiaux. Cette architecture performe par contre un peu moins bien que l'architecture U-Net qu'ils ont testé (1.13 ± 0.98 mm avec 12 valeurs aberrantes > 10 mm vs 0.87 ± 1.05 mm avec 15 valeurs aberrantes > 10 mm) pour les images 2D. Zhang, Liu & Shen (2017), pour leur part, ont développé une méthode constituée de deux réseaux neuronaux placés l'un à la suite de l'autre pour prédire la position des points d'intérêt (10 pour le cerveau et 7 pour la prostate) sur des images 3D du cerveau (IRM) et de la prostate (CT-scan). Le premier réseau est entraîné sur des imagettes dans le but d'apprendre le lien entre ces imagettes et les points d'intérêt, tandis que le second est entraîné sur les images entières dans le but de tenir compte de l'information spatiale entre les points d'intérêt durant la prédiction. Ils obtiennent une erreur moyenne de 2.86 ± 1.55 mm et de 3.34 ± 2.55 mm respectivement pour le cerveau et la prostate. Zheng, Liu, Georgescu, Nguyen & Comaniciu (2015), utilisent initialement un réseau peu profond avec une couche cachée pour générer plusieurs points d'intérêt candidats. Des imagettes centrées sur les repères sont ensuite extraites et classées via un réseau plus profond. L'inconvénient avec cette stratégie est que l'information de contexte sur l'emplacement des points d'intérêt n'est pas prise en compte par le réseau. Afin de tirer pleinement profit de la relation spatiale existant entre les points d'intérêt, Liu, Wang, Jiang, Chi, Zhang & Hua (2020) utilisent un réseau multitâches qui prédit à la fois la position des points d'intérêt et des arrêtes liant deux points d'intérêt avec une erreur moyenne de 2.14mm. Dans le but de localiser seize points d'intérêt des côtes sur des radiographies postéro-antérieures thoraciques, Mader, von Berg, Fabritz, Lorenz & Meyer (2018) ont utilisé un réseau de neurones convolutionnel régulé par un CRF (Conditional Random Fields). L'architecture neuronale FCN a donc été utilisée en premier temps pour prédire grossièrement la position des points d'intérêt par l'intermédiaire de cartes de chaleur, ce qui a par la suite été affiné par l'utilisation d'un CRF. Ils obtiennent une erreur moyenne de 4.7mm. Yang, Zhang, Yan, Tan, Li & Metaxas (2015) identifient des points d'intérêt

du fémur distal dans des images 3D IRM avec une erreur moyenne de 4.69 ± 2.30 mm. La méthode adoptée consiste à convertir dans un premier temps leurs images 3D en trois ensembles d'images 2D selon les plans frontal, sagittal et transversal. Chaque ensemble d'images 2D est ensuite entraîné séparément via trois classifieurs CNN (un classifieur pour chaque plan d'image). La position 3D du point d'intérêt est obtenue en combinant les résultats issus des classifieurs. Bier, Goldmann, Zaech, Fotouhi, Hegeman, Grupp, Armand, Osgood, Navab, Maier et al. (2019), quant à eux, détectent des points d'intérêt dans les images simulées du bassin, présentant des champs de vue arbitraires. Pour ce faire, des images radiographiques synthétiques sont générées et utilisées pour entraîner un réseau neuronal qui prédit par régression la position de 23 points d'intérêt de la hanche. Ce travail est intéressant du fait qu'il considère des images de la hanche avec différents points de vue. Par contre, la phase d'entraînement de leur réseau a uniquement été réalisée sur des images synthétiques ; ce qui a révélé de grandes erreurs lors de la validation sur les images radiographiques réelles. Ils obtiennent une erreur de détection de 18.9 mm pour les radiographies réelles et une erreur de 5.6 ± 4.5 mm pour les images synthétiques. Tsai (2021) prédit par régression à l'aide d'un réseau de neurones convolutionnel la position de 3 points d'intérêt des membres inférieurs (tête fémorale, condyle fémoral médial et centre du plafond tibial) sur des radiographies EOS antéro-postérieures avec une erreur d'identification de 3.6 ± 9.9 mm. Ce réseau considère par contre les images de façon individuelle et ne serait donc pas en mesure de prédire l'emplacement 3D de points stéréo-correspondants appartenant à deux images à la fois.

Une synthèse des méthodes présentées se retrouve dans le Tableau 2.3. Les approches basées sur des méthodes conventionnelles sont en premier lieu présentées suivies des approches basées sur des méthodes d'apprentissage machine. Nous pouvons remarquer dans le Tableau 2.3 que les approches basées sur les méthodes d'apprentissage machine en particulier celles basées sur les réseaux de neurones convolutionnels donnent des résultats intéressants. L'exploitation d'une méthode basée sur les réseaux de neurones convolutionnels pourrait aider à répondre à notre problématique de localisation de points d'intérêt.

Tableau 2.3 Synthèse de performance de différentes méthodes d’identification de points d’intérêt dans les images médicales (distance euclidienne utilisée comme métrique)

Travaux	Méthode	Type d’image	Structure identifiée	Résulat
Experiementation (Wang <i>et al.</i> , 2018b)	ASIFT+Hough SIFT+CART	Radio EOS Radio	tête fémorale crâne	$94.86 \pm 169.74\text{mm}$ $1.69 \pm 1.43 \text{ mm}$
(Guo <i>et al.</i> , 2021)	SIFT+MLP	Radio	pied	$1.85 \pm 1.61 \text{ mm}$
(Suzani <i>et al.</i> , 2015)	MLP+Otsu	IRM	vertébres	$3.0 \pm 2.4\text{mm}$
(Olory Agomma <i>et al.</i> , 2018)	DNI+CNN+DBSCAN	Radio EOS	Fémurs et Tibias	$29 \pm 18 \text{ mm}$
(Criminisi <i>et al.</i> , 2013)	Forêt aléatoire	CT-scan IRM	Différents organes	$13.5 \pm 13\text{mm} 14.95 \pm 11.33\text{mm}$
(Aubert <i>et al.</i> , 2019)	SSM+CNN	Radio EOS	vertèbres	$1.9 \pm 1.33 \text{ mm}$
(Yi <i>et al.</i> , 2020)	CNN	Radio	vertèbres	50.11
(Payer <i>et al.</i> , 2016)	CNN	Radio	mains	$1.13 \pm 0.98 \text{ mm}$, 12 outliers $> 10 \text{ mm}$
(Zhang <i>et al.</i> , 2017)	CNN	IRM	cerveau Prostate	$2.86 \pm 1.55\text{mm} 3.34 \pm 2.55\text{mm}$
(Zheng <i>et al.</i> , 2015)	MLP	CT-scan	Tête-cou	$2.64 \pm 4.98\text{mm}$
(Liu <i>et al.</i> , 2020)	CNN	Radio	Hanche	2.14mm
(Mader <i>et al.</i> , 2018)	CNN+CRF	Radio	Thorax	4.7mm
(Yang <i>et al.</i> , 2015)	CNN	IRM	Femur distal	$4.69 \pm 2.30\text{mm}$
(Bier <i>et al.</i> , 2019)	CNN	Radio Img synthétique	Bassin	$18.9 \text{ mm} 5.6 \pm 4.5 \text{ mm}$
(Tsai, 2021)	CNN	Radio EOS	Fémurs et Tibias	$3.6 \pm 9.9 \text{ mm}$

2.4 Méthodes d’augmentation de données

La mise en place d’une méthode entièrement automatisée requiert généralement l’accès à un large ensemble de données. Principalement, avec les méthodes d’apprentissage machine telles que les réseaux de neurones, l’exploitation d’un petit ensemble de données pourrait entraîner un sur-apprentissage du réseau ou baisser sa performance (Lemley *et al.*, 2017; Shorten & Khoshgoftaar, 2019). Dans la littérature, pour résoudre le problème de manque de données, la méthode d’augmentation de données est la plus souvent utilisée (Lemley *et al.*, 2017). Mis à part l’avantage lié à l’augmentation de la taille de l’ensemble d’entraînement, la méthode d’augmentation de données présente divers autres avantages. En effet, elle permet d’éviter le sur-apprentissage du réseau neuronal durant la phase d’entraînement, augmente la capacité de

généralisation et la performance de prédiction du réseau et le rend plus robuste pour apprendre des caractéristiques qui sont spatialement invariantes (Shorten & Khoshgoftaar, 2019) (Lemley *et al.*, 2017). Les méthodes d'augmentation de données peuvent être catégorisées en deux groupes (Shorten & Khoshgoftaar, 2019). Nous avons d'un côté les méthodes d'augmentation de données qui transforment l'image dans l'espace de données (exemples : les transformations géométriques, les méthodes d'augmentation des données d'espace colorimétrique) et de l'autre côté les méthodes d'augmentation de données basées sur les approches d'apprentissage machine (exemples : méthode d'augmentation de données GAN) (Shorten & Khoshgoftaar, 2019) qui transforment l'image dans l'espace de caractéristiques (Shorten & Khoshgoftaar, 2019). La modification de l'image dans l'espace des caractéristiques entraîne la génération d'images présentant des textures différentes de celles des images sources. Une mise en correspondance entre les images générées et l'image source pourrait donner de faibles correspondances vu qu'elles ont des textures différentes ; ce qui ne convient pas à notre application. Différentes stratégies d'augmentation de données dans l'espace de données telles que la rotation (Kordon *et al.*, 2019), la translation (Chen *et al.*, 2017), la mise en échelle (Kwolek *et al.*, 2019), la réflexion (ou symétrie) (Lianghui *et al.*, 2019), les fenêtres glissantes, l'extraction d'imagettes (Ciresan, Giusti, Gambardella & Schmidhuber, 2012), les déformations élastiques (Chen *et al.*, 2017) ou le rognage (Mittal *et al.*, 2018) sont utilisées dans la littérature pour générer de nouvelles données d'entraînement. Généralement, les images constituant l'ensemble de données d'entraînement (dans le cadre de la segmentation sémantique ou de l'identification de points) contiennent soit des structures complètes comme dans (Kordon *et al.*, 2019 ; Kwolek *et al.*, 2019 ; Lianghui *et al.*, 2019 ; Ronneberger *et al.*, 2015) ou soit uniquement des structures partielles (c'est-à-dire des sous-régions de l'image) comme dans (Chen *et al.*, 2017 ; Ciresan *et al.*, 2012 ; Wang *et al.*, 2018a).

- Dans le premier cas, des transformations globales (transformations sur toute l'image) telles que les transformations géométriques (exemples : la translation, la rotation, la mise à l'échelle, la réflexion, la déformation élastique) (Kordon *et al.*, 2019 ; Kwolek *et al.*, 2019 ; Lianghui *et al.*, 2019 ; Ronneberger *et al.*, 2015), les manipulations d'intensité de pixel (exemples :

ajout de bruit, correction gamma, augmentation ou réduction d'intensité, manipulation de contraste) (Perez, Vasconcelos, Avila & Valle, 2018) sont généralement utilisées pour générer de nouvelles images ; mais lorsque la taille de l'image est grande, une perte de détails fins peut être observée, car un redimensionnement des images peut être nécessaire pour s'assurer que les images sont gérables par le réseau neuronal. Par exemple, Kordon *et al.* (2019) ont utilisé comme stratégies d'augmentation de données, la rotation, la mise à l'échelle, la réflexion horizontale pour générer de nouvelles images. Les nouvelles images générées sont ensuite complétées par des zéros aux dimensions carrées, puis redimensionnées à 256×256 . Dans (Lianghui *et al.*, 2019), la stratégie de réflexion gauche-droite est utilisée. Kwolek *et al.* (2019), quant à eux, ont utilisé la rotation, le décalage (shift) horizontal et vertical, la mise à l'échelle et la réflexion horizontale comme stratégies d'augmentation des données.

- Dans le second cas, seules des images contenant des structures partielles (sous-régions de l'image) sont exploitées par le réseau. Des transformations locales permettant l'extraction d'imagettes (exemples : le recadrage (crop), les fenêtres glissantes) (Ciresan *et al.*, 2012) parfois combinées avec une transformation géométrique ou des manipulations d'intensité de pixel sont utilisées (Chen *et al.*, 2017). Par exemple, Ciresan *et al.* (2012) ont utilisé des fenêtres glissantes comme stratégie d'augmentation des données pour segmenter les membranes neuronales dans les images de microscopie électronique. Dans (Chen *et al.*, 2017), la translation, la rotation et la déformation élastique sont utilisées en premier pour générer de nouvelles images. Des imagettes de taille 480×480 extraient aléatoires à partir de ces nouvelles images générées sont par la suite utilisées comme entrée de leur réseau.

Afin d'avoir un ensemble de données d'images contenant à la fois des structures partielles et complètes, des transformations locales et globales peuvent être utilisées. Par exemple, Mittal *et al.* (2018) combinent la réflexion, la rotation et le recadrage aléatoire comme stratégies d'augmentation des données.

2.5 Synthèse

Le but dans ce travail est de pouvoir segmenter et identifier automatiquement des structures osseuses des membres inférieurs (fémurs et tibias) de même que localiser et identifier certains points d'intérêt des MI dans les radiographies EOS.

Les différentes méthodes d'extraction d'informations étudiées dans la littérature donnent des résultats prometteurs, mais ne sont pas toujours adaptées à notre cas d'application où l'on aimerait segmenter des os et localiser des points stéréo-correspondants sur des images présentant différents points de vue et orientations du patient. Les modèles déformables, utilisés pour la tâche de segmentation, sont rapides dans leur convergence, mais souffrent cependant d'un problème d'initialisation qui doit être faite proche de l'objet recherché (Medjeded & Mahmoudi, 2009). Ces méthodes sont souvent combinées avec des modèles de formes (exemple : modèle statistique, atlas) qui incorporent l'information *a priori* sur la forme de l'objet à segmenter afin d'avoir une segmentation robuste (Chen *et al.*, 2005 ; Linguraru *et al.*, 2010). Les méthodes basées sur les formes sont très utilisées dans la littérature et donnent des résultats prometteurs. Ces méthodes ne sont par contre pas capables d'estimer des variations de formes autres que celles présentes dans leur ensemble d'entraînement (Piazzese *et al.*, 2017). Dans le cas des images EOS, cela impliquerait de mettre en place un modèle de forme selon la région qu'on aimerait segmenter (exemples : MI, fémur proximal, tibia) ; ce qui n'est pas optimal dans notre cas étant donné la grande diversité d'images (différents points de vue et orientations) à traiter.

D'un autre côté, nous avons les méthodes basées sur les réseaux de neurones qui permettent de faire de la segmentation multi-classes et d'effectuer des tâches complexes. Nous avons abordé différentes méthodes permettant de faire de la segmentation de structures dans les images à partir de réseaux neuronaux. Les méthodes de segmentation d'instances (exemples : Mask R-CNN, YOLACT) sont intéressantes du fait qu'en plus de détecter les régions d'intérêt, elles les segmentent. La performance de ces méthodes dépend par contre de la performance de leurs méthodes de détection de ROI (Region of interest) car seules les régions détectées sont segmentées (Ruiz-Santaquiteria *et al.*, 2020).

D'un autre côté, nous avons les méthodes de segmentation sémantique avec des architectures en U du type encodeur-décodeur qui classifient chaque pixel dans l'image. Le réseau U-Net, un type de réseau en U, est largement utilisé dans la littérature et donne de bons résultats pour la tâche de segmentation d'objets dans l'image (Tableau 2.2). Une autre variante des réseaux en U est SegNet qui a une architecture similaire au U-Net. La principale différence entre SegNet et U-Net est que dans U-Net, l'intégralité des cartes des caractéristiques de la partie encodeur est transmise dans la partie décodeur, tandis que dans SegNet, seuls les indices de pooling sont transmis faisant en sorte qu'il utilise moins de mémoire. SegNet est donc plus rapide à entraîner et donne des résultats similaires à ceux de U-Net (Alam *et al.*, 2021 ; Hussain *et al.*, 2021 ; Oliveira & dos Santos, 2018). L'exploration de cette architecture dans cette thèse pourrait être intéressante.

Bien que les architectures du type encodeur-décodeur donnent les meilleurs résultats pour les tâches de segmentation sémantique, ils ne tirent pas toujours pleinement profit de l'information contextuelle présente dans l'image (Khanh, Dao, Ho, Yang, Baek, Lee, Kim & Yoo, 2020). Certains travaux, comme le travail de Shen *et al.* (2020) montrent de bons résultats de segmentation des fémurs et des tibias en améliorant l'information contextuelle dans le réseau U-Net. Améliorer le contextuel dans le réseau SegNet, pourrait donc améliorer sa performance de segmentation.

Nous avons également abordé différentes méthodes d'identification de points d'intérêt dans les images. Les méthodes de détection de points caractéristiques tels que les détecteurs de Harris, SIFT, etc, permettent de retrouver des points invariants dans les images avec un faible coût de calcul. Ces approches sont basées sur le calcul du gradient dans l'image, donc basées sur les valeurs d'intensité de pixels dans l'image. Une approche d'identification de points d'intérêt basée sur l'algorithme ASIFT et la transformée de Hough a d'ailleurs été explorée sur les radiographies EOS (voir ANNEXE I) et a révélé des erreurs assez grandes d'identification. Au terme de cette première expérimentation, la nécessité d'explorer une approche dépendant aussi de l'information a priori sur les structures ou les images et présentant une phase d'entraînement nous a paru intéressante.

Les méthodes d'apprentissage machine comme les réseaux de neurones ont une phase d'entraînement et sont très utilisées pour les problèmes d'identification de points d'intérêt dans les images. Ils sont par contre pour la plupart lents durant la phase d'entraînement, mais efficaces ou précis durant la phase de prédiction. Ils requièrent généralement d'avoir un grand ensemble de données d'images et de labélisation durant la phase d'entraînement ; ce qui n'est pas facile à avoir dans le domaine de l'imagerie médicale. Pour combler cette carence, les travaux dans la littérature (Lemley *et al.*, 2017 ; Lianghui *et al.*, 2019 ; Ronneberger *et al.*, 2015 ; Shorten & Khoshgoftaar, 2019) utilisent généralement des méthodes d'augmentation de données. Une approche (Olory Agomma *et al.*, 2018) combinant des méthodes de détection de points caractéristiques et d'apprentissage machine a été explorée sur les images EOS en utilisant l'extraction d'imagettes comme méthodes d'augmentation de données. Cette approche a révélé des erreurs d'identification de points d'intérêt assez grands, mais beaucoup moins que l'approche basée sur les détecteurs du type SIFT. Plusieurs travaux (Balntas *et al.*, 2016 ; Dosovitskiy *et al.*, 2016) ont prouvé la supériorité des réseaux de neurones par rapport aux méthodes de détection de points caractéristiques du type SIFT pour la tâche d'extraction de caractéristiques. Il serait peut-être optimal d'extraire les caractéristiques directement à partir d'un réseau neuronal plutôt que de passer par une première étape d'extraction de caractéristiques à partir d'approches de détections de points caractéristiques du type SIFT. Au terme de cette deuxième expérimentation, nous avons remarqué que les imagettes situées loin des points à prédire induisaient de grandes erreurs dans la prédiction peut-être parce qu'ils ne sont pas à proximité des régions à prédire ou parce que le contexte global entre le centre des imagettes et les points prédits n'étaient pas pris en compte par le réseau.

Certaines méthodes basées sur les réseaux de neurones comme OpenPose permettent de détecter des points et d'estimer la pose dans les images. Ces méthodes sont généralement rapides mais moins précises pour estimer des points d'intérêt ; ce qui les rend moins adaptées dans le cadre médical pour une évaluation de précision (Groos *et al.*, 2021). D'autres méthodes comme celle de Aubert *et al.* (2019), qui recèle, à l'aide d'un CNN, un modèle statistique de la vertèbre sur des paires d'images EOS dans le but de reconstruire en 3D la colonne vertébrale et par ailleurs

déterminer des points d'intérêt de la colonne vertébrale, ont donné des résultats intéressants. Cette méthode requiert par contre de mettre en place un modèle statistique de la structure d'intérêt sur laquelle l'on souhaite identifier les points. Dans notre cas cela reviendrait à mettre en place un modèle statistique selon la région (fémur ou tibia) considérée ; ce qui n'est pas optimal. D'autres méthodes comme celle de Tsai (2021) ou de Liu *et al.* (2020) utilisent une architecture neuronale pour identifier des points d'intérêt à partir d'une image en entrée. Dans notre cas, le réseau considéré prendrait soit une image de face ou de profil et prédirait l'emplacement 2D des points d'intérêt. Le réseau ne serait donc pas en mesure d'apprendre le lien existant entre les images de face et de profil pour prédire l'emplacement de points stéréo-correspondants appartenant aux deux images. Dans ce travail, on aimerait être en mesure de prédire des points uniques 3D visibles à la fois dans les images de face et de profil ; ce qui implique une détection simultanée des points dans les deux images pour une estimation 3D de leurs positions.

Une architecture neuronale très utilisée dans la littérature pour extraire des caractéristiques appartenant à la fois à plusieurs images et mesurer la similarité entre elles, est le réseau siamois. Plusieurs travaux (Rocco *et al.*, 2019 ; Yi *et al.*, 2016) ont prouvé la supériorité de ces méthodes par rapport aux méthodes du type SIFT pour les tâches d'extraction de caractéristiques communes à plusieurs images. C'est le cas de certains travaux (Han *et al.*, 2015 ; Rocco *et al.*, 2019 ; Yi *et al.*, 2016) qui extraient des caractéristiques communes à plusieurs images dans le but de faire de l'appariement d'images ou de trouver des similarités entre plusieurs images. Leur utilisation dans notre cas pourrait permettre d'apprendre des caractéristiques communes aux paires d'images EOS pour une prédiction optimale de l'emplacement 3D des points présents à la fois dans les paires d'images EOS.

Au vu de la revue effectuée et des différentes expérimentations menées, une piste à explorer dans le cadre de la segmentation des structures osseuses du MI serait l'utilisation des réseaux en U, en particulier le réseau SegNet du fait de leur performance démontrée en imagerie. Dans le cadre de l'identification des points d'intérêt 3D visibles à la fois dans les paires d'images EOS, les réseaux siamois pourraient être explorés à cet effet, du fait de leurs habilités à apprendre des caractéristiques communes à plusieurs images.

Dans le prochain chapitre, nous présenterons les objectifs de recherche, les choix méthodologiques de même que les contributions de ce travail.

CHAPITRE 3

OBJECTIFS ET CHOIX MÉTHODOLOGIQUES

3.1 Contexte et objectifs de recherche

Dans le cadre de cette thèse, nous travaillons en collaboration avec la compagnie EOS qui produit un logiciel commercial (SterEOS) de traitement d'images médicales. Les outils implantés dans SterEOS (Exemples : outil de mesures de paramètres cliniques 2D/3D, outil de reconstruction 3D) nécessitent l'intervention d'un humain pour numériser manuellement des points 2D/3D ou pour segmenter des structures osseuses. Ce processus manuel d'initialisation est chronophage (10 min en moyenne pour l'outil de reconstruction 3D des MI) ce qui accroît considérablement le délai d'exécution de ces outils. Réduire ce délai à l'aide d'une méthode permettant l'initialisation automatique de ces outils pourrait aider les utilisateurs.

Dans le cadre de cette thèse, nous avons accès à des images variables : images de corps entier, de membre inférieur et images contenant des structures partielles en orientation $0^\circ/90^\circ$, $20^\circ/70^\circ$ et $45^\circ/45^\circ$. La mise en place d'un processus automatique, capable de segmenter les os des MI, de localiser et d'identifier des points d'intérêt sur ces différentes images, serait une plus-value et permettrait d'initialiser les outils implantés dans SterEOS.

Comme mentionné précédemment, les méthodes basées sur les réseaux de neurones donnent les meilleurs résultats pour les tâches de segmentation d'os et d'identification de points dans les images. Elles pourraient donc être utiles dans le présent cas de figure. Mettre en place un processus automatique à l'aide de réseaux neuronaux requiert l'accès à un large ensemble de données représentatives. En imagerie médicale, il est souvent difficile d'avoir accès à un large ensemble de données. Particulièrement, dans cette thèse, nous avons accès à un ensemble de données limité d'images EOS. La mise en place d'une méthode d'augmentation de données sera explorée afin d'augmenter l'ensemble de données d'entraînement et de renforcer la capacité décisionnelle du réseau.

3.1.1 Objectifs

L'**objectif général** de cette thèse est de développer une méthode basée sur les réseaux de neurones capable de segmenter les structures osseuses des membres inférieurs (fémurs et tibias), de localiser et d'identifier des points d'intérêt des MI dans les images EOS du corps entier et de membres inférieurs avec le patient ayant les pieds décalés ou joints en orientation 0°, 45° et 90°. Cette méthode servira à initialiser les processus (mesures de paramètres cliniques 2D/3D, reconstruction 3D) implantés dans SterEOS.

3.1.1.1 Objectifs spécifiques

Objectif 1 : Développer une méthode entièrement automatique qui identifie et segmente les structures osseuses des membres inférieurs et qui classe des zones de superpositions entre ces structures osseuses dans les clichés radiographiques EOS.

Contrainte : La méthode développée devra fonctionner sur les images EOS présentant différents points de vue (images des membres inférieurs, images du corps entier, images contenant une partie de structure osseuse) et diverses orientations du patient (patient en orientation 0°, 20°, 45°, 70° et 90°).

Objectif 2 : Développer une méthode entièrement automatique basée sur les réseaux de neurones et capable de localiser et d'identifier des points d'intérêt stéréo-correspondants détectés simultanément dans les paires d'images contenant chaque structure osseuse des MI pour une estimation en 3D de leurs positions.

Contrainte : La méthode développée devra fonctionner sur les images EOS présentant différents points de vue (images des membres inférieurs et du corps entier) et orientations du patient (patient en orientation 0°/90° et 45°/45°). Dans la plateforme logicielle SterEOS, il est nécessaire d'extraire les points caractéristiques du MI sur des clichés radiographiques avec l'ensemble du MI pour estimer des paramètres cliniques sur l'ensemble du MI. L'objectif 2 répond à ce

besoin en proposant une approche qui permet d'initialiser l'ensemble de ces points sur le MI au complet. Les images à structures partielles ne seront donc pas considérées pour cet objectif.

3.2 Choix méthodologiques

Dans notre revue, nous avons vu que les méthodes basées sur les réseaux entièrement convolutionnels en forme de U tel que U-Net et SegNet donnent les meilleurs résultats pour les tâches de segmentation de structures (Tableau 2.2). Au terme de la revue, nous avons conclu qu'une piste intéressante à explorer dans le cadre de la segmentation des os (fémurs et tibias) du MI serait l'exploitation du réseau SegNet du fait qu'il est plus rapide à entraîner que U-Net tout en donnant des résultats similaires (Alam *et al.*, 2021 ; Hussain *et al.*, 2021 ; Oliveira & dos Santos, 2018). Néanmoins, les réseaux en U ne tirent pas toujours profit de l'information contextuelle présente dans l'image (Khanh *et al.*, 2020) ; ce qui pourrait engendrer des erreurs de segmentation (Cheng *et al.*, 2017). De la même façon que l'ont proposé Shen *et al.* (2020) en intégrant des convolutions dilatées dans le U-Net pour mieux considérer l'information contextuelle, une modification de l'architecture neuronale SegNet, de sorte qu'il tire pleinement avantage de l'information contextuelle permettrait donc d'éviter des erreurs de segmentation et d'obtenir un meilleur résultat de segmentation. Ce choix méthodologique répond à l'objectif 1.

Concernant la mise en place de la méthode d'identification de points d'intérêt, nous avons vu que les réseaux siamois étaient une piste intéressante à explorer à cause de leur capacité à extraire des caractéristiques communes à différentes images pour une prédiction optimale. Dans notre cas, le réseau siamois apprendrait des caractéristiques présentes à la fois sur les paires d'images EOS pour prédire la position tridimensionnelle des points d'intérêt appartenant à la fois aux images de face et de profil. Ce choix méthodologique répond à l'objectif 2.

Une problématique générale des méthodes d'apprentissage machine est leur besoin d'accès à un large ensemble de données pour une prise de décision optimale. Étant donné la taille limitée de notre ensemble de données et notre objectif de mettre en place une méthode fonctionnant sur des images présentant différents points de vue et orientations, l'utilisation d'une stratégie

d'augmentation de données qui génère des images contenant des structures partielles et complètes permettrait d'apprendre un grand nombre de configurations d'images pour permettre au réseau de mieux généraliser à tout type d'image. De plus, cette stratégie d'augmentation de données pourrait également permettre d'augmenter la précision de segmentation des os ou de localisation des points du réseau. Ce choix méthodologique répond aux objectifs 1 et 2.

Dans notre application, l'information de segmentation est importante. Une combinaison des approches de segmentation et d'identification de points d'intérêt semble donc plus logique. Dans le cadre de cette thèse, les régions contenant les os des MI (une région contient un os du MI) pourraient donc être extraites dans un premier temps sur la base de segmentations réalisées et les caractéristiques extraites de ces régions pourraient servir à prédire l'emplacement des points d'intérêt (contenus dans ces régions).

CHAPITRE 4

APPROCHE MÉTHODOLOGIQUE GLOBALE

L'objectif de cette thèse est de développer une méthode automatique d'identification (étiquetage : gauche et droite) et de segmentation du fémur et du tibia, de localisation et d'identification de points d'intérêt 3D stéréo-correspondants des MI dans les radiographies biplans EOS. Différentes expérimentations menées à cet effet nous ont permis de mettre en place une approche de résolution novatrice pour notre problématique. Une vue globale de l'approche méthodologique adoptée est présentée à la Figure 4.1. L'approche méthodologique adoptée est constituée de deux phases (Figure 4.1). Durant la première phase (voir phase 1 de la Figure 4.1), les fémurs et les tibias sont automatiquement identifiés et segmentés dans chaque vue EOS (considérée de manière indépendante) grâce à un nouveau réseau entièrement convolutionnel appelé RobustNet. Durant la deuxième phase (voir phase 2 de la Figure 4.1), des boîtes englobantes sont estimées autour de chaque structure osseuse segmentée sur les images de face et de profil pour définir des ROIs dans chacune des images. Un réseau siamois, modifié pour faire de la régression, est ensuite utilisé pour prédire la position 3D des points d'intérêt présents sur chacune des ROIs extraits. Il est à noter que les paires d'images (face et profil) à l'entrée du réseau siamois sont de profondeur 2. Le premier canal contient l'image ROI contenant la structure osseuse extraite de l'image EOS et le deuxième canal contient son masque. Chacune des deux phases de l'approche méthodologique proposée sera présentée dans des sections distinctes. Dans la section 4.1 une présentation de l'ensemble de données et des vérités terrain utilisées dans le cadre de ce travail sera faite. La section 4.2 sera consacrée à la présentation de la méthode d'identification et de segmentation des structures osseuses dans les radiographies EOS de face et de profil tandis que la section 4.3 sera consacrée à la présentation de la méthode de localisation et d'identification de points d'intérêt stéréo-correspondants des MI dans les radiographies EOS de face et de profil.

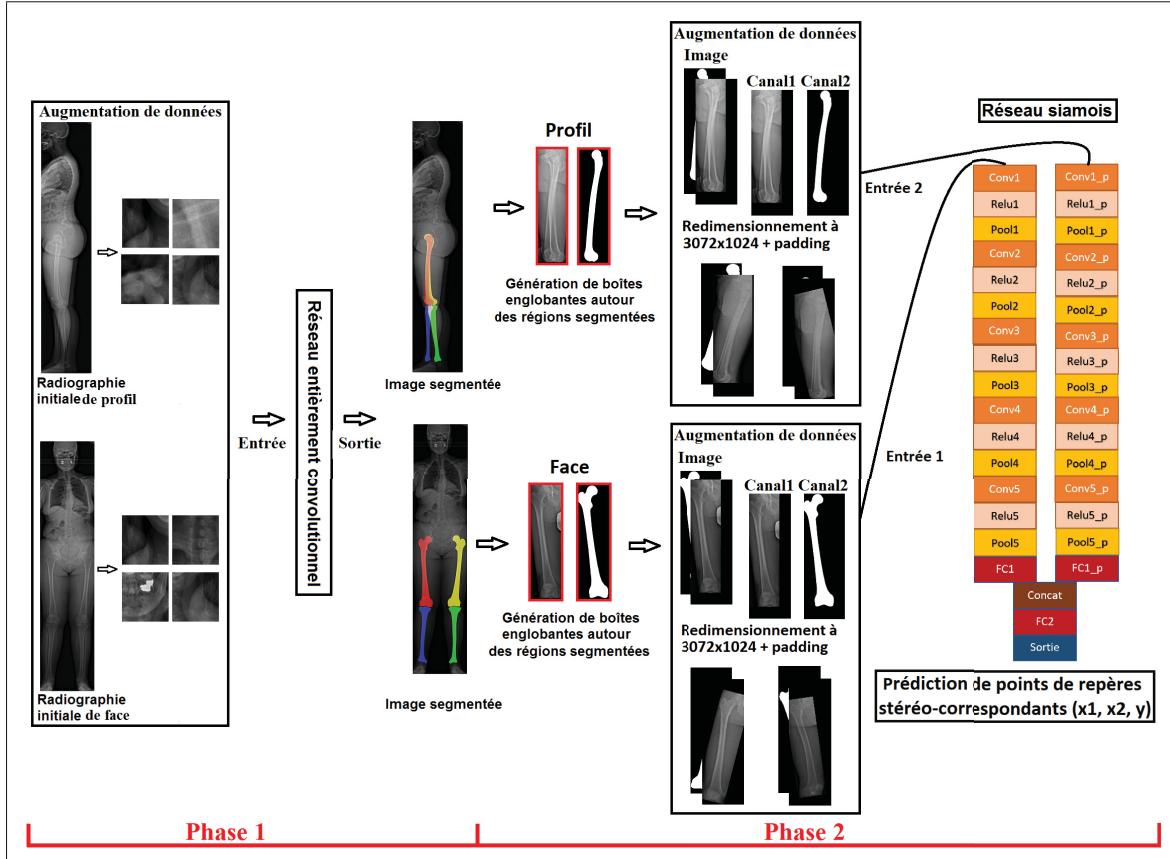


Figure 4.1 Processus d'entraînement de la méthode proposée. Phase 1 : Processus d'entraînement de la méthode d'identification et de segmentation des os proposée. Phase 2 : Processus d'entraînement de la méthode de localisation et d'identification de points 3D proposée

4.1 Ensemble de données et définition de la vérité terrain

4.1.1 Ensemble de données d'images

4.1.1.1 Ensemble de données pour la méthode d'identification et de segmentation des os des membres inférieurs

Les images utilisées dans notre ensemble de données présentent différents points de vue (Figure 4.2) : images des membres inférieurs (images ne contenant que les structures des membres inférieurs), images du corps entier (images contenant toutes les structures corporelles de la tête

aux pieds) et images contenant des structures partielles (images où l'os visé est partiellement visible). Ces images, acquises avec le patient en position debout (pieds joints ou décalés) présentent différentes orientations du patient. Nous pouvons distinguer : les images frontales en orientation 0° (Figure 4.2 (b) et (e)), 20° (Figure 4.2 (f)), 45° (Figure 4.2 (a)) et les images latérales en orientation 45° (Figure 4.2 (d)), 70° (Figure 4.2 (g)), 90° (Figure 4.2 (c)).

Notre ensemble de données est constitué de 220 radiographies EOS de face et de profil (110 de face et 110 de profil) en orientations 0° , 20° , 45° , 70° et 90° . Cet ensemble de données (avec les images de face et de profil regroupées de manière équivalente) a été subdivisé en trois sous-ensembles dans les proportions : 120 images en orientation 0° , 45° et 90° pour la phase d'entraînement, 30 images en orientation 0° , 45° et 90° pour la phase de validation et 70 images en orientation 0° , 20° , 45° , 70° et 90° pour la phase de test. Les 70 images de test sont constituées de 40 images entières EOS (images du corps entier et images des membres inférieurs) et de 30 images partielles contenant une portion de l'image entière. Les images en orientation 20° et 70° n'ont pas été incluses dans les ensembles d'entraînement et de validation car nous ne disposons que d'une seule paire.

Il est à noter que l'espacement des pixels (de formes carrées) dans les radiographies EOS est de 0,18 mm.

4.1.1.2 Ensemble de données pour la méthode d'identification de points d'intérêt

Les ensembles de données d'entraînement et de validation sont les mêmes que ceux utilisés dans la section 4.1.1.1, excepté l'ensemble de données de test qui ne contient plus d'images partielles. Nous avons utilisé un ensemble de données constitué de 60 paires d'images EOS (44 MI et 16 corps entier) pour l'entraînement, 15 paires d'images EOS pour la validation (10 MI et 5 corps entier) et 15 paires d'images EOS pour le test (12 MI et 3 corps entier). Des boîtes englobantes contenant les structures osseuses d'intérêt (fémur gauche, fémur droit, tibia gauche et tibia droit) et celles contenant les masques associés à ces structures sont extraites de ces paires d'images (le processus d'extraction des boîtes est présenté plus loin à la section 4.3.1.1). Un apprentissage a

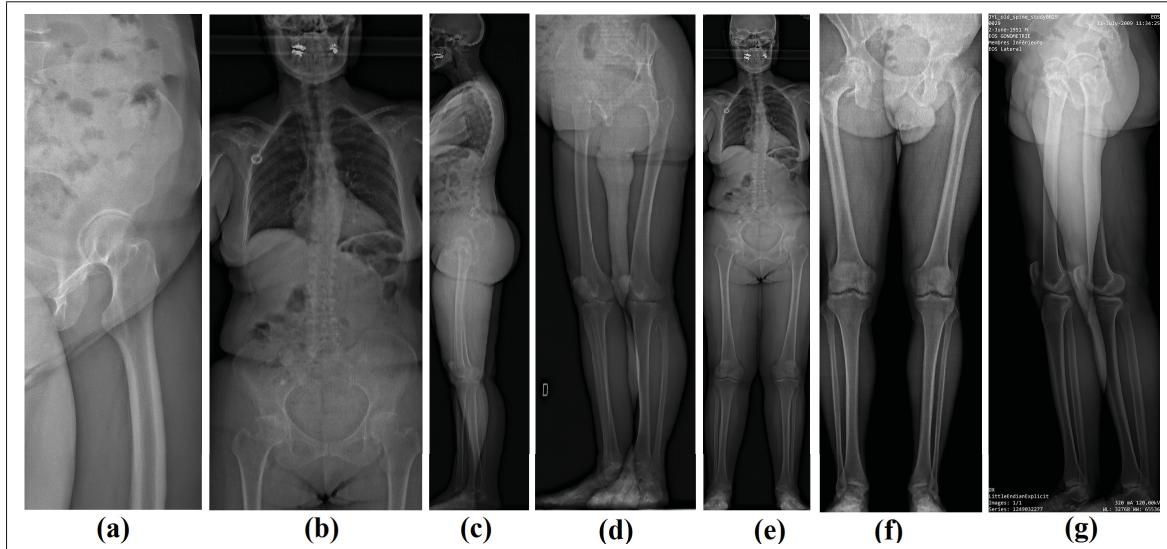


Figure 4.2 Images présentant différents points de vue et orientations. (a) Image contenant des structures partielles en orientation 45° . (b) Images contenant structures partielles en orientation 0° . (c) Images des MI en orientation 90° . (d) Images des MI en orientation 45° . (e) Images du corps entier en orientations 0° . (f) Image des MI en orientation 20° . (g) Image des MI en orientation 70°

été dédié pour le fémur (gauche et droit confondus) et un autre pour le tibia (gauche et droit confondus). L'ensemble de données pour le fémur est composé de :

- 60 paires (face et profil) de boîtes englobantes contenant le fémur gauche et 60 paires de boîtes englobantes contenant le fémur droit pour l'ensemble de données d'entraînement ;
- 15 paires (face et profil) de boîtes englobantes contenant le fémur gauche et 15 paires de boîtes englobantes contenant le fémur droit pour l'ensemble de données de validation ;
- 15 paires (face et profil) de boîtes englobantes contenant le fémur gauche et 15 paires de boîtes englobantes contenant le fémur droit pour l'ensemble de données de test.

L'ensemble de données pour le tibia est structuré pareillement à celle du fémur.

4.1.2 Définition de la vérité terrain

Pour nos méthodes de segmentation de structures osseuses et d'identification de points 3D, nous optons pour une vérité terrain construite à base de reconstructions 3D existantes faite à partir de la méthode semi-automatique de reconstruction 3D de Chaibi *et al.* (2012) de fémurs et de tibias.

4.1.2.1 Définition de la vérité terrain de la méthode d'identification et de segmentation des os du MI

Pour générer nos masques de vérité terrain (Figure 4.3(c)), nous projetons chacun des quatre modèles 3D osseux (fémur gauche, fémur droit, tibia gauche et tibia droit) précédemment reconstruit par l'expert suivant la méthode de Chaibi (2010), sur des radiographies EOS (face et profil) associées à la reconstruction 3D (Figure 4.3(b)). À chacune des régions délimitées par les contours obtenus après projection, est assignée une classe unique représentant une intensité de pixel unique pour former un masque étiqueté (Figure 4.3(c)). Nous pouvons distinguer jusqu'à sept classes à savoir : fémur gauche (FemG), fémur droit (FemD), tibia gauche (TibG), tibia droit (TibD), superposition entre fémur gauche et fémur droit (SupFem), superposition entre le tibia gauche et le tibia droit (SupTib) et l'arrière-plan (BackG) représentant toutes structures autres que les fémurs et les tibias. Le Tableau 4.1 ci-dessous présente les couleurs utilisées pour une représentation visuelle de chaque classe du membre inférieur de même que le pourcentage moyen de pixel par classe N_p . Nous pouvons remarquer que la classe correspondant à l'arrière-plan a plus de pixels que les quatre autres classes correspondant au fémur gauche, fémur droit, tibia gauche, tibia droit. Ces quatre classes ont à leur tour plus de pixels que les classes créées par la superposition des structures osseuses (superposition entre le fémur gauche et le fémur droit, superposition entre le tibia gauche et le tibia droit).

Tableau 4.1 Couleur et pourcentage moyen de pixels associés à chaque classe

Classes	BackG	FemG	FemD	TibG	TibD	SupFem	SupTib
Couleurs	Noir	Bleu	Jaune	Rouge	Vert	Orange	Blanc
N_p (%)	86.33	3.81	3.60	2.90	2.75	0.51	0.11

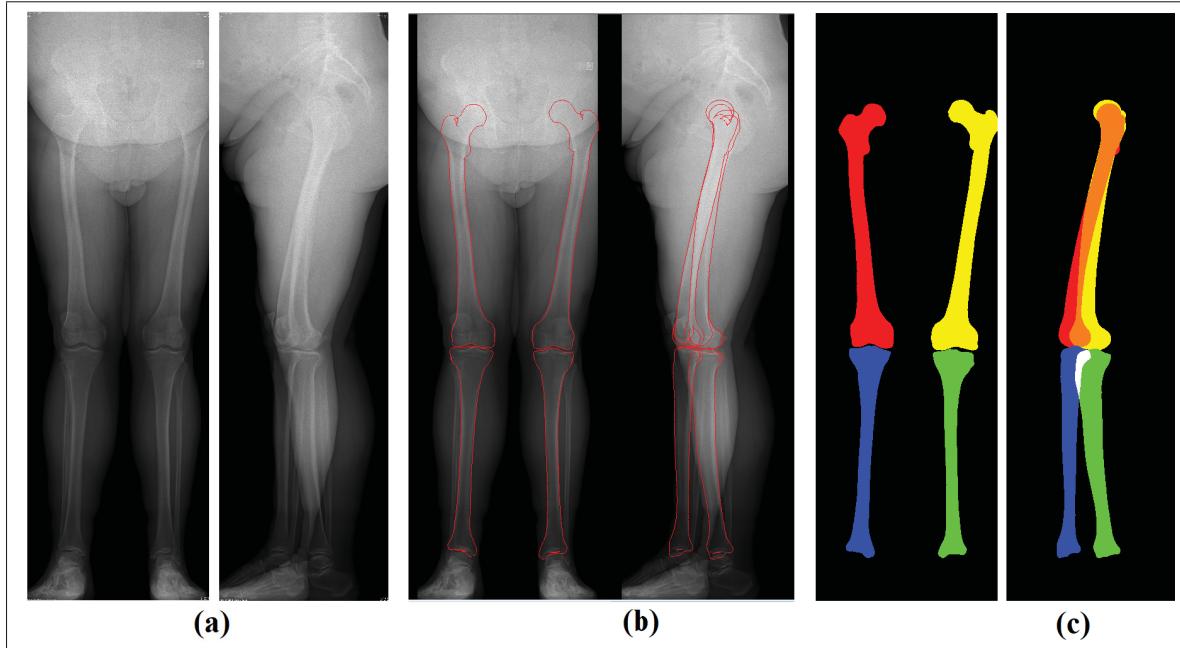


Figure 4.3 Radiographie EOS, résultat de projection des modèles 3D du MI et masques de vérité terrain associés.(a) Radiographie EOS. (b) Résultat de projection des modèles 3D du MI Tirée de SterEOS. (c) Masques associés aux radiographies EOS

4.1.2.2 Définition de la vérité terrain de la méthode d’identification de points d’intérêt

Les données de vérité terrain ont été générées en rétroprojetant des points 3D utilisés pour effectuer la reconstruction 3D et des mesures cliniques sur les paires (frontales et latérales) de radiographies 2D (Figure 4.4). Le nombre total de points d’intérêt employés dans ce travail est de vingt-trois dont treize points pour le fémur et dix points pour le tibia. Le choix de ces points a été fait d’un commun accord avec la compagnie EOS du fait de leurs pertinences d’utilisation autant pour mesurer des paramètres cliniques que pour initialiser le processus de reconstruction 3D. La liste de ces points est présentée dans le Tableau 4.2. Les Figures 4.5 et 4.6 donnent une bonne visualisation de ces points.

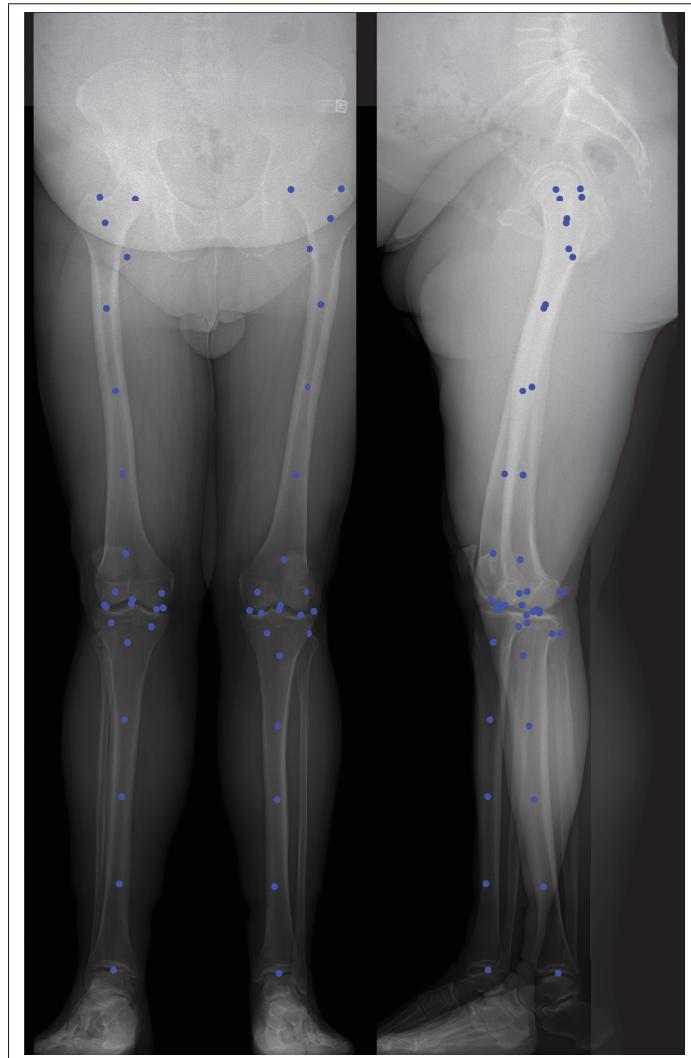


Figure 4.4 Résultat de projection des points 3D du MI sur de images EOS
Tirée de SterEOS

4.2 Méthode automatique pour l'identification et la segmentation de structures osseuses des membres inférieurs

Dans cette session, nous proposons une nouvelle solution basée sur les réseaux de neurones pour identifier et segmenter automatiquement les structures osseuses des membres inférieurs (fémurs et tibias) dans des radiographies EOS présentant différents points de vue (images de membres

Tableau 4.2 Liste des points d'intérêt et leurs numéros associés

Structure osseuse	Points d'intérêt	Numéro
Fémur	Centre de la tête fémorale	0
	Grand trochanter	1
	Petit trochanter	2
	Trochlée	3
	CondylePostExt	6
	CondylePostInt	7
	CondyleInfExt	8
	CondyleInfInt	9
	PointLDia	10
	SplineDia-1	4
	SplineDia-2	5
	SplineDia-3	11
	SplineDia-4	12
	PlateauPostInt	0
	PlateauPostExt	1
Tibia	PlateauInt	2
	PlateauExt	3
	SplineDiaT	4
	SplineDiaT-1	5
	SplineDiaT-2	6
	SplineDiaT-3	7
	SplineDiaT-4	8
	CentreEpines	9

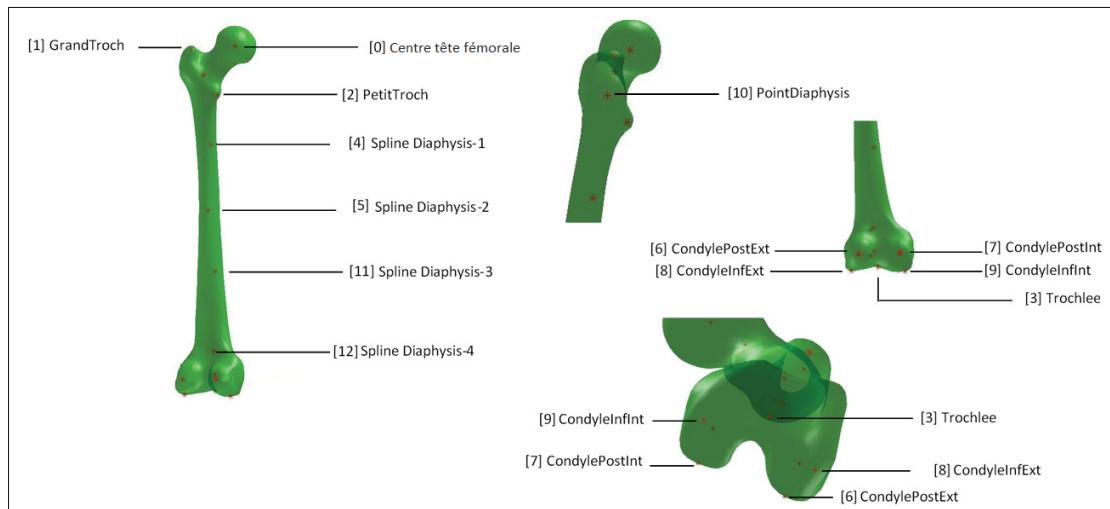


Figure 4.5 Position des points d'intérêt du fémur. Reproduite et adaptée avec l'autorisation de LIO (2017)

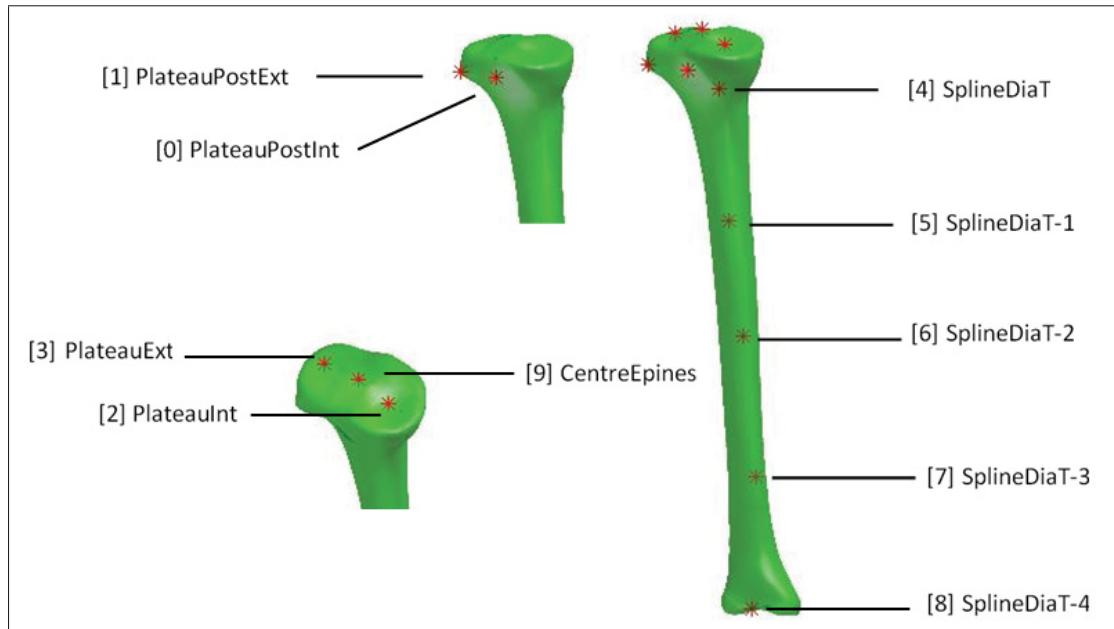


Figure 4.6 Position des points d'intérêt du tibia. Reproduite et adaptée avec l'autorisation de LIO (2017)

inférieurs, images corps entier et images avec structures partielles), différentes orientations du patient (0° , 20° , 45° et 70°) et une forte superposition de structures.

4.2.1 Détails méthodologiques

L'approche méthodologique (Figure 4.7) adoptée pour la résolution du problème d'identification et de segmentation des structures osseuses dans les radiographies EOS peut être subdivisée en deux phases, à savoir : la phase d'augmentation de données et la phase d'entraînement. L'objectif de la phase d'augmentation de données est de pallier au problème de manque de données en générant des images diversifiées présentant différents points de vue et orientations. Durant la phase d'augmentation de données, différentes transformations géométriques sont appliquées à un ensemble de données de départ pour produire un plus gros ensemble de données d'images contenant des structures partielles et complètes. Pendant la phase d'entraînement, chaque image de ce nouvel ensemble de données et sa vérité terrain correspondante sont transmises à un réseau

neuronal afin qu'il soit en mesure de produire en sortie un masque où les classes présentes dans l'image sont identifiées.

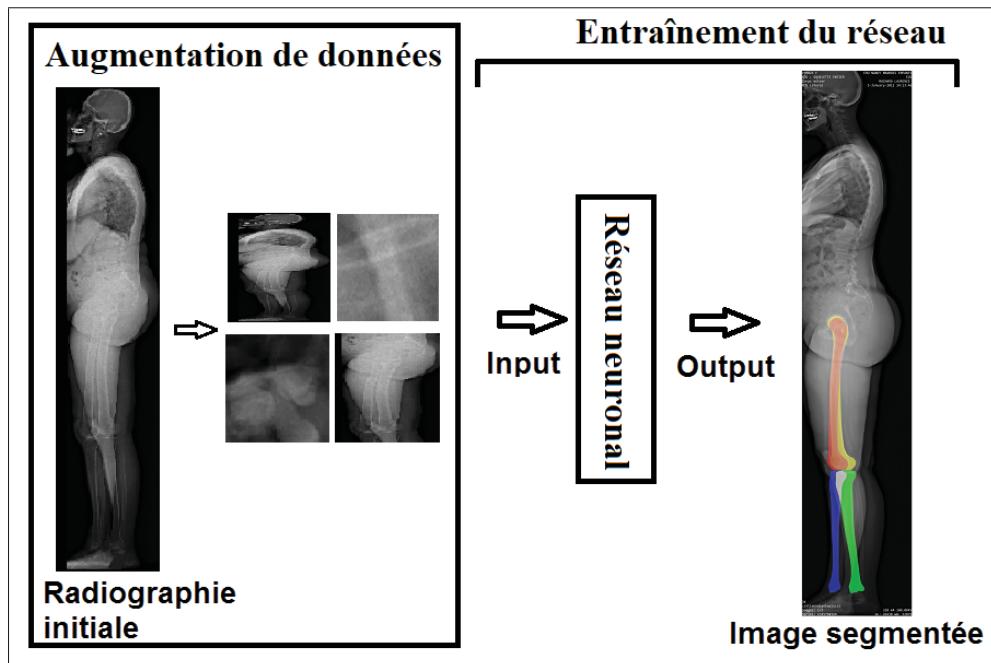


Figure 4.7 Processus d'entraînement de la méthode d'identification et de segmentation des os des MI

4.2.1.1 Phase 1 : Méthode d'augmentation de données

Quatre stratégies d'augmentation de données ont été appliquées dans ce travail pour agrandir notre base de données d'entraînement en générant des images contenant des structures complètes et partielles. Pour les différentes stratégies d'augmentation de données présentées ci-dessous, les images générées de tailles différentes de 360×480 pixels (taille par défaut proposée pour Segnet) sont redimensionnées à 360×480 par interpolation du plus proche voisin. Les images générées sont enregistrées sur 8 bits. Soient H et W respectivement la hauteur et la largeur de l'image source (image EOS).

- **Mise à l'échelle non uniforme :** Inspirée par la stratégie de déformation élastique (Ronneberger *et al.*, 2015)(Chen *et al.*, 2017), l'idée derrière l'utilisation de cette transformation

globale est de préserver la disposition des structures dans l'image tout en introduisant des déformations. L'introduction de déformations dans les réseaux neuronaux augmente leur capacité à déduire l'invariance aux transformations et les rend plus robustes. Dans notre cas, les déformations sont introduites dans les images en appliquant une mise à l'échelle non uniforme avec un facteur de redimensionnement $s \in [0.5, 1]$.

- **Recadrage** : Cette stratégie, utilisée dans la littérature (Chen *et al.*, 2017) (Mittal *et al.*, 2018), consiste à supprimer une partie de l'image. L'idée derrière l'utilisation de cette transformation locale est de générer des images contenant des structures partielles. Le recadrage est effectué en fixant une dimension et en recadrant sur la deuxième dimension et vice-versa. Dans le cas où la dimension fixée est W , l'image est recadrée suivant H de $H/2$ à H avec un pas de $0.025 * H$.
- **Fenêtre glissante** : Cette technique est souvent utilisée dans la littérature (Ciresan *et al.*, 2012) et consiste à faire glisser une fenêtre de taille fixe sur l'image afin d'extraire les imagettes. Cette transformation locale est utilisée afin de générer des images contenant des structures partielles. La différence de cette stratégie par rapport à la précédente (recadrage) est que les régions générées sont plus petites. Deux approches sont ici utilisées. Pour la première approche, l'image originale est parcourue par une fenêtre de taille, la moitié de l'image originale avec un pas de $W/10$ et $H/10$. Pour la deuxième approche, l'image d'origine est d'abord redimensionnée à $(360 \times 2; W/(H/(360 \times 2)))$ puis traversée par une fenêtre de taille 360×480 avec un pas de $360/2$ pour la dimension verticale et un pas de $480/2$ pour la dimension horizontale.
- **Rotation** : Cette technique d'augmentation de données consiste à faire pivoter l'image aléatoirement d'un angle compris entre -35° et 35° , chacune des images générées à partir des trois techniques d'augmentation de données ci-dessus présentées. Ce choix d'angle est fait pour représenter les inclinaisons existant au niveau du fémur ou du tibia chez certains patients. Cette transformation globale (la rotation) est utilisée pour permettre une invariance à la rotation par le réseau.

Cette stratégie d'augmentation de données, qui combine des transformations locales ou globales sur des images comme il est d'usage dans des cas similaires (Mittal *et al.*, 2018) permet d'obtenir un ensemble de données très diversifié, constitué d'une variété d'images contenant des structures partielles et complètes dans différentes orientations et champs de vue.

4.2.1.2 Phase 2 : Architecture neuronale proposée : RobustNet

Le problème à résoudre ici est un problème de classification où l'on cherche à classifier chaque pixel d'une image donnée. L'architecture neuronale proposée à cet effet (nommée RobustNet) est basée sur l'architecture encodeur-décodeur SegNet (Badrinarayanan *et al.*, 2017) conçue originellement pour segmenter des données de scène de routes. Nous avons dans un premier temps exploité cette architecture (non modifiée) pour segmenter les structures osseuses des membres inférieurs sur des radiographies EOS (Olory Agomma *et al.*, 2019) ; ce qui a donné globalement de bons résultats (coefficient de Dice moyen de $90.54 \pm 7\%$), mais a cependant révélé certaines erreurs de segmentation (Figure 4.8 (a)) où l'image segmentée détecte deux fois la même structure, dans ce cas le fémur. Cela est probablement dû au fait que l'architecture neuronale utilisée n'exploite pas suffisamment d'information contextuelle (Zhao *et al.*, 2018a) car comme nous le verrons plus tard dans le Tableau 4.3, le champ réceptif dans la partie encodeur de SegNet couvre environ 59% de la hauteur de l'image et environ 44% de sa largeur. Il pourrait donc ne pas couvrir entièrement une structure osseuse qui occupe environ la moitié de la largeur de l'image comme c'est le cas du fémur gauche dans la Figure 4.8 (a). Une possible solution serait d'augmenter ou d'améliorer l'information contextuelle présente dans le réseau de sorte que le champ réceptif couvre plus de la moitié de l'image dans chaque dimension permettant ainsi de couvrir au moins chaque structure d'intérêt dans l'image.

4.2.1.2.1 Description de l'architecture RobustNet et détails d'implémentation

Basé sur SegNet (Badrinarayanan *et al.*, 2017), RobustNet (Figure 4.9) est une architecture neuronale encodeur-décodeur développée pour identifier et segmenter automatiquement les os des membres inférieurs (fémur gauche, fémur droit, tibia gauche, tibia droit) à partir d'images

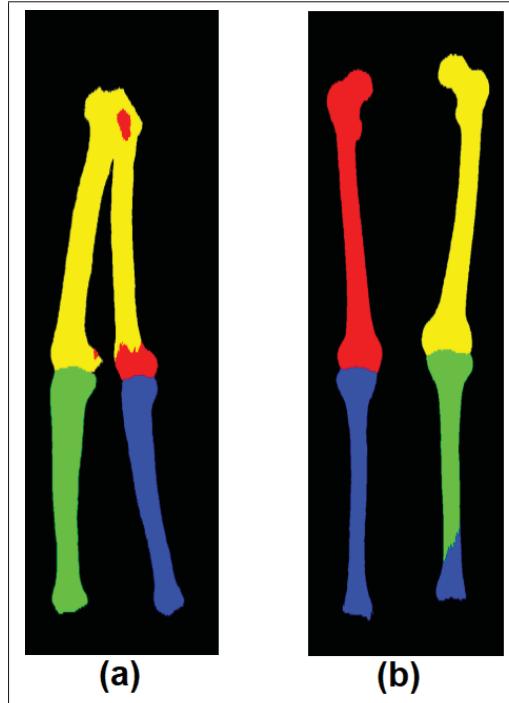


Figure 4.8 Exemples d'erreurs produites par SegNet (Image obtenue à partir de Olory Agomma *et al.* (2019))

présentant différents points de vue et orientations. L'objectif du développement de RobustNet est d'augmenter l'information contextuelle présente dans SegNet en augmentant le champ réceptif à l'intérieur et d'exploiter l'information locale et globale présente dans le réseau dans le but de réaliser une meilleure tâche de segmentation. Deux changements majeurs ont été apportés à SegNet (les modules ajoutés à SegNet sont encadrés par des lignes pointillées rouges sur la Figure 4.9) :

- Des convolutions dilatées avec un facteur de dilatation de 2 (2-convolutions dilatées) ont été intégrées dans la partie encodeur, notamment avant l'opération de pooling afin d'étendre de façon significative le champ réceptif dans le réseau. Nous avons appelé bloc contextuel le bloc constitué d'une couche de convolution dilatée suivie d'une couche de pooling (Figure 4.9).

- Les trois dernières cartes de caractéristiques à différentes échelles situées dans la partie décodeur ont été combinées (fusion de caractéristiques à différents niveaux) afin de générer des cartes de probabilité précises.

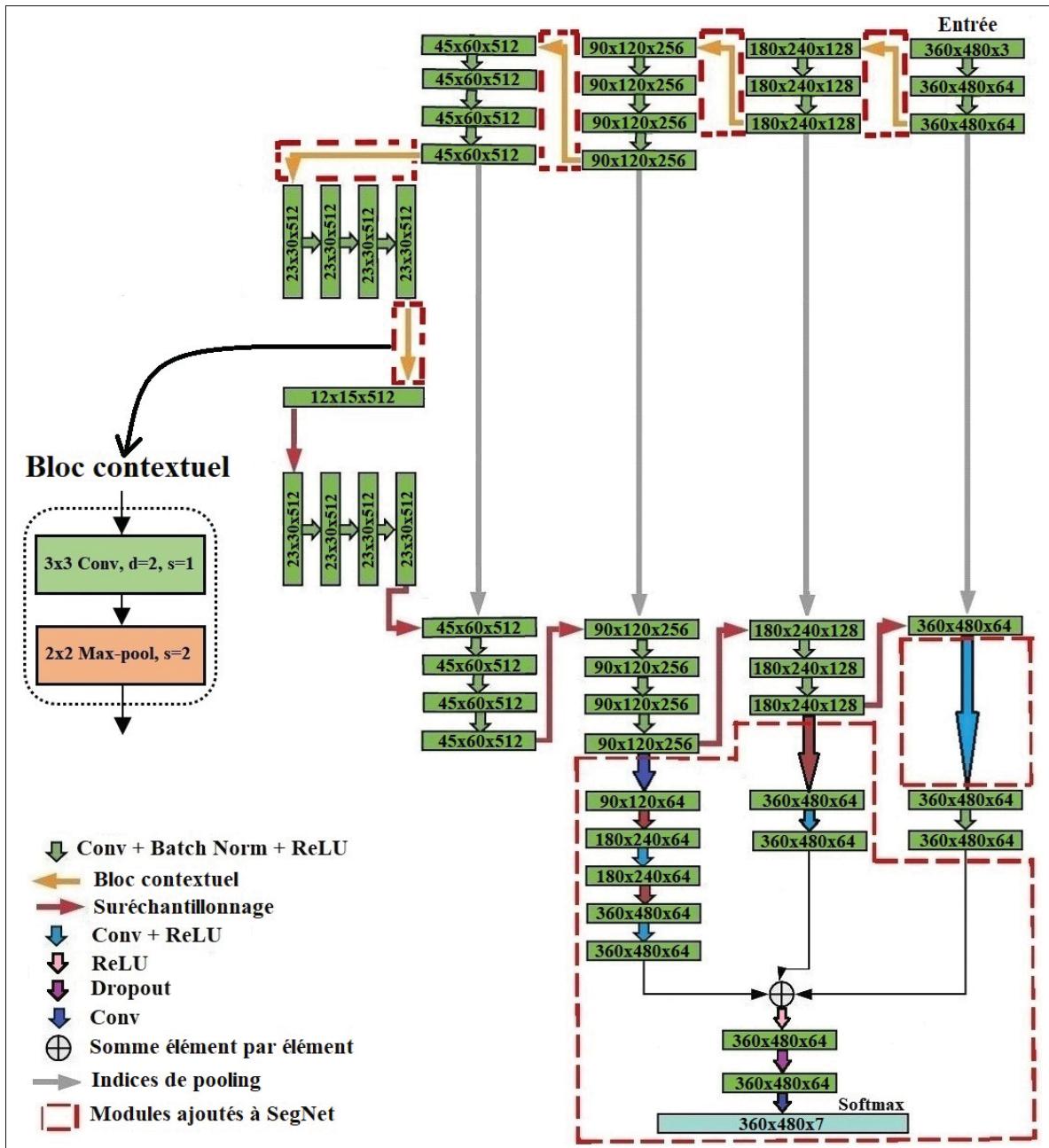


Figure 4.9 Architecture RobustNet

4.2.1.2.2 Bloc contextuel

Chaque couche de pooling présente dans l'architecture SegNet est remplacée par le bloc contextuel. Le bloc contextuel est constitué d'une convolution dilatée de 2 avec un noyau de 3×3 et un pas de 1 suivi d'un max-pooling avec un noyau de 2×2 et un pas de 2. La convolution dilatée est utilisée afin d'exploiter l'information contextuelle globale dans l'image. Un compromis a été fait sur le facteur de dilatation qui a été mis à 2 afin d'éviter que le noyau de convolution ne soit trop clairsemé, nous permettant ainsi de conserver également l'information locale (Wang *et al.*, 2018a). La couche de pooling a été placée après la couche de convolution dilatée dans le bloc contextuel, afin de préserver l'information locale pertinente tout en augmentant le champ réceptif (Haque & Neubert, 2020). Soit d , s et k respectivement le facteur de dilatation, le pas et la taille du noyau. Le champ réceptif de la couche l par rapport à la couche $l - m$ (avec $m \in \{1, \dots, l - 1\}$) est calculé comme suit (Yu & Koltun, 2016) :

$$RF\{l|l - m\} = (RF\{l|l - (m - 1)\} - 1) \times s + 1 + (k - 1) \times d \quad (4.1)$$

avec les paramètres $s = 1, k = 3, d = 1$ pour la convolution, $s = 1, k = 3, d = 2$ pour la convolution dilatée et $s = 2, k = 2, d = 1$ pour pooling. Le champ réceptif dans le bloc contextuel est représenté sur la Figure 4.10. Le Tableau 4.3 montre le champ réceptif présent dans chaque couche de la partie encodeur des architectures SegNet et RobustNet. Comme on peut le voir dans le Tableau 4.3, le champ réceptif dans SegNet est augmenté dans RobustNet d'environ 63%, démontrant qu'il a de fortes chances qu'au moins chaque structure osseuse d'intérêt dans l'image soit couverte par le champ réceptif. Ainsi, RobustNet préserve l'information générale de l'image tout en exploitant ses caractéristiques locales. L'augmentation du champ réceptif dans RobustNet a permis d'éviter des erreurs de segmentation observées avec SegNet. Cela sera présenté plus loin à la Figure 5.8.

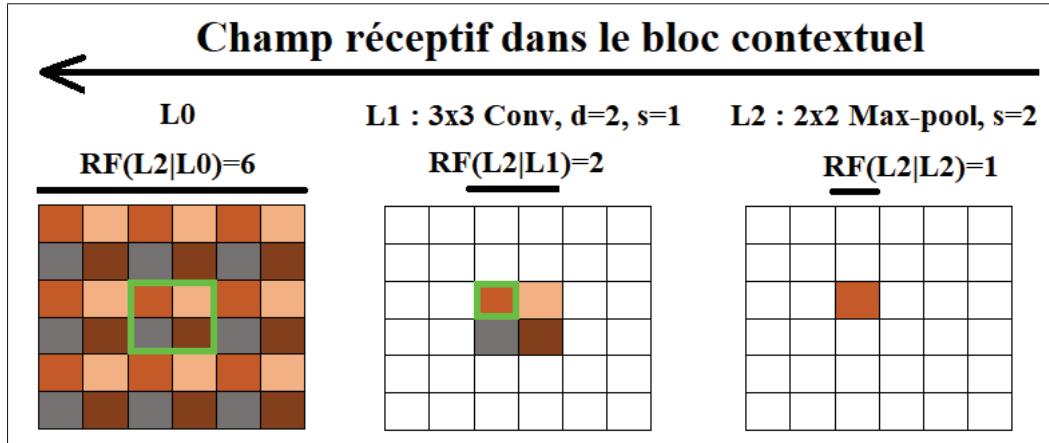


Figure 4.10 Visualisation du champ réceptif dans le bloc contextuel ($RF(L2|L0)$ signifie champs réceptif de la couche $L2$ par rapport à la couche $L0$)

4.2.1.2.3 Fusion de caractéristiques à différents niveaux

La fusion de cartes de caractéristiques à différents niveaux dans un réseau neuronal est une stratégie largement utilisée dans la littérature consistant à agréger des cartes de caractéristiques à différentes échelles dans le but de permettre au réseau neuronal d'exploiter à la fois les détails locaux (fins) et globaux (grossiers) et d'améliorer les performances de segmentation de l'image (Zhao *et al.*, 2017) (Chen *et al.*, 2017) (Long *et al.*, 2015) (Chen, Dou, Yu, Qin & Heng, 2018a). Quelques exemples de son utilisation peuvent être vus dans les architectures en forme de U comme U-Net (Ronneberger *et al.*, 2015) ou PDRU-NET (Shen *et al.*, 2020) où les cartes de caractéristiques de la partie encodeur sont intégrées dans leur correspondant dans la partie décodeur dans le but de combiner progressivement des caractéristiques à différentes échelles. D'autres exemples de son utilisation peuvent également être vus dans d'autres architectures (Chen *et al.*, 2017,1 ; Zhao *et al.*, 2017) où de nombreuses cartes de caractéristiques à différentes échelles sont fusionnées afin d'améliorer les performances du réseau. Cette stratégie a été utilisée dans RobustNet où ses trois dernières couches sont combinées par une sommation élément par élément afin d'intégrer des caractéristiques à différentes échelles.

RobustNet prend en entrée des images à trois canaux (contenant les mêmes informations en niveau de gris) de taille 360×480 et produit en sortie des masques de même taille pouvant

contenir un maximum de sept classes (un canal est assigné à chaque classe) à savoir : fémur gauche (FemG), fémur droit (FemD), tibia gauche (TibG), tibia droit (TibD), superposition entre fémur gauche et fémur droit (SupFem), superposition entre le tibia gauche et le tibia droit (SupTib) et l’arrière-plan (BackG). La fonction cross-entropie (Long *et al.*, 2015) est utilisée comme fonction objective durant l’entraînement. La fonction de perte cross-entropie pondérée basée sur l’équilibrage de fréquence médian (Eigen & Fergus, 2015) et la fonction de perte cross-entropie non pondérée ont été testées. Nous avons choisi d’utiliser la fonction cross-entropie non pondérée parce qu’elle nous a donné de meilleurs résultats expérimentalement.

Tableau 4.3 Champ réceptif de chaque couche de SegNet et de RobustNet dans la partie encodeur.

Caractéristiques de la couche	Champ réceptif SegNet	Champ réceptif RobustNet
Max-pool, k=2, s=2	$(1 - 1) \times 2 + 2 = 2$	$(1 - 1) \times 2 + 2 = 2$
Conv, d=2, k=3, s=1	N/A	$2 + (3 - 1) \times 2 = 6$
Conv, d=1, k=3, s=1	$(2 - 1) \times 1 + 3 = 4$	$(6 - 1) \times 1 + 3 = 8$
Conv, d=1, k=3, s=1	$(4 - 1) \times 1 + 3 = 6$	$(8 - 1) \times 1 + 3 = 10$
Conv, d=1, k=3, s=1	$(6 - 1) \times 1 + 3 = 8$	$(10 - 1) \times 1 + 3 = 12$
Max-pool, k=2, s=2	$(8 - 1) \times 2 + 2 = 16$	$(12 - 1) \times 2 + 2 = 24$
Conv, d=2, k=3, s=1	N/A	$24 + (3 - 1) \times 2 = 28$
Conv, d=1, k=3, s=1	$(16 - 1) \times 1 + 3 = 18$	$(28 - 1) \times 1 + 3 = 30$
Conv, d=1, k=3, s=1	$(18 - 1) \times 1 + 3 = 20$	$(30 - 1) \times 1 + 3 = 32$
Conv, d=1, k=3, s=1	$(20 - 1) \times 1 + 3 = 22$	$(32 - 1) \times 1 + 3 = 34$
Max-pool, k=2, s=2	$(22 - 1) \times 2 + 2 = 44$	$(34 - 1) \times 2 + 2 = 68$
Conv, d=2, k=3, s=1	N/A	$68 + (3 - 1) \times 2 = 72$
Conv, d=1, k=3, s=1	$(44 - 1) \times 1 + 3 = 46$	$(72 - 1) \times 1 + 3 = 74$
Conv, d=1, k=3, s=1	$(46 - 1) \times 1 + 3 = 48$	$(74 - 1) \times 1 + 3 = 76$
Conv, d=1, k=3, s=1	$(48 - 1) \times 1 + 3 = 50$	$(76 - 1) \times 1 + 3 = 78$
Max-pool, k=2, s=2	$(50 - 1) \times 2 + 2 = 100$	$(78 - 1) \times 2 + 2 = 156$
Conv, d=2, k=3, s=1	N/A	$156 + (3 - 1) \times 2 = 160$
Conv, d=1, k=3, s=1	$(100 - 1) \times 1 + 3 = 102$	$(160 - 1) \times 1 + 3 = 162$
Conv, d=1, k=3, s=1	$(102 - 1) \times 1 + 3 = 104$	$(162 - 1) \times 1 + 3 = 164$
Max-pool, k=2, s=2	$(104 - 1) \times 2 + 2 = 208$	$(164 - 1) \times 2 + 2 = 328$
Conv, d=2, k=3, s=1	N/A	$328 + (3 - 1) \times 2 = 332$
Conv, d=1, k=3, s=1	$(208 - 1) \times 1 + 3 = 210$	$(332 - 1) \times 1 + 3 = 334$
Conv, d=1, k=3, s=1	$(210 - 1) \times 1 + 3 = 212$	$(334 - 1) \times 1 + 3 = 336$

4.2.2 Détails d'implémentation

L'ensemble de données utilisé dans le cadre de ce travail est constitué de 220 images, soit 120 images d'entraînement, 30 images de validation et 70 images de test. La stratégie d'augmentation des données proposée à la section 4.2.1.1 a été utilisée sur nos données d'entraînement pour produire 132309 images d'entraînement. Durant la phase d'entraînement, les images et leurs masques correspondants sont envoyés au réseau pour qu'il apprenne le lien entre eux, tandis que pendant la phase de test, le réseau prend uniquement l'image en entrée et prédit en sortie son masque correspondant pouvant contenir jusqu'à sept classes. La librairie Caffe a été utilisée pour l'implémentation et l'entraînement du réseau qui est fait avec une carte graphique NVIDIA Titan RTX avec 24 Go de mémoire. Pour les deux architectures (RobustNet et SegNet), un gradient accéléré de Nesterov avec un momentum de 0.9 et un taux d'apprentissage fixe de 0.001 ont été utilisés. Nous avons utilisé un batch size de 6 pour la phase d'entraînement et le nombre maximal d'itérations est fixé à 350000. Les hyperparamètres utilisés ont été choisis expérimentalement.

4.3 Méthode automatique pour la localisation et l'identification de points d'intérêt des membres inférieurs

L'identification automatique des points d'intérêt stéréo-correspondants 3D des membres inférieurs dans les radiographies EOS de face et de profil est importante pour le calcul de paramètres cliniques et la reconstruction 3D des os des membres inférieurs. Identifier ces points d'intérêt stéréo-correspondants (qui permettent l'estimation d'un point 3D) manuellement sur une image radiographique biplan EOS est particulièrement difficile dans certaines régions (exemple : genou, en raison de la proximité des structures dans cette région) et coûteux en temps en raison de la complexité présente dans les images en termes de superposition des structures osseuses, de la grande variation d'intensité et de la grande variation anatomique présente dans les images radiographiques. Le but de ce chapitre est donc d'automatiser la tâche de localisation et d'identification de points d'intérêt stéréo-correspondants dans les radiographies EOS présentant différents points de vue et deux principales orientations afin d'initialiser les processus de calcul de paramètres cliniques ou pour la reconstruction 3D. Cette initialisation n'est pas définitive et

l'opérateur pourra la retoucher si les points ne sont pas assez précis. Ce premier positionnement automatique de ces points est cependant un immense avantage pour l'opérateur car la tâche d'identification manuelle est fastidieuse, chronophage et sujette à des erreurs.

La méthodologie adoptée exploite les résultats de segmentation obtenus dans la précédente méthode présentée dans la section 4.2 pour identifier des points d'intérêt stéréo-correspondants tridimensionnels dans les paires d'images EOS. Des boîtes englobantes contenant chaque structure osseuse du MI et leur masque sont utilisées pour régresser la position des points d'intérêt appartenant à cette structure de sorte que le contexte global régissant la position des points soit pris en compte durant l'entraînement.

Dans ce travail, nous nous concentrerons sur l'identification de points d'intérêt des membres inférieurs sur des radiographies 2D biplans du membre inférieur et du corps entier en orientation $0^\circ/90^\circ$ et $45^\circ/45^\circ$. Il est à noter qu'un tel travail n'a jamais été réalisé auparavant pour le membre inférieur sur cette variété d'images biplans.

4.3.1 Détails méthodologiques

L'approche méthodologique utilisée dans cette section peut être divisée en deux phases à savoir : la phase de génération de boîtes englobantes et la phase d'entraînement par l'intermédiaire d'un réseau siamois. La Figure 4.11 donne un aperçu global de l'approche méthodologique adoptée.

4.3.1.1 Phase 1 : Extraction de boîtes englobantes

Dans nos travaux précédents, les structures osseuses des membres inférieurs ont été segmentées sur des radiographies. Nous avons développé pour cela une méthode qui prend les images EOS en entrée et prédit des masques en sortie. Les masques prédits sont redimensionnés par interpolation bilinéaire à la taille de l'image réelle et superposés avec cette dernière pour délimiter chaque structure osseuse ciblée et ainsi l'encadrer d'un rectangle (boîte englobante) afin de l'extraire. Ainsi, sur la base des masques prédits, nous générerons des boîtes englobantes autour de chaque structure osseuse sur les images de face et de profil. Les boîtes de délimitation sont générées de

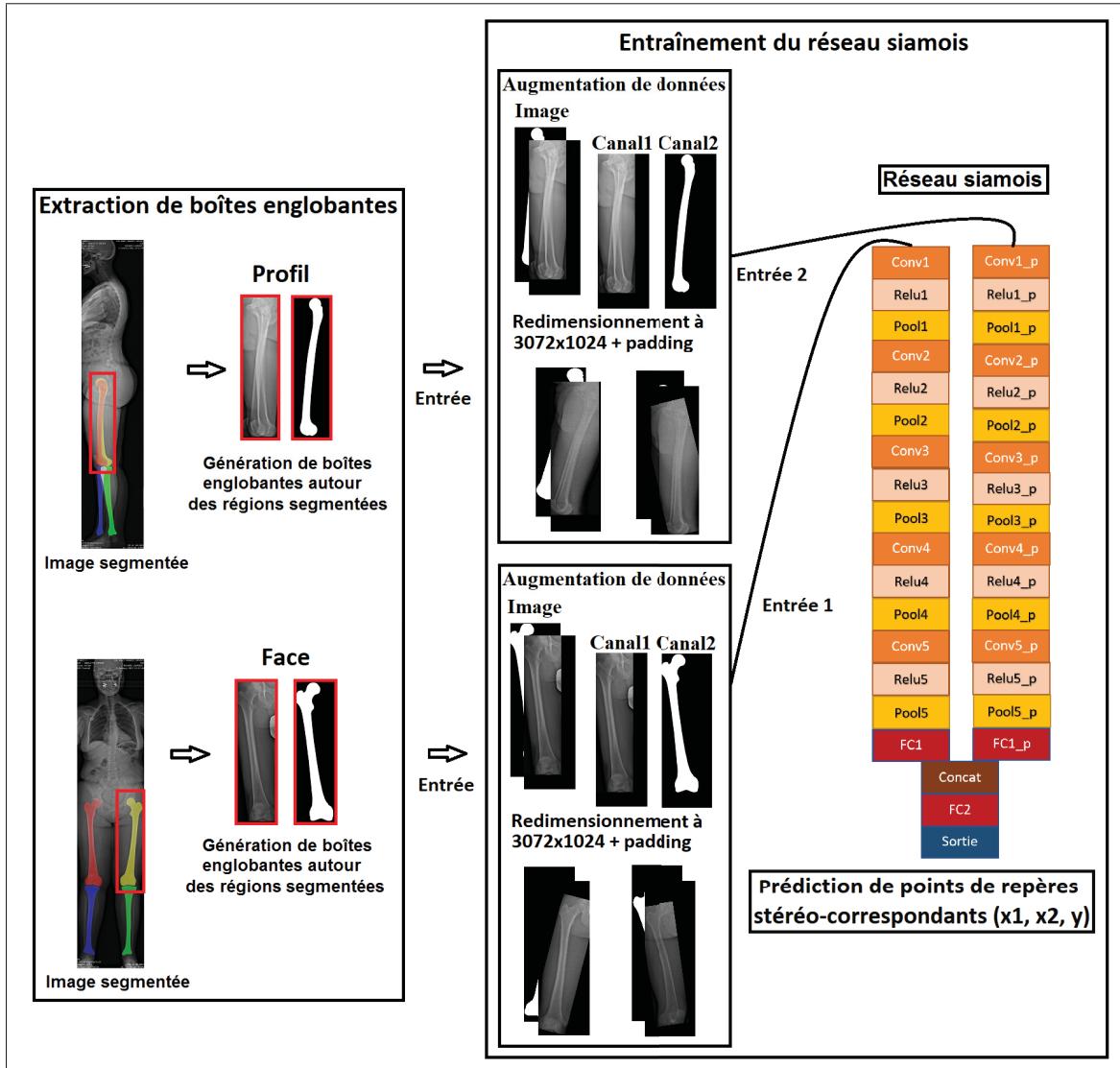


Figure 4.11 Processus d’entraînement de la méthode de localisation et d’identification des 13 points 3D du fémur. Le même processus est réalisé pour le tibia . avec 10 points 3D prédits

sorte que pour chaque paire d’images (face et profil) nous récupérons la même structure osseuse des membres inférieurs (par exemple, le fémur gauche de l’image frontale et le fémur gauche de l’image latérale). Ensuite, nous comparons les tailles de chaque boîte de délimitation potentielle et nous conservons la plus grande taille de chaque dimension. Soit :

- (X_{f_1}, Y_{f_1}) et (X_{f_2}, Y_{f_2}) les coordonnées du coin haut gauche et du coin bas droit de la boîte du fémur gauche sur l'image frontale avec W_f , H_f la largeur et la hauteur correspondantes.
- (X_{p_1}, Y_{p_1}) et (X_{p_2}, Y_{p_2}) les coordonnées du coin haut gauche et du coin bas droit de la boîte englobante du fémur gauche sur l'image de profil avec W_p et H_p la largeur et la hauteur correspondante.

Notre objectif est d'obtenir deux boîtes englobantes de même taille pour la paire (face et profil) avec des coordonnées en Z correspondantes. Supposons que la largeur W_f est supérieure à W_p et la hauteur H_p supérieure à H_f , il faut rappeler que l'image de face et celle de profil EOS partagent la même hauteur H puisqu'elles sont prises en position en charge (debout). Ainsi une même structure osseuse sur l'image de face et sur l'image de profil gardera la même délimitation sur l'axe vertical (sur les images de vérités terrain et dans le cas d'une prédiction sans erreur). Toutefois nous supposerons qu'il pourrait y avoir de petits décalages entre la délimitation verticale d'une structure osseuse sur les images de face et de profil (par exemple dans le cas d'une prédiction erronée). Ainsi sur l'axe vertical, nous prendrons la valeur minimale Y_1 entre Y_{f_1} et Y_{p_1} (du coin haut gauche) et la valeur maximale Y_2 entre Y_{f_2} et Y_{p_2} (du coin bas droit) ; ce qui nous permettra d'avoir une hauteur $H_{max} = Y_2 - Y_1$. Par contre pour la largeur, nous considérerons la largeur W_f (supérieure à W_p). Avec $\delta_w = W_f - W_p$, nous élargirons la boîte englobante de l'image de profil de $\delta_w/2$ à gauche et $\delta_w/2$ à droite. Finalement, nous obtenons deux boîtes englobantes de taille (W_f, H_{max}) . Soit :

- sur l'image de face : (X_{f_1}, Y_1) et (X_{f_2}, Y_2) sont les coordonnées du coin haut gauche et du coin bas droit de la boîte englobante.
- sur l'image de profil : $(X_{p_1} - \delta_w/2, Y_1)$ et $(X_{p_2} + \delta_w/2, Y_2)$ sont les coordonnées du coin haut gauche et du coin bas droit de la boîte englobante .

L'avantage d'avoir deux boîtes de même taille est que les transformations d'augmentation de données (redimensionnement, rotation) se feront de façon uniforme sur les deux images. Aussi, les deux images gardent le même positionnement vertical. Les portions d'image délimitées par

les boîtes englobantes sont alors extraites des images originales EOS (face et profil) pour obtenir une image contenant par exemple le fémur gauche (face ou profil).

4.3.1.2 Phase 2 : Entraînement du réseau siamois

Nous avons dans un premier temps extrait toutes les boîtes englobantes sur les images de tailles originelles EOS. Les boîtes englobantes extraites sont par la suite redimensionnées à 3072×1024 par ajout de zéro (compléter la dimension de taille inférieure avec des zéros pour atteindre la taille voulue). La taille 3072×1024 a été choisie considérant la hauteur et la largeur maximale dans la base de données de boîtes englobantes extraites arrondie au multiple de 256 supérieur pour maîtriser la décroissance de la taille de l'image dans le réseau de neurones. Comme stratégie d'augmentation de données, nous utilisons la rotation.

Nous avons créé une architecture siamoise dans le but de prédire par régression des points d'intérêt stéréo-correspondants des MI. Au lieu d'adopter une stratégie de mise en correspondance d'image, semblable à notre première expérience (voir ANNEXE I), nous l'avons réalisée via une architecture neuronale. Notre architecture siamoise prend en entrée des paires d'images de taille 3072×1024 et de profondeur 2 (deux canaux soit $W \times H \times 2$). Le premier canal contient l'image et le deuxième canal contient le masque permettant ainsi de contraindre le réseau à tenir compte à la fois des caractéristiques présentes dans l'image et du masque pour prédire la position des points d'intérêt. Il est à noter que la solution utilisant une image de profondeur 1 (boîte englobante extraite dans l'image EOS) a été testée. Nous avons retenu la solution utilisant une image de profondeur 2 du fait qu'elle donnait de meilleurs résultats d'identification (Section 5.2). Pour une structure d'intérêt donnée (fémur gauche par exemple), notre réseau siamois prend en entrée la boîte englobante extraite de la radiographie EOS frontale et celle extraite de la radiographie de profil et leurs masques, soit deux matrices $3072 \times 1024 \times 2$, et prédit en sortie des points à trois dimensions (13 points pour le fémur et 10 points pour le tibia). Il est à noter que les fémurs (en orientations $0^\circ/90^\circ$ et $45^\circ/45^\circ$) et les tibias (en orientations $0^\circ/90^\circ$ et $45^\circ/45^\circ$) sont entraînés séparément.

L'architecture du réseau siamois mise en place peut être visualisée sur la Figure 4.11. Elle est composée de cinq couches de convolutions (définies expérimentalement) intercalées par des couches ReLU et de pooling et de deux couches entièrement connectées. Les tailles des couches de convolution, de la couche d'entrée à la couche de sortie, sont 16, 32, 64, 128 et 256 avec un noyau de taille 9 et un pas de 1. La taille du noyau pour les couches de pooling est définie à 4 avec un pas de 2. Ces différents paramètres ont été choisis expérimentalement. Avant les couches entièrement connectées, nous avons une couche de concaténation (concat) qui concatène les caractéristiques de chaque branche du réseau siamois. La concaténation est faite sous forme de sommation élément par élément. Les deux couches entièrement connectées ont respectivement une taille de 512 et 39 (ou 30 dépendamment s'il s'agit du fémur ou du tibia). Comme fonction de perte, nous avons utilisé la fonction de perte euclidienne E qui mesure le carré de la distance entre les points prédis p par le réseau neuronal et les points de vérités terrain g . Elle est définie comme suit :

$$E = \frac{1}{2N} \sum_{n=1}^N \|p_n - g_n\|_2^2 \quad (4.2)$$

avec N l'ensemble de données d'entraînement.

4.3.2 Détail d'implémentation

L'ensemble de données utilisé dans le cadre de ce travail est constitué de 180 données EOS, soit 120 images d'entraînement, 30 images de validation et 30 images de test. Nous avons implémenté le modèle avec la librairie Caffe et effectué l'entraînement et les tests sur une carte graphique NVIDIA Titan RTX de 24Go de mémoire. Nous avons utilisé un gradient accéléré de Nesterov avec un momentum de 0.7 et un taux d'apprentissage fixe de 0.0001. Un batch size de 1 a été utilisé pour l'entraînement. Les hyperparamètres ont été choisis expérimentalement.

4.4 Méthode de validation

Dans le but de valider notre méthode de segmentation de structures osseuses et d'identification de points d'intérêt des membres inférieurs, une comparaison sera effectuée entre les résultats de

prédition de nos différentes méthodes et les vérités terrain. Les résultats seront présentés sous la forme moyenne (moy) \pm écart-type (std). Pour la méthodologie de segmentation de structures osseuses des membres inférieurs développée, nous allons dans un premier temps présenter les résultats de l'approche proposée dans la section 5.1.1.

Nous allons ensuite évaluer l'impact qu'a la technique d'augmentation de données proposée sur le résultat de segmentation de RobustNet. Pour cela, trois entraînements seront séparément effectués sur trois ensembles de données générés respectivement à partir de trois différentes techniques d'augmentation de données : la technique d'augmentation de données proposée, la technique générant des images ne contenant que des structures partielles, et celle générant des images contenant uniquement des structures complètes. Cette évaluation sera présentée dans la section 5.1.2.

Nous allons également évaluer l'impact que les modifications apportées au SegNet ont sur le résultat de la segmentation. Ceci sera présenté dans la section 5.1.3.

Pour la méthodologie de localisation et d'identification des points d'intérêt des membres inférieurs développée, nous allons présenter les résultats de l'approche proposée dans la section 5.2. Pour évaluer la précision de prédition de nos différentes méthodes, nous utiliserons différentes métriques.

4.4.1 Métrique de validation de la méthode d'identification et de segmentation des os du MI

Pour évaluer la méthode développée, nous vérifions dans un premier temps si les structures sont bien détectées dans l'image puis évaluons la qualité de segmentation des structures détectées à l'aide de métriques de segmentation. Nous supposons qu'une structure est bien détectée tant que nous avons un indice de Jaccard (Intersection over Union) supérieur ou égal à 50%. Nous utiliserons l'indice de Dice (D_c) et l'indice de Jaccard (J) comme métriques pour évaluer nos résultats de segmentation. Ces mesures sont définies comme suit : Soient T_d , N_d , S_t , G , P et $|.|$ représentant respectivement le taux de détection, le nombre de bonnes détections, la

taille de l'ensemble de données de test, la vérité terrain, la prédiction et le cardinal d'un ensemble.

$$T_d = \frac{|N_d|}{|S_t|} \quad (4.3)$$

$$D_c = \frac{2|G \cap P|}{|G| + |P|} \quad (4.4)$$

$$J = \frac{|G \cap P|}{|G \cup P|} \quad (4.5)$$

4.4.2 Métrique de validation de la méthode de localisation et d'identification des points d'intérêt

Dans le but d'évaluer la méthode de localisation et d'identification de points d'intérêt mise en place, nous allons utiliser la distance euclidienne comme métrique d'évaluation. Soit $d(p, g)$ la distance euclidienne entre un point prédict $p(p_x, p_y, p_z)$ par notre réseau neuronal et le point de vérité terrain $g(g_x, g_y, g_z)$ correspondant. La distance euclidienne d entre l et m est définie comme suit :

$$d(p, g) = \sqrt{(p_x - g_x)^2 + (p_y - g_y)^2 + (p_z - g_z)^2} \quad (4.6)$$

4.5 Résumé

Nous proposons une méthode entièrement automatique pour la localisation et l'identification d'un premier estimé de points d'intérêt stéréo-correspondants dans les images EOS. La méthodologie adoptée consiste à, dans un premier temps, segmenter les structures osseuses des membres inférieurs dans des images EOS présentant différents champs de vue (exemples : images corps entier, images du membre inférieur) et orientations (exemples : images en orientation 0° , 45° , 90°). Des paires de boîtes englobantes de même taille (une pour la face et une pour le profil) sont par la suite générées (à partir des masques) pour chacune des structures osseuses des membres

inférieurs. Ces paires de même que leurs masques associés servent par la suite d'entrée à notre réseau siamois qui est alors capable de prédire en sortie la position tridimensionnelle des points d'intérêt stéréo-correspondants appartenant à la fois aux images EOS de face et de profil.

CHAPITRE 5

RÉSULTATS

Dans ce chapitre, nous présenterons les résultats de la méthode proposée. Dans un premier temps, les résultats qualitatifs et quantitatifs de la méthode d'identification et de segmentation de structures osseuses des membres inférieurs proposée seront présentés. Par la suite, nous présenterons les résultats qualitatifs et quantitatifs de la méthode de localisation et d'identification de points employés pour la déformation des modèles génériques 3D du fémur et du tibia.

5.1 Résultats de la méthode d'identification et de segmentation de structures osseuses des membres inférieurs

5.1.1 Résultats RobustNet

RobustNet a été validé sur notre ensemble de données de test (70 images) constitué d'images radiographiques entières (images de corps entiers et images des membres inférieurs) et partielles (images contenant une partie de l'image EOS). Les Tableaux 5.1, 5.2 et 5.3 ci-dessous présentent les résultats quantitatifs obtenus respectivement sur l'entièreté de notre ensemble de données, sur notre ensemble de données d'images frontales (images en orientations 0° , 20° et 45°) et sur notre ensemble de données d'images latérales (images en orientations 45° , 70° , 90°). Sur l'entièreté de notre ensemble de données (Tableau 5.1), nous avons obtenu un taux de détection de 98.00%, un coefficient de Dice moyen de $95.25 \pm 9.02\%$ et un indice de Jaccard moyen de $91.93 \pm 11.86\%$. Sur notre ensemble de données d'images frontales (Tableau 5.2), nous avons obtenu un taux de détection de 100.00%, un coefficient de Dice moyen de $96.86 \pm 3.78\%$ et un indice de Jaccard moyen de $94.14 \pm 6.16\%$ tandis que sur notre ensemble de données d'images latérales (Tableau 5.3), nous avons obtenu un taux de détection de 96.00%, un coefficient de Dice moyen de $93.73 \pm 11.84\%$ et un indice de Jaccard moyen de $89.96 \pm 15.14\%$. Les Figures 5.1 et 5.2 présentent respectivement des résultats qualitatifs obtenus sur notre ensemble de données d'images entières et d'images partielles. La Figure 5.3 présente quelques exemples où

la méthode développée identifie de petits morceaux d'os dans des images partielles. La figure 5.4 montre quelques cas d'images où la segmentation comporte des erreurs significatives.

Tableau 5.1 Résultats par classe pour la totalité de notre ensemble de données de test

Méthodes	T_d	D_c	J
		<i>moy ± std</i>	<i>moy ± std</i>
BackG	100	99.43 ± 0.28	98.86 ± 0.55
FemG with superposition	98.57	94.34 ± 8.79	90.24 ± 11.82
FemD with superposition	97.14	94.17 ± 12.45	90.61 ± 14.29
TibG with superposition	97.14	94.02 ± 8.22	89.63 ± 11.99
TibD with superposition	97.14	93.99 ± 9.53	89.80 ± 12.88
Class average	98.00	95.25 ± 9.02	91.93 ± 11.86
SupFem	92.86	81.90 ± 23.51	74.01 ± 24.67
SupTib	92.86	57.68 ± 32.51	46.26 ± 28.33

Tableau 5.2 Résultats par classe pour les images frontales de notre ensemble de données

Méthodes	T_d	D_c	J
		<i>moy ± std</i>	<i>moy ± std</i>
BackG	100	99.48 ± 0.23	98.97 ± 0.46
FemG with superposition	100	96.51 ± 3.18	93.42 ± 5.24
FemD with superposition	100	97.02 ± 1.54	94.26 ± 2.83
TibG with superposition	100	95.89 ± 4.71	92.42 ± 7.40
TibD with superposition	100	94.94 ± 5.51	90.80 ± 8.61
Class average	100	96.86 ± 3.78	94.14 ± 6.16
SupFem	94.29	15.30 ± 6.34	8.34 ± 3.72
SupTib	100	N/A	N/A

Tableau 5.3 Résultats par classe pour les images latérales de notre ensemble de données

Méthodes	T_d	D_c	J
		<i>moy ± std</i>	<i>moy ± std</i>
BackG	100	99.37 ± 0.31	98.75 ± 0.62
FemG with superposition	94.14	92.23 ± 11.63	87.16 ± 15.27
FemD with superposition	94.29	91.48 ± 16.96	87.17 ± 19.24
TibG with superposition	94.29	92.33 ± 10.23	87.09 ± 14.67
TibD with superposition	94.29	93.16 ± 12.05	88.92 ± 15.81
Class average	96.00	93.73 ± 11.84	89.86 ± 15.14
SupFem	91.43	88.25 ± 10.96	80.27 ± 14.02
SupTib	85.71	57.68 ± 32.51	46.26 ± 28.33

5.1.2 Influence du choix de techniques d'augmentation des données sur la prédiction

Dans cette partie, la méthode d'augmentation de données adoptée dans ce travail, qui applique des transformations locales, mais aussi globales aux images (ADLG) pour générer 132309

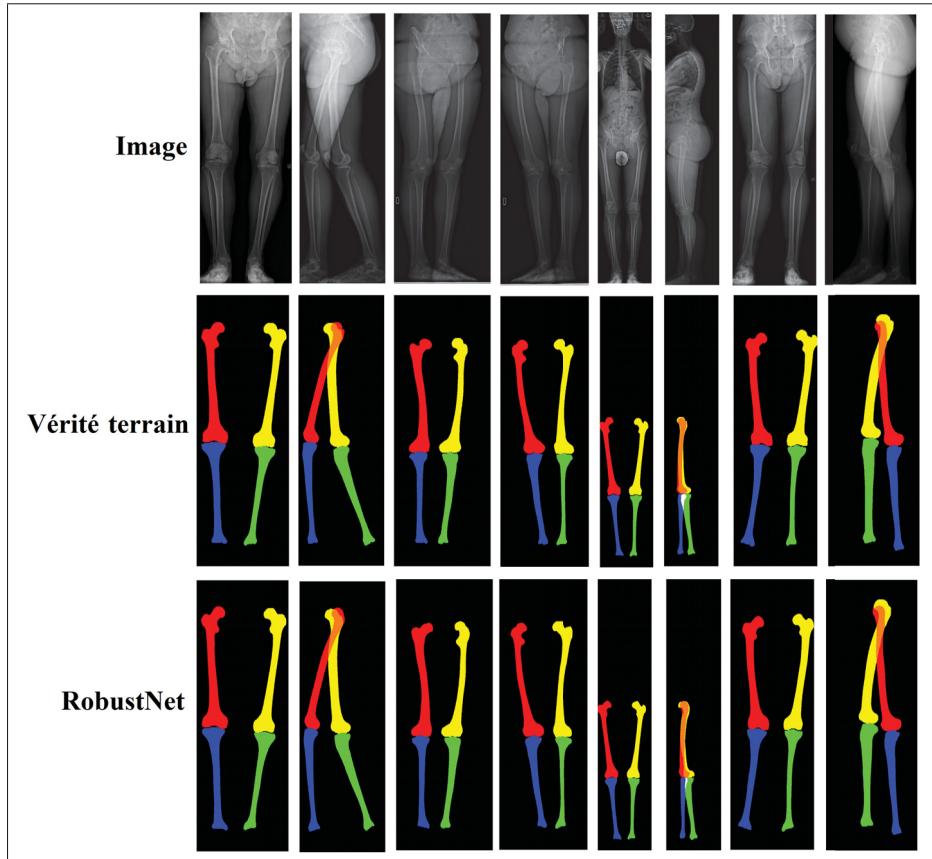


Figure 5.1 Résultats qualitatifs de radiographies entières. La première ligne représente les radiographies, la deuxième ligne représente la vérité terrain et la troisième ligne représente la prédiction avec RobustNet

images d’entraînement, est comparée à celle appliquant uniquement des transformations locales (ADL) et celle appliquant uniquement des transformations globales (ADG) aux images. Pour ce faire, la fenêtre glissante, le recadrage et la rotation sont utilisés comme stratégies d’augmentation des données pour générer 116829 images d’entraînement contenant uniquement des structures partielles, tandis que la mise à l’échelle non uniforme et la rotation sont utilisées pour générer 15480 images d’entraînement contenant uniquement des structures complètes. Le but des résultats (sur l’ensemble de données de test toutes images confondues) quantitatifs et qualitatifs présentés dans le Tableau 5.4 et la Figure 5.5 est de montrer l’influence qu’a le choix de la méthode d’augmentation des données sur les résultats de segmentation. En utilisant la transformation ADLG, nous avons obtenu un taux de détection de 98.00%, un coefficient de Dice moyen de

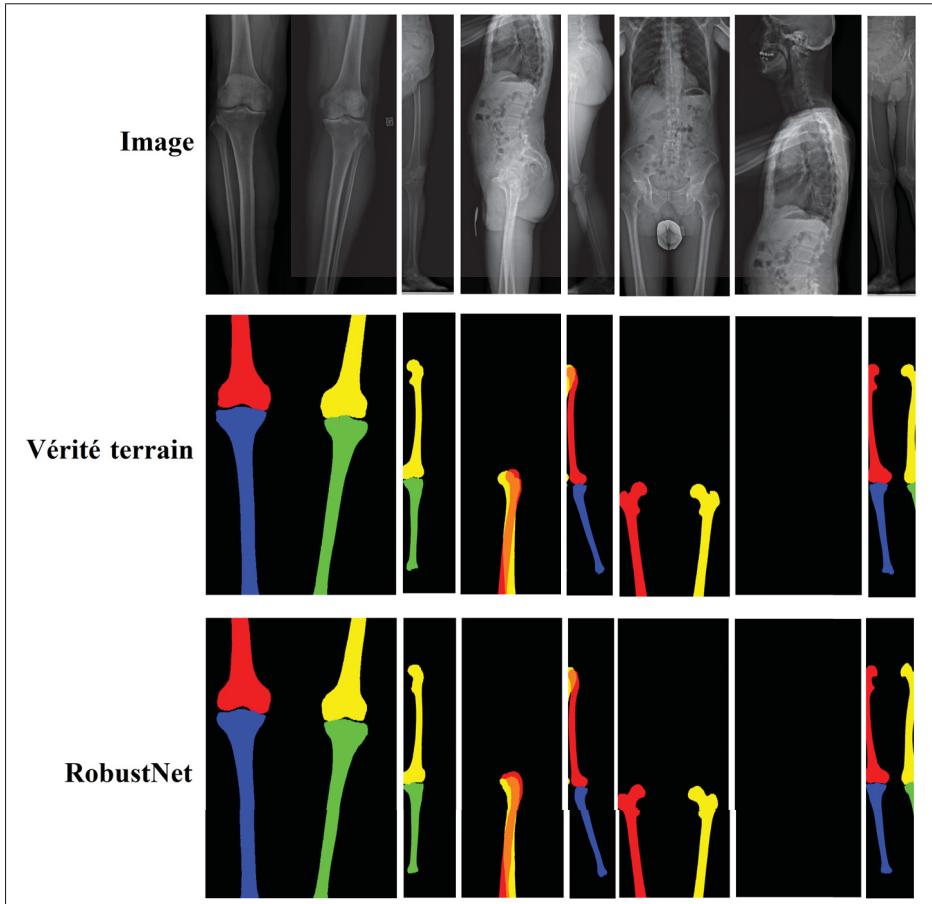


Figure 5.2 Résultats qualitatifs d'images avec structures partielles. La première ligne représente les radiographies, la deuxième ligne représente la vérité terrain et la troisième ligne représente la prédiction avec RobustNet

$95.25 \pm 9.02\%$ et un indice de Jaccard moyen de $91.93 \pm 11.86\%$ tandis qu'avec la transformation ADL nous avons obtenu un taux de détection de 96.00% , un coefficient de Dice moyen de $93.35 \pm 14.15\%$ et un indice de Jaccard moyen de $89.70 \pm 16.65\%$. En utilisant ADG, nous avons obtenu un taux de détection de 80.00% , un coefficient de Dice moyen de $83.20 \pm 22.67\%$ et un indice de Jaccard moyen de $76.26 \pm 26.41\%$. Pour déterminer si les différences entre (ADLG et ADL) et (ADLG et ADG) sont significatives, le test de Wilcoxon a été réalisé. Une p-value < 0.0001 a été obtenue (Tableau 5.4) dans les deux cas (ADLG vs ADL et ADLG vs ADG).

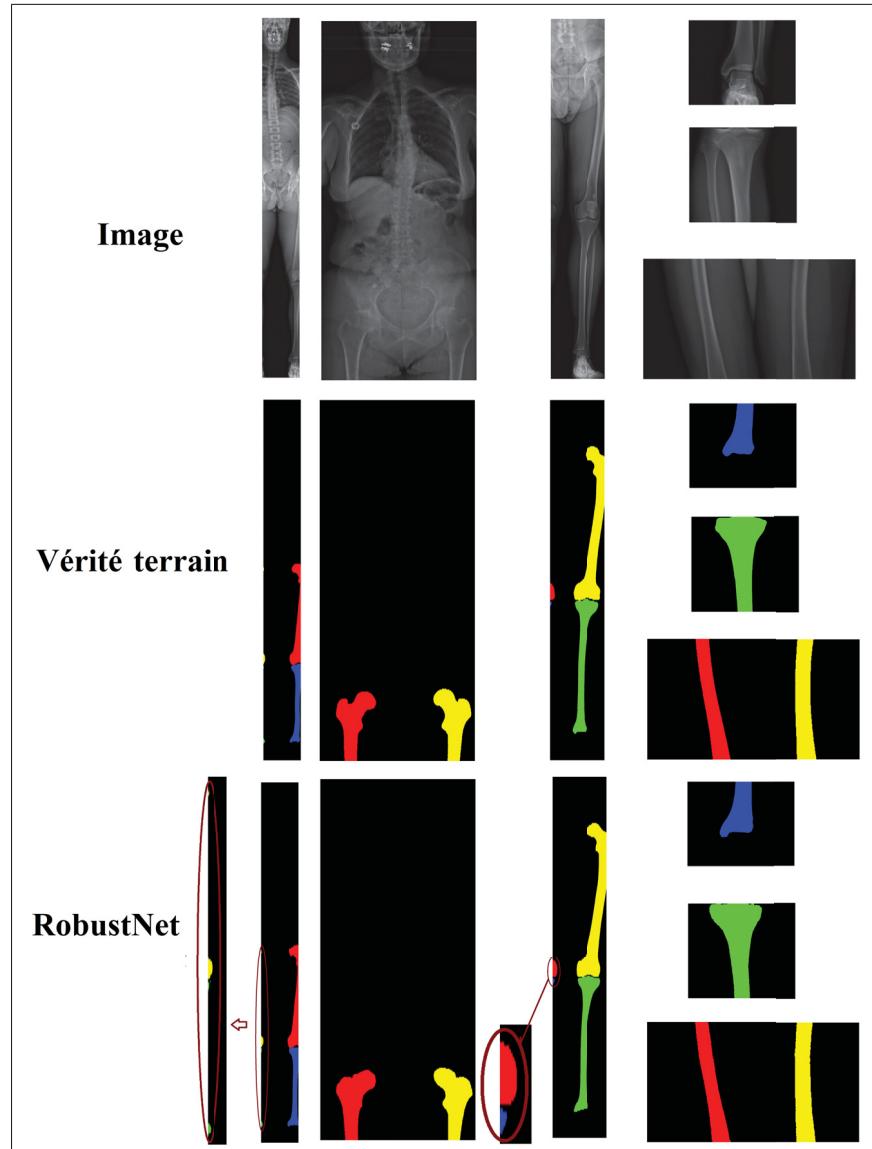


Figure 5.3 Images partielles contenant un petit morceau d'os. La première ligne représente les radiographies, la deuxième ligne représente la vérité terrain et la troisième ligne représente la prédiction avec RobustNet

Tableau 5.4 Comparaison des résultats de segmentation selon la méthode d'augmentation de données utilisée (%)

Méthodes	T_d	D_c	J	p-values
		$moy \pm std$	$moy \pm std$	
ADLG	98.00	95.25 ± 9.02	91.93 ± 11.86	
ADL	96.00	93.35 ± 14.15	89.70 ± 16.65	< 0.0001
ADG	80.00	83.20 ± 22.67	76.26 ± 26.41	< 0.0001

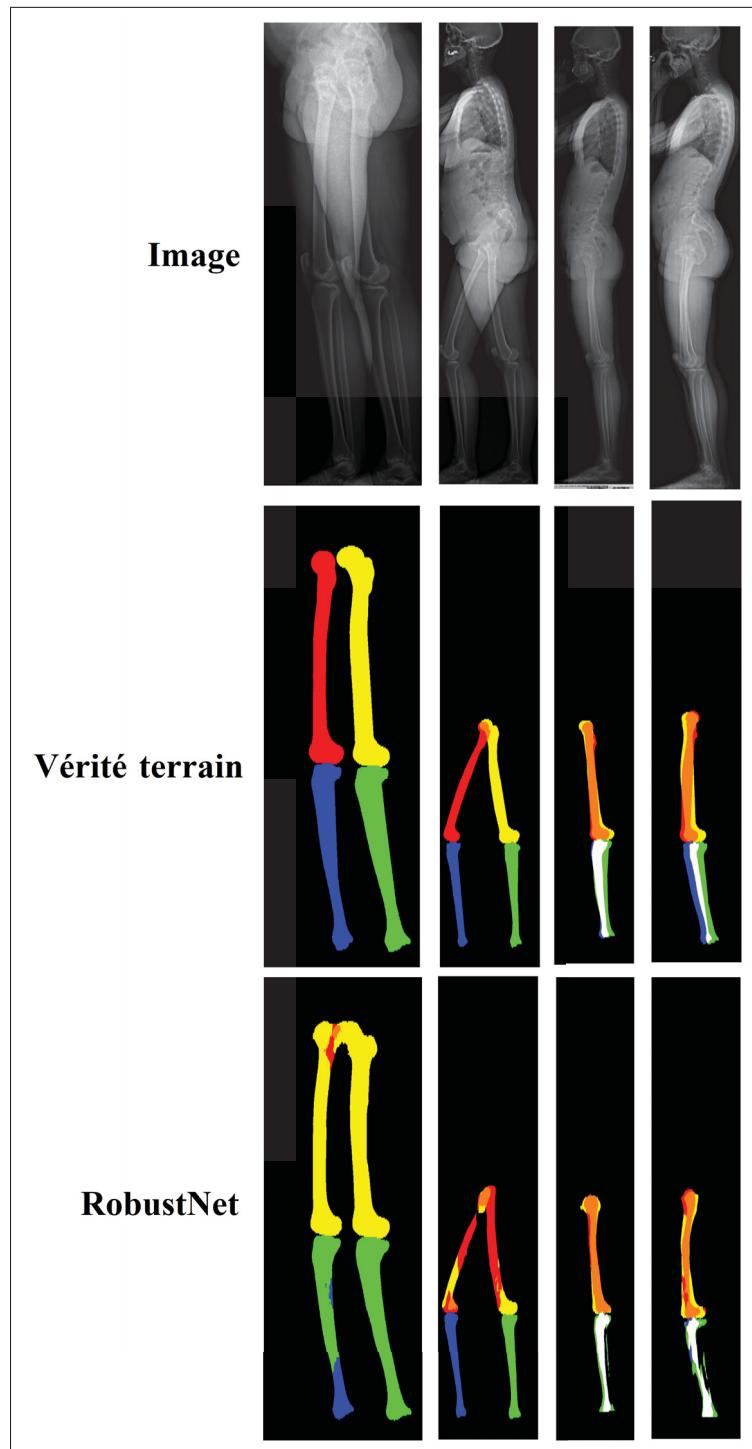


Figure 5.4 Cas où la segmentation comporte des erreurs importantes. La première ligne représente les radiographies, la deuxième ligne la vérité terrain et la troisième ligne la prédiction avec RobustNet

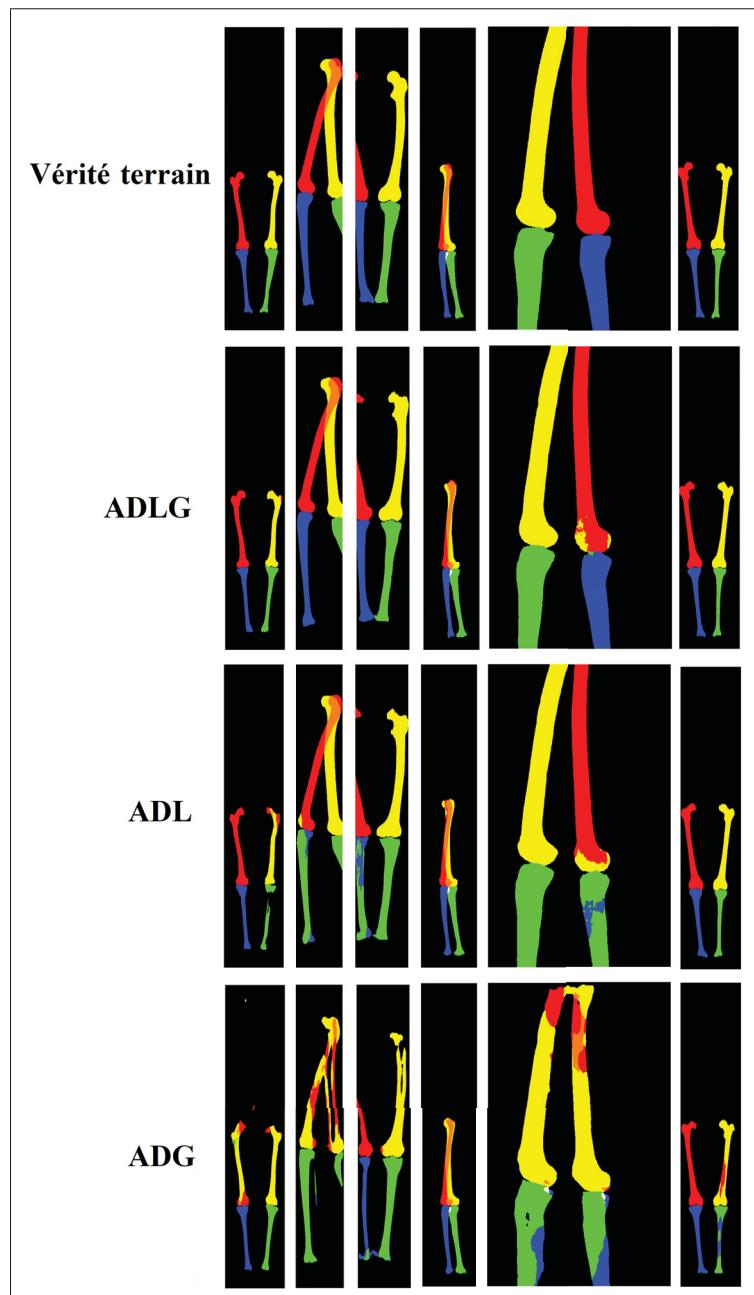


Figure 5.5 Prédiction de RobustNet utilisant ADLG, ADL et ADG. La première ligne représente la vérité terrain. Les deuxième, troisième et quatrième lignes représentent la prédiction de RobustNet utilisant respectivement la stratégie d'augmentation de données appliquant des transformations locales et globales (ADLG), la stratégie d'augmentation de données appliquant des transformations locales (ADL) et la stratégie d'augmentation de données appliquant des transformations globales (ADG)

5.1.3 RobustNet versus SegNet

Les courbes de convergence de RobustNet et de SegNet sont présentées à la Figure 5.6. Les résultats de RobustNet sont comparés quantitativement (Tableau 5.5) et qualitativement (Figure 5.7) aux résultats de SegNet sur l'ensemble de données de test (toutes images confondues). Nous avons obtenu avec RobustNet un taux de détection de 98.00%, un coefficient de Dice moyen de $95.25 \pm 9.02\%$ et un indice de Jaccard moyen de $91.93 \pm 11.86\%$ tandis qu'avec Segnet nous avons obtenu un taux de détection de 96.86%, un coefficient de Dice moyen de $94.39 \pm 12.07\%$ et un indice de Jaccard moyen de $91.00 \pm 14.47\%$. La p-value du test de Wilcoxon est calculée pour déterminer si la différence de distributions entre RobustNet et SegNet est quantitativement significative. Une p-value de 0.78 a été obtenue (Tableau 5.5).

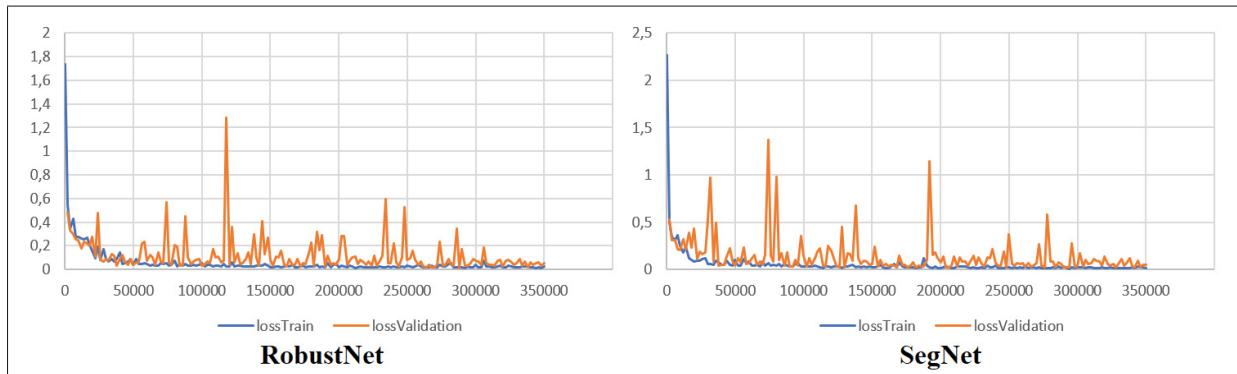


Figure 5.6 Courbes de convergence de RobustNet et de SegNet

Tableau 5.5 RobustNet vs SegNet pour l'ensemble de données de test (%)

Méthodes	T_d	D_c	J	p-value
		$moy \pm std$	$moy \pm std$	
RobustNet	98.00	95.25 ± 9.02	91.93 ± 11.86	
SegNet	96.86	94.39 ± 12.07	91.00 ± 14.47	0.78

La Figure 5.8 présente quelques cas de mauvaise classification obtenus avec SegNet. Sur cette figure, un zoom est effectué sur les cas d'erreurs observés à la Figure 5.7.

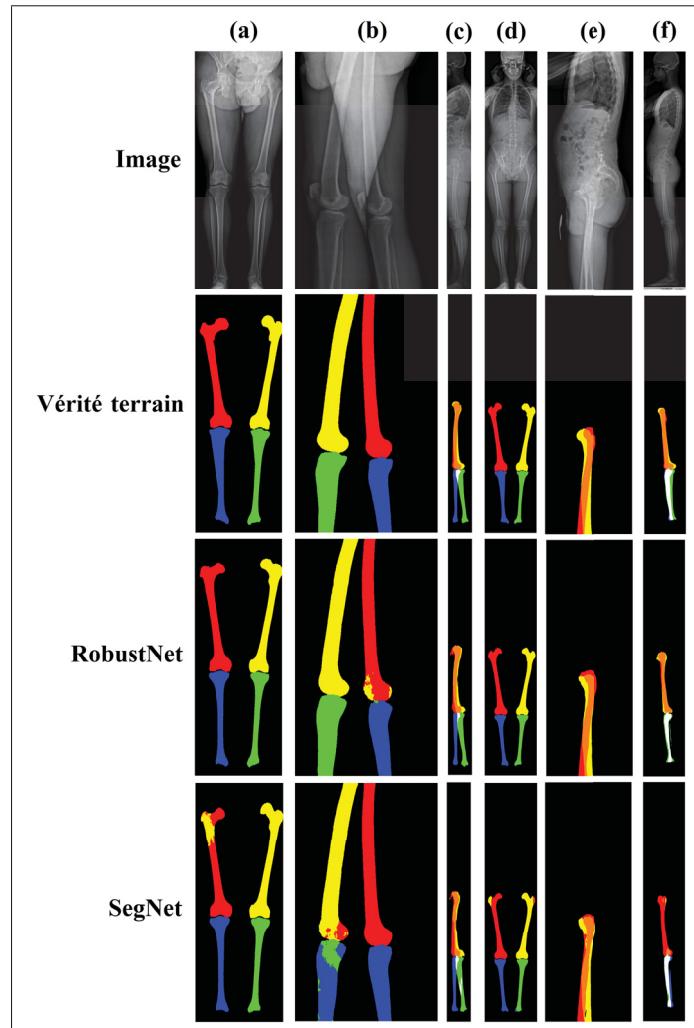


Figure 5.7 Comparaison qualitative entre RobustNet et SegNet

5.2 Résultats de la méthode de localisation et l'identification des points d'intérêt des membres inférieurs

Quinze paires d'images EOS ont été utilisées pour valider notre approche de localisation et d'identification de points d'intérêt stéréo-correspondants. Nous nous retrouvons avec 60 paires d'images ; soit 30 paires pour le fémur et 30 paires pour le tibia. Il est à noter que tous les points d'intérêt sont localisés et étiquetés automatiquement dans leurs régions anatomiques respectives. La distance euclidienne entre les points prédits et ceux de la vérité terrain est utilisée comme métrique de validation.

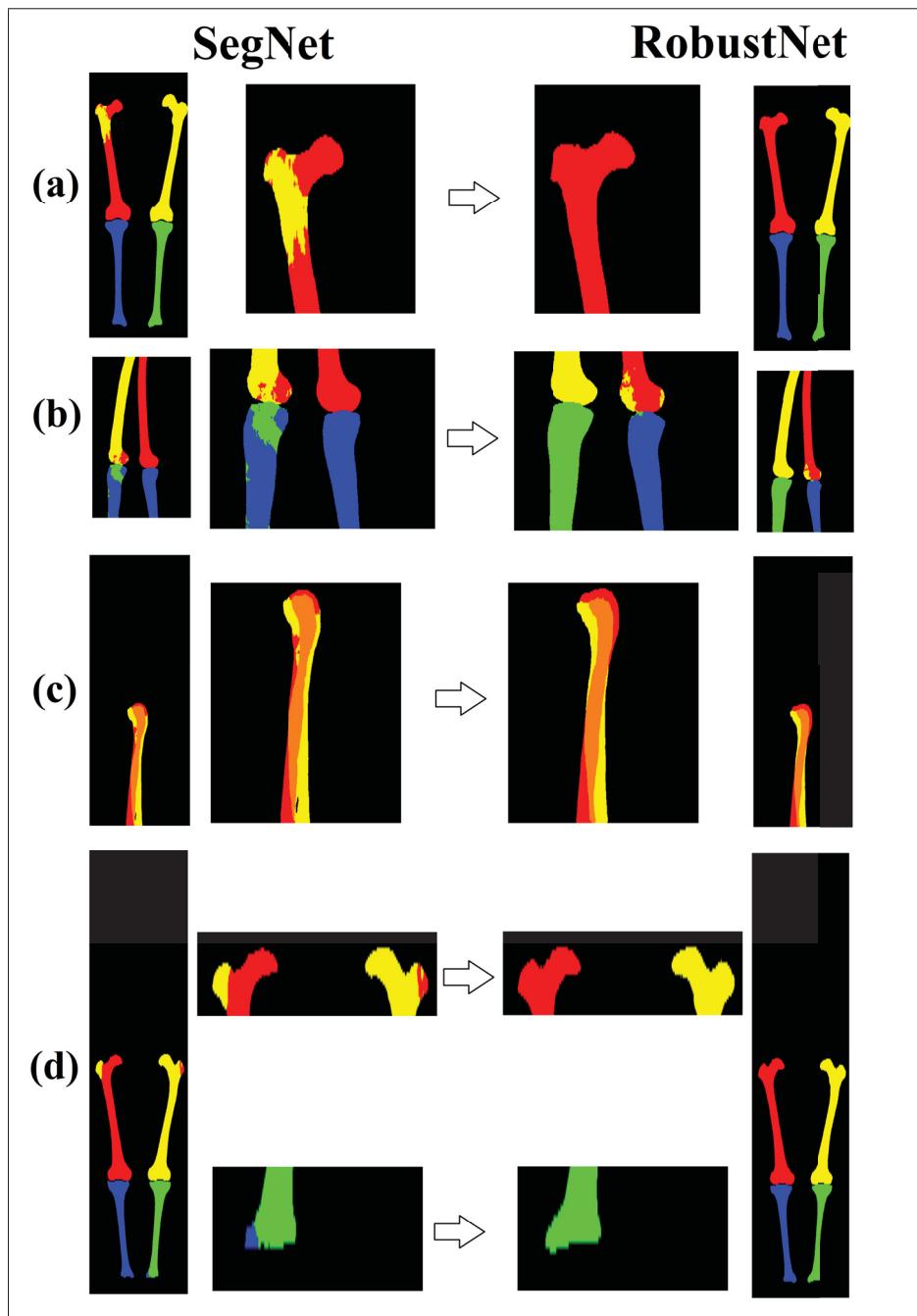


Figure 5.8 Quelques limitations de SegNet corrigées par RobustNet

5.2.1 Boîtes englobantes et masques comme données d'entrée

Les boîtes englobantes et leurs masques associés sont ici utilisés comme données d'entrée du réseau siamois. Les Tableaux 5.6 et 5.7 présentent respectivement pour nos 15 paires d'images EOS de test les erreurs quantitatives moyennes de localisation pour chaque point d'intérêt du fémur et du tibia en 2D et en 3D. Nous obtenons une erreur de localisation moyenne 2D (sur les points) de 6.16 ± 4.55 mm soit une erreur moyenne de 6.59 ± 4.86 mm pour le fémur et une erreur moyenne de 5.72 ± 4.24 mm pour le tibia. En 3D, nous avons obtenu une erreur moyenne de 8.14 ± 5.38 mm soit une erreur de 8.39 ± 5.61 mm pour le fémur et une erreur de 7.88 ± 5.14 mm pour le tibia. Des résultats qualitatifs de localisation des points d'intérêt du fémur et du tibia peuvent être visualisés aux Figures 5.9 et 5.10. Les Figures 5.11 et 5.12 présentent respectivement quelques cas de mauvaises localisations pour le fémur et le tibia.

Tableau 5.6 Erreurs (mm) par point pour le fémur (cas où boîtes englobantes et masques considérés comme entrée).

PPI (PlateauPostInt), PPE (PlateauPostExt), PI (PlateauInt), PE (Plat.
 TF (centre de la tête fémorale), GT (grand trochanter), PT (petit trochanter), Tc (trochlée), CPE (CondylePostExt), CPI (CondylePostInt), CIE (CondyleInfExt), CII (CondyleInfInt), PD (PointDia), SD₁ (SplineDia-1), SD₂ (SplineDia-2), SD₃ (SplineDia-3), SD₄ (SplineDia-4)

		TF	GT	PT	Tc	CPE	CPI	CIE	CII	PD	SD ₁	SD ₂	SD ₃	SD ₄	Moy	
Point 2D	Face	moy	6.87	8.95	6.20	5.72	9.10	8.09	10.11	6.92	5.73	4.64	4.67	4.56	5.92	6.73
		std	4.30	5.56	3.86	4.01	5.82	6.24	7.46	5.27	4.35	4.42	3.77	3.16	3.97	4.79
		min	1.55	1.59	0.53	1.27	1.17	0.66	0.75	1.27	0.22	0.51	0.26	0.88	0.63	0.87
		max	17.54	24.41	13.04	14.42	25.60	25.60	30.18	23.18	19.50	15.03	15.55	14.94	16.56	19.66
	Profil	moy	5.73	7.54	6.04	6.18	9.46	8.25	9.16	6.24	6.24	4.89	3.89	4.37	5.92	6.45
		std	3.99	4.97	3.82	4.33	7.70	6.69	6.68	6.20	4.20	4.09	3.26	3.01	5.09	4.92
		min	0.38	0.35	0.38	0.21	0.57	0.87	0.57	0.80	1.45	0.26	0.15	0.24	0.05	0.48
		max	15.16	22.27	14.75	15.13	31.66	27.74	28.05	33.54	17.80	14.68	13.74	14.10	18.55	20.55
	Point 3D	moy	8.20	10.46	7.90	7.14	12.08	11.25	12.56	8.99	7.31	5.86	5.01	5.09	7.20	8.39
		std	5.14	6.32	4.45	4.19	8.40	7.60	8.79	7.10	4.57	4.82	3.72	3.17	4.72	5.61
		min	1.60	1.62	0.58	1.34	1.20	1.74	0.85	1.35	1.54	0.96	0.28	1.19	1.25	1.19
		max	20.58	27.01	19.56	15.40	39.72	27.89	30.61	33.97	19.77	16.14	15.56	15.30	18.60	23.78

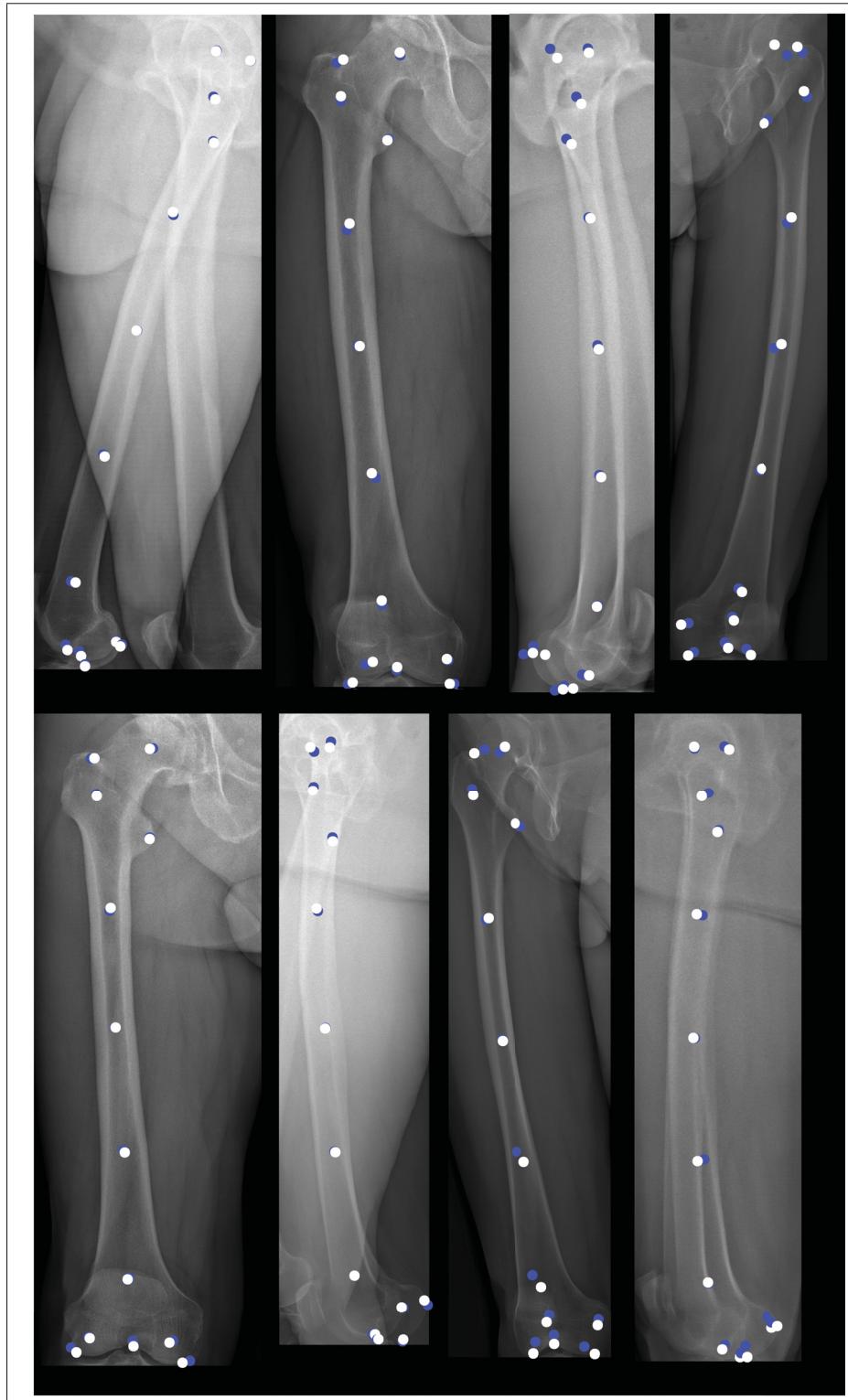


Figure 5.9 Identification de points d'intérêt stéréo-correspondants du fémur dans les radiographies EOS. Les points en blanc représentent la vérité terrain tandis que les points en bleu représentent les points prédicts

Tableau 5.7 Erreurs (mm) par point pour le tibia (cas où boîtes englobantes et masques considérés comme entrée).

PPI (PlateauPostInt), PPE (PlateauPostExt), PI (PlateauInt), PE (PlateauExt), SDT (SplineDiaT), SDT₁ (SplineDiaT-1), SDT₂ (SplineDiaT-2), SDT₃ (SplineDiaT-3), SDT₄ (SplineDiaT-4), CE(CentreEpines)

		PPI	PPE	PI	PE	SDT	SDP ₁	SDP ₂	SDP ₃	SDP ₄	CE	Moy	
Point 2D	Face	moy	7.32	12.50	9.68	10.67	4.30	3.80	2.99	4.10	4.02	3.70	6.31
		std	5.64	10.03	8.16	7.82	2.85	2.94	1.64	3.22	2.51	2.24	4.71
		min	1.56	0.64	0.40	1.78	0.18	0.71	0.51	0.11	0.26	0.60	0.67
		max	24.67	39.34	34.28	35.96	10.76	13.56	6.44	14.37	11.32	7.59	19.83
	Profil	moy	6.95	8.56	6.79	5.50	4.43	3.51	2.48	4.34	4.35	4.47	5.14
		std	7.21	7.37	4.74	3.19	2.34	2.73	1.30	3.06	2.63	3.14	3.77
		min	0.93	1.86	0.74	0.99	0.49	0.52	0.36	0.49	0.51	1.26	0.81
		max	40.23	32.24	20.51	12.97	10.11	13.54	5.76	13.82	10.23	14.40	17.38
	Point 3D	moy	10.46	15.80	12.24	11.93	5.42	4.37	3.35	4.59	5.22	5.44	7.88
		std	7.85	11.00	8.24	7.85	2.67	2.83	1.69	3.11	2.86	3.35	5.14
		min	1.81	2.89	2.00	2.13	1.14	1.24	0.62	0.50	1.34	1.41	1.51
		max	40.29	40.62	34.70	36.06	10.84	13.56	8.16	14.39	14.89	15.38	22.89

5.2.2 Boîtes englobantes comme données d'entrée

Uniquement les boîtes englobantes sont ici utilisées comme données d'entrée du réseau siamois. Les Tableaux 5.8 et 5.9 présentent respectivement pour nos 15 paires d'images EOS de test les erreurs quantitatives moyennes de localisation pour chaque point d'intérêt du fémur et du tibia en 2D et en 3D. Nous obtenons une erreur de localisation moyenne 2D sur les points de 9.93 ± 5.94 mm soit une erreur moyenne de 10.42 ± 6.48 mm pour le fémur et une erreur moyenne de 9.44 ± 5.39 mm pour le tibia. En 3D, nous obtenons une erreur moyenne de 12.68 ± 6.69 mm soit une erreur de 13.90 ± 7.47 mm pour le fémur et une erreur de 11.46 ± 5.91 mm pour le tibia.

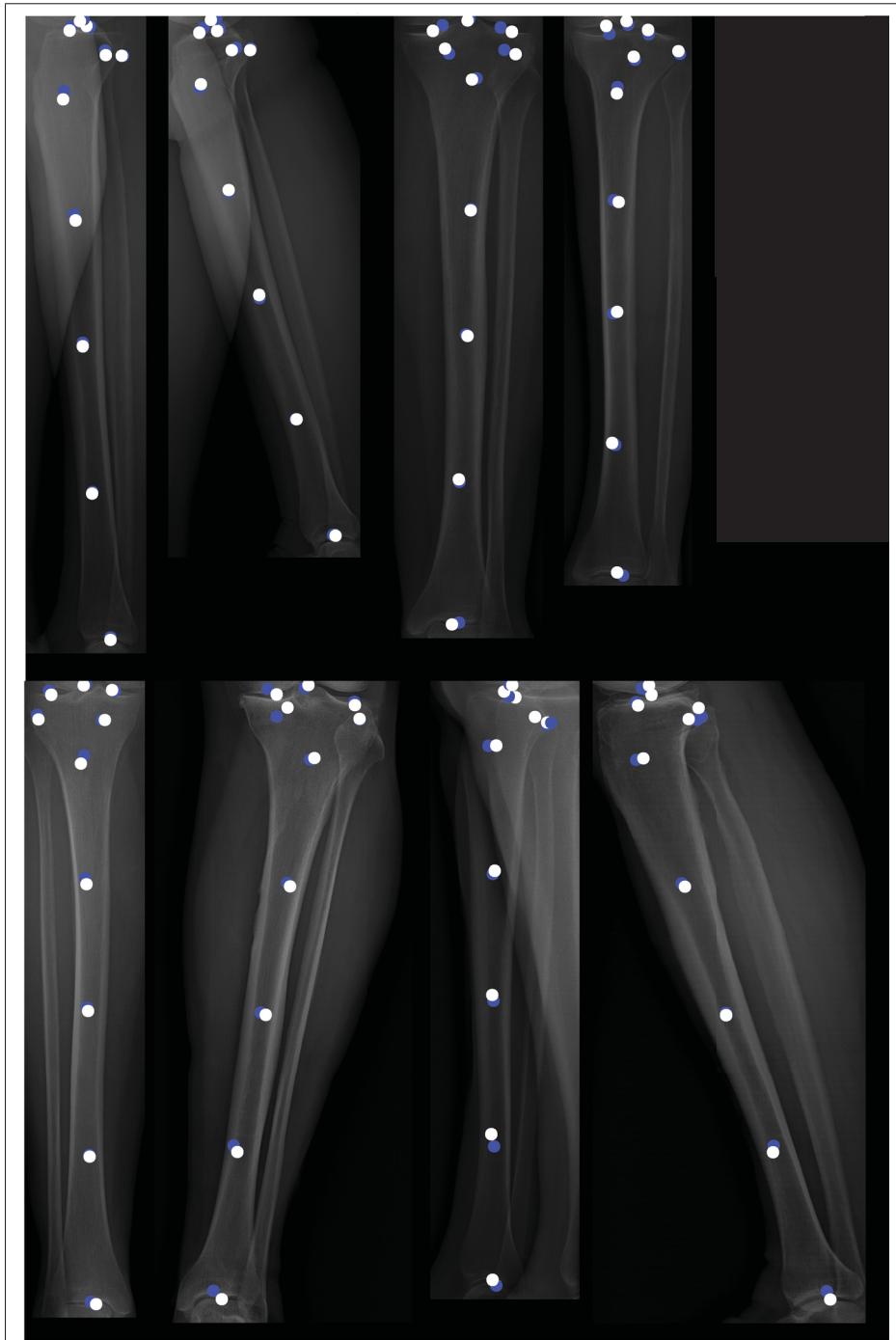


Figure 5.10 Identification de points d'intérêt stéréo-correspondants du tibia dans les radiographies EOS. Les points en blanc représentent la vérité terrain tandis que les points en bleu représentent les points prédicts

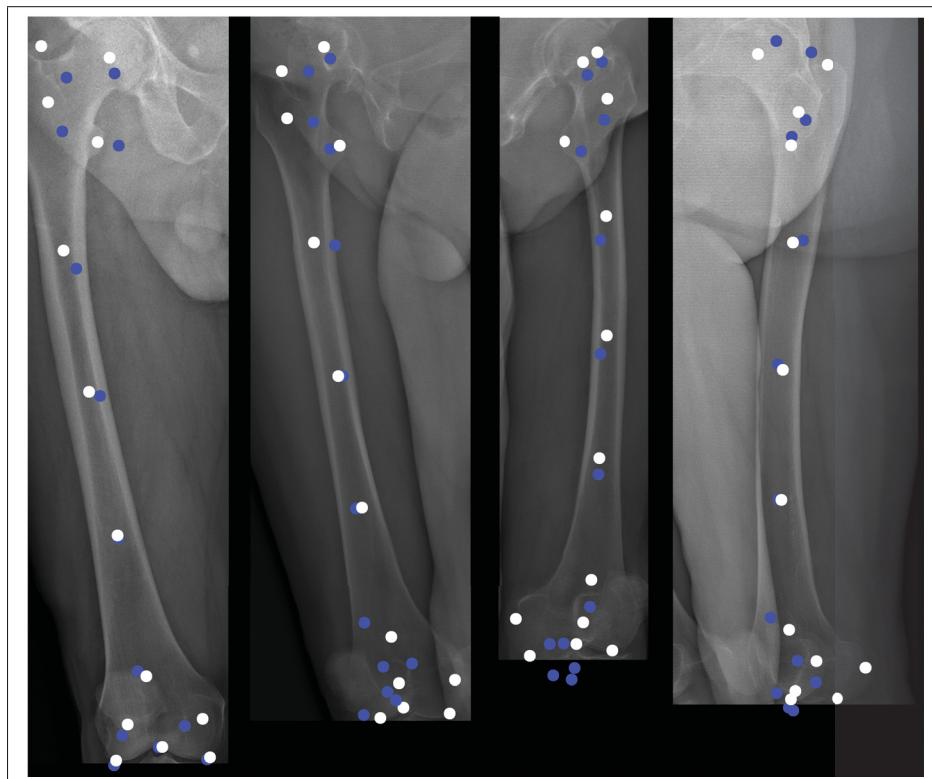


Figure 5.11 Images présentant de fortes erreurs de localisation de points d'intérêt stéréo-correspondants du fémur dans les radiographies EOS. Les points en blanc représentent la vérité terrain tandis que les points en bleu représentent les points prédits

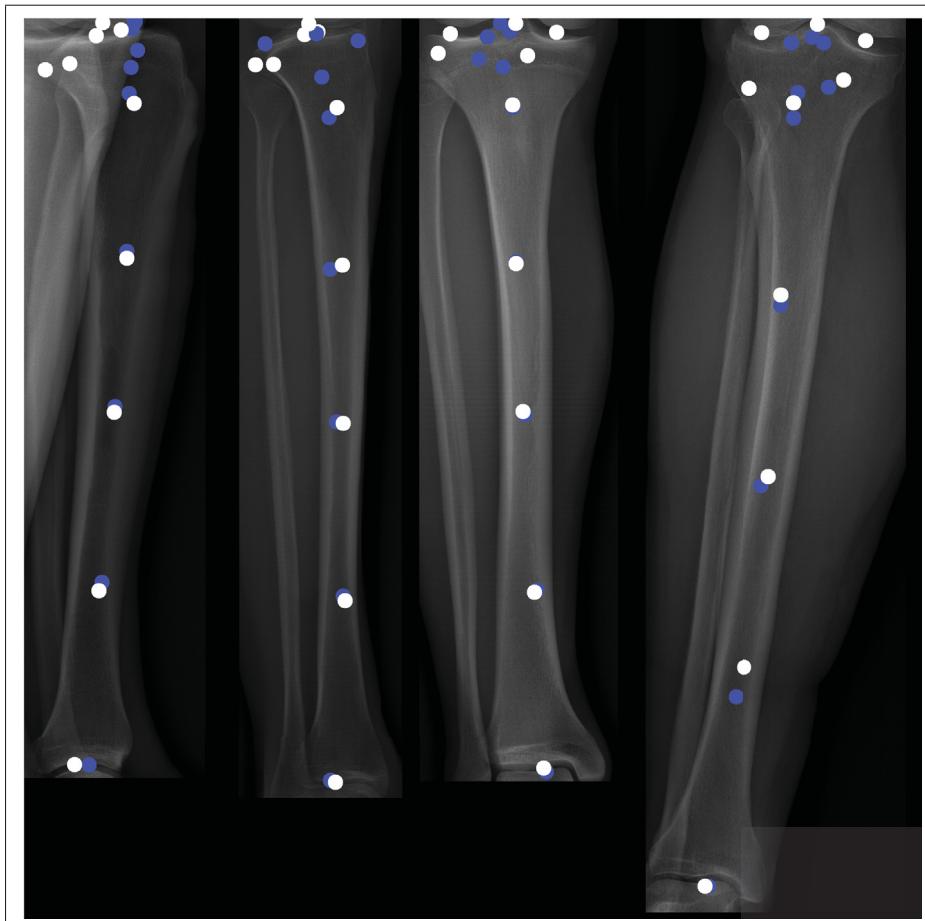


Figure 5.12 Images présentant de fortes erreurs de localisation de points d'intérêt stéréo-correspondants du tibia dans les radiographies EOS. Les points en blanc représentent la vérité terrain tandis que les points en bleu représentent les points prédits

Tableau 5.8 Erreurs (mm) par point pour le fémur (cas où uniquement boîtes englobantes considérées comme entrée).

TF (centre de la tête fémorale), GT(grand trochanter), PT (petit trochanter), Tc (trochlée), CPE (CondylePostExt), CPI (CondylePostInt), CIE (CondyleInfExt), CII (CondyleInfInt), PD (PointDia), SD₁ (SplineDia-1), SD₂ (SplineDia-2), SD₃ (SplineDia-3), SD₄ (SplineDia-4)

	TF	GT	PT	Tc	CPE	CPI	CIE	CII	PD	SD ₁	SD ₂	SD ₃	SD ₄	Moy	
Face	moy	9.16	12.24	9.18	8.48	9.84	11.07	10.75	11.54	9.48	7.25	6.25	6.26	7.62	9.16
	std	6.17	7.55	5.97	4.52	7.22	6.17	6.58	5.08	7.11	6.32	5.36	5.45	4.56	6.00
	min	1.80	1.73	0.65	2.24	1.54	1.64	1.01	3.35	1.95	0.33	0.56	0.60	1.30	1.44
	max	26.12	26.95	21.81	19.44	32.41	25.81	29.46	31.18	28.52	22.42	19.53	22.21	19.27	25.01
Point 2D	moy	13.78	17.87	13.85	11.26	13.12	14.06	11.88	12.11	11.92	10.15	6.99	6.41	8.51	11.69
	std	7.92	9.89	6.35	5.56	8.40	10.32	5.99	7.04	7.95	6.00	5.24	5.02	4.62	6.95
	min	2.44	3.36	3.65	5.12	3.10	0.60	3.01	1.93	0.69	1.54	0.37	0.70	2.26	2.21
	max	35.96	40.70	27.76	26.32	39.73	44.98	26.80	30.78	29.07	22.52	18.82	22.03	19.07	29.58
Point 3D	moy	16.24	21.05	15.93	12.38	15.51	16.40	14.66	15.11	14.96	11.91	8.44	7.92	10.20	13.90
	std	8.45	10.55	6.72	5.61	9.24	10.50	6.74	7.14	8.68	6.96	5.93	5.54	5.03	7.47
	min	2.97	3.40	5.43	5.15	3.17	1.81	6.07	5.96	3.38	2.06	0.60	1.03	2.27	3.33
	max	37.16	45.36	28.07	26.34	39.74	44.98	31.91	32.73	33.87	27.28	19.89	22.38	23.57	31.79

Tableau 5.9 Erreurs (mm) par point pour le tibia (cas où uniquement boîtes englobantes considérées comme entrée).

PPI (PlateauPostInt), PPE (PlateauPostExt), PI (PlateauInt), PE (PlateauExt), SDT (SplineDiaT), SDT₁ (SplineDiaT-1), SDT₂ (SplineDiaT-2), SDT₃ (SplineDiaT-3), SDT₄ (SplineDiaT-4), CE(CentreEpines)

	PPI	PPE	PI	PE	SDT	SDT ₁	SDT ₂	SDT ₃	SDT ₄	CE	Moy	
Face	moy	10.91	11.21	11.03	11.12	9.43	7.81	6.02	7.81	8.36	9.01	9.17
	std	6.05	5.85	6.89	5.86	5.80	4.83	4.34	3.95	4.82	5.04	5.34
	min	2.18	1.51	0.51	0.41	2.03	0.61	0.53	1.19	1.63	1.54	1.21
	max	26.40	30.70	27.10	30.49	23.55	15.84	16.59	19.89	21.56	18.65	23.08
Point 2D	moy	11.99	12.02	10.00	9.78	9.63	8.49	6.78	8.24	10.66	9.48	9.71
	std	7.93	8.21	5.36	4.94	5.44	4.31	4.09	3.98	5.42	4.64	5.43
	min	1.96	2.73	1.41	2.60	1.19	2.69	0.91	1.21	1.41	1.89	1.80
	max	30.78	33.49	21.84	24.25	23.64	18.04	16.36	19.25	21.63	18.83	22.81
Point 3D	moy	14.37	14.92	12.61	12.60	11.34	9.40	7.47	9.16	12.08	10.64	11.46
	std	7.98	9.04	6.87	6.18	5.56	4.49	4.28	4.03	5.74	4.95	5.91
	min	2.95	3.06	1.95	2.63	2.19	2.69	1.00	1.71	2.49	2.14	2.28
	max	30.87	39.22	31.51	32.02	23.65	18.28	16.60	19.94	23.26	18.83	25.42

CHAPITRE 6

DISCUSSION ET CONTRIBUTIONS DE LA THÈSE

Nous avons développé dans le cadre de cette thèse une méthode entièrement automatique permettant d'extraire des informations (masque de segmentation et points d'intérêt stéréo-correspondants) liées aux membres inférieurs dans des images EOS présentant différents points de vue et orientations. Nous avons pour cela développé une nouvelle méthode en deux phases. Durant la première phase, les structures osseuses des membres inférieurs sont segmentées via l'utilisation de notre architecture neuronale modifiée appelée RobustNet. Durant la deuxième phase, des points d'intérêt stéréo-correspondants présents sur chacune des structures osseuses segmentées sont localisés au moyen d'un réseau siamois modifié à cet effet. Dans la suite, une discussion sera effectuée pour chacune des phases de notre méthode. Les différentes contributions de cette thèse seront également présentées.

6.1 Discussion

6.1.1 Discussion sur la méthode d'identification et de segmentation de structures osseuses

Nous avons proposé une méthode entièrement automatique pour segmenter les fémurs et les tibias dans les images EOS présentant des champs de vision arbitraires (images du corps entier, images des membres inférieurs et image contenant une structure partielle) et différentes orientations du patient (0° , 20° , 45° , 70° et 90°). La méthode proposée est robuste (robuste au changement d'intensités, au chevauchement de tissus mous et à la superposition de structures osseuses), fonctionne sur des acquisitions EOS présentant différents champs de vue et orientations du patient (Figure 5.1 et Figure 5.2) et devrait être applicable à différents protocoles d'acquisition EOS (par exemple le bassin avec les fémurs proximaux, la pangonométrie). Elle est capable de distinguer les zones de superposition de structures osseuses, fonctionne bien sur les images contenant des structures partielles et parvient même à identifier et à segmenter les petites structures osseuses présentes dans l'image (Figure 5.3). Dans le cas où une structure osseuse du

membre inférieur ne serait pas présente dans l'image, notre méthode ne l'identifie pas (Figure 5.2) ; ce qui représente un avantage majeur, comparé aux méthodes basées sur les SSM qui obligent l'identification de la structure d'intérêt dans l'image même quand elle n'est pas présente. Nous avons obtenu de très bons résultats de segmentation sur l'entièreté de notre ensemble de test (voir Table 5.1, Figure 5.1 et Figure 5.1) avec un taux de détection de 98.00% et un coefficient de Dice moyen de 95.25%. Par contre, les résultats obtenus sur l'ensemble de test d'images latérales en orientations 45°, 70° et 90° (Tableau 5.3) sont inférieurs (4% en moyenne pour T_d et 3.13% en moyenne pour D_c) à ceux obtenus sur l'ensemble de données d'images frontales en orientation 0° et 45° (Tableau 5.2) principalement parce qu'ils sont plus complexes à traiter en raison de la superposition des structures osseuses. Dans certains cas (Figure 5.4), la segmentation contient des erreurs significatives, en particulier pour les images avec de fortes superpositions osseuses (Figure 5.4 troisième et quatrième colonnes) peut-être parce qu'elles sont sous-représentées dans notre ensemble de données (6 images sur 220).

Une comparaison entre différentes techniques d'augmentation de données est effectuée afin de mesurer l'impact qu'a le choix de l'ensemble de données d'entraînement sur le résultat de prédiction de RobustNet. En utilisant un ensemble de données d'images ne contenant que des structures partielles (116829 images) ou que des structures complètes (15480 images) à l'entrée de RobustNet, nous observons de nombreuses erreurs de classification dans le résultat final (Figure 5.5). En utilisant par contre un ensemble de données contenant à la fois des structures partielles et complètes (132309 images) comme entrée de RobustNet, nous observons une amélioration significative (Tableau 5.4) avec un Dice moyen de 95.25% versus 93.35% utilisant uniquement des images contenant des structures partielles versus 83.20% utilisant uniquement des images contenant des structures complètes. L'écart observé entre les résultats pourrait également s'expliquer par la différence de taille entre les ensembles de données. Ces résultats (95.25% vs 93.35% vs 83.20%) sont significativement différents avec une p-value < 0.0001 en utilisant le test de Wilcoxon (Tableau 5.4).

RobustNet a été comparé à SegNet, et certains résultats sont présentés à la Figure 5.7. Globalement, RobustNet est équivalent à SegNet (T_d de 98% contre 96.86%, D_c de 95.25 ± 9.02% contre

$94,39 \pm 12.07\%$, p-value 0,78) sur l'ensemble de données de test (Tableau 5.5). La Figure 5.8 présente pour les mêmes images les erreurs de segmentation de SegNet corrigées par RobustNet. Des améliorations significatives d'un point de vue applicatif peuvent être observées sur certaines parties spécifiques (Figure 5.8). Par exemple, l'un des paramètres cliniques importants à identifier en orthopédie est le point du grand trochanter. On peut voir sur les Figures 5.8(a) et (d) certains cas où les grands trochanters sont mal classés par SegNet ce qui implique une mauvaise caractérisation de ces derniers. Aussi, certains cas où l'image segmentée contient deux fois la même structure comme sur la Figure 5.8(b) peuvent être observés en utilisant SegNet. Cela conduit à une mauvaise caractérisation des structures. Cependant, ces erreurs de classification ont été corrigées avec RobustNet ; ce qui représente un avantage majeur, étant donné que la méthode a été développée pour une utilisation en orthopédie.

Les méthodes de segmentation utilisées dans la littérature ne sont pas appliquées à un ensemble de données aussi diversifié que le nôtre peut-être parce que ces méthodes sont développées pour une application spécifique ne nécessitant pas un ensemble varié de données. Cela rend difficile la comparaison de nos résultats avec plusieurs autres travaux de la littérature ayant les mêmes applications que les nôtres. Nous pouvons comparer les résultats obtenus sur notre ensemble de test d'images frontales à celui de Guillen *et al.* (2018) qui ont utilisé des modèles statistiques de forme et d'apparence pour segmenter le fémur proximal sur des radiographies pelviennes frontales antéro-postérieures. Nous avons obtenu un meilleur indice de Dice moyen (indice de Dice moyen de 97% vs 80%). En comparant les résultats obtenus sur l'ensemble de test d'images frontales avec ceux de Kwolek *et al.* (2019) qui segmente les os (fémur et tibia) des membres inférieurs sur des images frontales antéro-postérieures à partir de l'architecture U-Net, nous avons obtenu un résultat similaire de segmentation (indice de Dice moyen de 96.9% vs 96.8%). Il est à noter que dans (Kwolek *et al.*, 2019) les os des membres inférieurs sont segmentés, mais pas identifiés (c'est-à-dire que toutes les structures osseuses ont la même classe) contrairement à nous. Shen *et al.* (2020) ont développé l'architecture PDRU-Net, basée sur U-Net avec incorporation de la convolution dilatée et de blocs résiduels, pour segmenter les os des membres inférieurs sur des images radiographiques de face. En comparant nos résultats à

ceux de Shen *et al.* (2020), nous obtenons un résultat similaire (average Dice index of 96.86% vs. 97.26%). Il est à noter que dans (Shen *et al.*, 2020), uniquement des images en vue frontale sont traitées tandis que dans notre cas, des images présentant différents points de vue et orientations sont traitées. De plus dans (Shen *et al.*, 2020), les os des membres inférieurs sont segmentés, mais pas tous identifiés (le fémur gauche et le fémur droit ont la même classe de même que le tibia gauche et le tibia droit) contrairement à ce travail. Comparant nos résultats à notre précédent travail (Olory Agomma *et al.*, 2019) basé sur SegNet et ne tenant pas compte des superpositions de structures osseuses, nous avons également obtenu un meilleur résultat de segmentation (indice de Dice moyen $95.25 \pm 9\%$ vs $90.54 \pm 7\%$). En effet, dans notre précédent travail (Olory Agomma *et al.*, 2019), les pixels appartenant aux zones de chevauchement de deux classes ne pouvaient avoir que la valeur d'une des deux classes superposées. Ce problème a été résolu dans ce travail, ce qui implique qu'un pixel dans l'image puisse désormais appartenir à deux classes.

6.1.2 Discussion sur la méthode de localisation et d'identification de points d'intérêt

Nous avons proposé une nouvelle méthode entièrement automatique pour localiser et identifier des points d'intérêt stéréo-correspondants dans les images EOS présentant différents champs de vue (images du corps entier, images des membres inférieurs) et différentes orientations du patient ($0^\circ/90^\circ$ et $45^\circ/45^\circ$). L'approche développée est robuste et fonctionne sur différentes acquisitions EOS. Elle extrait dans un premier temps des paires de boîtes englobantes contenant des paires de structures osseuses des membres inférieurs puis identifie automatiquement l'emplacement des points d'intérêt présents dans chacune des paires extraites. Il est difficile de comparer nos résultats à ceux de la littérature, car à notre connaissance, il n'existe pas à ce jour dans la littérature de méthodes de localisation et d'identification des points d'intérêt stéréo-correspondants du membre inférieur sur des paires de radiographies présentant différentes orientations (images en orientation $0^\circ/90^\circ$ et $45^\circ/45^\circ$) comme la nôtre.

Plusieurs travaux de localisation de points d'intérêt ont été effectués sur des images IRM ou CT. Baek, Wang, Song, Lee, Lee & Koo (2013) ont obtenu pour la localisation de points d'intérêt du

fémur proximal et distal, une erreur moyenne comprise entre 2.80 ± 1.76 mm (erreur moyenne minimale) et 5.93 ± 3.77 mm (erreur moyenne maximale) sur des images CT en utilisant une méthode basée sur les modèles statistiques de formes tandis que Yang *et al.* (2015) obtiennent une erreur moyenne de 4.69 ± 2.30 mm pour la localisation des points d'intérêt du fémur proximal dans les images IRM, en utilisant un réseau de neurones convolution. Ces travaux ont été effectués uniquement sur le fémur.

Parmi les travaux traitant des membres inférieurs sur des radiographies, nous avons Liu *et al.* (2020), qui utilisent un réseau de neurones convolutionnel pour détecter des points d'intérêt de la hanche avec une erreur moyenne de 2.10 mm sur des radiographies antéro-postérieures de la hanche. Bier *et al.* (2019), pour leur part, localisent des points d'intérêt du bassin sur des images du bassin en vue frontale présentant différents points de vue grâce à un réseau de neurones convolutionnel. Ils obtiennent une erreur de 5.6 ± 4.5 mm sur des images synthétiques et une erreur moyenne de 18.9mm sur des radiographies réelles. Les images utilisées dans ce travail ne présentent par contre pas de superposition de structures osseuses comme dans notre cas.

Nous pouvons comparer nos résultats 2D (sur nos 23 points d'intérêt) obtenus sur les radiographies EOS du MI et du corps entier en orientation 0° , 45° et 90° à ceux de Tsai (2021) qui utilise un CNN pour prédire la position 2D du centre de la tête fémorale, du centre du condyle fémoral médial et du centre du plafond tibial dans des radiographies EOS du MI en orientation 0° . Nous obtenons une erreur moyenne plus grande (6.16 mm vs 3.6 mm) mais un écart-type plus faible (4.55 mm vs 9.9 mm) que Tsai (2021). Cela indique que les erreurs d'identification obtenues par Tsai (2021) sont plus dispersées par rapport à leur erreur moyenne contrairement à nous.

Dans (Olory Agomma *et al.*, 2019), nous avons développé une méthode automatique permettant de localiser treize points d'intérêt du membre inférieur sur des images EOS. Bien que ce travail n'ait été effectué que sur des radiographies EOS de face, il donne une plus grande erreur d'identification (28 ± 18 mm vs 6.16 ± 4.55 mm) comparé à ce présent travail qui exploite les informations se trouvant à la fois dans les images de face et de profil EOS pour identifier 23 points d'intérêt. Les erreurs les plus grandes observées (erreur moyenne en 2D de 9.64 ± 7.07 mm

pour CondyleInfExt, 9.28 ± 6.76 pour CondylePostExt, 10.53 ± 8.7 mm pour PlateauPostExt et 8.09 ± 5.51 pour PlateauExt) dans ce présent travail se situent dans la région du genou au niveau des condyles pour le fémur et au niveau des plateaux pour le tibia. Cela pourrait s'expliquer du fait de la proximité des structures dans cette région (Figure 6.1) et du fait que les structures se touchent ou parfois se chevauchent. De plus, la grande variabilité existant dans cette région tant au niveau de l'espacement interosseux, de l'intensité, de la texture pourrait rendre difficile l'apprentissage automatique de caractéristiques communes. Il arrive également dans quelques cas que certains points prédits se retrouvent à l'extérieur de la structure osseuse d'intérêt (Figure 5.11). Néanmoins, les points sont localisés à proximité des points de vérités terrain (Figure 5.9 et Figure 5.10) et les résultats obtenus sont comparables à ceux de la littérature traitant des structures des membres inférieurs dans les radiographies. Afin d'évaluer l'effet de la contrainte imposée par le masque sur le résultat, une comparaison des résultats obtenus, en considérant la boîte englobante avec masque et la boîte englobante sans masque comme entrée du réseau, a été faite. Nous avons obtenu une erreur moyenne de 6.16 ± 4.55 mm en 2D et de 8.14 ± 5.38 mm en 3D dans le cas où la boîte englobante et le masque sont considérés et une erreur moyenne de 9.93 ± 5.94 mm en 2D et 12.68 ± 6.69 mm en 3D dans le cas où uniquement la boîte englobante est considérée. L'apport du masque a donc permis de réduire l'erreur moyenne de localisation d'environ 4 mm en 2D et d'environ 5 mm en 3D.

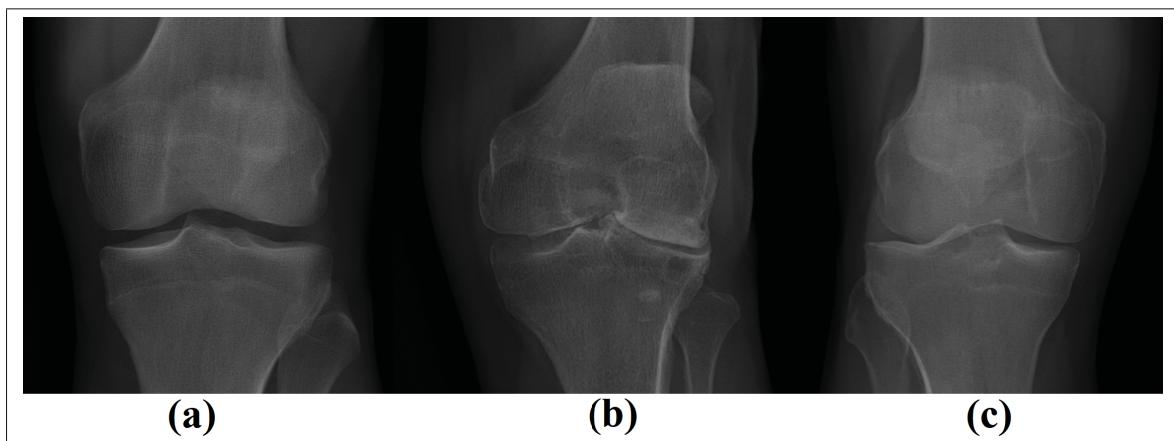


Figure 6.1 Images de genoux sain et arthrosique. (a) genou sain, (b) genou arthrosique avec structures se touchant, (c) genou arthrosique avec structures se chevauchant

La localisation et l'identification de ces points permettent une initialisation automatique des méthodes semi-automatiques de mesures cliniques et de reconstruction 3D actuellement utilisées en pratique clinique sur les radiographies EOS ; ce qui permet d'accélérer le temps d'exécution de ces méthodes.

6.2 Contributions de la thèse

Dans ce travail de thèse, nous proposons une approche automatique d'extraction d'informations dans les radiographies EOS applicable dans le cadre de l'application EOS. L'approche proposée est développée de sorte à fonctionner, selon certains points de vue et orientations des images (images frontales et sagittales du membre inférieur et de corps entier orientées à 0°, 90° et 45° avec le patient ayant les pieds joints ou décalés). Cela représente une nouveauté comparativement aux méthodes existantes (Guillen *et al.*, 2018 ; Kwolek *et al.*, 2019 ; Tsai, 2021) qui traitent spécifiquement un point de vue ou une orientation donnée d'images. La mise en place de cette approche est rendue possible grâce aux approches d'apprentissage machine.

Dans cette thèse, nous proposons une nouvelle méthode pouvant segmenter automatiquement des structures osseuses des membres inférieurs dans les radiographies EOS. Cette méthode est basée sur une nouvelle architecture neuronale (basée sur SegNet) que nous avons nommée RobustNet. L'architecture RobustNet tire pleinement avantage de l'information contextuelle grâce à l'intégration de blocs contextuels et la fusion de caractéristiques à plusieurs niveaux. L'amélioration de l'information contextuelle par RobustNet a permis d'éviter les erreurs de segmentation observées avec SegNet (Figure 5.8). La nouvelle architecture combinée avec une stratégie de création d'un jeu de données d'images contenant différents points de vue (images à structures partielles ou complètes) et orientations, permet d'apprendre un grand nombre de configurations d'images afin de mieux généraliser à tout type d'image. La méthode de segmentation proposée fonctionne sur des images avec une grande superposition d'os et est capable de distinguer les structures gauches des structures droites (par exemple, elle distingue le fémur gauche du fémur droit). Dans le cas de la superposition de deux structures osseuses (exemples : fémur gauche et fémur droit), nous identifions des parties superposées dont les

pixels appartiennent aux deux structures simultanément en les traitant comme de nouvelles classes. Même dans le cas d'images contenant des structures partielles ou de petits morceaux de structures osseuses, la méthode proposée parvient à les identifier (Figure 5.3). Ce type d'application (segmentation automatique des os du MI sur les images présentant différents points de vue et orientations) est une première tant pour la technologie EOS, que pour les méthodes déjà existantes.

Cette thèse propose également une nouvelle méthode de localisation et d'identification de points d'intérêt 3D des membres inférieurs dans les radiographies EOS de membre inférieur et de corps entier en orientation $0^\circ/90^\circ$ et $45^\circ/45^\circ$. La méthodologie proposée repose sur l'utilisation d'un réseau neuronal siamois capable d'apprendre le lien existant entre des paires d'images EOS frontales et sagittales pour prédire la position de points tridimensionnels stéréo-correspondants (points visibles sur les 2 vues); ce qui est inhabituel étant donné que dans la littérature, les réseaux siamois sont utilisés pour retrouver le degré de similitudes entre des images.

La nouvelle approche d'identification et de segmentation des os, de localisation et d'identification de points 3D proposée permet d'initialiser automatiquement les différents processus semi-automatiques du MI implantés dans la plateforme logiciel de la compagnie EOS Imaging.

CONCLUSION ET RECOMMANDATIONS

7.1 Conclusion

Afin de répondre à la problématique de la diversité d’algorithmes existants applicables à une structure spécifique des MI (exemple : fémur) ou à un type donné d’image EOS (exemple : image antéro-postérieure), nous avons développé une approche unificatrice permettant d’extraire des informations des MI sur différents types d’images EOS présentant différents points de vue et orientations du patient à savoir : les images EOS des membres inférieurs, de corps entier et les images EOS contenant des structures partielles avec le patient ayant les pieds décalés ou joints en orientation 0° , 45° et 90° .

La méthode de résolution adoptée dans le cadre de ce projet de thèse a consisté à segmenter dans un premier temps les structures osseuses des membres inférieurs, puis à identifier les points d’intérêt présents sur chacune des structures segmentées ; nous permettant ainsi de répondre à nos deux objectifs. L’approche développée est robuste au chevauchement de tissus mous, à la superposition de structures osseuses, au changement de points de vue et d’orientations et fonctionne sur différents types d’images EOS (images du corps entier et images des membres inférieurs en orientation 0° , 45° et 90°).

Nous avons développé de nouvelles architectures neuronales pour les tâches de segmentation, de localisation et d’identification de points d’intérêt des membres inférieurs. L’architecture neuronale RobustNet proposée a prouvé être similaire ou meilleure que certaines autres architectures proposées dans la littérature pour la tâche de segmentation sémantique. L’architecture neuronale RobustNet proposée a permis de segmenter les structures osseuses des membres inférieurs tout en identifiant les zones de superposition entre ces structures. L’utilisation de cette architecture, associée avec la méthode d’augmentation de données proposée, permet de segmenter de petites structures osseuses présentes dans l’image EOS et d’éviter des erreurs de classifications observées

en utilisant SegNet (Figure 5.8). Une architecture siamoise a également été proposée afin de prédire par régression la position de points d'intérêt appartenant à la fois aux images EOS de face et de profil. L'utilisation de cette architecture permet d'identifier et de pré-positionner automatiquement des points d'intérêt des MI nécessaires aux processus semi-automatiques de mesures cliniques et de reconstruction 3D avec des résultats comparables à certains travaux traitant des MI dans la littérature.

L'approche développée pourrait être utilisée pour d'autres applications hors du cadre de cette thèse comme l'identification automatique de structures osseuses, de points d'intérêt ou de régions des membres inférieurs dans des images médicales 2D (exemple : radiographies conventionnelles) autre que les radiographies EOS. Une autre application possible pourrait être d'identifier des structures osseuses ou des points d'intérêt des membres inférieurs dans les images radiographiques présentes dans les PACS (Picture Archiving and Communication System).

L'intégration de l'approche proposée dans la librairie d'analyse d'images du système EOS permettra d'automatiser la phase d'initialisation des processus de mesures cliniques et de reconstruction 3D des structures osseuses des membres inférieurs. Un transfert technologique vers la compagnie EOS a par ailleurs été fait.

7.2 Limitations et recommandations

Nous avons proposé une nouvelle méthode d'identification et de segmentation de structures osseuses, de localisation et d'identification de points d'intérêt stéréo-correspondants des membres inférieurs dans les radiographies EOS. Voici quelques limitations et directions futures en lien avec ce travail :

- L'ensemble de données présentement utilisé pourrait être plus important. En effet, dans le cadre de ce travail, nous avons accès à 190 images EOS (95 paires d'images EOS). De plus, cet ensemble de données pourrait être plus représentatif, car le nombre de paires d'images en orientation $0^\circ/90^\circ$ (71 paires d'images) est largement supérieur au nombre de paires d'images en orientation $45^\circ/45^\circ$ (23 paires d'images) qui est à son tour supérieur au nombre de paires d'images en orientation $20^\circ/70^\circ$ pratiquement inexistant (une seule paire d'images). Une augmentation de cet ensemble de données de sorte à avoir un même nombre d'images pour chaque orientation ($0^\circ/90^\circ$, $45^\circ/45^\circ$ et $20^\circ/70^\circ$) conduirait à de meilleurs résultats.
- Les vérités terrain utilisées dans ce travail sont obtenues en projetant des modèles 3D reconstruits par des experts sur les images EOS 2D. La précision de délimitation des os dans les images de vérités terrain dépend donc de la précision des reconstruction 3D effectuées. L'exploitation de vérités terrain labellisées manuellement (os segmentés directement dans les images) par des experts pour la méthode de segmentation proposée permettrait un meilleur apprentissage et une meilleure évaluation de cette dernière.
- Dans les cas d'arthrose du genou, les condyles fémoraux et les plateaux tibiaux sont partiellement superposés. La méthode de segmentation proposée n'identifie pas explicitement ces zones de superposition (ces zones sont actuellement considérées comme appartenant au tibia). Une classe pourrait être attribuée à ces zones de sorte à les identifier explicitement dans les images EOS.
- Des travaux futurs pourraient explorer l'utilisation de l'approche de segmentation développée sur des images médicales 2D autres que les images EOS.
- Dans ce travail, nous avons adopté une solution en deux étapes (segmentation puis régression) avec une méthode "diviser pour régner" ; ce qui requiert deux entraînements différents. Il serait envisageable d'explorer une méthode en une seule étape pouvant simultanément segmenter les os et prédire l'emplacement des points. Une architecture neuronale convolutionnelle

multitâches disposant de deux sorties (une sortie pour segmenter les os et l'autre pour prédire l'emplacement des points) pourrait être développée à cet effet.

- La méthode d'identification de points d'intérêt stéréo-correspondants peut être améliorée. Une piste pourrait être d'apprendre parallèlement, en plus des positions des points, la relation existante entre ces points (connectivité entre les points) grâce à un réseau multitâches. Un réseau siamois multitâches à deux branches avec une branche prédisant l'emplacement des points et la deuxième prédisant une carte de connectivité entre les points, pourrait être exploré à cet effet. Les sorties des deux branches pourraient être fusionnées de sorte que les connectivités apprises servent de contraintes lors de la prédiction de l'emplacement des points.
- La méthode d'identification de points d'intérêt proposée ne prend pas en compte les images contenant des structures partielles. L'intégration de ces images durant le processus d'entraînement pourrait être explorée. Le réseau actuel le permet. Par contre, étant donné que le nombre de points en sortie de notre réseau est fixé, ce même nombre sera également prédit par le réseau pour les images à structures partielles même si tous les points ne sont pas présents dans l'image. Les points non présents dans l'image auraient cependant des valeurs négatives.
- Ce travail pourrait être étendu à d'autres structures osseuses du corps humain comme le rachis, le bassin, la cheville ou le poignet.

7.3 Publications scientifiques

Dans cette section, la liste des différentes publications et communications réalisées dans le cadre de ce doctorat sera énumérée.

7.3.1 Revue scientifique et communications internationales

Olory Agomma, R., Vazquez, C., Cresson, T. and De Guise, J. (2021). Automatic Lower Limb Bones Segmentation in Radiographs with Different Orientations and Fields of View Based on a Contextual Network. *IEEE Transactions on Medical Imaging (TMI)*. (Article soumis).

Olory Agomma, R., Vazquez, C., Cresson, T. and De Guise, J. (2019). Detection and Identification of Lower-Limb Bones in Biplanar X-Ray Images with Arbitrary Field of View and Various Patient Orientations. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 1174–1177.

Olory Agomma, R., Vazquez, C., Cresson, T. and De Guise, J. (2018). Automatic detection of anatomical regions in frontal x-ray images: comparing convolutional neural networks to random forest. In *Medical Imaging 2018 : Computer-Aided Diagnosis* (Vol. 10575, p. 105753E). International Society for Optics and Photonics.

Olory Agomma, R., Vazquez, C., Cresson, T. and De Guise, J. (2018). Automatic recovery of EOS images' orientation via a convolutional neural network. In *2018 39th Orthopaedic World Congress*. International Society of Orthopedic Surgery and Traumatology. 10-13 October 2018, Montréal.

7.3.2 Congrès et communications locaux

Olory Agomma, R., Vazquez, C., Cresson, T. and De Guise, J. (2020). L'intelligence artificielle au service de l'imagerie orthopédique. *Substance ÉTS*. Juin 2020.

Olory Agomma, R., Vazquez, C., Cresson, T. and De Guise, J. (2020) Identification of lower-limb bones in x-ray images with arbitrary field of view and various patient orientations. *Journée*

académique du département de radiologie radio-oncologie et médecine nucléaire de l'Université de Montréal.

Olory Agomma, R., Vazquez, C., Cresson, T. and De Guise, J. (2019) Identification des os du membre inférieur dans les radiographies biplanaires présentant des champs de vision arbitraires et diverses orientations du patient. 39e journée de recherche du POES (programme orthopédique Édouard-Samson) et de la division d'orthopédie de l'Université de Montréal. Avril 2019.

Olory Agomma, R., Vazquez, C., Cresson, T. and De Guise, J. (2018). Classification automatique d'images radiographiques à l'aide d'un réseau de neurones convolutif. 20e Congrès des étudiants, stagiaires et résidents du CRCHUM. Mai 2018.

Olory Agomma, R., Vazquez, C., Cresson, T. and De Guise, J. (2018). Assister les chirurgiens au moyen de l'intelligence artificielle. Substance ÉTS. Juillet 2018.

Olory Agomma, R., Vazquez, C., Cresson, T. and De Guise, J. (2019). Méthode automatique pour la localisation et l'identification de structures osseuses dans les radiographies EOS. MEDTEQ.

Olory Agomma, R., Vazquez, C., Cresson, T. and De Guise, J. (2017). Détection automatique de régions anatomiques du membre inférieur dans les radiographies frontales basée sur un réseau de neurones convolutionnel. Congrès de l'APIBQ (Association des Physiciens et Ingénieurs Biomédicaux du Québec). Septembre 2017.

Olory Agomma, R., Vazquez, C., Cresson, T. and De Guise, J. (2017). Détection automatique de régions anatomiques dans les radiographies frontales basée sur un réseau de neurones convolutionnel. 37e journée de recherche du POES (programme orthopédique Édouard-Samson) et de la division d'orthopédie de l'Université de Montréal. Mai 2017.

7.3.3 Prix et distinction

- Lauréat du programme de reconnaissance Palmarès Féminin pluriel, ÉTS. 2021.
- Bourse CREATE, McGill University. 2018-2021.
- Bourse Substance de diffusion de la recherche, Substance ÉTS. 2020.
- Reconnaissance de l'engagement étudiant Génie+, ÉTS. 2017-2020.
- Bourse de la Fondation Arbour. 2018-2019.
- Bourse pour le rayonnement de la recherche pour les congrès SPIE 2018 et ISBI 2019. ÉTS. 2018-2019.
- 1er prix, "Une assiette exemplaire pour une terre prospère", Savoir Affaires. Compétition d'équipe. 2019.
- Prix Steinberg Center Simnovation, McGill Clinical Innovation. Compétition de startup. 2019.
- 1er prix, Programme d'innovation en chirurgie, Centech, ÉTS et McGill. Compétition d'équipe. 2019.
- 3ème prix, Dobson Practicum. Concordia University. Compétition de startup. 2019.
- 2ème prix, Dobson Cup. McGill University. Compétition de startup. 2019.
- Bourse d'excellence de la Fondation Desjardins. 2018.
- 1er prix à l'ÉTS et du 2ème prix à l'international, Les 24h d'innovation. Compétition d'équipe. 2018.
- 3ème prix du concours d'affiches du congrès de l'APIBQ. Septembre 2017.
- 3ème prix du Concours de résumés d'articles du Sara. Décembre 2017.

ANNEXE I

IDENTIFICATION DE POINTS D'INTÉRÊT DANS LES IMAGES EOS VIA L'ALGORITHME ASIFT ET LA TRANSFORMÉE DE HOUGH

Nous nous intéressons ici à la détection automatique du centre de la tête fémorale par recalage avec l'algorithme ASIFT et la transformée de Hough dans les radiographies EOS. Nous avons donc effectué une expérience basée sur les points ASIFT et la transformée de Hough dans le but de détecter le centre de la tête fémorale sur les radiographies EOS. L'objectif de cette combinaison vient du fait que, lorsque nous appliquons l'algorithme ASIFT durant la phase de mise en correspondance certaines fausses correspondances persistent. D'où l'idée d'ajouter la transformée de Hough afin de réduire ses fausses correspondances en retrouvant la transformée ayant le plus grand score. Cela permettra par la suite de calculer la matrice de transformation permettant de retrouver la position du centre de la tête fémorale sur l'image de test. Nous avons utilisé une base de données de 180 images frontales EOS, corps entier et membre inférieur, divisée en portion 3/4 et 1/4 soit 135 pour la base d'entraînement et 45 pour la base de test.

Étape 1 : Opérations sur les images d'entraînement

Nous générerons pour chaque image d'entraînement, les descripteurs ASIFT que nous stockons dans un fichier.

Étape 2 : Choix de l'image de training ayant beaucoup de similarité avec l'image de test

Pour chaque image de test nous générerons les descripteurs ASIFT. Nous faisons une mise en correspondance des descripteurs de l'image de test avec ceux de toutes les images de training et nous choisissons l'image de training à laquelle elle ressemble le plus. Le critère de ressemblance est le nombre de correspondances ASIFT. L'image d'entraînement ayant le plus fort taux de mise en correspondance avec l'image de test est donc retenue.

Étape 3 : Transformée de Hough appliquée aux mises en correspondance ASIFT pour la détection du centre de la tête fémorale

Voici les différentes étapes qui ont été suivies :

- Utiliser la transformée de Hough afin de supprimer les mauvaises correspondances obtenues par ASIFT
 - Pour chaque paire de points clés correspondants (x_1, y_1) et (x_2, y_2) , calculer les paramètres $d_x, d_y, d_\theta, d\sigma$ de la transformée de Hough où : $dx = x_1 - x_2, dy = y_1 - y_2, d\sigma = \sigma_1/\sigma_2, d\theta = \theta_1 - \theta_2$.
 - Comparer chaque transformée aux autres avec une marge d'erreur $(dx_1 - dx_2)^2 + (dy_1 - dy_2)^2 < dt_{max}^2, \log(\sigma_1/\sigma_2) < d\sigma_{max}, \log(\theta_1 - \theta_2) < d\theta_{max}$
 - Choisir celle qui a le plus grand score en comptant pour chaque transformée le nombre de transformées avec lesquelles elle correspond.
 - Supprimer les correspondances dont les transformées de Hough divergent énormément par rapport à la meilleure transformée choisie.
- Avec la transformée ayant le plus grand score, calculer la matrice de transformation linéaire
- Appliquer la transformation inverse au centre (x, y) du fémur de l'image d'entraînement pour obtenir le centre (x_1, y_1) de l'image de test.

Nous pouvons remarquer dans la Figure I-1 la ligne verte qui fait correspondre les deux centres (prédiction) et la ligne bleue qui représente la vérité terrain (ground truth). Au terme de cette expérience, nous obtenons 17 fausses détections sur 45 avec une erreur de $94.86 \pm 169.74mm$ (Figure I-2).

Au terme de cette expérience, nous nous sommes rendu compte que l'information de position (structure osseuse) doit être ajoutée afin de maximiser l'obtention de bonnes mises en correspondances, car des caractéristiques se situant à divers endroits du corps se ressemblent énormément. Ce qui a donc été intégré dans notre nouvelle méthodologie. L'ajout d'une phase d'entraînement pour apprendre les caractéristiques présentes dans l'image et pouvoir être en mesure de les retrouver sur de nouvelles images nous a également paru indispensable. Ce qui justifie notre deuxième expérience.

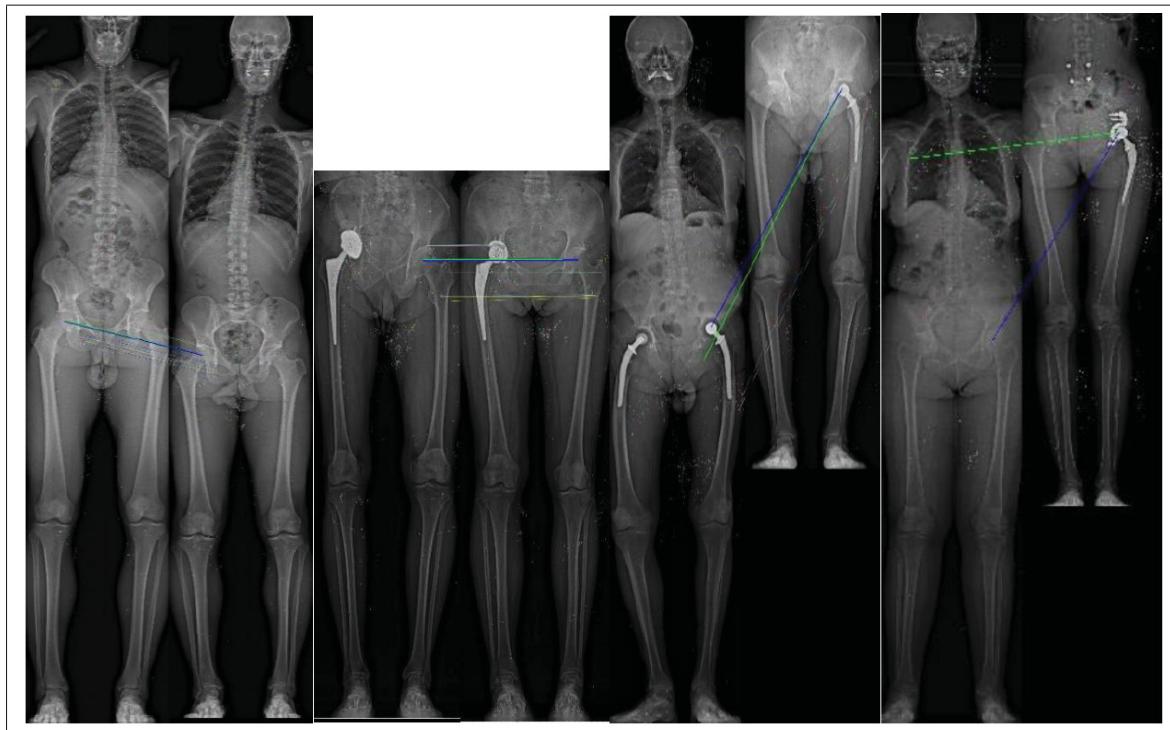


Figure-A I-1 Détection de bonnes et mauvaises correspondances
du centre de la tête fémorale

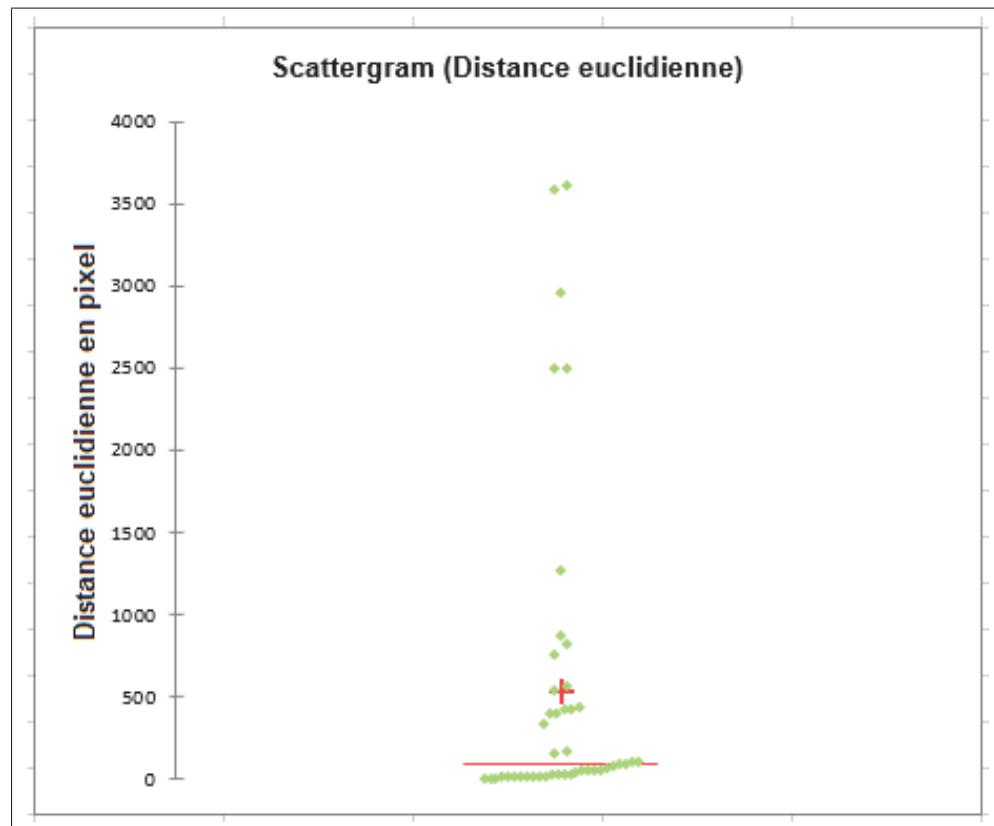


Figure-A I-2 Détection du centre de la tête fémorale sur 45 images de test

RÉFÉRENCES

- Abhyankar, A. & Schuckers, S. (2006). Active shape models for effective iris segmentation. *Biometric Technology for Human Identification III*, 6202, 62020H.
- Akbar, F. A., Maulana, H. & Yuniarti, A. (2018). Detection of Indonesian Vehicle Plate Location using Harris Corner Feature Detector Method. *International Conference on Science and Technology (ICST 2018)*, pp. 877–881.
- Akkoul Berkache, S. B. (2013). *Reconstruction 3D surfacique du fémur proximal à partir de quelques radiographies*. (Thèse de doctorat, Université d'Orléans). Repéré à <https://tel.archives-ouvertes.fr/tel-01124277>.
- Alam, M., Wang, J.-F., Guangpei, C., Yunrong, L. & Chen, Y. (2021). Convolutional Neural Network for the Semantic Segmentation of Remote Sensing Images. *Mobile Networks and Applications*, 26(1), 200–215. doi : 10.1007/s11036-020-01703-3.
- Albrecht, T., Lüthi, M. & Vetter, T. (2015). Deformable Models.
- Aljabar, P., Heckemann, R. A., Hammers, A., Hajnal, J. V. & Rueckert, D. (2009). Multi-atlas based segmentation of brain images : atlas selection and its effect on accuracy. *Neuroimage*, 46(3), 726–738. doi : 10.1016/j.neuroimage.2009.02.018.
- Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.
- Alves, M. & Lemberger, P. (2018). Le Deep Learning pas à pas. Accessed : 2021-03-18, Repéré à <https://www.technologies-ebusiness.com/enjeux-et-tendances/le-deep-learning-pas-a-pas>.
- Amrapur, B. et al. (2019). Active Contour Model Applied to Segmentation of Human Face. Dans *Innovations in Electronics and Communication Engineering* (pp. 31–36). Springer. doi : 10.1007/978-981-10-8204-7_4.
- Andriluka, M., Pishchulin, L., Gehler, P. & Schiele, B. (2014). 2D Human Pose Estimation : New Benchmark and State of the Art Analysis. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3686-3693.
- Ariji, Y., Yanashita, Y., Kutsuna, S., Muramatsu, C., Fukuda, M., Kise, Y., Nozawa, M., Kuwada, C., Fujita, H., Katsumata, A. et al. (2019). Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique. *Oral surgery, oral medicine, oral pathology and oral radiology*, 128(4), 424–430. doi : 10.1016/j.oooo.2019.05.014.

- ASPC. (2020). Larthrose au Canada. Accessed : 2021-03-18, Repéré à <https://www.canada.ca/fr/sante-publique/services/publications/maladies-et-affections/arthrose.html>.
- Aubert, B., Vazquez, C., Cresson, T., Parent, S. & de Guise, J. A. (2019). Toward automated 3D spine reconstruction from biplanar radiographs using CNN for statistical spine model fitting. *IEEE transactions on medical imaging*, 38(12), 2796–2806. doi : 10.1109/TMI.2019.2914400.
- Ayed, I. B., Chen, H.-m., Punithakumar, K., Ross, I. & Li, S. (2010). Graph cut segmentation with a global constraint : Recovering region distribution via a bound of the Bhattacharyya measure. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3288-3295.
- Badrinarayanan, V., Kendall, A. & Cipolla, R. (2017). Segnet : A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12), 2481–2495. doi : 10.1109/TPAMI.2016.2644615.
- Baek, S.-Y., Wang, J.-H., Song, I., Lee, K., Lee, J. & Koo, S. (2013). Automated bone landmarks prediction on the femur using anatomical deformation technique. *Computer-aided design*, 45(2), 505–510. doi : 10.1016/j.cad.2012.10.033.
- Balntas, V., Johns, E., Tang, L. & Mikolajczyk, K. (2016). *PN-Net : Conjoined triple deep network for learning local image descriptors*.
- Baltrusaitis, T., Zadeh, A., Lim, Y. C. & Morency, L.-P. (2018). OpenFace 2.0 : Facial Behavior Analysis Toolkit. *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pp. 59-66.
- Bélouard, T., Schmuck, H., Nageleisen, L.-M. & Guyon, D. (2018). Cartographie des défoliations du massif forestier du Pays des étangs en Lorraine : Apports potentiels de la télédétection. *Revue Forestière Française*. doi : 10.4267/2042/70132.
- Bier, B., Goldmann, F., Zaech, J.-N., Fotouhi, J., Hegeman, R., Grupp, R., Armand, M., Osgood, G., Navab, N., Maier, A. et al. (2019). Learning to detect anatomical landmarks of the pelvis in X-rays from arbitrary views. *International journal of computer assisted radiology and surgery*, 14(9), 1463–1473. doi : 10.1007/s11548-019-01975-5.
- Bolya, D., Zhou, C., Xiao, F. & Lee, Y. J. (2019). YOLACT : Real-Time Instance Segmentation. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9156-9165.
- Bonacorso, G. (2017). *Machine learning algorithms*. Packt Publishing Ltd.

- Bouvie, J. (2006). Notes on convolutional neural networks. Repéré à http://cogprints.org/5869/1/cnn_tutorial.pdf.
- Breiman, L. (2001). Random Forests Machine Learning, vol. 45.
- Bromley, J., Guyon, I., LeCun, Y., Säckinger, E. & Shah, R. (1994). Signature verification using a "siamese" time delay neural network. *Advances in neural information processing systems*, 737–737. doi : 10.1142/S0218001493000339.
- Brousseau, C. (2005). *CARACTERISATION LARGE BANDE DE FREQUENCE : APPLICATION AU RADAR ET AU SONDAGE DE CANAL*. (Thèse de doctorat, Université Rennes 1). Repéré à <https://tel.archives-ouvertes.fr/tel-00085138>.
- Butakoff, C. & Frangi, A. F. (2010). Multi-view face segmentation using fusion of statistical shape and appearance models. *Computer Vision and Image Understanding*, 114(3), 311–321. doi : 10.1016/j.cviu.2009.11.001.
- Buyssens, P. & Elmoataz, A. (2016). Réseaux de neurones convolutionnels multi-échelle pour la classification cellulaire. *RFIA 2016*.
- Cai, Z., Fan, Q., Feris, R. S. & Vasconcelos, N. (2016). A unified multi-scale deep convolutional neural network for fast object detection. *European conference on computer vision*, pp. 354–370.
- Candemir, S. & Akgul, Y. S. (2011). Statistical significance based graph cut segmentation for shrinking bias. *International Conference Image Analysis and Recognition*, pp. 304–313.
- Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E. & Sheikh, Y. (2021). OpenPose : Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1), 172-186. doi : 10.1109/TPAMI.2019.2929257.
- Ceron, J. C. A., Chang, L., Ochoa-Ruiz, G. & Ali, S. (2021). Assessing YOLACT++ for real time and robust instance segmentation of medical instruments in endoscopic procedures.
- Chaibi, Y., Cresson, T., Aubert, B., Hausselle, J., Neyret, P., Hauger, O., De Guise, J. & Skalli, W. (2012). Fast 3D reconstruction of the lower limb using a parametric model and statistical inferences and clinical measurements calculation from biplanar X-rays. *Computer methods in biomechanics and biomedical engineering*, 15(5), 457–466.
- Chaibi, Y. (2010). *Adaptation des méthodes de reconstruction 3D rapides par stéréoradiographie : Modélisation du membre inférieur et calcul des indices cliniques en présence de déformation structurale*. (Thèse de doctorat, Arts et Métiers ParisTech). Repéré à <https://pastel.archives-ouvertes.fr/pastel-00524999>.

- Chartrand, G. (2017). *Segmentation 3D du foie*. (Thèse de doctorat, École de technologie supérieure). Repéré à <https://espace.etsmtl.ca/id/eprint/1861>.
- Chen, C. W., Luo, J. & Parker, K. J. (1998). Image segmentation via adaptive K-mean clustering and knowledge-based morphological operations with biomedical applications. *IEEE transactions on image processing*, 7(12), 1673–1683. doi : 10.1109/83.730379.
- Chen, C. & Zheng, G. (2014). Fully automatic segmentation of AP pelvis X-rays via random forest regression with efficient feature selection and hierarchical sparse shape composition. *Computer Vision and Image Understanding*, 126, 1–10. doi : 10.1016/j.cviu.2014.04.015.
- Chen, H., Qi, X., Yu, L., Dou, Q., Qin, J. & Heng, P.-A. (2017). DCAN : Deep contour-aware networks for object instance segmentation from histology images. *Medical image analysis*, 36, 135–146. doi : 10.1016/j.media.2016.11.004.
- Chen, H., Dou, Q., Yu, L., Qin, J. & Heng, P.-A. (2018a). VoxResNet : Deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage*, 170, 446–455. doi : 10.1016/j.neuroimage.2017.04.041.
- Chen, J., Qiang, H., Wu, J., Xu, G. & Wang, Z. (2021). Navigation path extraction for greenhouse cucumber-picking robots using the prediction-point Hough transform. *Computers and Electronics in Agriculture*, 180, 105911. doi : 10.1016/j.compag.2020.105911.
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. (2018b). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European conference on computer vision (ECCV)*, pp. 801–818.
- Chen, Y., Ee, X., Leow, W. K. & Howe, T. S. (2005). Automatic extraction of femur contours from hip x-ray images. *International Workshop on Computer Vision for Biomedical Image Applications*, pp. 200–209.
- Cheng, R., Zhang, J., Yang, P., Liu, K. & Zhang, S. (2017). CNet : context-aware network for semantic segmentation. *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 67–72.
- Chiu, S.-H., Liaw, J.-J. & Lin, K.-H. (2010). A fast randomized Hough transform for circle/circular arc recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 24(03), 457–474. doi : 10.1142/S0218001410007956.
- Ciresan, D., Giusti, A., Gambardella, L. M. & Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. *Advances in neural information processing systems*, pp. 2843–2851.

- Cohen-Hadria, A. & Peeters, G. (2016). *Apprentissage de Structure Musicale par Réseaux Neuronaux Convolutionnels*. Repéré à http://www.atiam.ircam.fr/Archives/Stages1516/COHEN-HADRIA_Alice_Rapport_Stage.pdf.
- Cootes, T. F., Taylor, C. J., Cooper, D. H. & Graham, J. (1995). Active shape models-their training and application. *Computer vision and image understanding*, 61(1), 38–59. doi : 10.1006/cviu.1995.1004.
- Courbariaux, M. (2015). *Réduire la précision et le nombre des multiplications nécessaires à l'entraînement d'un réseau de neurones*. (Thèse de doctorat, École Polytechnique de Montréal). Repéré à <https://publications.polymtl.ca/1819/>.
- Criminisi, A., Robertson, D., Pauly, O., Glocker, B., Konukoglu, E., Shotton, J., Mateus, D., Möller, A. M., Nekolla, S. & Navab, N. (2013). Anatomy detection and localization in 3D medical images. Dans *Decision Forests for Computer Vision and Medical Image Analysis* (pp. 193–209). Springer. doi : 10.1007/978-1-4471-4929-3_14.
- Cristinacce, D. & Cootes, T. (2008). Automatic feature localisation with constrained local models. *Pattern Recognition*, 41(10), 3054–3067. doi : 10.1016/j.patcog.2008.01.024.
- Delalleau, O. (2008). Extraction hiérarchique de caractéristiques pour l'apprentissagea partir de données complexes en haute dimension. *Pre-doctoral report, University of Montreal*. Repéré à <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.387.4846&rep=rep1&type=pdf>.
- Delassus, R. (2018). *Apprentissage automatique pour la détection d'anomalies dans les données ouvertes : application à la cartographie*. (Thèse de doctorat, Université de Bordeaux). Repéré à <https://tel.archives-ouvertes.fr/tel-02100741>.
- Devalla, S. K., Renukanand, P. K., Sreedhar, B. K., Subramanian, G., Zhang, L., Perera, S., Mari, J.-M., Chin, K. S., Tun, T. A., Strouthidis, N. G. et al. (2018). DRUNET : a dilated-residual U-Net deep learning network to segment optic nerve head tissues in optical coherence tomography images. *Biomedical optics express*, 9(7), 3244–3265. doi : 10.1364/BOE.9.003244.
- Dey, S., Dutta, A., Toledo, J. I., Ghosh, S. K., Lladós, J. & Pal, U. (2017). *Signet : Convolutional siamese network for writer independent offline signature verification*.
- Ding, F., Leow, W. K. & Howe, T. S. (2007). Automatic segmentation of femur bones in anterior-posterior pelvis x-ray images. *International conference on computer analysis of images and patterns*, pp. 205–212.

- Ding, M., Kang, Y., Yuan, Z., Shan, X. & Cai, Z. (2021). Detection of facial landmarks by a convolutional neural network in patients with oral and maxillofacial disease. *International Journal of Oral and Maxillofacial Surgery*. doi : 10.1016/j.ijom.2021.01.002.
- Dolz, J., Desrosiers, C. & Ayed, I. B. (2018). 3D fully convolutional networks for subcortical segmentation in MRI : A large-scale study. *NeuroImage*, 170, 456–470. doi : 10.1016/j.neuroimage.2017.04.039.
- Dong, X. & Zheng, G. (2009). Automatic extraction of proximal femur contours from calibrated X-ray images using 3D statistical models : an in vitro study. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 5(2), 213–222.
- Dosovitskiy, A., Fischer, P., Springenberg, J. T., Riedmiller, M. & Brox, T. (2016). Discriminative Unsupervised Feature Learning with Exemplar Convolutional Neural Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(9), 1734-1747. doi : 10.1109/TPAMI.2015.2496141.
- Dubousset, J., Charpak, G., Dorion, I., Skalli, W., Lavaste, F., Deguise, J., Kalifa, G. & Ferey, S. (2005). Une nouvelle imagerie ostéo-articulaire basse dose en position debout : le système EOS. *Bulletin de l'académie nationale de médecine*, 189(2), 287–300. doi : 10.1016/S0001-4079(19)33584-8.
- Eigen, D. & Fergus, R. (2015). Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-scale Convolutional Architecture. *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 2650-2658.
- Eslami, M., Tabarestani, S., Albarqouni, S., Adeli, E., Navab, N. & Adjouadi, M. (2020). Image-to-images translation for multi-task organ segmentation and bone suppression in chest x-ray radiography. *IEEE transactions on medical imaging*, 39(7), 2553–2565. doi : 10.1109/TMI.2020.2974159.
- Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. (1996). Density-based spatial clustering of applications with noise. *Int. Conf. Knowledge Discovery and Data Mining*, 240, 6.
- Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J. & Zisserman, A. (2015). The pascal visual object classes challenge : A retrospective. *International journal of computer vision*, 111(1), 98–136. doi : 10.1007/s11263-014-0733-5.
- Feng, J., Wang, X. & Liu, W. (2021). Deep graph cut network for weakly-supervised semantic segmentation. *Science China Information Sciences*, 64(3), 130105. doi : 10.1007/s11432-020-3065-4.

- Fischler, M. A. & Bolles, R. C. (1981). Random Sample Consensus : A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM*, 24(6), 381–395. doi : 10.1145/358669.358692.
- Fu, C.-Y., Shvets, M. & Berg, A. C. (2019). *RetinaMask : Learning to predict masks improves state-of-the-art single-shot detection for free*.
- Gales, G. (2011). *Mise en correspondance de pixels pour la stéréovision binoculaire par propagation d'appariements de points d'intérêt et sondage de régions*. (Thèse de doctorat, Université de Toulouse). Repéré à <https://tel.archives-ouvertes.fr/tel-00622859>.
- Galisol, G., Brouard, T. & Ramel, J.-Y. (2016). Apprentissage et exploitation d'un graphe topologique d'atlas probabilistes locaux pour la segmentation d'images IRM. *RFIA 2016*.
- Gargouri, I. (2009). *Méthodes automatiques pour la pangonométrie, la morphométrie et la reconstruction 3D des os du membre inférieur*. (Thèse de doctorat, École de technologie supérieure). Repéré à <https://espace.etsmtl.ca/id/eprint/1063>.
- Gass, T., Szekely, G. & Goksel, O. (2014). Multi-atlas segmentation and landmark localization in images with large field of view. *International MICCAI Workshop on Medical Computer Vision*, pp. 171–180.
- Girshick, R. (2015). Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1440-1448. doi : 10.1109/ICCV.2015.169.
- Groos, D., Ramampiaro, H. & Ihlen, E. A. (2021). EfficientPose : Scalable single-person pose estimation. *Applied Intelligence*, 51(4), 2518–2533. doi : 10.1007/s10489-020-01918-7.
- Guillen, J. O. G., Cerquin, L. J. E., Obando, J. D. & Castro-Gutierrez, E. (2018). Segmentation of the proximal femur by the analysis of x-ray imaging using statistical models of shape and appearance. *International Conference on Artificial Intelligence and Soft Computing*, pp. 25–35.
- Guo, J., Mu, Y., Xue, D., Li, H., Chen, J., Yan, H., Xu, H. & Wang, W. (2021). Automatic analysis system of calcaneus radiograph : Rotation-invariant landmark detection for calcaneal angle measurement, fracture identification and fracture region segmentation. *Computer Methods and Programs in Biomedicine*, 206, 106124. doi : 10.1016/j.cmpb.2021.106124.
- Han, X., Leung, T., Jia, Y., Sukthankar, R. & Berg, A. C. (2015). MatchNet : Unifying feature and metric learning for patch-based matching. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3279-3286.

- Haque, I. R. I. & Neubert, J. (2020). Deep learning approaches to biomedical image segmentation. *Informatics in Medicine Unlocked*, 18, 100297. doi : 10.1016/j.imu.2020.100297.
- Harris, C. G. & Stephens, M. (1988). A Combined Corner and Edge Detector. *Proceedings of the Alvey Vision Conference, AVC, Manchester, UK*, pp. 1–6.
- He, H., Peng, B. & Chen, J. (2019). Image Segmentation Based on Similarity of Hierarchical Regions. *2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pp. 1001-1007.
- He, K., Gkioxari, G., Dollár, P. & Girshick, R. (2017). Mask R-CNN. *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2980-2988.
- Heimann, T. & Meinzer, H.-P. (2009). Statistical shape models for 3D medical image segmentation : a review. *Medical image analysis*, 13(4), 543–563. doi : 10.1016/j.media.2009.05.004.
- Hossenboccus, M. (2021). Ostéotomies autour du genou à Périgueux. Accessed : 2021-03-18, Repéré à <http://www.dr-hossenboccus.fr/chirurgie/chirurgie-du-genou/osteotomies-autour-du-genou/>.
- Hough, P. V. (1962). Method and means for recognizing complex patterns. Google Patents. US Patent 3,069,654.
- Huang, Z., Wang, X., Wei, Y., Huang, L., Shi, H., Liu, W. & Huang, T. S. (2020). CCNet : Criss-Cross Attention for Semantic Segmentation. pp. 1-1.
- Hussain, D., Naqvi, R. A., Loh, W.-K. & Lee, J. (2021). Deep Learning in DXA Image Segmentation. *Computers, Materials & Continua*, 66(3), 2587–2598. doi : 10.32604/cmc.2021.013031.
- Illés, T. & Somoskeöy, S. (2012). The EOS™ imaging system and its uses in daily orthopaedic practice. *International orthopaedics*, 36(7), 1325–1331. doi : 10.1007/s00264-012-1512-y.
- Ioannou, Y., Robertson, D., Zikic, D., Kotschieder, P., Shotton, J., Brown, M. & Criminisi, A. (2016). *Decision forests, convolutional networks and the models in-between*.
- Iqtait, M., Mohamad, F. & Mamat, M. (2018). Feature extraction for face recognition via Active Shape Model (ASM) and Active Appearance Model (AAM). 332, 012032.
- Irrera, P. (2015). *Low dose x-ray image processing : Joint denoising and contrast enhancement, and automatic detection of anatomical landmarks for image quality estimation*. (Thèse

- de doctorat). Repéré à <https://hal.archives-ouvertes.fr/tel-01218464>.
- Jenny, J. & Boeri, C. (2004). Low reproducibility of the intra-operative measurement of the transepicondylar axis during total knee replacement. *Acta Orthopaedica Scandinavica*, 75(1), 74-77. doi : 10.1080/00016470410001708150.
- Josse, J. (1832). *Quelques considérations générales sur les fractures du corps et du col du fémur, et sur un nouveau moyen d'appliquer l'extension soutenue à leur traitement ; thèse, etc.*
- Karsh, R. K., Das, A., Swetha, G. L., Medhi, A., Laskar, R. H., Arya, U. & Agarwal, R. K. (2016). Copy-move forgery detection using ASIFT. *2016 1st India International Conference on Information Processing (IICIP)*, pp. 1–5.
- Kass, M., Witkin, A. & Terzopoulos, D. (1988). Snakes : Active contour models. *International journal of computer vision*, 1(4), 321–331. doi : 10.1007/BF00133570.
- Kayalibay, B., Jensen, G. & van der Smagt, P. (2017). *CNN-based segmentation of medical imaging data.*
- Ketenci, U. G. (2013). *Modélisation agent de la perception visuelle humaine limitée appliquée à la simulation du comportement des conducteurs en carrefour.* (Thèse de doctorat, Université de Valenciennes et du Hainaut-Cambresis). Repéré à <https://tel.archives-ouvertes.fr/tel-00879042>.
- Khanh, T. L. B., Dao, D.-P., Ho, N.-H., Yang, H.-J., Baek, E.-T., Lee, G., Kim, S.-H. & Yoo, S. B. (2020). Enhancing U-Net with Spatial-Channel Attention Gate for Abnormal Tissue Segmentation in Medical Imaging. *Applied Sciences*, 10(17), 5729.
- Kido, S., Hirano, Y. & Hashimoto, N. (2018). Detection and classification of lung abnormalities by use of convolutional neural network (CNN) and regions with CNN features (R-CNN). *2018 International workshop on advanced image technology (IWAIT)*, pp. 1–4.
- Kordon, F., Fischer, P., Privalov, M., Swartman, B., Schnetzke, M., Franke, J., Lasowski, R., Maier, A. & Kunze, H. (2019). Multi-task localization and segmentation for x-ray guided planning in knee surgery. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 622–630.
- Krawczyk, Z. & Starzyński, J. (2018). Bones detection in the pelvic area on the basis of YOLO neural network. *19th International Conference Computational Problems of Electrical Engineering*, pp. 1–4.
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25,

1097–1105.

Kuwana, R., Ariji, Y., Fukuda, M., Kise, Y., Nozawa, M., Kuwada, C., Muramatsu, C., Katsumata, A., Fujita, H. & Ariji, E. (2021). Performance of deep learning object detection technology in the detection and diagnosis of maxillary sinus lesions on panoramic radiographs. *Dentomaxillofacial Radiology*, 50(1), 20200171. doi : 10.1259/dmfr.20200171.

Kwolek, K., Brychcy, A., Kwolek, B. & Marczyński, W. (2019). Measuring Lower Limb Alignment and Joint Orientation Using Deep Learning Based Segmentation of Bones. *International Conference on Hybrid Artificial Intelligence Systems*, pp. 514–525.

Langlais, F. & Thomazeau, H. (1999). Ostéotomies du tibia proximal. *Tec Chir Traumatol Encycl Med Chir Elsevier, Paris*, 44–830.

Lebore, T. (2007). *Segmentation d'image application aux documents anciens*. (Mémoire de maîtrise, Université de Nantes, France). Repéré à <https://docplayer.fr/10786453-Segmentation-d-image.html>.

LeCun, Y., Cortes, C. & Burges, C. J. (1994). The MNIST database of handwritten digits. Repéré à <http://yann.lecun.com/exdb/mnist/>.

LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436–444. doi : 10.1038/nature14539.

Lemley, J., Bazrafkan, S. & Corcoran, P. (2017). Smart augmentation learning an optimal data augmentation strategy. *Ieee Access*, 5, 5858–5869. doi : 10.1109/ACCESS.2017.2696121.

Lenc, L. & Král, P. (2015). Automatic face recognition system based on the SIFT features. *Computers & Electrical Engineering*, 46, 256–272. doi : 10.1016/j.compeleceng.2015.01.014.

Lermé, N., Malgouyres, F., Létocart, L. & Rocchisani, J.-M. (2011). Une méthode de réduction exacte pour la segmentation par graph cuts. *ORASIS - Congrès des jeunes chercheurs en vision par ordinateur*.

Li, M. D., Chang, K., Bearce, B., Chang, C. Y., Huang, A. J., Campbell, J. P., Brown, J. M., Singh, P., Hoebel, K. V., Erdoğmuş, D. et al. (2020). Siamese neural networks for continuous disease severity evaluation and change detection in medical imaging. *NPJ digital medicine*, 3(1), 1–9. doi : 10.1038/s41746-020-0255-1.

Lianghui, F., Gang, H. J., Yang, J. & Bin, Y. (2019). Femur segmentation in X-ray image based on improved U-Net. *IOP Conference Series : Materials Science and Engineering*, 533(1), 012061.

- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. & Zitnick, C. L. (2014). Microsoft coco : Common objects in context. *European conference on computer vision*, pp. 740–755.
- Lindner, C., Thiagarajah, S., Wilkinson, J. M., Wallis, G. A., Cootes, T. F., arcOGEN Consortium et al. (2013). Fully automatic segmentation of the proximal femur using random forest regression voting. *IEEE transactions on medical imaging*, 32(8), 1462–1472. doi : 10.1109/TMI.2013.2258030.
- Lindner, C., Bromiley, P. A., Ionita, M. C. & Cootes, T. F. (2014). Robust and accurate shape model matching using random forest regression-voting. *IEEE transactions on pattern analysis and machine intelligence*, 37(9), 1862–1874. doi : 10.1109/TPAMI.2014.2382106.
- Linguraru, M. G., Sandberg, J. K., Li, Z., Shah, F. & Summers, R. M. (2010). Automated segmentation and quantification of liver and spleen from CT images using normalized probabilistic atlases and enhancement estimation. *Medical physics*, 37(2), 771–783. doi : 10.1118/1.3284530.
- LIO. (2017). SterLIO – Plugin ReconstMI – Control points. Accessed : 2019-07-30, Repéré à <https://liosp.etsmtl.ca/liblio/Home%20Page.aspx>.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B. & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42, 60–88. doi : 10.1016/j.media.2017.07.005.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y. & Berg, A. C. (2016). Ssd : Single shot multibox detector. *European conference on computer vision*, pp. 21–37.
- Liu, W., Wang, Y., Jiang, T., Chi, Y., Zhang, L. & Hua, X.-S. (2020). Landmarks Detection with Anatomical Constraints for Total Hip Arthroplasty Preoperative Measurements. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 670–679.
- Long, J., Shelhamer, E. & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431-3440.
- López, V., Fernández, A., García, S., Palade, V. & Herrera, F. (2013). An insight into classification with imbalanced data : Empirical results and current trends on using data intrinsic characteristics. *Information sciences*, 250, 113–141. doi : 10.1016/j.ins.2013.07.007.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91–110. doi : 10.1023/B:VISI.0000029664.99615.94.

- Luo, W., Li, Y., Urtasun, R. & Zemel, R. (2016). Understanding the effective receptive field in deep convolutional neural networks. *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pp. 4905–4913.
- Mackechnie-Jarvis, A. (1983). Femoral neck fracture fixation : rigidity of five techniques compared. *Journal of the Royal Society of Medicine*, 76(8), 643–648. doi : 10.1177/014107688307600805.
- Mader, A. O., von Berg, J., Fabritz, A., Lorenz, C. & Meyer, C. (2018). Localization and labeling of posterior ribs in chest radiographs using a CRF-regularized FCN with local refinement. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 562–570.
- Mammeri, A., Boukerche, A. & Lu, G. (2014). Lane Detection and Tracking System Based on the MSER Algorithm, Hough Transform and Kalman Filter. *Proceedings of the 17th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, (MSWiM '14), 259–266.
- Martin, S. (2008). *Utilisation d'atlas anatomiques et de modèles déformables pour la segmentation de la prostate en IRM*. (Thèse de doctorat, Université Joseph-Fourier - Grenoble I). Repéré à <https://tel.archives-ouvertes.fr/tel-00374342>.
- McInerney, T. & Terzopoulos, D. (1996). Deformable models in medical image analysis : a survey. *Medical image analysis*, 1(2), 91–108. doi : 10.1016/S1361-8415(96)80007-7.
- Medjedé, M. & Mahmoudi, S. (2009). Les Modèles Déformables et leurs Applications dans le Traitement des Images Médicales. *CIIA*.
- Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R. et al. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging*, 34(10), 1993–2024. doi : 10.1109/TMI.2014.2377694.
- Mesejo, P., Valsecchi, A., Marrakchi-Kacem, L., Cagnoni, S. & Damas, S. (2015). Biomedical image segmentation using geometric deformable models and metaheuristics. *Computerized Medical Imaging and Graphics*, 43, 167–178. doi : 10.1016/j.compmedimag.2013.12.005.
- Mirikharaji, Z. & Hamarneh, G. (2018). Star shape prior in fully convolutional networks for skin lesion segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 737–745.

- Mittal, A., Hooda, R. & Sofat, S. (2018). LF-SegNet : A fully convolutional encoder-decoder network for segmenting lung fields from chest radiographs. *Wireless Personal Communications*, 101(1), 511–529. doi : 10.1007/s11277-018-5702-9.
- Molchanov, P., Yang, X., Gupta, S., Kim, K., Tyree, S. & Kautz, J. (2016). Online Detection and Classification of Dynamic Hand Gestures with Recurrent 3D Convolutional Neural Networks. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4207-4215.
- Morel, J.-M. & Yu, G. (2009). ASIFT : A new framework for fully affine invariant image comparison. *SIAM journal on imaging sciences*, 2(2), 438–469. doi : 10.1137/080732730.
- Nguyen, A.-D., Choi, S., Kim, W., Ahn, S., Kim, J. & Lee, S. (2019). Distribution padding in convolutional neural networks. *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 4275–4279.
- Nie, Y., Sommella, P., O’Nils, M., Liguori, C. & Lundgren, J. (2019). Automatic detection of melanoma with yolo deep convolutional neural networks. *2019 E-Health and Bioengineering Conference (EHB)*, pp. 1–4.
- Niu, X.-X. & Suen, C. Y. (2012). A novel hybrid CNN–SVM classifier for recognizing handwritten digits. *Pattern Recognition*, 45(4), 1318–1325. doi : 10.1016/j.patcog.2011.09.021.
- Nodé-Langlois, L. (2003). *3D analysis of the misalignment fo the lower extremity, a pre, per and postoperative study*. (Thèse de doctorat, Arts et Métiers ParisTech). Repéré à <https://pastel.archives-ouvertes.fr/pastel-00000981>.
- Oliveira, H. & dos Santos, J. (2018). Deep Transfer Learning for Segmentation of Anatomical Structures in Chest Radiographs. *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 204-211.
- Olory Agomma, R., Vázquez, C., Cresson, T. & De Guise, J. (2018). Automatic detection of anatomical regions in frontal x-ray images : comparing convolutional neural networks to random forest. *Medical Imaging 2018 : Computer-Aided Diagnosis*, 10575, 105753E.
- Olory Agomma, R., Vázquez, C., Cresson, T. & de Guise, J. (2019). Detection and identification of lower-limb bones in biplanar x-ray images with arbitrary field of view and various patient orientations. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 1174–1177.
- Osher, S. & Sethian, J. A. (1988). Fronts propagating with curvature-dependent speed : Algorithms based on Hamilton-Jacobi formulations. *Journal of computational physics*, 79(1), 12–49. doi : 10.1016/0021-9991(88)90002-2.

- Ouertani, F. (2015). *Segmentation de l'articulation de la hanche à partir de radiographies biplanes en utilisant une approche multi-structures*. (Mémoire de maîtrise, École de technologie supérieure). Repéré à <https://espace.etsmtl.ca/id/eprint/1485>.
- Ouertani, F., Vazquez, C., Cresson, T. & de Guise, J. (2015). Simultaneous extraction of two adjacent bony structures in x-ray images : Application to hip joint segmentation. *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 4555-4559.
- Parizeau, M. (2004). Réseaux de neurones. *GIF-21140 et GIF-64326*, 124. Repéré à <http://seborga1.free.fr/RNF.pdf>.
- Pasquier, D., Lacornerie, T., Vermandel, M., Rousseau, J., Lartigau, E. & Betrouni, N. (2007). Automatic segmentation of pelvic structures from magnetic resonance images for prostate cancer radiotherapy. *International Journal of Radiation Oncology* Biology* Physics*, 68(2), 592–600. doi : 10.1016/j.ijrobp.2007.02.005.
- Payer, C., Štern, D., Bischof, H. & Urschler, M. (2016). Regressing heatmaps for multiple landmark localization using CNNs. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 230–238.
- Perez, F., Vasconcelos, C., Avila, S. & Valle, E. (2018). Data augmentation for skin lesion analysis. Dans *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis* (pp. 303–311). Springer. doi : 10.1007/978-3-030-01201-4_33.
- Piazzese, C., Carminati, M. C., Pepi, M. & Caiani, E. G. (2017). Statistical shape models of the heart : Applications to cardiac imaging. Dans *Statistical Shape and Deformation Analysis* (pp. 445–480). Elsevier. doi : 10.1016/B978-0-12-810493-4.00019-5.
- Pinheiro, P. O., Collobert, R. & Dollár, P. (2015). *Learning to segment object candidates*.
- Pinto, A. & Brunese, L. (2010). Spectrum of diagnostic errors in radiology. *World journal of radiology*, 2(10), 377. doi : 10.4329/wjr.v2.i10.377.
- Polat, Ö. & Yıldırım, T. (2008). Hand geometry identification without feature extraction by general regression neural network. *Expert systems with Applications*, 34(2), 845–849. doi : 10.1016/j.eswa.2006.10.032.
- Poncelet, N. & Cornet, Y. (2010). Transformée de Hough et détection de linéaments sur images satellitaires et modèles numériques de terrain. *Bulletin de la Société Géographique de Liège*, 54, 145–156.

- Portaz, M. (2018). *Accès à de l'information en mobilité par l'image pour la visite de Musées : Réseaux profonds pour l'identification de gestes et d'objets*. (Thèse de doctorat, Université Grenoble Alpes). Repéré à <https://tel.archives-ouvertes.fr/tel-02020126>.
- Preim, B. & Botha, C. (2014). Image analysis for medical visualization. *Visual computing for medicine*. Elsevier. doi : 10.1016/B978-0-12-415873-3.00004-3.
- Razavian, A. S., Azizpour, H., Sullivan, J. & Carlsson, S. (2014). CNN Features Off-the-Shelf : An Astounding Baseline for Recognition. *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 512-519.
- Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016). You Only Look Once : Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788.
- Ren, S., He, K., Girshick, R. & Sun, J. (2016). Faster R-CNN : towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6), 1137–1149. doi : 10.1109/TPAMI.2016.2577031.
- Rocco, I., Arandjelović, R. & Sivic, J. (2019). Convolutional Neural Network Architecture for Geometric Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(11), 2553-2567. doi : 10.1109/TPAMI.2018.2865351.
- Ronneberger, O., Fischer, P. & Brox, T. (2015). U-net : Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241.
- Rosati, R., Romeo, L., Silvestri, S., Marcheggiani, F., Tiano, L. & Frontoni, E. (2020). Faster R-CNN approach for detection and quantification of DNA damage in comet assay images. *Computers in Biology and Medicine*, 123, 103912. doi : 10.1016/j.combiomed.2020.103912.
- Ruiz-Santaquiteria, J., Bueno, G., Deniz, O., Vallez, N. & Cristobal, G. (2020). Semantic versus instance segmentation in microscopic algae detection. *Engineering Applications of Artificial Intelligence*, 87, 103271. doi : 10.1016/j.engappai.2019.103271.
- Saeedan, F., Weber, N., Goesele, M. & Roth, S. (2018). Detail-preserving pooling in deep networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9108–9116.
- Şahin, N., Alpaslan, N. & Hanbay, D. (2021). Robust optimization of SegNet hyperparameters for skin lesion segmentation. *Multimedia Tools and Applications*, 1–21. doi : 10.1007/s11042-021-11032-6.

- Sarazin, A. (2018). *Segmentation précise des structures de l'articulation du genou sur des radiographies EOS de face*. (Mémoire de maîtrise, École de technologie supérieure). Repéré à <https://espace.etsmtl.ca/id/eprint/2258>.
- Sharma, A., Aggarwal, M., Roy, S. D., Gupta, V., Vashist, P. & Sidhu, T. (2018). Optic Disc Segmentation in Fundus Images Using Anatomical Atlases with Nonrigid Registration. *Workshop on Computer Vision Applications*, pp. 14–27.
- Shen, W., Xu, W., Zhang, H., Sun, Z., Ma, J., Ma, X., Zhou, S., Guo, S. & Wang, Y. (2020). Automatic segmentation of the femur and tibia bones from X-ray images based on pure dilated residual U-Net. *Inverse Problems & Imaging*. doi : 10.3934/ipi.2020057.
- Shorten, C. & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 60. doi : 10.1186/s40537-019-0197-0.
- Simo-Serra, E., Trulls, E., Ferraz, L., Kokkinos, I., Fua, P. & Moreno-Noguer, F. (2015). Discriminative Learning of Deep Convolutional Feature Point Descriptors. *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 118-126.
- Simonyan, K. & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*.
- Sirinukunwattana, K., Pluim, J. P., Chen, H., Qi, X., Heng, P.-A., Guo, Y. B., Wang, L. Y., Matuszewski, B. J., Bruni, E., Sanchez, U. et al. (2017). Gland segmentation in colon histology images : The glas challenge contest. *Medical image analysis*, 35, 489–502. doi : 10.1016/j.media.2016.08.008.
- Sjögren, J., Ubachs, J. F., Engblom, H., Carlsson, M., Arheden, H. & Heiberg, E. (2012). Semi-automatic segmentation of myocardium at risk in T2-weighted cardiovascular magnetic resonance. *Journal of Cardiovascular Magnetic Resonance*, 14(1), 1–11. doi : 10.1186/1532-429X-14-10.
- Soins-infirmiers. (2021). Anatomie du squelette. Accessed : 2021-03-29, Repéré à <https://www.soins-infirmiers.com/discipline/orthopedie-traumatologie/anatomie-du-squelette>.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014). Dropout : A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(56), 1929-1958. Repéré à <http://jmlr.org/papers/v15/srivastava14a.html>.
- Stolocescu-Crișan, C. & Holban, Ș. (2013). A comparison of X-ray image segmentation techniques. *Advances in Electrical and Computer Engineering Engineering*, 13(3). doi : 10.4316/AECE.2013.03014.

- Südhoff, I. (2007). *Modélisation géométrique et mécanique personnalisée de l'appareil locomoteur*. (Thèse de doctorat, École de technologie supérieure). Repéré à <https://espace.etsmtl.ca/id/eprint/236>.
- Suzani, A., Rasoulian, A., Seitel, A., Fels, S., Rohling, R. N. & Abolmaesumi, P. (2015). Deep learning for automatic localization, identification, and segmentation of vertebral bodies in volumetric MR images. *Medical Imaging 2015 : Image-Guided Procedures, Robotic Interventions, and Modeling*, 9415, 941514.
- Swiecicki, A., Said, N., O'Donnell, J., Buda, M., Li, N., Jiranek, W. A. & Mazurowski, M. A. (2020). Automatic estimation of knee joint space narrowing by deep learning segmentation algorithms. *Medical Imaging 2020 : Computer-Aided Diagnosis*, 11314, 113143N.
- Tan, M., Pang, R. & Le, Q. V. (2020). EfficientDet : Scalable and Efficient Object Detection. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10778-10787.
- Thang, N. D., Dung, N. V., Duc, T. V., Nguyen, A., Nguyen, Q. H., Anh, N. T., Cuong, N. N., Hanh, B. M., Phu, P. H., Phuong, N. H. et al. (2021). Building a X-ray Database for Mammography on Vietnamese Patients and automatic Detecting ROI Using Mask-RCNN. Dans *Soft Computing for Biomedical Applications and Related Topics* (pp. 315–329). Springer. doi : 10.1007/978-3-030-49536-7_27.
- Tortora, G. J. & Derrickson, B. (2018). *Anatomie et physiologie*. De Boeck supérieur.
- Truong, P. H. (2016). *Optimisation des performances de la machine synchrone à réluctance variable : approches par la conception et par la commande*. (Thèse de doctorat, Université de Haute Alsace - Mulhouse). Repéré à <https://tel.archives-ouvertes.fr/tel-01514165>.
- Tsai, A. (2021). Anatomical landmark localization via convolutional neural networks for limb-length discrepancy measurements. *Pediatric Radiology*, 1–17. doi : 10.1007/s00247-021-05004-z.
- Tsechpenakis, G. (2011). Deformable model-based medical image segmentation. Dans *Multi modality state-of-the-art medical image segmentation and registration methodologies* (pp. 33–67). Springer. doi : 10.1007/978-1-4419-8195-0_2.
- Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media. doi : 10.1007/s11263-014-0733-5.
- Vasquez Barahona, P. A. (2004). *Effet des contraintes résiduelles dues à la polymérisation du ciment dans le transfert de la charge d'une arthroplastie totale de la hanche*. (Thèse de

- doctorat, École de technologie supérieure). Repéré à <https://espace.etsmtl.ca/id/eprint/711>.
- Virole, B. (2001). Réseaux de neurones et psychométrie. *Editions du Centre de Psychologie Appliquée–ECPA*. Repéré à <https://virole.pagesperso-orange.fr/RNECPA.PDF>.
- Wang, C.-W., Huang, C.-T., Lee, J.-H., Li, C.-H., Chang, S.-W. et al. (2016). A benchmark for comparison of dental radiography analysis algorithms. *Medical image analysis*, 31, 63–76. doi : 10.1016/j.media.2016.02.004.
- Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X. & Cottrell, G. (2018a). Understanding convolution for semantic segmentation. *2018 IEEE winter conference on applications of computer vision (WACV)*, pp. 1451–1460.
- Wang, S., Li, H., Li, J., Zhang, Y. & Zou, B. (2018b). Automatic analysis of lateral cephalograms based on multiresolution decision tree regression voting. *Journal of healthcare engineering*, 2018. doi : 10.1155/2018/1797502.
- Wang, X. (2015). *Graph based approaches for image segmentation and object tracking*. (Thèse de doctorat, Ecole centrale de Lyon). Repéré à <http://www.theses.fr/2015ECDL0007>.
- Wang, Z.-S., Lee, J., Song, C. G. & Kim, S.-J. (2020). Efficient Chaotic Imperialist Competitive Algorithm with Dropout Strategy for Global Optimization. *Symmetry*, 12(4). doi : 10.3390/sym12040635.
- Wessel, J., Heinrich, M. P., von Berg, J., Franz, A. & Saalbach, A. (2019). *Sequential rib labeling and segmentation in chest X-ray using Mask R-CNN*.
- Wong, S. C., Gatt, A., Stamatescu, V. & McDonnell, M. D. (2016). Understanding data augmentation for classification : when to warp ? *2016 international conference on digital image computing : techniques and applications (DICTA)*, pp. 1–6.
- Yahiaoui, M. (2017). *Modèles statistiques avancés pour la segmentation non supervisée des images dégradées de l’iris*. (Thèse de doctorat, Université Paris-Saclay). Repéré à <https://tel.archives-ouvertes.fr/tel-01591556>.
- Yang, D., Zhang, S., Yan, Z., Tan, C., Li, K. & Metaxas, D. (2015). Automated anatomical landmark detection on distal femur surface using convolutional neural network. *2015 IEEE 12th international symposium on biomedical imaging (ISBI)*, pp. 17–21.
- Yau, W., Leung, A., Chiu, K., Tang, W. & Ng, T. (2005). Intraobserver errors in obtaining visually selected anatomic landmarks during registration process in nonimage-based navigation-assisted total knee arthroplasty : a cadaveric experiment. *The Journal of*

- arthroplasty*, 20(5), 591–601. doi : 10.1016/j.arth.2005.02.011.
- Yau, W., Leung, A., Liu, K., Yan, C., Wong, L. L. & Chiu, K. (2007). Interobserver and intra-observer errors in obtaining visually selected anatomical landmarks during registration process in non-image-based navigation-assisted total knee arthroplasty. *The Journal of arthroplasty*, 22(8), 1150–1161. doi : 10.1016/j.arth.2006.10.010.
- Yi, J., Wu, P., Huang, Q., Qu, H. & Metaxas, D. N. (2020). Vertebra-focused landmark detection for scoliosis assessment. *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pp. 736–740.
- Yi, K. M., Trulls, E., Lepetit, V. & Fua, P. (2016). Lift : Learned invariant feature transform. *European conference on computer vision*, pp. 467–483.
- Yu, F. & Koltun, V. (2016). Multi-Scale Context Aggregation by Dilated Convolutions. *International Conference on Learning Representations (ICLR)*, abs/1511.07122.
- Yu, F., Koltun, V. & Funkhouser, T. (2017). Dilated Residual Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 636-644.
- Yu, G. & Morel, J.-M. (2011). ASIFT : An algorithm for fully affine invariant comparison. *Image Processing On Line*, 1, 11–38. doi : 10.5201/ipol.2011.my-asift.
- Yuan, J. (2018). Learning Building Extraction in Aerial Scenes with Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(11), 2793-2798. doi : 10.1109/TPAMI.2017.2750680.
- Zadeh, A., Lim, Y. C., Baltrušaitis, T. & Morency, L.-P. (2017). Convolutional Experts Constrained Local Model for 3D Facial Landmark Detection. *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 2519-2528.
- Zagoruyko, S. & Komodakis, N. (2015). Learning to compare image patches via convolutional neural networks. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4353-4361.
- Zhang, J., Liu, M. & Shen, D. (2017). Detecting Anatomical Landmarks From Limited Medical Imaging Data Using Two-Stage Task-Oriented Deep Neural Networks. *IEEE Transactions on Image Processing*, 26(10), 4753-4764. doi : 10.1109/TIP.2017.2721106.
- Zhang, J., Hu, H., Chen, S., Huang, Y. & Guan, Q. (2016). Cancer cells detection in phase-contrast microscopy images based on Faster R-CNN. *2016 9th International Symposium on Computational Intelligence and Design (ISCID)*, 1, 363–367.

- Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D. N. & Zhou, X. S. (2012). Towards robust and effective shape modeling : Sparse shape composition. *Medical image analysis*, 16(1), 265–277. doi : 10.1016/j.media.2011.08.004.
- Zhang, Z., Liu, Q. & Wang, Y. (2018). Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5), 749-753. doi : 10.1109/LGRS.2018.2802944.
- Zhao, H., Shi, J., Qi, X., Wang, X. & Jia, J. (2017). Pyramid Scene Parsing Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6230-6239.
- Zhao, H., Qi, X., Shen, X., Shi, J. & Jia, J. (2018a). Icnet for real-time semantic segmentation on high-resolution images. *Proceedings of the European conference on computer vision (ECCV)*, pp. 405–420.
- Zhao, K., Han, Q., Zhang, C.-B., Xu, J. & Cheng, M.-M. (2021). Deep Hough Transform for Semantic Line Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-1. doi : 10.1109/TPAMI.2021.3077129.
- Zhao, K., Kang, J., Jung, J. & Sohn, G. (2018b). Building Extraction from Satellite Images Using Mask R-CNN with Building Boundary Regularization. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 242-2424.
- Zheng, L., Yang, Y. & Tian, Q. (2018). SIFT Meets CNN : A Decade Survey of Instance Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(5), 1224-1244. doi : 10.1109/TPAMI.2017.2709749.
- Zheng, Y., Liu, D., Georgescu, B., Nguyen, H. & Comaniciu, D. (2015). 3D deep learning for efficient and robust landmark detection in volumetric data. *International conference on medical image computing and computer-assisted intervention*, pp. 565–572.
- Zhou, Y., Zhang, W., Tang, X. & Shum, H. (2005). A Bayesian mixture model for multi-view face alignment. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2, 741-746 vol. 2.
- Zhuo, Y. & Gou, C. (2020). An Active Shape Model Parametric Approach for Facial Landmark Detection. *2020 Chinese Automation Congress (CAC)*, pp. 5676-5680.