



End of Master project report realized by

Patty Coupeau (UO279822)

DNI: 180253250683

Master's in Automation Engineering and Industrial Informatics (MAIIND)

Ankle's MRI images analysis by deep learning



Directed by Pr.François Rousseau

Supervised by Pr.Antonio M. López Rodríguez and Pr.Jean-Baptiste Fasquel

June 22, 2020 - September 18, 2020

Medical Information Processing Laboratory
655 Avenue du Technopôle, 29280 Plouzané, France



Declaration of anti-plagiarism

I, the undersigned Coupeau Patty,
declare to be fully aware that the plagiarism of documents or part of a
document published on any form of medium, including the Internet,
constitutes an infringement of copyright as well as gross fraud.

10/22/2020 in Laval (France)



Acknowledgments

I would like to thank all the people who contributed to the success of my end of Master work and who allowed me to have a great professional and personal end-of-study experience.

First of all, I would like to express my thanks to my current thesis directors, Mr Jean-Baptiste Fasquel and Mr Mickaël Dinomais who helped me find this work fully in line with my expectations and my future ambitions.

I would like to thank the supervisors of the MAIIND Master, Mr Antonio Miguel Lopez Rodriguez and Mr Felipe Mateos Martin, for their help and support throughout the year without which I would have had trouble keeping up mentally.

I warmly thank my supervisor, Mr. François Rousseau, professor at IMT Atlantique, for his help, guidance and recommendations which enabled me to consolidate my skills while adopting new work habits that will be very useful during my doctoral thesis. I also want to thank all the members of the GIS BeAChild of the hospital Morvan of Brest led by Mr. Sylvain Brochard for their welcome, their good humor and their time spent in presenting their profession to me. In this dynamic team, I was able to flourish and gain knowledge of the medical field and more particularly of cerebral palsy that I did not have.

Finally, I do not forget Mr. Jérémy Chopin from LaRIS for his invaluable help in the implementation of neural networks on PyTorch.

Within LaTIM, I took another step in the world of research by working with many doctoral students who were kind enough to help me and share their experience with me. I am now more motivated than ever to start my thesis.



Contents

1	Introduction	11
2	Glossary	13
3	Description of LaTIM and integration into the Imagine team	16
3.1	Description of the laboratory	16
3.2	The Imagine team and the GIS BeAChild	17
4	Scientific and medical context	19
4.1	The ankle joint	19
4.2	Cerebral palsy and its impact on the ankle joint	21
4.3	MRI: a technology for understanding the musculoskeletal system of the foot in motion	24
4.4	State of the art - Interest of the work	27
5	Processing of MRI data	28
5.1	MRI data acquisition	30
5.2	MRI data processing	31
5.2.1	<i>Data cleaning and correction</i>	<i>31</i>
5.2.2	<i>Foot segmentation on static and dynamic data</i>	<i>34</i>
5.2.3	<i>Dilation and blurring of bones segmentations</i>	<i>35</i>
5.2.4	<i>Creation of HR dynamic sequences by rigid registration</i>	<i>37</i>
6	Learning MRI data with neural networks	42
6.1	The neural network architectures tested	44
6.1.1	<i>Fully Convolutional Network (FCN)</i>	<i>44</i>
6.1.2	<i>U-Net network</i>	<i>44</i>
6.1.3	<i>Residual network (Resnet)</i>	<i>45</i>

6.2	Issue 1: Segmentation of the ankle bones on static HR MRI.....	47
6.2.1	<i>Network data extraction</i>	47
6.2.2	<i>Neural network training</i>	49
6.2.3	<i>Performance comparison of U-Net and ResNet networks</i>	55
6.3	Issue 2: HR reconstruction of dynamic volumes.....	58
6.3.1	<i>Network data extraction</i>	58
6.3.2	<i>Neural network training</i>	61
6.4	Issue 3: Segmentation of the ankle bones on dynamic LR volumes.....	65
6.4.1	<i>Network data extraction</i>	65
6.4.2	<i>Prospects</i>	66
7	Conclusion	67
A	Appendix	69
A.1	Gantt chart.....	69
A.2	Nibabel coordinate system.....	70
A.3	Some notions on registration.....	71
A.4	Convolutional neural networks.....	73
A.5	Technologies and tools used.....	76
	Bibliography	80

List of Figures

3.1	Laboratory organization chart from [3]	17
4.1	Illustration of the ankle bones from [6]	19
4.2	Illustration of foot movements from [5]	20
4.3	Representation of the different types of motor deficiency depending on the area of the brain affected from [17]	21
4.4	Chronology in cerebral palsy	22
4.5	Illustration of the typical foot and ankle deformities of an individual spastic CP from [34]	23
4.6	MRI working principle from [16]	24
4.7	Visualization of the 3 sections of a static MRI of the ankle with resolution 0.28x0.28x0.8mm using the ITKSnap tool. Left = axial, top right = sagittal, bottom right = coronal	25
4.8	Dynamic MRI of the ankle joint during dorso-plantar flexion movement. The sequence is composed of 15 3D volumes (only 6 represented here) corresponding to 15 moments allowing to visualize the movement of the ankle during the movement	26
5.1	MRI processing pipeline including data cleaning, blurring of bone masks, foot segmentation on static and dynamic data as well as rigid registration of static MRI and their segmentation on dynamic volumes	29
5.2	Device used for ankle fixation during MRI data acquisition	30
5.3	Differences between a left foot (to the left) and a right foot (to the right) on a static 3D T1 MRI	32
5.4	Histogram of the static 3D T1 MRI of a subject	34
5.5	Foot mask obtained for the static MRI (left) and 1 of the 15 volumes (right) which composed the dynamic MRI of one subject	35
5.6	Structuring element used for the dilation of the masks of the ankle bones (here with a radius of 5)	36
5.7	Dilated and blurred masks of the calcaneus (left), talus (center) and tibia (right) with a radius of dilation of 5	36
5.8	Static HR MRI (left) and initial LR dynamic volume (right) of a control subject	37

5.9	Dynamic HR volume obtained by registering the static MRI on a dynamic LR volume	38
5.10	Evolution of the correlation between dynamic LR volume and HR registration as a function of the DICE measured between the segmentations of each of the three bones	39
5.11	Results of rigid registration from left to right of the calcaneus, talus and tibia	40
5.12	Registration on dynamic LR volume from left to right of the segmentations of the calcaneus, talus and tibia	40
6.1	Nvidia Titan X graphics card used to perform deep learning calculations, image taken from [4]	42
6.2	Example of FCN architecture for image reconstruction, illustration from [20]	44
6.3	Example of U-Net architecture used for segmentation problem, illustration from [36]	45
6.4	Residual blocks of ResNet networks, illustration taken from [11]	46
6.5	2D patch extraction for the problem of automatic segmentation of static MRI	48
6.6	Training of the 4 architectures for the segmentation of the calcaneus (on the abscissa the number of batches)	50
6.7	Evolution of the average DICE according to the number of parameters for the segmentation of the calcaneus by the U-Net and ResNet networks on the training (points) and validation (cross) data	51
6.8	Model of U-Net and ResNet neural networks chosen for the segmentation problem of static HR MRI	52
6.9	Evolution of the average DICE according to the learning rate for the U-Net and ResNet networks on the training (in blue) and validation data (in orange)	52
6.10	Evolution of the average DICE obtained on the validation data according to the partitioning of control subjects for U-Net and ResNet networks	53
6.11	Examples of segmentations on 2D 128x128 patches extracted from equine subjects obtained with the U-Net (middle) and the ResNet (bottom) networks. The top image is the ground truth (manual segmentation).	55
6.12	Segmentation of the 3 ankle bones on a static HR MRI slice obtained with the U-Net (top) and ResNet (bottom) networks	57
6.13	Histogram of the number of LR dynamic sequence patches extracted as a function of the correlation of the latter with the corresponding patch of the registered HR dynamic sequence	59
6.14	2D patch extraction for the HR automatic reconstruction problem of dynamic volumes	59
6.15	Sigmoidal function used to define the influence of patches in deep learning according to their correlation (representation made on Solumaths, the red point represents $x = 0.7$)	61
6.16	Evolution of the performance of ResNet for reconstruction according to the number of parameters for training (in blue) and validation data (in red) . . .	62
6.17	Graphic representation of the distance measurements L1 (in purple) and L2 (in orange) (representation made on Solumaths)	63

6.18	Reconstruction results obtained with the ResNet network on the validation data. From top to bottom, the input LR image, the expected HR output and the network output.	63
6.19	Difference map obtained for reconstruction results on validation data with the ResNet network	64
6.20	2D patch extraction for the automatic segmentation problem of dynamic MRI	65
A.1	Gantt chart realized with <i>Office Timeline</i>	69
A.2	The basic components of the registration are two input images, a transform, a metric, an interpolator and an optimizer [2]	72
A.3	Illustration of the organization of the layers of a CNN from [28]	73
A.4	Representation of the main activation functions from [28]: from left to right sigmoid, tanh and ReLU	74
A.5	Principle of the full-connected layer, illustration taken from [28]	74
A.6	Principle of the convolutional layer, illustration from [28]	75
A.7	Principle of the pooling layer, illustration taken from [28]	75

List of Tables

6.1	Comparison of the different neural network architectures for the segmentation of HR MRI on validation data ($lr = 0.0005$)	50
6.2	Average DICE obtained with U-Net and ResNet models on training and validation data for the segmentation of the 3 ankle bones	53
6.3	Average DICE obtained with the U-Net and ResNet models on the test data for the segmentation of the 3 ankle bones on 2D 128x128 patches	56
6.4	DICE obtained for the segmentation of 3D HR static MRI sequences of equine subjects with U-Net and ResNet models	56
6.5	Average correlation obtained on training and validation data using L1 and L2 distance measurements for the loss function	63

1. Introduction

With the ambition to continue my engineering studies with a doctoral thesis combining image processing and deep learning in the medical field, I decided to do my end-of-study work in the Medical Information Processing Laboratory (LaTIM) in Brest, France. The laboratory develops a multidisciplinary research in which information and health sciences mutually enrich each other in order to participate in the optimization of therapeutic actions. I joined the Imagine team, whose objective is to develop and provide therapists with integrative decision support models.

Cerebral palsy is today the main cause of motor disability in children in France, affecting 2 births in 1000 [17]. It is estimated that 90% of deformities due to CP appear in the areas of the ankle and foot preventing the normal growth of the affected limbs. Medico-surgical intervention remains the best solution to stabilize the deformities. However, due to the lack of knowledge of joint and muscular bio-mechanics, it is difficult to implement effective therapies. For several years, research laboratories have tried to use medical imaging techniques and in particular the MRI in order to study the musculoskeletal system in motion. Unlike static MRI which is an image frozen at a given moment, dynamic MRI is acquired during a cycle of movement and thus offers a representation of the evolution of the musculoskeletal system. However, unlike the static MRI, dynamic MRI presents many shortcomings (low resolution, artefacts, etc.) [32]

Since cerebral palsy causes bone deformities, particularly in the region of the foot, it would be interesting to develop tools to automate the monitoring of the ankle bones. Yet, the segmentation of bones in order to obtain 3D representations of them is done entirely by hand nowadays, making segmentation a long and expensive operation.

In this context, my contribution was to combine the spatial high-resolution information of static imaging techniques with the temporal low-resolution information of dynamic sequences in order to respond to the problem: **How to automate by deep learning the analysis of static and dynamic MRI images of the ankle?** To do this, I started by discovering the scientific context and preparing the MRI data. Subsequently, from the processed data, I used deep learning to address three issues of increasing difficulty: automatic bone segmentation on [HR](#) static MRI, dynamic data reconstruction in order to improve the accuracy of dynamic



sequences and automatic bone segmentation on dynamic [LR MRI](#). If we manage to implement the automatic segmentation and reconstruction, it will be possible to perform a more precise study of the moving ankle which may contribute to a better understanding of the joint bio-mechanics and therefore to the implementation of therapies more adapted to children suffering cerebral palsy.

2. Glossary

Abbreviations

CP: Cerebral Palsy

MRI: Magnetic Resonance Imaging

NMR: Nuclear Magnetic Resonance

FFE: Fast-Field Echo

HR: High Resolution

LR: Low Resolution

INSERM: French national health and medical research body

BeAChild: Breizh center for research and technological innovation for the development and rehabilitation of children

GIS: Scientific Interest Group

ACTION: Therapeutic ACTION guided by multimodality Imaging in ONcology

IMAGINE: Multimodal information integration for decision making and interventional therapies optimisation

MD,PhD: Medical Doctor, Doctor of Philosophy

HDR: Habilitation to conduct researches

ICT: Information and Communication Technologies

NMSS: Neuro-Musculoskeletal System

LaTIM: Medical Image Processing Laboratory

CPU: Central Processing Unit

GPU: Graphics Processing Unit

CNN: Convolutional Neural Network

Nifti: Neuroimaging Informatics Technology Initiative

SSD: Sum of Squared Differences

SIFT: Scale-Invariant Feature Transform

ReLU: Rectified Linear Unit

FCN: Fully Convolutional Network

Technical terms

binarizing: Transforming a gray-scale image into an image whose pixel values are 1 or 0.

segmentation: Image processing operation which purpose is to gather pixels according to predefined criteria in order to form regions.

deep learning: A set of machine learning methods attempting to model with a high level of data abstraction through articulated architectures of different nonlinear transformations.

normalization: Process of changing the range of pixel intensity values of an image in order to bring this latter into a range more familiar to the senses.

dilation: Morphological operation aiming to extend the shapes contained in the input image using a structuring element.

registration: Technique which consists of the spatial matching of images in order to be able to compare or combine their respective information. This matching is done by looking for a geometric transformation allowing to go from one image to another.

convolutional neural networks: Class of deep neural networks, most commonly applied to the analysis of visual imagery. The pattern of connectivity between neurons looks like the organization of the animal visual cortex. CNNs use relatively little pre-processing compared to other image classification algorithms.

Nifti: Commonly used file format for storing brain imaging data obtained by MRI.

correlation: 2D or 3D optical method which measures the displacements between two images.

semantic segmentation: Semantic segmentation is a deep learning algorithm that associates a label or a category to each pixel of an image.

standardisation: Process of transforming one feature into another that will meet the normal distribution.

recurrent neural networks: Artificial neural network with recurrent connections made up of interconnected neurons interacting non-linearly. They are suitable for input data of varying size.

Medical terms

cerebral palsy: Cerebral palsy is linked to damage to the brain that occurs during pregnancy, childbirth, or in the first months of life. It affects a brain that is still maturing and causes a non-progressive motor disorder (paralysis, movement coordination disorders).

oncology: Branch of medicine devoted to the study, diagnosis and treatment of cancer.

multi-modal imagery: Use of several imaging technologies (nuclear imaging (PET, SPECT) as well as CT MRI imaging) to obtain additional anatomical, functional and molecular data.

CT scan: Examination also called a scanner giving cross-sectional images of an organ.

interventional radiology: All medical procedures performed under radiological control, allowing access to a lesion located inside the body to perform a diagnostic (sample for example) or therapeutic (treat, repair, close ...).

joint: Set of soft and hard parts by which two or more neighboring bones unite.

navicular assembly: Bones of the foot also called tarsal scaphoid.

inversion: Combined movement of the ankle coupling plantar flexion, supination and adduction.

eversion: Combined ankle movement coupling dorsiflexion, pronation and abduction.

brain damage: Injury to the brain which generally results in a more or less extensive destruction of nervous tissue leading to a deficit in perception, cognition, sensitivity or motor skills depending on the role played by the affected region. This lesion can be of various nature (ischemic, hemorrhagic, etc.).

muscle spasticity: Exaggerated tension of a muscle in the absence of voluntary contraction.

spin: Angular momentum of a particle. It is a quantum property that takes only whole or half-whole values.

precession: A spinning motion similar to that of a spinning top of atoms that tip over and remain in balance in rotation. The rate of precession is proportional to the strength of the magnetic field.

3. Description of LaTIM and integration into the Imagine team

3.1. Description of the laboratory

The medical information processing laboratory was born in 1996 from the creation of a scientific interest group [3]. It is recognized by the [INSERM](#) as a research unit in 2004. Since 2010, LaTIM has been known as a joint UMR 1101 research unit led by members of the University of Bretagne Occidentale, IMT Atlantique, INSERM and Brest CHRU (CHU).

Led by the [MD,PhD](#) Eric Stindel, the LaTIM develops a multidisciplinary research aimed at improving the medical service. At its creation, the laboratory's first major research activity was the development of computer-assisted surgery, mainly for orthopedic applications. Over time, the laboratory has seen its structure and its research activities evolve around themes strongly linked to the field of medical information processing and to associated areas of clinical activity. Thus, since January 2017, the scope of research carried out within the laboratory has been divided into two teams (Fig.3.1):

- **team ACTION**: therapeutic action guided by [multi-modal imaging](#) in [oncology](#).
- **team IMAGINE**: integration of multi-modal information for decision support systems and optimization of [interventional radiologies](#).

The environment is ideal for the integration into the world of research with the presence of 30 [HDR](#) and about 140 doctors, researchers and administrative staff.

The LaTIM has multiple office spaces and dedicated scientific platforms with a central location of 1000 m² at the Brest Biology and Health Research Institute. It also has 350 m² of offices in the premises of the hospital of Brest (CHRU Brest) and IMT-Atlantique located in Plouzané where I completed my work.

I joined the Imagine team, whose organization and objectives are detailed in the following section.

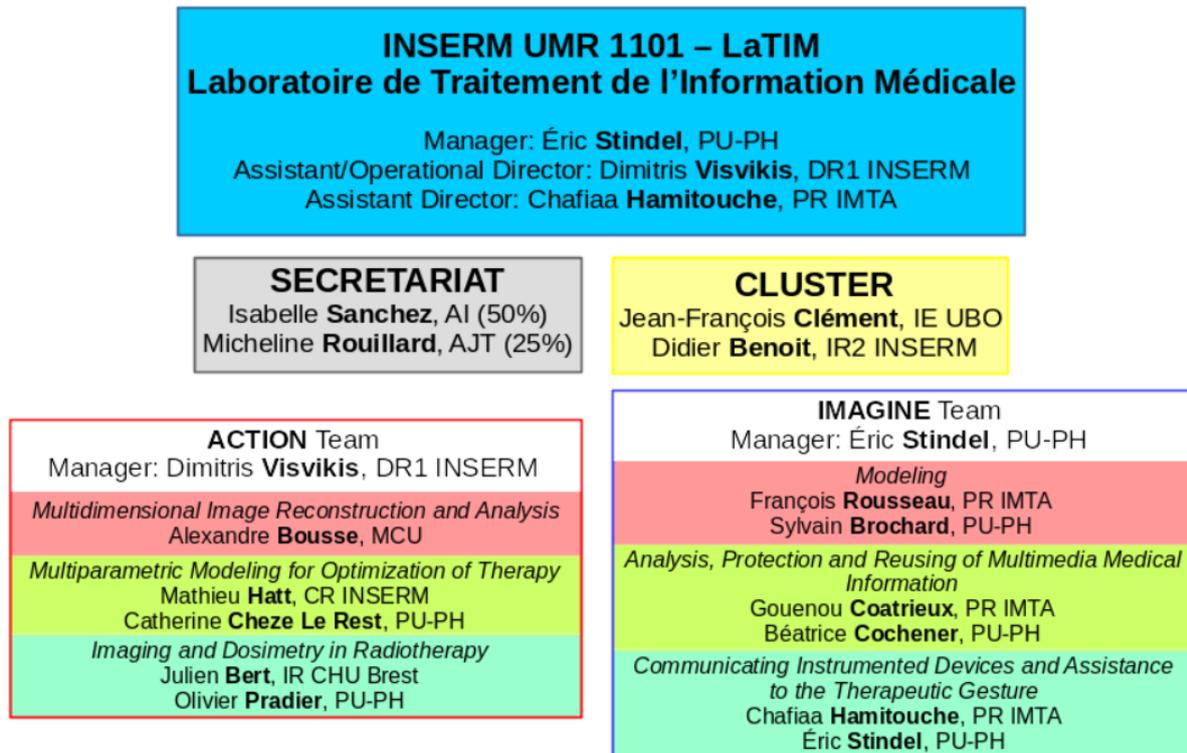


Figure 3.1: Laboratory organization chart from [3]

3.2. The Imagine team and the GIS BeAChild

The IMAGINE team aims to develop and provide therapists with integrative decision support models. The apparition of ICT in the field of health has had a lasting impact on the practice of interventional medicine and surgery. The proliferation of sensors, their communicating and on-board nature leads practitioners to optimize the therapeutic action in real time. Beyond the decision alone, the therapist's ability to implement it with precision is the second key to effective treatment. The development of new therapeutic models is therefore a major challenge for the team. To this end, the IMAGINE team is organized into three themes led by a clinician and a methodologist in order to maintain the balance between information and medical sciences (Fig.3.1):

- **Modeling:** study of the control/form/function relationship within the neuro-musculoskeletal system (NMSS) to support therapeutic decision making and therapy evaluation.
- **Therapeutic Gesture:** research for surgical management in orthopedics.
- **Analysis, Protection and Reusing of Multimedia Medical Information:** development of deep learning tools in ophthalmology, screening for ocular pathologies.

At the same time, the team continuously ensures the security of information during the clinical follow-up of the patient.

I evolved within the research axis **Modeling** led by Professor François Rousseau who was my supervisor and the MD,PhD Sylvain Brochard. The starting point of this axis is the acquisition of multi-modal spatial and temporal data with the corresponding processing methods. Its objective is to develop new 3D + temporal modeling techniques in a patient-specific approach in order to support therapeutic decision-making.

In parallel with Imagine, the scientific interest group (GIS) [BeaChild \[1\]](#) was created less than a year ago. It gathers some twenty engineers and health workers from the team around research focused on pediatric rehabilitation. The multidisciplinary nature of this group allowed me to deepen my skills in image processing and deep learning while getting involved in the world of health. Through discussions with the group's medical professionals (physiotherapists, occupational therapists), I have acquired medical knowledge which will be precious for me during my doctoral thesis started in November. During my experience within the Imagine team, I developed and improved essential research skills: being critical, communicating about my work, being imaginative and never stopping being curious.

4. Scientific and medical context

4.1. The ankle joint

The ankle [joint](#) is a complex anatomical structure that connects the foot to the lower leg [5]. It is made up of four main bones (Fig.4.1): tibia, fibula (fibula), talus and calcaneus.



Figure 4.1: Illustration of the ankle bones from [6]

The calcaneus, talus and tibia form several joints:

- **the talocrural joint:** this is the central ankle joint that connects the tibia to the talus. It governs the flexion/extension movements of the ankle also called plantar/dorsiflexion.
- **the subtalar joint:** this joint connecting the talus to the calcaneus directs the pronation/supination movements.
- **the mid-tarsal joint:** in this articular uniting the calcaneus with the [navicular assembly](#), the ankle performs adduction and abduction movements.

This joint complex, made up of several movements, carries out combined movements which are called: [inversion](#) /[eversion](#) (Fig.4.2).

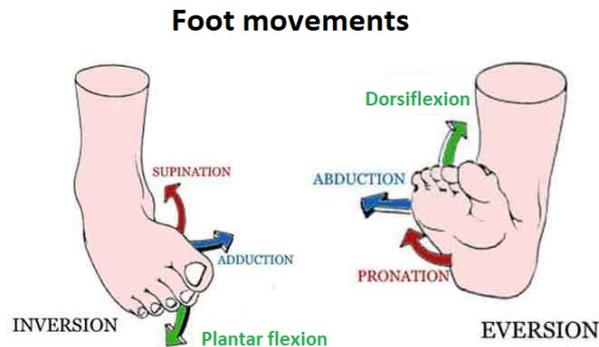


Figure 4.2: Illustration of foot movements from [5]

Ligaments are attached to both sides of the ankle to hold the ankle and foot together. On the outside of the ankle, one can distinguish the anterior talofibular ligament, the fibulocalcaneal ligament and the posterior talofibular ligament. Inside, the deltoid ligament is thicker and stronger.

Finally, these are the tendons that connect the muscles of the leg to the bones of the foot contributing to the stability of the ankle. We find:

- **Achilles tendon:** tendon connecting the calf muscles to the calcaneus. It plays a major role in walking, running and jumping.
- **fibular tendons:** tendons allowing the foot to turn outwards.
- **the anterior tibial and posterior tibial tendons:** tendons to lift the foot and turn the foot inward.

The ankle joint is a fragile complex. According to [17], 90% of deformities due to cerebral palsy appear in the ankle and foot regions. These deformities can have serious repercussions on the growth of the limbs. But what is this pathology?

4.2. Cerebral palsy and its impact on the ankle joint

Cerebral palsy is the leading cause of motor disability in children. It results from irreversible damage to the developing brain of the fetus or infant. These lesions cause a set of non-progressive movement or posture disorders that last a lifetime [17]. According to [21], in France:

- **17 million** people have a **CP**. It represents about 125000 people.
- **1 in 500 babies** is affected.

Although half of CP infants are born at term, it is in the population of premature infants that the risk of **brain damage** is the most important. Causes of CP are various:

- **antenatal factors**: stroke, brain malformation, infection of the fetus (cytomegalovirus, toxoplasmosis, etc).
- **neonatal or perinatal factors**: prematurity, neonatal stroke, infection, trauma, neonatal distress during difficult childbirths, etc.
- **post-natal factors**: infection, trauma, severe convulsions, metabolic disease, sudden infant death syndrome after resuscitation, etc. Post-natal causes represent only 10% of the known causes of CP.

The extent of symptoms varies from one individual to another depending on the location of the lesions (Fig.4.3), of their extent and the time of their appearance. Brain MRI is essential for etiologic diagnosis and often helps to define the causative lesion.

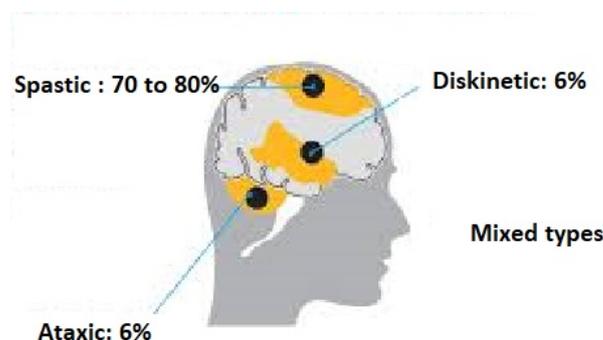


Figure 4.3: Representation of the different types of motor deficiency depending on the area of the brain affected from [17]

Forms of cerebral palsy are also classified according to the affected limbs [22]:

- **quadriplegia:** the 4 limbs affected.
- **diplegia:** the 4 limbs affected with predominance in the lower ones.
- **hemiplegia:** damage to one side of the body predominantly in the upper limb.
- **double hemiplegia:** the 4 limbs affected with predominance in the upper ones.
- **triplegia:** 3 limbs affected (often 2 upper limbs and one lower limb).
- **monoplegia:** only one limb affected.

The ankle and foot regions are most affected by deformities from CP. Although the brain injury is stable and not progressive, the bone deformities progress. Indeed, a lesion occurring on an immature brain is stable over time. However, it produces **muscle spasticity** (exaggerated muscle tension). As the child grows, spastic muscles fail to stretch to accommodate bone growth. To compensate, these muscles retract and exert abnormal forces on the skeleton which is deformed (Fig.4.4). Thus, individuals with spastic paralysis affecting the lower limbs tend to develop deformity of the feet and ankles in equine and valgus causing an abnormal gait (Fig.4.5). The equine incidence represents approximately 75 % of the deformities of the ankle due to CP [25].

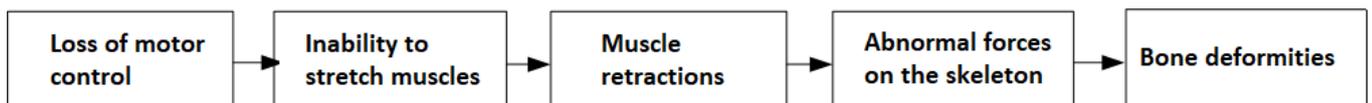


Figure 4.4: Chronology in cerebral palsy

Equinus is present when the dorsiflexion of the foot at the ankle is 90 degrees or less with the subtalar joint held in a neutral position and the knee extended [34].

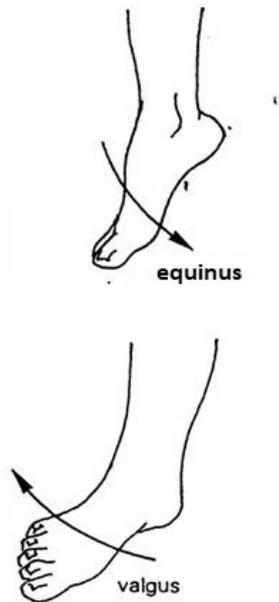


Figure 4.5: Illustration of the typical foot and ankle deformities of an individual spastic CP from [34]

The consequences of cerebral palsy on the regions of the foot and ankle are serious since it is estimated that **one in three CP person** can't walk [21]. Since brain plasticity is more important in the first years of life, it is essential to take care of the CP child as early as possible. Physiotherapy plays a key role in this care through various rehabilitation methods (intensive, participatory, etc.). Nevertheless, medico-surgical intervention remains the best solution to stabilize the lower limb. Unfortunately, due to the lack of knowledge of joint and muscular bio-mechanics, the therapies used do not have the expected effectiveness. This is the issue of current scientific research.

For several years, research laboratories have been trying to put in place methods to study the musculoskeletal system in motion, in particular by exploiting the techniques of medical magnetic resonance imaging (MRI).

4.3. MRI: a technology for understanding the musculoskeletal system of the foot in motion

Imaging methods have long played a crucial role in helping surgeons and clinicians diagnose various disorders [32].

In vivo imaging of moving articulated structures can be performed using various dynamic imaging techniques such as CT scan, ultrasound or MRI. However, ultrasound is limited to evaluating the soft tissue around the joint while CT scan only quantifies bone kinematics and exposes patients to ionizing radiation. MRI has the advantage of being a non-ionizing technique for studying the musculoskeletal system providing anatomical details of surrounding bones and soft tissues in static and dynamic contexts.

Magnetic Resonance Imaging (MRI) is a recent medical imaging technique (1973) allowing to visualize with great precision in 2D or 3D organs and soft tissues thanks to the properties of the magnetic field [16]. MRI is based on a physical phenomenon called Nuclear Magnetic Resonance (NMR) based on the possession by the nuclei of atoms of a kind of internal magnet called spin (Fig.4.6). The orientation of the latter can be changed by applying a magnetic field to it. This phenomenon concerns all atoms with an odd number of protons and neutrons.

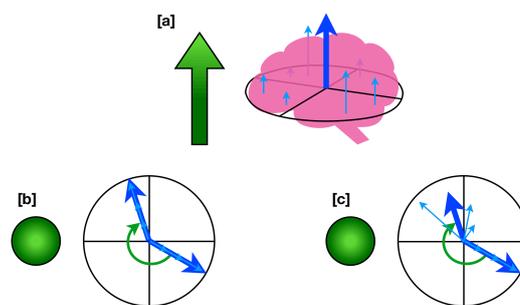


Figure 4.6: MRI working principle from [16]

To create an image, the ankle (pink) is subjected to a strong constant magnetic field (green arrow in image a) created by a large magnet: this is the tunnel in which the individual inserts his foot. Non-zero spin atoms begin to turn in such a way that their spin aligns with the magnetic field (light blue arrows). By adding the arrows, we get the magnetization of the ankle (big blue arrow). In order to collect a signal and obtain an image, it is necessary to switch this magnetization in the black plane. For this, the ankle is subjected for a short time to another magnetic field in addition to the initial green magnetic field. In image b (seen from below of image a) we can see the magnetization tilted in the black plane which begins to rotate around the axis of the magnetic field : it is the precession responsible of the

signal measured by the MRI.

When the oscillating field is stopped, the atoms will gradually regain their initial position and their nucleus releases the energy acquired during the excitation. The frequencies are then processed as electrical signals and provide the image.

Conventional **static MRI** is a 3-dimensional image taken at a given time. On each of the 3 dimensions, the static MRI presents several slices allowing to move spatially in this direction. These techniques are used for an accurate diagnosis of ankle and foot disorders given the complexity of their anatomy and the high-resolution image provided (Fig.4.7). On the latter, one notices the an-isotropic character of the resolution (different resolution according to the 3 acquisition axes: 576 slices for axial and coronal axes, 100 slices for the sagittal one). However, these techniques do not allow dynamic representation of the musculoskeletal system since there are static.

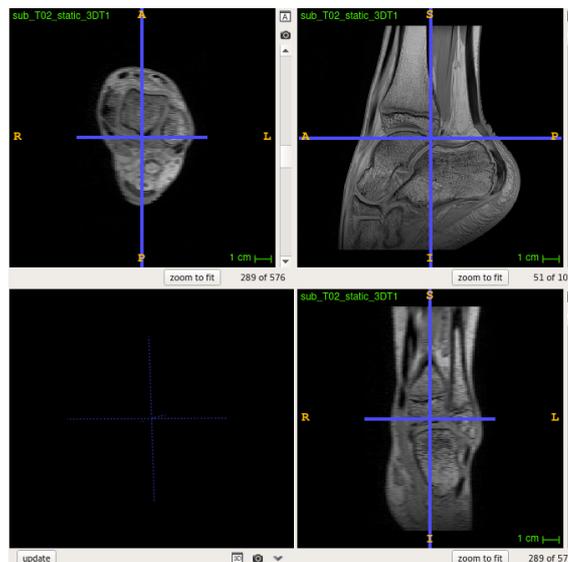


Figure 4.7: Visualization of the 3 sections of a static MRI of the ankle with resolution 0.28x0.28x0.8mm using the [ITKSnap tool](#). Left = axial, top right = sagittal, bottom right = coronal

Therefore, **dynamic MRI** is favored for study in motion. Indeed, this type of imaging is an assembly of 3D volumes acquired during a cycle of movement. Each 3D volume is a static image similar to the static MRI but this time of low resolution. We therefore say that dynamic MRI is in 4D, the fourth dimension being time. There are different dynamic MRI sequences [32]:

a) Fast-PC MRI

These sequences provide in vivo measurements of the joint velocity field. However, these data require long acquisition times with a high number of repeated cycles. A major clinical limitation of this type of sequences is that subjects with musculoskeletal disorders cannot perform a large number of repeated cycles of movement.

b) Ultrafast-MRI

This sequence is divided into two groups:

- **Ultrafast 3-T MRI:** Unlike the Fast-PC MRI technique, this imaging technique captures the joint trajectory during a single range of motion cycle by exploiting space-time redundancy.
- **Real-Time T1 Fast Field Echo:** Another reliable technique to assess joint movement with good contrast between ankle structures (Fig.4.8). This imaging modality allows in vivo quantification of joint kinematics where the scanning time to acquire a single dorso-plantar flexion cycle is reduced to only 18s (15 frames corresponding to 15 volumes, spatial resolution of $0.56 \times 0.56 \times 8$ mm). The work described in this report is based on this kind of MRI sequence.

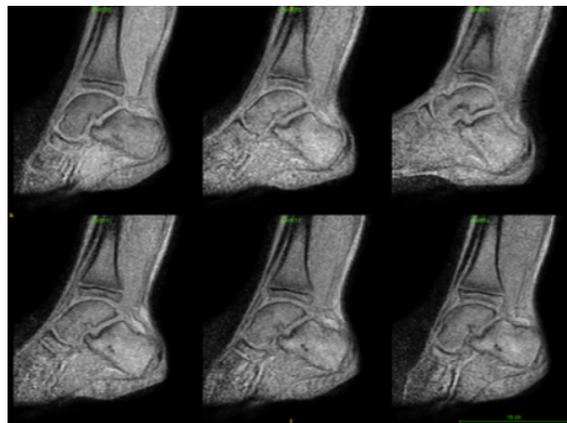


Figure 4.8: Dynamic MRI of the ankle joint during dorso-plantar flexion movement. The sequence is composed of 15 3D volumes (only 6 represented here) corresponding to 15 moments allowing to visualize the movement of the ankle during the movement

Dynamic MRI techniques allow to explore and study the musculoskeletal system in motion in the 3 dimensions of space. However, unlike high-resolution static MRI, dynamic MRI exhibit low resolution as well as many movement-related artifacts. All these drawbacks are as many problems on which current research is looking in order to allow a more precise analysis of the joint bio-mechanic and consequently the implementation of therapies more adapted to the child with CP.

4.4. State of the art - Interest of the work

Since 2005, many research projects have emerged in partnership with the Cerebral Palsy Foundation [17]. These projects cover various themes including:

- brain damage and pain study
- rehabilitation and analysis of posture
- development of new technologies for physiotherapy
- cognition and perception

Medical information processing has experienced considerable progress over the past twenty years. Several clinical applications have exploited the performance of Ultrafast-MRI: prenatal diagnosis of abnormalities [23], assessment of abusive head trauma [26] or the determination of the kinematics of the ankle joint in motion [15].

In 2015, LaTIM obtained funding to evaluate in vivo joint and muscle mechanics of the ankle in CP children with equine deformity [13]. This project, led by the MD,PhD Sylvain Brochard, aims to understand the mechanisms involved in the equine of the child with CP through dynamic imaging techniques. For this, CP children with unilateral fixed equine are compared to control children of the same age in order to characterize the underlying deformities and abnormal movements of the ankle bones.

In the same way, thesis works were carried out in order to develop new tools to improve the diagnosis of MRI [32]. The processing of volumetric data makes it possible to analytically follow the bones and the movements of the skeleton without having to identify anatomical landmarks. The measurement accuracy relies on the combination of spatial information from conventional static MRI with temporal information from dynamic four-dimensional MRI.

However, there is still no automatic tool to segment the ankle bones. The implementation of this type of tool would make it possible to obtain in no time a 3D representation of the bones necessary for the study of the latter.

Thus, this work aims to improve the monitoring of the bones of interest in the ankle (**calcaneus, talus and tibia**) while reducing human intervention. From MRI data acquired as part of the LaTIM project [13], I faced three challenges:

1. automatic segmentation of bones of interest in static HR MRI using deep learning.
2. automatic reconstruction of dynamic MRI using deep learning in order to create dynamic sequences of higher resolution and so easier to analyze.
3. automatic segmentation of bones of interest in dynamic LR MRI using deep learning.

Beforehand, it was necessary to implement an MRI data processing pipeline.

5. Processing of MRI data

The work was based on MRI data obtained in the context of the research project carried out by LaTIM. This latter seeks to dynamically evaluate in vivo the articular and muscular mechanics of the ankle in CP children with equine deformity [13]. Thirteen children with cerebral palsy aged 7 to 14 years and presenting a unilateral fixed equine are compared to ten unaffected children of the same age. In this section, the protocol for acquiring MRI data as well as cleaning and processing this data for deep learning will be described (Fig.5.1). It should be noted that the data were acquired before my arrival in the structure, so I did not participate.

This protocol consists of the following:

1. **Data cleaning and correction:** eliminating subjects for which all the necessary data are not available in the correct format (Nifti), checking the orientation of the images and decomposing dynamic 4D sequences into the 15 3D volumes which compose it.
2. **Foot segmentation on static and dynamic data:** [segmentation](#) which will be necessary to give [normalized](#) data in the foot area as an input of [deep learning](#).
3. **[Dilation](#) and blurring of segmentations of the calcaneus, talus and tibia**
4. **Rigid [registration](#) of static data (MRI and segmentations of the three bones of interest) on dynamic volumes:** last step in the creation of dynamic HR sequences and their segmentation.

A description of the different stages of the pipeline is given next.

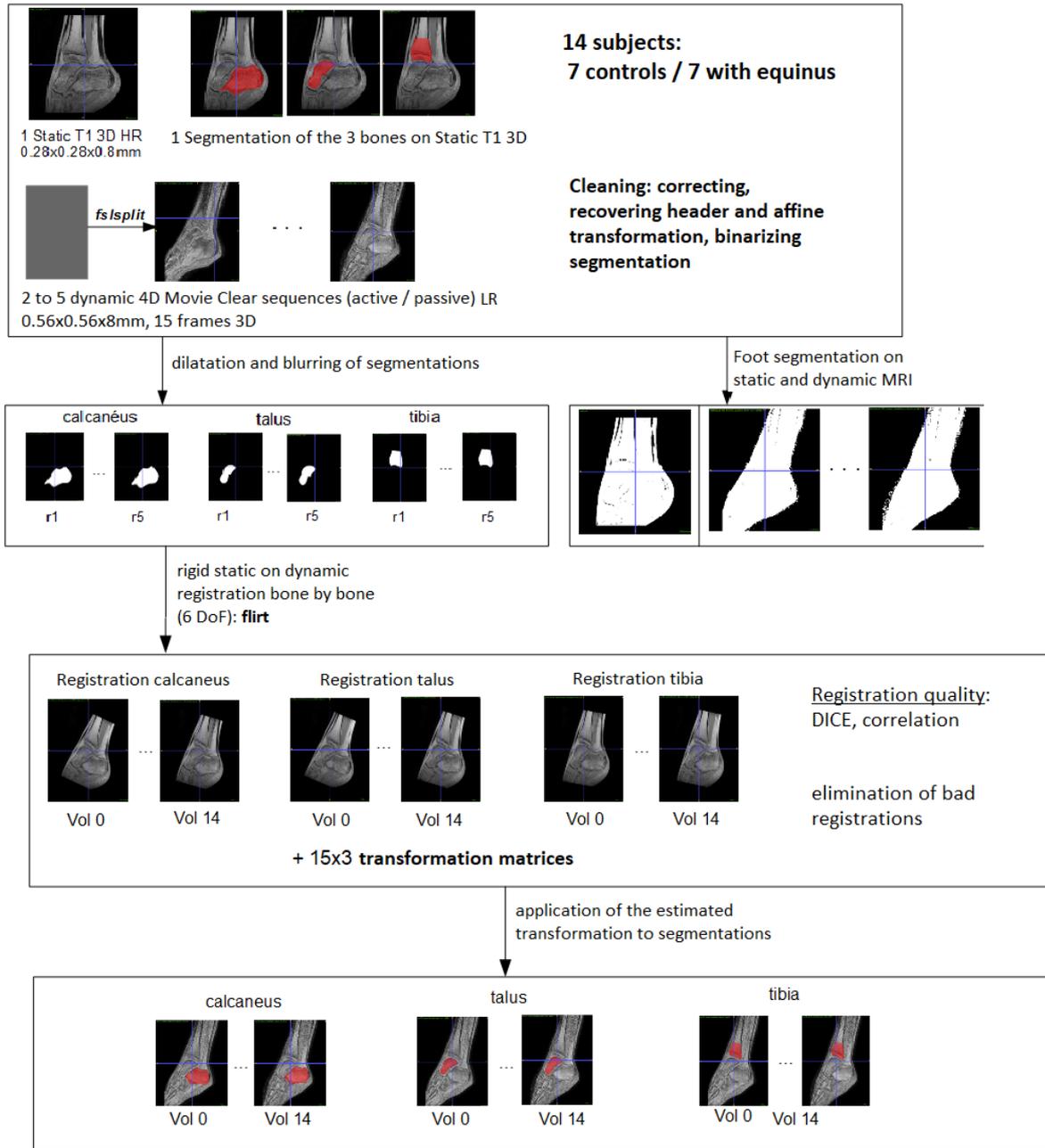


Figure 5.1: MRI processing pipeline including data cleaning, blurring of bone masks, foot segmentation on static and dynamic data as well as rigid registration of static MRI and their segmentation on dynamic volumes

5.1. MRI data acquisition

The acquisition of MRI data took place in the premises of the hospital Morvan in Brest before my arrival. The study involved ten control children and thirteen children with a unilateral fixed equine aged 7 to 14 years. Although this is a pathology to be followed from an early age, the age limit has been set at 7 years to ensure that the child remains focused during the entire protocol for about 1 hour 30 minutes.

The child was installed in a device such as the one shown Fig.5.2 and placed his ankle in the ankle rest. Once installed correctly, three MRI acquisitions were performed:

- **static MRI:** the child was watching a cartoon so that he relaxed his joint.
- **active dynamic MRI sequences:** the child had to voluntarily perform a plantar-dorsiflexion movement to the rhythm of a metronome.
- **passive dynamic MRI sequences:** a technician cyclically moved the ankle fixation using guide wires in a plantar-dorsiflexion motion to the rhythm of a metronome.

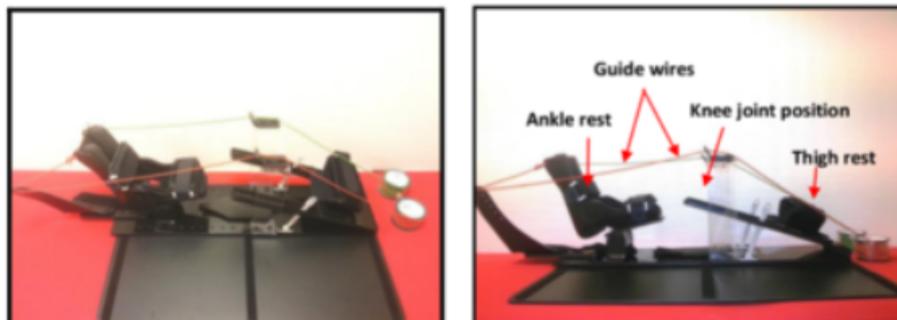


Figure 5.2: Device used for ankle fixation during MRI data acquisition

In the case of children with CP, an MRI scan was acquired of the ankle affected by an equine. For the control children, it was decided to study the ankle on the non-dominant side that is to say the left ankle if the child is right-handed and vice versa.

Once the MRI data has been acquired, it is necessary to sort them out in order to keep only the usable ones. In addition, it is necessary to think about a treatment of the data in order to use them as inputs of the neural networks intended to solve the three problems mentioned previously.

5.2. MRI data processing

Previously, a manual segmentation with the tool [ITKSnap](#) of the three ankle bones of interest (calcaneus, talus and tibia) on static MRI was performed by a former doctoral student. The principle of segmentation consists in drawing the ankle bones on each slice of the static MRI according to the 3 directions. In this way, the voxels belonging to the bone are set to 1 and the other voxels to 0.

The aim of this data processing is to create all the images that will feed the neural networks intended to solve the 3 deep learning challenges:

- **For the automatic segmentation of the ankle bones on static HR images:** The initial static MRI and the manual segmentation of the three bones on the static MRI.
- **For automatic reconstruction of dynamic volumes:** the initial LR volumes which composed the dynamic MRI and a HR reconstruction of these volumes which will be obtained by registering the HR static MRI on the LR dynamic volume. The principle of registration will be described later.
- **For the automatic segmentation of the ankle bones on dynamic LR images:** the initial LR dynamic volumes which composed the dynamic MRI and the segmentation of the three bones on these volumes obtained by applying the transformation estimated by registration during step 2 to the manual segmentation of the static MRI.

5.2.1. Data cleaning and correction

A subject is considered as exploitable for the study if the following data is available in the [Nifti](#) format :

- one static T1 3D MRI with a resolution of 0.28x0.28x0.8mm on which the 3 bones of interest are visible
- manual segmentation of the calcaneus, talus and tibia on static MRI
- at least one dynamic active and/or passive T1-FFE MRI sequence of 15 frames (volumes) with a resolution of 0.56x0.56x8mm

First of all, it is known for each subject whether the observed joint is the right or left ankle. Therefore, we must ensure that all of these images (MRI and segmentations) are not [flipped](#). This implies knowing how to recognize a left foot from a right one (Fig.5.3). When this is the case, one has to "reverse" the image along the good axis with a python script using the library [Nibabel](#).

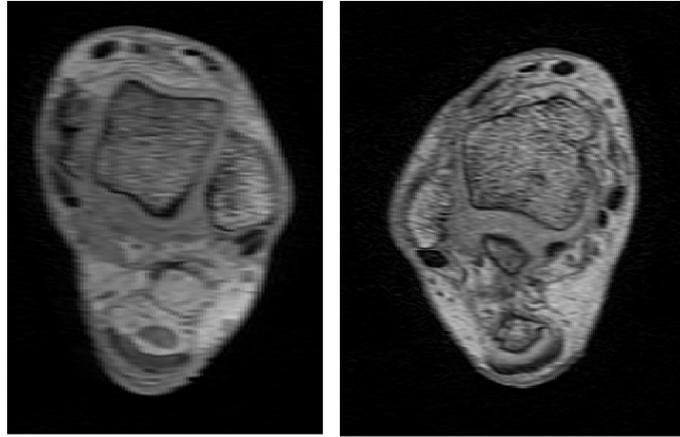


Figure 5.3: Differences between a left foot (to the left) and a right foot (to the right) on a static 3D T1 MRI

Moreover, it is necessary to ensure that all the data of the same subject are defined in the same spatial reference. In fact, the coordinates of the image do not give any information about the position in the world. Each image has its own spatial reference defined by an origin and three axes generally referenced in relation to the position of the patient:

- Origin = part of the patient being examined
- Left-right axis
- Anterior-posterior axis
- Inferior-superior axis

To compare two images, it is therefore essential to ensure that their system of reference is the same: ex. RAS scanner (right, anterior, posterior), LPI scanner (left, posterior, inferior). [Appendix](#) explains in more detail the Nibabel coordinate system [9].

Given that we are going to combine the static and dynamic intra-patient information, it is essential that all the images and segmentations are comparable. Otherwise, it is sufficient to copy the header of the static MRI (containing the common spatial reference of the latter) to the segmentations and dynamic sequences and to transfer the affine transformation of the latter in the correct coordinate system. All these manipulations are carried out using the library Nibabel.

Although the static and dynamic MRI are defined in the same spatial reference, their data are not yet fully comparable because they do not overlap. In fact, from one sequence to another, the origin of the space defined by the ankle's center of the patient is slightly different because the child obviously moves his ankle between each acquisition. To obtain perfectly comparable data, it will be necessary to carry out a registration of these.

Finally, registration is only possible between images of the same dimensions. However, static MRI has 3 dimensions and dynamic MRI 4D. Consequently, the dynamic 4D sequences (3D + time) should be broken down into the 15 3D volumes which composed it corresponding to the 15 instants of the movement cycle. The `fslsplit` tool of the library `fsl` is ideal to perform this type of operation.

At the end of this data cleaning and correction, the number of subjects usable for the study was reduced to **7 control children and 7 children with equine**. It is therefore possible to start processing the MRI data of these 14 subjects.

5.2.2. Foot segmentation on static and dynamic data

Segmentation of the foot is an important step in preparing data for deep learning. Indeed, depending on the subjects and the nature of the MRI, the intensities of the voxels in the image may present different orders of magnitude. This difference can lead to poorer performance during deep learning [7].

Therefore, when implementing neural networks for segmentation, we will normalize all the data by setting the normalization parameters (mean, standard deviation) to the region of the foot. In the case of the reconstruction problem, normalization will be in the region of the bone of interest. This will be explained in the next chapter on deep learning.

For this, it is necessary to extract a foot mask on the static MRI (for automatic segmentation on static data) and on each of the 15 volumes of the dynamic sequences (for automatic segmentation on dynamic data). The adapted strategy was the following:

1. Representing the histogram of the static MRI (or dynamic volume) of the 14 subjects (Fig.5.4)
2. Visually defining a threshold beyond which voxels are part of the foot: **250**
3. Binarizing static MRI (or dynamic volume) with this threshold

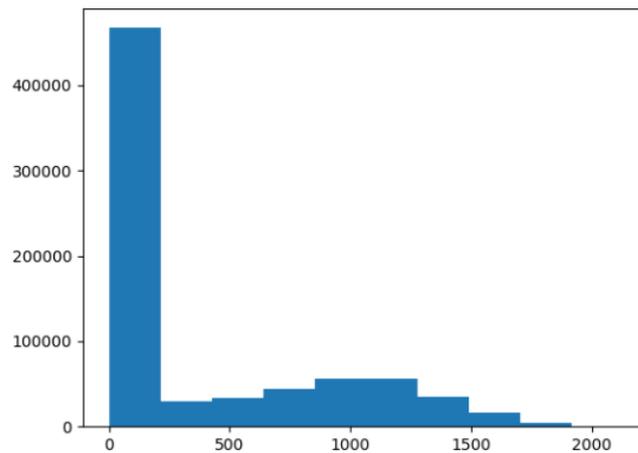


Figure 5.4: Histogram of the static 3D T1 MRI of a subject

An example of the masks obtained is proposed on Fig.5.5.



Figure 5.5: Foot mask obtained for the static MRI (left) and 1 of the 15 volumes (right) which composed the dynamic MRI of one subject

5.2.3. Dilation and blurring of bones segmentations

The dilation and blurring of the masks of the 3 ankle bones is essential to achieve a correct bone-to-bone registration. To understand it, it is advisable to recall its principle. The aim of the registration is to estimate the spatial transformation T to go from one image to another. For this, we seek to optimize a similarity criterion between the two images.

During this operation, it is possible to give more weight to some voxels of the image. In this way, we estimate more precisely the existing spatial transformation between the same specific region present on each of the two images. In our case, this property will be exploited since we want to obtain static and dynamic data which perfectly overlap in the bones regions.

Furthermore, to calculate the similarity between the two images in the region of the bone, it is necessary to know the contours of the bone in order to ensure that in both images it occupies the same space. Thus, a dilation is performed so that the segmentations of the bones that will be used during registration contain the contours of the latter.

Let X an image. For a structuring element B , the dilation of X by B is the result obtained by replacing each pixel p of X by its window B_p :

$$Dil_B(X) = \cup \{B_p | p \in X\}$$

The effect of dilation is to enlarge the image of a radius depending on the dimensions of the structuring element. In our case, we used a 3D structural element with the shape of a ball (Fig.5.6) so that the convex corners of the mask are rounded.

Dilation with radii ranging from 1 to 5 were performed. During registration, the most appropriate expansion radius will be determined for each bone depending on the accuracy of the registration.

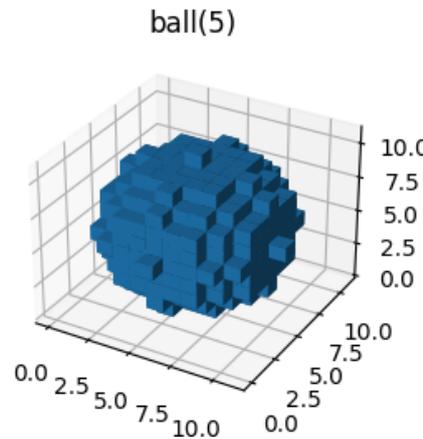


Figure 5.6: Structuring element used for the dilation of the masks of the ankle bones (here with a radius of 5)

Therefore, the dilated masks of the bones will be used to give more weight to the region of the bone of interest during the registration of the latter. However, we do not want the weight assigned to each voxel to be binary (1 if the voxel belongs to the studied bone, 0 otherwise). For this, a Gaussian filter of standard deviation $\sigma = 2$ is applied to the dilated mask in such a way that the weight of the voxels in the center of the bone is maximum and gradually decreases as they move away. Thus, during registration, the closer the voxels are to the center of the bone, the more impact they will have in the similarity measurement.

Examples of masks obtained after these operations for the 3 bones of interest is proposed on Fig.5.7.

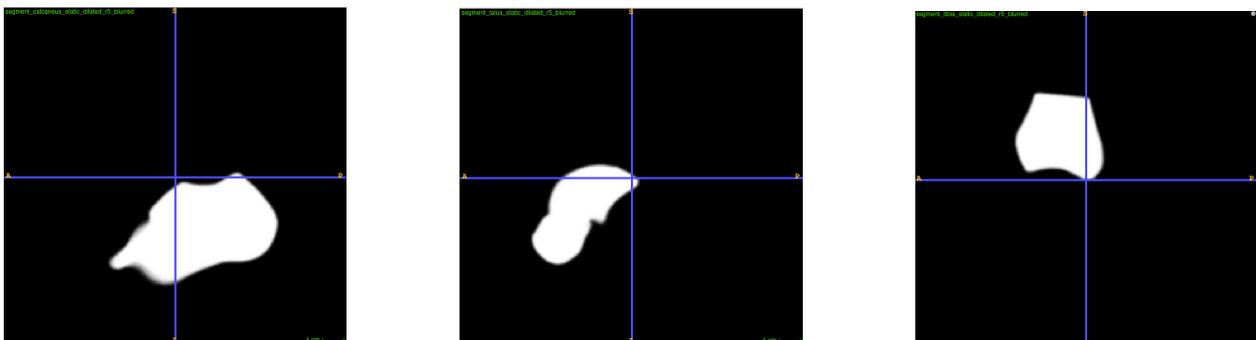


Figure 5.7: Dilated and blurred masks of the calcaneus (left), talus (center) and tibia (right) with a radius of dilation of 5

Once the blurred masks of the three ankle bones have been obtained, it is possible to move on to the registration step which is essential for obtaining HR dynamic sequences.

5.2.4. Creation of HR dynamic sequences by rigid registration

Before reading this section, it should be remembered what registration is and what is our objective. Registration is the matching of images assumed to represent the same object. Let I and J be two images and \mathcal{F} a family of spatial transformation. The registration problem is written as:

$$\arg \min_{T \in \mathcal{F}} \text{Simil}(T(I), J)$$

You can learn about registration in the [appendix](#) or through [12]. To perform the registration operations, the [FSL Flirt linear registration tool](#) was used. [8].

The objective of this operation is to obtain dynamic volumes of high resolution by registering the static HR MRI on the LR volumes composing the dynamic MRI (Fig.5.8). The HR volumes obtained at the end of this operation will be used during the deep learning to automate the reconstruction of the dynamic volumes (that is to say to increase the resolution of the dynamic volumes). Since the regions of the bones are of particular interest to this work, a registration for each bone will be carried out so as to obtain 3 dynamic HR volumes (one for each bone) where the regions of the same bone on the dynamic HR and LR volumes overlap perfectly.



Figure 5.8: Static HR MRI (left) and initial LR dynamic volume (right) of a control subject

First of all, it is important to adapt the registration parameters to our problem:

- **the similarity criterion:** The correlation coefficient, default parameter, was used because it is perfectly suited to an intra-patient registration between different types of MRI acquisition (static and dynamic).
- **the considered spatial transformations:** A rigid registration (6 DoF) was considered due to the rigid nature of the bones (from one image to another, bones do

not deform, they can only undergo translation and/or rotation). Moreover, the rigid registration is well suited for the intra-subject registration.

Firstly, a rigid registration of the entire static MRI on each of the volumes (without giving more importance to the bone regions) was performed in order to obtain a first dynamic HR volume not yet very precise (Fig.5.9).

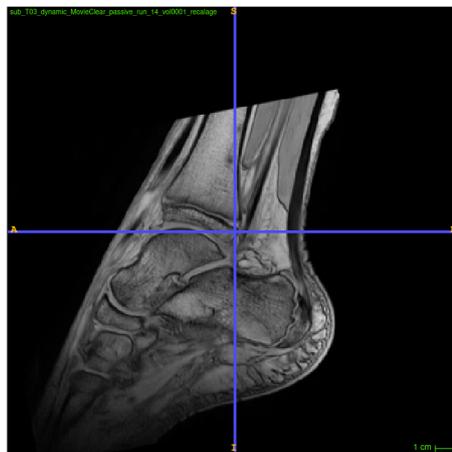


Figure 5.9: Dynamic HR volume obtained by registering the static MRI on a dynamic LR volume

At this stage, it is important to measure the quality of the registration. To this way, [correlation](#) between the registered image and the initial dynamic volume was calculated. However, it is difficult to estimate from which correlation value the registration is correct. How to estimate a validity threshold for the registration?

For one of the equine subjects, a bones segmentation on one dynamic MRI was available. Consequently, by applying the transformation estimated by registration to the three segmentations of the static MRI, it is possible to compare the masks of each bone and thus to assess the quality of the registration. For each ankle bone, the similarity between its segmentations (manual and obtained by registration) was measured by calculating the **DICE** coefficient.

Let X be the initial segmentation of the tibia on a dynamic sequence and Y the segmentation of the tibia registered on a dynamic volume. We denote by $|Z|$ the number of voxels of a volume Z . DICE is then calculated as:

$$dice(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|}$$

DICE offers a result between 0 and 1, 1 indicating perfect similarity. In our case, given that this first registration is not precise, we can estimate that the registration is acceptable if

the DICE between masks is greater than 0.85. Thus, by representing the value of the DICE as a function of the correlation between the LR dynamic volume and the registered image, it is possible to estimate a registration validity threshold at 0.64 (Fig.5.10).

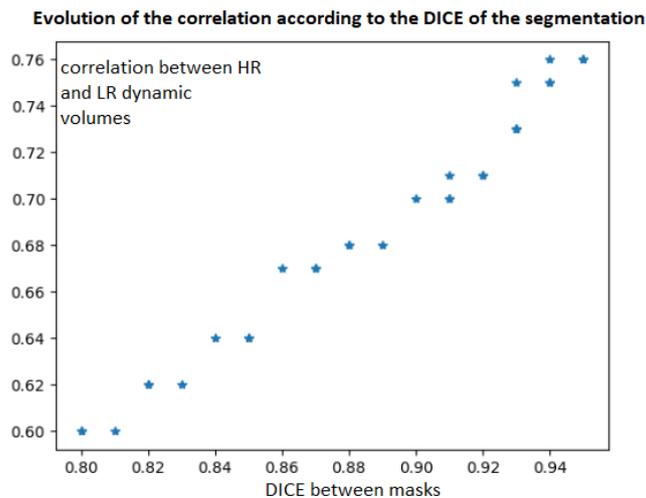


Figure 5.10: Evolution of the correlation between dynamic LR volume and HR registration as a function of the DICE measured between the segmentations of each of the three bones

When the registration is considered invalid, the dynamic volume is abandoned that is, it will not be used for deep learning.

Otherwise, a bone-to-bone registration of the volume is carried out in order to obtain three HR volumes registered with great precision in the regions of the calcaneus, the talus and the tibia. For that, more weight was assigned to voxels belonging to the bones. This is when the blurry masks created previously come into play. Thus, for the calcaneus for example, a new rigid registration of the static MRI on the dynamic LR volume is launched starting from the estimated transformation of the first registration (in order to accelerate the process and improve the final result) and attributing more weight to the voxels of the static MRI belonging to the blurred mask of the calcaneus. The greater the distance of the voxels from the center of the bone, the less importance they will have for registration. The voxels external to the blurred mask will not have an impact in the registration process.

In order to choose the most suitable dilation radius for each bone, a visual comparison of the registration results with the five dilation radii for all volumes of a dynamic sequence of a control subject was performed. The following decision was taken:

- calcaneus: radius of 1 (this bone is bigger than the other two)

- talus and tibia: radius of 2

On Fig.5.11 the result of the bone-to-bone registration for the static MRI and the dynamic volume represented on Fig.5.8.

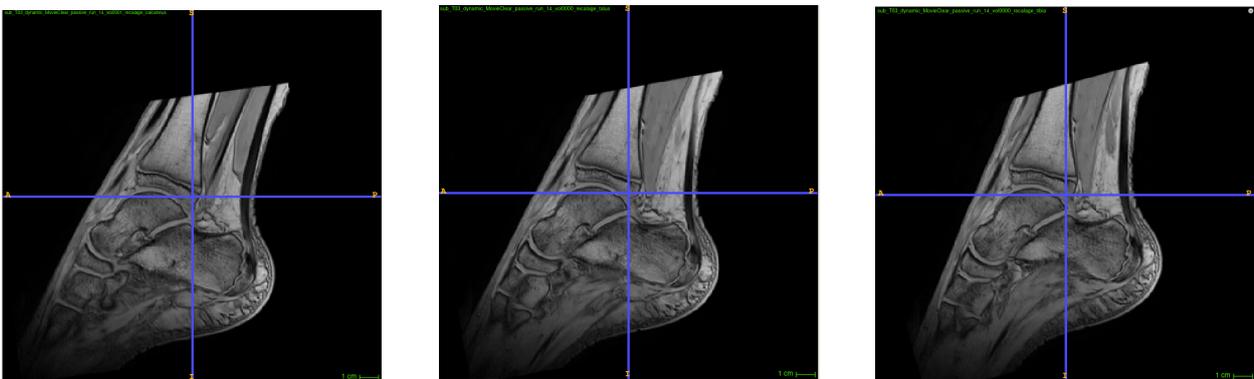


Figure 5.11: Results of rigid registration from left to right of the calcaneus, talus and tibia

Now there is only one step left to prepare the data for deep learning. Indeed, to automate the segmentation of the ankle bones on dynamic volumes (i.e the 3rd challenge), it is necessary to give the neural network segmentation images on LR volumes. For that, the initial manual segmentations on static MRI were used. Since we know thanks to the registration the existing spatial transformation between static MRI and dynamic volume, it suffices to apply this transformation to the initial segmentations to obtain a segmentation of the 3 bones on the LR volumes (Fig.5.12).

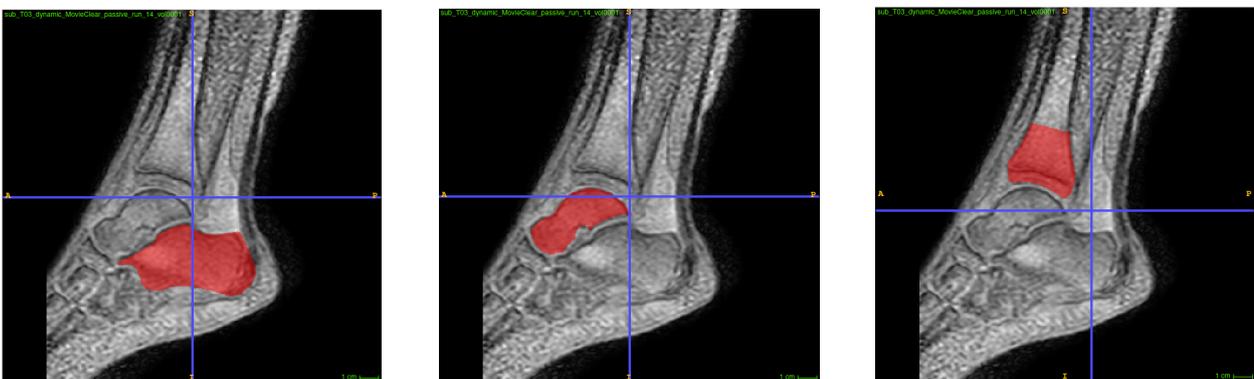


Figure 5.12: Registration on dynamic LR volume from left to right of the segmentations of the calcaneus, talus and tibia

The quality of the bone-to-bone registration is judged visually using the ITKSnap tool and stored in a spreadsheet for each volume of each dynamic sequence for each subject. Incorrect bone-to-bone registrations (that is where the registered segmentation of the bone is not visually overlaid with the bone on the dynamic volume) will be excluded from the deep learning data.

In total, for deep learning:

- 18 LR dynamic sequences of 15 volumes of control subjects were registered (or $3 \times 18 \times 15 = \mathbf{810}$ **registrations** bone-by-bone).
- 14 LR dynamic sequences of 15 volumes of subjects with equine were registered (or $3 \times 15 \times 15 = \mathbf{675}$ **registrations** bone-by-bone).

Now that the MRI data is corrected and prepared, it is possible to begin deep learning in order to attempt to meet the three challenges defined previously.

6. Learning MRI data with neural networks

The objective of this part is to use deep learning techniques to facilitate and improve the diagnosis of MRI data by automating the following three operations:

1. Segmentation of the three ankle bones on static high-resolution MRI.
2. Reconstruction of high-resolution dynamic volumes.
3. Segmentation of the three ankle bones on dynamic low-resolution MRI.

The neural networks were implemented using the framework **PyTorch** (see appendix). This latter, unlike TensorFlow, works with dynamic graphs that is to say it creates a new graph at each forward loop.



Figure 6.1: Nvidia Titan X graphics card used to perform deep learning calculations, image taken from [4]

In order to accelerate the computing times, they are carried out on the Nvidia Titan X GPU (Pascal 2016) of Fig.6.1 having:

- 3,584 CUDA cores clocked at 1.5 GHz for parallel operations
- 12 Go GDDR5X graphics memory

The programming of the GPU was carried out using CUDA 9.1 technology (see appendix) compatible with the nvidia 390.138 driver installed.

In order to have more training data, the strategy consists in extracting 2D patches of MRI data in gray levels and therefore in training neural networks in 2D. So, for each of the three problems, the input will be a 2D patch and the output will also be a 2D patch. Deep learning based on the extraction of 2D patches has already proved its worth, particularly for image denoising [24]. However, the issues we are talking about are a different challenge.

Regarding the partitioning of MRI data, we proceeded as such:

- MRI data of 6 control subjects for **training** neural network.
- The last control subject for **validation** in order to fix the hyperparameters of the network (architecture, number of layers, loss function, learning rate).
- **Performance evaluation** of neural networks with MRI data from the 7 equine subjects.

It was decided to test three different architectures described in the next section.

6.1. The neural network architectures tested

Three architectures were chosen because of their recognized efficiency in segmentation and reconstruction problems applied to the medical field [11].

6.1.1. Fully Convolutional Network (FCN)

In a **FCN**, each layer is a three-dimensional array of size $h \times l \times d$, where h and l are spatial dimensions and d is the channel dimension. Its basic components (convolutional layers and ReLU-type activation function) operate on local input regions and depend only on relative spatial coordinates ([annexe](#)).

The architecture tested is therefore a succession of CONV, BatchNormalization and ReLU activation function layers as illustrated on Fig.6.2.

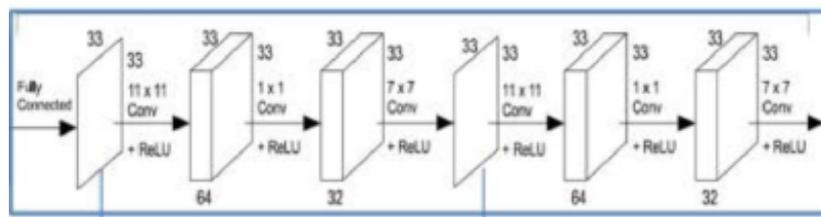


Figure 6.2: Example of FCN architecture for image reconstruction, illustration from [20]

This is a proven architecture for the segmentation of medical images [29] as well as for the reconstruction of medical MRI [18].

6.1.2. U-Net network

The U-Net architecture developed by Olaf Ronneberger et al [33] is based on a fully convolutional network model (**FCN**). It is an architecture widely used in biomedical imaging for **semantic segmentation**. It has the advantage of offering good results from a small data set.

The U-Net model uses various layers of neurons described in the [appendix](#):

- **CONV and MaxPool** layers which **downsample** the input data in order to obtain a lower resolution output volume and thus allow the model to learn what is present in the image.
- **Transposed convolutional** layers [35] which **upsample** the data in order to obtain an HR output volume from a LR input and thus allow the model to learn where each information is present in the image.

The network is divided into two paths. The first is the contraction path (also called **encoder**) which role is to capture the context or what the image contains. It combines CONV and MaxPool layers. The second path is the symmetric extension path (also called **decoder**) which uses transposed convolutional layers (Upsample) to locate information. Fig.6.3 shows an example of U-Net architecture adapted to the segmentation problem. To combine global and local contexts, the output of the transposed convolutional layers and the output of the encoder at the same level are concatenated at each step of the decoder. Because of this concatenation operation, it is necessary that the dimension of the input image be a power of two in order to avoid any rounding during the contraction which could provoke incorrect results in the dimension of the data and therefore prevent the concatenation.

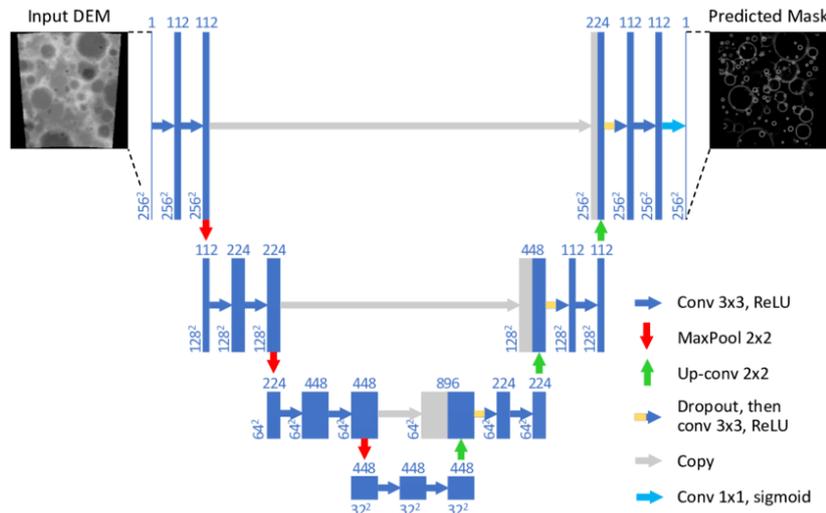


Figure 6.3: Example of U-Net architecture used for segmentation problem, illustration from [36]

6.1.3. Residual network (Resnet)

ResNet residual networks appeared in 2015 with the objective of providing a solution to the difficulty of optimizing very deep networks [28] by stacking residual blocks made up of 2 convolutional layers 3x3 (Fig.6.4).

Periodically, the number of filters is doubled and spatial subsampling is performed using a stride of 2 for convolution.

This type of architecture is highly appreciated in reconstruction and segmentation problems since it allows, unlike FCN, to take into account the difference between the input and the output of the layer. Due to this difference, the output evolutions are more progressive and more sensitive to details and textures.

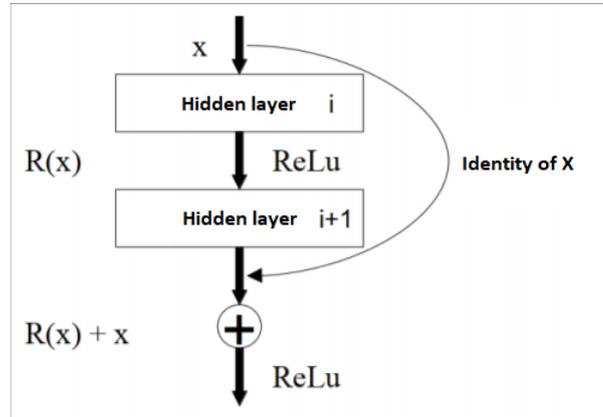


Figure 6.4: Residual blocks of ResNet networks, illustration taken from [11]

Two ResNet configurations will be tested. The first composed of classic residual blocks and the second using residual blocks combined with Bottlenecks (1x1 CONV layer) allowing a reduction in depth and therefore a reduction in the total number of parameters.

This is a proven architecture for the segmentation of medical images [19] as well as for the reconstruction of medical MRI [31].

It is now time to compare the performance of these different architectures for the segmentation and reconstruction of the MRI data.

6.2. Issue 1: Segmentation of the ankle bones on static HR MRI

The aim of this first problem is to automate the segmentation of the calcaneus, talus and tibia on static HR MRI. Thus, from a 2D patch of the HR image, we want to obtain the same 2D patch with the segmentation of the different bones.

6.2.1. Network data extraction

The 2D patches were obtained from the following data:

- Static HR MRI of each of the 14 subjects (control and with CP)
- The three segmentations (calcaneus, talus and tibia) associated with static MRI

However, the intensities of the MRI of the different subjects can present different orders of magnitude preventing any comparison between them. Therefore, the first step is to center and normalize the data to force the intensities of all the input images into one region. From the histogram (Fig.5.4) of static MRI, we observed that the intensities of the image corresponding to the region of the foot present a Gaussian distribution. Thus, from the foot mask previously extracted on the static MRI, it is possible to standardize the voxels of the image based on the distribution in the region of the foot (so as to obtain a distribution of zero mean and unit standard deviation for all subjects) according to the expression:

$$z = \frac{x - \mu[x \in M]}{\sigma[x \in M]}$$

M: foot mask

x: image voxel value

$\mu[x \in M]$: average of the intensities of the voxels belonging to the foot

$\sigma[x \in M]$: standard deviation of the intensities of the voxels belonging to the foot

Fig.6.5 represents the extraction of the 2D input (X) and the expected output patches (Y) for the neural network.

The size of the 2D patches to be extracted has been fixed at **128x128** in order to be large enough to contain several structures (bones, muscles, ...). Furthermore, only the patches for which at least 75% of the pixels belong to the foot were kept so as not to work from images of the background of the MRI.

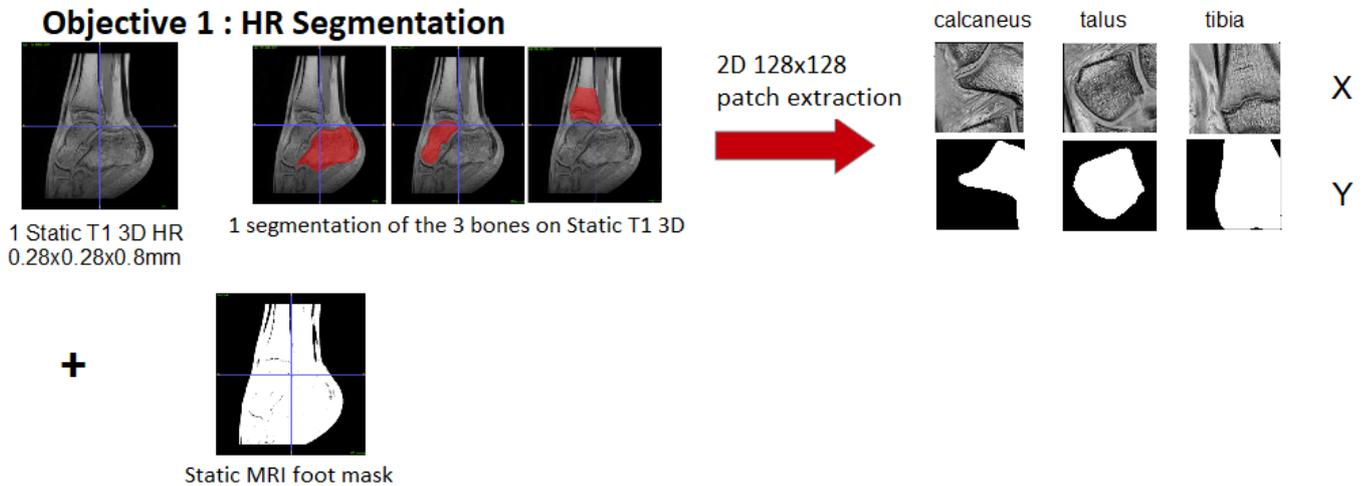


Figure 6.5: 2D patch extraction for the problem of automatic segmentation of static MRI

In total have been extracted for deep learning:

- **44924 patches of the static MRI of the 7 control subjects** and the patches of the associated segmentations which will be used for the training of the neural network and the validation of the associated hyperparameters.
- **43398 patches of the static MRI of the 7 equine subjects** and the patches of the associated segmentations which will be used to evaluate the performance of the neural network.

6.2.2. Neural network training

It was decided to implement a neural network for each bone. These three neural networks will have exactly the same configuration.

Each network was trained with the 2D patches of 6 control subjects. The last control subject was used for the validation of the network parameters. The **Binary Cross Entropy (BCE)** loss function was used because it has the advantage, unlike other functions implemented in PyTorch (CrossEntropyLoss, BCEWithLogitsLoss), of not integrating the sigmoid activation function. In fact, it is more natural to have probability data between 0 and 1 at the output of the network in order to then apply the loss function to these data.

The number of epochs and the size of the batches used for learning have been set to values that make it possible to obtain a convergence of the results while ensuring a relatively short training time (about 1 hour):

- number of epochs: 5
- size of batches: 8

Three other factors were compared in order to define the most suitable network for the HR segmentation problem:

1. The architecture
2. The network depth (defined according to the number of parameters)
3. The learning rate

In order to compare the quality of the results, the **DICE** (see section 5.2.4) between the expected segmentation and the segmentation obtained at the output of the network was calculated. We hope to obtain on the validation data an average DICE approaching 0.9. Moreover, the **complexity** of the network, represented by its number of parameters, is a good point of comparison. Indeed, we want to optimize the results of the segmentation while minimizing the complexity of the model.

a) Architecture

The three architectures defined in the previous point (FCN, U-Net and ResNet) were compared with:

- learning rate = 0.0005
- a standard depth for each architecture so as to reach a channel dimension equal to 128

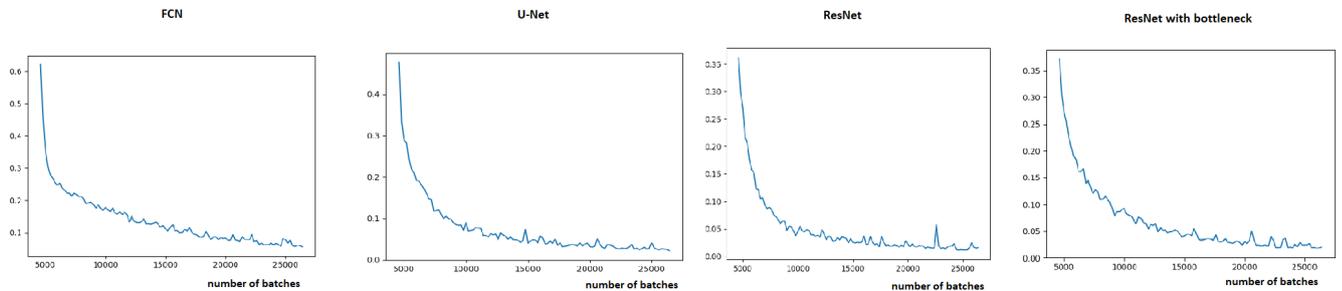


Figure 6.6: Training of the 4 architectures for the segmentation of the calcaneus (on the abscissa the number of batches)

Regarding ResNet, versions with and without bottleneck layers have been distinguished.

The evolution of the loss function during the training of the segmentation of the calcaneus for the 4 architectures is represented on Fig.6.6.

In the table 6.1 the comparison of the four architectures for the segmentation of the three bones is summarized.

Network	Number of parameters	DICE average validation calcaneus	DICE average validation talus	DICE average validation tibia
FCN	390 609	0.29	0.32	0.35
U-Net	1 179 569	0.8	0.81	0.92
ResNet with basic blocks	8 161 729	0.81	0.82	0.81
ResNet with bottlenecks	8 535 489	0.62	0.6	0.64

Table 6.1: Comparison of the different neural network architectures for the segmentation of HR MRI on validation data (lr = 0.0005)

From these results, we can say that the FCN model implemented (succession of CONV, BatchNormalization and ReLU layers) is too simple to solve the segmentation problem. Also, the use of bottleneck layers in the ResNet worsens its performance.

U-Net and ResNet architecture are the ones giving the best results. We can note the good performance of U-Net for the segmentation of the tibia. Although ResNet is a more complex architecture, it is interesting to continue the comparison with these two models given the results observed.

b) Network depth

For the two retained architectures (U-Net and Resnet with basic blocks), the performance of the network as a function of its depth were compared by fixing the learning rate at 0.0005.

The evolution of the DICE as a function of the number of parameters for the two architectures on the training and validation data is represented in Fig.6.7. This representation was obtained from the results of the segmentation of the calcaneus.

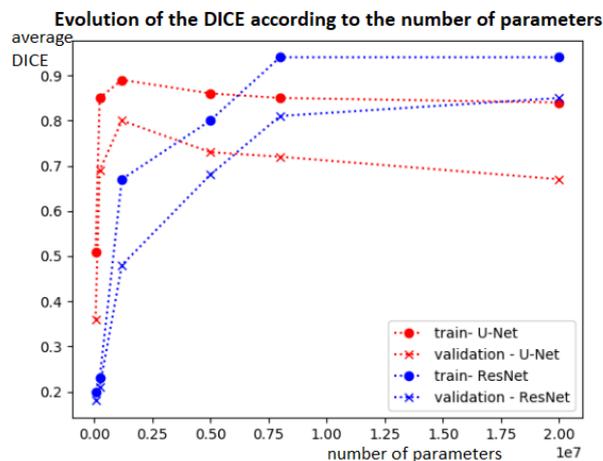


Figure 6.7: Evolution of the average DICE according to the number of parameters for the segmentation of the calcaneus by the U-Net and ResNet networks on the training (points) and validation (cross) data

From Fig.6.7, it is possible to define the number of parameters of each architecture maximizing the performance of the network while minimizing the complexity:

- U-Net: 1 179 569 parameters
- ResNet: 8 161 729 parameters

We can clearly see that ResNet needs more parameters to achieve equivalent results to U-Net. However, both architectures offer good performance. That's why it is worth continuing to study both architectures with the test data. The two preserved models are represented on Fig.6.8.

All that remains is to define the value of the learning rate optimizing the DICE on the validation data.

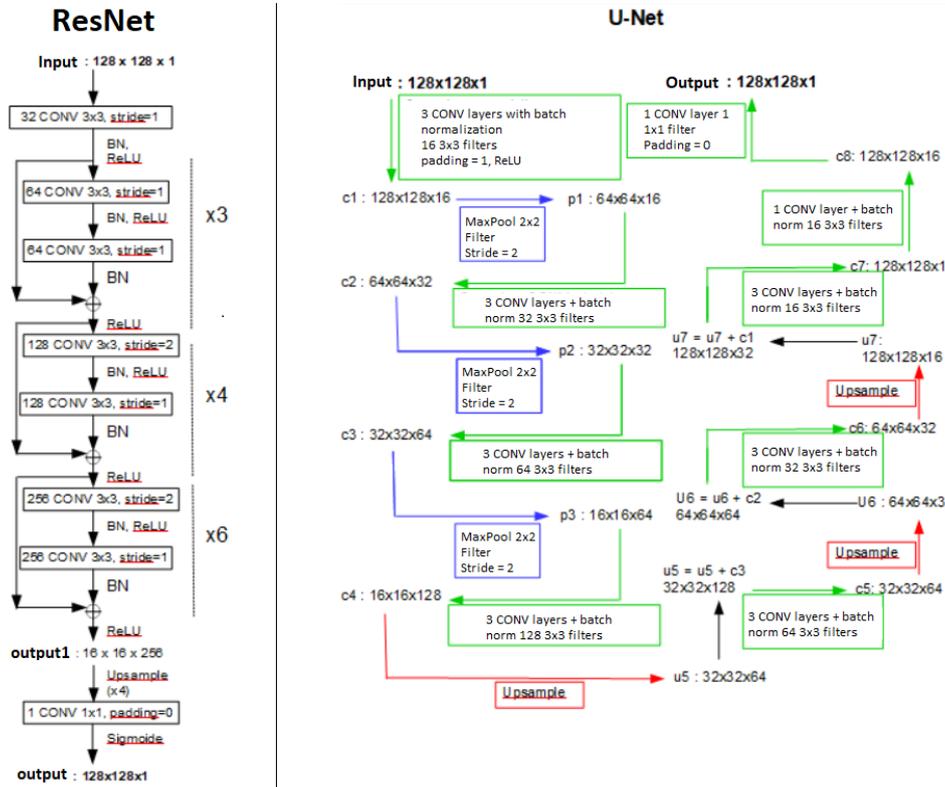


Figure 6.8: Model of U-Net and ResNet neural networks chosen for the segmentation problem of static HR MRI

c) Learning rate

The evolution of the DICE according to the learning rate on the training and validation data for the segmentation of the calcaneus with the two architectures is represented on Fig.6.9.

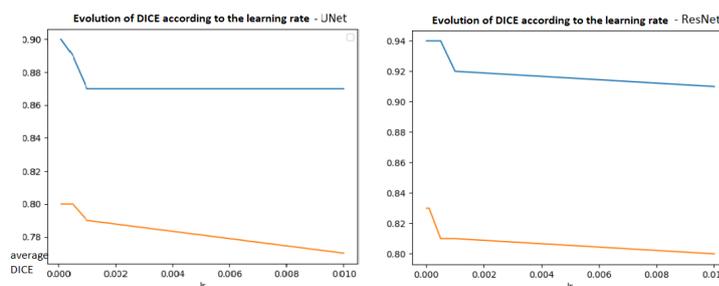


Figure 6.9: Evolution of the average DICE according to the learning rate for the U-Net and ResNet networks on the training (in blue) and validation data (in orange)

From this representation, the learning rate was set at **0.0005** for U-Net and **0.0001** for ResNet.

The average DICE obtained on the training and validation data is summarized in the table 6.2.

Data	Calcaneus	Talus	Tibia
U-Net Train/Validation	0.89/0.8	0.92/0.81	0.92/0.92
ResNet Train/Validation	0.94/0.83	0.94/0.82	0.94/0.82

Table 6.2: Average DICE obtained with U-Net and ResNet models on training and validation data for the segmentation of the 3 ankle bones

The results on the validation data for the three bones are all greater than 0.8. In order to verify that the two networks allow good learning and that these results do not depend on the control subject chosen for the validation, a cross validation was carried out.

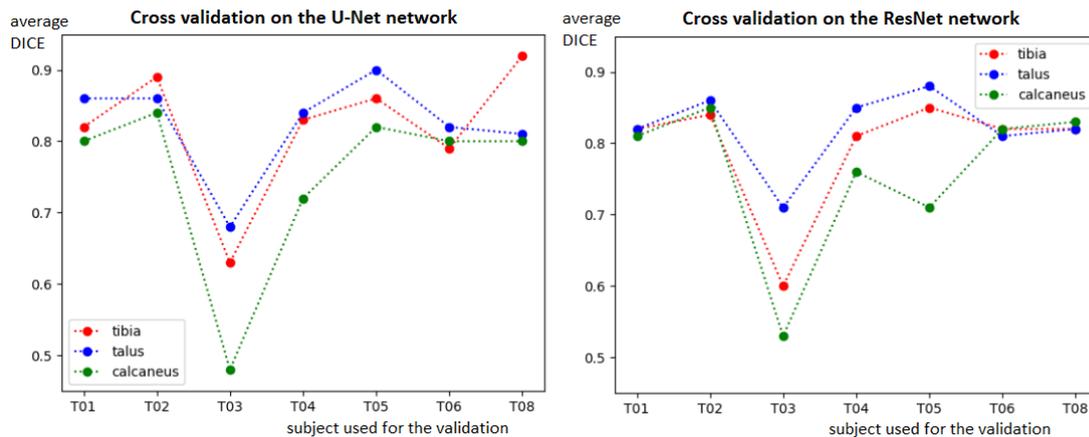


Figure 6.10: Evolution of the average DICE obtained on the validation data according to the partitioning of control subjects for U-Net and ResNet networks

On Fig.6.10, we can see the average DICE obtained for the validation data as a function of the control subject used. It can be seen that the results obtained for the three bones

remain satisfactory (around 0.8) whatever the partitioning used. We can note an exception for subject T03 with which validation fails. By observing the initial static MRI of this subject, I realized that his plantar flexion movement was much less marked than for the other subjects, thus inducing a different inclination of the ankle bones. We can therefore make the following statements:

- Both networks are able to correctly segment the ankle bones.
- It seems that the learning is too dependent on the orientation of the bones.

Now, we will compare the ability of the two networks to generalize over more 2D patches to segment the 3 ankle bones. For this, we will use the 43398 2D patches extracted from static MRI of subjects with an equine.

6.2.3. Performance comparison of U-Net and ResNet networks

Network performance was evaluated from 2D patches extracted from data from subjects with an equine. We also use the DICE to measure the quality of the segmentations obtained.

Examples of segmentation on test data with the two networks are shown on Fig.6.11.

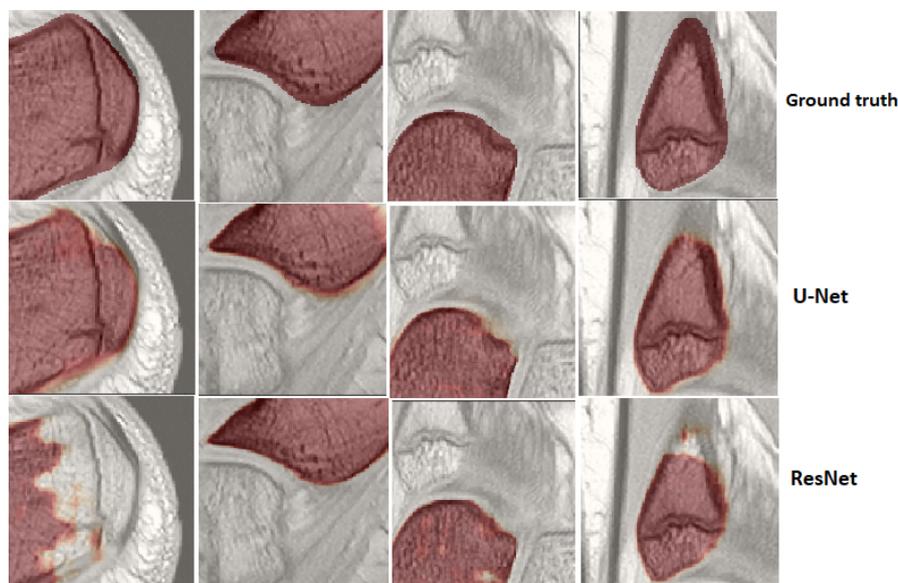


Figure 6.11: Examples of segmentations on 2D 128x128 patches extracted from equine subjects obtained with the U-Net (middle) and the ResNet (bottom) networks. The top image is the ground truth (manual segmentation).

We can observe on these images, that the two networks manage to correctly segment the different bones of the ankle on the 2D patches of dimension 128x128. However, it seems that U-Net draws the contours of the bone more accurately.

In order to have a more global visualization of the performance of the two neural networks on the test data, the 6.3 table lists the average DICE obtained for the segmentation of the 3 bones on the 43398 2D patches.

Network	Calcaneus	Talus	Tibia
U-Net	0.74	0.78	0.87
ResNet	0.8	0.81	0.8

Table 6.3: Average DICE obtained with the U-Net and ResNet models on the test data for the segmentation of the 3 ankle bones on 2D 128x128 patches

The average DICE obtained on the test data evolves around 0.8 for the three bones and for the two networks. This means that both models are able to generalize on other 2D patches of the same size as the patches used for training. One would tend to think that the U-Net model is more efficient for the segmentation of the tibia while the ResNet would be better for the calcaneus.

However, we would like the deep learning model to be efficient enough to segment entire 2D sections.

Consequently, for each of the 7 equine subjects, a 2D patch was extracted from each slice of the static MRI (dimension 576x576) and from the three associated segmentations. The 2D MRI patches were given as input to the two neural networks. In order to limit the impact of limit slices (on which the bones are not always visible) in the calculation of the mean DICE of the MRI sequence, the DICE was measured on the 3D volume by grouping all the 2D slices. The average DICE obtained for the 7 equine subjects as well as the minimum and maximum values are summarized in the table 6.4.

Network	Calcaneus	Talus	Tibia
U-Net average (min/max)	0.42 (0.26/0.53)	0.7 (0.52/0.79)	0.5 (0.27/0.63)
ResNet average (min/max)	0.41 (0.21/0.53)	0.42 (0.13/0.58)	0.6 (0.34/0.74)

Table 6.4: DICE obtained for the segmentation of 3D HR static MRI sequences of equine subjects with U-Net and ResNet models

Fig.6.12 allows to compare the segmentation of the three bones with the two neural networks on the same MRI slice.

It clearly appears from these results that neither of the two neural networks is efficient enough to generalize to 2D images reaching the size of an entire MRI slice (576x576). However, it seems by visualizing the Fig.6.12 and the overall network performance on table 6.4, that the U-Net network has more capacity to generalize especially for the talus (average

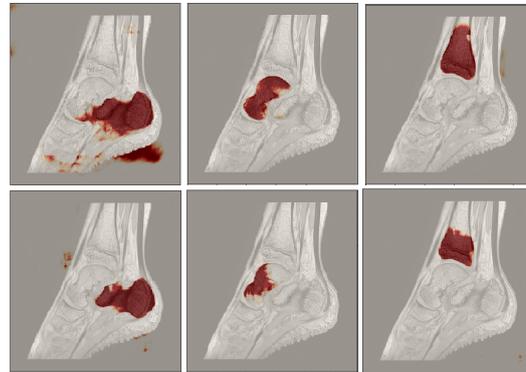


Figure 6.12: Segmentation of the 3 ankle bones on a static HR MRI slice obtained with the U-Net (top) and ResNet (bottom) networks

DICE of 0.7). However, the average DICE obtained for the calcaneus and the tibia is insufficient (0.42 for the calcaneus and 0.5 for the tibia).

At the end of this first study, my preferred choice for further work is the **U-Net** for different reasons:

1. Much less complex network and therefore faster to train
2. An acceptable generalization capacity on 2D slices for the talus
3. A more precise outline of the bones

In order to improve the results of the U-Net network, it would be interesting to try the following possibilities:

- Using data augmentation (zoom, translate, rotate, etc.) to reduce the importance of size and orientation during the training.
- Training the same neural network by gradually increasing the size of the patches until reaching a dimension of 576x576 while reducing the learning rate.
- Cutting the input image into overlapping 128x128 patches and applying the U-Net network on all the patches in order to reconstruct the segmentation of the entire 2D slice by grouping the results.

Thanks to deep learning, we set up an efficient automatic segmentation of the ankle bones on 2D patches of size 128x128 from static HR MRI. However, generalizing the segmentation to larger 2D images (the size of an MRI slice) is not good and remains a challenge to be pursued in future work.

6.3. Issue 2: HR reconstruction of dynamic volumes

The objective of this second problem is to automate the HR reconstruction of the ankle bone regions from dynamic LR MRI. Thus, from a 2D patch of the LR volume representing one of the bones of interest, we want to obtain the same 2D patch in high-resolution.

6.3.1. Network data extraction

The 2D patches were obtained from the following data:

- The volumes of dynamic LR MRI sequences of each of the 14 subjects
- HR volumes obtained by bone-to-bone registration of static MRI on dynamic volumes

The size of the 2D patches to be extracted has been fixed at **32x32** so as not to contain several different bones. The goal is to extract as many patches as possible from the three ankle bones of interest. In order to maximize the correlation between the HR patch and the LR patch, we will use the image registered with respect to the calcaneus to extract the patches from it, in the same way for the other bones.

Standardization of the data was, this time, carried out bone by bone given the small size of the patches extracted.

In addition, many tests have been carried out in order to define the proportion of pixels that must belong to each bone to consider the patch as correct. This threshold has to be large enough for the bone to occupy at least half of the image while still allowing a considerable amount of patches to be extracted from each bone. Depending on the size of the bone, this threshold varies:

- at least 70% of the pixels for the calcaneus
- at least 50% of the pixels for the talus
- at least 60% of the pixels for the tibia

Finally, with the objective of improving network performance, the correlation between the LR patch and the HR patch was calculated and stored with the input data of the neural network. The idea will be to include this measurement as a parameter of the network loss function in order to give more importance to the patches having a strong correlation between them. During the extraction of the patches, a minimum correlation threshold has been established so as to use for the training only patches that are well registered while ensuring the extraction of a high number of images. This threshold was set at 0.6 from the histogram Fig.6.13 representing the number of patches extracted from the dynamic sequences LR as a function of the correlation between the latter and the corresponding patch of the HR image.

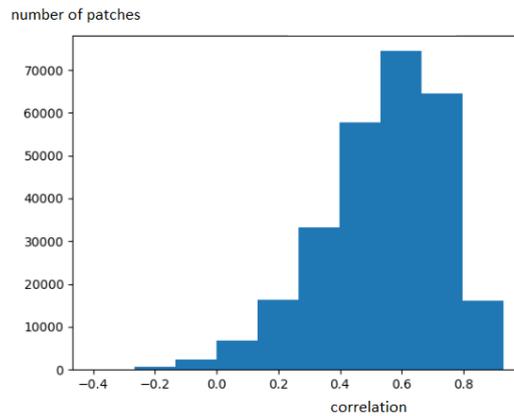


Figure 6.13: Histogram of the number of LR dynamic sequence patches extracted as a function of the correlation of the latter with the corresponding patch of the registered HR dynamic sequence

Fig.6.14 represents the extraction of the 2D patches of the LR input (X), of the expected HR output (Y) and of the correlation between the two corresponding patches (W) for the neural network.

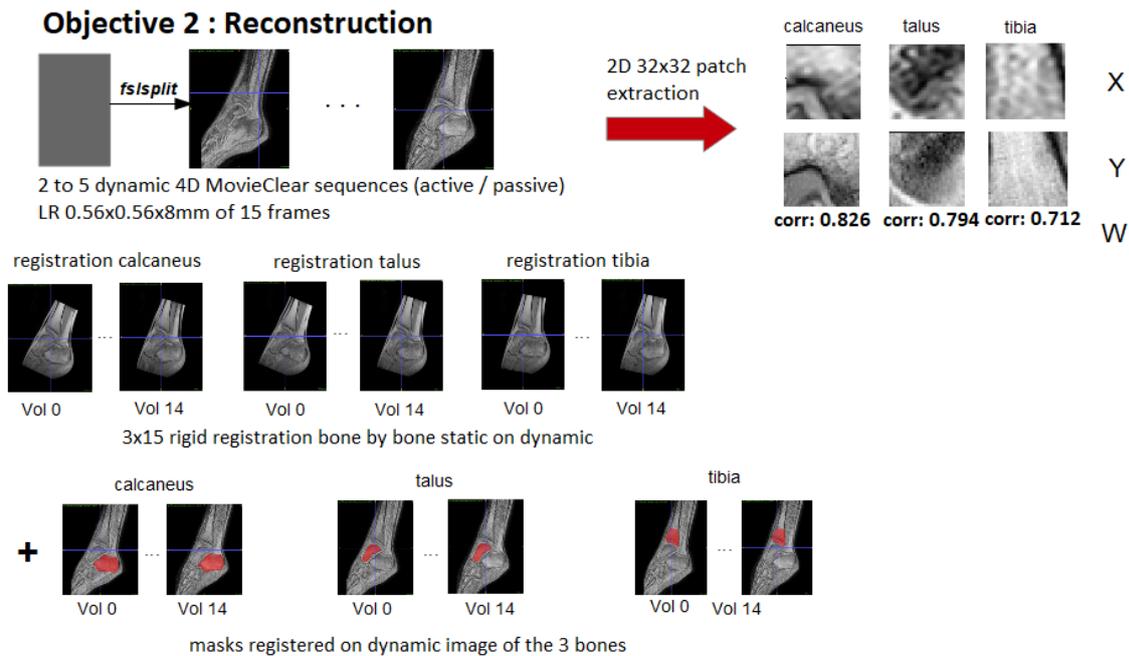


Figure 6.14: 2D patch extraction for the HR automatic reconstruction problem of dynamic volumes

In total have been extracted for deep learning:

1. **91402 32x32 patches of dynamic volumes of control subjects** and associated registrations which will be used for training the neural network. They are divided as such:
 - 37704 patches of the calcaneus
 - 15497 patches of the talus
 - 38201 patches of the tibia

2. **111865 32x32 patches of dynamic volumes of subjects with CP** and associated registrations which will be used for evaluating the performance of the neural network. They are divided as such:
 - 46043 patches of the calcaneus
 - 23633 patches of the talus
 - 42189 patches of the tibia

6.3.2. Neural network training

For this problem, only a residual network with basic blocks was tested because according to [11] and [31], it generally offers very good results for the reconstruction of medical images. The network was trained with 2D patches from 6 control subjects. The last control subject was used for network validation.

The number of epochs and the size of batches used for the training were set at values allowing to obtain a convergence of the results while ensuring a relatively short training time (approximately 45 minutes):

- number of epochs: 5
- size of batches: 16

The size of the batches has been increased compared to the segmentation problem because more training data is available for reconstruction. Moreover, since the patches are smaller, it is possible to double the size of the batches.

As for the segmentation problem, the Adam optimization algorithm was used because according to [28] it allows rapid convergence of the gradient while avoiding convergence towards local minima.

To compare the quality of the results, **correlation** between the expected reconstruction and the output of the network was calculated. In addition, a difference map has been represented in order to know the type of information lost during learning.

In order to give more importance during training to pairs of patches (HR and LR) having a strong correlation between them, I personalized the network loss function \mathcal{L} . The idea was that this impact should not be linear. A function g has been defined in such a way that for a correlation $x > 0.7$, $g(x)$ is close to 1 and otherwise close to 0. With the Solumaths tool, the following sigmoidal function represented on Fig.6.15 has been defined: $g(x) = \frac{1}{1+e^{-10*(x-0.5)}}$

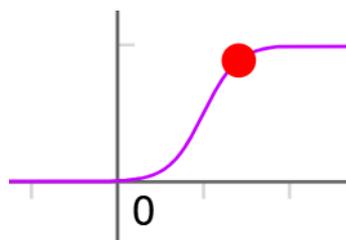


Figure 6.15: Sigmoidal function used to define the influence of patches in deep learning according to their correlation (representation made on Solumaths, the red point represents $x = 0.7$)

Thus, the loss function is of the form:

$$\mathcal{L} = \sum_{i=1}^N g(\text{corr}) * d(f(x_i), y_i) \text{ with}$$

$f(x_i)$: 2D patch obtained from the ResNet
 y_i : HR 2D patch expected

I started by studying the impact of the depth of the network on its performance by setting the learning rate to 0.001. The evolution of the correlation as a function of the number of parameters on the training and validation data is represented on Fig.6.16.

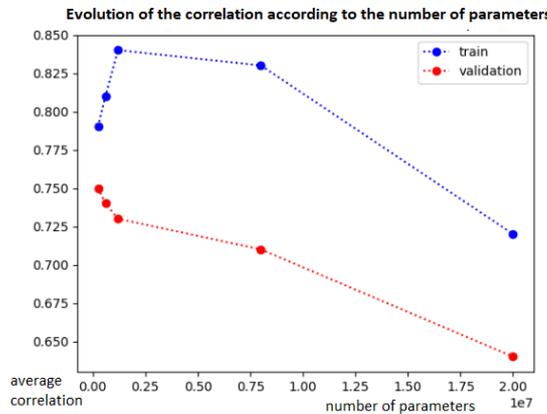


Figure 6.16: Evolution of the performance of ResNet for reconstruction according to the number of parameters for training (in blue) and validation data (in red)

We immediately see that the network does not learn correctly since the validation results are constantly decreasing. This poor learning can come from the fact that the initial feature extraction is carried out on very small regions which do not allow the LR images to be properly characterized.

However, we can retain from this study that it is not necessary to use a ResNet with more than 3 million parameters to obtain acceptable results.

The last test was to compare the use of two different image distance measurements for the loss function:

- **L1 distance** defined as $d(I_1, I_2) = \sum_{i=1}^p |I_1^p - I_2^p|$
- **Euclidean distance L2** defined as $d(I_1, I_2) = \sqrt{\sum_{i=1}^p (I_1^p - I_2^p)^2}$

Table 6.5 summarizes the results on training and validation data with the two measurements.

Distance measure	Training	Validation
L1	0.83	0.71
L2	0.84	0.71

Table 6.5: Average correlation obtained on training and validation data using L1 and L2 distance measurements for the loss function

Both distance measurements give similar results. However, due to their behavior depicted on Fig.6.17, the L1 was chosen because it smooths less the extreme values (ends of the parabola).

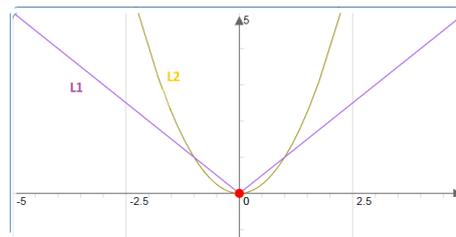


Figure 6.17: Graphic representation of the distance measurements L1 (in purple) and L2 (in orange) (representation made on Solumaths)

The current results on the validation data are similar to those of Fig.6.18.

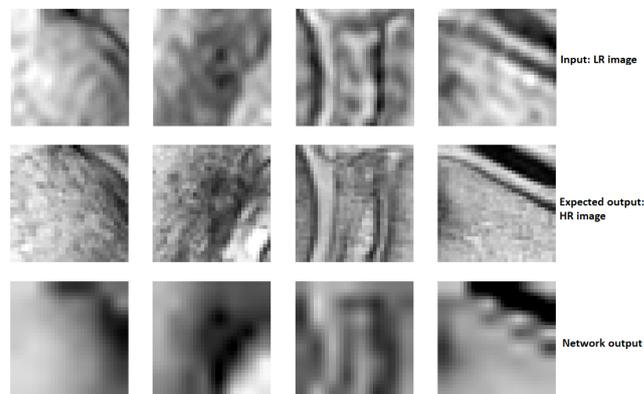


Figure 6.18: Reconstruction results obtained with the ResNet network on the validation data. From top to bottom, the input LR image, the expected HR output and the network output.

Results are far from being satisfactory since the image obtained (last line) is smooth and without texture. By drawing the difference map between the ground truth and the learning output (Fig.6.19), it is possible to view the information lost during learning. Overall, a lot of information is lost including a good part of the contours as well as the texture of tissues and bones.

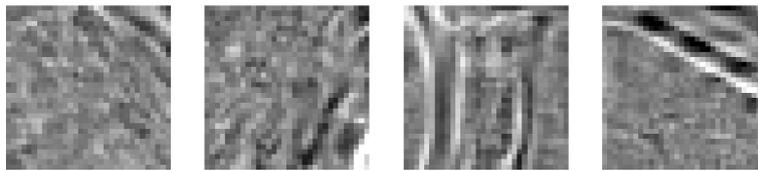


Figure 6.19: Difference map obtained for reconstruction results on validation data with the ResNet network

Since the current results on validation data are not good, the performance of the network on test data extracted from equine subjects has not been studied.

The reconstruction of dynamic HR sequences by deep learning is not successful at all. Different potential solutions should be tested to improve results:

1. Measuring the quality of the reconstruction with a PSNR (Peak Signal to Noise Ratio) more suited to assess the quality than the correlation.
2. Learning the function g from the loss function weighting the influence of the patches in the learning as a function of the correlation (avoiding the trivial case: $g(\text{corr}) = 0 \forall \text{corr}$).
3. Reviewing the first layers of ResNet in order to enlarge the feature extraction space (adding convolutional layers or increasing the size of the filters used).
4. Testing other architectures such as U-Net or FCN.

6.4. Issue 3: Segmentation of the ankle bones on dynamic LR volumes

The objective of this last problem is to automate the segmentation of the calcaneus, talus and tibia on dynamic LR MRI. Thus, from a 2D patch of the dynamic LR volume, we want to obtain the same 2D patch with the segmentation of the different bones present on it.

6.4.1. Network data extraction

The 2D patches were obtained from the following data:

- the volumes of dynamic MRI sequences of each of the 14 subjects
- the three initial segmentations of the static MRI registered on the dynamic volume by application of the transformation estimated during registration

The technique for normalizing the data for training was the same as for the HR segmentation problem. From the foot masks previously extracted on the dynamic volumes, the voxels of the image were standardized so as to obtain a zero mean and unit standard deviation distribution for all the volumes of all the subjects.

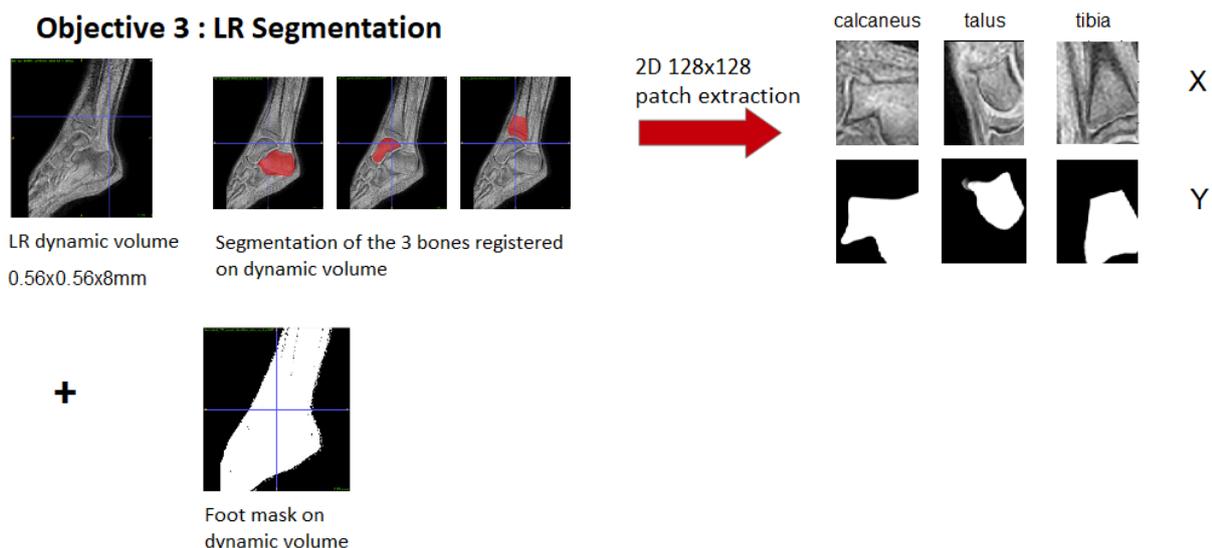


Figure 6.20: 2D patch extraction for the automatic segmentation problem of dynamic MRI

Fig.6.20 represents the extraction of the 2D input (X) and expected output (Y) patches for the neural network.

The size of the 2D patches to be extracted has been fixed at **128x128** so as to be large enough to contain several structures (bones, muscles, ...). Furthermore, only the patches for which at least 75% of the pixels belong to the foot were kept so as not to work from images of the background of the MRI. The difficulty here is based on the fact that for a 3D volume, we do not necessarily have the 3 segmentations. Indeed, it is possible that the registration with respect to one or more bones has failed. This decreases the amount of data for training.

In total, were extracted for learning:

1. Patches extracted from control subject data for training:

- 44802 patches of calcaneus
- 32617 patches of talus
- 37624 patches of tibia

2. Patches extracted from data from equine subjects to assess network performance:

- 29195 patches of calcaneus
- 28791 patches of talus
- 29224 patches of tibia

6.4.2. Prospects

Due to lack of time, deep learning for the ankle bones segmentation problem on dynamic MRI was not initiated. This problem is similar to the first one dealing with the segmentation of static MRI.

From the results obtained with the segmentation of static MRI, we can attempt to use a U-Net architecture similar to the one implemented and to train it with the patches extracted from dynamic LR MRI and associated segmentations.

7. Conclusion

At the end of this work, an automatic segmentation of the three main ankle bones on 2D static MRI patches was set up through two neural network architectures (U-Net and ResNet). These two models were compared on more than 40000 2D test patches using measurement tools recognized in the literature. The objective is now to generalize these performances to whole 2D slices of MRI.

In addition, all of the MRI data acquired at the CHU of Brest are now corrected, transported to a common spatial reference and registered in order to feed other neural networks intended for the automatic segmentation of the ankle bones on dynamic MRI or for the reconstruction of dynamic HR sequences. Although the work carried out is only a tiny step forward in the world of research, it marks the beginning of the development of new MRI image analysis tools for the study of the ankle in motion. In order to develop effective MRI analysis techniques, steps remain to be taken:

1. Generalizing the segmentation of the ankle bones on static MRI to full 2D slices by exploiting various approaches (data augmentation, U-Net training with larger patches, segmentation of 2D slices by dividing them into overlapping patches).
2. Improving reconstruction results by exploring new architectures and using more standardized quality metrics such as PSNR.
3. Starting the implementation of an automatic segmentation of the ankle bones on dynamic MRI from the extracted patches.
4. Implementing a unique network for the segmentation of the 3 bones on static or dynamic MRI.

In the long term, the robustness of these tools will contribute to an acceleration of analysis times and especially to a better understanding of the joint biomechanics essential for the implementation of therapies adapted to the deformities linked to cerebral palsy.

The experience lived within the LaTIM was very enriching in many aspects which will be essential to me in the success of my doctoral thesis undertaken since the month of November. The latter focuses on the use of deep learning and structural knowledge for the analysis of porcine cerebral MRI, once again intended for a better understanding of cerebral palsy.



Thus, thanks to these end-of-master work, I made a first discovery of deep learning applied to medical imaging by handling many tools such as PyTorch or FSL. Then, the ability acquired over the past few years to read scientific papers, to follow specific deep learning courses, to discuss with other doctoral students and to document myself on image processing applied to the field of health will be important to effectively start this thesis. Finally, evolving within a dynamic, generous team that shares its knowledge on cerebral palsy with pleasure allowed me to leave with a first medical background. I am more motivated than ever to continue the research on this pathology which affects many children every year.

A. Appendix

A.1. Gantt chart

Gantt chart

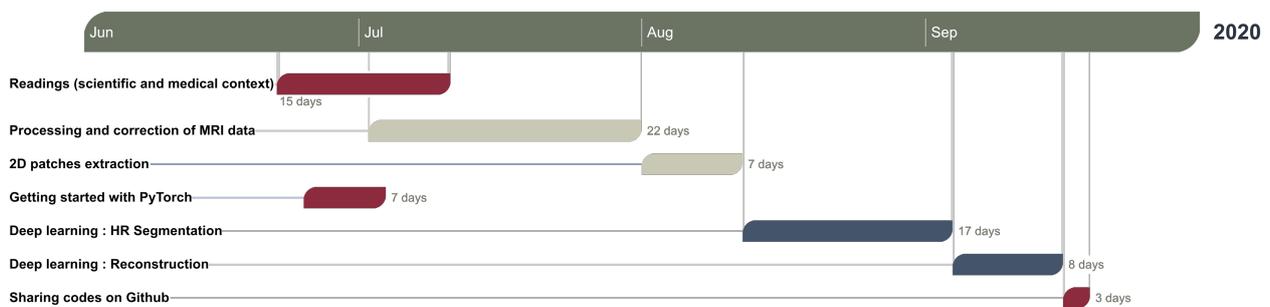


Figure A.1: Gantt chart realized with *Office Timeline*

A.2. Nibabel coordinate system

A Nibabel image is the association of 3 elements:

- 1 N-D array of the image data (intensity)
- a 4x4 matrix of the image coordinates in an affine reference coordinate system
- an header with the image metadata

The coordinates of the image do not give any information about the position in the world, that is, in the MRI. This means that it is impossible to compare the data of two subjects because they were not acquired under the same conditions. Two images are comparable only if they are defined in a **common reference system**.

The common spatial reference is defined by an origin and three axes generally referenced in relation to the position of the patient:

- Origin = part of the patient being examined
- Left-right axis
- Anterior-posterior axis
- Inferior-superior axis

To compare two images, it is therefore essential to ensure that their system of reference is the same: ex. RAS scanner (right, anterior, posterior), LPI scanner (left, posterior, inferior). Moreover, the origin of this system of reference is only common for an MRI series of a patient if the latter does not move. That's why, even if the system of reference is the same between two images of different sequences, it is often necessary to make a registration between them so that they overlap.

The 4x4 affine transformation matrix of the image is such that:

Let (x,y,z) be the coordinates in the common spatial system of reference and (i,j,k) the

coordinates of the voxels of the image.

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} i \\ j \\ k \\ 1 \end{bmatrix}$$

A.3. Some notions on registration

Registration is the matching of images assumed to represent the same object [27]. It is a fundamental tool in the analysis of medical images:

- reconstruction of a 3D volume from 2D sections
- analysis of temporal evolution (of brain development for example)
- subjects comparison
- overlay on an atlas
- fusion of modalities (MRI, CT, etc)

The aim of the registration is to estimate the spatial transformation T to go from one image to another. For this, we seek to optimize a similarity criterion between the two images [12]. Registration is based on three important choices:

1. **the similarity criterion**: which measures the "resemblance" between the two images. Several choices are possible, it is advisable to choose the most suitable for each problem:
 - **measuring the conservation of intensities**: SSD (to minimize). This solution is suitable for intra-patient mono-modal registration.
 - **measuring affine dependence**: correlation coefficient (to maximize). This solution is suitable for inter and intra-patient mono-modal registration.
 - **measuring statistical dependence**: mutual information (to maximize). This solution is suitable for multi-modal registration.
2. **the spatial transformations considered**: it is advisable to impose constraints on T which can be according to the number of degrees of freedom:
 - **homography** - 16 DoF (in 3D): suitable for multi-modal registration and to 3D / 2D registration.
 - **affine** - 12 DoF (in 3D): suitable for intra-subject multi-modal registration when the spatial resolution of the two images is different.
 - **rigid** (rotation+translation) - 6 DoF (in 3D): suitable for intra-subject registration.
3. **alignment techniques (optimization)**: There are several methods of estimating the transformation matrix T from two images:
 - **direct alignment methods**: exhaustive search, FFT registration, Lucas-Kanade [14].
 - **features methods**: search for common features in the two images SIFT [30]

- **hybrid methods:** block-matching [10]
- **non-linear transformations:** adapted to multi-modal, inter-patient or intra-patient registration over time (anatomical changes, tumor, etc): ANTS [37]

Let I and J be two images and \mathcal{F} a family of spatial transformation. The registration problem is written as:

$$\arg \min_{T \in \mathcal{F}} \text{Simil}(T(I), J)$$

Registration process is illustrated on Fig.A.2.

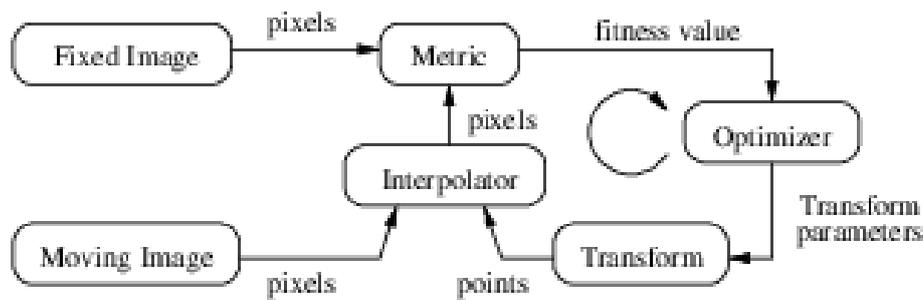


Figure A.2: The basic components of the registration are two input images, a transform, a metric, an interpolator and an optimizer [2]

A.4. Convolutional neural networks

Convolutional neural networks (CNN) are very similar to ordinary neural networks [28]. They are made up of neurons with evolutionary weights and biases. CNN architectures consider that the inputs are images, which allows encoding certain properties in the architecture and reducing the amount of parameters in the network. Thus, unlike an ordinary neural network, the layers of a CNN have neurons arranged in 3 dimensions: width, height (image dimension for example) and depth (number of channels for example) (Fig.A.3). Each layer transforms the 3D input volume into a 3D output volume for activating neurons. Moreover, the neurons of one layer are not necessarily connected to all the neurons of the next layer.

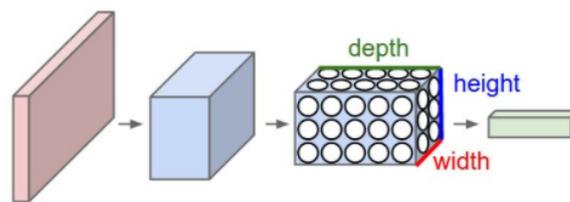


Figure A.3: Illustration of the organization of the layers of a CNN from [28]

Activation functions

This is a stimulation threshold which, once reached at the output of a neuron, results in a response from the latter. Most often, this is a non-linear function. There are several represented on Fig.A.4:

- **sigmoid:** $\frac{1}{1+e^{-x}}$
It has the drawback of saturating and not being centered at 0.
- **tanh:** It is centered at 0 but saturates.
- **Rectified Linear Unit:** $\max(0, x)$
The **ReLU** function has the advantage of increasing the gradient descent rate and better reproduces the behavior of neurons. But, if the learning rate is too high, it can cause the death of the ReLU and the inability for the neuron to activate.

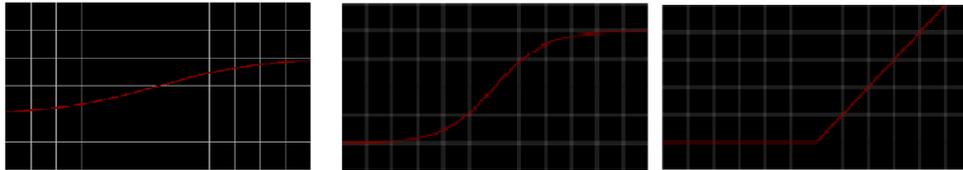


Figure A.4: Representation of the main activation functions from [28]: from left to right sigmoid, tanh and ReLU

Convolutional neural networks are made up of 4 main layers:

1. **Full-connected layers: FC**

Neurons of a full-connected layer have connections with all activations of the previous layer. These are the layers used in ordinary neural networks. They require a large number of parameters (weight and bias). These layers group the pixels of the image into a line vector (Fig.A.5).

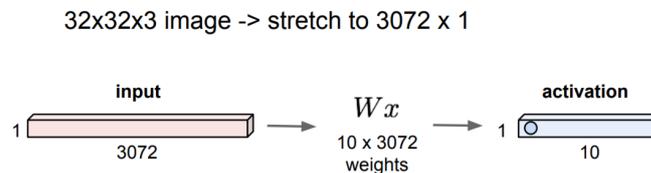


Figure A.5: Principle of the full-connected layer, illustration taken from [28]

2. **Convolutional layers: CONV**

Unlike the FC layer, the convolutional layer preserves the spatial structure of the image. It consists in dragging one or more filter(s) of the same thickness as the input image by locally calculating the scalar product (Fig.A.6) in order to produce an activation volume whose thickness is equal to the number of filters used. It requires a large number of parameters and has the following hyperparameters:

- stride: how many pixels the filter moves on each iteration
- the size of the filter
- zero-padding
- number of filters

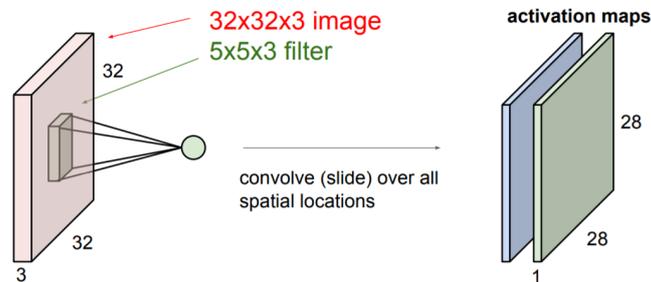


Figure A.6: Principle of the convolutional layer, illustration from [28]

3. Pooling layers: POOL

The POOL layer makes the representation smaller and therefore easier to handle (Fig.A.7). In general, Max Pooling is used. It consists in keeping the maximum value of the filtered pixels during downsampling. The POOL layer has no parameters but has the following hyperparameters:

- size of the filter
- stride: how many pixels the filter moves on each iteration

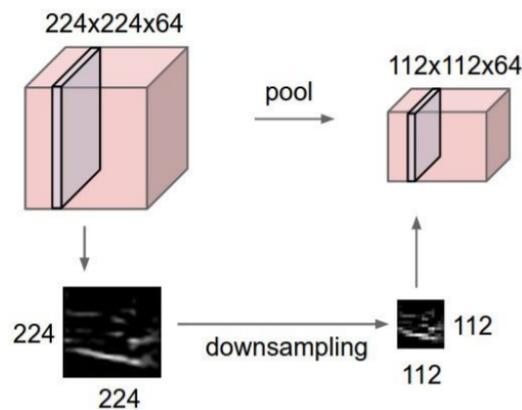


Figure A.7: Principle of the pooling layer, illustration taken from [28]

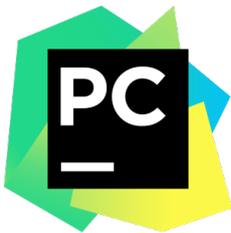
4. Batch Normalization layers

This is an additional layer placed after the FC or CONV layers and before the activation function in order to normalize the data so as to have a "zero-mean". This layer improves the gradient flow and the generalization of the model.

A.5. Technologies and tools used



Free and open source distribution of the Python and R programming languages for the development of applications dedicated to data science and machine learning. Package management is made easy with the conda package management system. [Anaconda](#)



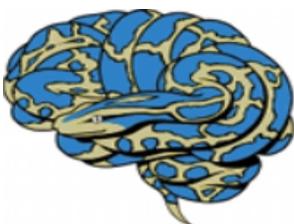
Integrated development environment used for programming in Python. [PyCharm](#)



Analysis tool library for FMRI, MRI and DTI brain imaging data. [FSL](#)



Software application for navigation and segmentation of structures in 3D medical images. [ITKSnap](#)



Python package for read/write access to some common medical and neuroimaging file formats such as ANALYZE, NIFTI or MINC. [Nibabel](#)



Technology that uses a graphics processor to perform calculations general instead of the processor. [Nvidia Cuda](#)



Torch-based open-source machine learning Python software library developed by Facebook. Performs in an optimized way tensor calculations necessary in particular for deep learning on processor (CPU) or if possible on graphics processor (GPU) supporting CUDA. [PyTorch](#)



Online, collaborative and real-time LaTeX editor. [Overleaf](#)



Git repository management tool. [GitHub](#)

Bibliography

- [1] Beachild. <https://beachild.fr/>. Consulted in August 2020.
- [2] Image registration. <https://www.orfeo-toolbox.org/SoftwareGuide/SoftwareGuidech9.html>. Consulted in October 2020.
- [3] Latim. <http://latim.univ-brest.fr/>. Consulted in July 2020.
- [4] Nvidia titan x. <https://www.materiel.net/produit/201608010129.html>. Consulted in August 2020.
- [5] Zoom sur l'articulation de la cheville. <https://www.sport-orthese.com/blog/zoom-sur-l-articulation-de-la-cheville-n31>, 2015.
- [6] Ankle joint : Anatomy, movement and muscle involvement. <https://www.howtorelief.com/ankle-joint-anatomy-overview/>, 2017.
- [7] Data preprocessing : Feature scaling avec python. <https://mrmint.fr/data-preprocessing-feature-scaling-python>, consulted in August 2020.
- [8] Flirt/user guide. <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLIRT/UserGuide>, consulted in July 2020.
- [9] Coordinate systems and affines. https://nipy.org/nibabel/coordinate_systems.html, consulted in June 2020.
- [10] A. Barjatya. Block matching algorithms for motion estimation. *Spring*, 2004.
- [11] P. Blanc-Durand. Réseaux de neurones convolutifs en médecine nucléaire : applications à la segmentation automatique des tumeurs gliales et à la correction d'atténuation en tep/irm. *Médecine humaine et pathologie*, 2018.
- [12] I. Bloch. Recalage et fusion d'images médicales. <https://perso.telecom-paristech.fr/bloch/VOIR/recalageImMed.pdf>. Consulted in July 2020.
- [13] S. Brochard. Evaluation dynamique in vivo de la mécanique articulaire et musculaire de la cheville chez des enfants pc présentant une déformation en équin. <https://www.fondationparalysiecerebrale.org/recherches-finances/evaluation-dynamique-vivo-de-la-mecanique-articulaire-et-musculaire-de-la>, 2015.

- [14] C.-H. Chang, C.-N. Chou, and E. Y. Chang. Clkn: Cascaded lucas-kanade networks for image alignment. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2213-2221, 2017.
- [15] E. Clarke, J. Martin, A. dâEntremont, M. Pandey, D. Wilson, and R. Herbert. A non-invasive, 3d, dynamic mri method for measuring muscle moment arms in vivo: demonstration in the human ankle joint and achilles tendon. *Medical Engineering & Physics*, vol.37, no.1, pp.93-99, 2015.
- [16] P. Coupeau. Implémentation d'un pipeline de reconstruction d'irms foetales, 2019.
- [17] A. de Lander. La fondation paralysie cérébrale, seul acteur français de la recherche sur la paralysie cérébrale. *Fondation Paralysie Cérébrale*, Décembre 2017.
- [18] M. A. Dedmari, S. Conjeti, S. Estrada, P. Ehses, T. Stöcker, and M. Reuter. Complex fully convolutional neural networks for mr image reconstruction. *Computer Science*, 2018.
- [19] M. Drozdal, G. Chartrand, E. Vorontsov, M. Shakeri, L. D. Jorio, A. Tang, A. Romero, Y. Bengio, C. Pal, and S. Kadoury. Learning normalized inputs for iterative estimation in medical image segmentation. *Medical Image Analysis*, vol.44, pp.1-13, 2018.
- [20] J. Du, X. Xie, C. Wang, G. Shi, X. Xu, and Y. Wang. Fully convolutional measurement network for compressive sensing image reconstruction. *Neurocomputing*, vol.328, pp.105-112, 2019.
- [21] Enorev'. Paralysie cérébrale, les symptômes. <https://enorev.fr/enorev/paralysie-cerebrale/>. Consulted in July 2020.
- [22] B. Henry and P.-L. Docquier. Neuro-orthopédie : la paralysie cérébrale, 2013. Article issu d'un manuel réalisé par le Service d'orthopédie des Cliniques universitaires Saint-Luc (Bruxelles-Woluwe).
- [23] I. A. Hosny and H. S. Elghawabi. Ultrafast mri of the fetus: an increasingly important tool in prenatal diagnosis of congenital anomalies. *Magnetic resonance imaging*, vol.28, no.10, pp.1431-1439, 2010.
- [24] A. Houdard. Quelques avancées dans le débruitage d'images par patches. https://houdard.wp.imt.fr/files/2019/01/presentation_Antoine_Houdard.pdf, Janvier 2019.
- [25] L. Houx, M. Lempereur, O. Rémy-Néris, and S. Brochard. Threshold of equinus which alters biomechanical gait parameters in children. *Gait & Posture*, 38, Issue 4, Pages 582-589, Septembre 2013.
- [26] S. Kralik, M. Yasrebi, N. Supakul, C. Lin, L. Netter, R. Hicks, R. Hibbard, L. Ackerman, M. Harris, and C. Ho. Diagnostic performance of ultrafast brain mri for evaluation of abusive head trauma. *American Journal of Neuroradiology*, vol.38, no.4, pp.807-813, 2017.
- [27] S. C. I. Lab. Imm-530, reconstruction et analyses d'images médicales: Recalage. http://scil.dinf.usherbrooke.ca/static/website/courses/imm530/Recalage_compressed.pdf. Consulted in July 2020.

- [28] F.-F. Li, J. Johnson, and S. Yeung. Convolutional neural networks for visual recognition. <https://cs231n.github.io/>. Consulted in June 2020.
- [29] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. *Computer Science*, 2015.
- [30] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, vol.60, no.2, pp.91-110, 2004.
- [31] T. Lu, X. Zhang, Y. Huang, D. Guo, F. Huang, Q. Xu, Y. Hu, L. Ou-Yang, J. Lin, Z. Yang, and X. Qu. pfista-sense-resnet for parallel mri reconstruction. *Journal of Magnetic Resonance*, vol.318, 2020.
- [32] K. Makki. *Développement de l'IRM dynamique pour l'étude de l'appareil Musculo-squelettique en mouvement*. PhD thesis, IMT Atlantique (Brest). Ecole doctorale MathSTIC, 2019.
- [33] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI. Lecture Notes in Computer Science*, vol.9351, pp.234-241, 2015.
- [34] J. P. Sees and F. Miller. Overview of foot deformity management in children with cerebral palsy. *Journal of Children s Orthopaedics*, 2013.
- [35] N. Shibuya. Up-sampling with transposed convolution. <https://medium.com/activating-robotic-minds/up-sampling-with-transposed-convolution-9ae4f2df52d0>, 2017.
- [36] A. Silburt, M. Ali-Dib, C. Zhu, A. Jackson, D. Valencia, Y. Kissin, D. Tamayo, and K. Menou. Lunar crater identification via deep learning. *Icarus*, 2018.
- [37] V. Siless. Multi-modal registration of t1 brain image and geometric descriptors of white matter tracts. *Medical Imaging*, 2014.



Analyse d'images IRM de la cheville par apprentissage profond

La paralysie cérébrale est la première cause de handicap moteur de l'enfant en France. 90% des déformations dues à la PC apparaissent dans la région de la cheville. La compréhension de la biomécanique articulaire est un enjeu médical afin de mettre en oeuvre des thérapies adaptées à la pathologie. Les laboratoires de recherche tentent d'exploiter l'IRM dynamique afin d'étudier l'appareil musculo-squelettique en mouvement. L'expérience vécue durant trois mois au Laboratoire de traitement de l'information médicale à Brest a participé à l'amélioration de l'analyse des IRMs de la cheville par apprentissage profond. Les travaux sont allés du traitement des données IRM à l'entraînement de divers réseaux de neurones destinés à la segmentation des os de la cheville ou encore à la reconstruction de séquences dynamiques.

Mots clés: LaTIM, IRM dynamique, apprentissage profond, segmentation, paralysie cérébrale

Analysis of MRI images of the ankle by deep learning

Cerebral palsy is the first cause of children's motor disability in France. 90% of deformities due to CP appear in the ankle region. Understanding joint biomechanics is a medical issue in order to implement therapies adapted to the pathology. Research laboratories are trying to explore dynamic MRI to study the musculoskeletal system in motion. The experience lived during three months at the Laboratory of medical information processing in Brest helped to improve the analysis of MRIs of the ankle with deep learning. Work went from processing MRI data to training multiple neural networks for ankle bones' segmentation or even the reconstruction of dynamic sequences.

Keywords: LaTIM, dynamic MRI, deep learning, segmentation, cerebral palsy

Análisis de imágenes por resonancia magnética del tobillo mediante aprendizaje profundo

La parálisis cerebral es la principal causa de discapacidad motora de los niños en Francia. El 90% de las deformidades por PC aparecen en la región del tobillo. Entender la biomecánica articular es un tema médico para poder implementar terapias adaptadas a la patología. Los laboratorios de investigación tratan de utilizar la IRM dinámica para estudiar el sistema musculoesquelético en movimiento. La experiencia vivida durante tres meses en el Laboratorio de procesamiento de información médica de Brest ayudó a mejorar el análisis de las IRM del tobillo mediante aprendizaje profundo. El trabajo se ha extendido del procesamiento de datos de IRM al entrenamiento de varias redes neuronales para la segmentación de los huesos del tobillo o incluso la reconstrucción de secuencias dinámicas.

Palabras clave: LaTIM, IRM dinámica, aprendizaje profundo, segmentación, parálisis cerebral