Computer Methods and Programs in Biomedicine 183 (2020) 105062

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Automated postoperative muscle assessment of hip arthroplasty patients using multimodal imaging joint segmentation

Marta B.M. Ranzini[a,c,d,*], Johann Henckel[b], Michael Ebner[a,c,d], M. Jorge Cardoso[c,d], Amanda Isaac[c,e], Tom Vercauteren[c,d], Sébastien Ourselin[c,d], Alister Hart[b], Marc Modat[c,d]

[a] *Centre for Medical Imaging Computing, University College London, London, UK*
[b] *Royal National Orthopaedic Hospital NHS Foundation Trust, London, UK*
[c] *School of Biomedical Engineering & Imaging Sciences, King's College London, King's Health Partners, St Thomas' Hospital, London SE1 7EH, United Kingdom*
[d] *Medical Physics and Biomedical Engineering Department, University College London, London WC1E 6BT, United Kingdom*
[e] *Radiology Department, Guys & St Thomas Hospitals NHS Foundation Trust, London SE1 7EH, UK*

ARTICLE INFO

Article history:
Received 5 July 2018
Revised 15 August 2019
Accepted 2 September 2019

Keywords:
Musculoskeletal imaging
Multimodal segmentation
Multimodal registration
Hip arthroplasty
Fat infiltration
Muscle atrophy

ABSTRACT

Background and objective: In patients treated with hip arthroplasty, the muscular condition and presence of inflammatory reactions are assessed using magnetic resonance imaging (MRI). As MRI lacks contrast for bony structures, computed tomography (CT) is preferred for clinical evaluation of bone tissue and orthopaedic surgical planning. Combining the complementary information of MRI and CT could improve current clinical practice for diagnosis, monitoring and treatment planning. In particular, the different contrast of these modalities could help better quantify the presence of fatty infiltration to characterise muscular condition and assess implant failure. In this work, we combine CT and MRI for joint bone and muscle segmentation and we propose a novel Intramuscular Fat Fraction estimation method for the quantification of muscle atrophy.

Methods: Our multimodal framework is able to segment healthy and pathological musculoskeletal structures as well as implants, and develops into three steps. First, input images are pre-processed to improve the low quality of clinically acquired images and to reduce the noise associated with metal artefact. Subsequently, CT and MRI are non-linearly aligned using a novel approach which imposes rigidity constraints on bony structures to ensure realistic deformation. Finally, taking advantage of a multimodal atlas we created for this task, a multi-atlas based segmentation delineates pelvic bones, abductor muscles and implants on both modalities jointly. From the obtained segmentation, a multimodal estimation of the Intramuscular Fat Fraction can be automatically derived.

Results: Evaluation of the segmentation in a leave-one-out cross-validation study on 22 hip sides resulted in an average Dice score of 0.90 for skeletal and 0.84 for muscular structures. Our multimodal Intramuscular Fat Fraction was benchmarked on 27 different cases against a standard radiological score, showing stronger association than a single modality approach in a one-way ANOVA F-test analysis.

Conclusions: The proposed framework represents a promising tool to support image analysis in hip arthroplasty, being robust to the presence of implants and associated image artefacts. By allowing for the automated extraction of a muscle atrophy imaging biomarker, it could quantitatively inform the decision-making process about patient's management.

1. Introduction

Hip arthroplasty consists of replacing a pathological hip joint with a prosthesis. It is the fourth most common surgical procedure in the US [1] and the second one in the UK [2]. Despite the reported success of these surgical procedures in improving life quality, hip implants are associated with a non-negligible failure rate −6.8% at 13 years from primary surgery in the UK [3] — and a consequent need of being replaced in revision surgery. The failure rate increases up to 14–27% when Metal-on-Metal (MoM) implants have been utilised in primary surgery, as this type of prosthesis is linked to adverse inflammatory reactions and muscular wastage due to fat infiltration [3,4].

---

[*] Corresponding author at: Medical Physics and Biomedical Engineering Department, University College London, London WC1E 6BT, UK
E-mail address: marta.ranzini.15@ucl.ac.uk (M.B.M. Ranzini).

https://doi.org/10.1016/j.cmpb.2019.105062

Diagnosis, monitoring and treatment of patients with hip implants routinely take advantage of medical imaging. However, the clinical use of advanced automated imaging analysis pipelines is currently hampered by the large inter-subject variability of the musculoskeletal structures, the lack of standardised acquisition protocols and the required large imaging Field-of-View (FOV) [5,6]. As a result, current decisions for patient post-surgical management in hip arthroplasty rely on separate, subjective and qualitative assessments of imaging data. Magnetic Resonance Imaging (MRI) is often used as preferred modality to assess muscle conditions on suspicion of inflammation or atrophy, such as in the case of patients with MoM implants [7].

Current approaches for quantification of muscle atrophy on MRI have several drawbacks. They discard the volumetric information, as single-slice scoring systems (e.g. Pfirmann score [8] or Bal and Lowe score [9]) are the current best practice. They have already been proven to be un-predictive of 3-dimensional measures in the rotator cuff muscle [10]. Most methods for volumetric estimation of fat infiltration require tailored MRI acquisition protocols such as the Dixon, based on chemical-shift MRI sequences [11]. However, such techniques are not routinely available in clinical practice, being very sensitive to magnetic field inhomogeneities in presence of metal implants, and typically rely on manual delineation of regions of interest (ROI).

Computed Tomography (CT) might represent a viable alternative for the quantification of fat infiltration. Its scanner-independent range of intensities provides a clear separation between fat and lean muscle [12,13] at -30 Hounsfield Units (HU). Additionally, CT is the most appropriate and clinically used modality for bone pathology diagnosis, surgical planning and post-surgical delivery assessments, due to its improved visualisation of bones and implants [14]. However, the lack of contrast between muscles and the presence of metal-induced noise and artefact complicate the segmentation of individual muscular volumes.

To diagnose and treat implant failure, as well as to monitor well functioning implants, a unique framework combining the complementary information of CT and MRI could therefore have clinical value for imaging biomarkers extraction and surgical planning. It would help better delineate both muscular and skeletal structures concurrently and their relative spatial localisation, towards a more accurate definition and visualisation of patient-specific anatomy. This would benefit the customisation of surgical planning as it would help minimise the damage to healthy musculoskeletal tissue and potentially result in longer-lasting implants after revision, especially relevant for younger patients [3].

The combination of multimodal information requires first to address several technical challenges arising from the low quality of clinical data and the presence of strong metal artefact induced noise. The difference of the patient's pose within the CT and MRI scanners also constitutes a major challenge when aligning both modalities from the same subject. Standard intra-subject registration for pelvic anatomy usually relies on global affine transformation, such as the robust and inverse-consistent block-matching algorithm presented by Rivest-Hénault et al. [15]. However, to cope with different scanner couches and patient's position, the applied transformation should allow for non-linear deformation of soft tissues while retaining the rigidity of the bones and implants. To tackle this issue, Staring et al. [16] proposed a non-linear registration framework where rigidity is promoted by using an additional penalty term as a soft constraint (during optimisation). This approach was extended for spine CT-MRI registration by Reaungamornrat et al. [17], who added a constraint to impose injectivity to the transformation model. In the context of CT synthesis for radiotherapy treatment planning, Dowling et al. [18] generated a well-aligned CT-MRI training dataset using the structure-guided nonrigid registration proposed by Rivest-Hénault et al. [19]. However, their algorithm relies on availability of accurate contours of the same anatomical structure in both modalities, which are matched in the registration through the addition of a soft constraint to the cost function. While being effective, all these methods need to find a trade-off between the terms in the cost function, which might still cause deviations from a strictly rigid transformation and therefore requires a careful selection of the rigidity penalty term weight. To avoid the need for prior rigid masks, Commowick et al. [20] proposed a two-step strategy that computes spatially sparse local rigid transformations through a modified block-matching approach, and then interpolates a dense velocity field between the rigid ones. An alternative solution was proposed by Haber et al. [21], who presented a mathematical formulation of the registration problem where a non-linear transformation model is enforced to be strictly rigid in specified areas as a hard constraint. The authors report a proof of concept on 2D examples only, and do not explicitly consider the integration of such formulation into a broader diffeomorphic registration framework.

As in pelvic musculoskeletal imaging multimodal non-linear registration still lacks robustness and generalisability, in the literature the segmentation of the hip joint and muscles has mostly been performed on single modality scenarios. CT is usually preferred to obtain bone segmentation and in particular to identify the acetabular space. The most successful methods rely on variations of statistical shape models such as Yokota et al. [22] or on hybrid approaches, combining atlas-based segmentation propagation with either statistical shape models [23] or graph-cut segmentation [24]. However, global intensity thresholding coupled with manual editing still remains the most commonly applied method for segmentation in CT [25], as automated methods are yet to reach satisfactory accuracy.

The segmentation of musculoskeletal structures in MR images is currently hampered by the lack of established and easily generalisable segmentation techniques [5]. Common approaches are atlas-based [26,27], or make use of deformable simplex meshes [28]. Semi-automated approaches have been introduced for either muscle or bone segmentation, requiring user manual initialisaton of an intra-subject 2D to 3D segmentation propagation [29] or user interactive corrections [30]. When considering only bone segmentation, robust statistical and active shape models have been applied for the hip joint [31,32], taking advantage of strong shape priors to cope with the MRI bone contrast variations. Similarly, Xia et al. [33] tested the use of CT-derived bone models to obtain the segmentation of T2 MR images either through multi-atlas segmentation propagation or active shape models, showing comparable robustness and accuracy. Most recently, deep learning approaches have been explored either for bone and cartilage segmentation in MR images [34,35], or for thigh muscle and fat segmentation [36]. Especially for hip arthroplasty, Klemt et al. [37] developed a framework for the automated segmentation of abductor muscles in MR images for both healthy and implanted hip sides, showing robustness against population variability. To the best of our knowledge, very little work has been proposed to combine the segmentation of all relevant musculoskeletal structures of the hip joint. The most recent example is the work by Yokota et al. [38], where a hierarchical multi-atlas approach is employed to obtain automatic segmentation of pelvis, femur and 19 muscles in CT images. However, with the exception of Klemt et al. [37], all the discussed methods are not directly applicable to hip arthroplasty patients, due to the presence of strong metal-induced noise and artefact in both imaging modalities. In this context, we hypothesise that the segmentation of musculoskeletal structures could be improved by combining CT and MRI, as one modality would compensate for the segmentation of those structures the other modality lacks contrast for.

Following this idea, we present the first automated pipeline for the joint segmentation of CT and MRI for patients treated with

MoM hip arthroplasty, inclusive of both musculoskeletal structures and implants. To obtain robust intra-subject multi-modality alignment, we introduce a novel non-linear, diffeomorphic registration algorithm that preserves the local rigidity of bones through hard constraints. Automated segmentation is obtained through a 4-dimensional multi-atlas segmentation propagation approach, which combines information from both CT and MRI and employs a new atlas of implanted hips we built for this purpose. In addition, we introduce a novel automated quantification of fatty infiltration, the Intramuscular Fat Fraction (IFF), which measures the percentage of fat in muscular volumes. Our derivation takes advantage of both imaging modalities and shows potential as a quantitative imaging biomarker of muscle atrophy.

The main contributions of our work are twofold. From a methodological perspective, the introduction of the proposed non-linear registration algorithm with rigidity constraints allows to manage patient's pose differences between scanner, producing biologically plausible deformation. The combination of this novel algorithm with state-of-the-art super-resolution reconstruction and metal artefact reduction techniques makes our pipeline robust to the low-quality of clinical data, characterised by highly anisotropic resolution and strongly affected by metal artefact induced noise. From a clinical perspective, our multimodal imaging biomarker of fatty infiltration provides a quantitative scoring system for abductor muscles atrophy. This arises from the ability to obtain accurate MRI-driven muscle boundary delineation also in the CT space, even in the presence of strong artefact induced noise. Moreover, the fused segmentation allows for patient-specific musculoskeletal anatomy visualisation and volumetric rendering.

This paper is an extension of the work presented by Ranzini et al. [39] at the MSKI "Computational Methods and Clinical Applications in Musculoskeletal Imaging" workshop, held in conjunction with MICCAI 2017. Together with a more detailed discussion of the proposed method (Section 2), this work introduces improved pre-processing steps involving the use of a metal artefact reduction technique for the CT and a multi-step bias field correction for the MRI. The methodological validation has also been extended to include comparisons for healthy gluteus muscle analysis in CT, in MRI and in our combined framework (Section 3.1). Finally, the clinical application discussed in Section 3.2 presents the extraction of the Intramuscular Fat Fraction as a muscle atrophy multimodal imaging biomarker from our framework.

2. Methods

The pipeline proposed in this work provides an automated tool for the multimodal segmentation of the pelvic bones, the implants and the abductor muscles, which are at greatest risk of developing atrophy or hypertrophy after hip arthroplasty. The segmentation is performed jointly for both CT and MRI and it is obtained through a multimodal multi-atlas based approach, which allows us to manage the large population variability of these anatomical structures.

This section is structured as follows. Firstly, we introduce our dataset and we describe the process we applied to generate our template images for the multi-atlas segmentation approach. Secondly, we provide a detailed description of our automated pipeline, and finally we introduce the Intramuscular Fat Fraction as a potential imaging biomarker for muscular atrophy.

2.1. Dataset

Our dataset comprises CT and MR images from 38 subjects that have been treated with MoM hip arthroplasty. Data was selected retrospectively, based on availability of images from both modalities and atrophy reporting, among patients that were referred to Charing Cross Hospital (London, UK) for unexplained hip pain between 2006 and 2012. Summary demographic statistics are reported in Table 1.

All the MR images were acquired on a Siemens MAGNETON Avanto 1.5T scanner, with the Metal Artefact Reduction Sequence (MARS) presented in Sabah et al. [40]. This is a standard protocol widely used in clinical practice and includes the collection of two T1-weighted Turbo Spin Echo highly anisotropic images: an axial acquisition (TE = 8 ms, TR = 509 ms, typical imaging resolution = $0.78 \times 0.78 \times 7.02$ mm^3) and a coronal acquisition (TE = 7.1 ms, TR = 627 ms, typical imaging resolution = $1.25 \times 1.25 \times 6.00$ mm^3). The CT acquisitions have been performed on a Siemens SOMATOM Sensation 16 scanner, with the exception of 8 cases acquired on a Siemens SOMATOM Definition AS machine. Tube voltage varied in the range [80,120] kVp. The reconstructed CT images have been corrected for metal artefact using the wxDICOM software,[1] based on the Refined Metal Artefact Reduction method proposed by Treece [41].

The available dataset was split as follows. Eleven subjects were randomly selected for manual segmentation in order to build the template dataset for the atlas-based segmentation (see Section 2.2). The remaining 27 were automatically segmented with the proposed pipeline and used for the extraction of imaging biomarkers.

2.2. Template dataset generation for automated segmentation

The template dataset consists of 11 subjects (10 unilateral, 1 bilateral), whose MR images and CT were acquired on the same day. For each subject, the CT and MR images were pre-processed for image quality enhancement as discussed in Section 2.3. Given its higher contrast for bony structures, the CT images were used to manually delineate[2] the pelvic bones, the femora and the implant(s). Manual segmentation of Gluteus Maximus (GMAX), Gluteus Medius (GMED), Gluteus Minimus (GMIN) and Tensor Fasciae Latae (TFL) was performed on the MR image after super-resolution reconstruction was applied (Section 2.3). The two modalities were then registered into the same reference frame (Section 2.3), and all the manual segmentation masks were merged into a single multi-label segmentation image. We finally performed further manual refinement, in order to guarantee non-overlapping regions between the segmented structures.

We therefore organised the template dataset into two subsets, the implanted hip sides (12 sides) and the non-implanted ones (10). For the sake of simplicity, we refer to these latter as the healthy sides, although we acknowledge that the absence of the implant does not imply absence of pathology nor implant-induced artefact. In order to separate and group all the implanted and the healthy hip sides, each template set of CT, MR and label images was split along the CT-derived sagittal axis of symmetry, and reoriented according to the presence or absence of an implant (Fig. 1). For this task, we developed a symmetry-detection algorithm that, in analogy with the mechanics of rigid body, computes the inertia tensor of the CT image using its intensities as mass values. Specifically, to balance the high intensity values of the implant typically present only on one side, the inertia tensor is obtained from the input CT after averaging it with its left-right flipped image. The mid-sagittal plane of symmetry can then be extracted by the inertia tensor. Simple thresholding at 3000 HU is then exploited to assess the presence of implants in each hip side of the original CT. All implanted sides are then oriented to be on the left and the healthy sides on the right.

Within each subset, all the templates were rigidly aligned into their mid-space through the robust group-wise registration

---

[1] https://mi.eng.cam.ac.uk/Main/GMT_wxDicom.

[2] Manual segmentations were performed using ITKSnap tools. www.itksnap.org.

Table 1
Summary of the dataset demographic statistics. Unilateral cases refer to patients with only one implanted side, while bilateral cases to subjects with implants on both hip joints. The mean age at the time of CT acquisition is also shown, together with the number of cases grouped according to the time difference between the CT and the MR acquisition.

	Females	Males	Total
Number of subjects	25	13	38
Unilateral cases	17	12	29
Bilateral cases	8	1	9
Mean [Range] age	54.92 [35, 74]	57.08 [37, 69]	55.66 [35, 74]
MRI within 1 month from CT	21	9	30
MRI within 6 months from CT	2	0	2
MRI within 24 months from CT	1	4	5
MRI within 37 months from CT	1	0	1

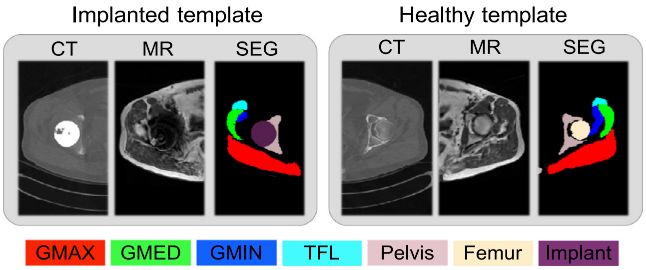


Fig. 1. Examples from the template sets, composed of CT, the respective non-linearly registered MRI (see Section 2.3 for processing details), and their fused segmentation of hip joint bones and abductor muscles. An implanted (left) and a healthy (right) template sides are shown. A single slice is displayed for illustration purposes. However, we emphasise that we created template images in the 3D space.

framework proposed by Klemt et al. [37]. Only the header orientation matrices were updated to avoid intensity resampling.

2.3. Pipeline for automated segmentation

The proposed segmentation framework processes the inputs in three sequential blocks of steps: image quality enhancement, intra-subject MR-CT image registration with rigidity constraints on bone, and multi-atlas based multimodal segmentation.

A schematic representation of this pipeline is shown in Fig. 2. Taking advantage of the NiPype framework for the implementation [42], our pipeline makes use and extends image processing utilities from `NiftyReg`,[3] `NiftySeg`,[4] FSL[5] and `SimpleITK`.[6]

Image quality enhancement. The purpose of the first processing block is to enhance the quality of the input images, to deal with routinely acquired clinical data. The steps we are presenting here were specifically implemented for the available data types (i.e. one CT image and two anisotropic MR acquisitions). However, our modular formulation can be tailored to manage only partial data - e.g only one MR image available - or extended to include different acquisitions - e.g. other sequences MR images.

The two modalities are initially processed separately. Both the axial and the coronal MRI acquisitions in our datasets are heavily affected by bias field inhomogeneities. Hence, to each image we first apply a global N4 bias field correction [43]. The adipose tissue regions are then extracted automatically using the Expectation-Maximisation segmentation algorithm [44] and used as masks to estimate the residual bias field, which is further corrected on the whole FOV with the N4 algorithm [43]. After this correction, the intensity histograms of the two MRI acquisitions are matched by means of SimpleITK HistogramMatchingImageFilter. The two views are then combined into a single volume at $1 \times 1 \times 1$ mm^3 resolution using a super-resolution reconstruction (SRR) algorithm [45]. No slice-to-volume registration was required as motion was negligible. The use of a SRR algorithm has several advantages. Firstly, it allows us to compensate for the highly anisotropic resolution of the acquired data, which would adversely affect the subsequent registration to the CT image; secondly, it compensates for the lack of 3D MRI acquisition, as 2D multislice sequences are usually preferred in clinical routine due to scanning time constraints. An example of super-resolution reconstructed MR image is shown in Fig. 3.

To improve the quality of the CT images, we applied a metal artefact correction [41]. A cubic interpolation scheme is utilised to resample the corrected CT to the same isotropic resolution of the SRR MRI. At this stage, we also extract binary masks for each hip bone that are needed for the subsequent non-linear registration, as we want to enforce a rigid transformation on the bony structures. The binary masks are generated by multi-atlas based segmentation using our template CT images and the respective bone segmentations. For each hip side, templates CT are registered to the target image through affine [46] and following free-form non-linear registration [47] with default parameters in `NiftyReg`; the template label images are then propagated with the respective estimated transformation and final consensus is obtained using the STEPS label fusion algorithm [48].

Intra-Subject MR-CT Registration with bone rigid constraint. The second block of our pipeline aims at aligning the CT and the SRR MRI of the same subject. This requires a multimodal non-linear registration to compensate for different patient's pose within the two scanners while respecting the rigidity of bony structures.

As a first step, an initial global alignment is achieved by affinely registering the SRR MRI to the CT using the symmetric block-matching algorithm proposed by Modat et al. [46]. A more refined local alignment is then achieved through the non-linear registration step. In order to prevent implausible deformation of bones, we used an intensity-based non-linear registration framework that applies a strictly rigid transformation to all the voxels within specified masks, extending the mathematical formulation of hard rigid constraints presented by Haber et al. [21]. Given a reference image $R : X \to R$, defined in the reference space $X \in \mathbb{R}^3$, and a floating image $F : Y \to F$, defined in the floating space $Y \in \mathbb{R}^3$, we want to optimise a transformation model $\phi : X \to Y$ which is enforced to be rigid within specific areas defined by a set of masks $M_j \subset X$ and non-linear elsewhere. Therefore, given the cost function

$$\mathcal{C}(R, F; \phi) = (1 - \lambda)\ \mathcal{S}(F(\phi(X)),\ R(X))\ - \lambda Reg(\phi), \qquad (1)$$

where $\mathcal{S}$ is a measure of similarity between the reference and the warped floating image and *Reg* a regularisation term weighted by

---

[3] https://github.com/KCL-BMEIS/NiftyReg.

[4] https://github.com/KCL-BMEIS/NiftySeg.

[5] https://fsl.fmrib.ox.ac.uk/fsl/fslwiki.

[6] http://www.simpleitk.org/.

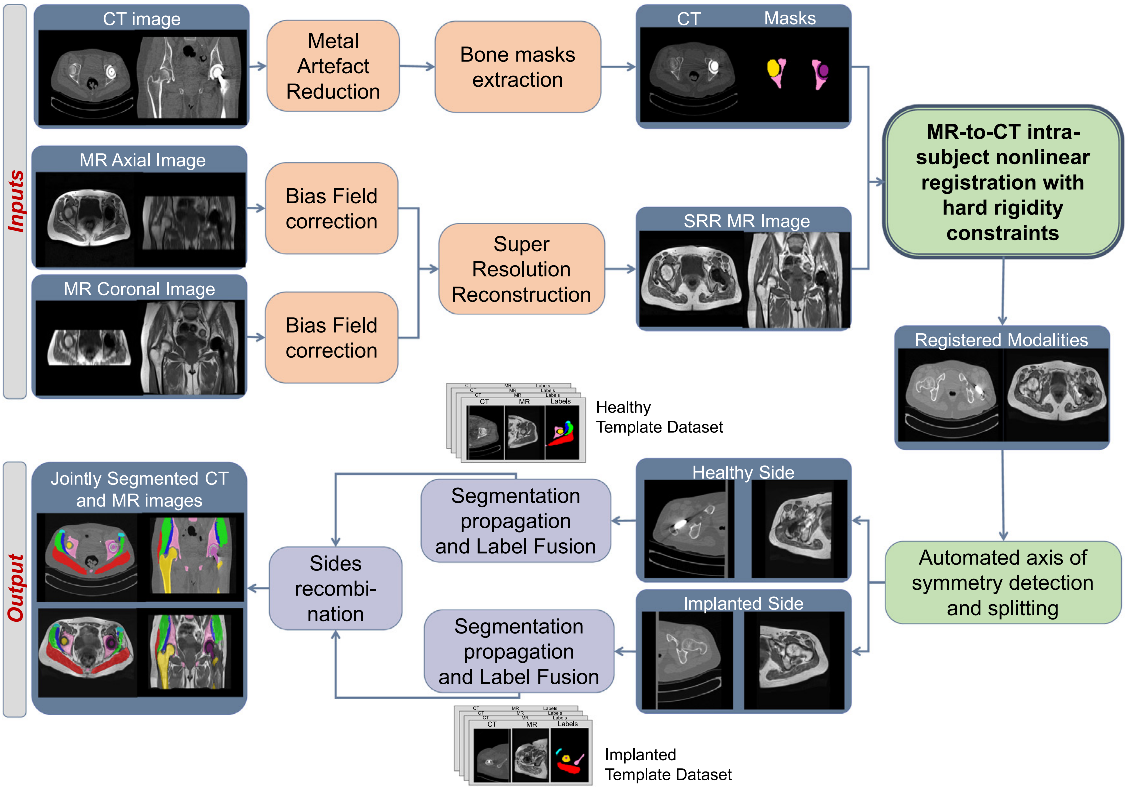


Fig. 2. Proposed pipeline for joint automated segmentation of CT and MR pelvic images. The three blocks composing the pipeline are highlighted: image quality enhancement (orange), where the two modalities are firstly processed independently; our novel intra-subject multimodal registration (green), where the proposed rigidly constrained non-linear registration provides alignment of the CT and MRI while guaranteeing a rigid behaviour in bones; multi-atlas based automated segmentation (purple), where the joint CT and MRI segmentation of the two hip sides are separately obtained and then recombined in the full FOV.

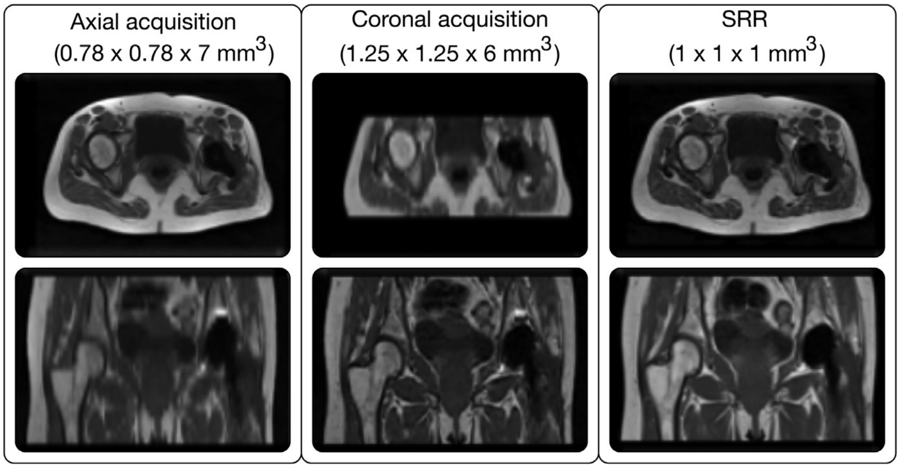


Fig. 3. Example of the available input axial and coronal MRI acquisitions and the respective super-resolution reconstructed (SRR) MRI, showing how low-resolution information is combined to maintain high resolution in both the imaging planes.

a coefficient λ, we want to solve:

$$\max_{\phi} \mathcal{C}(R, F; \phi) \quad \text{subject to } \phi(\mathbf{x}) - \mathcal{R}_j(\mathbf{x}) = \mathbf{0} \quad \forall \mathbf{x} \in M_j \subset X \tag{2}$$

with $\mathcal{R}_j : \mathbb{R}^3 \to \mathbb{R}^3$ being a rigid transformation applied to all voxels $\mathbf{x}$ within the j-th mask. Using this formulation, the rigid constraint in the bony structures is embedded directly into the transformation model as a “hard constraint”. This differs from currently proposed approaches based on “soft constraints” such as regularisation terms in the cost function, which simply penalise large deviations from a rigid behaviour within specified regions - e.g. Staring et al. [16], Reaungamornrat et al. [17].

We used a stationary velocity field parametrisation for the transformation model, computed over a set of control points $\{\mu\}$ through cubic B-Splines. The scaling-and-squaring exponentiation of the stationary velocity field allows us to derive diffeomorphic and symmetric deformation fields, where the direct and the inverse transformations are optimised simultaneously [49]. The optimisation is performed through a conjugate gradient scheme as

Algorithm 1 Control point parametrisation update with rigid constraints.

```
Compute the gradient, $G$, of the current cost function value for each control point $\mu$:
    $G(\mu) = \nabla\mathcal{C}(R, F; \mu), \quad \forall\mu \in \{\mu\}$
Perform a line search along the direction of G:
for each step in line search do
    Update the current control point parameters:
        $\mu \leftarrow \mu + G(\mu), \quad \forall\mu \in \{\mu\}$
    if stationary velocity field parametrisation then
        Down-scale the control point parameters
            $\mu \leftarrow \mu/2^n, \quad \forall\mu \in \{\mu\}$
    end if
    for each mask $M_j$ do
        Define the subset $\{\mu\}_j$ of the control points within $M_j$
        LTS estimation of the rigid transformation $\mathcal{R}_j$:
            $\mathcal{R}_j = \mathrm{LTS}(\{\mu\}_j)$
        Update the parameters with the estimated rigid displacement:
            $\mu \leftarrow \mathcal{R}_j(\mu) \quad \forall\mu \in \{\mu\}_j$
    end for
    if stationary velocity field parametrisation then
        Up-scale control point parameters
            $\mu \leftarrow \mu * 2^n, \quad \forall\mu \in \{\mu\}$
    end if
end for
if stationary velocity field parametrisation then
    Scaling-and-squaring exponentiation ($n$ steps)
end if
```

by default in `NiftyReg`. However, we modified the scheme to update the transformation parameters as in Algorithm 1 to introduce the rigidity constraints. Specifically, at each iteration the gradient of the cost function is used to update the new parameters of each control point. Within each mask associated with a rigid region, we extract the parameters of all the control points and use a Least Trimmed Square (LTS) regression to robustly estimate the rigid transformation that best fits the displacement associated with each control points, discarding 50% of points with highest residuals as outliers. The obtained rigid transformation is subsequently applied to update the current estimate of the parameters associated with each control point within the mask. With the velocity field parametrisation, the rigidity constraints are applied to the velocity field after scaling, which represents a small deformation field. The integration to the final deformation is then approximated through the composition of this small field by itself several times (squaring step of the scaling-and-squaring exponentiation). As the scaled deformation is set to rigid, due to the properties of composition the final deformation is maintained rigid within the masks.

Note that our approach provides a formulation for the gradient computation for both rigid and non-rigid areas that can be easily incorporated in a gradient-based optimisation scheme, and can be used in both a velocity field parametrisation and standard free-form deformation one. In order to ensure smooth transitions in the deformation field, the proposed algorithm is iterated on a five-level course-to-fine pyramidal approach, using a final control point spacing of 5 mm. To account for the local support of the cubic B-spline parametrisation, the rigid masks are dilated by 1 at each level of the resolution pyramid.

Multi-atlas-based automated segmentation. The final block of our pipeline uses the template datasets in a multi-atlas segmentation propagation framework to estimate the final segmentation of each hip side. Given the selected dataset (either healthy or implanted), the templates CT-MR images are registered to the target CT-MRI (stacked into a 4D volume) through the following steps:

1. *Rigid registration*: All the templates are rigidly aligned to the target independently, using symmetric block-matching [46].
2. *LTS average of rigid transformations*: The optimised rigid matrices are averaged in the log-Euclidean space, discarding 50% of them as outliers through least trimmed square regression. Since all the templates are co-registered to their mid-space, this guarantees robustness against potentially failed template registrations.
3. *Affine registration*: All the templates are affinely aligned to the target. Instead of initialising the registration with the previously computed rigid matrix for each template, we employ the LTS rigid average for all of them.
4. *LTS average of affine transformations*: Similarly to the rigid case, a LTS average affine transformation is computed.
5. *Non-linear registration*: Using the LTS average affine matrix as initialisation, each template is non-linearly registered to the target using a multi-channel approach that aligns simultaneously both imaging modalities. Both CT and MRI contribute equally to the cost function, hence the optimisation will converge to a final transformation that balances the alignment of the CTs and the alignment of the MR images. In this approach, we make use of a standard free-form deformation registration algorithm, parametrised through cubic B-splines [47]. A bending energy penalty term is also added to the cost function to regularise the optimised deformation field.

The final estimated non-linear transformations are deployed to resample the label image of the respective templates onto the target space, using a nearest-neighbour interpolation scheme. As for the bone-mask creation, a final consensus is obtained from the candidate segmentations by means of the STEPS label fusion algorithm, which we modified to use a multi-channel version of the local normalized correlation coefficient (LNCC) to rank the templates, defined as the sum of LNCC values from each channel.

This procedure is performed for both hip sides separately. The two sides are finally reoriented back and recombined to their original full FOV, providing a final multi-label image that can be overlaid on both CT and MRI, highlighting all the segmented musculoskeletal structures.

2.4. Imaging biomarkers for muscle atrophy

The proposed segmentation pipeline provides an automated delineation of regions of interest (ROI) for the extraction of clinically relevant information. In the context of muscle atrophy, two quantities are of main interest to define muscular condition: the muscle gross volume and the quantification of fat infiltration within the muscle [50]. Whilst the segmentation provides a direct quantification of muscle gross volume, further differentiation is needed to classify intra-muscular fat and lean tissue within each muscular ROI. From the multi-label segmentation output by our pipeline, we use each muscle label independently as a mask to identify the ROI where to estimate the fat infiltration. Each mask is eroded by three voxels to ignore inter-muscular fat [11]. Within each eroded ROI mask, the CT image is segmented by thresholding at -30 HU [12], classifying each voxel as either adipose tissue or lean muscle. From this classification, we compute the percentage of fat tissue in the muscle gross volume as the Intramuscular Fat Fraction (IFF):

$$\mathrm{IFF} = \frac{V_{\mathrm{fat}}}{V_{\mathrm{lean}} + V_{\mathrm{fat}}}, \tag{3}$$

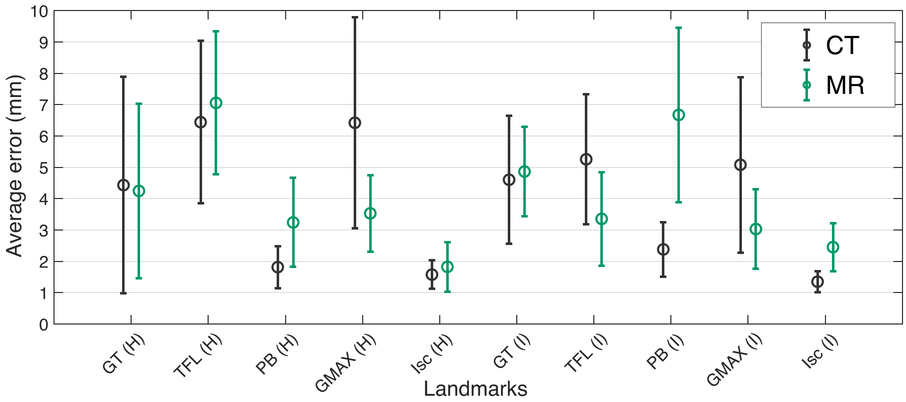


Fig. 4. Manual selection reproducibility error for the 10 landmarks in CT and in MR images (mean and 95% confidence interval are reported). Each landmark is categorized as healthy (H) or implanted (I) according to the hip side it belongs to. Landmarks abbreviations: Greater Trochanter (GT), anterior tip of the Tensor Fasciae Latae (TFL), Pelvic Brim (PB), Gluteus Maximus sacrum origin (GMAX) and Ischium (Isc).

where the volume V of fat/lean muscle is simply the total number of voxels classified as fat/lean muscle multiplied by the voxel volume. In other words, MRI helps define the boundaries of the different muscles, i.e. the different ROIs, while CT helps differentiate fat and lean muscle tissue within each ROI.

3. Results

3.1. Segmentation pipeline validation

In order to assess the performance of the proposed pipeline, we designed two sets of experiments. The first aims at evaluating our intra-subject registration algorithm, selecting the optimal registration parameters and comparing it to the corresponding fully non-linear approach without using rigidity constraints. The second set consists of a leave-one-out cross validation (LOOCV) framework to both optimise the multi-atlas-based segmentation parameters and compare the segmentation results for the CT, the MRI and the combined CT/MRI cases. As an extension to this, a nested LOOCV was also employed to test the generalisability of the proposed approach to unseen data.

Only the 11 template subjects were used in these experiments, as ground truth information was available.

3.1.1. Intra-subject registration with rigid constraints

For this experiment, the Normalised Mutual Information (NMI) was used as a measure of similarity, being the most commonly adopted solution for multimodal registration, and the bending energy was selected as a regularisation term [47] in order to promote smooth deformations. We registered the full-FOV CT and SRR MR images of each template subject by using our rigidity constraint on bones and implants, as well as with fully non-linear deformation at varying regularisation weight λ.

The quantification of the registration accuracy was obtained by computing the Target Registration Error (TRE) on 10 anatomical landmarks - 5 for each hip side - manually selected in both skeletal (3 landmarks) and muscular (2 landmarks) structures. The landmarks were chosen according to three criteria: (1) being easily identifiable in both imaging modalities; (2) being spread across the full FOV; (3) being located on the structures we aimed to segment. In addition, the selected landmarks are not gender related, not age related and very little susceptible to normal variants (a more detailed description can be found in Appendix A). With this choice, we tried to better characterise the registration error on the structures of interest, although we acknowledge that our landmarks are not directly linked to clinical relevance. As the identification of clear landmarks on extended and generally homogeneous structures such as muscles is not trivial, to account for choice bias the manual selection was performed twice for each image at different times by a non-clinical expert, and the selection protocol was reviewed by an expert musculoskeletal radiologist. The average reproducibility error across all the subjects is reported in Fig. 4.

Given a transformation ϕ, the TRE for a specific landmark i was calculated in a symmetric form as:

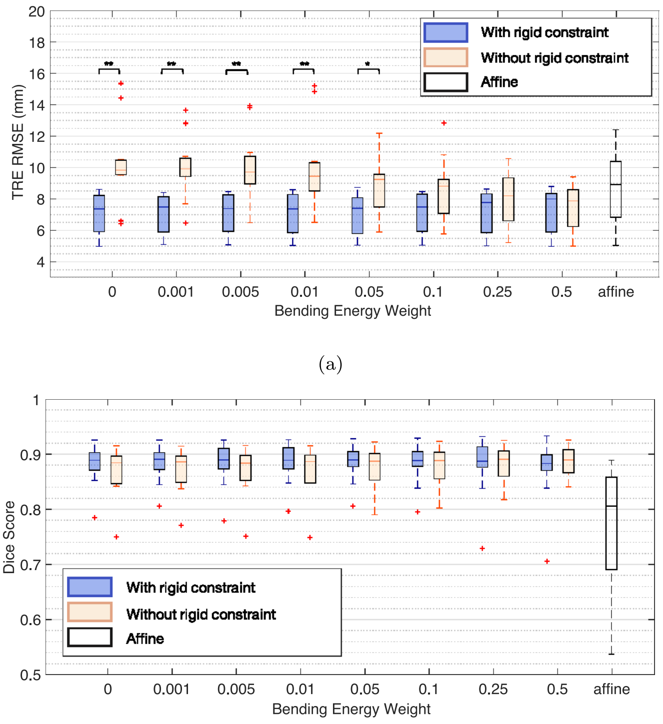
$$\mathrm{TRE}_i = \frac{1}{2}\left(\|\mathbf{x_i} - \phi^{-1}(\mathbf{y_i})\|_{\ell^2} + \|\phi(\mathbf{x_i}) - \mathbf{y_i}\|_{\ell^2}\right), \qquad (4)$$

where $\mathbf{x_i}$ is the landmark position in the CT space and $\mathbf{y_i}$ is the corresponding landmark position in the MRI.

For each landmark and each subject we computed the average of the TRE values from the different manually selected landmark sets. For each subject and for each registration approach, we reported the root mean square error (RMSE) of the TRE values across the ten landmarks. Fig. 5(a) shows the RMSE TRE distributions at varying bending energy weights for the registration with and without rigid constraints. Overall, the rigidly-constrained registration algorithm not only provides clinically plausible deformations, but it also outperforms the standard non-linear one in effectively reducing the TRE RMSE, producing therefore a more accurate alignment at the landmark locations. We also observed a reduced sensitivity to the choice of the regularisation parameter. Moreover, the introduced algorithmic steps for the rigidity constraint did not impact on the total computation time (on the same machine, the rigidly constrained non-linear registration required 93 minutes on average, compared to 96 minutes for the fully non-linear approach).

As a further comparison, we additionally manually delineated the gluteus medius of healthy hip sides on the template CT images and we computed the Dice score between the registered CT segmentation and MRI based segmentation. In particular, to obtain a symmetric estimation we computed the average between the Dice score in the CT space (by warping the manual MRI segmentation with the inverse transformation) and the Dice score in the MRI space (by warping the manual CT segmentation with the direct transformation). In Fig. 5(b), the distributions of the Dice score values are reported with and without the use of the rigid constraint using the same set of λ. In agreement with the previous result, higher Dice scores are obtained with the proposed method, with similar improved performances for regularisation weights lower than 0.1.

Both experiments have been used to select an optimal bending energy weight for further analysis. We selected $\lambda = 0.001$, being the value that yields to the best TRE RMSE, while still providing high degree of overlap for the segmented gluteus medius muscles



(a)

(b)

Fig. 5. (a) Comparison of TRE RMSE values obtained from the rigidly-constrained non-linear registration and the standard fully non-linear one with varying weights of the bending energy regularisation term. The TRE RMSE for the affine registration is reported as well. (b) Dice Score for gluteus medius segmentation overlap between registered CT and MRI (only the healthy hip sides are considered). Wilcoxon rank sum test was performed between pairs with same registration parameters, and statistically significant differences are reported (* indicates p-value $p < 0.05$, ** indicates $p < 0.01$).

(median Dice score [minimum, maximum] = 0.89 [0.81, 0.93]). A visual comparison of the registration results for one case is presented in Fig. 6. At equal regularisation ($\lambda = 0.001$), the Jacobian determinant maps show that the use of the rigid constraints enforces a volume-preserving deformation within bones and implant, ensuring anatomical plausibility of the applied transformation. For a fair comparison, we also reported the outcome of the standard non-linear registration with its optimal regularisation weight ($\lambda = 0.5$). While it promotes smoother transformations, it does not provide sufficient local deformation in the soft tissue to compensate for the different patient's position, especially on the implanted side. Also, the Jacobian determinant map still shows volume variations in the bony area, which are instead impeded by design with our method.

3.1.2. Leave-one-out cross validation on template data sets

To assess the performance of the automated segmentation approach, we designed a LOOCV experiment on the template datasets. Given the healthy ($N = 10$) or implanted ($N = 12$) dataset, we performed the multi-atlas based automated segmentation step for each template image using the remaining $N - 1$ templates and varying the segmentation propagation and the label fusion parameters. The obtained segmentations were then compared to the ground truth by computing the Dice Score for each label. Three examples of the automated results are shown in Fig. 7.

This test was performed with three different settings: (1) using only the CT images; (2) using only the MR images; and (3) using the registered 4D CT-MR images. No muscles segmentations were available for the CT, and similarly no bones and implant labelling were available on the MRI. Only the available labels were considered in the single-modality experiments. For each setting, we selected the optimal set of multi-atlas-based automated segmentation parameters as the ones maximising the lowest Dice Score across subjects and across labels; when different sets of parameters had values within 0.05 difference, we kept the one with the highest median Dice Score. The median Dice score values for bones, muscles and implants obtained with the best set of parameters for each experiment setting are reported in Table 2. As reported in our earlier work [39], overall the bones and the implant are better segmented by the multi-atlas based segmentation framework with respect to muscles, given their lower shape and texture variability. Compared to our previously reported results, however, the introduction of a metal artefact reduction correction on the CT images as well as the improvements in the MR pre-processing led to an increase of Dice score values. By improving the quality of the images, we facilitated the intra- and inter-subject registrations, therefore improving the performance of our pipeline. When testing the hypothesis of same underlying distributions for Dice Score values obtained with single- or multi-modality (Wilcoxon rank-sum test, 5% significance level), significant difference was found only for the segmentation of the bony structures, although the sample size is too small to draw any definite conclusion. The implanted side provided slighlty lower values with our multimodal pipeline. This side is more affected by residual metal artefact in CT and lack of metal intensity information in the MR, which hamper both the intra- and inter-subject registration. Moreover, in the multimodal experiment the multi-atlas segmentation parameters are selected based on all the musculoskeletal structures, finding a trade-off between the segmentation accuracy of bones, muscles and implant. This differs from the single-modality cases, where the parameters are selected only on bones and implant (CT) or muscles (MRI), therefore being more performant on such structures.

Nonetheless, our proposed 4D framework is able to provide a consistent and unified solution to the segmentations of both the CT and the MRI. Our approach guarantees no overlap between the segmented ROIs, which cannot be ensured by the use of independent approaches for muscular segmentation on the MRI and bone or implant on the CT. On the template datasets, we verified that on average 2% of the voxels labeled as muscle on the MRI overlapped with CT-labeled bone voxels in our manual segmentations. In addition to this, without the use of a registration framework able to combine the two modalities while maintaining their biological plausibility, the joint segmentation of both muscular and skeletal structures would be more challenging and less accurate on the single modality. To test this hypothesis, we adopted the same LOOCV framework to obtain an automated segmentation of the Gluteus Medius on CT images, exploiting the available manual segmentations of the healthy side as in Section 3.1.1. The comparison of the Dice Score values from CT, MR and combined multimodal framework is reported in Fig. 8. A visual example is also shown in Fig. 9, reporting the automated segmentation of GMED and the ground truth. While our multimodal approach provides a reasonable result, the single modality results are less satisfactory: the CT-only automated segmentation is not able to clearly recover the boundary between fat and muscle; in the MRI-only, parts of the femur are wrongly classified as muscle. Overall, the lack of contrast for muscular structures in CT images hampers the templates manual delineation process, introducing larger variability in the segmentations propagated to a target space and therefore hindering the accuracy of the final consensus. In the MRI space, the manual segmentation of the templates is more reliable, but the qualitative nature of this modality makes the inter-subject registrations more difficult in the multi-atlas approach. Our 4D multimodal framework takes advantage of both modalities, as the

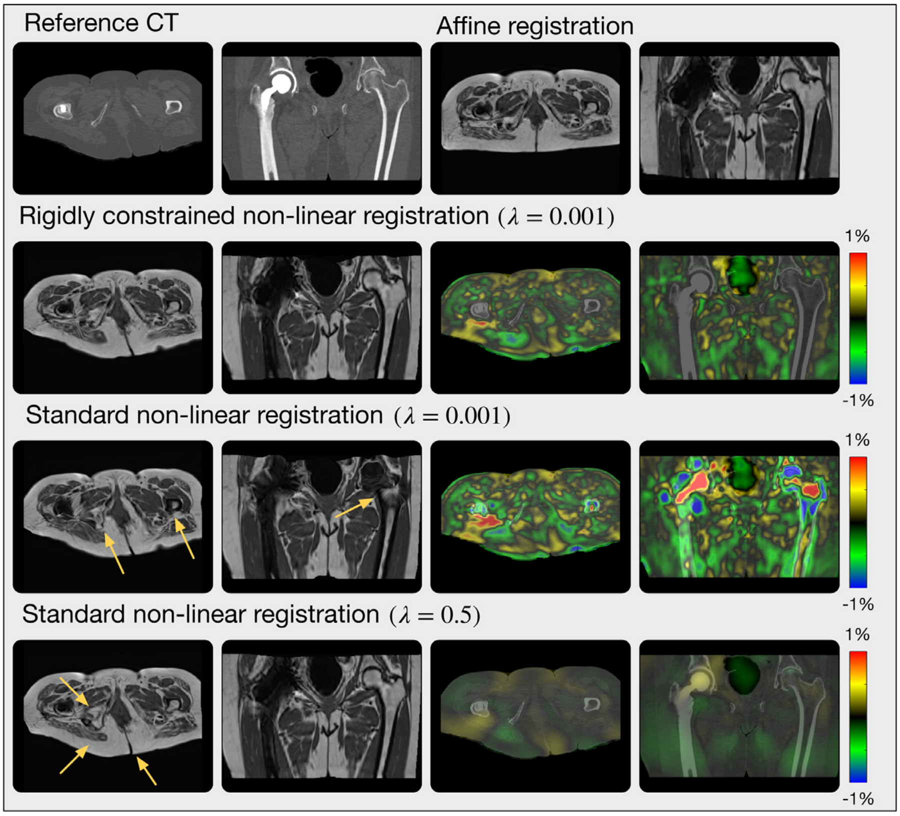


Fig. 6. Example of registration results. The central axial and coronal views for the reference CT, the affinely registered MRI, the non-linearly registered MRI with and without the use of the rigidity constraints are reported. The coloured insets show the Jacobian determinant maps (in percentage of volume change) for each non-linear deformation. Results are displayed with the rigidly-constrained registration optimal regularisation weight ($\lambda = 0.001$) to showcase the sole effect of the rigidity constraints. The result of the optimal fully non-linear regularisation weight ($\lambda = 0.5$) is also shown. Yellow arrows point at areas where the standard non-linear registration fails to recover a good alignment. Differently from this latter, our rigidly constrained non-linear registration allows for more localised deformation and better soft tissue alignment while preserving the volume and shape of bones and implant.

Table 2
Median Dice Score values and 95% confidence intervals for bones, implant and muscles: comparison between single- and multi-modality results. Wilcoxon rank sum test was performed to test the null hypothesis of same distribution for the multi-modality- and the respective single-modality-derived Dice Scores (obtained *p*-values are reported and starred are the cases of rejection of the null hypothesis with 5% significance level). N.A. indicates cases where the manual segmentation was not available.

Leave-One-Out-Cross-Validation for model parameter selection				
Healthy side				
	CT	MR	Multimodal	p-value
Bones	0.95 [0.72, 0.97]	N.A.	0.93 [0.75, 0.96]	0.036*
Muscles	N.A.	0.84 [0.59, 0.93]	0.85 [0.66, 0.92]	0.333
Implanted side				
	CT	MR	Multimodal	p-value
Bones	0.91 [0.70, 0.94]	N.A.	0.89 [0.69, 0.91]	0.005*
Muscles	N.A.	0.83 [0.60, 0.93]	0.81 [0.61, 0.92]	0.170
Implant	0.92 [0.81, 0.95]	N.A.	0.91 [0.84, 0.95]	0.371

CT helps regularising the templates-to-target registration while exploiting the more trustworthy manual segmentation from the MR, leading therefore to more robust results.

3.1.3. *Generalisability analysis*

As a final test of robustness of the proposed segmentation pipeline, we extended the LOOCV experiment to a nested version, in order to verify the generalisability of our approach to unseen data. In particular, we used N-1 subjects to perform the parameter selection in the same LOOCV fashion as described in the previous paragraph, and then we tested the performance of the model with the selected parameters on the hold-out subject. By iterating this over all the template cases, we obtained the summary Dice score statistics reported in Table 3. These results are in line with the full dataset LOOCV analysis, which represents the optimal performance, showing the robustness of our approach on a more realistic setting. The median Dice Scores are unchanged or within 1% difference, while a very slight reduction of the Dice score range in some of the classes is imputable to the use of less templates for the parameter selection. Overall, we can conclude that our approach is able to generalise to unseen data, as it does not show a significant drop in performance when tested on the hold-out set.

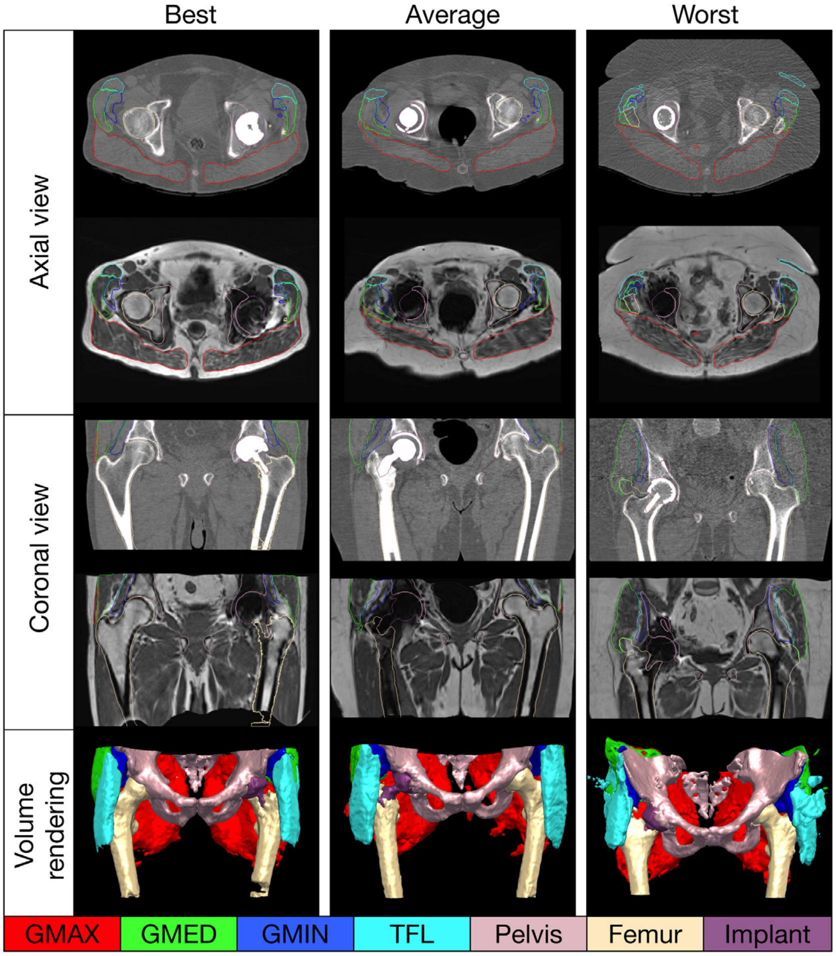


Fig. 7. Examples of automated segmentation results obtained with the proposed multimodal pipeline. Left, middle and right columns correspond to the subjects with best, average and worst mean Dice score across the segmented structures. They also demonstrate the large variability of musculoskeletal shapes, implant types, patient's position and degree of artefacts present in this type of images. For the first case our algorithm achieved plausible segmentation throughout the field of view, while in the second case the muscular structures are less well-defined. In the third case, instead, the strong metal artefact in the MR image affected the templates-to-target registration and the consequent segmentation accuracy.

Table 3
Median Dice Score values and 95% confidence intervals for bones, implant and muscles obtained from the nested LOOCV (*p*-values for 5% significance level of Wilcoxon rank sum test between monomodal and multimodal comparisons are shown). N.A. indicates that manual segmentation was not available.

Nested Leave-One-Out-Cross-Validation for Generalisability				
Healthy side				
	CT	MR	Multimodal	p-value
Bones	0.94 [0.65, 0.97]	N.A.	0.94 [0.75, 0.96]	0.298
Muscles	N.A.	0.85 [0.65, 0.92]	0.86 [0.71, 0.92]	0.179
Implanted side				
	CT	MR	Multimodal	p-value
Bones	0.91 [0.64, 0.95]	N.A.	0.88 [0.64, 0.92]	0.002*
Muscles	N.A.	0.83 [0.60, 0.93]	0.80 [0.58, 0.91]	0.051
Implant	0.92 [0.70, 0.96]	N.A.	0.91 [0.82, 0.94]	0.285

3.2. Clinical application: Muscle atrophy assessment

The remaining 27 subjects in our dataset were employed to test the accuracy of the proposed Intramuscular Fat Fraction as a measure of muscle atrophy. For these subjects, fatty atrophy of gluteus medius was graded by a radiologist according to the Bal and Lowe visual scoring system [9]: grade 0 corresponds to no atrophy, grade 1 is less than 30% change in muscle, grade 2 between 30% and 70%, and grade 3 more than 70% fatty change. Such radiological scores were available only for the implanted hip sides, and only the symptomatic side was considered in case of bilateral hip replacement.

The CT and MR images of these subjects were processed with the presented segmentation pipeline using the best sets of parameters from Section 3.1. The IFF for the pathological gluteus medius was then derived as in Section 2.4, and the association

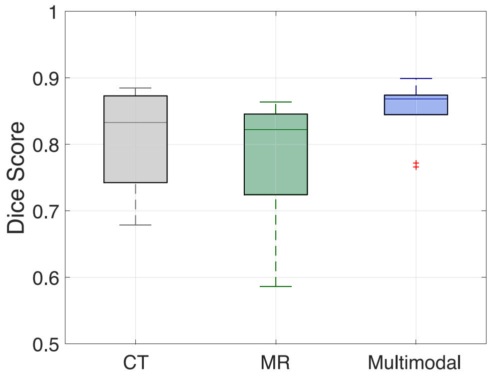


Fig. 8. Dice Score values for healthy-side GMED automated segmentation using only CT, only MR or our 4D CT-MR framework. The Dice Score is computed against the manual segmentation on the CT space for the CT framework, on the MR space for the MR framework, and on the MR space then registered to the CT space for the multimodal framework.

between our computed IFF and the gold-standard radiological score was tested using the one-way ANOVA F-test. Examples for each radiological scores of our obtained fat segmentation and the respective IFF values are reported in Fig. 10.

For the sake of comparison, we estimated the IFF from the MR images, i.e. discarding the extra information that the CT could provide. In this case, we manually identified a threshold for each subject to segment fat and lean muscle within the gluteus medius mask and we computed the IFF from this segmentation. The same analysis was performed with an automated thresholding on MRI, defined as the average percentile obtained from the manual thresholding. On our dataset, this corresponded to the 38th percentile of the full range of intensities in the masked MRI.

Fig. 11 displays the association of the IFF values and the radiological scores for the three analysed cases. The results of the one-way ANOVA F-test are reported in Table 4. These results show stronger association between the radiological score and IFF when this latter is computed from the CT image, as it provides improved separation between the different atrophy groups. It can be noticed that the group with radiological score equal to 1 is the most overlapping with other groups. As our analysis does not account for the inter- nor intra-rater variability of this visual score, this result is not unexpected as this class represents the intermediate situation where no clear atrophy pattern can be distinguished in the image. Further analysis will therefore be needed to better represent this intermediate class.

Table 4
One-way ANOVA F-test results. The table reports the p-values for the pair-wise comparison between the different classes, grouped according to the radiological score, and the *p*-value for the full test (total). Significance level is set at 5% (starred values).

Group pair	MRI manual	MRI automated	Multimodal automated (proposed)
0 vs 1	4.27E-01	7.79E-01	5.07E-01
0 vs 2	1.08E-01	6.80E-02	6.32E-02
0 vs 3	2.82E-05*	4.86E-06*	1.11E-07*
1 vs 2	5.20E-01	1.18E-01	2.61E-01
1 vs 3	3.47E-05*	1.01E-06*	3.13E-08*
2 vs 3	4.27E-03*	1.11E-03*	1.45E-05*
Total	1.85E-05*	9.58E-07*	2.55E-08*

4. Discussion

In this work, we presented an automated processing framework to register and jointly segment same-subject pelvic CT and MR images for hip arthroplasty assessment. Our framework allows to highlight both muscular and skeletal regions of interest on all available modalities. The implant, for example, can be contoured on the MR image, where the metal artefact completely shadows it. Visualising the spatial position of the implant with respect to muscular structures could help better determine the muscles at greater risk of developing atrophy or the presence of other inflammatory lesions. Also, by reducing the burden of manually selecting regions of interest through an automated segmentation scheme, 3D rendering of patient-specific anatomy as well as volumetric imaging biomarkers can be more easily derived.

We propose a flexible framework which can be readily adjusted to various types of input data. Despite relying on a dataset of registered CT-MRI atlases, the pipeline could be applied to a

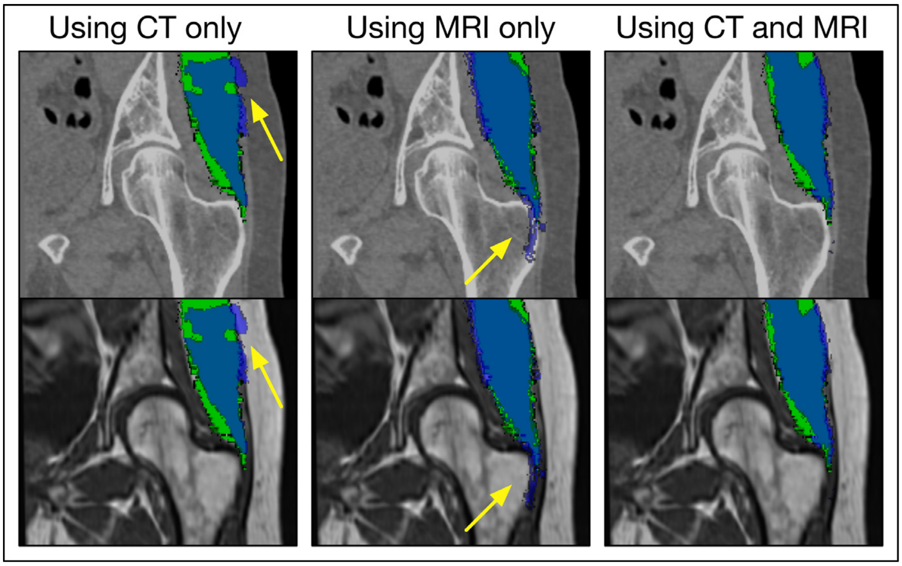


Fig. 9. An example of GMED automated segmentation from the LOOCV experiments. The automated segmentation (blue) is overlaid onto the manual ground truth (green). Although the first two columns show results for the single-modality experiments, we reported both images for clearer visualisation of failure. For this case, Dice Score values are: 0.68 (CT only), 0.72 (MR only), 0.77 (our multimodal approach).

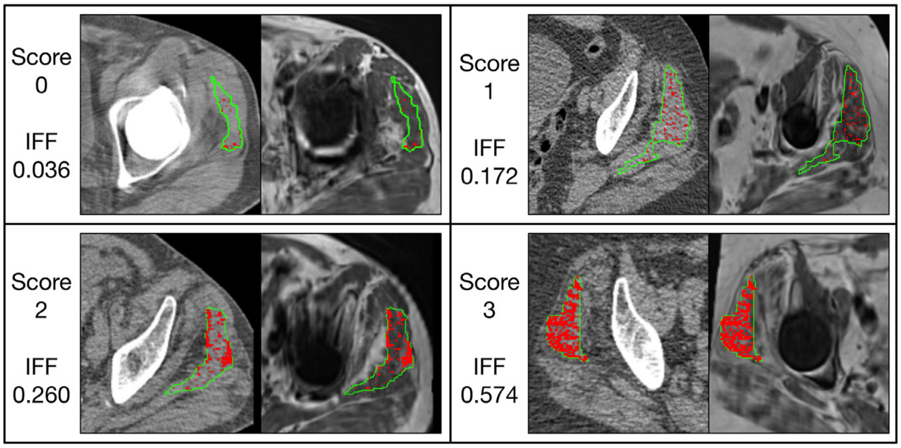


Fig. 10. Examples of registered CT and MRI for each radiological score, and the automated fat segmentation from the proposed method. The eroded gross muscle boundary is reported in green, while in red is the fat segmentation. The derived IFF is also reported, showing an increase of fat infiltration in agreement with the radiological score.

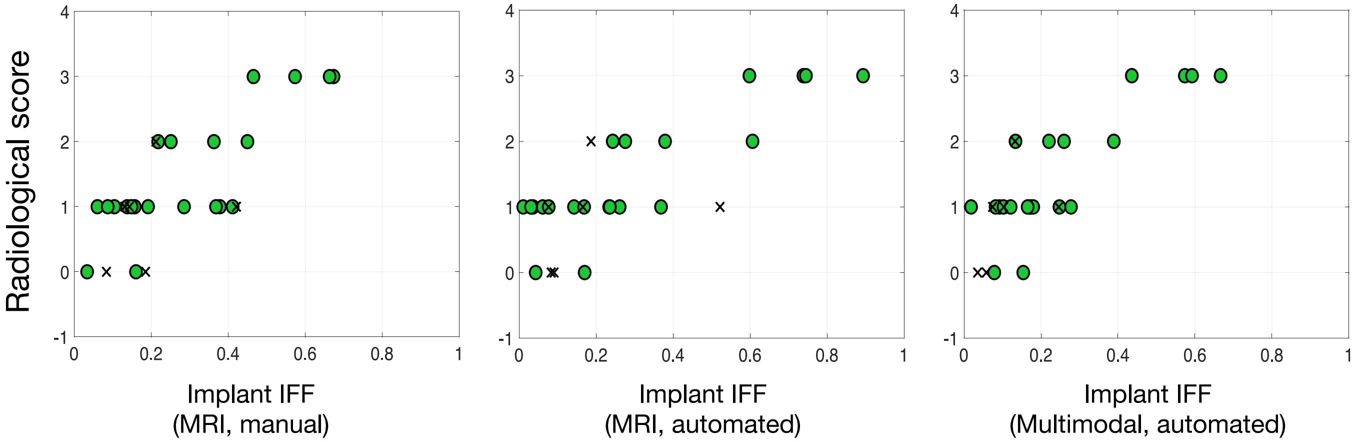


Fig. 11. Intramuscular Fat Fraction of Gluteus Medius (implanted side), estimated with different thresholding methods and plotted against the corresponding visual radiological score. In the right-most plot, cases where CT and MRI were acquired with more than 6-month difference are reported as crosses.

single modality and still provide the whole musculoskeletal range of structures while requiring only little modification. However, further analysis should be performed to test the robustness of the pipeline when used on single image or on other acquisition types.

We assessed the impact of all steps in our proposed pipeline and show their importance to overall better outcomes. The image quality enhancement step makes the pipeline suitable for routinely acquired clinical data, heavily corrupted by noise and artefacts. The use of a tailored bias field correction scheme and of the super resolution algorithm helps obtain high quality MR images with isotropic resolution, without the need for longer volumetric MR acquisitions. The introduction of a Metal Artefact Reduction (MAR) technique for the CT also improved the accuracy of the final segmentation by facilitating the registration tasks.

As the results in Section 3.1 show, our novel CT-MR intra-subject registration algorithm proved to be more robust to choice of regularisation parameters and more accurate than standard same-parametrisation non-linear registration. Although we acknowledge that other non-linear registration frameworks could be tested, we underline that the purpose of these experiments was not the identification of the most performing framework, but rather to assess the impact of the local rigidity constraint. Its introduction for bony and implant structures is crucial to obtain clinical trust, as it guarantees fidelity in the applied anatomical deformations. Indeed, differently from other approaches such as Staring et al. [16] and Reaungamornrat et al. [17], our hard-constraint formulation limits the optimisation only to strictly rigid transformations within rigid structures, preserving their shape and volume. We would also like to underline that our approach does not include any a priori knowledge on the considered anatomical area. Therefore, while it was applied on preservation of bone and implant rigidity in the proposed pipeline, our registration algorithm is already applicable to other medical image registration problems, provided that masks for the structures to be rigidly deformed are available.

For this specific application, we automatically extracted the bone masks using a segmentation propagation approach. As this could represent a weakness in our approach, we tested whether inaccuracies in the automated segmentation would adversely influence the subsequent pipeline steps (Appendix B). We found the method to be fairly robust, and no significant differences in the TRE estimation, in the final Dice Scores and in the derived IFF was identified when comparing the use of automated masks with manual ones.

Despite the demonstrated improvements on intra-subject registration accuracy and transformation plausibility, the proposed method still presents some limitations. Firstly, the non-linear step is initialised with global registration. If this registration fails or is not sufficiently accurate, the non-linear step would not be able to compensate for global misalignment due to limited capture range and local optima and, consequently, the resulting final segmentation would not be reliable. Secondly, the current parametrisation does not model sliding effects, which can cause large discontinuities in the deformation field within the acetabular

Table 5
Comparison with literature results. Mean Dice score and standard deviation are reported for different musculoskeletal structures. Literature methods were tested on non implanted patients, hence only “healthy” hip sides were considered for the proposed method. In bold is the highest value for each structure.

	Yokota et al. [38] (CT only)	Hiasa et al. [51] (MR+synthetic CT)	Proposed (MR + CT)
Pelvis	–	0.808 ± 0.036	**0.882** ± 0.072
Femur	–	0.883 ± 0.029	**0.949** ± 0.013
GMAX	**0.921**	–	0.906 ± 0.024
GMED	**0.875**	0.804 ± 0.040	0.850 ± 0.045
GMIN	0.697	0.669 ± 0.054	**0.826** ± 0.040
TFL	**0.807**	–	0.797 ± 0.088

space when large femoral rotation occurs between the patient's position in the MR and in the CT scanners. When compared to rigid registration focussed on individual bones only, we found that the proposed technique was able to correctly recover rotations up to 8° along any axis (templates dataset median rotation angle: 3°). Future work will therefore address the modelling of sliding effects in our registration framework, to improve the rotation range captured by our algorithm.

Should the intra-subject registration succeed, we do not expect any of the subsequent steps to fail but rather to be non-optimal in the worst case. This is due to the robustness of the pipeline. First, a robust affine scheme is used for inter-subject registration, which allows up to 50% of the pairwise affine registrations to fail without compromising the final outcome. Second, the fusion algorithm automatically selects the template images that appear to be the most similar to the template and thus the best registered. This second aspect also enables the pipeline to deal with several non-linear inter-subject registration failures. In the leave-one-out cross validation experiment, the presented multi-channel multi-atlas segmentation approach proved to achieve good accuracy with median Dice Score of 0.90 for skeletal and 0.84 for muscular structures, in line with the results obtained from the single modality experiments (Table 2). The implanted sides overall reported a slight reduction in accuracy with respect to the single modality, but it is reasonable to expect an improvement on this performance following further advances in CT-MR registration quality, such as the introduction of the aforementioned sliding modelling and increased template dataset size.

A comparison with two of the most recent works presented in the literature is summarised in Table 5, reporting the mean Dice Score and standard deviation (if available) for the available segmented structures. Yokota et al. [38] applied a hierarchical multi-atlas approach for muscle segmentation in CT, while Hiasa et al. [51] proposed a multimodal framework for CT-MR image synthesis, followed by a 2D UNet for segmentation of musculoskeletal structures. Although definite conclusions cannot be drawn due to differences in the data, field of view and especially the presence/absence of metal artefact induced noise, overall our method reported higher segmentation accuracy. It outperformed the multimodal segmentation network proposed by Hiasa et al. in all the compared structures, though we acknowledge that their use of synthetically generated data for training might not be representative of performance on real data. While Yokota *et al.* reported slightly higher dice score for GMAX, GMED and TFL (still within one standard deviation from our result), we obtained improved performance on the smaller GMIN, typically characterised by higher shape variability. Together with the experiment on the gluteus medius (Fig. 8), this comparison supports the usefulness of a multimodal approach, as this combines the reliability of manually delineated MR templates with CT-facilitated templates-to-target registration.

With the advent of highly performing deep learning approaches in medical imaging, it can be anticipated that novel segmentation methods will outperform atlas-based approaches in musculoskeletal imaging as well, both in accuracy and speed. However, the success of most deep learning algorithms still depends on the availability of large annotated datasets that can be used for training. Hence, together with the already mentioned advantages, the proposed multi-atlas based segmentation pipeline can play an important part in helping the development of such datasets.

The identification of reliable and quantitative imaging biomarkers that could be automatically extracted would facilitate the interpretation of the image and the assessment of implant failure. In the case of muscle atrophy, current clinical practice is based on a visual and qualitative assessment of few 2D slices, which is therefore strongly subjective and does not account for the whole muscular volume. The proposed Intramuscular Fat Fraction estimation takes advantage of the multimodality information to automatically and more robustly delineate regions of interest. Whilst retaining the 3D nature of muscles, our approach also avoids the need for manual ROI selection, which is currently the most common practice for 3D fatty infiltration estimation [11,52]. Differently from MR-based approaches, the use of a standardised range of intensities makes the automated classification into lean muscle and fat straightforward in CT. We acknowledge that a large time gap between the acquisition of the CT and the MRI would jeopardize the reliability of our comparison between the multimodal IFF estimation and the MRI radiological score, as muscle change could have occured in between. However, clinical literature shows that significant gluteus medius atrophy change appears after a mean interval of 11 months [4]. In our dataset only six cases had a time difference greater than 6 months, and except for one case with grade 2 atrophy, all the others were associated with no or very little fatty infiltration. For this reason we included all the available cases in our analysis, in order to maximise the sample size. A very similar trend was indeed found even when the aforementioned six cases were excluded. However, our analysis is still constrained by the limited number of cases available for each radiological score. It would be of interest to extend the IFF analysis to the other gluteal muscles, including both an estimation of inter-rater variability of the current visual scoring system and ideally increasing the dataset to better analyse the role of confounding factors, such as gender or time from primary surgery. Further analysis will also be required in future to better characterise the effect of partial volume as well as any residual metal artefact hyper- or hypo-intensities in the estimation of IFF. Nonetheless, the proposed multimodal muscle atrophy estimation is advantageous as it combines the robustness of MRI-driven gross muscle volume segmentation with a consistent and reproducible fat quantification from CT.

5. Conclusions

In this paper, we have introduced a novel framework for the simultaneous automated segmentation of CT and MR images of patients with hip implants and for a multimodal-driven postoperative muscle assessment. The proposed framework represents a promising tool to support image analysis in hip arthroplasty, being robust to the presence of implants and associated image artefacts which heavily degrade the image quality. Our framework enables the automated extraction of multiple imaging biomarkers (e.g. Intramuscular Fat Fraction, volume asymmetries detection, etc), encouraging research that could quantitatively inform the decision-making process about patient's management.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

This work is supported by the EPSRC-funded UCL Centre for Doctoral Training in Medical Imaging [EP/L016478/1], the Royal National Orthopaedic Hospital NHS Trust, the EPSRC [NS/A000027/1], an Innovative Engineering for Health award by the Wellcome Trust [WT101957], the Wellcome Flagship Programme [WT213038/Z/18/Z] and the Wellcome/EPSRC Centre for Medical Engineering [WT 203148/Z/16/Z, NS/A000049/1]. This research is also supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. TV is supported by a Medtronic / Royal Academy of Engineering Research Chair [RCSRF1819/7/34]. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2019.105062.

References

[1] K.R. Fingar, C. Stocks, A.J. Weiss, C.A. Steiner, Most frequent operating room procedures performed in U.S. hospitals, HCUP Stat. Brief 186 (2014) 1–15.

[2] Royal College of Surgeons, Surgery and the NHS in numbers, 2014. Retrieved from https://goo.gl/csnf94.

[3] M. Green, N. Wishart, E. Young, V. Mccormack, M. Swanson, National joint registry for england, wales, northern ireland and the isle of man, 14th Annual Report, 1821(December 2016) (2017).

[4] R. Berber, M. Khoo, E. Cook, A. Guppy, J. Hua, J. Miles, R. Carrington, J. Skinner, A. Hart, Muscle atrophy and metal-on-metal hip implants: a serial MRI study of 74 hips, Acta Orthop 86 (3) (2015) 351–357, doi:10.3109/17453674.2015.1006981.

[5] V. Pedoia, S. Majumdar, T.M. Link, Segmentation of joint and musculoskeletal tissue in the study of arthritis, Magn. Reson. Mater Phy. 29 (2) (2016) 207–221, doi:10.1007/s10334-016-0532-9.

[6] S.S. Blemker, D.S. Asakawa, G.E. Gold, S.L. Delp, Image-based musculoskeletal modeling: applications, advances, and future opportunities, J. Magn. Res. Imaging 25 (2) (2007) 441–451, doi:10.1002/jmri.20805.

[7] MHRA of the Department of Health, Medical device alert, MHRA Database (2012) 1–7.

[8] C.W.A. Pfirrmann, H.P. Notzli, C. Dora, J. Hodler, M. Zanetti, Abductor tendons and muscles assessed at MR imaging after total hip arthroplasty in asymptomatic and symptomatic patients, Radiology 235 (3) (2005) 969–976, doi:10.1148/radiol.2353040403.

[9] B.S. Bal, J.A. Lowe, H. Arthroplasty, M.R.I. Evidence, A.E. Hillard, Muscle damage in minimally invasive total hip arthroplasty: MRI evidence that it is not significant, Inst Course Lect 57 (2008) 223–229.

[10] M.E. Vidt, A.C. Santago, C.J. Tuohy, G.G. Poehling, M.T. Freehill, R.A. Kraft, A.P. Marsh, E.J. Hegedus, M.E. Miller, K.R. Saul, Assessments of fatty infiltration and muscle atrophy from a single magnetic resonance image slice are not predictive of 3-dimensional measurements, Arthrosc. J. Arthrosc. Related Surg. 32 (1) (2016) 128–139, doi:10.1016/j.arthro.2015.06.035.

[11] D.C. Karampinos, T. Baum, L. Nardo, H. Alizai, H. Yu, J. Carballido-Gamio, S.P. Yap, A. Shimakawa, T.M. Link, S. Majumdar, Characterization of the regional distribution of skeletal muscle adipose tissue in type 2 diabetes using chemical shift-based water/fat separation, J. Magn. Reson. Imaging 35 (4) (2012) 899–907, doi:10.1002/jmri.23512.

[12] H. Lee, F.M. Troschel, S. Tajmir, G. Fuchs, J. Mario, F.J. Fintelmann, S. Do, Pixel-Level deep segmentation: artificial intelligence quantifies muscle on computed tomography for body morphometric analysis, J. Digital Imaging 30 (4) (2017) 487–498, doi:10.1007/s10278-017-9988-z.

[13] Y. Otake, F. Yokota, N. Fukuda, M. Takao, S. Takagi, N. Yamamura, L.J. O'Donnell, C.-F. Westin, N. Sugano, Y. Sato, Patient-specific skeletal muscle fiber modeling from structure field of clinical ct images, in: M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D.L. Collins, S. Duchesne (Eds.), Medical Image Computing and Computer Assisted Intervention MICCAI, Springer International Publishing, Cham, 2017, pp. 656–663.

[14] E.A. Bogner, C.M. Sofka, CT Evaluation of total hip arthroplasty complication: dissociation of acetabular component, HSS J. 3 (1) (2007) 112–114, doi:10.1007/s11420-006-9026-z.

[15] D. Rivest-Hénault, N. Dowson, P.B. Greer, J. Fripp, J.A. Dowling, Robust inverse-consistent affine CT-MR registration in MRI-assisted and MRI-alone prostate radiation therapy, Med. Image Anal. 23 (1) (2015) 56–69, doi:10.1016/j.media.2015.04.014.

[16] M. Staring, S. Klein, J.P.W. Pluim, A rigidity penalty term for nonrigid registration, Med. Phys. 34 (11) (2007) 4098–4108, doi:10.1118/1.2776236.

[17] S. Reaungamornrat, a.S. Wang, a. Uneri, Y. Otake, a.J. Khanna, J.H. Siewerdsen, Deformable image registration with local rigidity constraints for cone-beam CT-guided spine surgery, Phys. Med. Biol. 59 (14) (2014) 3761–3787, doi:10.1088/0031-9155/59/14/3761.

[18] J.A. Dowling, J. Sun, P. Pichler, D. Rivest-Hénault, S. Ghose, H. Richardson, C. Wratten, J. Martin, J. Arm, L. Best, S.S. Chandra, J. Fripp, F.W. Menk, P.B. Greer, Automatic substitute computed tomography generation and contouring for magnetic resonance imaging (MRI)-alone external beam radiation therapy from standard MRI sequences, Int. J. Radiat. Oncol. Biol. Phys. 93 (5) (2015) 1144–1153, doi:10.1016/j.ijrobp.2015.08.045.

[19] D. Rivest-Hénault, P. Greer, J. Fripp, J. Dowling, Structure-Guided nonrigid registration of CTMR pelvis scans with large deformations in MR-Based image guided radiation therapy david, CLIP, LNCS 8361 (2014) 65–73, doi:10.1007/978-3-319-46472-5.

[20] O. Commowick, N. Wiest-Daesslé, S. Prima, Automated diffeomorphic registration of anatomical structures with rigid parts: application to dynamic cervical MRI, in: Proceedings of the 15th International Conference on Medical Image Computing and Computer Assisted Intervention, LNCS, 15, Springer, Nice, France, 2012, pp. 163–170, doi:10.1007/978-3-642-33418-4_21.

[21] E. Haber, S. Heldmann, J. Modersitzki, A computational framework for image-based constrained registration, Linear Algebra Appl. 431 (3) (2009) 459–470, doi:10.1016/j.laa.2009.03.020.

[22] F. Yokota, T. Okada, M. Takao, N. Sugano, Y. Tada, N. Tomiyama, Y. Sato, Automated CT segmentation of diseased hip using hierarchical and conditional statistical shape models, in: K. Mori, I. Sakuma, Y. Sato, C. Barillot, N. Navab (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 190–197, doi:10.1007/978-3-642-40763-5_24.

[23] C. Chu, C. Chen, L. Liu, G. Zheng, FACTS: Fully automatic CT segmentation of a hip joint, Ann. Biomed. Eng. 43 (5) (2015) 1247–1259, doi:10.1007/s10439-014-1176-4.

[24] C. Chu, J. Bai, X. Wu, G. Zheng, MASCG: Multi-Atlas segmentation constrained graph method for accurate segmentation of hip CT images, Med. Image Anal. 26 (1) (2015) 173–184, doi:10.1016/j.media.2015.08.011.

[25] M. van Eijnatten, R. van Dijk, J. Dobbe, G. Streekstra, J. Koivisto, J. Wolff, CT Image segmentation methods for bone used in medical additive manufacturing, Med. Eng. Phys. 0 (2017) 1–11, doi:10.1016/j.medengphy.2017.10.008.

[26] E. Ahmad, M.H. Yap, H. Degens, J.S. McPhee, Atlas-registration based image segmentation of MRI human thigh muscles in 3D space, Proc. SPIE, Med. Imaging 9037 (2014) 90371L, doi:10.1117/12.2043606.

[27] A. Le Troter, A. Fouré, M. Guye, S. Confort-Gouny, J.P. Mattei, J. Gondin, E. Salort-Campana, D. Bendahan, Volume measurements of individual muscles in human quadriceps femoris using atlas-based segmentation approaches, Mag. Reson. Mater. Phys. Biol. Med. 29 (2) (2016) 245–257, doi:10.1007/s10334-016-0535-6.

[28] B. Gilles, N. Magnenat-Thalmann, Musculoskeletal MRI segmentation using multi-resolution simplex meshes with medial representations, Med. Image Anal. 14 (3) (2010) 291–302, doi:10.1016/j.media.2010.01.006.

[29] A. Ogier, M. Sdika, A. Foure, A. Le Troter, D. Bendahan, Individual muscle segmentation in MR images: a 3D propagation through 2D non-linear registration approaches, Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS(2017) 317–320. doi:10.1109/EMBC.2017.8036826.

[30] F. Ozdemir, N. Karani, P. Fürnstahl, O. Goksel, Interactive segmentation in MRI for orthopedic surgery planning: bone tissue, Int. J. Comput. Ass. Radiol. Surgery 12 (6) (2017) 1031–1039, doi:10.1007/s11548-017-1570-0.

[31] J. Schmid, J. Kim, N. Magnenat-Thalmann, Robust statistical shape models for MRI bone segmentation in presence of small field of view, Med. Image Anal. 15 (1) (2011) 155–168, doi:10.1016/j.media.2010.09.001.

[32] S.S. Chandra, Y. Xia, C. Engstrom, S. Crozier, R. Schwarz, J. Fripp, Focused shape models for hip joint segmentation in 3D magnetic resonance images, Med. Image Anal. 18 (3) (2014) 567–578, doi:10.1016/j.media.2014.02.002.

[33] Y. Xia, J. Fripp, S.S. Chandra, R. Schwarz, C. Engstrom, S. Crozier, Automated bone segmentation from large field of view 3D MR images of the hip joint, Phys. Med. Biol. 58 (20) (2013) 7375–7390, doi:10.1088/0031-9155/58/20/7375.

[34] F. Liu, Z. Zhou, H. Jang, A. Samsonov, G. Zhao, R. Kijowski, Deep convolutional neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging, Mag. Reson. Med. 00 (2017), doi:10.1002/mrm.26841.

[35] G. Zeng, X. Yang, J. Li, L. Yu, P.-A. Heng, G. Zheng, 3D U-net with multi-level deep supervision: fully automatic segmentation of proximal femur in 3D MR Images, in: Q. Wang, Y. Shi, H.-I. Suk, K. Suzuki (Eds.), Machine Learning in Medical Imaging, Springer International Publishing, Cham, 2017, pp. 274–282.

[36] J. Yao, W. Kovacs, N. Hsieh, C.-Y. Liu, R.M. Summers, Holistic segmentation of intermuscular adipose tissues on thigh MRI, in: M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D.L. Collins, S. Duchesne (Eds.), Medical Image Computing and Computer Assisted Intervention MICCAI, Springer International Publishing, Cham, 2017, pp. 737–745.

[37] C. Klemt, M. Modat, J. Pichat, M.J. Cardoso, J. Henckel, A. Hart, S. Ourselin, Automatic assessment of volume asymmetries applied to hip abductor muscles in patients with hip arthroplasty, Proc. SPIE Med. Imaging 9413 (2015) 94131M, doi:10.1117/12.2082341.

[38] F. Yokota, Y. Otake, M. Takao, T. Ogawa, T. Okada, N. Sugano, Y. Sato, Automated muscle segmentation from CT images of the hip and thigh using a hierarchical multi-atlas method, Int. J. Comput. Ass. Radiol. Surg. (1) (2018), doi:10.1007/s11548-018-1758-y.

[39] M.B.M. Ranzini, M. Ebner, M.J. Cardoso, A. Fotiadou, T. Vercauteren, J. Henckel, A. Hart, S. Ourselin, M. Modat, Joint multimodal segmentation of clinical ct and mr from hip arthroplasty patients, in: B. Glocker, J. Yao, T. Vrtovec, A. Frangi, G. Zheng (Eds.), Computational Methods and Clinical Applications in Musculoskeletal Imaging, Springer International Publishing, Cham, 2018, pp. 72–84.

[40] S.A. Sabah, A.W.M. Mitchell, J. Henckel, A. Sandison, J.A. Skinner, A.J. Hart, Magnetic resonance imaging findings in painful metal-On-Metal hips. a prospective study, J. Arthropl. 26 (1) (2011) 71–76.e2, doi:10.1016/j.arth.2009.11.008.

[41] G. Treece, Refinement of clinical x-ray computed tomography (ct) scans containing metal implants, Comput. Med. Imaging Graph. 56 (Supplement C) (2017) 11–23, doi:10.1016/j.compmedimag.2017.01.005.

[42] K. Gorgolewski, C. Burns, C. Madison, D. Clark, Y. Halchenko, M. Waskom, S. Ghosh, Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python, Front. Neuroinf. 5 (2011) 13, doi:10.3389/fninf.2011.00013.

[43] N.J. Tustison, B.B. Avants, P.A. Cook, Y. Zheng, A. Egan, P.A. Yushkevich, J.C. Gee, N4ITK: Improved N3 bias correction, IEEE Trans. Med. Imaging 29 (6) (2010) 1310–1320, doi:10.1109/TMI.2010.2046908.

[44] K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens, Automated model-based bias field correction of MR images of the brain, IEEE Trans. Med. Imaging 18 (10) (1999) 885–896, doi:10.1109/42.811268.

[45] M. Ebner, M. Chouhan, P.A. Patel, D. Atkinson, Z. Amin, S. Read, S. Punwani, S. Taylor, T. Vercauteren, S. Ourselin, Point-spread-function-aware slice-to-volume registration: application to upper abdominal MRI super-resolution, in: M.A. Zuluaga, K. Bhatia, B. Kainz, M.H. Moghari, D.F. Pace (Eds.), Reconstruction, Segmentation, and Analysis of Medical Images, Springer International Publishing, Cham, 2017, pp. 3–13.

[46] M. Modat, D.M. Cash, P. Daga, G.P. Winston, J.S. Duncan, S. Ourselin, Global image registration using a symmetric block-matching approach, J. Med. Imaging 1 (2) (2014) 024003, doi:10.1117/1.JMI.1.2.024003.

[47] M. Modat, G.R. Ridgway, Z.A. Taylor, M. Lehmann, J. Barnes, D.J. Hawkes, N.C. Fox, S. Ourselin, Fast free-form deformation using graphics processing units, Comput. Methods Progr. Biomed. 98 (3) (2010) 278–284, doi:10.1016/j.cmpb.2009.09.002.

[48] J.M. Cardoso, K. Leung, M. Modat, S. Keihaninejad, D. Cash, J. Barnes, N.C. Fox, S. Ourselin, STEPS: Similarity and truth estimation for propagated segmentations and its application to hippocampal segmentation and brain parcelation, Med. Image Anal. 17 (6) (2013) 671–684, doi:10.1016/j.media.2013.02.006.

[49] M. Modat, P. Daga, M.J. Cardoso, S. Ourselin, G.R. Ridgway, J. Ashburner, Parametric non-rigid registration using a stationary velocity field, in: Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis, 2012, pp. 145–150, doi:10.1109/MMBIA.2012.6164745.

[50] A. Karlsson, J. Rosander, T. Romu, J. Tallberg, A. Grönqvist, M. Borga, O. Dahlqvist Leinhard, Automatic and quantitative assessment of regional muscle volume by multi-atlas segmentation using whole-body water-fat MRI, J. Mag. Resona. Imaging 41 (6) (2015) 1558–1569, doi:10.1002/jmri.24726.

[51] Y. Hiasa, Y. Otake, M. Takao, T. Matsuoka, K. Takashima, J. L. Prince, N. Sugano, Y. Sato, Cross-modality image synthesis from unpaired data using CycleGAN: Effects of gradient consistency loss and training data size, arXiv e-prints, (2018), arXiv:1803.06629.

[52] V. Ugarte, U. Sinha, V. Malis, R. Csapo, S. Sinha, 3D Multimodal spatial fuzzy segmentation of intramuscular connective and adipose tissue from ultrashort TE MR images of calf muscle, Mag. Resona. Med. 77 (2) (2017) 870–883, doi:10.1002/mrm.26156.