Automated Multi-Atlas Segmentation of Gluteus Maximus from Dixon and T1-Weighted Magnetic Resonance Images

Martin A. Belzunce, PhD[1], Johann Henckel, MD[1], Anastasia Fotiadou, PhD,[1], Anna Di Laura, PhD[1], and Alister Hart, MD[1,2]

1 Royal National Orthopaedic Hospital, Stanmore HA7 4LP, UK

2 Institute of Orthopaedics and Musculoskeletal Science, University College London, Stanmore HA7 4LP, UK

Corresponding author: Martin A. Belzunce at Research, Royal National Orthopaedic Hospital (RNOH), Brockley Hill, Stanmore, Middlesex, HA7 4LP. E-mail: martin.belzunce@nhs.net.

Authors' Contribution

Belzunce M. A.: Study conception and design, acquisition of data, analysis and interpretation of data, drafting of manuscript, critical revision

Johann Henckel: Study conception and design, acquisition of data, analysis and interpretation of data, critical revision

Anastasia Fotiadou: Acquisition of data, critical revision

Anna Di Laura: Study conception and design, Acquisition of data, critical revision

Alister Hart: Study conception and design, acquisition of data, analysis and interpretation of data, critical revision

Abstract

Objective: To design, develop and evaluate an automated multi-atlas method for segmentation and volume quantification of gluteus maximus from Dixon and T1-weighted images.

Materials and Methods: The multi-atlas segmentation method uses an atlas library constructed from 15 Dixon MRI scans of healthy subjects. A non-rigid registration between each atlas and the target, followed by majority voting label fusion, is used in the segmentation. We propose a region of interest (ROI) to standardize the measurement of muscle bulk. The method was evaluated using the dice similarity coefficient (DSC) and the relative volume difference (RVD) as metrics, for Dixon and T1-weighted target images.

Results: The mean(±SD) DSC was 0.94±0.01 for Dixon images, while 0.93±0.02 for T1-weighted. The RVD between the automated and manual segmentation had a mean(±SD) value of 1.5±4.3% for Dixon and 1.5±4.8% for T1-weighted images. In the muscle bulk ROI the DSC was 0.95±0.01 and the RVD was 0.6±3.8%.

Conclusion: The method allows an accurate fully automated segmentation of gluteus maximus for Dixon and T1-weighted images and provides a relatively accurate volume measurement in shorter times (~20 min) than the current gold-standard manual segmentations (2 hours). Visual inspection of the segmentation would be required when higher accuracy is needed.

Introduction

The gluteal muscles play an important role in daily living activities, including walking, running, stair climbing and lifting activities [1, 2]. Gluteus maximus is the largest of the gluteal muscles and has many different functions such as providing sacroiliac joint stability, strength for lifting and, in the running gait cycle, flexes the trunk on the stance-side and decelerate the swing leg [3, 4]. The assessment of gluteus maximus is of interest in a wide number of applications for example, in patients with osteoarthritis (OA) differences of 15% in the gluteus maximus volume between affected and unaffected limbs have been reported [5, 6]; in the same context, an increase of up to 40% in gluteus maximus volume was observed in a longitudinal study looking at patients that went through hip arthroplasty [7]. Gluteus maximus volume is also of interest in sports science [8, 9], rehabilitation [10] and plastic surgery [11].

Magnetic resonance imaging (MRI) is the best imaging modality to quantitatively assess muscles since it provides better soft-tissue contrast than other imaging techniques and additionally it does not require to expose the patient to radiation as in CT scans. Volume is one of the most important quantitative metrics to assess muscles since it determines muscle power [12, 13], which is associated with functional ability to perform daily activities [14, 15] and also with performance in sports practice [16]. Another important metric in muscle assessment is the fat-muscle ratio that, in MRI imaging, can be quantified using water-fat separation techniques such as chemical-shift based acquisition methods, also known as Dixon sequences [17–20]. These type of sequences have been successfully used to assess intramuscular fat or muscle fat infiltration [21–24].

Despite the importance of measuring volume, shape and fat-muscle ratio of individual muscles, 3D segmentation and labelling of muscles is hardly performed in clinical studies because of the limited availability of robust automated musculoskeletal segmentation tools. Manual

segmentation is the gold standard method for labelling muscles, which is a tedious, extremely time-demanding and not cost-effective task. For this reason, researchers usually estimate volume from cross-sectional areas (CSA) [25–28], where only a very small number of slices are labelled and surrogate volume metrics are used. However, 3D labelling of muscles provides not only volume, but also shape and the possibility of extracting other quantitative metrics (i.e. intramuscular fat). Therefore, automated methods to segment gluteus maximus are an important unsatisfied need.

Automated segmentation of skeletal muscle from MR images is a challenging task since individual muscles share similar intensity/contrast values [29]. Intensity based methods have been proposed for the thigh muscles thanks to its simpler anatomy, although with limited accuracy [30–32]. Multi-atlas segmentation, originally proposed for brain segmentation [33], have the ability to overcome the limitations of intensity-based methods by using a set of manually segmented images (atlases) to encode the relationship between the segmentation labels and the voxel intensities [34]. A number of multi-atlas methods have proved to successfully segment the thigh muscles [23, 35–38], where the anatomy between individuals is more homogeneous. Due to the high anatomical variability of the hip among the population and its complex anatomy, only a small number of works based on multi-atlas approaches have been investigated for the hip muscles, including gluteus maximus [22, 39, 40] . The main challenge in multi-atlas segmentation is building a library with high-quality atlases that are representative of the population.

In this work, we present an automated multi-atlas segmentation method that uses an atlas library constructed from 15 high resolution Dixon MRI scans, acquired with a standardized protocol exclusively designed for this purpose. The method automatically labels gluteus maximus from an MRI scan of the pelvis, which allows the measurement of muscle volume and the generation of 3D models of the muscles. In addition, we propose a region of interest

(ROI) to standardize the measurement of muscle bulk. The segmentation quality was evaluated with a leave-one-out strategy using the dice similarity coefficient (DSC) and relative volume difference (RVD) between automated and manual segmentations, for both Dixon and T1-weighted images.

Materials and Methods

A Multi-Atlas segmentation method is proposed for a fully automated segmentation of gluteus maximus from MR images. In this approach, a library with multiple atlases is employed. Each atlas is an MR image of the pelvis with left and right gluteus maximus labelled manually by a trained operator. The method was evaluated using each atlas of the library with a leave-one-out cross validation scheme. In the following subsections we describe the acquisition of the MRI scans used in this work, the generation of the atlas library, the multi-atlas segmentation method and implementation, and the segmentation evaluation.

Study Subjects and Data Acquisition

MRI scans from a group of 15 healthy subjects, recruited for a study looking at the effects of marathon running in the hip joints, were used in this work. The demographic characteristic of the group is available in Table 1. The local Institutional Review Board approved the study and all subjects gave informed consent.

The MR images were acquired on a 3T scanner (Siemens Magneton Vida, Erlangen, Germany) using a body coil. The scanning protocol consisted of standard clinical sequences for the hips; axial Dixon (slice thickness/gap, 1.5 mm/0.45 mm; repetition time (TR) 4570 msec, echo time (TE) 45 msec; number of excitations 1, number of echoes 14; flip angle 120°) and axial T1-weighted turbo spin echo (slice thickness/gap, 3.0 mm/0.3 mm; TR/TE, 895 msec/8.9 msec) sequences of the pelvis. Dixon and T1-weighted sequences were specially designed to assess gluteal muscles and had a field of view (FOV) that covered axially from 3 cm below the lesser trochanter to the top of the iliac crest. The voxel sizes were $0.47{\times}0.47{\times}1.95$ mm^3 and $1.09{\times}1.09{\times}3.3$ mm^3.The total scanning protocol took in average 30 minutes. Figure 1 shows an example of the acquired Dixon and T1-weighted images, where the landmarks used to define the FOV are highlighted with arrows. For the Dixon image, the in-phase and out-of-phase images are shown, from which water and fat images are generated.

Atlas Library Construction

The 15 Dixon scans were used to build the atlas library for the multi-atlas segmentation. Each atlas was generated by manually segmenting right and left gluteus maximus muscles in the in-phase Dixon image. The in-phase image was used as its contrast is closer to the contrast in T1-weighted images, giving the option to use the library for the segmentation of both Dixon and T1-weighted images. Each atlas A_k in the library consists of an intensity image (Dixon in-phase) (I_{A_k}) and a labels image (L_{A_k}): $\boldsymbol{A_k} = \left\{ I_{A_k}, L_{A_k} \right\}$.

The manual segmentation was carried out in Simpleware™ ScanIP (Version 2018.12; Synopsys, Inc., Mountain View, USA), an FDA and CE marked 3D image processing software for medical scan data. To accelerate the segmentation process, every other axial slice was manually segmented using the magnetic lasso tool and then it was completed using the interpolation toolbox available in the software. A different label was used for right and left muscles. Intramuscular fat was included as part of the muscle label, while intermuscular was excluded.

Automated Multi-Atlas Segmentation of Gluteus Maximus

In multi-atlas segmentation, a selection of atlases or every atlas in the library is registered to the image to be segmented (target image). The labels of the atlases are then propagated to the target image coordinates by applying the transform obtained in the image registration process. Finally, the propagated atlas labels are combined to generate the output of the multi-atlas segmentation process (label fusion stage). Different strategies can be used in each of these steps. We described our proposed method and the implementation of each of these stages, which we also summarized in Figure 2.

This fully automated segmentation method was implemented in a plugin for Simpleware™ ScanIP (Version 2018.12; Synopsys, Inc., Mountain View, USA). It was developed in C# and

uses SimpleITK and SimpleElastix for image processing and image registration respectively. The plugin requires a full pelvis MRI scan as target image. The segmentation process consists of a pre-processing step, the actual multi-label segmentation and a final post-processing step.

Pre-Processing of the Target Image

The image pre-processing stage consists of a bias field inhomogeneity correction filter [41] followed by an Otsu multi-threshold algorithm. The former is used to correct for low frequency intensity non-uniformities present in the images, which can affect the Otsu method and the image registration in the multi-atlas segmentation. The Otsu algorithm is applied to generate a soft-tissue mask that restricts the voxels where labels can be transferred during the multi-atlas segmentation, excluding the background and fat tissue. Ideally, the Otsu algorithm would be used to segment the image into four classes: background, soft-tissue, cancellous bone and fat. However, there is a good amount of cross-talk between the classes even after correcting for bias field inhomogeneities. In order to avoid excluding any soft-tissue voxel with the soft-tissue mask, the Otsu algorithm was used with five intensity classes: background, soft-tissue, soft-tissue/bone, bone/fat and fat; and then the two classes including soft-tissue voxels were fused to generate the final mask. Finally, a morphological closing operator was used to avoid excluding intramuscular fat as we want to include it as part of the muscle labels.

Atlas Registration in Multi-Atlas Segmentation

We used an image registration scheme that consists of a rigid registration to achieve an initial rough alignment followed by a *B-spline* non-rigid image registration [42]. It was implemented using SimpleElastix [43], a Python and C# interface for the open source and validated image registration software Elastix [44]. The image registration problem is formulated as an optimization problem where a cost function C is minimized:

$$\widehat{T}_\mu = \operatorname*{argmin}_{T_\mu} C\big(T_\mu; I_F, I_M\big)$$

where T_μ is a parametrized spatial transform with parameters μ, I_F is the fixed image and I_M is the moving image. In our implementation of the multi-atlas segmentation, I_F is the target image (I_T) and the I_M is the intensity image of each atlas k (I_{A_k}).

In both rigid and non-rigid registration, the normalized correlation coefficient (NCC) was used as the similarity measure to maximize, which it is well-suited for intra-modality images and it has proved to be a reliable metric in atlas-based methods [34, 45]. The rigid registration was initialized using the geometrical centre of the images as all the images have similar FOV and are centred in the pelvis. The B-spline non-rigid registration was initialized with the output of the rigid registration and the optimization problem was solved with the adaptive stochastic gradient descent algorithm [46] with 2048 spatial samples and 1000 iterations. These hyper-parameters were empirically selected to achieve optimal segmentation performance. A pyramidal scheme of 4 layers with down-sampling factors of 8, 4, 2 and 1 was employed to improve the registration.

Label Propagation and Fusion in Multi-Atlas Segmentation

Only the labels of the 5 most similar atlases (after image registration) are propagated. This reduced number of atlases was selected due to the large heterogeneity of the population anatomy in the pelvis anatomy. The similarity values between the target and each registered atlas, obtained by computing the global NCC, were sorted in descending order before doing the atlas selection:

$$\mathbf{o} = sort\{s_1, s_2 \dots, s_n\}$$

$$s_k = NCC\big(\widehat{T}_{\mu_k}; I_T, I_{A_k}\big)$$

where s_k is the similarity value between the target image I_T and the intensity image of atlas k after applying the spatial transform $\hat{T}_{\mu_k}$. **o** is a vector with the atlas indices sorted by their similarity values in descending order.

A majority voting scheme [47, 48] was used to fuse the labels of atlases Ao_{1-5} into the segmented image S, where the label of each pixel or voxel is selected as the label that most of the selected atlases (Ao_{1-5}) agree on. The labels involved in the voting were left gluteus maximus, right gluteus maximus and background. In our implementation, the voxels where there was not a winning label were excluded. However, we did not find such cases since we are using an odd number of selected atlases and left and right gluteus maximus labels are not close enough to have voxels with votes from the three labels (left/right gluteus maximus and background).

Post-Processing

A post-processing chain was applied in order to add spatial consistency by removing poorly connected regions and filling holes. To achieve this, an opening morphological filter was applied to the output of the multi-atlas segmentation S and then the largest connected object was obtained using a three-dimensional six-connected neighbourhood. Finally, a flood fill operator was used to fill holes in the segmented mask.

Segmentation Evaluation

The performance of the muscle segmentation of gluteus maximus was evaluated with each of the 15 atlases in the library using a leave-one-out cross validation, where each atlas is removed from the library before executing the automated segmentation. The in-phase Dixon image was used as input to the automated segmentation method. As left and right gluteus maximus were labelled independently, two labels per atlas and a total of 30 muscles were evaluated. The Dice Similarity Coefficient (DSC) was used as the main segmentation performance metric:

$$DSC_l = \frac{2TP}{2TP + FP + FN}$$

where *DSC$_l$* is the dice score for label *l*; *TP*, *FP* and *FN* are the number of true positive, false positive and false negative voxels respectively for the label *l* between the segmented image (*S*) and the manually segmented image (*MS*).

In addition, the relative volume difference (*RVD*) was used as a task specific metric as we are particularly interested in measuring muscle volume:

$$RVD_l = \frac{V_{S_l} - V_{MS_l}}{V_{MS_l}} 100\%$$

where *RVD$_l$* is the relative volume difference for label l, V_{S_l} the gluteus maximus volume obtained from the automated segmentation and V_{MS_l} the muscle volume from the manually segmented images for label *l*. Mean, median, standard deviation and interquartile range (IQR) values were computed for DSC and RVD.

The same evaluation test was performed using the T1-weighted images available for every subject. The manually segmented labels from the Dixon image were used as a reference in the comparison. Because T1-weighted and Dixon images were acquired during the same scanning protocol, the images were already aligned in the same image space. However, there were some small differences in their FOV and in some cases a small misalignment was found. For this reason, the T1-weighted image were rigidly registered to the Dixon image using NCC as similarity metric before performing the segmentation evaluation. The image registration was successful in all the cases, correcting for small misalignments when needed and not introducing any change in those cases already aligned.

Muscle Bulk Evaluation

We propose a region of interest (ROI) to quantify muscle bulk with the goal of standardizing the muscle volume comparison and to focus on the main functional part of gluteus maximus. Our definition of the muscle bulk for gluteus maximus covers the region from the axial slice where the lesser trochanter tip is found to the slice that corresponds to the anterior superior iliac spine (ASIS). This definition of the muscle bulk was done to avoid regions where the uncertainty and variability in muscle volume is higher as in either top and bottom extremes of gluteus maximus, where the muscle shrinks or where tendinous tissue is found (i.e. inferior attachment to the femur).

The ROIs were defined by manually obtaining the slice indices for top and bottom landmarks (Figure 3). The top (superior) landmark is defined by the ASIS, which can be identified by scrolling slices in the coronal view until the iliac spine disappears. In Figure 3-a, the selected coronal slice is shown in blue and the axial slice where the ASIS is found is highlighted in red. The bottom (inferior) landmark is defined by the most medial point of the lesser trochanter which can be seen in the axial slice with a red border in Figure 3-a). In Figure 3-b), the defined muscle bulk for gluteus maximus is painted in red and the top and bottom slices are shown with red lines.

The DSC and RVD were computed for the muscle bulk ROI. An assessment of the landmark selection process and its impact on the volume measurement was performed by comparing the outcomes for two different independent operators.

Results

Gluteus Maximus Segmentation

Figure 4 shows boxplots of the DSC and RVD values for both Dixon and T1-weighted target images. In Figure 4-a), a boxplot of the DSC for the 30 gluteus maximus assessed (15 subjects)

is shown. The mean (±SD) DSC was 0.94 ± 0.01 (median=0.94; IQR 0.94-0.95; min=0.92, max=0.96) when using in-phase Dixon images as input, the same type of image as in the library. All of the 30 segmentations had a DSC higher than 0.9. 3D models from the manual and automated segmentation for a single case are shown in Figure 5 from posterior-anterior and anterior-posterior views. The models are also shown superimposed to illustrate their differences. The manual segmentation was labelled in red, the automated segmentation in cyan and the regions where both segmentations overlapped in yellow (true positives for the automated segmentation).

The mean (±SD) volume for the 30 assessed gluteus maximus was $7.6\times10^5 \pm 1.0\times10^5$ mm^3 in the manually segmented library, while we obtained $7.7\times10^5 \pm 9.2\times10^4$ mm^3 with the proposed automated method. The error for the population mean volume was 1.2%.

When using T1-weighted images as target images, the mean (±SD) DSC was 0.93 ± 0.02 (median=0.93; IQR 0.92-0.94; min=0.88, max=0.95). These results show that using the in-phase Dixon image is flexible enough to use the segmentation method for both Dixon and T1-weighted images, although a small penalty in performance is observed for the latter.

The RVD between the automated and manual segmentation, used to assess volume error, had a mean (±SD) value of 1.5 ± 4.3% (median=1.8%; IQR -1.2- 3.4%; min=-8.2%, max=10.6%). When using T1-weighted images, instead of the in-phase Dixon image, a mean (±SD) value of 1.5 ± 4.8% (median=1.7%; IQR -1.9-3.8%; min=-8.9%, max=10.4%) was achieved. A boxplot for the RVD for both Dixon and T1-weighted images is presented in Figure 4-b). In addition, we performed a modified Bland-Altman analysis to assess volume error for the Dixon images, where the RVD is plotted against the manual segmentation volume (Figure 6). In this figure, we can see that the RVD values distribute equally for positive and negative errors, although in

larger muscles (higher volume values) a negative bias was observed. We only found two cases out of the 95% confidence interval (1.96 SD).

In Figure 7, we show segmentation results for a case with low RVD (a-c images) and the case with highest RVD (b-d images). The true positives (intersection between manual and automated segmentation) are shown in yellow, the false negatives (manual segmentation only) in red and the false positives (automated segmentation only) in cyan. It can be seen that in overall the segmentation was accurate in both cases with a thin layer of false negatives in the edge of the muscles. In the case with high RVD, the main source of errors were false positives in the lower part (b) and in a set of slices of the medial section (c) of gluteus maximus. Both regions have high variability across the atlases.

Muscle Bulk Evaluation

When using the proposed ROI to evaluate muscle bulk, the mean (±SD) DSC was 0.95 ± 0.01 (median=0.95; IQR 0.94-0.96; min=0.92, max=0.96; 30/30 with DSC>0.9) and the RVD between the automated and manual segmentations had a mean (±SD) value of 0.6 ± 3.8% (median=0.6%; IQR -1.6-2.6%; min=-7.7%, max=9.7%). The mean (±SD) muscle bulk volume for the 30 gluteus maximus assessed was $5.9{\times}10^5 \pm 9.8{\times}10^4$ mm^3 in the manually segmented library, while we obtained $5.9{\times}10^5 \pm 8.9{\times}10^4$ mm^3 with the proposed automated method.

The muscle bulk volume depends on the selection of bottom and top slices for the ROI. Two independent operators obtained the slice numbers for the proposed landmarks and there was a mean (±SD) difference of -1.0 ± 1.5 and 0.6 ± 1.3 slices for the lesser trochanter and ASIS landmarks respectively, which translated into ROIs larger in 1.6 ± 1.8 slices for operator 1. The mean (±SD) volume difference between the two operators was 0.8 ± 1.7%.

Discussion

A multi-atlas segmentation method that provides accurate and automated segmentation of gluteus maximus from a pelvis MRI is presented in this work. The segmentation accuracy was evaluated using two metrics: DSC and RVD. A good mean DSC of 0.94 was achieved. When quantifying muscle volume, we obtained a mean (±SD) error of 1.5 ± 4.3% for the whole muscle. These results show that the method can be used to assess both shape (related to the Dice score) and volume, especially for cases where the change in volume to measure is large (i.e. patients with OA).

The measurement of the muscle volume did not reveal a particular bias and errors were distributed both negative and positively. Based on visual inspection of the images, a minor but general source of error was that the automated segmentation does not extend to the same edge as the manual segmentation. This mismatch was due to the limited accuracy of the image registration as the intensity mask did extend to the same edge (or outer) as the manual segmentation. The main source of error was the anatomical differences between the selected atlases and the target image that the image registration was not able to completely compensate for it (Figure 7-b and d shows an example). In the case of larger muscles, negative errors were obtained that could be related to the fact that there were not enough atlases in the library with larger muscles.

We found that the regions with poorer segmentation were the gluteus maximus anterior border with the gluteus medius fascia, and the inferior and superior extremes of the muscle. The segmentation in the gluteus maximus / medius border is challenging because in that region the intermuscular fat is not thick enough to guide the image registration process.

Muscle Bulk Evaluation

In order to standardize the volume measurement of gluteus maximus, we propose a ROI delimited by the lesser trochanter and the ASIS that focus on the muscle bulk of gluteus maximus. The ROI looks also to avoid differences in the image FOV across scans for either inter-subject or intra-subject (longitudinal) studies. In the muscle bulk ROI, the segmentation was marginally more accurate but when tested with two different operators, we found that their landmarks were slightly different, adding a small volume error to the volume segmentation error. A better set of instructions to find the landmarks could reduce this error. In summary, the use of the proposed ROI should be considered in the analysis of gluteus maximus as it avoids some of the regions with lower segmentation accuracy; restricts the muscle evaluation to a region delimited by two anatomical landmarks instead of the variable FOV; and concentrates the analysis in the muscle bulk, where the gluteus maximus is larger and free of tendons.

Applications

Tools that can measure volume, shape and are able to label muscles from MR images are increasingly important thanks to the wider availability of MRI scanners. Manual delineation and labelling of muscles is considered the gold-standard method for muscle segmentation however it is extremely time consuming; while surrogates metrics for volume such as cross-sectional areas [25–28], provide limited accuracy and information. This method provides good segmentation accuracy although introducing a mean error of -4.3% when compared to the gold-standard manual segmentation. In terms of time and resources, our method does not need user interaction except for loading the target image and it takes a computation time of approximately 20 minutes; while for the manual segmentation an average of 100 slices needs to be delineated

(half of them if interpolation is used) that represents approximately a 2 hours user intensive task.

Measurement of volume changes in longitudinal studies and detection of asymmetries in the context of clinical research is one of the main applications for the segmentation method presented in this work. For example, gluteus maximus wasting and atrophy is observed in patients with osteoarthritis (OA). Zacharias et al [5] assessed abductors muscles, including gluteus maximus, in patients with osteoarthritis (OA) by labelling MR images in only a small set of slices. They reported a difference in gluteus maximus volume between the affected and unaffected limb in the range of 4-15% depending on the grade of OA and greater differences in OA-control groups comparison. Similar results were observed in [6]. The volume of gluteus maximus is also of interest in sports science, where volume differences of 7% were found between groups of footballers with and without hamstring injuries, although not statically significant [8].

Tools like the one presented in this work allows the execution of equivalent studies but using fully 3D volume segmentation, where the muscle volume is not extrapolated from cross-sectional areas, adding the possibility of generating 3D models of the muscles and the estimation of shape metrics. However, given the accuracy of the presented method and the requirements of some of the applications, a visual inspection and correction stage could be necessary in some cases. In this regard, an automated atlas segmentation combined with a manual correction procedure have been previously used to measure the volume of the lower limb muscles [49].

Dixon sequence

We decided upon a Dixon sequence in our atlas library because water-fat separation techniques have proved to be an excellent tool to assess muscles. The segmentation of Dixon images

allows not only the estimation of muscle volume and shape, but also the estimation of the fat fraction in each muscle [21–24] which is associated with muscle health [50]. We are aware that these scanning sequences are not widely available or not frequently chosen due to their long acquisition time, and this would limit the number of studies where our segmentation method could be applied. To address this problem, we used the in-phase image of the Dixon sequence in the atlas library to extend the method to T1-weighted images as they have relatively similar contrast. When segmenting T1-weighted images, we observed an overall good segmentation performance with just a marginal decrease in performance compared to the Dixon in-phase images due to a less accurate image registration. We aimed to maximize the segmentation performance for the Dixon images, however a successful segmentation of T1-weighted images is still important as this sequence is widely included in clinical protocols and useful to obtain muscle volume and shape (but not quantitative analysis of fat infiltration).

Limitations

The method presented in this paper has two main limitations: the number of atlases in the library and the label fusion method. Using scans of a group of heterogeneous healthy subjects and being able to design the acquisition protocol prior to the construction of the library allowed us to build a high-quality library with homogeneous scan parameters. The group was heterogeneous as we included subjects of different gender, height and weight, although people with overweight are underrepresented since only two subjects had a BMI higher than 25. In addition, patients with larger muscles are also underrepresented as it can be seen in Figure 6 and the moderate number of atlases (15) included in the library not necessarily accounts for the wide population heterogeneity of the pelvic region. These problems can be addressed in the future by adding new atlases to the library focusing on underrepresented groups. However, only data collected with the same sequence and parameters would be suitable in order to ensure that the similarity metrics are equivalent across the atlas members.

Regarding the label fusion method, we have selected a simple majority voting approach as we prioritized speed and robustness over complexity. Statistical iterative label fusion methods such as STAPLE [51] or STEPS [52] could marginally increase accuracy as Dice scores are already high. In relation with the label fusion and propagation stage, a minor limitation of the proposed method is the simple approach used to generate the intensity-based muscle mask used to restrict the voxels where the labels from the atlases can be propagated. More sophisticated methods to classify the tissue type of each voxel could improve the segmentation in some regions of the image. In this sense, the use of machine learning have been proposed to improve tissue type classifiers for the hip and the thigh. However, their implementation is challenging as they require custom MRI sequences [53] or a very large training data set [54].

Approaches based on deep learning (i.e. convolutional neural networks (CNN)) have potential to overcome some of the limitations of multi-atlas methods. They have been very successful in the segmentation of tumours or other relatively small structures [55, 56], however its application in the musculoskeletal field has been less popular. It has shown promising results in specific applications where the area of interest is small, such as in the segmentation of the knee [57, 58]. However, in the segmentation of gluteus maximus this approach is more challenging due to the large number of manually segmented images and memory needed to train a CNN and address the variability of the hip anatomy. For this reason, we consider that currently a multi-atlas segmentation method is the best practical solution to build a gluteus maximus segmentation tool to be used in clinical research.

Conclusion

We presented a multi-atlas method for automated segmentation of gluteus maximus based on an atlas library with Dixon images, as we aimed to develop a tool for quantitative evaluation of the gluteal muscles. The method allows an accurate fully automated segmentation of gluteus

maximus for either Dixon or T1-weighted images and provides a relative accurate volume measurement in shorter times than the current accurate gold-standard manual segmentations. We also proposed a ROI to assess muscle bulk that aims to standardize inter and intra-subject comparison of gluteus maximus volume. In this region, we obtained similar accuracy to the full muscle.

Compliance with Ethical Standards

Conflict of Interest: All the authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in this work involving human participants were approved by the local Institutional Review Board. All subjects have given written informed consent.

Bibliography

1. Anderson FC, Pandy MG (2003) Individual muscle contributions to support in normal walking. Gait Posture 17:159–169.

2. Bartlett JL, Sumner B, Ellis RG, Kram R (2014) Activity and functions of the human gluteal muscles in walking, running, sprinting, and climbing. Am J Phys Anthropol 153:124–131.

3. Lieberman DE, Raichlen DA, Pontzer H, Bramble DM, Cutright-Smith E (2006) The human gluteus maximus and its role in running. J Exp Biol 209:444–454.

4. Ward SR, Winters TM, Blemker SS (2010) The Architectural Design of the Gluteal Muscle Group: Implications for Movement and Rehabilitation. J Orthop Sport Phys Ther 40:95–102.

5. Zacharias A, Pizzari T, English DJ, Kapakoulakis T, Green RA (2016) Hip abductor muscle volume in hip osteoarthritis and matched controls. Osteoarthr Cartil 24:1727–1735.

6. Grimaldi A, Richardson C, Stanton W, Durbridge G, Donnelly W, Hides J (2009) The association between degenerative hip joint pathology and size of the gluteus medius, gluteus minimus and piriformis muscles. Man Ther 14:605–610.

7. Uemura K, Takao M, Sakai T, Nishii T, Sugano N (2016) Volume Increases of the Gluteus Maximus, Gluteus Medius, and Thigh Muscles After Hip Arthroplasty. J Arthroplasty 31:906-912.e1.

8. Smith MMF, Bonacci J, Mendis MD, Christie C, Rotstein A, Hides JA (2017) Gluteus medius activation during running is a risk factor for season hamstring injuries in elite footballers. J Sci Med Sport 20:159–163.

9. Sakamaki M, G Bemben M, Abe T (2011) Legs and trunk muscle hypertrophy following walk training with restricted leg muscle blood flow. J Sports Sci Med 10:338–40.

10. Ward SR, Winters TM, Blemker SS (2010) The Architectural Design of the Gluteal Muscle Group: Implications for Movement and Rehabilitation. J Orthop Sport Phys Ther 40:95–102.

11. Mendieta CG, Sood A Classification System for Gluteal Evaluation Revisited. doi: 10.1016/j.cps.2017.12.013

12. O'Brien TD, Reeves ND, Baltzopoulos V, Jones DA, Maganaris CN (2009) Strong relationships exist between muscle volume, joint power and whole-body external mechanical power in adults and children. Exp Physiol 94:731–738.

13. Akagi R, Suzuki M, Kawaguchi E, Miyamoto N, Yamada Y, Ema R (2018) Muscle size-strength relationship including ultrasonographic echo intensity and voluntary activation level of a muscle group. Arch Gerontol Geriatr 75:185–190.

14. Dawes H, Smith C, Collett J, Wade D, Howells K, Ramsbottom R, Izadi H, Sackley C (2005) A pilot study to investigate explosive leg extensor power and walking performance after stroke. J Sports Sci Med 4:556–62.

15. Bassey EJ, Fiatarone MA, O'Neill EF, Kelly M, Evans WJ, Lipsitz LA (1992) Leg extensor power and functional performance in very old men and women. Clin Sci (Lond) 82:321–7.

16. Cronin J, Sleivert G (2005) Challenges in Understanding the Influence of Maximal Power Training on Improving Athletic Performance. Sport Med 35:213–234.

17. Bley TA, Wieben O, François CJ, Brittain JH, Reeder SB (2010) Fat and water magnetic resonance imaging. J Magn Reson Imaging 31:4–18.

18. Xiang QS, An L Water-fat imaging with direct phase encoding. J Magn Reson Imaging 7:1002–15.

19. Hernando D, Liang Z-P, Kellman P (2010) Chemical shift-based water/fat separation: A comparison of signal models. Magn Reson Med 64:811–822.

20. Dixon WT (1984) Simple proton spectroscopic imaging. Radiology 153:189–194.

21. Wokke BH, Bos C, Reijnierse M, Van Rijswijk CS, Eggers H, Webb A, Verschuuren JJ, Kan HE (2013) Comparison of dixon and T1-weighted MR methods to assess the degree of fat infiltration in duchenne muscular dystrophy patients. J Magn Reson Imaging 38:619–624.

22. Barnouin Y, Butler-Browne G, Voit T, Reversat D, Azzabou N, Leroux G, Behin A, McPhee JS, Carlier PG, Hogrel JY (2014) Manual segmentation of individual muscles of the quadriceps femoris using MRI: A reappraisal. J Magn Reson Imaging 40:239–247.

23. Karlsson A, Rosander J, Romu T, Tallberg J, Grönqvist A, Borga M, Dahlqvist Leinhard O (2015) Automatic and quantitative assessment of regional muscle volume by multi-atlas segmentation using whole-body water-fat MRI. J Magn Reson Imaging 41:1558–1569.

24. Kovanlikaya A, Mittelman SD, Ward A, Geffner ME, Dorey F, Gilsanz V (2005) Obesity and fat quantification in lean tissues using three-point Dixon MR imaging. Pediatr Radiol 35:601–607.

25. Belavý DL, Miokovic T, Rittweger J, Felsenberg D (2011) Estimation of changes in volume of individual lower-limb muscles using magnetic resonance imaging (during bed-rest). Physiol Meas 32:35–50.

26. Nordez A, Jolivet E, Südhoff I, Bonneau D, De Guise JA, Skalli W (2009) Comparison of methods to assess quadriceps muscle volume using magnetic resonance imaging. J Magn Reson Imaging 30:1116–1123.

27. Voronov A V. (2003) Anatomical Cross-Sectional Areas and Volumes of the Muscles of the Lower Extremities. Hum Physiol 29:201–211.

28. Mersmann F, Bohm S, Schroll A, Arampatzis A (2014) Validation of a simplified method for muscle volume assessment. J Biomech 47:1348–1352.

29. Pedoia V, Majumdar S, Link TM (2016) Segmentation of joint and musculoskeletal tissue in the study of arthritis. Magn Reson Mater Physics, Biol Med 29:207–221.

30. Baudin PY, Azzabou N, Carlier PG, Paragios N (2012) Automatic skeletal muscle segmentation through random walks and graph-based seed placement. Proc. - Int. Symp. Biomed. Imaging. IEEE, pp 1036–1039

31. Orgiu S, Lafortuna CL, Rastelli F, Cadioli M, Falini A, Rizzo G (2016) Automatic muscle and fat segmentation in the thigh from T 1-Weighted MRI. J Magn Reson Imaging 43:601–610.

32. Brunner G, Nambi V, Yang E, Kumar A, Virani SS, Kougias P, Shah D, Lumsden A, Ballantyne CM, Morrisett JD (2011) Automatic quantification of muscle volumes in magnetic resonance imaging scans of the lower extremities. doi: 10.1016/j.mri.2011.02.033

33. Heckemann RA, Hajnal J V., Aljabar P, Rueckert D, Hammers A (2006) Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage 33:115–126.

34. Iglesias JE, Sabuncu MR (2015) Multi-atlas segmentation of biomedical images: A

survey. Med Image Anal 24:205–219.

35. Le Troter A, Fouré A, Guye M, Confort-Gouny S, Mattei JP, Gondin J, Salort-Campana E, Bendahan D (2016) Volume measurements of individual muscles in human quadriceps femoris using atlas-based segmentation approaches. Magn Reson Mater Physics, Biol Med 29:245–257.

36. Scheys L, Loeckx D, Spaepen A, Suetens P, Jonkers I (2009) Atlas-based non-rigid image registration to automatically define line-of-action muscle models: A validation study. J Biomech 42:565–572.

37. Ghosh S, Ray N, Boulanger P (2017) A Structured Deep-Learning Based Approach for the Automated Segmentation of Human Leg Muscle from 3D MRI. 2017 14th Conf. Comput. Robot Vis. IEEE, pp 117–123

38. Ogier A, Sdika M, Foure A, Le Troter A, Bendahan D (2017) Individual muscle segmentation in MR images: A 3D propagation through 2D non-linear registration approaches. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS. IEEE, pp 317–320

39. Ranzini MBM, Ebner M, Cardoso MJ, Fotiadou A, Vercauteren T, Henckel J, Hart A, Ourselin S, Modat M (2018) Joint multimodal segmentation of clinical CT and MR from hip arthroplasty patients. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). Springer, Cham, pp 72–84

40. Klemt C, Modat M, Pichat J, Cardoso MJ, Henckel J, Hart A, Ourselin S (2015) Automatic assessment of volume asymmetries applied to hip abductor muscles in patients with hip arthroplasty. 94131M.

41. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC (2010)

N4ITK: Improved N3 bias correction. IEEE Trans Med Imaging 29:1310–1320.

42. Rueckert D, Sonoda LI, Hayes C, Hill DLG, Leach MO, Hawkes DJ (1999) Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans Med Imaging 18:712–721.

43. Marstal K, Berendsen F, Staring M, Klein S (2016) SimpleElastix: A User-Friendly, Multi-lingual Library for Medical Image Registration. 2016 IEEE Conf. Comput. Vis. Pattern Recognit. Work. IEEE, pp 574–582

44. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW (2010) elastix: A Toolbox for Intensity-Based Medical Image Registration. IEEE Trans Med Imaging. doi: 10.1109/TMI.2009.2035616

45. Avants BB, Yushkevich P, Pluta J, Minkoff D, Korczykowski M, Detre J, Gee JC (2010) The optimal template effect in hippocampus studies of diseased populations. Neuroimage 49:2457–2466.

46. Klein S, Pluim JPW, Staring M, Viergever MA (2009) Adaptive Stochastic Gradient Descent Optimisation for Image Registration. Int J Comput Vis 81:227–239.

47. Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D (2009) Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. Neuroimage 46:726–738.

48. Xu L, Krzyzak A, Suen CY (1992) Methods of combining multiple classifiers and their applications to handwriting recognition. IEEE Trans Syst Man Cybern 22:418–435.

49. Kolk S, Klawer EME, Schepers J, Weerdesteyn V, Visser EP, Verdonschot N (2015) Muscle Activity during Walking Measured Using 3D MRI Segmentations and [18F]-Fluorodeoxyglucose in Combination with Positron Emission Tomography. Med Sci

Sports Exerc 47:1896–1905.

50. Addison O, Marcus RL, Lastayo PC, Ryan AS (2014) Intermuscular fat: A review of the consequences and causes. Int J Endocrinol. doi: 10.1155/2014/309570

51. Warfield SK, Zou KH, Wells WM (2004) Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. IEEE Trans Med Imaging 23:903–921.

52. Jorge Cardoso M, Leung K, Modat M, Keihaninejad S, Cash D, Barnes J, Fox NC, Ourselin S (2013) STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcelation. Med Image Anal 17:671–684.

53. Bojorquez JZ, Bricq S, Walker PM, Lalande A (2015) Automatic classification of tissues using T1 and T2 relaxation times from prostate MRI: A step towards generation of PET/MR attenuation map. 2015 IEEE Int. Conf. Image Process. IEEE, pp 1185–1189

54. Yang YX, Chong MS, Tay L, Yew S, Yeo A, Tan CH (2016) Automated assessment of thigh composition using machine learning for Dixon magnetic resonance images. Magn Reson Mater Physics, Biol Med 29:723–731.

55. Wang G, Li W, Ourselin S, Vercauteren T (2018) Automatic Brain Tumor Segmentation Using Cascaded Anisotropic Convolutional Neural Networks. pp 178–190

56. Perone CS, Calabrese E, Cohen-Adad J (2018) Spinal cord gray matter segmentation using deep dilated convolutions. Sci Rep 8:5966.

57. Zhou Z, Zhao G, Kijowski R, Liu F (2018) Deep convolutional neural network for segmentation of knee joint anatomy. Magn Reson Med 80:2759–2770.

58. Liu F, Zhou Z, Jang H, Samsonov A, Zhao G, Kijowski R (2018) Deep convolutional

neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging. Magn Reson Med 79:2379–2391.

Tables

Demographics of the Scanned Subjects

Volunteers	N	Age	Height	Weight	BMI
Female	8	29.8 (20-43)	169.2 (155-183)	69.4 (53.0-79.0)	24.1 (22.1-27.1)
Male	7	30.3 (22-43)	177.9 (172-185)	74.6 (62.5-82.0)	23.6 (21.0-27.4)
Total	15	30.0	173.2	71.8	23.9

Table 1. Demographics of the 15 volunteers that were scanned under the MRI protocol. The age, height and BMI correspond to mean (min-max) values.

Figures
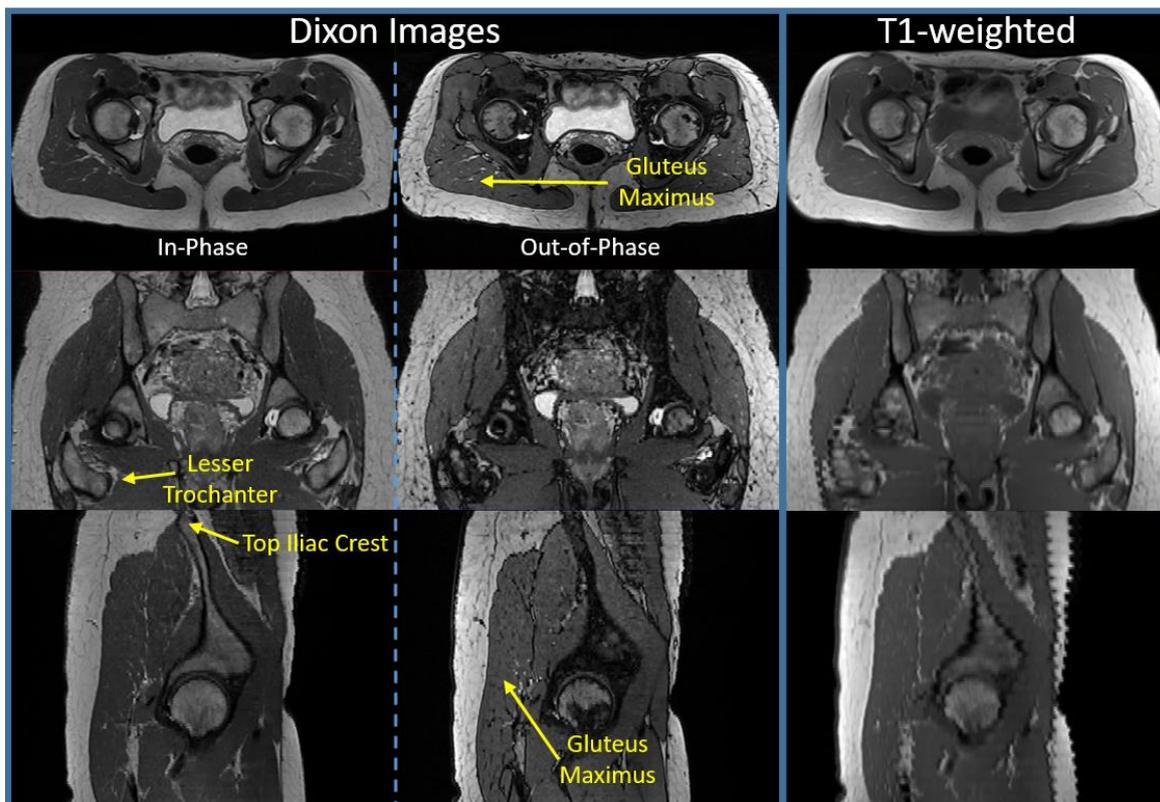
Figure 1. Dixon and T1-weighted full pelvis images acquired to build the atlas library. Axial, coronal and sagittal views are shown in the top, middle and bottom rows respectively. The top iliac crest and lesser trochanter used to define the FOV are highlighted with arrows.
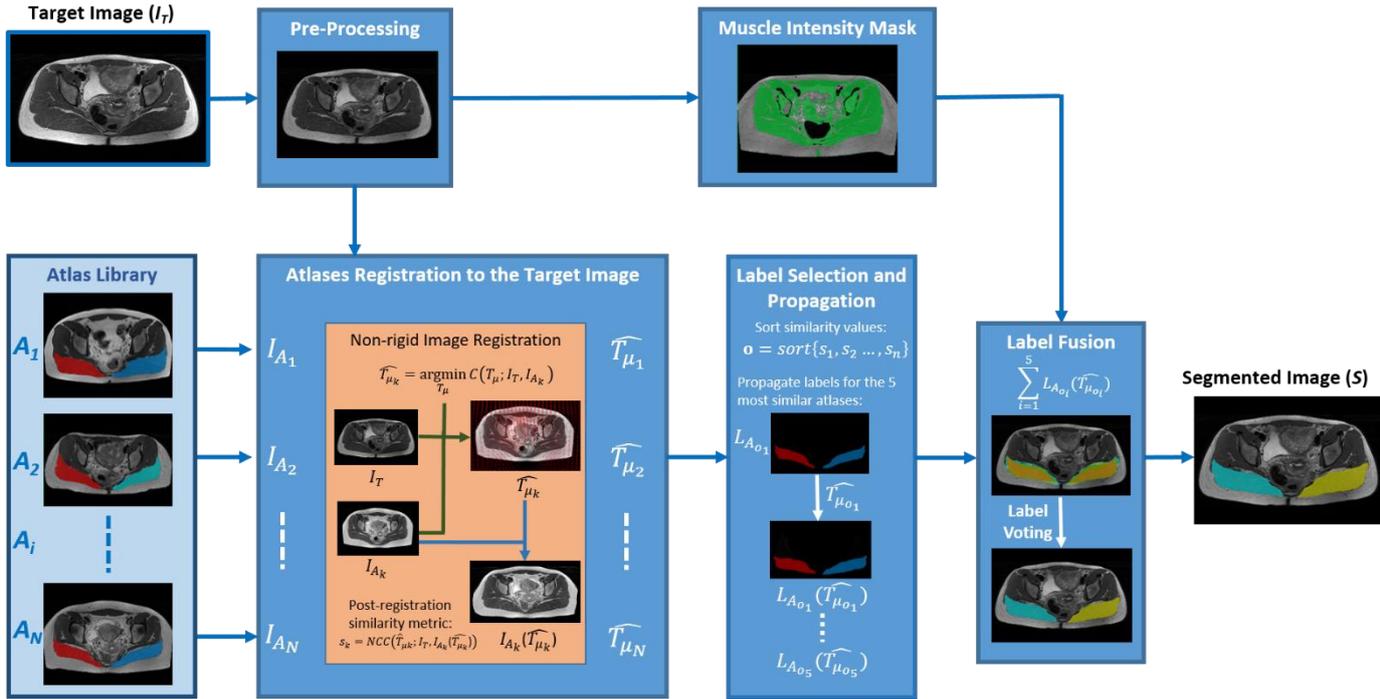
Figure 2. Schematic description of the automated segmentation. The target image is pre-processed with a bias correction filter before being used in the multi-atlas segmentation process, where each atlas in the atlas library is registered to the target image, the labels of the 5 most similar atlases are propagated to the target image space and, finally, the labels are fused using label voting and a muscle intensity mask.
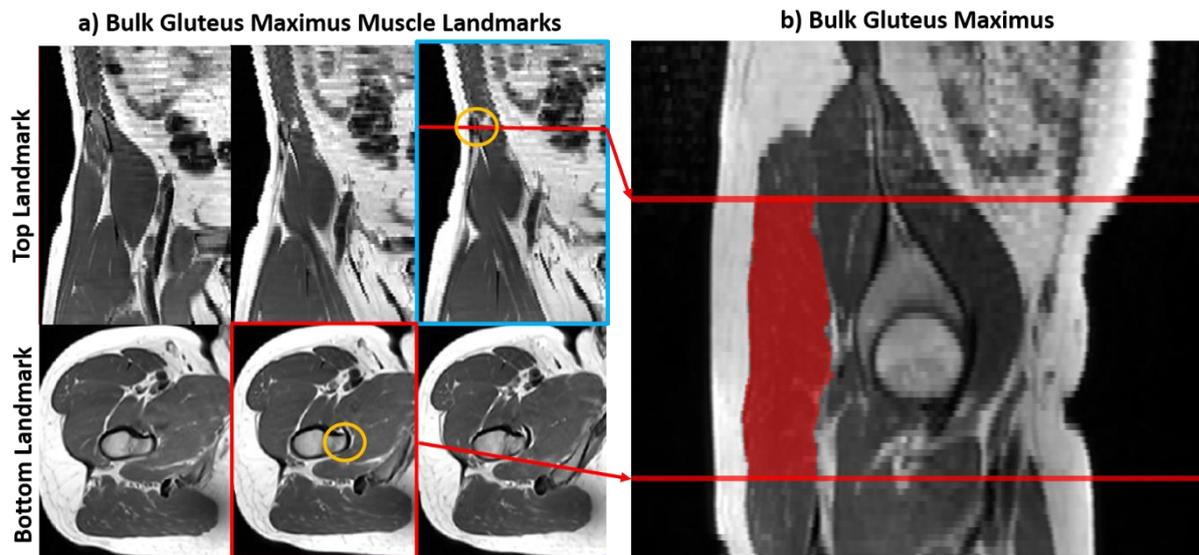
Figure 3. a) Top and bottom landmarks that define the slices delimiting the proposed region to assess the muscle bulk of gluteus maximus. The top landmark is defined by the ASIS, which can be identified in the coronal view. The bottom landmark is defined by the highest and most medial point of the lesser trochanter which can be seen in the axial slice with a red border. b) ROI used for the muscle bulk is shown in red.
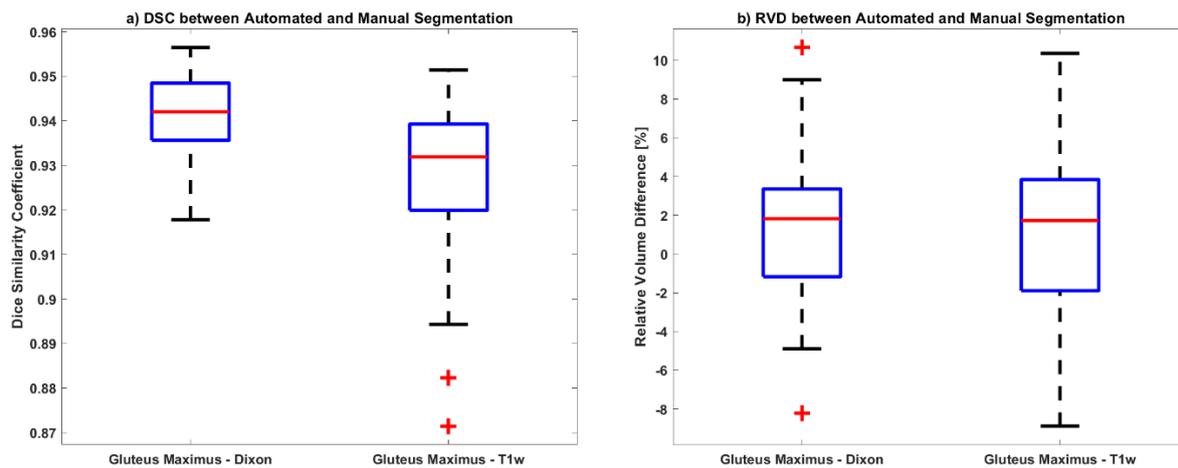
Figure 4. Boxplot of Dice similarity coefficient (a) and relative volume difference (b) between automated and manual segmentation of gluteus maximus for a sample of 30 (15 subjects).
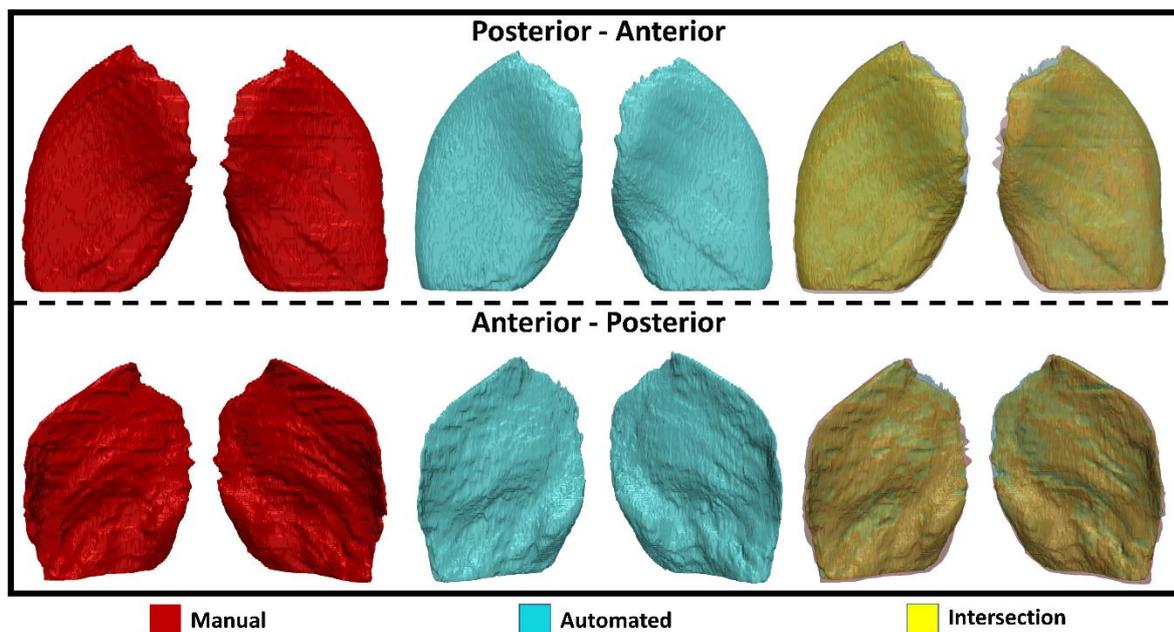
Figure 5. Posterior-anterior (top) and anterior-posterior (bottom) views of the 3D models of the segmented gluteus maximus for manual segmentation (left, red), automated segmentation (middle, cyan) and automated-manual superimposed masks (right). In the superimposed images, the regions were both segmentations matched are shown in yellow.
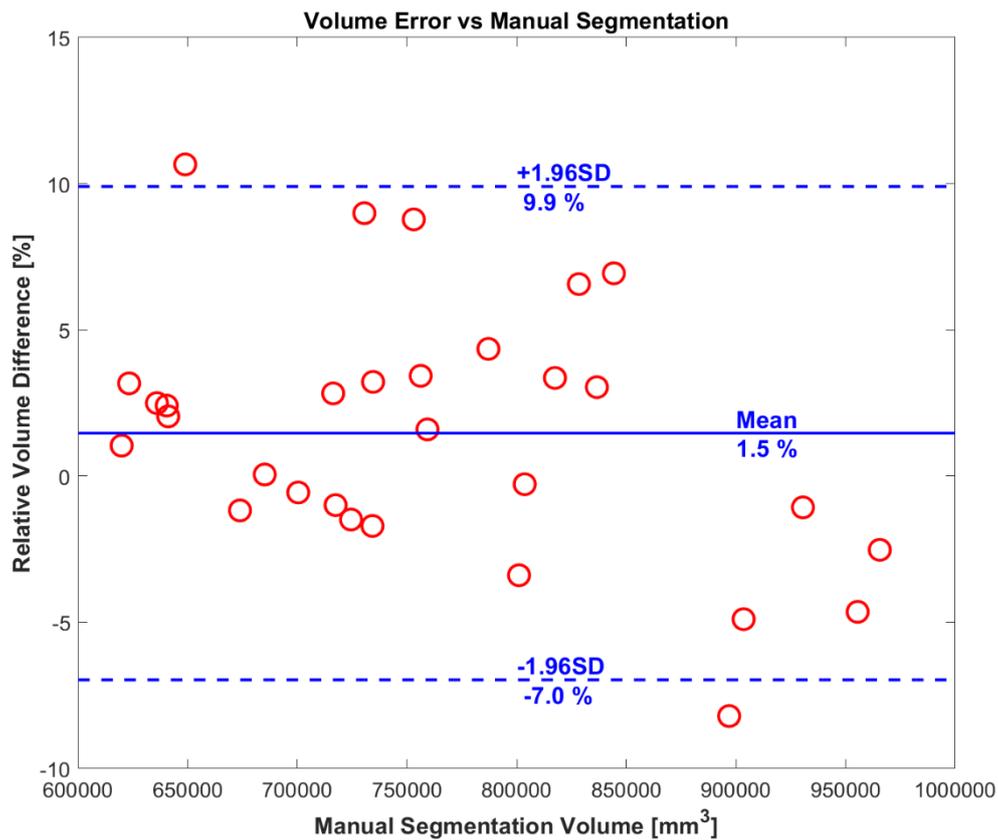
Figure 6. Relative volume difference as metric of volume error plot as a function of the gluteus maximus volume of the manually segmented images.
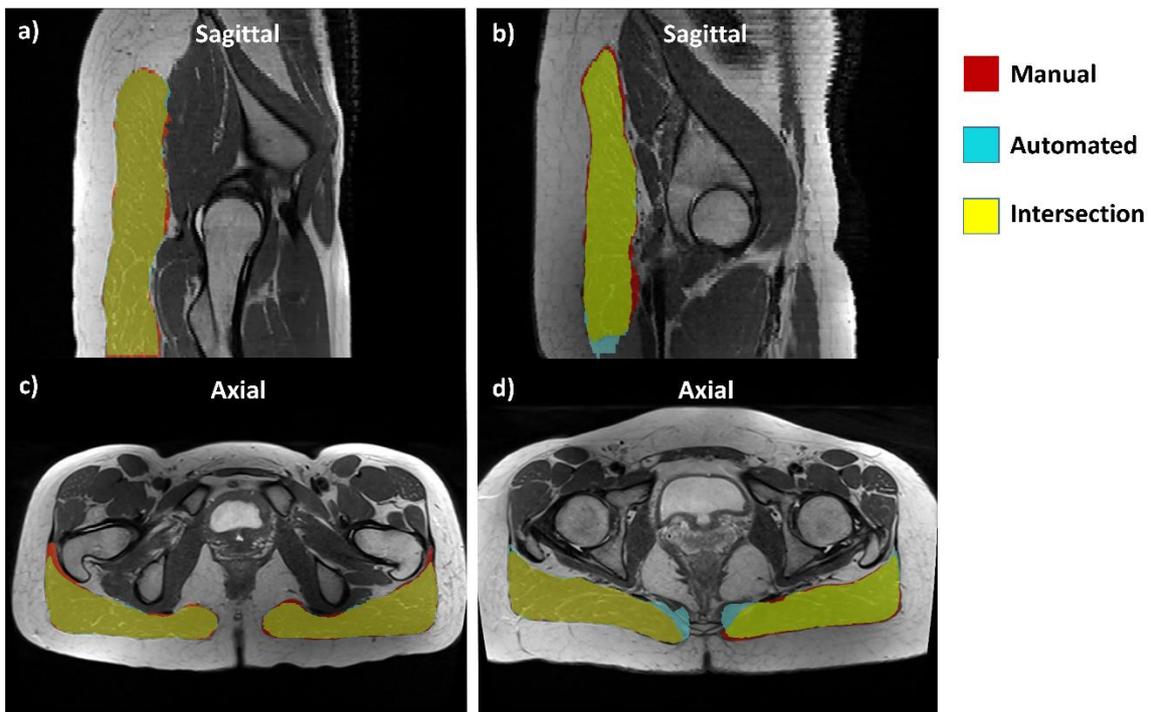
Figure 7. Set of images showing segmentation accuracy in a sagittal and axial slice for a case good case (a-c) and a. The intersection between the automated and manual segmentation is shown in yellow (true positives), in red the manual segmentation only regions (false negatives) and in cyan the automated segmentation only regions (false positives).